STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Rapport S 200 (C 8)

Syllabus van een oriënterende cursus Mathematische Statistiek

door

Prof.Dr J. Hemelrijk

1956/1957

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 trough the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

Hoofdstuk 1. Inleiding.

Gaandeweg is de wiskunde als hulpwetenschap doorgedrongen in velerlei gebieden van wetenschap en practijk, gedeeltelijk is het ontstaan ervan zelfs direct aan bepaalde practische problemen verbonden. Meetkunde, algebra, analyse, enz. worden als routine-middelen gehanteerd in de natuurkunde, sterrekunde, scheepvaart-kunde, en waar al niet meer. In de regel - en tot voor betrekkelijk kort vrijwel steeds - wordt de wiskunde gebruikt om problemen van deterministische aard op te lossen. De elementaire natuurkunde b.v. houdt zich bezig met verschijnselen, die zich onder gelijke omstandigheden steeds op dezelfde wijze voordoen. Voor toepassing van wiskundige theorieën bedient men zich dan van een wiskundig model van dat (kleine) deel van de waarneembare werkelijkheid, dat men onderzoekt. Zo is een rechte lijn b.v. de model-voorstelling van de draad van een schietlood, van een muur of van een schrikdraad; men werkt met getalwaarden voor het gewicht of de lengte van een voorwerp, de grootte van een kracht, de lading van een electron, alsof deze begrippen ondubbelzinnig bepaald zijn en een exact bepaalbare grootte bezitten.

Dergelijke modellen leiden dan tot exact gedetermineerde uitkomsten, hetgeen in de practijk neerkomt op pertinente uitspraken en voorspellingen: "Op die dag zal van zo tot zo laat een zonsverduistering optreden", "de omtrek van de aard-aequator is 40.000 km", etc. Weliswaar bezitten deze uitspraken slechts een beperkte precisie - waar men zich ook wel van bewust is - maar zolang deze voldoende is voor het doel, waarop het onderzoek gericht is, zijn de uitspraken in hun deterministische vorm bruikbaar.

Vele verschijnselen echter laten zich niet - of althans niet in voldoende mate - in een dergelijk model vangen. Deterministische methoden leiden tot gedetailleerde voorspellingen en vele zaken laten zich niet - of althans nog niet - in detail voorspellen. Voorbeelden hiervan zijn:

1. Het weer van morgen (of, in sterkere mate, van volgende week).
2. Het aantal verkeersongelukken in Amsterdam gedurende de eerstvolgende Paasweek.
3. De wachttijd van een auto bij de Hembrugpont.

4. Het jaarlijkse nationale inkomen.
5. De werkzaamheid van een geneesmiddel.
6. Het cijfer voor meetkunde van een examinandus,
enz.

Hoewel dit soort "onzekere" verschijnselen zich niet in detail laat voorspellen, tast men er toch niet geheel over in het duister. Indien men voorspelt, dat het morgen zal regenen, heeft men in Nederland vaak - zij het niet altijd - gelijk en nog vaker indien men voorspelt, dat het weer morgen "hetzelfde" zal zijn als vandaag. Indien men beschikt over de ongevallenstatistieken van vorige jaren kan men een goede slag slaan naar het aantal ongelukken in de niet te verre toekomst. Onderzoekingen omtrent de werkzaamheid van een geneesmiddel verschaffen een globale indruk daarvan; maar of het in een bepaald geval zal helpen of niet kan men niet met zekerheid voorspellen.

De statistiek houdt zich bezig met dit soort onzekere verschijnselen. Het is de wetenschap van het "hoe vaak" als de vraag "wanneer" niet zou worden beantwoord. Zo kan men b.v. bij een reeks worpen met een dobbelsteen niet voorspellen bij welke worpen een 6 zal optreden, maar wel (bij goede benadering) hoe vaak dit in een lange reeks zal gebeuren. Onvoorspelbaarheid in detail impliceert nog niet onvoorspelbaarheid in het groot. Dit is het fundamentele (experimentele) feit, waarop de statistiek gebaseerd is.

Het toepassingsgebied van de wiskunde is door het ontwikkelen van wiskundige methoden voor de behandeling van onzekere verschijnselen zeer aanzienlijk uitgebreid. Om slechts enkele gebieden te noemen waarin de statistiek een belangrijk hulpmiddel is gebleken: biologie, medicijnen, landbouw, industrie, geordsie, astronomie, verzekering, economie, sociologie, bevolkingsleer, strategie en kernphysica.

Het is duidelijk, dat men in detail onvoorspelbare verschijnselen niet in een deterministisch model kan vangen. De statistiek bedient zich dan ook vaak van modellen, waarin de onzekerheid een essentiëel element is, nl. van dat deel van de zuivere wiskunde, dat de kansrekening of waarschijnlijkheidsrekening genoemd wordt en dat een onderdeel van de maat-theorie is. Het essentiële practische verschil met de "klassieke" beschouwingswijze is, dat men

bij toepassing van de statistiek afziet van de eis, dat iedere uitspraak of voorspelling juist is - iets wat men bij "klassieke" methoden weliswaar evenmin bereikt, maar waarop deze methoden toch gericht zijn - , maar dat men tevreden is als dit in de regel, b.v. in 95% of 99% van de gevallen, waarin men statistiek toepast, het geval is. De onzekerheid, van de menselijke kennis wordt in de statistiek erkend, opgenomen en gekwantificeerd.

Verdere uitwerking van deze grondgedachte vindt plaats in latere hoofdstukken. Ter illustratie geven wij hieronder enkele voorbeelden van problemen, die met behulp van statistische methoden onderzocht kunnen worden. Bij ieder van deze problemen zal, zoals eigenlijk vanzelf spreekt, de statistische analyse van het probleem, gebaseerd moeten worden op waarnemingen, die de numerieke gegevens voor de berekeningen moeten leveren.

1. Bij het ontwikkelen van een methode voor het doen van weersvoorspellingen is statistische analyse van grote hoeveelheden waarnemingsmateriaal een belangrijk hulpmiddel.

2. Het werk van de Delta-commissie ter bepaling van de hoogte van dijken, die nodig zijn om in de toekomst de kans op overstromingen zeer gering te maken, berust voor een deel op statistische analyse van de hoogwaterstanden, die in het verleden waargenomen zijn.

3. Voor het onderzoeken van de al of niet werkzaamheid van een geneesmiddel en bij het vergelijken van twee of meer geneesmiddelen zijn experimenten nodig, waarvan de uitkomsten statistisch geanalyseerd dienen te worden.

4. Hetzelfde geldt voor talrijke technische vraagstukken, b.v. het onderzoeken van de invloed, die verschillende factoren, zoals de soort en de hoeveelheid van de brandstof en snelheid en hoeveelheid van de toegevoegde lucht, hebben op de straling van een vlam in een vlamoven.

5. Landbouwkundig onderzoek van de invloed van factoren, zoals grondbewerking, bemesting en variëteit op de opbrengst van een akker berusten tegenwoordig vrijwel steeds op statistische methoden, waarbij vooral ook aandacht wordt besteed aan een doeltreffende opzet der - gewoonlijk zeer tijdrovende - experimenten.

Hoewel nog vele andere voorbeelden genoemd zouden kunnen worden - van de walvisvangst tot het breken van garen in de textielindustrie, van de maximum snelheid van bromfietsen tot het schatten van de ouderdom van aardlagen - volstaan wij hier met de bovenstaande.

Tot slot van deze inleiding merken wij nog slechts op, dat de eerste stap van de statistische verwerking van waarnemingsmateriaal gewoonlijk bestaat uit het maken van overzichtelijke samenvattingen van dit materiaal in de vorm van tabellen en grafieken en in het globaal karakteriseren van de waarnemingen in de vorm van kengetallen (zoals gemiddelden en standaardafwijkingen), iets dat vooral bij omvangrijk waarnemingsmateriaal reeds van grote betekenis kan zijn. Bij de bevolkingsstatistiek b.v. vormt dit beschrijvende statistische werk het grootste deel van de statistische werkzaamheid. De techniek van het verzamelen van gegevens door middel van enquêtes - ter wille van een enorme tijdsbesparing - neemt daarbij de plaats in van de techniek van het opzetten van experimenten op gebieden als landbouw, biologie en techniek. Voor beide technieken zijn aparte onderdelen van de statistische theorie ontwikkeld.

Literatuur

Elementaire inleidingen

J. NEYMAN, First course in probability and statistics, H. Holt and Co, New York 1950 (geschreven voor 1e jaars studenten, die aan een statistische opleiding beginnen).

H. CRAMÉR, The elements of probability theory and some of its applications, Wiley and sons, N.Y., Almquist and Wiksell, Stockholm, 1955 (eveneens bedoeld als eerste cursus).

W.J. DIXON and F. MASSEY, An introduction to statistical analysis, Mac Graw Hill Book Co, New York 1951 (een "receptenboek" zonder theorie, waarin de gebruikte stellingen langs experimentele weg aannemelijk gemaakt worden).

Wiskundige leerboeken en handboeken

M.G. KENDALL, Advanced theory of statistics, I, II, C. Griffin and Co, London 1948.(een uiterst belangrijk handboek).

H. CRAMÉR, Mathematical methods of statistics, Princeton University Press, Princeton 1946 (één der belangrijkste boeken over mathematische statistiek).

Hoofdstuk 2. Het experimentele fundament van de statistiek

Het experimentele fundament van de statistiek is tot op zekere hoogte hetzelfde als dat van het dobbelspel: hoewel men bij een worp met een dobbelsteen niet kan voorspellen wat de uitkomst zal zijn, weet men, dat in een lange reeks worpen met een zorgvuldig gemaakte dobbelsteen alle 6 mogelijke uitkomsten ongeveer even vaak voorkomen. (Men is gewend dit feit in het kort uit te drukken door te zeggen, dat alle 6 mogelijke uitkomsten dezelfde kans bezitten.) Op grond van deze (experimenteel vastgestelde) gelijkwaardigheid der 6 mogelijke uitkomsten kan men dan langs volkomen elementaire weg van allerlei dobbelspelen de eigenschappen berekenen, b.v.: zal het vaker, even vaak of minder vaak gelukken in 4 worpen met één dobbelsteen minstens éénmaal 6 te werpen dan in 24 worpen met twee dobbelstenen dubbel 6? Dit soort van vraagstukken heeft in de 17e eeuw in Frankrijk geleid tot het ontstaan van de waarschijnlijkheidsrekening (CHEVALIER DE MÉRÉ, de speler; PASCAL en FERMAT, de wiskundigen).

Een soortgelijke regelmaat als bij geluksspelen vindt men echter ook op andere terreinen. F.W.J. SÜSSKIND (1746) en A. QUETELET (1827) hebben reeds gewezen op het merkwaardige verschijnsel, dat de frequentie per jaar per 1000 inwoners van geboorte, sterfte en huwelijk, van moord en andere misdaden en ook van andere maatschappelijke verschijnselen vrij nauwkeurig constant blijft in de tijd, zolang er althans geen veranderingen in de maatschappelijke toestanden optreden. Toch kan men doorgaans van de mensen individueel niet voorspellen of ze binnen een jaar zullen sterven, een moord zullen begaan, etc. Men noemt het beschreven verschijnsel de <u>experimentele wet der grote aantallen</u>. Een iets nauwkeurigere formulering is de volgende.

Laat E een experiment voorstellen, dat een aantal (N) malen wordt herhaald. Laat S één (willekeurige) der mogelijke uitkomsten van E voorstellen en $n(S)$ het aantal malen, dat S bij de N uitvoeringen van E, inderdaad optreedt. Dan wordt $n(S)/N$ het frequentiequotiënt[1] van S in deze reeks experimenten genoemd. Verricht men verschillende van deze reeksen van experimenten, met dezelfde E, S en N, dan zullen in de regel verschillende waarden van $f_q(S)$ gevonden worden. De experimentele wet van de grote aantallen zegt nu, dat voor een grote klasse van experimenten E geldt, dat deze verschillen tussen de voor $f_q(S)$ gevonden waarden voor grote waarden van N zeer klein worden.

---

1) afkorting: f_q, meervoud: f_{qn}; notatie $f_q(S)$.

De term "experiment" vatte men hierbij zeer ruim op. Bij de bovenstaande voorbeelden valt daar niet alleen een worp met een dobbelsteen onder, maar ook het verrichten van een waarneming, b.v. betrekking hebbende op het al of niet in leven zijn van een bepaalde persoon op een bepaalde datum.

De kansrekening en de statistiek houden zich nu bezig met al die verschijnselen, waarvoor de experimentele wet der grote aantallen geldt of geacht wordt te gelden. De kansrekening is de zuiver wiskundige kern, die los van problemen van toepassing en verificatie op axiomatische wijze wordt opgebouwd en de statistiek houdt zich bezig met het opbouwen van begrippen en theorieën, die de brug moeten slaan tussen de practische problemen en deze theorie.

Een practisch voorbeeld van de experimentele wet der grote aantallen vindt men in de onderstaande tabellen, die betrekking hebben op het geslacht van in 1954 in 6 wijken van Amsterdam geboren kinderen.

Tabel I

Aantallen jongens- en meisjesgeboorten in 6 wijken in Amsterdam (1954)

wijk nr	aantal geborenen			f_{iq} (jongens)
	jongens	meisjes	totaal (N)	
1	23	35	58	0,40
2	300	269	569	0,53
3	277	272	549	0,50
4	25	27	52	0,48
5	281	289	570	0,49
6	68	61	129	0,53

De f_{qn} der jongensgeboorten vertonen in tabel I reeds ongeveer dezelfde orde van grootte, terwijl het fq, dat het meest van de overige afwijkt, optreedt bij een kleine waarde van N. Vergroten wij de waarden van N nu door de wijken achtereenvolgens bij elkaar te nemen, dan worden de verschillen reeds vanaf de tweede regel veel geringer. Dit is in tabel II gedaan.

Tabel II

Aantallen jongens- en meisjesgeboorten in 6 wijken in Amsterdam, cumulatief

wijk nr	aantal geborenen			f_q (jongens)
	jongens	meisjes	totaal (N)	
1	23	35	58	0,40
1 en 2	323	304	627	0,52
1 t/m 3	600	576	1176	0,51
1 t/m 4	625	603	1228	0,51
1 t/m 5	906	892	1798	0,50
1 t/m 6	974	953	1927	0,51

Hierbij valt nog op te merken, dat de f_{qn} , die in de tabellen slechts in twee decimalen zijn opgegeven, wel "nauwkeuriger" berekend kunnen worden, doch dat het - juist vanwege de statistische variabiliteit - zinloos is dit te doen. Tegen deze regel, die betrekking heeft op alle verschijnselen, die statistische variabiliteit vertonen, wordt nogal eens gezondigd, waardoor dan een onverantwoorde suggestie van precisie ontstaat.

In aansluiting hierop doet zich de vraag voor, of wij het f_q, verkregen uit alle gegevens tezamen, eigenlijk niet op 0,50 af zouden moeten ronden, d.w.z. of de tweede decimaal in dit resultaat "nog wel zin heeft". Deze vage uitdrukking kan men vervangen door de iets suggestievere: "Wijst de uitkomst 0,51 erop, dat er in Amsterdam systematisch meer jongens dan meisjes geboren worden, of is de afwijking van 0,50 slechts een toevallige?" Vraagt men zich nu af, wat men met de termen "systematisch" en "toevallig" bedoelt, dan kan het antwoord hierop als volgt luiden. Indien niet alleen in het onderzochte jaar, maar ook in voorafgaande en volgende jaren bijna altijd meer jongens dan meisjes geboren worden, spreken wij van een systematisch afwijking van 0,50. In dat geval kan men dus ook voorspellen, dat er in een komend jaar meer jongens dan meisjes geboren zullen worden en die voorspelling zal dan in de regel uitkomen. Is dit niet zo, dus is er geen systematisch verschil tussen het aantal jongens- en meisjesgeboorten, dan noemt men de gevonden afwijking van 0,50 een toevallige afwijking. Een nadere precisering van deze, ook in bovenstaande terminologie nog verre van exacte, formulering, benevens de ontwikkeling van methoden, om op grond van gegevens van de in de tabellen I en II vermelde aard uit te maken in welk geval men verkeert, is nu juist de taak van de statistiek. Daarbij wordt

dan gebruik gemaakt van een exact mathematisch model, dat een vereenvoudiging en schematisering van het onderzochte probleem inhoudt en waarin zowel voor de "toevallige" als voor de "systematische" verschillen exacte begrippen worden ingevoerd. Waargenomen verschillen, zoals hier het verschil tussen 0,50 en 0,51, worden in een dergelijk model gesplitst in twee delen, nl. een systematisch en een toevallig (de statistische vakterm is "stochastisch") verschil, terwijl dan op grond van de gevonden aantallen getoetst kan worden of het systematische deel wellicht gelijk aan 0 is. Tevens worden de voorspellingsmogelijkheden uitgewerkt en gepreciseerd.

Hoofdstuk 3. Eigenschappen van frequentiequotiënten

Statistiek en kansrekening houden zich, zoals in het voorafgaande betoogd is, in de eerste plaats bezig met fqn en het is dan ook van belang enkele eenvoudige eigenschappen van fqn af te leiden, die verderop van fundamentele betekenis zullen blijken te zijn.

Wij beschouwen daartoe een eindige verzameling Λ van N elementen λ en een aantal eigenschappen (of "kenmerken") A,B,C,.... Ieder element bezit geen, één of meer van deze kenmerken. Uit gegeven kenmerken kunnen door negatie ($\bar{A}$ = niet A), conjunctie ($A \wedge B$ = A en B) en disjunctie ($A \vee B$ = A of/en B) en door herhaling van deze operaties nieuwe kenmerken gevormd worden, waarvoor uiteraard hetzelfde opnieuw geldt.

Voor een willekeurig kenmerk S geven wij het aantal elementen λ, dat S bezit, aan met N(S). Vervolgens definiëren wij het fq van S door

$$(3;1) \qquad fq(S) \overset{def}{=} \frac{N(S)}{N}$$

en het voorwaardelijke fq van S, onder de voorwaarde T, waarin T een kenmerk is met $N(T) \neq 0$, door

$$(3;2) \qquad fq(S|T) \overset{def}{=} \frac{N(S \wedge T)}{N(T)} \quad .$$

Nemen wij voor T het kenmerk "tot Λ te behoren", dan gaat fq(S|T) over in fq(S). Anderzijds kan men zeggen: de voorwaardelijke fqn onder voorwaarde T worden verkregen als gewone (onvoorwaardelijke) fqn indien men Λ vervangt door de verzameling van die elementen λ, die het kenmerk T bezitten.

De volgende eigenschappen van fqn zijn nu evident.

$$(3;3) \quad \begin{cases} \text{I} & 0 \leq fq(S) \leq 1 & \text{voor ieder der beschouwde kenmerken.} \\ \text{II} & fq(S) = 0 & \text{dan en slechts dan als geen } \lambda \in \Lambda \text{ het kenmerk S bezit.} \\ \text{III} & fq(S) = 1 & \text{dan en slechts dan als iedere } \lambda \in \Lambda \text{ het kenmerk S bezit.} \\ \text{IV} & fq(S \vee U) = fq(S) + fq(U) - fq(S \wedge T) & \text{voor ieder paar S,U van beschouwde kenmerken.} \end{cases}$$

Eigenschap IV volgt direct uit

$$N(S \vee U) = N(S) + N(U) - N(S \wedge U).$$

Deling door N geeft het gewenste resultaat.

Volgens definitie (3;2) en het daaronder vermelde gelden de eigenschappen (3;3) ook voor voorwaardelijke fqn. Dan gaat b.v. IV over in:

$$fq(S \vee U|T) = fq(S|T) + fq(U|T) - fq(S \wedge U|T).$$

Verder volgt uit (3;1) en (3;2):

$$(3;4) \qquad fq(S|T) = \frac{fq(S \wedge T)}{fq(T)} \text{ of: } fq(S \wedge T) = fq(S|T)fq(T).$$

Nu kunnen <u>op grond van (3;3) en (3;4) alleen</u>[1], <u>dus zonder terug te grijpen op de definities (3;1) en (3;2)</u>, een aantal stellingen bewezen worden, die ook weer zowel voor onvoorwaardelijke als voor voorwaardelijke fqn gelden. Bij de bewijzen van deze stellingen wordt bovendien slechts gebruik gemaakt van

$$(3;3') \quad \begin{cases} \text{II} \quad fq(S) = 0 \text{ als geen } \lambda \in \Lambda \text{ het kenmerk } S \text{ bezit} \\ \text{III} \quad fq(S) = 1 \text{ als iedere } \lambda \in \Lambda \text{ het kenmerk } S \text{ bezit} \end{cases}$$

in plaats van (3;3;II) en (3;3;III).

De te bewijzen stellingen gelden, zoals gezegd, ook voor voorwaardelijke fqn, d.w.z. alle stellingen blijven gelden, indien men aan alle fqn dezelfde voorwaarde toevoegt; dit kan alleen spaak lopen doordat er voorwaardelijke fqn ontstaan, die onbepaald zijn omdat geen enkele λ meer aan de bij het fq behorende voorwaarde voldoet.

<u>Stelling 3;1</u> (algemene optelregel).

Zijn $S_1, S_2, \ldots, S_k$ kenmerken en is

$$(3;5) \qquad \bigvee_{i=1}^{k} S_i \overset{\text{def}}{=\!=} S_1 \vee S_2 \vee \ldots \vee S_k$$

en

$$(3;6) \qquad \bigwedge_{i=1}^{k} S_i \overset{\text{def}}{=\!=} S_1 \wedge S_2 \wedge \ldots \wedge S_k,$$

dan is

$$(3;7) \quad fq\left(\bigvee_{i=1}^{k} S_i\right) = \sum_{i} fq(S_i) - \sum_{i<j} fq(S_i \wedge S_j) + \sum_{i<j<h} fq(S_i \wedge S_j \wedge S_h) - \ldots + (-1)^{k-1} fq\left(\bigwedge_{i=1}^{k} S_i\right).$$

---

1) En onder gebruikmaking van stellingen uit de logica met betrekking tot de conjunctie en disjunctie.

Bewijs door volledige inductie.

Voor k=2 gaat (3;7) over in (3;3;IV).

Voor k+1 geldt, indien (3;7) voor k juist wordt ondersteld, volgens (3;3;IV):

$$fq(\bigvee_{i=1}^{k+1} S_i) = fq(\bigvee_{i=1}^{k} S_i \vee S_{k+1}) = fq(\bigvee_{i=1}^{k} S_i) + fq(S_{k+1}) - fq(\bigvee_{i=1}^{k} S_i \wedge S_{k+1}).$$

Hierin is

$$\bigvee_{i=1}^{k} S_i \wedge S_{k+1} \equiv \bigvee_{i=1}^{k} (S_i \wedge S_{k+1}),$$

zodat, met behulp van (3;7) voor de waarde k, volgt

$$fq(\bigvee_{i=1}^{k+1} S_i) = \sum_{i=1}^{k+1} fq(S_i) - \sum_{i<j}^{k} fq(S_i \wedge S_j) + \ldots + (-1)^{k-1} fq(\bigwedge_{i=1}^{k} S_i) +$$

$$- \sum_{i=1}^{k} fq(S_i \wedge S_{k+1}) + \sum_{i<j}^{k} fq(S_i \wedge S_j \wedge S_{k+1}) - \ldots +$$

$$-(-1)^{k-1} fq(\bigwedge_{i=1}^{k} S_i \wedge S_{k+1}) = \sum_{i=1}^{k+1} fq(S_i) - \sum_{i<j}^{k+1} fq(S_i \wedge S_j) + \ldots$$

$$+(-1)^{k} fq(\bigwedge_{i=1}^{k+1} S_i).$$

q.e.d.

Stelling 3;2 (bijzondere optelregel).

Vormen de kenmerken $S_1, S_2, \ldots, S_k$ op Λ een exclusief systeem, d.w.z. dat géén $\lambda \in \Lambda$ meer dan één der kenmerken S_i bezit, dan geldt

$$(3;8) \qquad fq(\bigvee_{i=1}^{k} S_i) = \sum_{i=1}^{k} fq(S_i).$$

Bewijs

Daar $S_1, S_2, \ldots, S_k$ een exclusief systeem vormen is er, voor iedere $i_1, i_2, \ldots, i_h$ met $h \geqslant 2$, geen $\lambda \in \Lambda$ die het kenmerk

$$S_{i_1} \wedge S_{i_2} \wedge \ldots \wedge S_{i_h}$$

bezit. Uit (3;3';II) volgt dan dat voor iedere $i_1, i_2, \ldots, i_h$ met $h \geqslant 2$

$$fq(S_{i_1} \wedge S_{i_2} \wedge \ldots \wedge S_{i_h}) = 0.$$

Vult men dit in (3;7) in dan vindt men (3;8).

Stelling 3;3.

Vormen de kenmerken $S_1, S_2, \ldots, S_k$ op Λ een kategorisch systeem, d.w.z. dat iedere $\lambda \in \Lambda$ precies één der kenmerken S_i bezit, dan geldt:

(3;9) $$\sum_{i=1}^{k} fq(S_i) = 1.$$

Bewijs

Daar ieder kategorisch systeem een exclusief systeem is, geldt volgens (3;8)

$$\sum_{i=1}^{k} fq(S_i) = fq(\bigvee_{i=1}^{k} S_i).$$

Verder bezit bij een kategorisch systeem iedere $\lambda \in \Lambda$ het kenmerk $\bigvee_{i=1}^{k} S_i$. Uit (3;3;III) volgt dan

$$fq(\bigvee_{i=1}^{k} S_i) = 1.$$

Bijzonder geval

(3;10) $$fq(S) + fq(\bar{S}) = 1.$$

Stelling 3;4

Is $S_1, S_2, \ldots, S_k$ een symmetrisch kategorisch systeem op Λ, d.w.z. een kategorisch systeem met

(3;11) $$fq(S_1) = fq(S_2) = \ldots = fq(S_k)$$

dan is

(3;12) $$fq(S_i) = \frac{1}{k} \qquad (i = 1,2,\ldots,k).$$

Bewijs

Dit volgt uit stelling 3;3.

Stelling 3;5

Zijn S en U twee willekeurige kenmerken, dan geldt

(3;13) $$fq(S \wedge U) \leqslant fq(S)$$

en

(3;14) $$fq(S \vee U) \geqslant fq(S).$$

Bewijs

Uit

$$S \equiv (S \wedge U) \vee (S \wedge \bar{U})$$

en (3;8) volgt

$$fq(S) = fq(S \wedge U) + fq(S \wedge \bar{U}),$$

daar $S \wedge U$ en $S \wedge \bar{U}$ op Λ een exclusief systeem vormen. Verder is, volgens (3;3;I)

$$fq(S \wedge \bar{U}) \geqslant 0,$$

dus

$$fq(S) \geqslant fq(S \wedge U).$$

Verder is volgens (3;3;IV)

$$fq(S \vee U) = fq(S) + fq(U) - fq(S \wedge U).$$

Daar volgens (3;13): $fq(S \wedge U) \leqslant fq(U)$ is dus

$$fq(S \vee U) \geqslant fq(S).$$

<u>Stelling 3;6</u> (algemene vermenigvuldigingsregel)

Als

$$(3;15) \qquad fq(\bigwedge_{i=1}^{k-1} S_i) > 0,$$

dan geldt

$$(3;16) \quad fq(\bigwedge_{i=1}^{k} S_i) = fq(S_1)fq(S_2|S_1)fq(S_3|S_1 \wedge S_2)\ldots fq(S_k|\bigwedge_{i=1}^{k-1} S_i).$$

<u>Bewijs</u>

Uit (3;13) volgt

$$fq(S_1) \geqslant fq(S_1 \wedge S_2) \geqslant fq(S_1 \wedge S_2 \wedge S_3) \geqslant \ldots \geqslant fq(\bigwedge_{i=1}^{k-1} S_i).$$

Daar volgens (3;15) $fq(\bigwedge_{i=1}^{k-1} S_i) > 0$ bestaan dus alle voorwaardelijke fqn in (3;16).

De stelling wordt nu bewezen met volledige inductie. Voor k=2 gaat de stelling over in (3;4)

$$fq(S_1 \wedge S_2) = fq(S_1)fq(S_2|S_1) \quad \text{als } fq(S_1) > 0.$$

Onderstel nu de stelling juist voor de waarde k-1, dan volgt de stelling voor k uit

$$fq(\bigwedge_{i=1}^{k} S_i) = fq(\bigwedge_{i=1}^{k-1} S_i \wedge S_k) = fq(\bigwedge_{i=1}^{k-1} S_i)fq(S_k|\bigwedge_{i=1}^{k-1} S_i) \text{ als } fq(\bigwedge_{i=1}^{k-1} S_i) > 0,$$

daar $fq(\bigwedge_{i=1}^{k-2} S_i) \geqslant fq(\bigwedge_{i=1}^{k-1} S_i)$.

<u>Stelling 3;7</u>

Is $T_1, T_2, \ldots, T_k$ een kategorisch systeem op $\wedge$ met $fq(T_i) > 0$ voor alle i en is S een willekeurig kenmerk, dan geldt

$$(3;17) \qquad fq(S) = \sum_{i=1}^{k} fq(S|T_i)\, fq(T_i).$$

<u>Bewijs</u>

Daar $T_1, T_2, \ldots, T_k$ een kategorisch systeem op $\wedge$ vormen, vormen $S \wedge T_i$ voor $i = 1,2,\ldots,k$ een exclusief systeem op $\wedge$ en

$$S \equiv \bigvee_{i=1}^{k} (S \wedge T_i).$$

Uit (3;8) volgt dan

$$fq(S) = fq(\bigvee_{i=1}^{k}(S \wedge T_i)) = \sum_{i=1}^{k} fq(S \wedge T_i).$$

De stelling volgt dan uit

$$fq(S \wedge T_i) = fq(S|T_i)\, fq(T_i) \quad \text{als} \quad fq(T_i) > 0.$$

<u>Stelling 3;8</u> (Stelling van BAYES)

Is $T_1, T_2, \ldots, T_k$ een kategorisch systeem op Λ en is S een willekeurig kenmerk met $fq(S) \neq 0$ en $fq(T_i) \neq 0$ voor iedere i, dan geldt

$$(3;18) \qquad fq(T_i|S) = \frac{fq(S|T_i)fq(T_i)}{\sum_{i=1}^{k} fq(S|T_i)fq(T_i)} \qquad (i = 1,2,\ldots,k)$$

<u>Bewijs</u>

Volgens (3;4) is

$$fq(S \wedge T_i) = fq(S)fq(T_i|S) = fq(T_i)fq(S|T_i) \qquad (i = 1,2,\ldots,k),$$

dus

$$fq(T_i|S) = \frac{fq(S|T_i)fq(T_i)}{fq(S)} .$$

De stelling volgt dan uit (3;17).

<u>Opmerking</u>

De stellingen 3;7 en 3;8 gelden ook als $T_1, T_2, \ldots, T_k$ alleen kategorisch zijn met betrekking tot S, d.w.z. op de deelverzameling van Λ bestaande uit alle $\lambda \in \Lambda$, die S bezitten. Ook dan vormen nl. $S \wedge T_i$ voor $i = 1,2,\ldots,k$ een exclusief systeem op Λ en is

$$S \equiv \bigvee_{i=1}^{k} (S \wedge T_i).$$

Bij hoofdstuk III

Voorbeeld ter verduidelijking van de begrippen voorwaardelijke en gewone frequenties.

Gegevens betreffende 928 Nederlandse mannen:

kleur haar / kleur ogen	rood	blond	bruin	zwart	totaal
bruin	2	7	167	35	211
intermediair	1	36	164	8	209
blauw	9	184	307	8	508
totaal	12	227	638	51	928

Onvoorwaardelijke frequenties: $fq(\text{bruine ogen}) = \frac{211}{928} = 0,227$

$fq(\text{blond haar}) = \frac{227}{928} = 0,245$

Voorwaardelijke frequenties:

$fq(\text{br.o.}|\text{br.h.}) = \frac{167}{638} = 0,262$

$fq(\text{bl.o.}|\text{br.h.}) = \frac{307}{638} = 0,481$

$fq(\text{br.h.}|\text{br.o.}) = \frac{167}{211} = 0,791$

$fq(\text{bl.h.}|\text{br.o.}) = \frac{7}{211} = 0,033$

$fq(\text{bl.h.}|\text{bl.o.}) = \frac{184}{508} = 0,362$

$fq(\text{br.h.}|\text{bl.o.}) = \frac{307}{508} = 0,604$

Toepassing algemene optelregel (st.3.1)

$fq(\text{br.h. òf br.o.}) = \frac{211}{928} + \frac{638}{928} - \frac{167}{928} = \frac{2+7+167+35+164+307}{928} = 0,735$

Algemene vermenigvuldigingsregel (st 3.2.)

$fq(\text{bl.o. èn bl.h.}) = fq(\text{bl.o.})fq(\text{bl.h.}|\text{bl.o.}) = \frac{508}{928} \cdot \frac{184}{508} = \frac{184}{928} = 0,198$

Bizondere optelregel (st 3.3.)

$fq(\text{r.h. of bl.h.}) = \frac{12}{928} + \frac{227}{928} = 0,258$

$fq(\text{r.h.of bl.h.}|\text{bl.o.}) = fq(\text{r.h.}|\text{bl.o.}) + fq(\text{bl.h.}|\text{bl.o.}) =$

$\frac{9}{508} + \frac{184}{508} = 0,380$

st.3.6. Neem voor $T_1, \ldots . T_k$ de haarkleur en voor S: blauwe ogen, dan is:

$fq(\text{bl.o.}) = fq(\text{bl.o.}|\text{r.h.})fq(\text{r.h.}) + \ldots . + fq(\text{bl.o.}|\text{zw.h.})fr(\text{zw.h.})$

$= 0,547$

st.3.7. $T_1 \ldots . T_k$ haarkleur, S blauwe ogen

$$fq(\text{bl.h.}|\text{bl.o.}) = \frac{fq(\text{bl.o.}|\text{bl.h.})\ fq(\text{bl.h.})}{fq(\text{bl.o.})} = \frac{\frac{184}{227} \cdot \frac{227}{928}}{\frac{508}{928}} = 0,362$$

Hoofdstuk 4. Het kansbegrip

Het kansbegrip is oorspronkelijk ontwikkeld naar aanleiding van problemen omtrent "zuivere" kansspelen, zoals het dobbelspel (mits gespeeld met "zuivere" dobbelstenen). Als algemene beschrijving van één enkele maal spelen met een dergelijk spel kan het volgende gelden: er zijn n mogelijke "elementaire" uitkomsten $A_1, A_2, \ldots, A_n$, waarvan er precies één moet optreden en het spel is symmetrisch, d.w.z. hoewel de betekenis der A_i voor de regels van het spel voor verschillende i sterk kan verschillen, verandert het spel niet van eigenschappen, indien in de verlies- en winstregels de uitkomsten $A_1, A_2, \ldots, A_n$ aan een willekeurige permutatie onderworpen worden. Dit is dan een gevolg van het feit, dat $A_1, A_2, \ldots, A_n$ in principe alle even vaak optreden. Oorspronkelijk drukte men dit uit door te zeggen, dat $A_1, A_2, \ldots, A_n$ "gelijkelijk mogelijk" zijn.

De voor deze situatie door LAPLACE gegeven definitie van een kans is nu:

De kans op een gebeurtenis S is gelijk aan het quotiënt van het aantal (elementaire) mogelijkheden, waarbij S gerealiseerd wordt en het totale aantal mogelijkheden.[1)]

Zo is b.v. bij één worp met een ("zuivere") dobbelsteen de kans op het werpen van een even aantal ogen gelijk aan 3/6; de kans op het trekken van een schoppenkaart uit een ("goed geschud") kaartspel is 13/52 = 1/4; algemeen: de kans op ieder der elementaire uitkomsten $A_1, A_2, \ldots, A_n$ is 1/n.

Deze definitie heeft twee bezwaren. In de eerste plaats is hij kennelijk circulair zolang men geen definitie van "gelijkelijk mogelijk" geeft en de boven vermelde eis van symmetrie is zonder verdere uitwerking te vaag, om daartoe voldoende geacht te worden. Ook physische eisen van symmetrie zijn niet voldoende. Fabriceert men b.v. een zo zuiver mogelijk symmetrische munt van zeer homogeen materiaal, doch legt men deze telkens met "kruis" boven op tafel,

---

1) LAPLACE en verschillende andere schrijvers (b.v. HUYGENS) maakten een onderscheid tussen "kans" en "waarschijnlijkheid". Zij spraken van "het aantal kansen" waar wij "het aantal mogelijkheden" zeggen en gebruikten voor een samengestelde gebeurtenis S het woord "waarschijnlijkheid" in plaats van "kans". Het korte woord "kans" wint het echter langzamerhand van het lange woord "waarschijnlijkheid" en wij laten daarom dit onderscheid tussen de beide termen vallen.

dan kan men desgewenst de munt "zuiver" noemen, maar het zo uitgevoerde "kruis- of munt"-spel zeker niet. Men zal dus ook aan de wijze van werpen eisen moeten stellen, om tot een "zuiver" spel te komen en het is niet zo eenvoudig deze eisen te formuleren.

Een tweede bezwaar is de beperktheid van de definitie van LAPLACE. In lang niet alle situaties, waarop de statistiek van toepassing is, laten zich gelijkwaardige mogelijkheden aanwijzen, op grond waarvan de kans op een bepaalde gebeurtenis uitgerekend kan worden. VON MISES demonstreert dit aan het begrip "sterftekans": wat zijn daarbij de mogelijkheden, waaraan men gelijke kansen toe kan kennen? Hetzelfde doet zich trouwens reeds voor bij worpen met een "valse" dobbelsteen of munt, b.v. een punaise.

Wij zullen verderop zien, hoe deze bezwaren opgeheven kunnen worden.

Beschouwen wij voorlopig nog even de definitie van LAPLACE, dan is het verband met de stellingen over fqn, die in het vorige hoofdstuk gegeven zijn, duidelijk. Een verzameling Λ van n elementen λ_i $(i=1,2,\ldots,n)$, waarbij aan λ_i het kenmerk A_i wordt toegevoegd, is nu een wiskundig model van één uitvoering van het spel en de fqn op Λ zijn precies de boven gedefiniëerde kansen. In plaats van de abstracte elementen λ_i kunnen ook de A_i zelf als de elementen van Λ beschouwd worden.

Wij kunnen nu dan ook de in het begin van hoofdstuk 2 gestelde vraag trachten te beantwoorden met behulp van een stelling uit hoofdstuk 3. Deze vraag luidde als volgt. "Wat zal vaker gelukken: met 4 worpen van een dobbelsteen minstens éénmaal 6 te werpen of met 24 worpen met twee dobbelstenen minstens éénmaal dubbel 6?"[1] Bij de behandeling van dit vraagstuk gaan wij uit van de onderstelling, dat de dobbelstenen "zuiver" zijn, zodat het spel symmetrisch is en de kansdefinitie van LAPLACE kan worden toegepast. Geven wij de gebeurtenis "minstens éénmaal 6 bij 4 worpen" aan met S_1 en "minstens éénmaal (6,6) bij 24 worpen met twee stenen" met S_2, dan behoeven wij dus, volgens het bovenstaande, slechts de fqn van deze twee kenmerken op de bij de spelen behorende verzamelingen Λ_1, resp. Λ_2 te berekenen.

---

1) Daar er bij één dobbelsteen 6 mogelijkheden zijn en bij 2 stenen 36, terwijl 4:6 = 24:36, verwachtte de steller van deze vraag (CHEVALIER DE MÉRÉ), dat de beide spelen gelijke winstkansen zouden geven. In de practijk won hij echter met het eerste, maar verloor met het tweede. Hij concludeerde hieruit, dat de wiskunde niet deugde.

De elementaire gebeurtenissen, die de elementen van Λ_1 vormen, zijn de 6^4 mogelijke viertallen uitkomsten van 4 worpen met een dobbelsteen. Wij geven ze aan met (A_1, A_2, A_3, A_4), waarin A_i het aantal ogen bij de i^e worp voorstelt. De uitkomst $A_i = 6$ geven wij kortweg aan met b_i en de uitkomst $A_i \neq 6$ met $\bar{b}_i$. Dan is

$$S_1 \equiv b_1 \vee b_2 \vee b_3 \vee b_4,$$

en

$$\bar{S}_1 \equiv \bar{b}_1 \wedge \bar{b}_2 \wedge \bar{b}_3 \wedge \bar{b}_4.$$

Volgens (3;10) is

$$fq(S_1) = 1 - fq(\bar{S}_1),$$

zodat het voldoende is $fq(\bar{S}_1)$ te berekenen. Volgens (3;16) is:

$$fq(\bar{S}_1) = fq(\bar{b}_1)\, fq(\bar{b}_2|\bar{b}_1)\, fq(\bar{b}_3|\bar{b}_1 \wedge \bar{b}_2)\, f(\bar{b}_4|\bar{b}_1 \wedge \bar{b}_2 \wedge \bar{b}_3).$$

Nu is gemakkelijk na te gaan, dat de fqn van het rechterlid op Λ_1 alle gelijk aan 5/6 zijn, zodat

$$fq(\bar{S}_1) = (5/6)^4, \text{ dus } \quad fq(S_1) = 1-(5/6)^4 = 0{,}518$$

is. Geheel analoog kan men $fq(S_2)$ berekenen op Λ_2, een verzameling bestaande uit de 36^{24} mogelijke 24-tallen van uitkomsten van het tweede spel. De uitkomst is

$$fq(S_2) = 1-(35/36)^{24} = 0{,}491 .$$

Dit zijn dus tevens de kansen volgens LAPLACE en indien men ervan uitgaat, dat bij vele malen herhalen van de spelen voor ieder daarvan alle elementaire gebeurtenissen (ongeveer) even vaak zullen optreden, dan geven deze kansen aan, hoe vaak S_1 resp. S_2 ongeveer zullen optreden; S_1 treedt dan dus vaker dan in de helft der gevallen op en S_2 minder vaak.

Men kan nu verschillende wegen volgen, om deze opzet der kansrekening tot een goed gefundeerde theorie uit te breiden. Bij één daarvan, waarvan wij slechts heuristisch de gedachtengang zullen schetsen, wordt expliciet vastgehouden aan het uitgangspunt van gelijkwaardigheid van elementaire gebeurtenissen. Een tweede, die ondanks het feit, dat daarbij niet van gelijkwaardigheid uitgegaan wordt, in feite aequivalent is met deze eerste, is een axiomatische opzet van de theorie en deze methode is tegenwoordig de meest gebruikelijke, mede vanwege de eenvoudige wijze, waarop langs deze weg exactheid kan worden bereikt.

Bij beide voorstellingswijzen is het uitgangspunt van alle toepassingen, dat symmetrie of gelijke kansen betekenen, dat de betrokken gebeurtenissen bij veelvuldig herhaalde uitvoering van het experiment (of bij herhaalde waarneming) ongeveer even vaak optreden. Beide leiden ook tot dezelfde interpretatie van een kans: een kans p op een gebeurtenis S betekent dat het fq van S in een lange reeks waarnemingen ongeveer gelijk aan p zal zijn. Zij steunen derhalve direct op de experimentele wet der grote aantallen (hoofdstuk 2) en stemmen hierin overeen met de definitie van LAPLACE.

De eerste gedachtengang zullen wij slechts aan een voorbeeld toelichten en wel aan dat van het geslacht van pas geborenen. Dit is één van de in hoofdstuk 2 vermelde voorbeelden van de experimentele wet der grote aantallen, waarbij het echter op het eerste gezicht niet duidelijk is, wat de gelijkwaardige mogelijkheden zijn. Immers er worden - zoals uit vele statistieken blijkt - meer jongens dan meisjes geboren, zodat de twee mogelijke uitkomsten "jongen" en "meisje" (ook) in dit opzicht niet als gelijkwaardig beschouwd kunnen worden. De theorie zal moeten toelaten, dat de kans op een jongen groter is dan die op een meisje. Het is echter wèl mogelijk Λ samen te stellen uit de geboorten als elementen (vroegere en toekomstige inbegrepen) en deze elementen als gelijkwaardig te beschouwen. Het fq van het kenmerk "jongen" resp. "meisje" op deze Λ is dan de kans op een jongensgeboorte. Deze verzameling Λ kan dan als basis voor de verdere ontwikkeling van op dit probleem toe te passen statistische methoden gebruikt worden en in feite kan men zodoende de gehele statistiek opbouwen zonder het kansbegrip in te voeren, daar dit nu slechts een ander woord voor fq is. De statistiek wordt dan volledig <u>frequentie-quotiëntenrekening</u>.

Bij de axiomatische opzet van de kansrekening volgt men een andere weg, waarbij men juist de gelijkwaardigheid van mogelijkheden - die de principiële moeilijkheid vormt bij de definitie van LAPLACE - laat vallen. Overigens is de opbouw vrijwel analoog aan die van een regelrechte fqn-rekening, op één belangrijk verschil na. Indien men nl. wil werken met waargenomen fqn, moet men rekening houden met hun variabiliteit, waardoor telkens in alle formuleringen het woord "ongeveer" insluipt. De relatie "is ongeveer gelijk aan" (notatie: $\approx$) is echter niet transitief ($10^6 \approx 10^6 - 1 \approx \ldots \approx 1$), hetgeen het rekenen zeer bemoeilijkt.

Anderzijds hebben fqn in lange reeksen waarnemingen een neiging tot stabiliteit (experimentele wet der grote aantallen) en de kansrekening verdisconteert dit feit - daarbij tegelijkertijd bovengenoemde moeilijkheid oplossend - door aan de gebeurtenis, waarvan het fq beschouwd wordt, een (in de regel onbekend, maar constant) getal toe te voegen, dat de kans op (of waarschijnlijkheid van) die gebeurtenis heet. Deze kans is een wiskundig analogon van het in een lange reeks waarnemingen optredende fq, waarbij - in eerste instantie - van de onzekerheid van het fq afgezien is. Later blijkt, dat in het kansmodel het fq als zodanig, met zijn statistische fluctuaties, op natuurlijke wijze weer ingevoerd kan worden. Het kansbegrip wordt vastgelegd door axioma's, niet door een definitie.

Deze gang van zaken, die verderop uitvoeriger wordt beschreven, is vrijwel analoog aan die bij de opbouw en toepassing van de meetkunde. Ook daar wordt eerst afgezien van verschillen, die in de practijk steeds bestaan. Immers twee voorwerpen bezitten nooit precies gelijke afmetingen, het is zelfs nauwelijks mogelijk een definitie van "precies gelijk" te geven, die niet reeds een modelkarakter bezit. Lijnstukken van gelijke lengte en congruente lichamen zijn echter belangrijke begrippen voor de opbouw en de toepassing van de meetkunde. Ook de begrippen "punt", "lijn" en "vlak" zijn wiskundige modelbegrippen, die niet aan definities, doch slechts aan axioma's onderworpen worden en die aan de werkelijkheid ontleend worden door schematisering. De toepassing van de meetkunde maakt dan tenslotte de waarneming en beheersing van verschillen, waarvan eerst geabstraheerd is, beter mogelijk dan oorspronkelijk het geval was. Een soortgelijke kringloop valt waar te nemen bij de statistiek.

Bij het begrip kans heeft men dus in gedachte de "constante kern" van het fq te karakteriseren. Het ligt daarom voor de hand voor de axioma's een aantal eenvoudige eigenschappen van fqn te nemen. Daarvoor blijken nu de eigenschappen (3;3) uitermate geschikt te zijn.

Aan deze eigenschappen lag een verzameling Λ ten grondslag, waarop kenmerken gedefinieerd waren. Een dergelijke verzameling hebben wij nu ook nodig en deze verzameling, die de basis van het wiskundige model vormt, is voor elk practisch probleem verschillend. Een practisch probleem, waarop de statistiek van toepassing is, heeft steeds betrekking op een situatie, die verschillende uitkomsten kan hebben. Voorbeelden daarvan hebben wij reeds herhaaldelijk genoemd. Deze verschillende mogelijke uitkomsten zijn nu

de elementen van de basisverzameling, die wij met Γ aan zullen geven. Zij behoeven nu niet meer "gelijkelijk mogelijk" te zijn. Op Γ zijn nu weer kenmerken gedefiniëerd, waaruit door negatie, conjunctie en disjunctie nieuwe kenmerken afgeleid kunnen worden. Voorbeeld: bij één worp met een dobbelsteen ("zuiver" of "vals", dat doet nu niet ter zake) bestaat Γ uit 6 elementen, de 6 zijden van de dobbelsteen. Als kenmerken treden op: de aantallen ogen 1,2,...,6; even, oneven, 5, etc. Deze kenmerken worden ook "<u>eventualiteiten</u>" genoemd, een term (ingevoerd door Prof.Dr D. VAN DANTZIG) die nauw aansluit bij hun karakter: eventueel optredende gebeurtenissen.

Is Γ nu eindig, dan kunnen wij de eigenschappen (3;3) letterlijk overnemen als axioma's voor de kansrekening. Vaak heeft men echter met oneindige Γ te maken. Een eenvoudig voorbeeld is het werpen met een munt tot voor het eerst kruis verkregen wordt. Het aantal worpen, dat bij dit experiment optreedt, kan dan alle waarden 1,2,... ad infinitum aannemen en Γ moet dus oneindig veel elementen bevatten. Dit maakt een kleine wijziging in de eigenschappen II en III wenselijk, terwijl een 5e axioma toegevoegd wordt, om het werken met oneindig veel mogelijkheden vlot te doen verlopen.

De kans op een eventualiteit S wordt aangegeven met $P[S]$. Aan alle op Γ gedefiniëerde eventualiteiten wordt nu een constant getal als kans toegevoegd en deze kansen (die in de regel van onbekende grootte zijn) voldoen aan de volgende <u>axioma's</u>.

(4;1)

Axioma I. $0 \leqq P[S] \leqq 1$ voor iedere beschouwde eventualiteit S.

Axioma II. $P[S]=0$ als geen element van Γ het kenmerk S bezit.

Axioma III. $P[S]=1$ als ieder element van Γ het kenmerk S bezit.

Axioma IV. $P[S \vee U] = P[S] + P[U] - P[S \wedge U]$ voor ieder paar S,U van beschouwde eventualiteiten.

<u>Opmerking</u>

De axioma's II en III worden ook vaak als volgt uitgedrukt: $P[S]=0$ als S onmogelijk is; $P[S]=1$ als S zeker is.

De op grond van de eigenschappen (3;3) voor onvoorwaardelijke fqn afgeleide stellingen gelden nu dus ook voor kansen, daar bij het bewijs van de stellingen in hoofdstuk 3 geen gebruik werd gemaakt van (3;3;II) en (3;3;III), maar slechts van (3;3';II) en (3;3';III), die overeenkomen met (4;1;II) en (4;1;III).

Eén van de stellingen (stelling 3;2) luidt:

Vormen de eventualiteiten $S_1, S_2, \ldots, S_k$ op Γ een exclusief systeem, dan geldt

$$(4;2) \qquad P[\bigvee_{i=1}^{k} S_i] = \sum_{i=1}^{k} P[S_i] \ .$$

Deze eigenschap geldt voor iedere eindige k. Als 5e axioma wordt nu de overeenkomstige eigenschap voor oneindige exclusieve systemen genomen:

Axioma V:

Vormen $S_1, S_2, \ldots$ een oneindig exclusief systeem op Γ, dan geldt:

$$(4;3) \qquad P[\bigvee_{i=1}^{\infty} S_i] = \sum_{i=1}^{\infty} P[S_i] \ .$$

Het axiomasysteem is nu volledig. Een verzameling Γ van eventualiteiten met bijbehorende kansen, die aan de axioma's voldoen, wordt een kansveld of waarschijnlijkheidsveld genoemd.

Wij gaan nu over tot de invoering van voorwaardelijke kansen. De definitie daarvan is analoog aan (3;4): de voorwaardelijke kans, $P[S|T]$, op de eventualiteit S onder voorwaarde van het optreden van de eventualiteit T, die een kans $P[T] > 0$ bezit, wordt gedefinieerd door

$$(4;4) \qquad P[S|T] \overset{\text{def}}{=} \frac{P[S \wedge T]}{P[T]} \ .$$

Het is nu niet zonder meer duidelijk, dat de axioma's ook gelden voor voorwaardelijke kansen. Bij fqn volgde dit direct uit definitie (3;2), maar nu moeten wij deze eigenschappen bewijzen. Is dit gebeurd, dan gelden ook alle in hoofdstuk 3 afgeleide stellingen voor voorwaardelijke kansen.

Stelling 4;1

Indien $P[T] > 0$ is, geldt:

(4;5)
$$\begin{cases} \text{I} & 0 \leqslant P[S|T] \leqslant 1 \quad \text{voor iedere eventualiteit S.} \\ \text{II} & P[S|T] = 0 \quad \text{als geen der elementen van } \Gamma \text{, die het kenmerk T bezitten, ook S bezit.} \\ \text{III} & P[S|T] = 1 \quad \text{als ieder element, dat het kenmerk T bezit ook S bezit.} \\ \text{IV} & P[S \vee U|T] = P[S|T] + P[U|T] - P[S \wedge U|T] \quad \text{voor ieder paar kenmerken S en U.} \\ \text{V} & P\left[\bigvee_{i=1}^{\infty} S_i|T\right] = \sum_{i=1}^{\infty} P[S_i|T] \text{ als } S_1, S_2, \ldots \text{ een oneindig exclusief systeem op } \Gamma \text{ vormen.} \end{cases}$$

Volgens stelling 3;5 geldt

(4;6) $$P[T \wedge S] \leqslant P[T] .$$

Dus

$$P[S|T] = \frac{P[T \wedge S]}{P[T]} \leqslant 1.$$

Teller en noemer zijn verder beide niet negatief.

II. Als geen der elementen van Γ zowel T als S bezit, is volgens axioma II: $P[T \wedge S] = 0$.

III. Als ieder element, dat T bezit, ook S bezit, is volgens axioma II: $P[T \wedge \bar{S}] = 0$. Uit (4;6) volgt dan: $P[T] = P[T \wedge S]$.

IV.
$$P[(S \vee U) \wedge T] = P[(S \wedge T) \vee (U \wedge T)] =$$
$$= P[S \wedge T] + P[U \wedge T] - P[S \wedge U \wedge T].$$
$$\therefore P[S \vee U|T] = \frac{P[(S \vee U) \wedge T]}{P[T]} =$$
$$= \frac{P[S \wedge T]}{P[T]} + \frac{P[U \wedge T]}{P[T]} - \frac{P[S \wedge U \wedge T]}{P[T]} =$$
$$= P[S|T] + P[U|T] - P[S \wedge U|T].$$
q.e.d.

V. Als $S_1, S_2, \ldots$ een oneindig exclusief systeem op Γ vormen dan is

$$P\left[\bigvee_{i=1}^{\infty} S_i|T\right] = \frac{P\left[\bigvee_{i=1}^{\infty} S_i \wedge T\right]}{P[T]} = \frac{P\left[\bigvee_{i=1}^{\infty} (S_i \wedge T)\right]}{P[T]} = \frac{\sum_{i=1}^{\infty} P[S_i \wedge T]}{P[T]} = \sum_{i=1}^{\infty} P[S_i|T] .$$

Tenslotte voeren wij nog het begrip <u>stochastische onafhankelijkheid</u> in. De eventualiteit S is per definitie stochastisch onafhankelijk van de eventualiteit T, als

(4;7) $$P[S|T] = P[S] .$$

Volgens definitie (4;4) is dan

(4;8) $$P[S \wedge T] = P[S].P[T],$$

zodat dan ook geldt:

(4;9) $$P[T|S] = P[T] .$$

De relatie van stochastische onafhankelijkheid is dus wederkerig. Tevens geldt, als S en T stochastisch onafhankelijk zijn:

(4;10) $$\begin{cases} P[S|T] = P[S|\bar{T}] = P[S], \\ P[\bar{S}|T] = P[\bar{S}|\bar{T}] = P[\bar{S}] \end{cases}$$

en analoog met verwisseling van S en T. Het bewijs van deze eigenschappen en van de hieronder volgende laten wij aan de lezer over.

(4;11) als $S \rightarrow T$ (d.w.z. als S T impliceert), dan is $P[S] \leq P[T]$ en $P[S \wedge T] = P[S]$, $P[S \vee T] = P[T]$,

(4;12) is $P[S] = 1$, dan is iedere T stochastisch onafhankelijk van S.

Stochastische onafhankelijkheid voor meer dan twee eventualiteiten wordt als volgt gedefinieerd. De eventualiteiten S_1, $S_2,\ldots,S_k$ (k mag ∞ zijn) zijn stochastisch onafhankelijk, indien voor ieder h-tal ($h \leq k$) $S_i, S_{i_2},\ldots,S_{i_h}$ daaruit geldt:

(4;13) $$P[\bigwedge_{j=1}^{h} S_{i_j}] = \prod_{j=1}^{h} P[S_{i_j}] .$$

Dit impliceert o.a. dat ieder tweetal S_a, S_b stochastisch onafhankelijk is en ook, dat

(4;14) $$P[\bigwedge_{i=1}^{k} S_i] = \prod_{i=1}^{k} P[S_i]$$ (<u>bijzondere productregel</u>; voor onafhankelijke eventualiteiten)

Paarsgewijze onafhankelijkheid en het vervuld zijn van (4;14) zijn echter niet voldoende om (4;13) te garanderen. Een tegenvoorbeeld voor het tweede is als volgt te construeren. Laat Γ uit 30 elementen bestaan, aangegeven door de nummers 1,...,30 en laat deze alle een kans 1/30 bezitten. Beschouw nu de volgende kenmerken:

A: $1 \vee 2 \vee \ldots \vee 10$,

B: $1 \vee 2 \vee \ldots \vee 20$,

C: $1 \vee 2 \vee 11 \vee 12 \vee 21 \vee 22 \vee 23 \vee 24 \vee 25$.

A en B zijn dan afhankelijk, want A impliceert B .
Toch is

$$P[A].P[B].P[C] = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{9}{30} = \frac{1}{15}$$

en

$$P[A \wedge B \wedge C] = \frac{2}{30} = \frac{1}{15} .$$

Een voorbeeld van een drietal paarsgewijze onafhankelijke eventualiteiten, waarvoor echter de productregel niet opgaat, is een Γ met 12 elementen $(1,\ldots,12)$, die ieder een kans 1/12 bezitten en

$$A: \quad 1 \vee 2 \vee 3 \vee 4,$$

$$B: \quad (1 \vee 2) \vee 5 \vee 6 \vee 7 \vee 8 ,$$

$$C: \quad (1 \vee 2) \vee 5 \vee 9 \vee 10 \vee 11 .$$

De verificatie van de juistheid van deze bewering laten wij aan de lezer over.

Het is van belang na te gaan, wat de practische interpretatie van het onafhankelijkheidsbegrip is, d.w.z. wanneer wij twee of meer mogelijk optredende gebeurtenissen S en T in het model zullen voorstellen door onafhankelijke eventualiteiten. Volgens (4;10) houdt onafhankelijkheid van S en T in, dat de kans op S niet beïnvloed wordt door het al of niet optreden van T . Of anders gezegd: dat de informatie of T wel of niet optreedt ons geen informatie verschaft over het al of niet optreden van S . Dit is echter reeds min of meer een practische interpretatie en indien het practisch onaannemelijk geacht wordt, dat het al of niet optreden van T dat van S beïnvloedt, kan men deze twee dus door onafhankelijke eventualiteiten voorstellen. Men kan zich daarbij uiteraard vergissen, doch dat is bij modelkeuze altijd het geval. Verkeert men in twijfel, dan late men in het model een eventuele afhankelijkheid toe. Deze kan dan, zoals wij later zullen merken, statistisch onderzocht worden.

Voorbeelden

Op elkaar volgende worpen met een dobbelsteen, met de nodige voorzorgen uitgevoerd, worden gewoonlijk als onafhankelijk beschouwd, daar men aanneemt dat een dobbelsteen "geen geheugen heeft". Door verschillende waarnemers uitgevoerde wegingen van een zelfde object, zonder dat de waarnemers van elkanders uitkomsten op de hoogte zijn, zijn onafhankelijk (zouden zij elkanders uitkomsten kennen, dan is beïnvloeding, dus afhankelijkheid, van psychologische aard niet uitgesloten). Anderzijds betekent

afhankelijkheid van twee eventualiteiten nog niet, dat er een direct oorzakelijk verband bestaat; er kan ook een derde verschijnsel zijn, dat de beide eerste beïnvloedt. Beschouwt men b.v. voor ieder der na-oorlogse jaren het aantal verkochte auto's en het aantal schoolkinderen, dan vertonen deze beide kenmerken een duidelijke afhankelijkheid: in de jaren met veel schoolkinderen waren er ook veel auto's. Toch schaft men gewoonlijk geen auto aan omdat men zijn kinderen naar school moet brengen; ook stuurt men niet meer kinderen naar school omdat men een auto heeft. Het is de parallelle stijging met de tijd, die hier het statistische verband veroorzaakt. Over een jaar of wat kan dit er trouwens weer heel anders uitzien. Men zij dus voorzichtig met de interpretatie van afhankelijkheden.

De definitie van LAPLACE komt nu als speciaal geval te voorschijn. Als model voor een symmetrisch spel (b.v. één worp met een zuivere dobbelsteen) wordt nl. uiteraard een Γ genomen, waarvan alle elementen (de elementaire mogelijke uitkomsten van het spel) gelijke kans bezitten. Voor n elementen dus ieder kans $1/n$. Uit stelling 3;2 (de bijzondere optelregel) volgt dan direct, dat de kans op een willekeurige eventualiteit in zo'n geval berekend kan worden volgens de definitie van LAPLACE.

Opmerkingen

De hier beschreven practische interpretatie van het kansbegrip, als een model-analogon van een fq, wordt de frequentie-interpretatie genoemd. Naast deze interpretatie zijn nog verschillende andere ontwikkeld, waaronder ook van subjectivistische aard: een kans wordt dan opgevat als een "graad van redelijk geloof" in het optreden van een eventualiteit. Wij houden ons in deze cursus geheel aan de frequentie-interpretatie, die wegens zijn objectief en verifiëerbaar karakter een hechtere basis voor de toepassingen geeft dan de andere interpretaties.

Literatuur

De axiomatiek van de kansrekening is afkomstig van
A. KOLMOGOROV, Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlijn 1933; Chelsea Publishing Co, New York 1946.

Hoofdstuk 5. Voorbeelden van het berekenen van kansen

Ter illustratie van het gebruik van de axioma's geven wij in dit hoofdstuk een aantal elementaire voorbeelden van berekeningen van kansen op een gegeven kansveld.

Voorbeeld 1

Wij beschouwen twee onafhankelijke uitvoeringen van een experiment, waarbij een "succes" (S) of een "mislukking" (M) op kan treden. De kans op succes is beide keren gelijk aan p $(0 < p < 1)$. Gevraagd wordt de voorwaardelijke kansen $P[SM \mid SM \vee MS]$ en $P[MS \mid SM \vee MS]$ te berekenen.

Oplossing

Wij hebben: $P[S] = p$, $P[S \vee M] = 1$ (ax. III) dus volgens (3;11): $P[M] = 1-p$. Daar de beide uitvoeringen van het experiment onafhankelijk zijn, is (4;8) van toepassing, dus is, met $q = 1-p$,

$$P[SS] = p^2 \,,\; P[SM] = P[MS] = pq \,,\; P[MM] = q^2 .$$

Verder is volgens (3;9)

$$P[SM \vee MS] = P[SM] + P[MS] = 2pq ,$$

zodat definitie (4;4) geeft:

$$P[SM \mid SM \vee MS] = \frac{P[SM \wedge (SM \vee MS)]}{P[SM \vee MS]} = \frac{P[SM]}{P[SM \vee MS]} = \frac{pq}{2pq} = \frac{1}{2} .$$

Hetzelfde geldt voor de andere voorwaardelijke kans, dus is de uitkomst:

(5;1) $$P[SM \mid SM \vee MS] = P[MS \mid SM \vee MS] = \frac{1}{2} .$$

Opmerkingen

Deze berekening is geheel analoog aan de in hoofdstuk 3 gegeven berekening, die tot (3;19) leidde. Nemen wij voor het experiment een worp met een knoop, waarbij S resp. M voorstelt, dat de holle resp. de bolle kant boven komt (zodat p niet $\frac{1}{2}$ behoeft te zijn) en voeren wij paren worpen uit, tot er ofwel SM ofwel MS komt, dan is dus de kans op deze beide uitkomsten gelijk aan $\frac{1}{2}$. Langs deze eenvoudige weg kan dus een kans $\frac{1}{2}$ "geconstrueerd" worden. De in hoofdstuk 3 beschreven procedure, die tot (3,21) leidde, stelt ons dan, analoog uitgevoerd, in staat willekeurige rationale kansen te construeren, dus reële willekeurig dicht te benaderen.

Voorbeeld 2

Een doosje bevat n briefjes, genummerd van 1 tot n. Deze briefjes worden, nadat zij eerst grondig dooreen geschud zijn, één voor één blindelings uit de doos genomen en in volgorde van trekking neergelegd. Hoe groot is de kans, dat zij in een bepaalde van tevoren gegeven volgorde zullen komen te liggen.

Oplossing

Als model nemen wij in dit geval, dat bij elke trekking ieder der nog in de doos aanwezige briefjes gelijke kans bezit om bij de eerstvolgende trekking getrokken te worden. Als de briefjes even groot, even zwaar, even glad en op dezelfde wijze gevouwen (of niet gevouwen) zijn, kortom als men de voorzorg neemt ze - - afgezien van hun nummer - vrijwel identiek te maken, voldoet dit model - naar experimenteel aangetoond kan worden - goed.[1] Door dit model is het kansveld volledig beschreven.

De kans, dat de briefjes b.v. in de volgorde $1,2,3,\ldots,n$ getrokken zullen worden is nu gemakkelijk te berekenen met behulp van stelling 3;2. Geven wij het nummer van de trekking aan met een index, dan is, daar in dit geval de definitie van LAPLACE toepasbaar is:

$$P[1_1] = \frac{1}{n}\ ,\ P[2_2 \mid 1_1] = \frac{1}{n-1}\ ,\ P[3_3 \mid 1_1 \wedge 2_2] = \frac{1}{n-2}\ ,\quad \text{enz.},$$

dus volgens stelling 3;2

$$(5;2)\qquad P[1_1 \wedge 2_2 \wedge \ldots \wedge n_n] = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \ldots \cdot \frac{1}{2} \cdot 1 = \frac{1}{n!}\ .$$

Ditzelfde geldt echter ook voor iedere andere volgorde, m.a.w. alle $n!$ permutaties hebben gelijke kans $\frac{1}{n!}$.

Opmerking

Trekkingen, waarbij alle nog aanwezige briefjes (of in het algemeen elementen) gelijke kans hebben om getrokken te worden, worden aselecte trekkingen[2] genoemd. Wij hebben hier met aselecte trekkingen zonder teruglegging te maken. Wordt ieder getrokken briefje voor de uitvoering der volgende trekking weer in de doos gelegd, dan zijn het aselecte trekkingen met terugleggingg; maar dan is schudden voor iedere trekking nodig.

---

1) Daarbij valt op te merken, dat het dan niet nodig is tussen ieder tweetal trekkingen opnieuw te schudden; éénmaal grondig schudden van tevoren is genoeg.
2) Deze term is geïntroduceerd door Prof.Dr D. VAN DANTZIG.

Voorbeeld 3

Uit de doos met n briefjes worden aselect en zonder terugleggin$g$ k ($k \leqq n$) briefjes getrokken. Hoe groot is de kans, dat deze de nummers 1,...,k in één of andere volgorde bevatten?

Oplossing

De kans, dat deze k briefjes de nummers 1,..., k in de natuurlijke volgorde dragen is, zoals direct uit het vorige voorbeeld blijkt

$$P[1_1 \wedge 2_2 \wedge \ldots \wedge k_k] = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \ldots \cdot \frac{1}{n-k} = \frac{(n-k)!}{n!}.$$

Voor iedere andere gegeven volgorde van deze k nummers geldt hetzelfde en daar er slechts één volgorde uit kan komen, is volgens stelling 3;3 (de bijzondere optelregel), de kans dat deze nummers er in de één of andere volgorde uit zullen komen gelijk aan de som van deze kansen. Daar er $k!$ verschillende volgorden mogelijk zijn, is de gezochte kans gelijk aan

$$(5;3) \qquad \frac{k!\,(n-k)!}{n!} = \binom{n}{k}^{-1}.$$

Opmerking

Deze uitkomst geldt uiteraard niet alleen voor de nummers 1,..., k, doch voor iedere k-tal. De kans op ieder k-tal is dus dezelfde, zoals uit symmetrie-overwegingen reeds zonder meer te zien is. Daar er $\binom{n}{k}$ verschillende k-tallen zijn, volgt (5;3) hieruit met behulp van stelling 3;5. Een aselect zonder teruggelegging getrokken k-tal wordt, indien de volgorde van trekking buiten beschouwing gelaten wordt, een steekproef van de omvang k genoemd.

Voorbeeld 4

Uit de doos met n briefjes worden er k ($k \leqq n$) aselect en zonder terugleggin$g$ getrokken en in volgorde van trekking neergelegd. Hoe groot is de kans, dat er op ieder van h ($h \leqq k$) van tevoren aangegeven plaatsen een van tevoren gegeven nummer terecht komt?

Oplossing

Dit vraagstuk laat zich b.v. langs de volgende directe weg oplossen. Onder de k getrokken nummers moeten zich, wil de genoemde gebeurtenis, die wij met G aangeven, kunnen optreden, de h gegeven nummers bevinden. Er zijn $\binom{n}{k}$ k-tallen, die gelijke kans bezitten (voorbeeld 3) en hieronder zijn er $\binom{n-h}{k-h}$, die de h gegeven nummers bevatten. Derhalve is de kans, dat het

getrokken k-tal de h nummers bevat gelijk aan

$$\binom{n-h}{k-h} \Big/ \binom{n}{k} .$$

Is nu aan deze voorwaarde voldaan, dan zijn alle permutaties der getrokken lootjes even waarschijnlijk (voorbeeld 2). Er zijn k! permutaties, waaronder $(k-h)!$, waarbij G optreedt. De voorwaardelijke kans op G is dus

$$\frac{(k-h)!}{k!} ,$$

zodat

$$P[G] = \frac{(k-h)!}{k!} \quad \frac{\binom{n-h}{k-h}}{\binom{n}{k}} = \frac{(n-h)!}{n!} .$$

Deze eenvoudige uitkomst doet vermoeden, dat er een eenvoudigere oplossing bestaat. Dit is inderdaad het geval. Na het trekken van de eerste k briefjes kan men nl. rustig doorgaan en alle briefjes uit de doos trekken. Het al of niet optreden van G wordt daardoor immers niet meer beïnvloed. Nu zijn er in totaal n! even waarschijnlijke permutaties en daarvan zijn er $(n-h)!$ waarbij G optreedt. Waaruit de gevonden uitkomst direct volgt.

Voorbeeld 5

Uit een doos met M witte en N zwarte knikkers worden aselect k knikkers genomen. Hoe groot is de kans, dat de k^e knikker wit is.

A. als de trekkingen met terugleggin; geschieden,

B. als de trekkingen zonder terugleggin; geschieden.

Oplossing

Bij trekkingen met terugleggin; is de k^e trekkin; een aselecte trekkin; uit dezelfde verzamelin; knikkers als de eerste, zodat volgens de definitie van LAPLACE de gevraagde kans gelijk is aan

(5;4) $$\frac{M}{M+N} .$$

Bij trekkingen zonder terugleggin; is de samenstellin; van de knikker-verzamelin; bij de k^e trekkin; niet dezelfde als bij de eerste, maar hangt van de resultaten der eerste k-1 trekkingen af. Toch is de uitkomst dezelfde als eerst, zoals op de volgende wijze in te zien is. Als wij na de k^e trekkin; doorgaan tot alle M+N knikkers getrokken zijn, beïnvloeden wij het resultaat van de k^e trekkin; niet. Denken wij de knikkers genummers van 1,..., M+N , dan zijn alle $(M+N)!$ permutaties der trekkingsvolgorde even waarschijnlijk, zodat alle knikkers dezelfde kans bezitten om de

k^e getrokkene te zijn. Er zijn dus voor de k^e trekking weer $M+N$ mogelijkheden met gelijke kansen, waaruit volgt dat de kans op een witte k^e knikker weer gelijk is aan

$$\frac{M}{M+N}.$$

Opmerking

De berekende kans is de onvoorwaardelijke kans, dat de k^e knikker wit is. De voorwaardelijke kans hierop, bij gegeven resultaat van de eerste $k-1$ trekkingen is afhankelijk van dit resultaat en dus in het algemeen niet gelijk aan $M/M+N$. Het verschil tussen aselecte trekkingen met en zonder teruglegging is dus, dat in het eerste geval de trekkingen stochastisch onafhankelijk zijn (anders gezegd: de kans is bij iedere trekking opnieuw, wat ook de vorige voor resultaat geven, dezelfde), terwijl in het laatste geval de trekkingen stochastisch afhankelijk zijn, ook al is de onvoorwaardelijke kans voor alle trekkingen dezelfde.

Voorbeeld 6

Uit een doos met M witte en N zwarte knikkers worden er n aselect met teruglegging getrokken. Hoe groot is de kans, P_x, dat er x witte knikkers onder de n getrokkene voorkomen?

Oplossing

Geven wij "wit" en "zwart" aan met W en Z, dan zijn, bij iedere trekking, de kansen daarop

$$P[W] = \frac{M}{M+N} \ (= p) \quad , \quad P[Z] = \frac{N}{M+N} \ (= q = 1-p).$$

De trekkingen zijn onafhankelijk (zie opmerking bij voorbeeld 5), zodat wij de bijzondere productregel (4;16) kunnen toepassen. Daaruit volgt, dat de kans op een rij uitkomsten van de vorm

$$W\ W\ M\ W\ M\ M\ \ldots\ M$$

gelijk is aan

$$p\ p\ q\ p\ q\ q\ \ldots\ q,$$

dus, als er in die rij x witte en $n-x$ zwarte voorkomen, gelijk aan $p^x q^{n-x}$. Het aantal verschillende rijen met x witte en $n-x$ zwarte is gelijk aan het aantal x-tallen plaatsen onder n plaatsen, dus $\binom{n}{x}$. Volgens de bijzondere optelregel is dus

(5;5) $$P_x = \binom{n}{x} p^x q^{n-x} \qquad (\text{met } \binom{n}{0} = 1).$$

Opmerking

Dezelfde uitkomst wordt verkregen bij iedere rij van n onafhankelijke experimenten met kans p op "succes" voor ieder daarvan; x is dan het aantal successen. Zijn b.v. de knikkers in de doos niet alle even groot of even zwaar en wordt volgens een bepaalde procedure getrokken, dan is p, de kans op een witte knikker, niet noodzakelijk meer gelijk aan $M/M+N$, maar (5;5) blijft geldig, alleen met een andere (wellicht onbekende) waarde van p. De grootheid x kan de waarde $0,1,\ldots,n$ aannemen. Daar deze $n+1$ mogelijkheden een katagorisch systeem vormen, geldt volgens stelling 3;4

$$(5;6) \qquad \sum_{x=0}^{n} P_x = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = 1,$$

hetgeen overeenkomt met de binomiale ontwikkeling van

$$(p+q)^n.$$

Voorbeeld 7

Uit een doos met M witte en N zwarte knikkers worden er n aselect zonder teruglegging getrokken. Hoe groot is de kans P'_x, dat er x witte knikkers onder de n getrokkene voorkomen?

Oplossing

Volgens voorbeeld 3 hebben alle $\binom{M+N}{n}$ n-tallen gelijke kans om getrokken te worden. Het aantal verschillende n-tallen, die samengesteld zijn uit x witte en $n-x$ zwarte knikkers, bedraagt

$$\binom{M}{x}\binom{N}{n-x},$$

zodat volgens de definitie van LAPLACE geldt:

$$(5;7) \qquad P'_x = \frac{\binom{M}{x}\binom{N}{n-x}}{\binom{M+N}{n}}.$$

Opmerking

De grootheid x neemt nu gehele waarden aan tussen $\max(n-N,0)$ en $\min(M,n)$, deze grenzen inbegrepen. Buiten deze grenzen is $P'_x = 0$ en het rechterlid van (5;7) ook volgens de afspraak dat $\binom{a}{b} = 0$ als $b<0$ of $>a$. Derhalve geldt volgens stelling 3;4

$$(5;8) \qquad \sum_{x} \binom{M}{x}\binom{N}{n-x} = \binom{M+N}{n}.$$

Als n klein is t.o.v. N en M is

(5;9) $$P'_x \approx P_x \qquad \text{met} \quad p = \frac{M}{M+N}.$$

Dit is als volgt in te zien.

$$P'_x = \frac{M!}{x!(M-x)!}\,\frac{N!}{(n-x)!(N-n+x)!}\,\frac{n!(M+N-n)!}{(M+N)!}.$$

hierin is, als $x << M$,

$$\frac{M!}{(M-x)!} = M(M-1)\cdots(M-x+1) \approx M^x$$

en, als $n-x << N$,

$$\frac{N!}{(N-n+x)!} = N(N-1)\cdots(N-n+x+1) \approx N^{n-x}.$$

Voor $n << M+N$ is verder

$$\frac{(M+N)!}{(M+N-n)!} = (M+N)(M+N-1)\cdots(M+N-n+1) \approx (M+N)^n$$

en indien deze drie benaderingen ingevuld worden, komt er

(5;10) $$P'_x \approx \binom{n}{x}\left(\frac{N}{N+M}\right)^x\left(\frac{M}{N+M}\right)^{n-x}.$$

Is $n << N, M$, dan is ook $x << N$, $n-x << M$ en $n << N+M$.

Hieruit blijkt - wat ook zonder bewijs wel duidelijk is - dat het niet teruglegggen der getrokken elementen wel verwaarloosd kan worden zolang het er slechts weinig zijn in vergelijking met de aanwezige elementen, die dezelfde kenmerken dragen.

Voorbeeld 8

Een reeks van onafhankelijke experimenten, ieder met kans p op succes, wordt voortgezet tot precies k successen verkregen zijn. Hoe groot is de kans, Q_n, dat hiervoor n uitvoeringen van het experiment nodig zijn?

Oplossing

De gezochte kans is gelijk aan de kans, dat bij de eerste $n-1$ experimenten $k-1$ successen verkregen worden en bij het n^{de} ook een succes. De kans op het laatste is p en die op het eerste wordt door (5;5) gegeven met $n-1$ in plaats van n en $k-1$ in plaats van x.

Wegens de onafhankelijkheid der experimenten moeten deze beide kansen met elkaar vermenigvuldigd worden, hetgeen leidt tot

(5;11) $$Q_n = \binom{n-1}{k-1} p^k q^{n-k} \qquad (q = 1-p).$$

Opmerking

Derhalve is, wegens axioma V,

(5;12) $$\sum_{n=k}^{\infty} Q_n = \sum_{n=k}^{\infty} \binom{n-1}{k-1} p^k q^{n-k} = 1.$$

Voorbeeld 9 [1]

Een dictator, die van oordeel is, dat de verhouding van het aantal jongensgeboorten tot het aantal meisjesgeboorten niet groot genoeg is, vaardigt een wet uit, die inhoudt dat in een gezin, waarin een meisje geboren is verder geen kinderen ter wereld mogen komen. Wat is de invloed van deze maatregel, indien aangenomen mag worden dat het geslacht van een nieuwe baby onafhankelijk is van dat van eventuele vroegere baby's van het gezin, terwijl de kans op een jongensgeboorte voor alle gezinnen dezelfde is?

Oplossing

Daar het door deze maatregel aan gezinnen, waarin een meisje geboren is, verboden wordt nog meer kinderen te krijgen, zal het totale aantal geboorten dalen, indien daar geen compensatie tegenover wordt gesteld. Onder de gestelde omstandigheden wordt echter de verhouding tussen de aantallen jongens- en meisjes-geboorten niet beïnvloed, daar bij iedere nieuwe geboorte de kans op een jongen gelijk aan p is, in welk gezin deze geboorte ook plaats vindt. Men kan het ook zo stellen; de ooievaar, die de kinderen brengt, neemt telkens een baby aselect uit een voorraad, waarin een vaste verhouding der geslachten aanwezig is. Door de invoering van de nieuwe wet brengt hij sommige baby's op andere adressen dan anders het geval zou zijn geweest, maar hij put ze op dezelfde wijze uit zijn voorraad, zodat er, afgezien van een zekere vertraging in de aflevering, niets aan de situatie verandert.

Opmerking

Indien de kans op een jongensgeboorte van gezin tot gezin verschilt heeft de maatregel wel invloed op de geslachts-samen-stelling van de bevolking. Immers dan wordt de omvang van gezinnen met een kleinere kans op jongens sterker verkleind dan die met een grotere kans op jongens, omdat in de regel in de eerstgenoemde gezinnen eerder een meisje geboren wordt dan in de laatsgenoemde.

---

1) Dit voorbeeld werd mij medegedeeld door Prof.Dr N.G. de Bruijn. Het heeft niet direct betrekking op het berekenen van kansen, maar dient ter illustratie van het onafhankelijkheidsbegrip.

Hoofdstuk 6. Het meten van kansen

In het vorige hoofdstuk zijn een aantal voorbeelden gegeven van het berekenen van kansen. Dit kan geschieden, zoals in voorbeeld 6, door de berekening te baseren op een gegeven kans p, die dan ook in het resultaat (5;5) voorkomt; in voorbeeld 1 is uitgegaan van een onbekende kans p, die geëlimineerd werd, zodat een bekend getal als uitkomst verkregen werd; ook bij de overige voorbeelden, behalve bij het laatste, konden de gevraagde kansen expliciet berekend worden, op grond van bepaalde gegevens, die de aard van symmetrie-voorwaarden bezitten. In feite werd in al die gevallen, waarin numerieke uitkomsten verkregen zijn, uitgegaan van de gelijkheid van een aantal kansen.

Voor de practische toepassing van de kansrekening kan men echter met de beheersing van dergelijke eenvoudige situaties niet volstaan. Algemeen geldt, dat het voor de toepasbaarheid van een axiomatisch uitgevoerd maatbegrip nodig is dit ook practisch meetbaar te maken, d.w.z. experimentele middelen aan te geven om er in practijkproblemen een (min of meer) bepaalde waarde aan toe te kennen. Denkt men b.v. aan het begrip "lengte", dan zal men een practisch uitvoerbare procedure aan moeten geven, die aan de volgende eisen voldoet.

a. Men moet na kunnen gaan of twee voorwerpen even lang zijn en, zo neen, welk het langst is.

b. Men voert een materiële standaard (standaardmeter) in.

c. Op grond van a en b moet aan ieder voorwerp een getal als lengte toegevoegd kunnen worden.

Pas als aan deze eisen voldaan is, kan de theorie ook in toepassing gebracht worden en deze toepassing strekt zich dan ook alleen uit tot objecten, waarvoor instrumenten ontworpen zijn, waarmee aan deze eisen voldaan kan worden.

Hetzelfde geldt voor het kansbegrip en de theorie daarvan (de kansrekening). Het instrumentarium, dat aan de toepasbaarheid ten grondslag ligt, wordt grotendeels geleverd door de statistiek, waarbij gebruik gemaakt wordt van de kansrekening, o.a. in de vorm van eenvoudige principes, die in de behandelde voorbeelden besloten zijn.

Punt b is daarbij het eenvoudigst, daar een absolute maat voor het kansbegrip reeds in de axioma's II en III vervat is. Wij kunnen dan ook, zoals uit voorbeeld 1 van hoofdstuk 5 blijkt, langs eenvoudige weg en met simpele middelen een kans van willekeurige gegeven grootte "construeren".

De punten a en c eisen meer voorbereiding. Wij zullen verderop in deze cursus de oplossing van a resp. c ontmoeten in de vorm van statistische toetsen voor de hypothesen dat twee onbekende kansen gelijk zijn resp. dat een onbekende kans een gegeven waarde bezit. De theorie der betrouwbaarheidsintervallen - ten nauwste verbonden met de toetsingstheorie - maakt het mogelijk grenzen aan te geven voor het verschil tussen twee onbekende kansen resp. de waarde van één onbekende kans, terwijl de schattingstheorie de middelen beschrijft om een zo goed mogelijke getalwaarde ("schatting") van kansen en verschillen daarvan te verkijgen en de nauwkeurigheid daarvan te onderzoeken. Al deze methoden berusten op de verwerking van waarnemingen, die verricht dienen te worden onder zodanige voorwaarden, dat de wiskundige modellen, waarop deze statistische methoden berusten, een getrouwe afspiegeling van de werkelijkheid vormen.

Deze methoden, die alle verderop uitvoerig ter sprake komen, vormen uiteraard slechts een klein onderdeel van de statistiek, een eerste stap op de goede weg.

Hoofdstuk 7. Stochastische grootheden, kansverdelingen

In dit en het volgende hoofdstuk worden een aantal belangrijke begrippen en stellingen uit de kansrekening behandeld.

Definitie 7;1

Een grootheid $\underline{x}$, die verschillende waarden aan kan nemen, wordt een stochastische grootheid[1] genoemd indien voor ieder gegeven reëel getal x de kans, dat $\underline{x}$ een waarde aanneemt, die hoogstens daaraan gelijk is, gedefinieerd is.

Voorbeelden

Grootheden, die in een wiskundig model door stochastische grootheden voorgesteld kunnen worden, zijn de volgende. Het aantal successen in een reeks onafhankelijke experimenten met ieder een kans p op succes. Het aantal worpen met een munt tot voor het eerst kruis verkregen wordt. Maar ook: de gewichtstoename van een proefdier tijdens een proefperiode; de dikte van de speklaag van de 5^e walvis, die op een bepaalde dag door een walvisvaarder gevangen wordt; het aantal telefoongesprekken, dat gedurende een bepaald tijdsinterval over een bepaalde lijn gevoerd wordt.

Notatie

Stochastische grootheden geven wij aan met onderstreepte latijnse letters: $\underline{x}, \underline{y}, \underline{z}, \underline{x}_1, \underline{x}_2, \underline{x}_i$ enz., terwijl getalwaarden, die zij aan kunnen nemen, vaak beschouwd als algebraische variabelen, door dezelfde letters zonder onderstreping aangegeven worden.

(7;1) $\qquad P[\underline{x} \leqq x] \stackrel{def}{=}$ de kans dat $\underline{x}$ een waarde $\leqq x$ aanneemt.

In plaats van onderstreepte letters worden vaak ook vet gedrukte letters of hoofdletters gebruikt.

Bijbehorende termen

De kansverdeling van een stochastische grootheid $\underline{x}$ is de verzameling van alle waarden x, die $\underline{x}$ aan kan nemen met de bijbehorende kansen $P[\underline{x} \leqq x]$. De verdelingsfunctie van $\underline{x}$ is de functie

(7;2) $\qquad F(x) \stackrel{def}{=} P[\underline{x} \leqq x],$

opgevat als functie van de algebraïsche variabele x.

---

1) στοχαζειν = raden; een grootheid, naar de waarde waarvan men slechts gissen kan.

Iedere verdelingsfunctie is een functie, die monotoon niet dalend is en uitsluitend waarden in het gesloten interval $[0,1]$ aanneemt, terwijl

(7;3) $$F(-\infty) = 0 \quad , \quad F(+\infty) = 1.$$

Definitie 7;2

Discrete[1] kansverdelingen zijn kansverdelingen, waarbij de beschouwde stochastische grootheid een eindig of aftelbaar aantal waarden $x_1, x_2, \ldots$ aan kan nemen, terwijl

(7;3) $$p_i \stackrel{def}{=} P[\underline{x} = x_i] > 0 \quad , \quad \sum_i p_i = 1.$$

De bijbehorende verdelingsfuncties zijn trapfuncties:

(7;4) $$F(x) = \sum_{x_i \leqq x} p_i .$$

Voorbeelden

De binomiale verdeling (voorbeeld 6 van hoofdstuk 5):

(7;5) $$P[\underline{x} = x] = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, \ldots, n;\ q = 1-p).$$

De hypergeometrische verdeling (voorbeeld 7 van hoofdstuk 5):

(7;6) $$P[\underline{x} = x] = \frac{\binom{M}{x}\binom{N}{n-x}}{\binom{M+N}{n}}$$ (x geheel tussen $\max(n-N, 0)$ en $\min(M, n)$).

De negatief binomiale verdeling (voorbeeld 8 van hoofdstuk 5):

(7;7) $$P[\underline{x} = x] = \binom{x-1}{k-1} p^k q^{x-k} \quad (x = k, k+1, \ldots;\ q = 1-p).$$

Definitie 7;3

Continue kansverdelingen zijn kansverdelingen, waarbij de stochastische grootheid een continuum van waarden (bijv. alle waarden van een eindig of oneindig interval) aan kan nemen, terwijl de verdelingsfunctie $F(x)$ voor iedere x, op hoogstens eindig veel na, continu differentiëerbaar is. De afgeleide

(7;8) $$f(x) \stackrel{def}{=} \frac{d}{dx} F(x)$$

---

1) Of: discontinue.

wordt dan de <u>verdelingsdichtheid</u> genoemd. Derhalve geldt

$$(7;9) \qquad F(x) = \int_{-\infty}^{x} f(v)\,dv \quad 1)$$

en, zoals op grond van de axioma's gemakkelijk te bewijzen is

$$(7;10) \qquad P[x_1 \leqq \underline{x} \leqq x_2] = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx,$$

$$(7;11) \qquad \int_{-\infty}^{+\infty} f(x)\,dx = 1,$$

$$(7;12) \qquad P[\underline{x} = x] = 0 \qquad \text{voor iedere } x\ .$$

Deze kansverdelingen treden op als limiet van discrete kansverdelingen (zoals wij later zullen zien; zij worden daarom vaak voor benaderingsdoeleinden gebruikt) en als wiskundig model bij grootheden, die quantitatief op een bij benadering continue schaal afgelezen worden. Voorbeelden daarvan zijn: de lengte van een recruut, de tijdsduur van een treinrit, de snelheid van een vliegtuig.

<u>Voorbeelden</u>

De belangrijkste continue kansverdeling is de <u>normale</u>, waarvoor de volgende formules gelden:

$$(7;13) \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad 2) \qquad (-\infty < x < \infty)$$

waarin $\sigma\,(>0)$ en μ constanten ("<u>parameters</u>") zijn. De verdelingsfunctie is

$$(7;14) \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-\mu)^2}{2\sigma^2}}\,dv.$$

<u>Notatie</u>

Bezit $\underline{x}$ een normale verdeling met parameters μ en σ, dan wordt dit aangegeven met: $\underline{x}$ is $N(\mu,\sigma)$ -verdeeld.

De vorm van de verdeling is in figuur 7;1 aangegeven; σ is de afstand van het punt $x = \mu$ tot de abscis van het buigpunt van $f(x)$.

---

1) Als de verdeling in werkelijkheid "begint" bij een punt $x = a$ kan men ook a als ondergrens van de integraal nemen; daar dan echter $F(x) = 0$ voor $x \leqq a$, dus ook $f(x) = 0$ in dat gebied, is de ondergrens $-\infty$ steeds correct. Analoge overwegingen gelden voor de bovengrens, indien deze het "einde" van de verdeling overschrijdt.

2) Dat voor deze functie inderdaad de relatie (7;11) geldt vindt men bewezen in § 1 van de appendix.

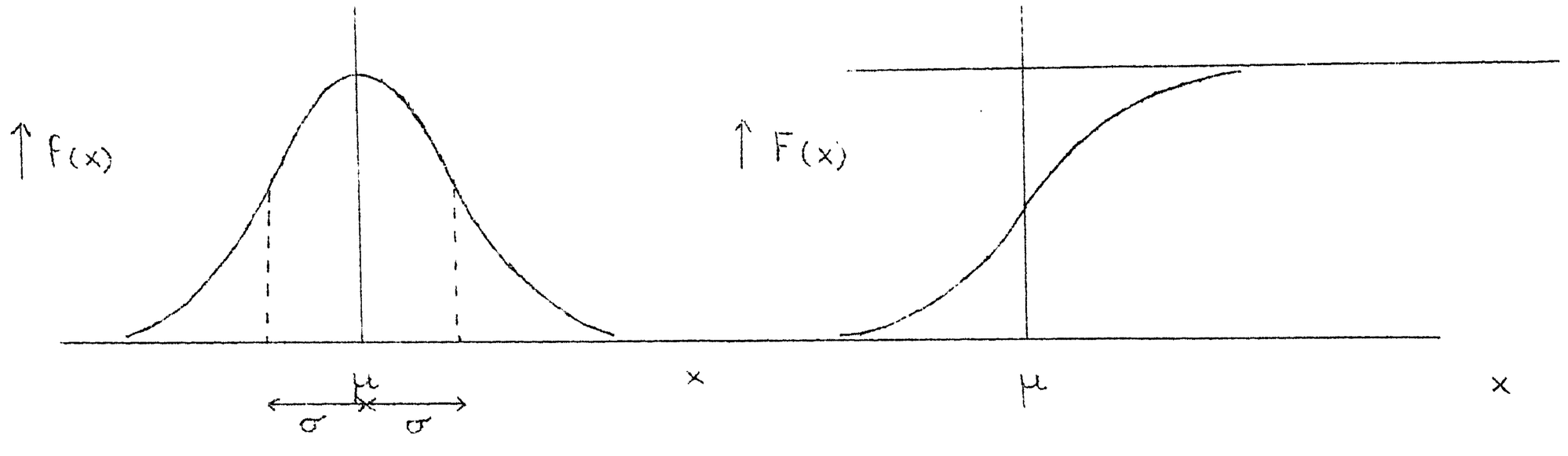


Fig. 7;1. Normale verdeling

Is $\mu=0$ en $\sigma=1$, dus is $\underline{x}$ $N(0,1)$-verdeeld, dan spreekt men van de gestandaardiseerde normale verdeling. Voor grootheden, die deze verdeling bezitten, gebruiken wij het symbool $\underline{u}$. De gestandaardiseerde normale verdeling is uitvoerig getabelleerd.

Een andere continue verdeling, die bijv. van belang is voor de theorie van het afronden van meet- of berekeningsresultaten, is de homogene verdeling, waarvoor de volgende formules gelden.

$$(7;15) \qquad f(x) = \begin{cases} \dfrac{1}{\theta_2-\theta_1} & \text{voor} \quad \theta_1 \leqq x \leqq \theta_2, \\ 0 & \text{elders} \end{cases}$$

$$(7;16) \qquad F(x) = \begin{cases} 0 & \text{voor} \quad x < \theta_1, \\ \dfrac{x-\theta_1}{\theta_2-\theta_1} & \text{voor} \quad \theta_1 \leqq x \leqq \theta_2, \\ 1 & \text{voor} \quad x > \theta_2. \end{cases}$$

De "eindpunten" θ_1 en θ_2 treden hier als parameters op.[1)]

Definitie 7;4

Gemengde kansverdelingen zijn kansverdelingen, die verkregen kunnen worden als gewogen gemiddelden van een discrete en een continue, d.w.z. waarvoor geldt:

$$(7;17) \qquad F(x) = \lambda F_1(x) + (1-\lambda) F_2(x), \qquad (0<\lambda<1)$$

waarin $F_1(x)$ een trapfunctie en $F_2(x)$ continu differentieerbaar is.

---

1) Parameters van kansverdelingen worden meestal door griekse letters aangegeven. In bepaalde gevallen, zoals bij de parameter p van een binomiale verdeling, zijn andere symbolen zozeer ingeburgerd, dat van deze regel afgeweken wordt.

Voorbeeld

De schadeuitkering, die in één jaar op een verzekeringspolis moet geschieden. Er is een positieve kans op uitkering 0 (geen verhaalbare schade), terwijl verder een (vrijwel) continue schaal van mogelijke uitkeringen bestaat.

Opmerking

De mogelijkheden zijn hiermede nog niet uitgeput. Men kan continue verdelingsfuncties beschouwen, die een niet continue afgeleide bezitten of die in oneindig veel punten niet differentiëerbaar zijn. Deze pathologische gevallen zijn voor de practijk van weinig of geen belang en blijven hier buiten beschouwing.

Definitie 7;5

Een aantal stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ bezit een simultane kansverdeling als voor ieder stel reële getallen $x_1, x_2, \ldots, x_n$ de (simultane) verdelingsfunctie

$$(7;18)\quad F(x_1, x_2, \ldots, x_n) \stackrel{def}{=} P[\underline{x}_1 \leqq x_1 \wedge \underline{x}_2 \leqq x_2 \wedge \ldots \wedge \underline{x}_n \leqq x_n]$$

gedefiniëerd is.

Voorbeelden

Het aantal worpen met een munt tot voor de tweede maal kruis verkregen wordt en het nummer van de worp, waarbij in die rij worpen de eerste keer kruis viel. De lichaamslengte en de borstomvang van een aselect uit de Nederlandse mannelijke bevolking genomen man. In deze beide gevallen is $n=2$ (tweedimensionale verdelingen). Een n-dimensionale verdeling wordt bijv. verkregen indien men de temperaturen op verschillende punten van Nederland beschouwt, gemeten op een aselect gekozen dag om 12 uur.

Definitie 7;6

Een discrete n-dimensionale kansverdeling is een verdeling, waarbij in de n-dimensionale ruimte R_n met $x_1, x_2, \ldots, x_n$ als coördinaten een eindig of aftelbaar aantal punten $P_j \equiv (x_{1,j}, x_{2,j}, \ldots, x_{n,j})$ $(j = 1, 2, \ldots)$ aanwezig zijn, waarvoor geldt

$$(7;19)\qquad P[\underline{P} = P_j] = p_j > 0\ ,\quad \sum_j p_j = 1,$$

waarin $\underline{P} \equiv (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$ een stochastisch punt in R_n voorstelt.

De bijbehorende verdeling is dan een terrasfunctie:

$$F(x_1, x_2, \ldots, x_n) = \sum_{\substack{x_i,j \leqq x_i \\ (i=1,2,\ldots,n)}} p_j . \tag{7;20}$$

Voorbeelden

Stellen $\underline{x}_1$ resp. $\underline{x}_2$ de nummers voor van de worpen, in een rij onafhankelijke worpen met een munt, waarbij voor het eerst resp. voor de tweede maal kruis verkregen wordt, dan is, zoals gemakkelijk te zien is, als de kans op kruis p is:

$$P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2] = p^2 q^{x_2 - 2} \qquad (x_2 > x_1). \tag{7;21}$$

Deze kans is dus voor iedere $x_1 < x_2$ dezelfde. Dit doet vermoeden, dat bij gegeven x_2 alle mogelijke waarden van x_1 dezelfde kans zullen bezitten. Dat dit zo is kan als volgt bewezen worden (vgl. 7;7):

$$P[\underline{x}_1 = x_1 | \underline{x}_2 = x_2] = \frac{P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2]}{P[\underline{x}_2 = x_2]} = \tag{7;22}$$

$$= \frac{p^2 q^{x_2 - 2}}{\binom{x_2 - 1}{1} p^2 q^{x_2 - 2}} = \frac{1}{x_2 - 1} .$$

Onder de voorwaarde $\underline{x}_2 = x_2$ is $\underline{x}_1$ dus homogeen verdeeld over de getallen $1, 2, \ldots, x_2 - 1$.

Laat, als tweede voorbeeld, $\underline{x}_i$ $(i = 1, 2, \ldots, n)$ het aantal malen kruis voorstellen bij de i^e worp uit een rij onafhankelijke worpen met een munt. Iedere $\underline{x}_i$ kan dus de waarde 0 of 1 aannemen en de kansen hierop stellen wij q resp. p . Dan geldt dus

$$P[\underline{x}_i = 1] = p \quad , \quad P[\underline{x}_i = 0] = q \qquad (i = 1, 2, \ldots, n)$$

en wegens de onafhankelijkheid

$$P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2 \wedge \ldots \wedge \underline{x}_n = x_n] = p^{\sum x_i} q^{n - \sum x_i}, \tag{7;23}$$

waarin x_i $(i = 1, 2, \ldots, n)$ de waarde 0 of 1 bezit. Stellen wij

$$\underline{x} = \sum_{i=1}^{n} \underline{x}_i , \tag{7;24}$$

dan is $\underline{x}$ het aantal malen kruis bij n worpen, waarvoor dus (7;5), de binomiale verdeling, geldt.

Deze opbouw van een binomiaal verdeelde grootheid $\underline{x}$ met parameters n en p als som van een aantal onafhankelijke grootheden $\underline{x}_i$ met dezelfde p en $n = 1$ zal ons later nog te pas komen.

De lezer berekene de voorwaardelijke kans

$$P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2 \wedge \ldots \wedge \underline{x}_n = x_n \mid \underline{x} = x]$$

en interpretere de verkregen uitkomst.

Definitie 7;7

Een continue n-dimensionale kansverdeling is een kansverdeling, waarvoor de functie

$$(7;25) \qquad f(x_1, x_2, \ldots, x_n) \stackrel{def}{=} \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}$$

overal bestaat en continu is (discontinuïteit van f in eindig veel punten of op één of meer krommen is toegestaan; een exacte formulering hiervan zou te ver voeren). De functie f heet de (simultane) verdelingsdichtheid. De omgekeerde relatie van (7;24) is

$$(7;26) \quad F(x_1, x_2, \ldots, x_n) = \int_{\substack{v_i \leq x_i \\ (i=1,2,\ldots,n)}} \cdots \int f(v_1, v_2, \ldots, v_n)\, dv_1 dv_2 \ldots dv_n.$$

Is G een gebied in de ruimte R_n met coördinaten $x_1, x_2, \ldots, x_n$ en is $\underline{P} \equiv (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$, dan geldt

$$(7;27) \qquad P[\underline{P} \in G] = \int_G \cdots \int f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n.$$

Het bewijs van deze relatie is elementair als G een n-dimensionaal rechthoekig parallelopipedum is. Is G van een andere vorm, dan leidt overdekking van G met inkrimpende parallelopipeda en limietovergang tot het doel. Dit maattheoretische bewijs en de eisen, die aan G opgelegd moeten worden, behandelen wij hier niet.

Voorbeelden

Twee grootheden $\underline{x}$ en $\underline{y}$ bezitten een tweedimensionale normale verdeling, als hun verdelingsdichtheid de volgende vorm heeft:

$$(7;28) \quad f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}.$$

Hierin komen 5 parameters voor: $\mu_1, \mu_2, \sigma_1, \sigma_2$ en ρ.

Deze verdeling kan vaak als model gebruikt worden indien aan één aselect uit een grote verzameling gekozen object twee afmetingen waargenomen worden, bijv. lengte en breedte van een meeuwenei.

Een tweedimensionale homogene verdeling op een cirkelschijf wordt gegeven door

$$(7;29) \qquad f(x,y) = \begin{cases} \frac{1}{\pi r^2} & \text{voor } x^2 + y^2 \leqq r^2, \\ 0 & \text{voor } x^2 + y^2 > r^2. \end{cases}$$

Dit model treedt o.a. op bij het onderzoek van de verdeling van bacteriën of bacterie-koloniën in een Petri-schaal.

Ook in het meerdimensionale geval kunnen gemengde kans-verdelingen optreden, die wij echter niet beschouwen.

Marginale verdelingen en voorwaardelijke verdelingsdichtheden

Bezitten $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een n-dimensionale discrete verdeling (vgl. def. 7;6) dan geldt voor iedere i

$$(7;30) \qquad F_i(x_i) \overset{\text{def}}{=} P[\underline{x}_i \leqq x_i] = \sum_{\substack{j \\ x_{i,j} \leqq x_i}} p_j .$$

(Uit welke stelling volgt dit?)

$F_i(x_i)$ heet de marginale verdelingsfunctie van $\underline{x}_i$.

De marginale kansen worden gegeven door

$$(7;31) \qquad P[\underline{x}_i = x_i] = \sum_{\substack{j \\ x_{i,j} = x_i}} p_j$$

en analoog voor meerdimensionale marginale verdelingen.

Voorbeeld

De marginale verdeling van $\underline{x}_1$, behorende bij de door (7;21) gegeven verdeling, is volgens (7;31):

$$P[\underline{x}_1 = x_1] = \sum_{x_2 > x_1} p^2 q^{x_2 - 2} = \sum_{x_2 - x_1 > 0} p^2 q^{x_2 - x_1 - 2 + x_1} =$$

$$= p q^{x_1 - 1} \sum_{i=1}^{\infty} p q^{i-1} = p q^{x_1 - 1},$$

daar $\sum_{i=1}^{\infty} p q^{i-1} = 1$. Deze marginale verdeling is dus dezelfde als (7;7) met $k = 1$. (Was dit ook direct in te zien? De lezer berekene de marginale verdeling van $\underline{x}_2$).

Bezitten $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een n-dimensionale continue verdeling (vgl. def. 7;7), dan geldt volgens (7;27) voor de marginale verdelingsfunctie van $\underline{x}_i$:

$$(7;32) \qquad F_i(x_i) = \int_{v_i = -\infty}^{x_i} \int_{-\infty}^{+\infty} \cdots \int f(v_1, v_2, \ldots, v_n)\, dv_1\, dv_2 \ldots dv_n .$$

De marginale verdelingsdichtheid is dus

$$(7;33) \qquad f_i(x_i) = \frac{d}{dx_i} F_i(x_i) = \int_{-\infty}^{+\infty} \cdots \int f(v_1, v_2, \ldots, x_i, \ldots, v_n) \prod_{j \neq i} dv_j .$$

Meerdimensionale marginale kansverdelingen worden op analoge wijze uit de simultane verdeling afgeleid.

Voorbeelden

De marginale verdeling van $\underline{x}$ behorende bij (7;28) laat zich eenvoudig als volgt berekenen. Stellen wij

$$(7;34) \qquad u = \frac{x-\mu_1}{\sigma_1} \qquad \text{en} \qquad v = \frac{y-\mu_2}{\sigma_2}$$

dan is $dx = \sigma_1\, du$ en $dy = \sigma_2\, dv$.

Substitueren wij dit in (7;28) en stelt :: voor: "op een constante factor na gelijk aan", dan is dus

$$f(x,y) :: e^{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)}.$$

Volgens (7;33) geldt dus voor de marginale verdelingsdichtheid $f_1(x)$ van $\underline{x}$:

$$f_1(x) :: e^{-\frac{u^2}{2(1-\rho^2)}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}\{(v^2-2\rho uv+\rho^2u^2)-\rho^2u^2\}}\, dv =$$

$$= e^{-\frac{1}{2}u^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}(v-\rho u)^2}\, dv.$$

Door substitutie van $w = v-\rho u$ in de integraal blijkt, dat deze - daar de grenzen door de substitutie niet veranderen - niet meer van u afhangt, zodat dus geldt

$$f_1(x) :: e^{-\frac{1}{2}u^2} = e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}}.$$

Dit betekent echter, dat de marginale verdeling van $\underline{x}$ een $N(\mu_1,\sigma_1)$-verdeling is. De nog ontbrekende factor kan aangevuld worden met behulp van de voorwaarde (7;11), die uiteraard ook voor marginale verdelingen geldt. Uit formule (7;13) blijkt, dat het resultaat moet zijn

$$(7;35) \qquad f_1(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}}.$$

Een analoge formule geldt voor $f_2(y)$, de marginale verdelingsdichtheid van $\underline{y}$.

Voorwaardelijke verdelingen kunnen, in het discrete geval direct volgens de definitie van voorwaardelijke kansen berekend worden. Een voorbeeld hiervan was (7;22). In het continue geval doet zich de complicatie voor, dat de voorwaarde, waaronder men werkt, zelf een kans 0 bezit, zodat de definitie niet zonder meer bruikbaar is. Bezitten $\underline{x}_1$ en $\underline{x}_2$ een continue 2-dimensionale

verdeling, dan is dus $P[\underline{x}_2 = x_2] = 0$ voor iedere x_2, zodat voorwaardelijke kansen van de vorm

$$P[\underline{x}_1 \leqq x_1 | \underline{x}_2 = x_2]$$

vooralsnog niet gedefinieerd zijn. Deze kans wordt nu gedefinieerd als

$$(7;36) \quad F(x_1|x_2) = P[\underline{x}_1 \leqq x_1 | \underline{x}_2 = x_2] = \lim_{dx_2 \searrow 0} P[\underline{x}_1 \leqq x_1 | x_2 \leqq \underline{x}_2 \leqq x_2 + dx_2].$$

Volgens de definitie van een voorwaardelijke kans en (7;27) is dit gelijk aan

$$\lim_{dx_2 \searrow 0} \left\{ \int_{-\infty}^{x_1} dv_1 \int_{x_2}^{x_2+dx_2} dv_2 f(v_1,v_2) \Big/ \int_{-\infty}^{+\infty} dv_1 \int_{x_2}^{x_2+dx_2} dv_2 f(v_1,v_2) \right\}$$

en hierin is, na deling van teller en noemer door dx_2:

$$\lim_{dx_2 \searrow 0} \int_{-\infty}^{x_1} dv_1 \int_{x_2}^{x_2+dx_2} dv_2 f(v_1,v_2) \Big/ dx_2 = \frac{\partial}{\partial x_2} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(v_1,v_2)\, dv_1\, dv_2 = \frac{\partial}{\partial x_2} F(x_1,x_2),$$

terwijl voor de noemer geldt volgens (7;32):

$$\lim_{dx_2 \searrow 0} \int_{-\infty}^{+\infty} dv_1 \int_{x_2}^{x_2+dx_2} dv_2 f(v_1,v_2) \Big/ dx_2 = \frac{d}{dx_2} \int_{-\infty}^{+\infty} dv_1 \int_{-\infty}^{x_2} dv_2 f(v_1,v_2) =$$

$$= \frac{d}{dx_2} F_2(x_2) = f_2(x_2),$$

als F_2 en f_2 de marginale verdelingsfunctie resp. -dichtheid van $\underline{x}_2$ voorstellen. Derhalve is

$$(7;37) \qquad F(x_1|x_2) = \frac{\frac{\partial}{\partial x_2} F(x_1,x_2)}{f_2(x_2)}$$

en

$$(7;38) \qquad f(x_1|x_2) = \frac{\partial}{\partial x_1} F(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)}.$$

(De lezer controlere of $f(x_1|x_2)$ inderdaad een kansverdeling is, d.w.z. of $\int_{-\infty}^{+\infty} f(x_1|x_2) = 1$ is).

Voor meer dan 2 dimensies kan op analoge wijze de volgende formule afgeleid worden:

$$(7;39) \qquad f(x_1,x_2,\ldots,x_k|x_{k+1},x_{k+2},\ldots,x_n) =$$

$$= f(x_1,x_2,\ldots,x_n) \Big/ \int_{-\infty}^{+\infty} \cdots \int f(x_1,x_2,\ldots,x_n)\, dx_1\, dx_2 \ldots dx_k.$$

Voorbeeld

Wij berekenen nu de voorwaardelijke verdelingsdichtheid $f(x|y)$ van de tweedimensionale normale verdeling (7;28) Deze is, volgens (7;38) gelijk aan

$$\frac{f(x,y)}{f_2(y)}$$

waarin, volgens (7;35),

$$f_2(y) :: e^{-\frac{1}{2}\frac{(y-\mu_2)^2}{\sigma_2^2}}$$

is (:: betekent: op een constante factor na gelijk aan).

Voeren wij weer de substitutie (7;34) uit, dan komt er

$$f(x|y) :: e^{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)+\frac{1}{2}v^2} =$$

$$= e^{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+\rho^2v^2)} = e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2} =$$

$$= e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}-\rho\frac{y-\mu_2}{\sigma_2}\right)^2} =$$

$$= e^{-\frac{\left[x-\left\{\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)+\mu_1\right\}\right]^2}{2(1-\rho^2)\sigma_1^2}},$$

waarin nu y, wegens de voorwaarde $y=y$, als een gegeven constante optreedt. Derhalve volgt uit (7;13):

$$(7;40)\quad f(x|y) = \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{\left[x-\left\{\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)+\mu_1\right\}\right]^2}{2(1-\rho^2)\sigma_1^2}}.$$

Hoofdstuk 8. Onafhankelijkheid van stochastische grootheden

In hoofdstuk 4 is het begrip stochastische onafhankelijkheid gedefiniëerd voor gebeurtenissen, door middel van de productformules (4;8) en (4;15). Aequivalent hiermee is, dat voorwaardelijke en onvoorwaardelijke kansen gelijk zijn (formule (4;7)). Wij sluiten nier bij die definities aan, waarbij dan de beschouwde gebeurtenissen zijn: $\underline{x}_1 \leqq x_1$, $\underline{x}_2 \leqq x_2$, enz.

Definitie 8;1. Een aantal stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ zijn onafhankelijk verdeeld (stochastische onafhankelijk) indien zij een simultane kansverdeling bezitten, waarvan de verdelingsfunctie voldoet aan

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)\, F_2(x_2) \ldots F_n(x_n), \tag{8;1}$$

waarin F de simultane verdelingsfunctie is en F_i ($i = 1, 2, \ldots, n$) de marginale verdelingsfunctie van $\underline{x}_i$.

Opmerking

Vergelijking (8;1) is identiek met:

$$P[\underline{x}_1 \leqq x_1 \wedge \underline{x}_2 \leqq x_2 \wedge \ldots \wedge \underline{x}_n \leqq x_n] = \prod_{i=1}^{n} P[\underline{x}_i \leqq x_i]. \tag{8.2}$$

Voor de eenvoud beschouwen wij verder in het bijzonder het geval van een tweedimensionale verdeling van stochastische grootheden $\underline{x}_1$ en $\underline{x}_2$. De uitbreiding tot meer dan 2 dimensies is gemakkelijk aan te geven en de bewijzen verlopen analoog. Definities (8;1) en (8;2) gaan voor $n = 2$ over in

$$F(x_1, x_2) = F_1(x_1)\, F_2(x_2) \tag{8.1'}$$

en

$$P[\underline{x}_1 \leqq x_1 \wedge \underline{x}_2 \leqq x_2] = P[\underline{x}_1 \leqq x_1]\, P[\underline{x}_2 \leqq x_2]. \tag{8.2'}$$

Stelling 8;1

Zijn $\underline{x}_1$ en $\underline{x}_2$ stochastisch onafhankelijk, dan geldt

$$P[\underline{x}_1 \leqq x_1 \mid \underline{x}_2 \leqq x_2] = P[\underline{x}_1 \leqq x_1]. \tag{8;3}$$

Bewijs

Volgens de definitie van een voorwaardelijke kans, tezamen met (8;1'), geldt

$$P[\underline{x}_1 \leqq x_1 \mid \underline{x}_2 \leqq x_2] = \frac{F(x_1, x_2)}{F_2(x_2)} = \frac{F_1(x_1)\, F_2(x_2)}{F_2(x_2)} = F_1(x_1) = P[\underline{x}_1 \leqq x_1].$$

Stelling 8;2.

Zijn $\underline{x}_1$ en $\underline{x}_2$ discreet en onafhankelijk verdeeld, dan geldt

$$(8;4)\qquad \begin{cases} P[\underline{x}_1 \leqq x_1 \wedge \underline{x}_2 = x_2] = P[\underline{x}_1 \leqq x_1]\, P[\underline{x}_2 = x_2]\,, \\ P[\underline{x}_1 = x_1 \wedge \underline{x}_2 \leqq x_2] = P[\underline{x}_1 = x_1]\, P[\underline{x}_2 \leqq x_2] \end{cases}$$

en

$$(8;5)\qquad P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2] = P[\underline{x}_1 = x_1]\, P[\underline{x}_2 = x_2].$$

Bewijs

Is $x_2' < x_2$ de grootste waarde $< x_2$, die door $\underline{x}_2$ aangenomen kan worden, dan is $P[\underline{x}_2 < x_2] = P[\underline{x}_2 \leqq x_2']$ en daar (8;2') voor alle mogelijke waarden van x_1 en x_2 geldt is dus tevens

$$(8;6)\qquad P[\underline{x}_1 \leqq x_1 \wedge \underline{x}_2 < x_2] = P[\underline{x}_1 \leqq x_1]\, P[\underline{x}_2 < x_2].$$

Is er geen dergelijke waarde x_2', dan zijn linker- en rechterlid van (8;6) beide = 0, zodat de geldigheid behouden blijft.

Laat nu

A_1 resp. A_2 voorstellen $\underline{x}_1 \leqq x_1$ resp. $\underline{x}_2 \leqq x_2$,
B_1 resp. B_2 $\underline{x}_1 < x_1$ resp. $\underline{x}_2 < x_2$,
C_1 resp. C_2 $\underline{x}_1 = x_1$ resp. $\underline{x}_2 = x_2$.

Dan is $A_2 \equiv B_2 \vee C_2$, dus $A_1 \wedge A_2 \equiv (A_1 \wedge B_2) \vee (A_1 \wedge C_2)$
zodat

$$(8;7)\qquad \begin{aligned} P[A_1 \wedge A_2] &= P[A_1]\, P[A_2] = P[A_1]\, P[B_2 \vee C_2] = \\ &P[A_1]\{P[B_2] + P[C_2]\} = P[A_1]\, P[B_2] + P[A_1]\, P[C_2]. \end{aligned}$$

Anderzijds is volgens (8;6):

$$P[A_1 \wedge B_2] = P[A_1]\, P[B_2],$$

zodat

$$P[A_1 \wedge A_2] = P[(A_1 \wedge B_2) \vee (A_1 \wedge C_2)] = P[A_1 \wedge B_2] + P[A_1 \wedge C_2] = P[A_1]\, P[B_2] + P[A_1 \wedge C_2].$$

Uit deze relatie en (8;7) volgt direct, dat

$$(8;8)\qquad P[A_1 \wedge C_2] = P[A_1]\, P[C_2],$$

hetgeen overeenkomt met de eerste der twee vergelijkingen (8;4).
De tweede wordt uiteraard op dezelfde wijze bewezen.

Om (8;5) te bewijzen ontwikkelen wij nu $P[A_1 \wedge C_2]$ op analoge wijze met behulp van $A_1 \equiv B_1 \vee C_1$. Dit geeft:

$$P[A_1 \wedge C_2] = P[(B_1 \wedge C_2) \vee (C_1 \wedge C_2)] = P[B_1 \wedge C_2] + P[C_1 \wedge C_2] = P[B_1]\, P[C_2] + P[C_1 \wedge C_2],$$

daar (8;4) niet alleen geldt met $\underline{x}_1 \leqq x_1$, maar ook met $\underline{x}_1 \leqq x_1'$ als x_1' de grootste waarde $< x_1$ is, die $\underline{x}_1$ aan kan nemen (vergelijk (8;6)). Anderzijds is volgens (8;8)

$$P[A_1 \wedge C_2] = P[C_2]\,P[A_1] = P[C_2]\{P[B_1] + P[C_1]\} =$$

$$P[B_1]\,P[C_2] + P[C_1]\,P[C_2],$$

wat, tezamen met de vorige relatie, leidt tot

$$P[C_1 \wedge C_2] = P[C_1]\,P[C_2]$$

dus tot (8;5).

Stelling 8;3

Zijn $\underline{x}_1$ en $\underline{x}_2$ continu en onafhankelijk verdeeld, dan geldt

(8;9) $$f(x_1, x_2) = f_1(x_1)\,f_2(x_2),$$

waarin f de simultane en f_1 en f_2 de marginale verdelingsdichtheden voorstellen.

Bewijs

Deze stelling volgt onmiddelijk uit definitie 7;7:

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1\,\partial x_2} = \frac{\partial^2\{F_1(x_1)F_2(x_2)\}}{\partial x_1\,\partial x_2} = f_1(x_1)\,f_2(x_2).$$

Opmerkingen

In hoofdstuk 4 (blz. 23) is opgemerkt, dat de bijzondere productregel (4;16) niet voldoende is om stochastische onafhankelijkheid te garanderen. Daartoe was (4;15) nodig. In definitie 8;1 ligt echter een analoge relatie als (4;15) opgesloten. Immers vult men in (8;1) bijv. in $x_n = \infty$, dan gaat deze relatie over in dezelfde relatie, maar nu voor de eerste $n-1$ stochastische grootheden. Wij hebben hier dus aan (8;1) genoeg voor de definitie van stochastische onafhankelijkheid. Anderzijds is paarsgewijze onafhankelijkheid niet voldoende voor volledige onafhankelijkheid van meer dan twee stochastische grootheden. Dit is gemakkelijk door een tegenvoorbeeld aan te tonen. Op blz. 24 (hoofdstuk 4) hebben wij een voorbeeld aangetroffen van 3 eventualiteiten A, B en C , die paarsgewijze onafhankelijk waren, maar waarvoor

$$P[A \wedge B \wedge C] \neq P[A]\,P[B]\,P[C]$$

was. Definiëren wij nu $\underline{x}$, $\underline{y}$ resp $\underline{z}$ als grootheden, die de waarde 1 aannemen, als A , B resp. C optreden en de waarde 0 als A , B , resp. C niet optreden, dan is $P[\underline{x} = 1] = P[A]$, $P[\underline{y} = 0] = P[\bar{B}]$ enz., en onafhankelijkheid van de gebeurtenissen is aequivalent met die van de stochastische grootheden. Deze laatste zijn dus

ook paarsgewijze onafhankelijk, maar het drietal is dat niet, daar

$$P[A \wedge B \wedge C] = P[\underline{x} = 1 \wedge \underline{y} = 1 \wedge \underline{z} = 1] \neq P[A]\,P[B]\,P[C] =$$
$$= P[\underline{x} = 1]\,P[\underline{y} = 1]\,P[\underline{z} = 1].$$

De relaties (6;5) en (8;9) (voor n dimensies opgeschreven) kunnen ook dienen om stochastische onafhankelijkheid te <u>definiëren</u>. Dan kan (8;1) daaruit gemakkelijk afgeleid worden. De hierboven gevolgde methode bezit het voordeel, dat het continue en het discrete geval (en ook het hier niet behandelde gemengde geval) in één universele formule samengevat worden.

Uit de stellingen 8;2 en 8;3 volgt, dat voor onafhankelijke grootheden geldt:

(8;10) $$P[\underline{x}_1 = x_1 \mid \underline{x}_2 = x_2] = P[\underline{x}_1 = x_1]$$ (discontinue geval)

en (volgens (7;38))

(8;11) $$F(x_1 \mid x_2) = F_1(x_1)$$ (continue geval).

Of, algemeen gezegd, <u>bij stochastisch onafhankelijke grootheden is de voorwaardelijke verdeling van een aantal daarvan, onder voorwaarden, die alleen op de overige grootheden betrekking hebben, gelijk aan de marginale verdeling van de eerstgenoemde grootheden</u>.

<u>Voorbeelden</u>

Bij vele experimentele onderzoekingen richt men de experimenten zo in, dat in het wiskundige model met onafhankelijke stochastische grootheden gewerkt kan worden. De wiskundige moeilijkheden worden daardoor sterk verminderd, omdat dan de eenvoudige relaties, die in dit hoofdstuk zijn afgeleid, gelden. Als voorbeeld van de vereenvoudigingen, die optreden, beschouwen wij de tweedimensionale normale verdeling (7;28). De marginale verdelingen van $\underline{x}$ en $\underline{y}$ zijn, volgens (7;35)

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}} \quad \text{en} \quad f_2(y) = \frac{1}{\sigma_2\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(y-\mu_2)^2}{\sigma_2^2}}.$$

Zijn $\underline{x}$ en $\underline{y}$ stochastisch onafhankelijk, dan geldt dus volgens (8;9)

(8.12) $$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2}\, e^{-\frac{1}{2}\left\{\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right\}},$$

hetgeen overeenkomt met (7;28) wanneer daarin $\rho = 0$ gesteld wordt. De simultane verdelingsdichtheid is nu dus veel eenvoudiger dan in het algemene geval: ten eerste is er één parameter minder en ten tweede kunnen allerlei berekeningen

gemakkelijker uitgevoerd worden, omdat $f(x,y)$ uit twee factoren bestaat, die ieder slechts één der beide grootheden bevatten. Voor meer dan 2 grootheden is de vereenvoudiging van dezelfde aard, terwijl het effect bij het oplossen van vele problemen nog sterker is.

Dat $\underline{x}$ en $\underline{y}$, als zij een simultane normale verdeling bezitten, dan en slechts dan onafhankelijk verdeeld zijn als $\rho=0$ is, volgt ook uit de formules (7;35) en (7;40), die de marginale en de voorwaardelijke verdelingen van $\underline{x}$ voorstellen. Deze zijn dan en slechts dan identiek, als $\rho=0$ is.

Een tweede voorbeeld van stochastisch onafhankelijke grootheden verkrijgt men door de resultaten te beschouwen van n op dezelfde wijze uitgevoerde worpen met een dobbelsteen. In dat geval kan men de worp-resultaten aangeven met $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en het ligt voor de hand, daar een dobbelsteen geen geheugen heeft, als modelveronderstelling in te voeren, dat deze grootheden stochastisch onafhankelijk zijn en alle dezelfde (marginale) verdeling bezitten. Wij merken hierbij terloops op, dat men een dergelijke onderstelling niet klakkeloos behoeft te aanvaarden, doch experimenteel kan onderzoeken. De statistiek beschikt dan over de middelen om op grond van waarnemingsresultaten de gemaakte onderstelling op zijn juistheid te toetsen.

Indien wij voor de eenvoud veronderstellen, dat de dobbelsteen zuiver is, en indien $x_1, x_2, \ldots, x_n$ een willekeurig stel mogelijke uitkomsten is, geldt volgens (8;5)

$$(8;13)\quad P[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2 \wedge \ldots \wedge \underline{x}_n = x_n] = \prod_{i=1}^{n} P[\underline{x}_i = x_i] = 6^{-n}$$

en deze kans is dus voor alle mogelijke n-tallen uitkomsten dezelfde. Op grond van deze uitkomst kan dan desgewenst voor verdere berekeningen gebruik gemaakt worden van de kansdefinitie van LAPLACE, die op blz. 15 vermeld is. Hiervan is in feite in hoofdstuk 4 (blz. 17) reeds impliciet gebruik gemaakt.

Als derde voorbeeld beschouwen wij een reeks worpen met een munt. Worden deze telkens op dezelfde wijze uitgevoerd, onafhankelijk van de bij vorige worpen verkregen resultaten, dan kan men de worpresultaten als stochastisch onafhankelijk beschouwen, met eenzelfde kans p op kruis voor iedere worp. Stellen nu $\underline{x}_1, \underline{x}_2,$ enz. de nummers voor van de worpen, waarbij voor de eerste maal, voor de tweede maal, enz. kruis verkregen wordt, dan geldt uiteraard

$$(8;14)\qquad P[\underline{x}_1 < \underline{x}_2 < \ldots] = 1,$$

daar de tussen de vierkante haken genoteerde eventualiteit zeker is. Deze stochastische grootheden zijn dus stochastisch afhankelijk, daar de waarden, die bijv. $\underline{x}_2$ aan kan nemen, afhangen van de waarde, die $\underline{x}_1$ aanneemt (en andersom). Beschouwt men echter de grootheden:

(8;15) $\qquad \underline{y}_1 = \underline{x}_1 \;,\; \underline{y}_2 = \underline{x}_2 - \underline{x}_1 \;,\; \underline{y}_3 = \underline{x}_3 - \underline{x}_2 \;,$ enz.,

dan zijn deze grootheden weer onafhankelijk en zij bezitten alle dezelfde negatief binomiale verdeling (formule (7;7), hier met $k=1$). Dit volgt ogenblikkelijk uit de overweging, dat na iedere worp, waarbij kruis verkregen is, een "nieuwe" rij van onafhankelijke worpen "begint", waarop formule (7;7) van toepassing is. Bij verschillende berekeningen omtrent de kansverdeling van $\underline{x}_k$ (met $k>1$) is het dan ook eenvoudiger om van de $\underline{y}$'s uit te gaan. Wij zullen daar verderop gebruik van maken.

Tenslotte een strikvraag[1], die vaak niet juist wordt beantwoord. In fig. 8;1 is een 2-dimensionale discrete verdeling getekend, waarbij de kansen op de drie mogelijke uitkomsten nog onbepaald zijn gelaten.

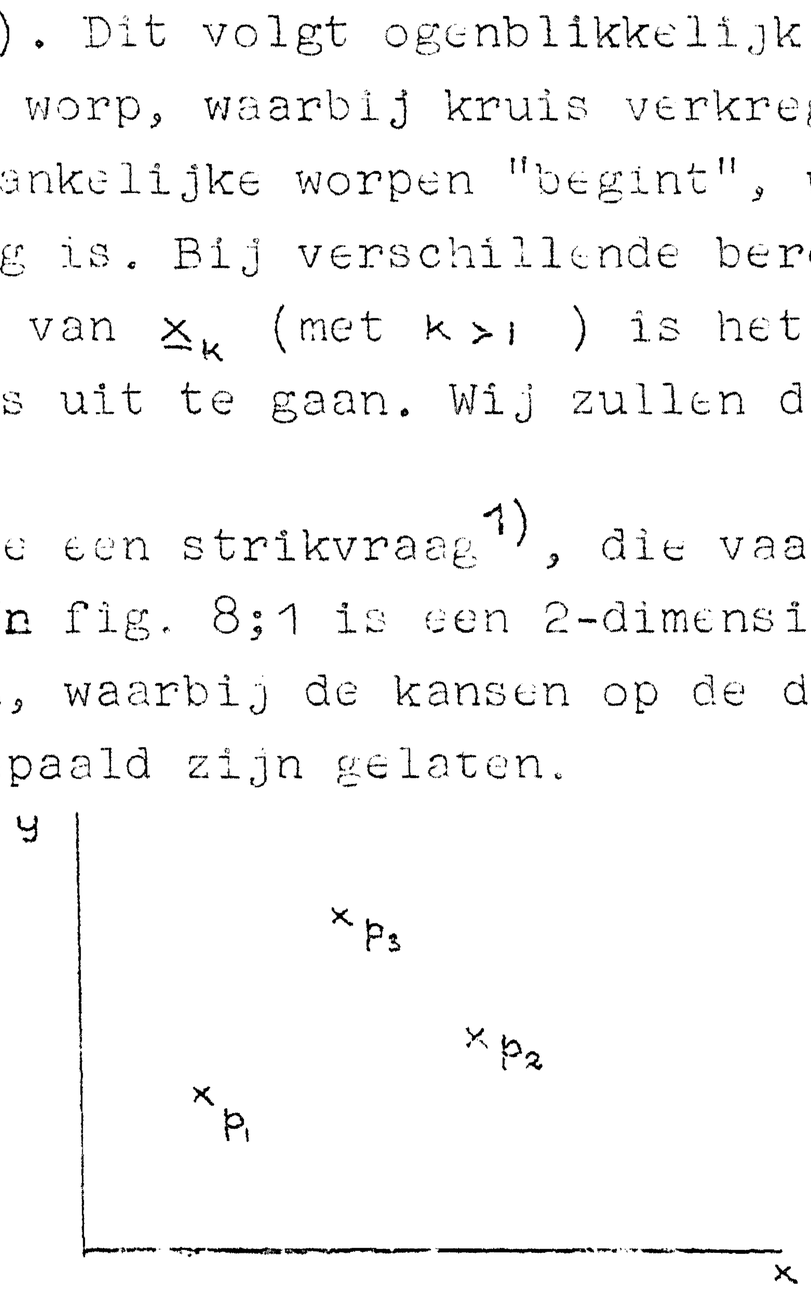


Fig. 8;1. Discrete tweedimensionale verdeling

Gevraagd wordt of het mogelijk is de kansen p_1, p_2 en p_3 zo te bepalen, dat de grootheden $\underline{x}$ en $\underline{y}$ stochastisch onafhankelijk zijn.

Opgave[2]

Twee stochastische grootheden $\underline{x}$ en $\underline{y}$, die een simultane verdeling bezitten, zijn onderling onafhankelijk verdeeld. Beide kunnen uitsluitend de waarden 1,2 en 3 aannemen. De volgende kansen zijn gegeven:

---

1) Dit is opgave 75 van J. HEMELRIJK en DORALINE WABEKE, Elementaire statistische opgaven met uitgewerkte oplossingen, Noorduijn en zoon N.V., Gorinchem 1957.

2) Opgave 81 l.c.

x	y	$P[\underline{x} = x \wedge \underline{y} = y]$
1	3	1/30
2	1	1/10
2	2	3/10
.	2	1/5

Er zijn twee kansverdelingen, waarbij deze kansen optreden.
Bereken deze.

Hoofdstuk 9. Mathematische verwachting en momenten

Het begrip mathematische verwachting, dat in dit hoofdstuk wordt besproken, is zowel voor de theorie als voor de toepassingen van groot belang. De momenten van een kansverdeling, die o.a. gebruikt kunnen worden om een globale beschrijving van ligging en vorm daarvan te geven, zijn speciale gevallen.

Wij behandelen eerst het ééndimensionale geval.

Definitie 9;1

Is $\underline{x}$ een discreet verdeelde stochastische grootheid, die de waarden $x_1, x_2, \ldots$ aan kan nemen, met

$$p_i = P[\underline{x} = x_i] \qquad (\textstyle\sum_i p_i = 1) \tag{9;1}$$

en is $\varphi(x)$ een functie van x, dan wordt de mathematische verwachting van $\varphi(\underline{x})$ - notatie: $\mathcal{E}\varphi(\underline{x})$ - gedefinieerd door

$$\mathcal{E}\varphi(\underline{x}) \stackrel{def}{=} \sum_i p_i \varphi(x_i). \tag{9;2}$$

Bezit $\underline{x}$ een continue verdeling met verdelingsdichtheid $f(x)$, dan luidt de definitie

$$\mathcal{E}\varphi(\underline{x}) \stackrel{def}{=} \int_{-\infty}^{+\infty} \varphi(x) f(x)\, dx. \tag{9;3}$$

Opmerking

De functie $\varphi(\underline{x})$ van $\underline{x}$ is, daar $\underline{x}$ stochastisch is, zelf in het algemeen ook een stochastische grootheid. De verwachting $\mathcal{E}$ [1] is een "operator", die deze stochastische grootheid in een getal omzet. Immers de rechterleden van (9;2) en (9;3) zijn getallen.

Voor gemengde kansverdelingen (vergelijk bldz. 39) is $\mathcal{E}\varphi(\underline{x})$ een lineaire combinatie, met coëfficiënten λ en $(1-\lambda)$ van de rechterleden van (9;2) en (9;3).

Definitie 9;2

Het k^{e} moment ($k = 0, 1, 2, \ldots$) van een stochastische grootheid $\underline{x}$ is de verwachting van $\underline{x}^k$.

Notatie: μ_k of, indien nodig, $\mu_k(\underline{x})$. Dus

$$\mu_k \stackrel{def}{=} \mathcal{E}\underline{x}^k. \tag{9;4}$$

---

1) Ter verkorting wordt de toevoeging "mathematische" vaak weggelaten.

Opmerkingen

Voor iedere kansverdeling geldt: $\mu_0 = 1$. Stellen wij de kansverdeling van $\underline{x}$ aanschouwelijk voor als een massaverdeling over de $\underline{x}$-as, in het discrete geval met massapunten van massa p_i in de punten x_i en in het continue geval als een continu aangebrachte massa met dichtheid $F(x)$, dan stelt μ_0 de totale massa voor. Het eerste moment μ_1 is de abscis van het zwaartepunt van de massaverdeling en μ_2 is het traagheidsmoment om de oorsprong. De beschrijvende waarde van deze momenten voor de kansrekening is dezelfde als die in de mechanica. Aan deze analogie is het gebruik van de term "momenten" toe te schrijven.

Bij μ_1 wordt de index 1 vaak weggelaten.

Voorbeelden

Indien $\underline{x}$ kansen p resp. q ($p + q = 1$) bezit op de waarden 1 resp. 0 (dichotomie), dan geldt:

(9;5) $$\mu_k = p \quad \text{voor} \quad k = 1, 2, \ldots .$$

Immers dan is

$$\mu_k = p \cdot 1^k + q \cdot 0^k = p .$$

Indien $\underline{x}$ een negatief binomiale verdeling bezit met $k = 1$ (d.i. als $\underline{x}$ het aantal experimenten is tot en met het eerste succes; vergelijk voorbeeld 8 van hoofdstuk 5), dan is

$$P[\underline{x} = x] = p q^{x-1} \qquad (x = 1, 2, \ldots)$$

en dus

$$\mathcal{E}\underline{x} = \sum_{x=1}^{\infty} x \, P[\underline{x} = x] = \sum_{x=1}^{\infty} x \, p q^{x-1} = p \sum_{x=1}^{\infty} x q^{x-1} .$$

Nu is

$$\sum_{x=0}^{\infty} q^x = \frac{1}{1-q}$$

en nemen wij links en rechts de afgeleide, dan komt er

$$\sum_{x=1}^{\infty} x q^{x-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2} .$$

Derhalve is

(9;6) $$\mu = \mathcal{E}\underline{x} = \frac{1}{p} .$$

Op analoge wijze valt af te leiden:

(9;7) $$\mu_2 = \mathcal{E}\underline{x}^2 = \frac{1+q}{p^2} .$$ [1]

---

1) Vergelijk voor volledige afleiding opgave 27 uit het opgavenboekje (zie voetnoot op blz. 52).

Is $\underline{x}$ <u>homogeen verdeeld</u> (vergelijk pag. 39), dan is

$$\mathcal{E}\underline{x} = \int_{\theta_1}^{\theta_2} \frac{x\,dx}{\theta_2 - \theta_1} = \tfrac{1}{2}(\theta_1 + \theta_2)$$

en

$$\mathcal{E}\underline{x}^2 = \int_{\theta_1}^{\theta_2} \frac{x^2 dx}{\theta_2 - \theta_1} = \tfrac{1}{3}(\theta_1^2 + \theta_1\theta_2 + \theta_2^2).$$

Dus

$$(9;8) \qquad \mu = \tfrac{1}{2}(\theta_1 + \theta_2) \quad , \quad \mu_2 = \tfrac{1}{3}(\theta_1^2 + \theta_1\theta_2 + \theta_2^2).$$

Is x <u>normaal verdeeld</u> volgens formule (7;13), dan is

$$(9;9) \qquad \mathcal{E}\underline{x} = \mu \qquad \text{en} \qquad \mathcal{E}\underline{x}^2 = \sigma^2 + \mu^2,$$

waarin μ en σ de parameters uit formule (7;13) voorstellen. Voor het bewijs hiervan vergelijke men par. 1 van de appendix.

<u>Definitie 9;3</u>

Het <u>k^{de} gereduceerde moment</u> ($k = 1, 2, \ldots$) van een stochastische grootheid $\underline{x}$ is de verwachting van $(\underline{x} - \mu)^k$.
<u>Notatie</u>: $\tilde{\mu}_k$ of, indien nodig $\tilde{\mu}_k(\underline{x})$. Verder $\tilde{\underline{x}} = \underline{x} - \mu$, de <u>gereduceerde</u> $\underline{x}$. Dus:

$$(9;10) \qquad \tilde{\mu}_k = \mu_k(\tilde{\underline{x}}) = \mathcal{E}\tilde{\underline{x}}^k = \mathcal{E}(\underline{x} - \mu)^k.$$

<u>Opmerkingen</u>

Het eerste gereduceerde moment is altijd 0. Het tweede gereduceerde moment wordt de <u>variantie</u> van $\underline{x}$ genoemd en met σ^2, zo nodig $\sigma^2(\underline{x})$, aangegeven:

$$(9;11) \qquad \sigma^2(\underline{x}) = \mathcal{E}(\underline{x} - \mu)^2 = \mathcal{E}\tilde{\underline{x}}^2 = \tilde{\mu}_2.$$

De variantie komt overeen met het traagheidsmoment om het zwaartepunt. Voor iedere kansverdeling is $\sigma^2 \geqq 0$ en alleen dan 0 als $\underline{x}$ één waarde met kans 1 aanneemt. De wortel uit de variantie wordt de <u>spreiding</u> of <u>standaardafwijking</u> van $\underline{x}$ genoemd.

Het derde gereduceerde moment is 0 voor symmetrische verdelingen, maar het omgekeerde behoeft niet juist te zijn.

De berekening van de gereduceerde momenten, waarvan vooral de variantie belangrijk is, geschiedt het snelst met behulp van verderop af te leiden stellingen over de mathematische verwachting.

Het geval van meerdimensionale verdelingen behandelen wij voor het gemak in hoofdzaak voor 2 dimensies. Voor meer dan 2 verloopt het geheel analoog.

Definitie 9;4

Bezitten $\underline{x}$ en $\underline{y}$ een simultane discrete verdeling, waarbij het punt $\underline{P} \equiv (\underline{x}, \underline{y})$ met kans p_j $(j = 1, 2, \ldots)$ in het punt $P_j \equiv (x_j, y_j)$ terecht komt en is $\varphi(x, y)$ een functie van x en y, dan is de mathematische verwachting van $\varphi(\underline{x}, \underline{y})$:

$$(9;12) \qquad \mathcal{E}\,\varphi(\underline{x}, \underline{y}) \stackrel{def}{=} \sum_j p_j\, \varphi(x_j, y_j).$$

Is de verdeling continu, met verdelingsdichtheid $f(x, y)$ dan luidt de definitie

$$(9;13) \qquad \mathcal{E}\,\varphi(\underline{x}, \underline{y}) \stackrel{def}{=} \int_{-\infty}^{+\infty}\int \varphi(x, y)\, f(x, y)\, dx\, dy.$$

Definitie 9;5

De voorwaardelijke verwachting van $\varphi(\underline{x}, \underline{y})$, onder de voorwaarde $\underline{y} = y$, is de verwachting van $\varphi(\underline{x}, y)$ met betrekking tot de voorwaardelijke verdeling van $\underline{x}$ onder dezelfde voorwaarde.
Notatie: $\mathcal{E}(\varphi(\underline{x}, \underline{y}) \mid \underline{y} = y)$ of $\mathcal{E}(\varphi(\underline{x}, \underline{y}) \mid y)$, of kortweg: $\mathcal{E}\,\varphi(\underline{x}, y)$.

Dus, volgens (7;38), voor het continue geval:

$$(9;14) \qquad \mathcal{E}\,\varphi(\underline{x}, y) = \int_{-\infty}^{+\infty} \varphi(x, y)\, \frac{f(x, y)}{f_2(y)}\, dx.$$

Voorbeeld

Volgens (7;40) is de voorwaardelijke verdeling van $\underline{x}$, onder de voorwaarde $\underline{y} = y$, als $\underline{x}$ en $\underline{y}$ een simultane normale verdeling (7;28) bezitten, ook normaal. Uit (7;40) en (9;9) volgt nu direct:

$$(9;15) \qquad \mathcal{E}(\underline{x} \mid y) = \mu_1 + \rho\, \frac{\sigma_1}{\sigma_2}\, (y - \mu_2).$$

Definitie 9;6

De marginale verwachting van $\varphi(\underline{x}, \underline{y})$ met betrekking tot $\underline{x}$, is de verwachting van $\varphi(\underline{x}, \underline{y})$ genomen over de marginale verdeling van $\underline{x}$.
Notatie: $\mathcal{E}_{\underline{x}}\,\varphi(\underline{x}, \underline{y})$.

Opmerking

$\mathcal{E}\,\varphi(\underline{x},y)$ is een functie van y en $\mathcal{E}_{\underline{x}}\,\varphi(\underline{x},\underline{y})$ is een functie van $\underline{y}$. De laatstgenoemde is derhalve een stochastische grootheid. De eerstgenoemde gaat in een stochastische grootheid over, indien men daarin y weer als zodanig beschouwt. De definities en de hieronder gegeven stellingen gelden natuurlijk ook onder verwisseling van x en y .

Stelling 9;1

Is φ een functie van x alleen: $\varphi(x)$, dan is de verwachting van $\varphi(\underline{x})$ gelijk aan de marginale verwachting.

Bewijs

Wij geven het bewijs alleen voor het continue geval; voor het discrete verloopt het analoog.

Volgens (9;13) is

$$\mathcal{E}\,\varphi(\underline{x}) = \int_{-\infty}^{+\infty}\int \varphi(x)\,f(x,y)\,dx\,dy =$$

$$= \int_{-\infty}^{+\infty}\varphi(x)\,dx \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_{-\infty}^{+\infty}\varphi(x)\,f_1(x)\,dx,$$

waarin $f_1(x)$ (vergelijk (7;33)) de marginale verdelingsdichtheid van $\underline{x}$ is.

Voorbeeld

Voor de tweedimensionale normale verdeling, die gegeven wordt door (7;28), is de marginale verdeling van $\underline{x}$ volgens (7;35) een normale, met parameters μ_1 en σ_1 . Volgens (9;9) en stelling 9;1 is dus voor deze verdeling

(9;16) $$\mathcal{E}\,\underline{x} = \mu_1 \quad , \quad \mathcal{E}\,\underline{x}^2 = \sigma_1^2 + \mu_1^2$$

en analoog voor $\underline{y}$.

Stelling 9;2

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan geldt voor de verwachting van een product $\varphi(\underline{x})\,\psi(\underline{y})$ van functies van x en y afzonderlijk:

(9;17) $$\mathcal{E}\,\varphi(\underline{x})\,\psi(\underline{y}) = \mathcal{E}\,\varphi(\underline{x})\;\mathcal{E}\,\psi(\underline{y}).$$

Bewijs (voor het continue geval)

Met behulp van (8;9) en stelling 9;1 verkrijgen wij:

$$\mathcal{E}\,\varphi(\underline{x})\,\psi(\underline{y}) = \int_{-\infty}^{+\infty}\int \varphi(x)\,\psi(y)\,f(x,y)\,dx\,dy =$$

$$= \int_{-\infty}^{+\infty} \varphi(x)\,f_1(x)\,dx \int_{-\infty}^{+\infty} \psi(y)\,f_2(y)\,dy = \mathcal{E}\,\varphi(\underline{x})\,\mathcal{E}\,\psi(\underline{y}).$$

Voorbeeld

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan is

$$\mathcal{E}\,\underline{x}\,\underline{y} = \mathcal{E}\,\underline{x}\;\mathcal{E}\,\underline{y}. \tag{9;18}$$

Opmerking

Het omgekeerde geldt niet, zoals door een tegenvoorbeeld gemakkelijk aan te tonen is. Beschouwen wij bijv. de verdelingsdichtheid (7;29), een homogene verdeling op een cirkelschijf, dan is gemakkelijk in te zien, dat

$$\mathcal{E}\,\underline{x}\,\underline{y} = \mathcal{E}\,\underline{x}\,\mathcal{E}\,\underline{y} = 0$$

is, zodat (9;18) geldt. Dit volgt direct uit de symmetrie van de verdeling ten opzichte van de assen. Toch zijn $\underline{x}$ en $\underline{y}$ stochastisch afhankelijk (waarom?).

Stelling 9;3

De operator $\mathcal{E}$ is een lineaire operator, d.w.z.

$$\mathcal{E}\{a\,\varphi(\underline{x},\underline{y}) + b\,\psi(\underline{x},\underline{y}) + c\} = a\,\mathcal{E}\,\varphi(\underline{x},\underline{y}) + b\,\mathcal{E}\,\psi(\underline{x},\underline{y}) + c, \tag{9;19}$$

als $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten, φ en ψ functies en a, b en c constanten zijn. De stelling geldt analoog voor meer dan 2 stochastische grootheden.

Bewijs

De geldigheid van (9;19) volgt onmiddelijk uit definitie 9;1.

Toepassing

Deze stelling maakt het mogelijk de gereduceerde momenten in de gewone uit te drukken. Bijvoorbeeld (vergelijk (9;11)):

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

$$= \mathcal{E}\,\underline{x}^2 - 2\mu\,\mathcal{E}\,\underline{x} + \mu^2 = \mu_2 - 2\mu^2 + \mu^2 = \mu_2 - \mu^2,$$

hetgeen dus leidt tot de volgende zeer nuttige formule ter berekening van de variantie

$$(9;20) \qquad \sigma^2 = \mu_2 - \mu^2 \qquad \text{of} \qquad \sigma^2(\underline{x}) = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2.$$

Opmerking

Uit (9;20) volgt direct de ongelijkheid

$$(9;21) \qquad \mu_2 \geqq \mu^2 \qquad \text{of} \qquad \mathcal{E}\underline{x}^2 \geqq (\mathcal{E}\underline{x})^2,$$

waarbij de gelijkheidstekens alleen gelden als $\underline{x}$ met kans 1 een bepaalde waarde aanneemt.

Stelling 9;4. Optelregel voor varianties

Zijn $\underline{x}$ en $\underline{y}$ stochastisch onafhankelijk verdeeld[1)], dan geldt

$$(9;22) \qquad \sigma^2(\underline{x} + \underline{y}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y}).$$

De stelling geldt ook voor meer dan twee grootheden.

Bewijs

Volgens (9;19) is

$$\mathcal{E}(\underline{x} + \underline{y}) = \mathcal{E}\underline{x} + \mathcal{E}\underline{y},$$

dus

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

Verder is, eveneens volgens (9;19)

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

daar, wegens de onafhankelijkheid (9;18) toegepast kan worden. Volgens (9;20) is

$$\sigma^2(\underline{x} + \underline{y}) = \mathcal{E}(\underline{x} + \underline{y})^2 - \{\mathcal{E}(\underline{x} + \underline{y})\}^2$$

en substitueren wij de bovenstaande resultaten hierin, dan komt er

$$\sigma^2(\underline{x} + \underline{y}) = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 + \mathcal{E}\underline{y}^2 - (\mathcal{E}\underline{y})^2 = \sigma^2(\underline{x}) + \sigma^2(\underline{y})$$

weer volgens (9;20)

Opmerking

De voorwaarde van onafhankelijkheid is voldoende, doch niet noodzakelijk. Nodig en voldoende is blijkens het bewijs, dat (9;18) geldt. Deze stelling geeft tevens een belangrijk verschil aan tussen verwachtingen en varianties. Voor optelbaarheid van verwachtingen behoeft (9;18) niet te gelden.

---

1) Deze uitdrukking houdt stilzwijgend in, dat $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten.

Toepassing

De stellingen 9;3 en 9;4 stellen ons in staat op eenvoudige wijze de verwachting en de variantie van de binomiale en de negatief-binomiale verdeling te berekenen evenals de verwachting van de hypergeometrische.

Volgens blz. 41 onderaan kan een binomiaal verdeelde grootheid $\underline{x}$ (vergelijk (7;5)) beschouwd worden als de som van n onafhankelijk verdeelde grootheden $\underline{x}_i$, die ieder een kans p resp. q bezitten op de waarden 1 resp. 0.

Volgens (9;5) geldt voor ieder daarvan

$$\mathcal{E}\underline{x}_i = p \quad \text{en} \quad \mathcal{E}\underline{x}_i^2 = p,$$

dus (vergelijk (9;20))

$$\sigma^2(\underline{x}_i) = p - p^2 = pq.$$

Voor

$$\underline{x} = \sum_{i=1}^{n} \underline{x}_i$$

geldt dus volgens stelling 9;3

(9;23) $$\mu = \mathcal{E}\underline{x} = np$$

en volgens (9;22)

(9;24) $$\sigma^2 = \sigma^2(\underline{x}) = npq.$$ [1)]

Voor de negatief-binomiale verdeling (formule (7;7)) verloopt de berekening analoog. Stelt $\underline{x}$ het aantal experimenten voor tot en met het k^{de} succes, dan kan $\underline{x}$ beschouwd worden als de som van k onafhankelijke grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ waarvan $\underline{x}_1$ het aantal experimenten voorstelt tot en met het eerste succes, $\underline{x}_2$ het daarop volgende aantal experimenten tot en met het tweede succes, enz. Daar alle experimenten onafhankelijk zijn, geldt dit ook voor de grootheden $\underline{x}_i$ $(i = 1, 2, \ldots, k)$ en deze bezitten bovendien alle dezelfde verdeling als $\underline{x}_1$. De momenten daarvan worden gegeven door (9;6) en (9;7), dus

$$\mathcal{E}\underline{x}_i = \frac{1}{p} \quad \text{en} \quad \mathcal{E}\underline{x}_i^2 = \frac{1+q}{p^2},$$

---

1) Deze variantie wordt veelal berekend door $\mathcal{E}\underline{x}(\underline{x}-1)$ direct uit de definiërende formule $\mathcal{E}\underline{x}(\underline{x}-1) = \sum_{x=0}^{n} x(x-1)p^x q^{n-x}\binom{n}{x}$ te berekenen en vervolgens hieruit $\mathcal{E}\underline{x}^2$ af te leiden. Deze berekeningswijze is ingewikkelder dan de hier gevolgde.

waaruit volgens (9;20) volgt

$$\sigma^2(\underline{x}_i) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Derhalve leiden stelling 9;3 en 9;4 tot

(9;25) $$\mu = \mathcal{E}\underline{x} = \frac{k}{p} \quad \text{en} \quad \sigma^2 = \sigma^2(\underline{x}) = \frac{kq}{p^2}.$$

Voor de <u>hypergeometrische verdeling</u> (vergelijk (7;6)) kan deze redenering slechts gedeeltelijk gevolgd worden. Volgens voorbeeld 7 van hoofdstuk 5 wordt deze verdeling verkregen als die van het aantal witte knikkers bij aselecte trekking zonder teruglegging van n knikkers uit een doos met M witte en N zwarte knikkers. Noemen wij nu het aantal witte knikkers bij de ie trekking $\underline{x}_i$ (met mogelijke waarden 0 en 1), dan geldt weer

$$\underline{x} = \sum_{i=1}^{n} \underline{x}_i$$

en volgens voorbeeld 5 (blz. 29) bezitten alle $\underline{x}_i$ dezelfde (marginale) verdeling, met een kans $\frac{M}{M+N}$ op de waarde 1. Zij zijn nu echter niet stochastisch onafhankelijk (waarom niet?), zodat (9;22) niet toegepast kan worden. Wel echter stelling 9;3 en deze geeft, daar $\mathcal{E}\underline{x}_i = \frac{M}{M+N}$ is

(9;26) $$\mu = \mathcal{E}\underline{x} = \frac{Mn}{M+N}.$$

Voor de <u>normale verdeling</u> volgt uit (9;9) en (9;22)

(9;27) $$\sigma^2(\underline{x}) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Verwachting en variantie van de normale verdeling zijn dus juist de parameters μ en σ^2, die in formule (7;13) voorkomen. Een normale verdeling wordt dus door verwachting en variantie volledig bepaald.

<u>Stelling 9;5</u>

Zijn a en b constanten en is $\underline{x}$ een stochastische grootheid, dan is

(9;28) $$\sigma^2(a\underline{x} + b) = a^2\sigma^2(\underline{x}).$$

Bewijs

$$\mathcal{E}(a\underline{x}+b) = a\,\mathcal{E}\underline{x} + b.$$

De bij $a\underline{x}+b$ behorende gereduceerde grootheid is dus

$$a\underline{x}+b - a\,\mathcal{E}\underline{x} - b = a(\underline{x} - \mathcal{E}\underline{x}) = a\tilde{\underline{x}}.$$

Dus

$$\sigma^2(a\underline{x}+b) = \mathcal{E}(a\tilde{\underline{x}})^2 = \mathcal{E}\,a^2\tilde{\underline{x}}^2 = a^2\,\mathcal{E}\,\tilde{\underline{x}}^2 = a^2\sigma^2(\underline{x}).$$

De belangrijkste toepassing hiervan is vervat in:

Stelling 9;6

Is $\underline{x}$ een stochastische grootheid met verwachting μ en variantie σ^2, dan bezit de stochastische grootheid

(9;29)
$$\underline{x}^* \overset{\text{def}}{=\!=} \frac{\underline{x}-\mu}{\sigma} = \frac{\tilde{\underline{x}}}{\sigma}$$

verwachting 0 en variantie 1.

Het bewijs wordt aan de lezer overgelaten.

Opmerkingen

De grootheid $\underline{x}^*$ wordt de bij $\underline{x}$ behorende gestandaardiseerde grootheid of kortweg de gestandaardiseerde $\underline{x}$ genoemd (vergelijk ook blz. 39).

Dit standaardiseringsproces stelt ons o.a. in staat om iedere normale verdeling te beheersen met behulp van een tabel van de gestandaardiseerde normale verdeling.

Voorbeeld

$\underline{x}$ is $N(4;0,5)$-verdeeld; dus normaal met verwachting 4 en spreiding 0,5. Gevraagd wordt te berekenen

$$P[\underline{x} \geqq 4,75].$$

De berekening verloopt als volgt:

$$P[\underline{x} \geqq 4,75] = P\left[\frac{\underline{x}-\mu}{\sigma} \geqq \frac{4,75-\mu}{\sigma}\right] = P[\underline{u} \geqq 1,5],$$

waarin $\underline{u}$ de gestandaardiseerde normaal verdeelde grootheid voorstelt. Uit de tabel van de normale verdeling (tabel 1) blijkt, dat de gezochte kans 0,067 bedraagt. Verdere opgaven van deze aard zijn in het opgavenboekje(zie voetnoot op blz. 52) te vinden.

Stelling 9;7

Indien $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten en $\varphi(x,y)$ is een functie van x en y, terwijl

(9;30)
$$g(y) \overset{\text{def}}{=\!=} \mathcal{E}(\varphi(\underline{x},\underline{y}) \mid \underline{y} = y)$$

de voorwaardelijke verwachting van $\varphi(\underline{x},\underline{y})$ onder de voorwaarde $\underline{y}=y$ is, dan geldt

(9;31) $$\mathcal{E}\,\varphi(\underline{x},\underline{y}) = \mathcal{E}\,g(\underline{y}).$$

Bewijs (voor het continue geval)

Volgens definitie 9;4 en 9;5 en stelling 9;1 geldt

$$\mathcal{E}\,\varphi(\underline{x},\underline{y}) = \int_{-\infty}^{+\infty}\int \varphi(x,y)\,f(x,y)\,dx\,dy =$$

$$= \int_{-\infty}^{+\infty} f_2(y)\,dy \int_{-\infty}^{+\infty} \varphi(x,y)\,\frac{f(x,y)}{f_2(y)}\,dx =$$

$$= \int_{-\infty}^{+\infty} g(y)\,f_2(y)\,dy = \mathcal{E}\,g(\underline{y}).$$

Toepassingen

Met behulp van deze stelling kunnen wij nu de variantie van een hypergeometrische verdeling berekenen.[1] Laat daartoe $\underline{x}_i$ resp. $\underline{x}_j$ de aantallen witte knikkers bij de i^e resp. j^e trekking zonder teruglegging voorstellen (beide nemen dus één van de waarden 0 of 1 aan), dan berekenen wij eerst $\mathcal{E}\,\underline{x}_i\,\underline{x}_j$. Op analoge wijze als bij voorbeeld 5 van hoofdstuk 5 valt in te zien, dat de verdeling van $\underline{x}_i$, bij gegeven $\underline{x}_j$, voor alle $i \neq j$ dezelfde is en wel:

Als $\underline{x}_j = 1$ is, is $P[\underline{x}_i = 1] = \frac{M-1}{M+N-1}$,

en als $\underline{x}_j = 0$ is, is $P[\underline{x}_i = 1] = \frac{M}{M+N-1}$.

Derhalve geldt:

$$\mathcal{E}(\underline{x}_i \mid \underline{x}_j = 1) = \frac{M-1}{M+N-1} \quad ; \quad \mathcal{E}(\underline{x}_i \mid \underline{x}_j = 0) = \frac{M}{M+N-1}.$$

Dus is

$$\mathcal{E}(\underline{x}_i\,\underline{x}_j \mid \underline{x}_j = 1) = \mathcal{E}(\underline{x}_i \mid \underline{x}_j = 1) = \frac{M-1}{M+N-1}$$

en

$$\mathcal{E}(\underline{x}_i\,\underline{x}_j \mid \underline{x}_j = 0) = \mathcal{E}(0 \mid \underline{x}_j = 0) = 0.$$

Volgens stelling 9;7 is dus

(9;32) $$\mathcal{E}\,\underline{x}_i\,\underline{x}_j = \frac{M-1}{M+N-1}\,P[\underline{x}_j = 1] + 0 = \frac{M-1}{M+N-1}\cdot\frac{M}{M+N}.$$

---

1) Ook deze variantie kan, evenals die van de binomiale verdeling, berekend worden door de definitie 9;1 toe te passen op $\mathcal{E}\,\underline{x}(\underline{x}-1)$.

Voor $\underline{x} = \sum_{i=1}^{n} \underline{x}_i$ geldt nu

$$\mathcal{E}\,\underline{x}^2 = \mathcal{E}\left(\sum_{i=1}^{n} \underline{x}_i\right)^2 = \mathcal{E}\sum_{i=1}^{n} \underline{x}_i^2 + 2\,\mathcal{E}\sum_{i<j} \underline{x}_i \underline{x}_j ,$$

maar hierin is volgens (9;5) voor iedere i

$$\mathcal{E}\,\underline{x}_i = \mathcal{E}\,\underline{x}_i^2 = \frac{M}{M+N}$$

en volgens (9;32) voor ieder paar i, j met $i \neq j$

$$\mathcal{E}\,\underline{x}_i \underline{x}_j = \frac{M(M-1)}{(M+N)(M+N-1)} .$$

Dus is

$$\mathcal{E}\,\underline{x}^2 = \frac{nM}{M+N} + \frac{n(n-1)M(M-1)}{(M+N)(M+N-1)} .$$

Toepassing van (9;20), tezamen met (9;26) geeft dan, na enige herleiding

(9;33) $$\sigma^2(\underline{x}) = \frac{nMN(M+N-n)}{(M+N)^2(M+N-1)} .$$

Een grootheid, die evenals verwachting en variantie van groot belang is, is de <u>covariantie</u> van twee stochastische grootheden $\underline{x}$ en $\underline{y}$. Deze wordt geschreven als $\mathrm{cov}(\underline{x}, \underline{y})$ en gedefiniëerd door

(9;34) $$\mathrm{cov}(\underline{x}, \underline{y}) \overset{\text{def}}{=\!=} \mathcal{E}\,\tilde{\underline{x}}\,\tilde{\underline{y}} = \mathcal{E}(\underline{x} - \mathcal{E}\underline{x})(\underline{y} - \mathcal{E}\underline{y}).$$

<u>Stelling 9;8</u>

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan is

(9;35) $$\mathrm{cov}(\underline{x}, \underline{y}) = 0.$$

<u>Bewijs</u>

Men houde in het oog, dat $\mathcal{E}\underline{x}$ en $\mathcal{E}\underline{y}$ constanten zijn. Volgens stelling 9;3 is

$$\mathcal{E}(\underline{x} - \mathcal{E}\underline{x})(\underline{y} - \mathcal{E}\underline{y}) = \mathcal{E}(\underline{x}\underline{y} - \underline{x}\,\mathcal{E}\underline{y} - \underline{y}\,\mathcal{E}\underline{x} - \mathcal{E}\underline{x}\,\mathcal{E}\underline{y}) =$$

$$\mathcal{E}\,\underline{x}\underline{y} - \mathcal{E}\underline{x}\,\mathcal{E}\underline{y} - \mathcal{E}\underline{y}\,\mathcal{E}\underline{x} + \mathcal{E}\underline{x}\,\mathcal{E}\underline{y} = \mathcal{E}\,\underline{x}\underline{y} - \mathcal{E}\,\underline{x}\,\mathcal{E}\,\underline{y} = 0$$

volgens (9;18).

Wij berekenen nu, met behulp van stelling 9;7 de covariantie van $\underline{x}$ en $\underline{y}$, voor het geval van de tweedimensionale normale verdeling (formule (7;28)).

Volgens (9;15) is

$$\mathcal{E}(\underline{x}-\mu_1|y)=\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),$$

dus

$$g(y)=\mathcal{E}\{(\underline{x}-\mu_1)(\underline{y}-\mu_2)|\,y\}=\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)^2.$$

Volgens (9;31) is dus

$$\mathcal{E}(\underline{x}-\mu_1)(\underline{y}-\mu_2)=\mathcal{E}\,\rho\frac{\sigma_1}{\sigma_2}(\underline{y}-\mu_2)^2=\rho\frac{\sigma_1}{\sigma_2}\mathcal{E}(\underline{y}-\mu_2)^2=$$
$$=\rho\frac{\sigma_1}{\sigma_2}\sigma_2^2=\rho\sigma_1\sigma_2.$$

Dus

(9;36) $$\operatorname{cov}(\underline{x},\underline{y})=\rho\sigma_1\sigma_2.$$

Stelling 9;9

Bezitten $\underline{x}$ en $\underline{y}$ een simultane verdeling, dan geldt

(9;37) $$\sigma^2(a\underline{x}+b\underline{y})=a^2\sigma^2(\underline{x})+b^2\sigma^2(\underline{y})+2ab\operatorname{cov}(\underline{\tilde{x}},\underline{\tilde{y}}).$$

Bewijs

De bij $a\underline{x}+b\underline{y}$ behorende gereduceerde grootheid is $a\underline{\tilde{x}}+b\underline{\tilde{y}}$, daar $\mathcal{E}(a\underline{x}+b\underline{y})=a\,\mathcal{E}\underline{x}+b\,\mathcal{E}\underline{y}$ is.

Derhalve is

$$\sigma^2(a\underline{x}+b\underline{y})=\mathcal{E}(a\underline{\tilde{x}}+b\underline{\tilde{y}})^2=a^2\mathcal{E}\underline{\tilde{x}}^2+b^2\mathcal{E}\underline{\tilde{y}}^2+2ab\,\mathcal{E}\underline{\tilde{x}}\underline{\tilde{y}}.$$

Stelling 9;10

Bezitten $\underline{x}$ en $\underline{y}$ een simultane verdeling, dan geldt

(9;38) $$|\operatorname{cov}(\underline{x},\underline{y})|\leqq\sigma(\underline{x})\sigma(\underline{y}).$$

Bewijs

Stel in (9;37) $b=1$. Daar een variantie steeds $\geqq 0$ is, geldt

$$0\leqq\sigma^2(a\underline{x}+\underline{y})=\sigma^2(\underline{x})a^2+2\operatorname{cov}(\underline{x},\underline{y})a+\sigma^2(\underline{y}).$$

Dit geldt voor iedere a , dus moet de discriminant van deze kwadratische vorm in a niet positief zijn:

$$\tfrac{1}{4}D=\{\operatorname{cov}(\underline{x},\underline{y})\}^2-\sigma^2(\underline{x})\sigma^2(\underline{y})\leqq 0.$$

Hieruit volgt (9;38).

De <u>correlatiecoëfficiënt</u> van $\underline{x}$ en $\underline{y}$ wordt gedefiniëerd door

$$\rho(\underline{x},\underline{y}) \stackrel{\text{def}}{=} \frac{\text{cov}(\underline{x},\underline{y})}{\sigma(\underline{x})\sigma(\underline{y})} . \tag{9;39}$$

Volgens (9;38) geldt

$$-1 \leqq \rho(\underline{x},\underline{y}) \leqq +1 \tag{9;40}$$

en volgens (9;36) is $\rho(\underline{x},\underline{y})$ voor een tweedimensionale normale verdeling gelijk aan de parameter ρ uit formule (7;28).

<u>Opgave</u>

Twee stochastische grootheden $\underline{x}$ en $\underline{y}$ bezitten een simultane kansverdeling met $|\rho(\underline{x},\underline{y})| = 1$. Dit betekent, dat het stochastische punt $(\underline{x},\underline{y})$ met kans 1 op een bepaalde rechte lijn ligt, en omgekeerd. Bewijs dit.(Dit is opgave 88 van het opgavenboekje. Zie voor de oplossing blz. 131-132 daarvan).

Hoofdstuk 10. De ongelijkheid van BIENAYMÉ-TCHEBYCHEFF en de theoretische wet van de grote aantallen

Een vooral voor theoretische doeleinden belangrijke ongelijkheid is door BIENAYMÉ en door TCHEBYCHEFF (onafhankelijk van elkaar) gevonden.

Stelling 10;1

Voor een stochastische grootheid $\underline{x}$ met verwachting μ en spreiding σ geldt voor iedere $k>0$

$$(10;1) \qquad P[\,|\underline{x}^*| \geqq k\,] \leqq \frac{1}{k^2} \;,$$

waarin

$$(10;2) \qquad \underline{x}^* = \frac{\underline{x}-\mu}{\sigma}$$

de gestandaardiseerde $\underline{x}$ voorstelt.

Bewijs

Wij geven het bewijs voor een discreet verdeelde $\underline{x}$; voor een continue verdeling verloopt het analoog. Per definitie is

$$\sigma^2 = \sum_i p_i (x_i-\mu)^2,$$

als x_i ($i = 1,2,\ldots$) de waarden voorstellen, die $\underline{x}$ aan kan nemen. Dus

$$1 = \sum_i p_i \frac{(x_i-\mu)^2}{\sigma^2} = \sum_i p_i x_i^{*2}.$$

Deze som splitsen wij in twee deelsommen Σ_1 en Σ_2 , waarbij

Σ_1 loopt over die i , waarvoor $|x_i^*| \geqq k$

en Σ_2 " " " i , " $|x_i^*| < k$.

Dan is uiteraard

$$1 \geqq \Sigma_1 p_i x_i^{*2} \geqq k^2 \Sigma_1 p_i.$$

Hierin is echter

$$\Sigma_1 p_i = P[\,|\underline{x}^*| \geqq k\,],$$

waaruit (10;2) direct volgt.

Voorbeelden

Laat $\underline{x}$ de waarden 1 en 0 met kansen p en $q = 1-p$ aannemen, zodat $\mu = p$ en $\sigma^2 = pq$ is. Dan geldt volgens (10;1)

$$P[|\underline{x} - p| \geqq k\sqrt{pq}] \leqq \frac{1}{k^2}.$$

Vullen wij $p = \frac{1}{2}$ en $k = 1$ in, dan komt er

$$P[|\underline{x} - \tfrac{1}{2}| \geqq \tfrac{1}{2}] \leqq 1,$$

hetgeen in dit geval in een gelijkheid overgaat, daar de kans in het linkerlid = 1 is. In dit speciale geval is de ongelijkheid dus scherp: verkleining van het rechterlid is niet mogelijk zonder tot een onjuist resultaat te komen.

In de regel is de ongelijkheid echter zeer grof. Bezit $\underline{x}$ bijv. een $N(0,1)$ -verdeling, dan geldt volgens (10;1)

$$P[|\underline{x}| \geqq 2] \leqq \frac{1}{4},$$

terwijl volgens tabel 1 de kans in het linkerlid gelijk is aan 0,046.

Voor theoretische doeleinden is dit gebrek aan scherpte veelal van weinig belang, zoals uit de hieronder vermelde stellingen blijkt.

Stelling 10;2

Als $\underline{x}_1, \underline{x}_2, \ldots$ onafhankelijk verdeelde stochastische grootheden zijn, die alle dezelfde verwachting μ en variantie σ^2 bezitten, dan convergeert, voor $n \to \infty$, het gemiddelde

$$(10;3) \qquad \overline{\underline{x}}_n \overset{\text{def}}{=\!=} \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i$$

stochastisch naar μ, d.w.z. dat voor iedere $\varepsilon > 0$ geldt

$$(10;4) \qquad \lim_{n \to \infty} P[|\overline{\underline{x}}_n - \mu| > \varepsilon] = 0. \quad ^{1)}$$

Bewijs

Volgens stelling 9;3 is

$$\mathcal{E} \sum_{i=1}^{n} \underline{x}_i = n\mu,$$

dus

$$(10;5) \qquad \mathcal{E}\, \overline{\underline{x}}_n = \mu.$$

---

1) Men noemt dit ook wel: convergentie in waarschijnlijkheid.

Volgens stelling 9;4 is

$$\sigma^2\left(\sum_{i=1}^{n} \underline{x}_i\right) = n\sigma^2$$

en dus volgens stelling 9;5

(10;6) $$\sigma^2(\bar{\underline{x}}_n) = \frac{\sigma^2}{n}.$$

Invullen in (10;1) geeft

$$P\left[\left|\frac{\bar{\underline{x}}_n - \mu}{\sigma}\sqrt{n}\right| \geq k\right] = P\left[\left|\bar{\underline{x}}_n - \mu\right| \geq k\frac{\sigma}{\sqrt{n}}\right] \leq \frac{1}{k^2}.$$

Dit geldt voor iedere k, zodat wij dus mogen stellen

$$\varepsilon = k\frac{\sigma}{\sqrt{n}} \quad \text{of} \quad k = \frac{\varepsilon\sqrt{n}}{\sigma}.$$

Dit geeft

$$P\left[\left|\bar{\underline{x}}_n - \mu\right| \geq \varepsilon\right] \leq \frac{\sigma^2}{n\varepsilon^2}$$

en de limiet, voor $n \to \infty$, van het rechterlid is 0.

Een speciaal geval van deze stelling wordt verkregen door voor de $\underline{x}_i$ het aantal successen (0 of 1) te nemen bij een experiment met kans p op succes. De stelling neemt dan de volgende vorm aan.

Stelling 10;3. (Theoretische wet der grote aantallen)

In een reeks van n onafhankelijke experimenten, ieder met kans p op succes, convergeert de fractie successen $\underline{x}/n$ voor $n \to \infty$ in waarschijnlijkheid naar p, d.w.z. voor iedere $\varepsilon > 0$ is

(10;7) $$\lim_{n\to\infty} P\left[\left|\underline{x}/n - p\right| \geq \varepsilon\right] = 0.$$

Opmerkingen

Deze stelling heeft een grote practische betekenis. Het is de theoretische tegenhanger van de experimentele wet van de grote aantallen (zie hoofdstuk 2). Het zou niet juist zijn te zeggen, dat deze laatste nu "bewezen" is, daar een stelling, die op de werkelijkheid betrekking heeft, nooit op grond van een axiomastelsel bewezen kan worden. Alleen experimentele bevestiging is daar geschikt voor. Wel kunnen wij nu echter opmerken, dat de uitermate eenvoudige axioma's, waarvan wij zijn uitgegaan, adaequaat blijken te zijn ter beschrijving van de nogal ingewikkelde experimentele wet van de grote aantallen, die wij in hoofdstuk 2 fundamenteel voor statistische verschijnselen hebben genoemd. Dit betekent, dat de interpretatie van een kans p als - bij benadering - een f_q in een lange reeks

waarnemingen bevestigd is, terwijl bovendien dit succes van de theorie het vertrouwen in verdere resultaten - d.w.z. het vertrouwen dat de toepassingen daarvan in de practijk bruikbaar zullen blijken te zijn - vergroot. De op blz. 19 gemaakte opmerking, "dat in het kansmodel het f_q als zodanig, met zijn statistische fluctuaties, op natuurlijke wijze weer ingevoerd kan worden", is hiermede waar gemaakt.

De practische betekenis van stelling 10;2 is de volgende. Indien men van een stochastische grootheid $\underline{x}$ met verwachting μ en variantie σ^2, een aantal onafhankelijke waarnemingen verricht, kan de i^e waarneming voorgesteld worden door $\underline{x}_i$ en is de stelling toepasbaar, daar alle $\underline{x}_i$ nu dezelfde verdeling (nl. dezelfde verdeling als $\underline{x}$) bezitten . Formule (10;3) betekent nu, dat het rekenkundige gemiddelde der waarnemingen, naarmate n groter wordt, steeds minder kans bezit om ver van μ verwijderd te zijn. Als men n maar groot genoeg maakt, kan men - behoudens een willekeurig kleine kans - zorgen willekeurig dicht bij μ terecht te komen.

Indien men bijv. een physische of chemische constante (gewicht, soortelijk gewicht, titer van een vloeistof, lading van een electron, enz.) nauwkeurig wenst te bepalen, kan men een aantal onafhankelijke waarnemingen daarvan verrichten. Ieder daarvan is aan meetfouten onderhevig en kan daarom als een stochastische grootheid beschouwd worden. Worden de waarnemingen op dezelfde wijze (door dezelfde personen en met dezelfde apparatuur) verricht, dan bezitten ze alle dezelfde kansverdeling. Worden bovendien de uitkomsten van latere waarnemingen niet beïnvloed door die van de voorafgaande - gevaar voor psychologische beïnvloeding is vaak in sterke mate aanwezig -, dan kunnen zij als stochastisch onafhankelijk beschouwd worden. In deze omstandigheden is de stelling toepasbaar. Veelal zal men n dan niet zeer groot kunnen (of willen) maken, maar ook bij kleine n doet de invloed van formule (10;6) zich reeds voor: de variabiliteit van een gemiddelde van een aantal onafhankelijke waarnemingen is aanzienlijk kleiner, dus de nauwkeurigheid aanzienlijk groter dan van één enkele waarneming.

De grootheid μ, waarbij $\overline{\underline{x}}_n$ in de buurt komt, hangt vaak mede van de waarnemingsapparatuur af. Men dient zich daarvan bij het verrichten van metingen steeds bewust te zijn (hetgeen niet altijd het geval is). Bij de bepaling van een titer van een vloeistof bijv. kunnen afwijkingen in de gebruikte buret

en pipet systematische verschillen veroorzaken tussen μ en de titer, die men eigenlijk wenst te meten. Het is dan zinloos de nauwkeurigheid van de bepaling, door n op te voeren, zeer groot te maken, daar de systematische fout dan gaat overheersen. Men kan daaraan tegemoet komen door de apparatuur te wisselen. Zolang de spreiding van de met verschillende apparaturen verkregen gemiddelden $\bar{x}_n$ nog groot zijn in vergelijking met de systematische verschillen tussen de bij de apparaturen behorende μ's, is vergroting van het aantal waarnemingen bij ieder der apparaturen in principe zinvol. Zijn de systematische verschillen groot, dan is groter nauwkeurigheid alleen te bereiken door verbetering van de apparatuur.

Laat men verschillende waarnemers werken met dezelfde of verschillende apparatuur, maar zijn er geen (of nauwelijks) systematische verschillen, zodat μ voor alle waarnemingen gelijk is, dan is het van belang op te merken, dat stelling 10;2 ook blijft gelden als de varianties ongelijk zijn, mits deze begrensd zijn. Is $\sigma^2(x_i) \leqq \sigma_0^2$ voor alle i, dan is nl. het bewijs, met een kleine wijziging, aan te passen. De voorwaarden kunnen trouwens nog verder verzwakt worden, doch dit punt zullen wij hier buiten beschouwing laten.

De uitwerking van de hierboven geschetste gedachtengang behoort tot de verderop te behandelen schattingstheorie.

Hoofdstuk 11. Functies van stochastische grootheden; centrale limietstelling

In de vorige hoofdstukken is reeds opgemerkt, dat een functie van één of meer stochastische grootheden zelf weer een stochastische grootheid is. Eén der belangrijkste voorbeelden hiervan is het gemiddelde van een aantal stochastische grootheden, waarvan in hoofdstuk 10 sprake is geweest.

Wij gaan nu na, hoe in het algemeen de kansverdeling van een dergelijke functie berekend kan worden.

Stelling 11;1

Laat $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een simultane verdeling bezitten met $f(x_1, x_2, \ldots, x_n)$ als kansdichtheid en laat $\varphi(x_1, x_2, \ldots, x_n)$ een functie van $x_1, x_2, \ldots, x_n$ zijn. Laat verder, in de n-dimensionale ruimte R_n met $x_1, x_2, \ldots, x_n$ als coördinaten, G_a het gebied voorstellen van al die punten, waarvoor

$$\varphi(x_1, x_2, \ldots, x_n) \leqq a \tag{11;1}$$

is. Dan is

$$F_{\underline{\varphi}}(a) = P[\underline{\varphi} \leqq a] = P[\varphi(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n) \leqq a] = P[\underline{P} \in G_a] = \int\cdots\int_{G_a} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n. \tag{11;2}$$

Hierin stelt $\underline{P}$ het stochastische punt $(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n)$ voor. In geval van een discrete verdeling wordt in het laatste lid van (11;2) de meervoudige integraal door een som vervangen. De stelling volgt uit (7;27).

Een toepassing van de stelling is de volgende.

Stelling 11;2

Als $\underline{x}$ een $N(\mu,\sigma)$-verdeelde grootheid is, is

$$\underline{u} = \frac{\underline{x} - \mu}{\sigma} \tag{11;3}$$

$N(0,1)$-verdeeld.

Bewijs

Het gebied G_a, waarin $u \leqq a$ is, wordt gegeven door $\frac{x-\mu}{\sigma} \leqq a$ of $x \leqq \mu + a\sigma$. Of, direct opgeschreven,

$$F_{\underline{u}}(u) \overset{\text{def}}{=} P[\underline{u} \leqq u] = P[\underline{x} \leqq \mu + u\sigma] = F_{\underline{x}}(\mu + u\sigma).$$

De verdelingsdichtheid $f_{\underline{u}}(u)$ wordt hieruit verkregen door differentiëren:

$$f_{\underline{u}}(u) = \frac{d}{du} F_{\underline{u}}(u) = \frac{d}{du} F_{\underline{x}}(\mu + u\sigma) = \sigma f_{\underline{x}}(\mu + u\sigma).$$

Vullen wij echter in $f_{\underline{x}}(x)$ voor x in: $\mu + u\sigma$, dan komt er weer een kwadratische e-macht, dus weer een normale verdelingsdichtheid. Volgens stelling 9;6 heeft $\underline{u}$ bovendien verwachting 0 en spreiding 1, waarmede het bewijs geleverd is.

Het onder stelling 9;6 gegeven voorbeeld is een toepassing van stelling 11;2 die daar reeds (zonder vermelding) is gebruikt.

Numerieke voorbeelden van toepassing van stelling 11;1 zijn vervat in de volgende twee opgaven, die betrekking hebben op discrete verdelingen.

Opgave 11:1

Twee stochastische grootheden $\underline{x}$ en $\underline{y}$ zijn beide binomiaal verdeeld; $\underline{x}$ met parameters $n = 3$ en $p = \frac{1}{3}$, $\underline{y}$ met parameters $n = 4$ en $p = \frac{1}{4}$. Verder is gegeven, dat $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld zijn.

Bepaal de kansverdeling van $\underline{z} = \underline{x} + \underline{y}$ en bereken verwachting en spreiding van deze verdeling.

(Opgave 72 van het Opgavenboekje, zie voor de oplossing blz. 109-111 daarvan.)

Opgave 11;2

Een tentamen bestaat uit 5 vragen, die goed of fout beantwoord kunnen worden. Voor ieder goed antwoord wordt 1 punt gegeven, voor een fout antwoord 0.

a) Candidaat A heeft zich zodanig op het tentamen voorbereid, dat hij bij iedere vraag een kans $\frac{3}{5}$ heeft op het geven van het goede antwoord. Bereken, in de onderstelling van onafhankelijkheid der antwoorden, de kansverdeling van zijn eindcijfer en de verwachting en spreiding daarvan.

b) Doe hetzelfde voor candidaat B, die voor de eerste 3 vragen dezelfde kans op een goed antwoord heeft als A, maar voor de laatste twee slechts een kans $\frac{1}{2}$.
(Opgave 73 van het Opgavenboekje, zie voor de oplossing blz. 111-113 daarvan.)

Een belangrijke toepassing van stelling 11;1 is de volgende.

Stelling 11;3

Zijn $\underline{x}$ en $\underline{y}$ continu en onafhankelijk verdeelde stochastische grootheden met verdelingsfunctie $F(x)$ en $G(y)$ en verdelingsdichtheden $f(x)$ en $g(y)$, dan geldt voor verdelingsfunctie $H(z)$ en

verdelingsdichtheid $h(z)$ van $\underline{z} = \underline{x} + \underline{y}$

$$(11;4) \qquad H(z) = \int_{-\infty}^{+\infty} F(x)\, G(z-x)\, dx$$

en

$$(11;5) \qquad h(z) = \int_{-\infty}^{+\infty} F(x)\, g(z-x)\, dx.$$

Bewijs

Volgens (11;2) en stelling 8;3 is

$$H(z) = \iint_{x+y \leqq z} f(x)\, g(y)\, dx\, dy = \int_{-\infty}^{+\infty} f(x)\, dx \int_{-\infty}^{z-x} g(y)\, dy = \int_{-\infty}^{+\infty} f(x)\, G(z-x)\, dx.$$

Hieruit wordt $h(z)$ verkregen door naar z te differentiëren, hetgeen onder de integraal mag geschieden. Daarbij is dan echter x als een constante op te vatten, zodat (11;5) direct volgt.

Een toepassing hiervan is de volgende.

Stelling 11;4

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk normaal verdeeld, dan is $\underline{z} = \underline{x} + \underline{y}$ eveneens normaal verdeeld.

Bewijs

De verdelingsdichtheden $f(x)$ en $g(y)$ zijn nu van de vorm

$$f(x) :: e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \qquad \text{en} \qquad g(y) :: e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2},$$

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

Het product $f(x)\, g(z-x)$ heeft dus de vorm

$$f(x)\, g(z-x) :: e^{-\frac{1}{2}\left\{\left(\frac{x}{\sigma_1}\right)^2 - \frac{2\mu_1 x}{\sigma_1^2} + \left(\frac{z-x}{\sigma_2}\right)^2 - \frac{2\mu_2(z-x)}{\sigma_2^2}\right\}}.$$

De exponent kan (vergelijk blz. 44) gesplitst worden in twee delen, waarvan het ene een kwadratische vorm in z is (zonder x) en het tweede een volkomen kwadraat van de vorm $(\alpha x + \beta z + \gamma)^2$. Het eerste deel kan als een e-macht voor de integraal van (11;5) gebracht worden en deze integraal zelf is een constante (substitueer daarin nl. $v = \alpha x + \beta z + \gamma$, dan veranderen de grenzen niet; deze blijven $-\infty$ en $+\infty$). Daaruit volgt echter de stelling

Opgave 11;3

a) Geef de formule voor h(z) uit stelling 11;4.

b) Bereken de correlatiecoëfficiënt van de grootheden $\underline{x}$ en $\underline{z}$ uit stelling 11;4.

c) Geef de formule voor de simultane verdelingsdichtheid van $\underline{x}$ en $\underline{z}$.

Oplossing van b)

De correlatiecoëfficiënt wordt gedefinieerd door (9;39). Zijn $\tilde{\underline{x}}$ en $\tilde{\underline{y}}$ de gereduceerde van $\underline{x}$ en $\underline{y}$, dus

$$\tilde{\underline{x}} = \underline{x} - \mu_1 \quad \text{en} \quad \tilde{\underline{y}} = \underline{y} - \mu_2$$

en is $\tilde{\underline{z}}$ de gereduceerde $\underline{z}$, dan is volgens (9;19)

$$\tilde{\underline{z}} = \underline{z} - \mathcal{E}\,\underline{z} = \underline{z} - \mathcal{E}(\underline{x}+\underline{y}) = \underline{z} - \mu_1 - \mu_2 = \tilde{\underline{x}} + \tilde{\underline{y}}.$$

Derhalve is (vergelijk blz. 65)

$$\operatorname{cov}(\underline{x}, \underline{z}) = \mathcal{E}\,\tilde{\underline{x}}\,\tilde{\underline{z}} = \mathcal{E}\,\tilde{\underline{x}}(\tilde{\underline{x}}+\tilde{\underline{y}}) = \mathcal{E}\,\tilde{\underline{x}}^2 = \sigma_1^2,$$

daar $\mathcal{E}\,\tilde{\underline{x}}\,\tilde{\underline{y}} = 0$ wegens de onafhankelijkheid van $\underline{x}$ en $\underline{y}$.

Dus is, met gebruik maken van stelling 9;4:

$$\rho(\underline{x}, \underline{z}) = \frac{\operatorname{cov}(\underline{x}, \underline{z})}{\sigma(\underline{x})\,\sigma(\underline{y})} = \frac{\sigma_1^2}{\sigma_1\sqrt{\sigma_1^2+\sigma_2^2}} = \sqrt{\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}}\,.$$

$\underline{x}$ en $\underline{z}$ zijn dus niet onafhankelijk verdeeld.

Opmerkingen

Stelling 11;4 is direct (door inductie) uit te breiden tot meer dan 2 onafhankelijke normaal verdeelde grootheden.

De stelling geldt ook als deze grootheden niet stochastisch onafhankelijk zijn, doch een simultane normale verdeling bezitten. Het bewijs verloopt op geheel analoge wijze met behulp van de formules (7;28) en (11;2), waarin voor G_a het gebied $x+y \leqq z$ genomen wordt.

Tevens is gemakkelijk in te zien, dat vermenigvuldiging van de stochastische grootheden met constante factoren de normaliteit van de simultane verdeling niet verstoort (vergelijk stelling 11;2), evenmin als het optellen van constante termen, zodat de volgende algemene stelling geldt.

Stelling 11;5

Bezitten $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een simultane normale verdeling, dan is ook iedere lineaire combinatie

$$\underline{z} = a_0 + a_1\underline{x}_1 + \ldots + a_n\underline{x}_n \tag{11;6}$$

normaal verdeeld.

Opgave 11;4

Een leverancier van flessen slaolie vermeldt als netto inhoud van zijn flessen: 350 gram olie. Een winkelier wenst deze bewering te controleren zonder de flessen te openen. Hij weegt daartoe een zeer groot aantal gevulde flessen en hij vindt voor dit gewicht een normale verdeling met gemiddelde 585,2 gram en spreiding 12,8 gram. Vervolgens weegt hij een groot aantal lege flessen die hij van zijn klanten teruggekregen heeft, met de bijbehorende sluitcapsules. Voor dit gewicht vindt hij eveneens een normale verdeling met gemiddelde 228,3 gram en spreiding 11,3 gram.

Bereken, in de veronderstelling, dat het gewicht van de olie in de flessen normaal verdeeld is en onafhankelijk is van het gewicht van fles + sluiting:

a) de kansverdeling van het gewicht aan olie in de fles,
b) het percentage der flessen olie, die minder dan 300 gram olie bevatten,
c) de covariantie van het gewicht van een gevulde fles en dat van de olie, die erin zit.
d) Indien men nu, door zorgvuldiger vullen van de flessen, de spreiding van fles + inhoud, bij gelijkblijvend gemiddelde, zoveel kleiner wil maken, dat slechts 2,5% der flessen minder dan 350 gram olie bevat, tot welk bedrag zal men dan de spreiding der gevulde flessen moeten verminderen?

(Opgave 67 uit het Opgavenboekje. Zie voor de oplossing blz. 106-107 daarvan.)

Opmerking

Deze opgave geeft een typerend beeld van enigszins primitieve toepassingen van de statistiek, zoals deze in de practijk vaak voorkomen. Het aantal metingen wordt zo groot verondersteld, dat het verschil tussen de uit deze metingen gevonden gemiddelden en de daardoor geschatte verwachtingen (eigenlijk: gemiddelden over "alle" flessen olie van die fabrikant) te verwaarlozen is. Hetzelfde geldt voor de spreidingen. De normaliteit van de kansverdelingen is een onderstelling, die onderzocht wordt door de vorm van de frequentieverdeling van de metingen na te gaan. Vertoont deze een "redelijke" gelijkenis met een normale verdelingsdichtheid, dan wordt de normaliteit aanvaard. De uitkomsten van dit soort berekeningen dienen steeds als slechts bij benadering juist te worden beschouwd. De nauwkeurigheid van dergelijke benaderingen zullen wij later leren berekenen.

Voor sommen van niet-normaal verdeelde stochastische grootheden geldt de volgende limiet-stelling.

Stelling 11;6 (Centrale limietstelling voor onderling onafhankelijke stochastische grootheden)

Laat $\underline{x}_1, \underline{x}_2, \ldots$ een rij stochastisch onafhankelijke grootheden voorstellen. Laat μ_i en σ_i verwachting en spreiding van $\underline{x}_i$ $(i = 1,2,\ldots)$ zijn en laat verder

$$\delta_i^3 \stackrel{\text{def}}{=\!=} \mathcal{E}|\underline{x}_i - \mu_i|^3 \qquad (11;7)$$ 1)

en

$$(11;8) \qquad \begin{cases} \underline{x}(n) \stackrel{\text{def}}{=\!=} \sum_{i=1}^{n} \underline{x}_i, \\ \mu(n) \stackrel{\text{def}}{=\!=} \sum_{i=1}^{n} \mu_i, \\ \sigma^2(n) \stackrel{\text{def}}{=\!=} \sum_{i=1}^{n} \sigma_i^2, \\ \delta^3(n) \stackrel{\text{def}}{=\!=} \sum_{i=1}^{n} \delta_i^3. \end{cases}$$

Indien nu de voorwaarde

$$(11;9) \qquad \lim_{n\to\infty} \frac{\delta(n)}{\sigma(n)} = 0$$

vervuld is, dan is de grootheid

$$(11;10) \qquad \underline{x}(n)^* \stackrel{\text{def}}{=\!=} \frac{\underline{x}(n) - \mu(n)}{\sigma(n)}$$

voor $n \to \infty$ asymptotisch $N(0,1)$ verdeeld,[2] d.w.z. dat voor iedere a geldt:

$$(11;11) \qquad \lim_{n\to\infty} P[\underline{x}(n)^* \leq a] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}u^2} du.$$

Opmerkingen

1. De grootheid $\underline{x}(n)^*$ is de gestandaardiseerde van $\underline{x}(n)$, want deze heeft $\mu(n)$ als verwachting en $\sigma(n)$ als spreiding.
2. Het bewijs van deze stelling valt buiten het kader van deze cursus. Het kan bijv. gevonden worden in H. CRAMÉR, "Mathematical methods of statistics", Princeton Univ. Press, Princeton 1946, p. 215 e.v.

---

1) Dit is het z.g. absolute derde moment.
2) In plaats van deze exacte wijze van uitdrukken zullen wij ook vaak de volgende gebruiken: "$\underline{x}(n)$ is asymptotisch normaal verdeeld" (waarbij dus nog geen standaardisering toegepast is).

Behalve in de hier gegeven vorm bestaan er nog verschillende andere vormen van de centrale limietstelling; de verschillen hebben dan betrekking op de voorwaarden, waaronder de asymptotische normaliteit geldt.

Als $\underline{x}_1$, $\underline{x}_2$, ... alle dezelfde verdeling bezitten, geldt de volgende stelling:

Stelling 11;7 (Centrale limietstelling voor onderling onafhankelijke identiek-verdeelde stochastische grootheden)

Zijn $\underline{x}_1, \underline{x}_2, \ldots$ onderling onafhankelijk verdeeld volgens dezelfde verdeling, dan is

$$(11;12) \qquad \frac{\sum_{i=1}^{n} \underline{x}_i - n\mu}{\sigma\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i - \mu}{\sigma\sqrt{n}},$$

waarin μ en σ verwachting en spreiding van de verdeling der $\underline{x}_i$ voorstellen, voor $n \to \infty$ asymptotisch $N(0,1)$ verdeeld.

Opmerkingen

1. Het bewijs van stelling 11;7 is bijv. te vinden in H. CRAMÉR, "Mathematical methods of statistics", p. 214-215.

2. Voorwaarde (11;9) van stelling 11;6 gaat, voor het speciale geval dat $\underline{x}_1, \underline{x}_2, \ldots$ alle dezelfde verdeling bezitten, over in de voorwaarde

$$(11;9') \qquad \delta^3 \overset{\text{def}}{=} \mathcal{E}|\underline{x}_i - \mu|^3 \qquad \text{is eindig}$$

(Bewijs dit.) Uit stelling 11;7 volgt dat de asymptotische normaliteit in dit speciale geval ook geldt, als niet aan (11;9') voldaan is.

Een belangrijke toepassing van 11;7 is de volgende.

Stelling 11;8 (Stelling van De Moivre en Laplace)

Bezit x een binomiale verdeling met parameters n en p, d.w.z. is voor $x = 0, 1, \ldots, n$

$$(11;13) \qquad P[\underline{x} = x] = \binom{n}{x} p^x q^{n-x} \qquad (q = 1-p),$$

dan is $\underline{x}$ voor $n \to \infty$ asymptotisch normaal verdeeld. Dus voor iedere a geldt

$$(11;14) \qquad \lim_{n\to\infty} P\left[\frac{\underline{x} - np}{\sqrt{npq}} \leq a\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}u^2}\, du.$$

Bewijs

Op blz. 61 hebben wij er reeds gebruik van gemaakt, dat $\underline{x}$ beschouwd kan worden als de som van n onafhankelijk verdeelde stochastische grootheden $\underline{x}_i$, die "het aantal successen bij het i^e experiment" voorstellen. Met behulp daarvan worden daar de formules $\mu(\underline{x}) = np$ en $\sigma^2(\underline{x}) = npq$ afgeleid. Daarmede is de stelling echter teruggebracht tot een speciaal geval van stelling 11;7, met $\mu = \mu(\underline{x}_i) = p$ en $\sigma^2 = \sigma^2(\underline{x}_i) = pq$.

Opmerkingen

1. Stelling 11;8 volgt ook uit stelling 11;6. Voor de grootheden $\underline{x}_i$ geldt nl. dat δ^3 (zie (11;9')) eindig is. (Bewijs dit.)
2. Deze stelling werd door DE MOIVRE en LAPLACE langs geheel andere weg afgeleid. DE MOIVRE gaf de stelling niet voor de verdelingsfunctie, maar bewees de analoge stelling voor ieder der termen apart. Deze krijgt dan de vorm

$$(11;15) \qquad P[\underline{x}=x] \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}\frac{(x-np)^2}{npq}},$$

waarin $\approx$ betekent: "asymptotisch, voor $n \to \infty$, gelijk aan".

De oorspronkelijke publicaties zijn

A. DE MOIVRE, Approximatio ad summam terminorum binomii $(a+b)^n$ in seriem expansi, 1733.

P.S. DE LAPLACE, Théorie analytique des probabilités, 1812.

Op grond van deze stelling kan de normale verdeling gebruikt worden als benadering van de binomiale. Zijn de parameters van de binomiale verdeling n en p, dan is volgens (9;23) en (9;24) de verwachting np en de spreiding $\sqrt{npq}$. De normale verdeling met dezelfde verwachting en spreiding wordt nu de aangepaste normale verdeling genoemd.

In fig. 11;1 is, voor p=0,1 en p=0,36 en voor n=4,9,25,100 zowel de binomiale verdeling als de aangepaste normale verdeling geschetst. De x-schaal is daarbij in ieder der 8 figuren zo gekozen, dat de normale verdeling dezelfde vorm bezit (de spreidingen zijn dus (in cm) gelijk gemaakt; hoe wordt dit bereikt?).

p = 0,1

n = 4

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

σ=0,6

0 0,4 1 2 x

p = 0,36

n = 4

σ=0,96

0 1 2 3 4 x

1,44

n = 9

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

σ=0,9

0 1 2 3 4 x

0,9

n = 9

σ=1,44

0 1 2 3 4 5 6 7 x

3,24

n = 25

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

σ=1,5

0 1 2 3 4 5 6 7 x

2,5

n = 25

σ=2,4

3 5 7 9 11 13 15 x

n = 100

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

σ=3

4 7 10 13 16 x

n = 100

σ=4,8

24 28 32 36 40 44 48 x

Fig. 11;1 Convergentie van de binomiale verdeling naar de normale.

Stelling 11;8 wordt geïllustreerd door het feit, dat de benadering duidelijk beter wordt bij toenemende n. Tevens blijkt uit de figuren, dat bij dezelfde waarde van n de benadering voor p=0,36 beter is dan voor p=0,1. De aanpassing is op zijn best voor $p=\frac{1}{2}$, hetgeen men reeds kan vermoeden op grond van het feit, dat de binomiale verdeling dan symmetrisch is, evenals de normale.

De laatste twee stellingen zijn van groot belang voor het berekenen van bepaalde kansen. Wij geven hiervan twee voorbeelden.

Voorbeeld 11;1

Op blz. 7 hebben wij gezien, dat in 1954 in 6 wijken van Amsterdam tezamen op een totaal van 1927 geboorten 974 jongens voorkwamen. Wij vragen ons nu af, hoe groot de kans op een zo groot of nog groter aantal jongens zou zijn als bij ieder der geboorten de kans op een jongen gelijk is aan $\frac{1}{2}$.

Het aantal jongens, $\underline{x}$, heeft dan een binomiale verdeling met n=1927 en $p=\frac{1}{2}$, dus met

$$\mathcal{E}\underline{x} = np = 963{,}5 \text{ en} \qquad \sigma(\underline{x}) = \sqrt{npq} = \tfrac{1}{2}\sqrt{1927} = 21{,}9.$$

De gevraagde kans is

$$P\left[\underline{x} \geqq 974 \,|\, n = 1927,\ p = \tfrac{1}{2}\right],$$

waarin nu achter de streep geen voorwaarden, maar gegevens staan aangegeven. Wij berekenen deze kans nu - bij benadering - met behulp van (11;14) op de volgende wijze (zonder de gegevens over n en p telkens weer te vermelden).

$$P[\underline{x} \geqq 974] = P\left[\frac{\underline{x}-np}{\sqrt{npq}} \geqq \frac{974-np}{\sqrt{npq}}\right] \approx$$

$$\approx P\left[\underline{u} \geqq \frac{974-963{,}5}{21{,}9}\right] = P[\underline{u} \geqq 0{,}48],$$

waarin $\underline{u}$ een N(0,1) verdeelde grootheid voorstelt en $\approx$ betekent, dat wij met een benadering te maken hebben.

Volgens tabel 1 van de N(0,1) verdeling is de kans in het laatste lid gelijk aan 0,32. Onze uitkomst is dus

$$P\left[\underline{x} \geqq 974 \,|\, n = 1927,\ p = \tfrac{1}{2}\right] \approx 0{,}32.$$

Het belang van kansen van dit type, die overschrijdingskansen genoemd worden, zullen wij later leren kennen.

Voorbeeld 11;2

Van een verpakkingsmachine, die thee in pakjes van nominaal a gram verpakt is uit vroegere waarnemingen bekend dat de werkelijk verpakte hoeveelheid $\underline{x}$ een normale verdeling bezit met sprei-

ding σ gram. De verwachting μ van de verdeling kan men regelen met een instellings-mechanisme.

Volgens de warenwet moet elk pakje minstens het nominale gewicht bevatten. Om de bereiken dat slechts een fractie α (met $\alpha < \frac{1}{2}$) van de pakjes minder dan a gram bevat zal men μ groter dan a moeten nemen.

Gevraagd wordt met hoeveel procent het overwicht μ -a verminderd kan worden indien de controle slechts zou eisen dat het gemiddelde van een asclecte steekproef van n pakjes minstens a gram moet bedragen, terwijl weer een kans α toegelaten wordt dat hieraan niet voldaan is?

Oplossing

Bij verwachting μ geldt

$$P[\underline{x} \leq a] = P\left[\frac{\underline{x}-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right] = P\left[\underline{u} \leq \frac{a-\mu}{\sigma}\right],$$

waarin $\underline{u}$ een $N(0,1)$ verdeelde grootheid voorstelt. Deze kans moet gelijk zijn aan α. Als nu ξ_α gedefiniëerd wordt door

$$P[\underline{u} \geq \xi_\alpha] = \alpha$$

dan moet dus gelden

$$\frac{a-\mu}{\sigma} = -\xi_\alpha \quad \text{of} \quad \mu = a + \xi_\alpha \sigma.$$

Het overwicht moet dus $\xi_\alpha \sigma$ bedragen om ervoor te zorgen dat slechts een fractie α van de pakjes minder dan a gram bevat.

Stelt $\underline{\bar{x}}$ het gemiddelde gewicht van n aselect gekozen pakjes voor dan is (vergelijk (10;5) en (10;6)) $\mathcal{E}\,\underline{\bar{x}} = \mathcal{E}\,\underline{x} = \mu$ en

$$\sigma(\underline{\bar{x}}) = \frac{\sigma(\underline{x})}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}}.$$

Derhalve is

$$P[\underline{\bar{x}} \leq a] = P\left[\underline{u} \leq \frac{a-\mu}{\sigma}\sqrt{n}\right]$$

en deze kans is $= \alpha$ als

$$\frac{a-\mu}{\sigma}\sqrt{n} = -\xi_\alpha \text{ of } \mu = a + \xi_\alpha \frac{\sigma}{\sqrt{n}}$$

Het overwicht moet nu dus $\xi_\alpha \frac{\sigma}{\sqrt{n}}$ bedragen. De besparing in overwicht is dus, in procenten van het oorspronkelijk overwicht

$$100\,\frac{\xi_\alpha \sigma - \xi_\alpha \frac{\sigma}{\sqrt{n}}}{\xi_\alpha \sigma} = 100\left(1 - \frac{1}{\sqrt{n}}\right).$$

Deze uitkomst is dus onafhankelijk van a, σ en α.

Opgave 11;5

Een automatische sorteermachine sorteert eieren op gewicht. Het gewicht van de eieren is normaal verdeeld met verwachting

μ = 42,0 gr en spreiding σ = 5,2 gr. De sorteermachine scheidt de eieren in 5 klassen.

klasse	gewicht in gr
A	$\geqq 55$
B	$45 \leqq B < 55$
C	$35 \leqq C < 45$
D	$25 \leqq D < 35$
E	< 25

Gevraagd: hoeveel procent van de eieren bevat ieder der klassen?

(Opgave 25 van het Opgavenboekje; oplossing op blz. 70-71 daarvan.)

Opgave 11;6

Van een partij assen is de diameter normaal verdeeld met verwachting 14,82 mm en spreiding 0,03 mm. Van een partij boringen is de diameter eveneens normaal verdeeld en wel met verwachting 14,89 mm en spreiding 0,04 mm.

a) Bereken welk percentage assen te groot zal zijn voor de boringen indien as en boring aselect aan elkaar worden toegevoegd.

b) Bereken ook op welke waarde de verwachting van de diameter van de boringen (bij gelijkblijvende spreiding en zonder verandering bij de assen) ingesteld moet worden, opdat slechts 1% van de assen te groot voor de boringen zal zijn, indien wederom as en boring aselect aan elkaar worden toegevoegd.
(Opgave 78 van het Opgavenboekje; oplossing op blz. 117-118 daarvan.)

Opgave 11;7

$\underline{x}$ en $\underline{y}$ zijn onderling onafhankelijk N(0,1) verdeeld. Bereken de kansverdeling van $|\underline{x} - \underline{y}|$.
(Opgave 93 van het Opgavenboekje; oplossing op blz. 137-138 daarvan.)

Opgave 11;8

Bewijs de volgende stelling.

Bezit $\underline{x}$ een negatief binomiale verdeling met parameters k en p, d.w.z. is voor x=k, k+1, ...

11;16) $$P[\underline{x}=x] = \binom{x-1}{k-1} p^k q^{x-k} \qquad (q=1-p),$$

dan is $\underline{x}$ voor $k \to \infty$ asymptotisch normaal verdeeld.

Voor iedere a geldt:

$$(11;17)\qquad \lim_{k\to\infty} P\left[\frac{px-k}{\sqrt{kq}} \leq a\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}u^2}\,du.$$

Aanwijzing:

Vergelijk blz. 61-62 en de centrale limietstelling 11;7.

Opgave 11;9

Hoe groot is bij benadering de kans, dat in een reeks van onafhankelijke experimenten met ieder een kans 0,36 op succes, het 100e succes niet voor het 320e experiment verkregen wordt.

Oplossing

a) Met behulp van (11;17), met p=0,36 en k=100; $\underline{x}$ stelt het aantal experimenten tot en met het 100e succes voor.

$$P[\underline{x} \geqq 320] = P\left[\frac{0{,}36\,\underline{x} - 100}{\sqrt{100 \cdot 0{,}64}} \geqq \frac{0{,}36 \cdot 320 - 100}{\sqrt{100 \cdot 0{,}64}}\right] \approx$$

$$\approx P[\underline{u} \geqq 1{,}9] = 0{,}029.$$

($\underline{u}$ stelt een N(0,1) verdeelde grootheid voor)

b) Met behulp van de binomiale verdeling.
Laat, om verwarring te voorkomen, $\underline{y}$ het aantal successen onder de eerste 319 experimenten voorstellen. Als $y < 100$ is, is het aantal experimenten tot en met het 100e succes $\geqq 320$ en omgekeerd. Dus berekenen wij (nu met behulp van stelling 11;8):

$$P[\underline{y} \leqq 99] = P\left[\frac{\underline{y} - np}{\sqrt{npq}} \leqq \frac{99 - np}{\sqrt{npq}}\right] \approx P\left[\underline{u} \leq \frac{99 - 319 \cdot 0{,}36}{\sqrt{319\ 0{,}36\ 0{,}64}}\right] =$$

$$= P[\underline{u} \leqq -1{,}85] = P[\underline{u} \geqq 1{,}85] = 0{,}032.$$

Opmerking

De beide benaderingen geven niet precies dezelfde uitkomst, hetgeen voor 2 verschillende benaderingsmethoden ook niet nodig is. Zou men de beide berekeningen exact uitvoeren, dan zou de uitkomst wel dezelfde zijn.

Hoofdstuk 12. Schattingstheorie; meest aannemelijke schattingen

De schattingstheorie is een belangrijk onderdeel van de statistiek, dat wij hier wegens de grote uitgebreidheid van het onderwerp slechts oppervlakkig kunnen behandelen. Wij geven eerst een aantal voorbeelden met een intuïtief voor de hand liggende oplossing en behandelen daarna het principe van een algemene theorie, waarop deze oplossingen gebaseerd kunnen worden.

Voorbeeld 12;1. Onderzoek van de eigenschappen van een serieproduct

Eén der belangrijkste eigenschappen van gloeilampen is hun levensduur. Dit geldt niet alleen voor gloeilampen, doch voor zeer vele andere producten, in het bijzonder voor onderdelen van machines en apparaten. De hier te beschrijven statistische bewerkingen komen verder op ieder gebied voor. Wij nemen hier, om de gedachten te bepalen, gloeilampen als voorbeeld. De levensduur van een gloeilamp kan men bepalen door hem te laten branden tot hij doorbrandt. De levensduur is dan bekend, maar de lamp is stuk. Het is daarom duidelijk, dat men in dit geval gedwongen is de eigenschappen van een bepaald soort gloeilampen steekproefsgewijs te onderzoeken, een volledig onderzoek schiet zijn doel voorbij.

De levensduur van verschillende gloeilampen van één soort is verschillend; dit blijkt bij beproeving. Een geschikt beschrijvingsmodel van deze eigenschap is daarom de levensduur als een stochastische grootheid $\underline{x}$ te beschouwen, die voor iedere lamp een bepaalde waarde aanneemt. De bij een steekproef van lampen gevonden waarden vormen dan een reeks van stochastisch onafhankelijke waarnemingen van $\underline{x}$. Veelal wordt bij dergelijke problemen verondersteld, dat $\underline{x}$ een (bij benadering) normale verdeling bezit. Soms geldt dit niet voor $\underline{x}$ zelf, maar wel voor $\log \underline{x}$. Een dergelijke veronderstelling kan, op grond van de waarnemingen, ook statistisch worden onderzocht. Wij zullen deze onderstelling hier hanteren alsof zij reeds onderzocht en bevestigd is.

Is nu $\underline{x}$ normaal verdeeld, dan zijn twee parameters, nl. $\mu = \mathcal{E}\underline{x}$ en $\sigma^2 = \sigma^2(\underline{x})$, voldoende om de verdeling volledig te karakteriseren en het schattingsprobleem is dan deze twee parameters uit een reeks waarnemingen te schatten.

Laat de steekproef van waarnemingen voorgesteld worden door de onafhankelijke grootheden

(12;1) $$\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n,$$

alle met dezelfde verdeling als $\underline{x}$, dan hebben wij (vgl. (10;5)) reeds gezien, dat voor het gemiddelde

(12;2) $$\overline{\underline{x}} = \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i$$

geldt:

(12;3) $$\mathcal{E}\, \overline{\underline{x}} = \mu.$$

Het ligt daarom voor de hand om $\overline{\underline{x}}$, het gemiddelde der waarnemingen, als schatting voor μ te gebruiken.

Definitie 12;1

Een schatting $\underline{t}$ van een onbekende parameter θ wordt zuiver[1)] genoemd als zijn verwachting gelijk is aan deze onbekende parameter

(12;4) $$\mathcal{E}\, \underline{t} = \theta.$$

Definitie 12;2

Een schatting van een onbekende parameter wordt asymptotisch raak[2)] genoemd, als nij voor een onbegrensd stijgend aantal waarnemingen stochastisch convergeert naar die onbekende parameter:

(12;5) $$\lim_{n \to \infty} P[\,|\underline{t} - \theta| > \varepsilon\,] = 0 \qquad \text{voor iedere } \varepsilon > 0.$$

Volgens stelling 10;2 is $\overline{\underline{x}}$ zowel zuiver als asymptotisch raak met betrekking tot μ , ook als $\overline{\underline{x}}$ niet normaal verdeeld is.

Vervolgens gaan wij ons beraden over een schatting voor de variantie σ^2 . Daartoe overwegen wij, dat bij de berekening van een schatting in zekere zin de steekproef van waarnemingen in de plaats wordt gesteld van de gehele onbekende kansverdeling. De waarnemingen van de steekproef zijn verder alle gelijkwaardig. Zijn de aangenomen waarden: $x_1, x_2, \ldots, x_n$, dan kunnen wij - op grond van deze heuristische redenatie - een nieuwe stochastische grootheid $\underline{x}'$ beschouwen, die deze n waarden met gelijke kansen aanneemt:

(12;6) $$P[\,\underline{x}' = x_i\,] = \frac{1}{n} \qquad (i = 1, 2, \ldots, n).$$

---

1) Engels: unbiased.
2) " : consistent.

Voor deze stochastische grootheid geldt

$$(12;7) \qquad \mathcal{E}\,\underline{x}' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

de bij deze steekproef behorende waarde van $\bar{\underline{x}}$, de schatting van μ. Dus: de verwachting van $\underline{x}'$ kan dienen als schatting voor de verwachting van $\underline{x}$. Naar analogie hiervan kunnen wij verwachten, dat de variantie van $\underline{x}'$ ons een schatting voor de variantie van $\underline{x}$ zal geven.

Volgens de definitie van een variantie is:

$$\sigma^2(\underline{x}') = \mathcal{E}(\underline{x}' - \mathcal{E}\underline{x}')^2 = \mathcal{E}(\underline{x}' - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Zodat wij, als schatting voor σ^2 de grootheid

$$(12;8) \qquad \underline{s}'^2 \overset{\text{def}}{=\!=} \frac{1}{n}\sum_{i=1}^{n}(\underline{x}_i - \bar{\underline{x}})^2$$

op zijn merites kunnen onderzoeken.

Voorlopig verstaan wij daaronder, dat wij nagaan of deze schatting zuiver en asymptotisch raak is. Eerst de <u>zuiverheid</u>. De verwachting van $\underline{s}'^2$ berekenen wij als volgt.

$$\sum_{i=1}^{n}(\underline{x}_i - \bar{\underline{x}})^2 = \sum_{i=1}^{n}(\underline{x}_i - \mu + \mu - \bar{\underline{x}})^2 =$$

$$= \sum_{i=1}^{n}(\underline{x}_i - \mu)^2 - 2(\underline{x} - \mu)\sum_{i=1}^{n}(\underline{x}_i - \mu) + n(\bar{\underline{x}} - \mu)^2.$$

Hierin is

$$\sum_{i=1}^{n}(\underline{x}_i - \mu) = n(\bar{\underline{x}} - \mu),$$

zodat

$$\sum_{i=1}^{n}(\underline{x}_i - \bar{\underline{x}})^2 = \sum_{i=1}^{n}(\underline{x}_i - \mu)^2 - n(\underline{x} - \mu)^2.$$

Daar

$$\mathcal{E}(\underline{x}_i - \mu)^2 = \sigma^2 \quad \text{en} \quad \mathcal{E}(\bar{\underline{x}} - \mu)^2 = \sigma^2(\bar{\underline{x}}) = \frac{\sigma^2}{n},$$

geldt

$$(12;9) \qquad \mathcal{E}\,\underline{s}'^2 = \frac{n-1}{n}\sigma^2.$$

De schatting $\underline{s}'^2$ is dus geen zuivere schatting van σ^2, maar

$$(12;10) \qquad \underline{s}^2 \overset{\text{def}}{=\!=} \frac{1}{n-1}\sum_{i=1}^{n}(\underline{x}_i - \bar{\underline{x}})^2$$

is dat wel. Voor practische toepassingen wordt daarom veelal $\underline{s}^2$ als schatting van σ^2 gebruikt.

Opmerking 12;1

Het argument, dat $\underline{s}^2$ een zuivere schatting van σ^2 is en $\underline{s}'^2$ niet, verliest zijn waarde, indien men niet σ^2 maar σ wenst te schatten, hetgeen in de practijk gewoonlijk het geval is. Immers beschouwt men $\underline{s}$ als schatting van σ , dan is $\underline{s}$ weer onzuiver, zoals uit (9;21) volgt. Vullen wij daarin $\underline{s}$ in voor $\underline{x}$, dan komt er

$$\mathcal{E}\,\underline{s}^2 > (\mathcal{E}\,\underline{s})^2,$$

dus

$$\sigma^2 > (\mathcal{E}\,\underline{s})^2 \qquad \text{of} \quad \sigma > \mathcal{E}\,\underline{s},$$

zodat de verwachting van $\underline{s}$ kleiner is dan de te schatten parameter σ .

Het gebruik van $\underline{s}^2$ in plaats van $\underline{s}'^2$ berust dan ook in feite op traditie. Het bovenstaande geldt nog steeds zonder de veronderstelling van normaliteit. Voor normaal verdeelde $\underline{x}$ kan de factor

$$c_n \overset{\text{def}}{=} \frac{\sigma}{\mathcal{E}\,\underline{s}}$$

berekend worden. De schatting $c_n\,\underline{s}$ is dan een zuivere schatting voor σ .

Zowel $\underline{s}$ als $\underline{s}'$ zijn asymptotisch raak; voor $n \to \infty$ zijn zij aequivalent, daar dan $\frac{n-1}{n} \to 1$. Beschouw nu

$$\underline{s}'^2 = \frac{1}{n} \sum_{i=1}^{n} (\underline{x}_i - \mu)^2 - (\bar{\underline{x}} - \mu)^2.$$

Volgens stelling 10;2 convergeert $\bar{\underline{x}}$ stochastisch naar μ , dus $|\bar{\underline{x}} - \mu|$ naar 0, dus ook $(\bar{\underline{x}} - \mu)^2$ naar 0. Deze term kan dus verwaarloosd worden: behoudens een willekeurig kleine kans wordt hij zelf willekeurig klein. Noemen wij verder

dan is

$$\left.\begin{aligned} \underline{y}_i &\overset{\text{def}}{=} (\underline{x}_i - \mu)^2, \\ \mathcal{E}\,\underline{y}_i &= \sigma^2 \end{aligned}\right\} \qquad (i = 1, 2, \ldots, n)$$

en, wederom volgens stelling 10;2 convergeert nu

$$\frac{1}{n} \sum_{i=1}^{n} \underline{y}_i = \frac{1}{n} \sum_{i=1}^{n} (\underline{x}_i - \mu)^2$$

stochastisch naar de verwachting, dus naar σ^2 .[1] Daarmede is echter het gestelde (zij het niet volledig exact) bewezen.

---

1) Volgens stelling 10;2 moet de voorwaarde vervuld zijn dat $\sigma^2(\underline{y}_i)$ eindig is; deze voorwaarde kan echter vermeden worden.

Numeriek voorbeeld 12;1

Van 10 lampen van een bepaald type, aselect gekozen uit de productie, is de levensduur (in uren) gemeten. De uitkomsten zijn vermeld in de eerste kolom van tabel 12;1.[1)]

Tabel 12;1. Berekening van gemiddelde en variantie van een reeks waarnemingen

x_i	$x_i - 1000$	$(x_i - 1000)^2$
982	- 18	324
996	- 4	16
1047	47	2209
980	- 20	400
1057	57	3249
1025	25	625
990	- 10	100
1058	58	3364
994	- 6	36
1036	36	1296
totaal	165	11619

Om de berekening te vereenvoudigen kiezen wij eerst een z.g. "voorlopig gemiddelde", d.i. een getal x_0 (het doet er niet veel toe welk precies), dat vermoedelijk in de buurt van het werkelijke gemiddelde $\bar{x}$ ligt. In dit geval ligt het voor de hand $x_0 = 1000$ te nemen. In de tweede kolom staat $x_i - 1000$ berekend, in de derde $(x_i - 1000)^2$. Nu is

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} (x_i - x_0) + n x_0 ,$$

dus

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} (x_i - x_0) + x_0 .$$

In ons geval

$$\bar{x} = \frac{1}{10} \cdot 165 + 1000 = 1016{,}5 .$$

Verder is (vgl. blz. 89)

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i - x_0)^2 - n (x_0 - \bar{x})^2 .$$

---

1) Deze gegevens zijn niet aan de practijk ontleend.

In ons geval

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 11619 - 10(16,5)^2 = 8896,5 .$$

Derhalve is

$$s^2 = \frac{1}{9} . 8896,5 = 988,5, \quad s = 31,4 .$$

Willen wij nu bijv. schatten, hoe groot het percentage lampen is, dat een levensduur < 950 uur bezit, dan berekenen wij

$$\frac{1016,5 - 950}{31,4} = 2,12$$

en zoeken dit getal op in de tabel van de $N(0,1)$ -verdeling. De uitkomst is 0,017, dus 1,7%. Van de steekproef is dan ook geen enkele uitkomst zo laag.

Voorbeeld 12;2. Kurken tellen

Een aardig historisch voorbeeld van een eenvoudig schattingsprobleem is het volgende.[1] In een kurkenfabriek, die millioenen kurken in voorraad had, waarvan dagelijks duizenden in kleinere en grotere partijen moesten worden afgeleverd, waren voortdurend een aantal arbeiders bezig met de hand kurken te tellen. Dit was een vermoeiende en zenuwslopende bezigheid, waarmee het bovendien nog niet mogelijk was aan het einde van het jaar de voorraad op te nemen. Men wenste daarom een kurkentel-apparaat te ontwikkelen.

Aangezien het bij aflevering van partijen kurken niet op een paar meer of minder aankomt, kon Ir VAN ETTINGER de oplossing in statistische richting zoeken. Zijn methode komt hierop neer, dat uit een zak kurken van één maat een steekproef genomen wordt, waarvan het totale gewicht (g) en het aantal (n) bepaald wordt. Het gemiddelde gewicht van één kurk is dan dus g/n . Is nu het gewicht van de gehele zak kurken gelijk aan G , dan is

$$\frac{n\,G}{g}$$

een schatting van het aantal kurken uit de zak.

Deze methode werd bovendien als volgt gemechaniseerd. Een weegschaal werd vervaardigd, die in evenwicht is als het gewicht aan kurken op de éne schaal 99 maal zo groot is als dat op de andere. Men zet nu de zak kurken op de ene schaal en legt daaruit zoveel kurken op de andere, dat de weegschaal in evenwicht is.

---

1) J. VAN ETTINGER, Statistiek en productiviteit, Statistica 6 (1952), pp. 19-28, pag. 20.

Het totale aantal kurken is dan, bij benadering, gelijk aan 100 maal het aantal op de tweede schaal overgebrachte kurken. De tijdsbesparing is enorm, de investering gering en de arbeiders zijn trots op de eenvoudige wijze waarop zij kurken tellen. Het systeem is, in die fabriek, reeds bijna 30 jaar in gebruik.

Voorbeeld 12;3. Enquête

Wij beschouwen vervolgens een enquête, die erop gericht is de kennis van een bepaalde groep mensen te onderzoeken. Voor de eenvoud vatten wij één vraag in het oog, die juist of onjuist beantwoord kan worden. Deze vraag wordt voorgelegd aan een steekproef van n personen, aselect genomen uit de te onderzoeken groep. In statistisch jargon wordt deze te onderzoeken groep personen de populatie genoemd. Als model nemen wij aan, dat een fractie p van de personen uit deze groep de vraag goed zou beantwoorden en de overige (fractie $q=1-p$) fout. Het gaat er nu om deze fracties te schatten op grond van de antwoorden van de n personen van de steekproef.

Kiezen wij aselect één persoon uit de populatie, dan is de kans, dat wij een juist antwoord verkrijgen gelijk aan p . Bestaat de populatie uit N personen, dan bevinden zich daaronder dus Np , die een juist antwoord zouden geven en Nq , die de vraag onjuist zouden beantwoorden. Hieruit zijn er nu n aselect en zonder terugleggingg genomen en wij bevinden ons dus in de situatie van voorbeeld 7 van hoofdstuk 5. Het aantal juiste antwoorden in de steekproef van omvang n , dat wij met $\underline{x}$ aangeven, heeft dus een hypergeometrische verdeling.

$$(12;11) \qquad P[\underline{x}=x]=\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}},$$

met, volgens (9;26),

$$(12;12) \qquad \mathcal{E}\,\underline{x}=np,$$

dus

$$(12;13) \qquad \mathcal{E}\,\frac{\underline{x}}{n}=p.$$

De fractie juiste antwoorden in de steekproef is dus een zuivere schatting voor de onbekende fractie p . De vraag van asymptotisch gedrag doet zich hier niet op natuurlijke wijze voor, daar $n>N$ niet op kan treden. Voor $n=N$ vindt men p reeds exact.

Wat wel kan gebeuren is, dat N niet bekend is. Volgens (12;12) kan $\underline{x}/n$ dan toch als schatting gebruikt worden. Is $N >> n$, dan is volgens (5;10) de kansverdeling van $\underline{x}$ bij benadering <u>binomiaal</u>:

$$(12;14) \qquad P[\underline{x} = x] \approx \binom{n}{x} p^x q^{n-x},$$

waarbij dan (12;12) geldig blijft. Aan de schatting verandert dus niets. Volgens stelling 10;3 (de theoretische wet van de grote aantallen) is de schatting in het geval van een binomiale verdeling <u>asymptotisch raak</u>. (In het hier beschouwde geval impliceert $n \to \infty$ dat ook $N \to \infty$; is p bijv. de onbekende kans op succes bij een experiment, dat willekeurig vaak herhaald kan worden, dan doet deze complicatie zich niet voor.)

<u>Voorbeeld 12;4</u>. <u>Schieten op de kermis</u>

Een enthousiast kermisganger schiet in de schiettent op Goudse pijpen. Bij ieder schot kan hij alleen zien of hij het pijpje, waar hij op mikte geraakt heeft of niet, maar als hij mist kan hij niet ziet, waar zijn kogel terecht gekomen is. Hij kan bij een volgend schot dus geen gebruik maken van het resultaat van het vorige schot om zijn richten te verbeteren. Als model kunnen wij daarom stellen, dat hij bij ieder schot dezelfde kans p op succes heeft en dat de schoten onafhankelijk zijn.

De schutter neemt zich nu voor net zo lang door te gaan met schieten tot hij k pijpjes stukgeschoten heeft. Hiervoor heeft hij $\underline{x}$ schoten nodig. Gevraagd wordt uit deze getallen een schatting voor p af te leiden.

Op het eerste gezicht ligt het in dit geval wellicht voor de hand om, net als bij het vorige voorbeeld, de fractie successen, dus hier $k/\underline{x}$, als schatting voor p te nemen.

De verdeling van $\underline{x}$ is in dit geval <u>negatief binomiaal</u> (formule (7;7)) en volgens (9;25) geldt

$$\mathcal{E}\,\underline{x} = k/p \qquad \text{dus} \qquad p = k/\mathcal{E}\,\underline{x},$$

hetgeen een verdere rechtvaardiging in schijnt te houden.

Dit is echter niet geheel juist, daar uit deze formules niet volgt dat de schatting zuiver is. Daartoe zou nl. moeten gelden

$$\mathcal{E}\,k/\underline{x} = k\,\mathcal{E}\,\frac{1}{\underline{x}} = p,$$

dus zou

$$\mathcal{E}\,\frac{1}{\underline{x}} = \frac{1}{\mathcal{E}\,\underline{x}}.$$

moeten zijn en dit is niet het geval. In de oplossing van opgave 85 van het opgavenboekje (blz. 129-130) wordt bewezen, dat wel geldt

$$(12;15) \qquad \mathcal{E}\,\frac{k-1}{\underline{x}-1} = p,$$

zodat $\frac{k-1}{\underline{x}-1}$ wel een zuivere schatting van p is. Deze correctie van -1 in teller en noemer wordt intuïtief min of meer begrijpelijk als men overweegt, dat het laatste schot bij deze proefopzet altijd een succes is, daar het k^{e} succes het sein tot beëindiging van de serie is. Laat men nu dit laatste schot geheel weg, dan gaat de serie over in een reeks van $\underline{x}-1$ schoten met $k-1$ successen, die nu overal in de serie geplaatst kunnen zijn, juist als bij een binomiale verdeling het geval is.

Voor $k \to \infty$ zijn de beide schattingen $\frac{k}{\underline{x}}$ en $\frac{k-1}{\underline{x}-1}$ aequivalent. Nu kan (vgl. blz. 61 en 62) $\underline{x}$ beschouwd worden als de som van k onafhankelijke grootheden $\underline{x}_i$, die alle dezelfde (negatief binomiale) verdeling (met $k=1$) bezitten. Volgens stelling 10;2 convergeert dan het gemiddelde

$$\overline{\underline{x}} = \frac{1}{k}\sum_{i=1}^{k} \underline{x}_i$$

van deze grootheden stochastisch naar hun verwachting, die $\frac{1}{p}$ is. Met andere woorden $\underline{x}/k$ zal, voor voldoende grote k, willekeurig dicht tot $\frac{1}{p}$ naderen. Dit betekent echter dat $k/\underline{x}$ willekeurig dicht bij p komt. Beide schattingen zijn dus asymptotisch raak.

Opmerking

Voor $k=1$ is het niet mogelijk de beschreven correctiemethode toe te passen, daar de teller van de schatting dan = 0 zou worden. Voor dit geval is geen zuivere schatting voor p beschikbaar.

De verwachting van $\frac{1}{\underline{x}}$ voor $k=1$ laat zich als volgt berekenen.

$$\mathcal{E}\,\frac{1}{\underline{x}} = \sum_{x=1}^{\infty} \frac{1}{x}\, p\, q^{x-1} = \frac{p}{q}\sum_{x=1}^{\infty} \frac{q^{x}}{x} = -\frac{p}{q}\,{}^{e}\log(1-q).$$

Voor $p=q=\frac{1}{2}$ wordt dit ${}^{e}\log 2 = 0{,}69$. De verwachting van de schatting verschilt dan dus aanzienlijk van de te schatten waarde.

Wij behandelen vervolgens de elementen van de theorie der meest aannemelijke schattingen, waarvan de grondslag gelegd is door R.A. FISHER. Hoewel deze theorie niet de enige theorie over schattingen is, is het momenteel toch wel de belangrijkste en wij zullen ons tot deze theorie beperken.

Het principe van de theorie zetten wij uiteen voor continue verdelingen. Voor discrete is het analoog, met kansen in plaats van verdelingsdichtheden.

Laat

(12;16) $$\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$$

n (al of niet onafhankelijke) waarnemingen voorstellen, die een simultane verdeling bezitten, met verdelingsdichtheid

(12;17) $$f(x_1, x_2, \ldots, x_n | \theta_1, \theta_2, \ldots, \theta_h),$$

waarin $\theta_1, \theta_2, \ldots, \theta_h$ onbekende parameters voorstellen, die uit de waarnemingen geschat moeten worden.

Stellen nu $x_1, x_2, \ldots, x_n$ de waarden voor, die in werkelijkheid gevonden zijn, zodat het dus getallen zijn, geen variabelen meer, dan is (12;17) een functie van de onbekende parameters θ_j $(j = 1, 2, \ldots, h)$.

Deze functie wordt de <u>aannemelijkheidsfunctie</u> of kortweg <u>aannemlijkheid</u> genoemd en aangegeven als

(12;18) $$L(\theta_1, \theta_2, \ldots, \theta_h) \stackrel{\text{def}}{=} f(x_1, x_2, \ldots, x_n | \theta_1, \theta_2, \ldots, \theta_h).$$

Gewoonlijk (doch niet voor alle verdelingsdichtheden f) bezit L een maximum; het punt, waarin dit wordt aangenomen, wordt aangegeven met $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_h$ en deze waarden - die functies van de x_i $(i = 1, 2, \ldots, n)$ zijn - worden de <u>meest aannemelijke schattingen</u> van $\theta_1, \theta_2, \ldots, \theta_h$ genoemd.

<u>Toelichting aan voorbeelden</u>

<u>Voorbeeld 12;1. Normale verdeling</u> (zie ook blz. 87)

In dit voorbeeld heeft $\underline{x}$ een normale verdeling met onbekende μ en σ. De waarnemingen zijn $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. Daar deze stochastisch onafhankelijk zijn, is

(12;19) $$L(\mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} = e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}}.$$

In deze vorm beschouwen wij σ^2 (en niet σ) als onbekende parameter, daar dit de berekeningen vereenvoudigt. Om het maximum van L te vinden, bij gegeven waarden der x_i, differentiëren wij, ter verdere vereenvoudiging, niet L zelf, maar de logarithme van L naar μ en naar σ^2.

$$(12;20)\qquad \log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2},$$

$$(12;21)\qquad \frac{d\log L}{d\mu} = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2},$$

$$(12;22)\qquad \frac{d\log L}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^4}.$$

Stellen wij deze afgeleiden gelijk aan 0, dan verkrijgen wij dus 2 vergelijkingen voor $\hat{\mu}$ en $\hat{\sigma}^2$, en wel

$$(12;23)\qquad \sum_{i=1}^{n}(x_i-\hat{\mu}) = 0$$

en

$$(12;24)\qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2.$$

Uit de eerste volgt

$$(12;25)\qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x} \qquad \text{(vgl. (12;2))}$$

en vervolgens uit de tweede

$$(12;26)\qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s'^2 \qquad \text{(vgl. (12;8))}.$$

De meest aannemelijke schattingen komen dus in dit geval overeen met de boven op intuïtieve wijze gekozene.

Opmerkingen

12;2. Indien wij niet σ^2, maar σ zelf als onbekende parameter gekozen hadden, zou dezelfde uitkomst verkregen zijn. Dit is een speciaal geval van een algemene stelling, inhoudende dat de meest aannemelijke schatting van een continue 1,1-functie van een onbekende parameter verkregen wordt door dezelfde functie te nemen van de meest aannemelijke schatting van die parameter. Dus: is $\hat{\theta}$ de meest aannemelijke schatting van θ, dan is $\varphi(\hat{\theta})$ de meest aannemelijke schatting van $\varphi(\theta)$ als φ een continue en ondubbelzinnig omkeerbare functie is.

12;3. Wij zien tevens, dat meest aannemelijke schattingen niet zuiver behoeven te zijn. Dit volgt trouwens reeds uit opmerking 12;2, daar zuiverheid, zoals wij in opmerking 12;1 gezien hebben, bij transformatie van de parameter verloren kan gaan.

Voorbeeld 12;2. Binomiale verdeling (zie ook blz. 94)

Is $\underline{x}$ binomiaal verdeeld met onbekende p , maar bekende n , dan geldt dus

$$L(p) = P[\underline{x} = x \mid p] = \binom{n}{x} p^x q^{n-x}. \tag{12;27}$$

Dus

$$\log L(p) = \log \binom{n}{x} + x \log p + (n-x) \log (1-p). \tag{12;28}$$

Differentiëren naar p geeft

$$\frac{d \log L}{dp} = \frac{x}{p} - \frac{n-x}{1-p}, \tag{12;29}$$

zodat wij, om $\hat{p}$ te vinden, de vergelijking

$$\frac{x}{\hat{p}} = \frac{n-x}{1-\hat{p}} \tag{12;30}$$

moeten oplossen. Dit geeft

$$\hat{p} = \frac{x}{n}. \tag{12;31}$$

Opmerking 12;4

Ook hier vinden wij dus de vroeger reeds vermelde schatting. Voor de hypergeometrische verdeling is dit niet het geval; daarbij treden complicaties op, o.a. doordat Np in dat geval geheel moet zijn, zodat p niet alle waarden tussen 0 en 1 kan bezitten. Wij zullen dit geval hier buiten beschouwing laten.

Voorbeeld 12;4. Negatief binomiale verdeling (zie ook blz. 94)

Is $\underline{x}$ negatief binomiaal verdeeld met onbekende p , maar bekende k , dan is

$$L(p) = \binom{x-1}{k-1} p^k q^{x-k}, \tag{12;32}$$

dus

$$\log L(p) = \log \binom{x-1}{k-1} + k \log p + (x-k) \log (1-p) \tag{12;33}$$

en

$$\frac{d \log L}{dp} = \frac{k}{p} - \frac{x-k}{1-p}. \tag{12;34}$$

De op te lossen vergelijking is nu dus

$$\frac{k}{\hat{p}} = \frac{x-k}{1-\hat{p}}, \tag{12;35}$$

waaruit volgt

$$\hat{p} = \frac{k}{x}. \tag{12;36}$$

Voorbeeld 12;5. Hoogwaterstanden; exponentiële verdeling

Indien men de frequentieverdeling van hoogwaterstanden waargenomen in Hoek van Holland, beschouwt, blijkt dat deze, voor zoverre het niet te kleine waarden betreft, bij goede benadering exponentieel verdeeld zijn. Dit betekent, dat voor iedere niet te lage waarde x_o geldt

$$P[\xi \geqq x \mid \xi \geqq x_o] = e^{-\lambda(x-x_o)} \qquad (x > x_o). \tag{12;37}$$ [1)]

Geldt dit voor één waarde x_o, dan ook voor iedere grotere waarde $x_o' > x_o$. Immers uit (12;37) volgt

$$e^{-\lambda(x-x_o)} = \frac{P[\xi \geqq x \wedge \xi \geqq x_o]}{P[\xi \geqq x_o]}.$$

Daar $x \geqq x_o$ is de teller gelijk aan $P[\xi \geqq x]$. De noemer geven wij aan met P_o. Dan geldt dus

$$P[\xi \geqq x] = P_o\, e^{-\lambda(x-x_o)}. \tag{12;38}$$

Voor een willekeurige waarde $x_o' > x_o$ geldt dus

$$P[\xi \geqq x_o'] = P_o\, e^{-\lambda(x_o'-x_o)}. \tag{12;39}$$

Voor $x \geqq x_o'$ is nu

$$P[\xi \geqq x \mid \xi \geqq x_o'] = \frac{P[\xi \geqq x \wedge \xi \geqq x_o']}{P[\xi \geqq x_o']}.$$

Hierin is de teller weer gelijk aan $P[\xi \geqq x]$ en voor de noemer geldt (12;39). Dus is

$$P[\xi \geqq x \mid \xi \geqq x_o'] = \frac{P_o \cdot e^{-\lambda(x-x_o)}}{P_o \cdot e^{-\lambda(x_o'-x_o)}} = e^{-\lambda(x-x_o')} \quad (x \geqq x_o'). \tag{12;40}$$

---

1) P.J. WEMELSFELDER, Wetmatigheden in het optreden van stormvloeden, "De Ingenieur", 3 Maart 1939.

Dit betekent, dat de keuze van het "beginpunt" x_o (of x_o') de vorm van de verdeling - en in het bijzonder de waarde van λ - niet beïnvloedt, hetgeen het mogelijk maakt de lage waarden van x, waar de verdeling niet bekend is, uit te schakelen zonder dat de betrekkelijk willekeurige keuze van het beginpunt veel invloed op de uitkomst uitoefent.

Deze eigenschap van de exponentiële verdeling, dat de voorwaardelijke verdeling onder de voorwaarde $\underline{x} \geqq x_o$ niet van de waarde van x_o afhangt is daarom voor het statistische onderzoek van de hoogwaterstanden aan de Nederlandse kust - en daarmede voor de bepaling van dijkhoogten, die een voldoende bescherming tegen overstromingen bieden - van veel belang. Een verklaring voor het experimenteel vastgestelde feit, dat de verdeling der hoogwaterstanden exponentieel is, is (nog?) niet bekend.

Dat de verdeling deze vorm bij goede benadering bezit kan experimenteel vastgesteld worden door gebruik te maken van logarithmisch grafiekenpapier. Noemen wij het linkerlid van (12;37) P, dan geldt

$$\log P = -\lambda(x - x_o). \tag{12;41}$$

Zetten wij dus $\log P$ verticaal uit tegen $x - x_o$ horizontaal, dan wordt een rechte lijn verkregen, zoals in fig. 12;1 geschetst is.

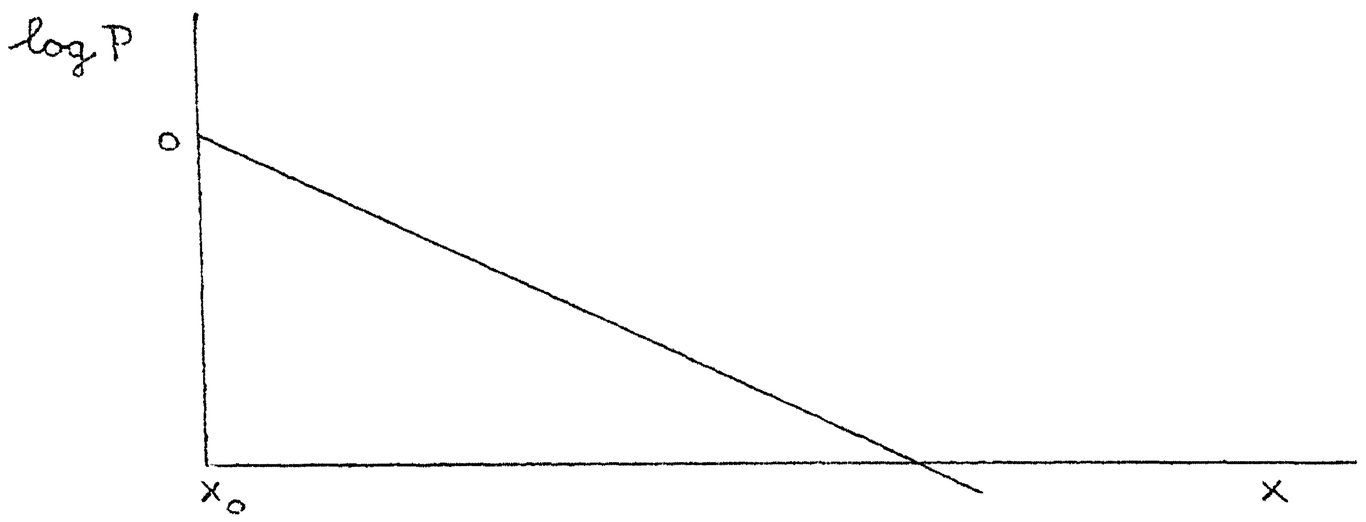


Fig. 12;1. Verdeling van hoogwaterstanden in Hoek van Holland

Deze lijn is, uiteraard, onbekend en moet uit waarnemingen geschat worden. Beschouwen wij daartoe alle hoogwaterstanden boven x_o, die in een bepaalde periode zijn opgetreden, dan kunnen deze (indien bepaalde voorzorgen genomen worden, die hier onbesproken blijven) beschouwd worden als onafhankelijke waarnemingen van een stochastische grootheid $\underline{x}$, die de door (12;37) gegeven verdeling bezit.

De enige onbekende parameter is λ en daarvan berekenen wij nu de meest aannemelijke schatting.

De verdelingsdichtheid, steeds onder de voorwaarde $\underline{x} \geqq x_0$, is

$$(12;42) \qquad f(x) = F'(x) = \frac{d(1-P)}{dx} = \lambda e^{-\lambda(x-x_0)},$$

dus als $x_1, x_2, \ldots, x_n$ de waarnemingen zijn, is

$$(12;43) \qquad L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda(x_i - x_0)} = \lambda^n e^{-\lambda \sum_{i=1}^{n} (x_i - x_0)}.$$

Verder

$$(12;44) \qquad \log L = n \log \lambda - \lambda \sum_{i=1}^{n} (x_i - x_0)$$

en

$$(12;45) \qquad \frac{d \log L}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} (x_i - x_0).$$

Derhalve is

$$(12;46) \qquad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} (x_i - x_0)} = \frac{1}{\bar{x} - x_0},$$

dus $\hat{\lambda}^{-1}$ is het gemiddelde der waarnemingen boven x_0 verminderd met x_0 .

De schatting (12;46) is niet zuiver. Beschouwen wij de schatting weer als een stochastische grootheid, dan is

$$\underline{\hat{\lambda}} = \frac{1}{\underline{\bar{x}} - x_0},$$

terwijl steeds de voorwaarde $\underline{x} \geqq x_0$, dus ook $\underline{\bar{x}} \geqq x_0$ geldt.

In par. 2 van de appendix wordt bewezen, dat

$$(12;47) \qquad \mathcal{E}\left(\frac{1}{\underline{\bar{x}} - x_0} \,\middle|\, \underline{\bar{x}} \geqq x_0 \right) = \frac{n}{n-1} \lambda,$$

zodat $\frac{n-1}{n} \frac{1}{\underline{\bar{x}} - x_0} = \frac{n-1}{\sum_{i=1}^{n} (\underline{x}_i - x_0)}$ wel een zuivere schatting van λ is.

Voor $n \to \infty$ zijn de twee schattingen aequivalent. Verder is

$$\mathcal{E}\left(\underline{\hat{\lambda}}^{-1} \,\middle|\, \underline{\bar{x}} \geqq x_0 \right) = \mathcal{E}\left(\underline{\bar{x}} - x_0 \,\middle|\, \underline{\bar{x}} \geqq x_0 \right) = \mathcal{E}\left(\underline{x} - x_0 \,\middle|\, \underline{x} \geqq x_0 \right)$$

en de (voorwaardelijke) verdelingsdichtheid van $\underline{x}$ is (zie (12;42))

$$f(x) = \lambda e^{-\lambda(x-x_0)},$$

zodat

$$\mathcal{E}\left(\underline{x} - x_0 \,\middle|\, \underline{x} \geqq x_0 \right) = \lambda \int_{x_0}^{\infty} (x - x_0) e^{-\lambda(x-x_0)} dx = \lambda \int_{0}^{\infty} v\, e^{-\lambda v} dv.$$

Nu is echter (partiëel integreren)

$$\lambda \int_0^\infty v\, e^{-\lambda v}\, dv = \left[-v\, e^{-\lambda v}\right]_0^\infty + \int_0^\infty e^{-\lambda v}\, dv = \left[-\tfrac{1}{\lambda}\, e^{-\lambda v}\right]_0^\infty = \tfrac{1}{\lambda}.$$

Dus is $\frac{1}{\underline{x}}$ wel een zuivere schatting van $\frac{1}{\lambda}$. De schatting $\frac{1}{\underline{x} - x_0}$ (en dus ook $\frac{n-1}{n(\underline{x} - x_0)}$) is wel asymptotisch raak. Uit stelling 10;2 volgt nl. dat $\underline{x} - x_0$ stochastisch convergeert naar $\mathcal{E}(\underline{x} - x_0 \mid \underline{x} \geq x_0) = \frac{1}{\lambda}$ en dus convergeert $\hat{\underline{\lambda}} = \frac{1}{\underline{x} - x_0}$ stochastisch naar λ .

Opmerking 12;5

Bij het werkelijke onderzoek van deze waterstanden werden niet alle hoogwaterstanden gebruikt, doch slechts een in overleg met het K.N.M.I. op grond van meteorologische overwegingen uitgekozen groep, behorende bij bepaalde depressies van een potentieel gevaarlijk karakter.

Voorbeeld 12;6. Regressie; theorie der kleinste kwadraten

Wij beschouwen, als laatste voorbeeld, een stochastische grootheid $\underline{y}$, waarvan de verdeling afhangt van een andere, niet stochastische grootheid x . Zo kan $\underline{y}$ bijv. de in een week door een koe geleverde hoeveelheid melk voorstellen (die van week tot week en van koe tot koe verschilt) en x de leeftijd (in jaren) van de koe. Of $\underline{y}$ kan de lengte van de remweg van een auto voorstellen en x de snelheid bij het begin van het remmen. In deze beide gevallen kan men x zelf kiezen - in het eerste geval door een koe van de gewenste leeftijd uit te zoeken, in het tweede door bij de gewenste snelheid te gaan remmen -, terwijl $\underline{y}$ waargenomen wordt.

In beide gevallen gaat het erom te onderzoeken hoe de kansverdeling van $\underline{y}$ van x afhangt.

Wij beschouwen hier alleen het eenvoudigste geval, nl. dat alleen de verwachting van $\underline{y}$ van x afhankelijk is en wel lineair:

$$\mathcal{E}(\underline{y} \mid x) = \alpha + \beta x, \tag{12;48}$$

waarin α en β onbekende parameters zijn, die geschat moeten worden. Verder veronderstellen wij, dat de verdeling van $\underline{y}$, bij gegeven x, normaal verdeeld is met een onbekende spreiding σ , die niet van x afhangt.

Op dit probleem gaan wij nu de methode der meest aannemelijke schattingen toepassen. De waarnemingen, die verondersteld worden stochastisch onafhankelijk te zijn, zijn nu van de vorm

$$(12;49) \qquad (x_1, \underline{y}_1), (x_2, \underline{y}_2), \ldots, (x_n, \underline{y}_n),$$

daar iedere waar te nemen waarde $\underline{y}_i$ bij een bepaalde waarde x_i van x behoort. De waarnemingsuitkomsten (x_i, y_i) $(i = 1, 2, \ldots, n)$ kunnen als punten in een vlak worden uirgezet, zoals in fig. 12;2 gedaan is.

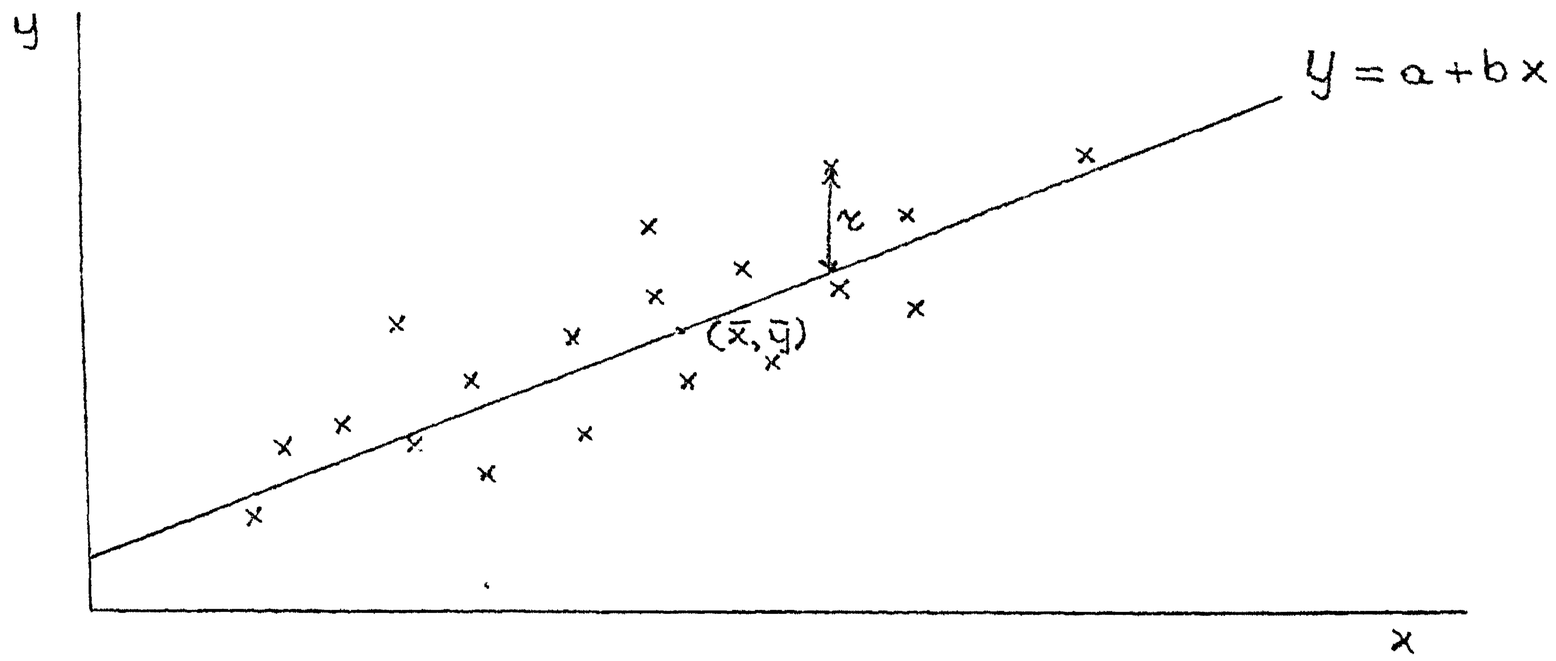


Fig. 12;2. Lineaire regressie van y op x .

De verdelingsdichtheid van $\underline{y}_i$ is

$$(12;50) \qquad f(y_i | x_i) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{\{y_i - \mathcal{E}(\underline{y}_i | x_i)\}^2}{\sigma^2}}$$

Wegens de onafhankelijkheid der $\underline{y}_i$ is

$$(12;51) \qquad L(\alpha, \beta, \sigma^2) = \prod_{i=1}^{n} f(y_i | x_i)$$

en dus (vgl. (12;48))

$$(12;52) \qquad \log L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

Voor de schatting van α en β is alleen de laatste term van belang. Deze geven wij (onder weglaten van de factor $-\frac{1}{2\sigma^2}$) aan met Q .

Daar $\log L$ gemaximaliseerd moet worden, dient Q geminimaliseerd te worden. Voor de eenvoud van de berekeningen voeren wij het gemiddelde $\bar{x}$ der x_i in en schrijven

$$(12;53) \qquad \alpha + \beta x_i = \alpha' + \beta(x_i - \bar{x}), \quad \text{dus} \qquad \alpha = \alpha' - \beta\bar{x}.$$

Nu wordt Q :

$$(12;54) \qquad Q = \sum_{i=1}^{n} \{ y_i - \alpha' - \beta (x_i - \bar{x}) \}^2.$$

Door uitvermenigvuldigen van de kwadraten kan hiervoor geschreven worden

$$Q = \sum_{i=1}^{n} (y_i - \alpha')^2 - 2\beta \sum_{i=1}^{n} (y_i - \alpha')(x_i - \bar{x}) + \beta^2 \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Voor de tweede term kan geschreven worden

$$-2\beta \sum_{i=1}^{n} y_i (x_i - \bar{x}) + 2\beta\alpha' \sum_{i=1}^{n} (x_i - \bar{x})$$

en daar $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$ valt de laatste term weg, zodat verkregen wordt

$$(12;55) \qquad Q = \sum_{i=1}^{n} (y_i - \alpha')^2 + \beta^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\beta \sum_{i=1}^{n} y_i (x_i - \bar{x}).$$

Q bestaat dus uit 2 stukken, waarvan het éne alleen α' en het andere alleen β bevat, zodat deze afzonderlijk geminimaliseerd kunnen worden. Beschouw eerst de eerste term, die α' bevat.

$$\sum_{i=1}^{n} (y_i - \alpha')^2 = \sum_{i=1}^{n} (y_i - \bar{y} + \bar{y} - \alpha')^2 =$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + 2(\bar{y} - \alpha') \sum_{i=1}^{n} (y_i - \bar{y}) + n (\bar{y} - \alpha')^2 =$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 + n (\bar{y} - \alpha')^2,$$

daar $\sum_{i=1}^{n} (y_i - \bar{y}) = 0$.

De eerste der overgebleven termen bevat α' niet en de tweede term wordt minimaal als voor α' de waarde $\bar{y}$ ingevuld wordt. Derhalve is de meest aannemelijke schatting van α' :

$$(12;56) \qquad \alpha' \overset{\text{def}}{=} \hat{\alpha}' = \bar{y}.$$

De twee laatste termen van Q in (12;55), die β bevatten, worden, zoals bekend, tezamen minimaal als voor β de waarde $\sum_{i=1}^{n} y_i (x_i - \bar{x}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$ ingevuld wordt, dus is

$$(12;57) \qquad b \overset{\text{def}}{=} \hat{\beta} = \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

De schattingen voor α' en β zijn dus verkregen en uit (12;48) volgt nu, tezamen met (12;53), als schatting y voor $\mathcal{E}(\underline{y} \mid x)$

$$(12;58) \qquad y = a' + b(x_i - \bar{x}) = a + b x_i \qquad (a \overset{\text{def}}{=} a' - b\bar{x}).$$

y wordt de bij x behorende <u>regressiewaarde</u> genoemd; (12;58) stelt een rechte lijn voor, die door het punt $(\bar{x}, \bar{y})$ gaat (het "zwaartepunt" van de "puntenwolk" in fig. 12;2) en die de (geschatte) <u>regressielijn</u> van y op x genoemd wordt. De "werkelijke" regressielijn wordt door (12;48) gegeven.

Uit (12;52) valt nu ook de meest aannemelijke schatting van σ^2 te berekenen, na in deze vorm voor α en β de reeds gevonden schattingen a en b ingevuld te hebben.

Noemen wij

$$(12;59) \qquad Q_0 = \sum_{i=1}^{n} (y_i - a - b x_i)^2,$$

dan gaat $\log L$ dus over in

$$-\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{Q_0}{2\sigma^2}$$

en de afgeleide hiervan naar σ^2 is

$$-\frac{n}{2\sigma^2} + \frac{Q_0}{2\sigma^4}.$$

Stellen wij deze vorm gelijk aan 0, dan volgt daaruit de schatting $\hat{\sigma}^2$ van σ^2 :

$$(12;60) \qquad \hat{\sigma}^2 = \frac{Q_0}{n}.$$

Vervolgens onderzoeken wij de <u>zuiverheid</u> der verkregen schattingen, door de $\underline{y}_i$ weer als stochastisch te beschouwen. Eerst $\underline{a}'$.

$$\mathcal{E}\,\underline{a}' = \mathcal{E}\,\underline{\bar{y}} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}\,(\underline{y}_i \mid x_i) = \frac{1}{n}\sum_{i=1}^{n}\{\alpha' + \beta(x_i - \bar{x})\} =$$

$$= \alpha' + \frac{\beta}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = \alpha',$$

daar $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. Derhalve is $\underline{a}'$ een zuivere schatting van α' .

Vervolgens $\underline{b}$.

$$\mathcal{E}\sum_{i=1}^{n}\underline{y}_i(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i - \bar{x})\,\mathcal{E}\,\underline{y}_i =$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})\{\alpha' + \beta(x_i - \bar{x})\} = \alpha'\sum_{i=1}^{n}(x_i - \bar{x}) + \beta\sum_{i=1}^{n}(x_i - \bar{x})^2 =$$

$$= \beta\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Dus

$$\mathcal{E}\,\underline{b} = \frac{\beta\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta,$$

zodat $\underline{b}$ een zuivere schatting van β blijkt te zijn.

Uit deze twee resultaten volgt, dat ook $\underline{a}$ een zuivere schatting van α is, immers (zie (12;58) en (12;53)):

$$\mathcal{E}\,\underline{a} = \mathcal{E}(\underline{a}' - \underline{b}\,\bar{x}) = \alpha' - \beta\,\bar{x} = \alpha.$$

De berekening van $\mathcal{E}\,\underline{\hat{\sigma}}^2$, die iets ingewikkelder is, zullen wij hier niet geven. De uitkomst is

(12;61) $$\mathcal{E}\,\underline{\hat{\sigma}}^2 = \frac{n-2}{n}\,\sigma^2,$$

zodat $\underline{\hat{\sigma}}^2$ geen zuivere schatting is, maar

(12;62) $$\underline{s}^2 \overset{\text{def}}{=} \frac{\underline{Q}_0}{n-2} = \frac{\sum_{i=1}^{n}(\underline{y}_i - \underline{a} - \underline{b}\,x_i)^2}{n-2}$$

wel.

De schattingen zijn ook <u>asymptotisch raak</u>, doch het bewijs daarvoor laten wij achterwege.

<u>Opmerkingen</u>

12;6. De hier toegepaste methode komt, meetkundig, daarop neer dat die lijn $y = a + b\,x$ in fig. 12;2 gekozen wordt, waarvoor de som van de kwadraten van de afstanden der waargenomen punten tot die lijn, in verticale richting gemeten, minimaal is. Deze afstanden worden de <u>residuen</u> genoemd. Eén residu is in fig. 12;2 door de letter r aangegeven. Het bij (x_i, y_i) behorende residu is

(12;63) $$r_i \overset{\text{def}}{=} y_i - a - b\,x_i,$$

dus

(12;64) $$Q_0 = \sum_{i=1}^{n} r_i^2.$$

De methode ontleent daaraan de naam: <u>methode der kleinste kwadraten</u>. Onder deze naam is zij reeds zeer lang bekend als de basis van de z.g. <u>foutentheorie</u>, die veel toepassing vindt op het gebied van de natuur- en sterrekunde, ook voor veel gecompliceerdere problemen dan de hier behandelde.

12;7. Soms zijn ook de $\underline{x}_i$ stochastisch. De methode kan dan onveranderd toegepast worden, daar men dan gebruik kan maken van de voorwaardelijke verdelingen der y_i onder de voorwaarden $\underline{x}_i = x_i \; (i = 1, 2, \ldots, n)$; deze voorwaardelijke verdelingen moeten dan voldoen aan de boven voor de $\underline{y}_i$ geformuleerde voorwaarden. Dit is bijv. het geval als $\underline{x}$ en $\underline{y}$ een 2-dimensionale normale verdeling bezitten, waaruit de waarnemingspunten (x_i, y_i) een steekproef vormen. Voorbeeld: de lengtes van aselect gekozen vaders en hun oudste zonen. Naast de regressie van $\underline{y}$ op $\underline{x}$ kan dan ook de regressie van $\underline{x}$ op $\underline{y}$ onderzocht worden, die een andere uitkomst geeft. Wij gaan hier niet nader op in.

12;8. Ten onrechte wordt veelal ondersteld, dat de methode der kleinste kwadraten alleen voor normaal verdeelde $\underline{y}_i$ bruikbaar zou zijn. Uit het bovenstaande blijkt, dat bij de bewijzen der zuiverheid geen gebruik is gemaakt van die veronderstelling, zodat zij algemener gelden. Wel geldt speciaal in het normale geval, dat de methode der kleinste kwadraten ook de meest aannemelijke schattingen geeft. In andere gevallen kunnen de volgens de beide methoden verkregen schattingen verschillend zijn.

Nauwkeurigheid van schattingen

Een schatting wordt uit waarnemingen berekend. Deze waarnemingen worden als stochastische grootheden beschouwd, zodat niet alleen zij, maar ook de schatting een kansverdeling bezit. De uitkomst, die verkregen wordt door bepaalde waargenomen waarden in de schattingsformule te substitueren - een concrete waarde van de schatting - kan zelf als één waarneming van die schatting beschouwd worden.

De kansverdeling van een schatting is een volledige beschrijving van de eigenschappen van die schatting, waarvan bijzondere eigenschappen, zoals zuiverheid en asymptotische raakheid speciale aspecten zijn. Een andere, in het voorafgaande nog niet beschouwde eigenschap is de nauwkeurigheid van een schatting, die grofweg door zijn spreiding gegeven wordt.

Voorbeeld 12;1. Spreiding van het gemiddelde van een steekproef

Wij beschouwden bij dit voorbeeld een steekproef van n waarnemingen

(12;1) $$\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$$

uit een normale verdeling met verwachting μ en variantie σ^2.
Als schatting voor μ namen wij

(12;2) $$\overline{\underline{x}} = \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i ,$$

hetgeen ook (vergelijk (12;25)) de meest aannemelijke schatting is. Volgens stelling 11;4 is $\overline{\underline{x}}$ zelf ook normaal verdeeld met (vergelijk (10;5) en (10;6))

(12;65) $$\mathcal{E}\,\overline{\underline{x}} = \mu \qquad \text{en} \qquad \sigma^2(\overline{\underline{x}}) = \frac{\sigma^2}{n} .$$

Voor de spreiding van $\overline{\underline{x}}$ geldt dus

(12;66) $$\sigma(\overline{\underline{x}}) = \frac{\sigma}{\sqrt{n}}$$

en men kan bewijzen (hetgeen wij hier niet zullen doen), dat er in het geval van een steekproef uit een normale verdeling geen enkele zuivere schatting van μ bestaat, die een kleinere spreiding bezit. Men noemt daarom $\overline{\underline{x}}$ de meest doeltreffende zuivere schatting van μ.

De grootheid

(12;67) $$\overline{\underline{x}}^* = \frac{\overline{\underline{x}} - \mu}{\sigma} \sqrt{n} ,$$

de gestandaardiseerde van $\bar{x}$, is dus $N(0,1)$-verdeeld, zodat, volgens tabel 1 bijv. geldt:

$$P[\,|x^*| \leqq 1{,}96\,] = 0{,}95$$

of, na omrekenen van de ongelijkheid tussen de vierkante haken:

(12;68) $$P\left[\,\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}} \leqq \mu \leqq \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\,\right] = 0{,}95.$$

Bij het numerieke voorbeeld van blz. 91 vonden wij $\bar{x} = 1016{,}5$. De ongelijkheid tussen de vierkante haken van (12;68) gaat, bij substitutie van deze waarde over in

$$1016{,}5 - 1{,}96\frac{\sigma}{\sqrt{10}} \leqq \mu \leqq 1016{,}5 + 1{,}96\frac{\sigma}{\sqrt{10}},$$

waarmee wij nog niet veel opschieten als σ onbekend is. Nemen wij even aan, dat σ op grond van vroegere waarnemingen bekend is en bijv. 33 bedraagt, dan vinden wij na invullen van deze waarde

(12;69) $$996 \leqq \mu \leqq 1037\,.$$

Deze uitspraak, waarbij de onbekende parameter μ tussen twee grenzen ingesloten wordt, kan juist of onjuist zijn. Op grond van (12;68) zou men de neiging kunnen hebben te zeggen, dat er een kans 0,95 is, dat hij juist is. Strikt genomen is dit niet waar: (12;69) is òfwel juist (en dan is de kans, dat hij juist is, gelijk aan 1) òfwel onjuist (en dan is die kans gelijk aan 0). Maar wèl volgt uit (12;68), dat de gevolgde methode een kans 0,95 bezit op een juiste uitkomst. Beperkt men dus zijn beschouwingen niet tot één geval (zoals de uitkomst (12;69)), maar beschouwt men de methode als zodanig, of een lange reeks toepassingen daarvan, dan kan men zeggen, dat de kans op een juiste uitspraak 0,95 bedraagt.

In algemene vorm wordt (12;68):

(12;70) $$P\left[\,\bar{x} - u_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}} \leqq \mu \leqq \bar{x} + u_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}}\,\right] = 1-\alpha,$$

waarin $u_{\frac{1}{2}\alpha}$ die waarde van u in tabel 1 voorstelt, waarvoor $k_u = \frac{1}{2}\alpha$ is.

Het interval tussen de vierkante haken van (12;70) wordt een betrouwbaarheidsinterval voor μ genoemd, $1-\alpha$ heet de betrouwbaarheidscoëfficiënt daarvan (dit is dus de kans, dat de methode tot een juiste uitspraak zal leiden) en α heet de onbetrouwbaarheid.

Meestal is σ niet bekend. In eerste instantie substitueert men dan veelal een schatting van σ , bijv. $s = \sqrt{s^2}$ (zie (12;10)) voor σ . Zoals wij op blz. 90 hebben gezien is deze schatting niet zuiver: gemiddeld wordt σ door $\underline{s}$ onderschat. Dit doet vermoeden, dat de grenzen, waartussen μ op deze wijze ingesloten wordt wel eens te dicht bij elkaar kunnen komen, of ook dat de kans op een verkeerde uitspraak bij toepassing van deze methode groter dan α zal zijn. Dit is inderdaad juist en men kan de substitutie van s voor σ dan ook slechts als een benadering beschouwen. Voor grote steekproeven ($n \to \infty$) bestaat dit bezwaar niet. Dan convergeert nl. $\underline{s}$ stochastisch naar σ . Voor kleine steekproeven bestaat - in het normale geval - een exacte methode, die op de z.g. verdeling van STUDENT berust. Wij gaan daar momenteel niet op in.

Uit (12;70) is (nogmaals) te zien, dat μ in principe met willekeurige nauwkeurigheid bepaald kan worden door n voldoende groot te nemen. Immers voor $n \to \infty$ naderen de beide grenzen tot elkaar. Dit is aequivalent met het vroeger bewezene, dat $\bar{x}$ voor $n \to \infty$ stochastisch tot μ convergeert.

In de practijk dient deze stelling echter met voorzichtigheid gehanteerd te worden. Indien men bijv. de lengte van een voorwerp een aantal malen opmeet met een meetinstrument dat niet nauwkeuriger dan in cm afleesbaar is en de werkelijke lengte van het voorwerp is gelijk aan 10,9 cm, dan zullen alle waarnemingen (of vrijwel alle) de waarde 11 cm aangeven en men behoeft er niet op te hopen, dat het gemiddelde tot 10,9 cm zal convergeren. De kansverdeling der waarnemingen is dan ook niet normaal, zelfs niet continu, daar alleen gehele waarden aangenomen kunnen worden. In de practijk zijn echter waarnemingen altijd van discreet karakter. De moraal daarvan is, dat het in de regel zinloos is het gemiddelde van een aantal metingen in meer decimalen op te geven dan op het meetinstrument afgelezen kunnen worden. Hiertegen wordt vaak gezondigd.

Voorbeeld 12;7. De mediaan in plaats van de verwachting

De mediaan van een continu verdeelde stochastische grootheid $\underline{x}$ is die waarde x_M , waarvoor geldt

$$(12;71) \qquad P[\underline{x} \leqq x_M] = P[\underline{x} \geqq x_M] = \tfrac{1}{2} .$$

Deze waarde is niet altijd ondubbelzinnig bepaald, in principe kan er een geheel interval van waarden zijn, die aan (12;71) voldoen. In dat geval wordt dit gehele interval de mediaan genoemd.

Bij discrete verdelingen kan (12;71) op dezelfde wijze gebruikt worden, als er minstens één waarde x_M is, die daaraan voldoet. Dit is niet altijd het geval. De algemene definitie, die voor alle verdelingen van toepassing is, luidt als volgt.

Definitie 12;3

De mediaan van een stochastische grootheid $\underline{x}$ is de verzameling van alle waarden x_M, die voldoen aan

(12;72) $\quad P[\underline{x} \leqq x_M] \geqq \frac{1}{2}$ en $P[\underline{x} \geqq x_M] \geqq \frac{1}{2}$.

Voor symmetrische verdelingen (zoals de normale en de binomiale met $p = \frac{1}{2}$) is de mediaan gelijk aan de verwachting. Voor de meeste verdelingen, die scheef naar rechts (resp. naar links) zijn, ligt de mediaan links (resp. rechts) van de verwachting (zie fig. 12;1).

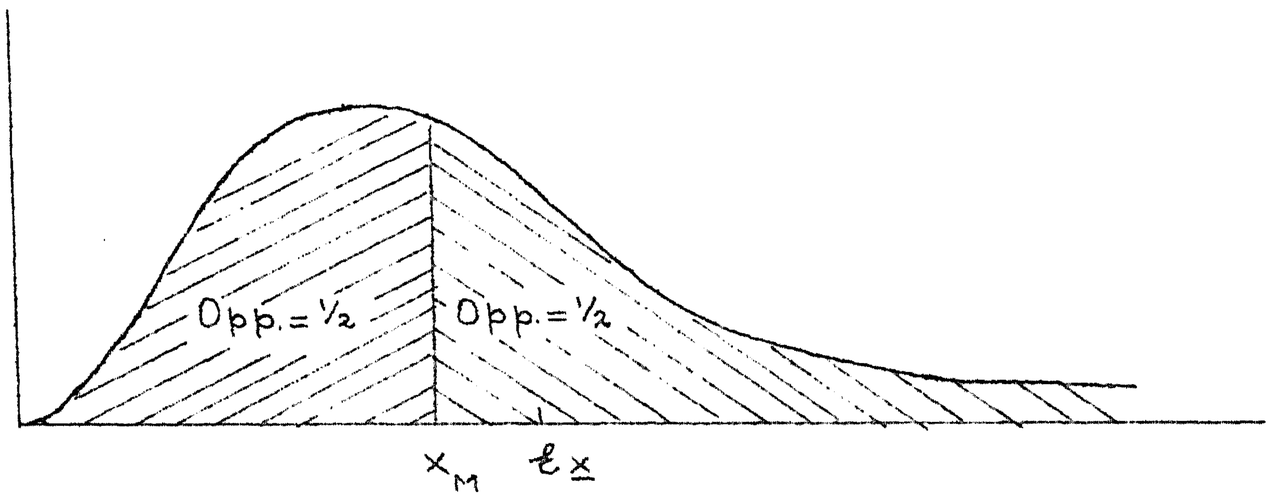


Fig. 12;1. Ligging van mediaan en verwachting bij een continue verdeling, die scheef naar rechts is.

Er is soms reden om de mediaan in plaats van de verwachting te gebruiken als een parameter, die grofweg de ligging van de verdeling aangeeft. Een voorbeeld daarvan treft men bij het onderzoeken van de studieduur bij studenten. Een aantal daarvan studeert nl. in het geheel niet af en heeft dus een oneindige studieduur. Is echter bij een stochastische grootheid de kans, dat de waarde ∞ aangenomen wordt, positief, dan is ook de verwachting $= \infty$. Voor de mediaan geldt dit niet, die wordt in het geheel niet beïnvloed door het gedrag van de verdeling "in de staarten", maar alleen door het middelste gedeelte.

Ook als men alleen die groep studenten beschouwt, die wel hun studie voltooien, dan behoudt de mediaan nog zekere voordelen boven de verwachting; er zijn nl. altijd studenten, die om één of andere reden hun studie gedurende lange tijd onderbreken en pas veel later afstuderen en het is vaak moeilijk vast te stellen wat men voor die gevallen de studieduur moet noemen. Dat deze lang is staat vrijwel altijd vast, maar op hoe lang men hem precies moet waarderen is vaak niet bij benadering te zeggen. Deze gevallen hebben echter juist een zeer grote invloed op het gemiddelde (als schatting van de verwachting), maar geen enkele op de mediaan.

Als schatting voor de mediaan x_M wordt gewoonlijk de mediaan van de waarnemingen genomen; dit is de middelste waarneming na rangschikking naar grootte als het aantal oneven is of het gemiddelde der twee middelste waarnemingen als het aantal even is. Notatie: $\underline{x}_m$.

Het onderzoek van de stochastische eigenschappen van deze schatting is vrij moeilijk en valt grotendeels buiten het bestek van de cursus.

Is $\underline{x}$ symmetrisch verdeeld om zijn verwachting (tevens mediaan) μ, dan is $\underline{\tilde{x}} = \underline{x}-\mu$ symmetrisch om 0, d.w.z. dat $-\underline{\tilde{x}}$ en $+\underline{\tilde{x}}$ dezelfde verdeling bezitten. Uit symmetrie-overwegingen volgt dan direct, dat ook de gereduceerde mediaan $\underline{\tilde{x}}_m = \underline{x}_m - \mu$ van een steekproef van waarnemingen symmetrisch om 0 verdeeld is, zodat $\underline{x}_m$ zelf een zuivere schatting van x_M is. Dat de mediaan $\underline{x}_m$ altijd asymptotisch goed is als schatting van x_M is vrij gemakkelijk in te zien. Wij bewijzen het alleen voor het geval van een continue verdeling met $f(x) > 0$ voor x in een omgeving van x_M. Laat $\varepsilon > 0$ een klein getal zijn en noem

$$p = P\left[\underline{x} \leq x_M - \varepsilon\right] \quad \text{en} \quad p' = P\left[\underline{x} \geq x_M + \varepsilon\right],$$

dan is $p < \frac{1}{2}$ en $p' < \frac{1}{2}$. De fractie waarnemingen, die $\leq x_M - \varepsilon$ zijn zal nu, voor $n \to \infty$, stochastisch convergeren naar p en de fractie, die $\geq x_M + \varepsilon$ zijn, naar p'. Behoudens een willekeurig kleine kans zal dus $\underline{x}_m$ tussen $x_M - \varepsilon$ en $x_M + \varepsilon$ vallen, als n voldoende groot is. Dit geldt echter voor iedere $\varepsilon > 0$, waaruit de te bewijzen eigenschap volgt.

In de meeste in de practijk voorkomende gevallen is de mediaan onnauwkeuriger dan het gemiddelde, d.w.z. dat hij een grotere spreiding bezit. In het geval van een normale verdeling geldt, dat $\underline{x}_m$ voor $m \to \infty$ asymptotisch normaal verdeeld is met μ als verwachting en

(12;73) $$\sigma(\underline{x}_m) \approx \sigma\sqrt{\frac{\pi}{2m}},$$

als σ de spreiding van de verdeling is. Wij hebben reeds opgemerkt, dat $\bar{\underline{x}}$ in dat geval de schatting met de kleinste mogelijke spreiding is. Deze is $\sigma/\sqrt{n}$. De limiet van het quotiënt der varianties

(12;74) $$\lim_{n\to\infty} \frac{\sigma^2(\bar{\underline{x}})}{\sigma^2(\underline{x}_m)} = \frac{2}{\pi} = 0{,}63$$

heet de asymptotische doeltreffendheid van $\underline{x}_m$.

Ondanks deze kleinere doeltreffendheid wordt bij routinetoepassingen (in het bijzonder controles op lopende-band-productie) van de statistiek in de industrie vaak gebruik gemaakt van de mediaan in plaats van het gemiddelde, ook bij normaal verdeelde grootheden. De reden daarvan is, dat de bepaling van de mediaan geen berekening nodig maakt, zodat deze ook door ongeschoold personeel zonder moeite bepaald kan worden (men gebruikt dan bijv. telkens 5 waarnemingen, waarvan de middelste gemakkelijk te bepalen is). De geringere doeltreffendheid kan dan vaak ondervangen worden door de controle wat vaker te doen plaatsvinden.

Voorbeeld 12;3. Binomiale verdeling

Bij een binomiale verdeling met parameters n en p vonden wij $\underline{x}/n$ als meest aannemelijke schatting. De variantie hiervan volgt direct uit (9;24): $\sigma^2(\underline{x}) = npq$. Wegens stelling 9;5 is nl.

(12;75) $$\sigma^2(\underline{x}/n) = \frac{1}{n^2}\sigma^2(\underline{x}) = \frac{pq}{n}.$$

De variantie, die weer de kleinst mogelijke voor een zuivere schatting van p is, is weer in de regel onbekend, doch kan uit de waarnemingen geschat worden. Een zuivere schatting ervan is

(12;76) $$\frac{\underline{x}(n-\underline{x})}{n^2(n-1)},$$

zoals uit de volgende berekening blijkt.

Volgens (9;20), (9;23) en (9;24) is

$$\mathcal{E}\,\underline{x}^2 = \sigma^2(\underline{x}) + (\mathcal{E}\,\underline{x})^2 = npq + n^2p^2.$$

Dus

$$\mathcal{E}\,\underline{x}(n-\underline{x}) = n\,\mathcal{E}\,\underline{x} - \mathcal{E}\,\underline{x}^2 = n^2p - npq - n^2p^2 = n(n-1)pq.$$

Verder weten wij uit stelling 11;8, dat $\underline{x}$ asymptotisch normaal verdeeld is. Dit geldt dan ook voor $\underline{x}/n$. Voor voldoende grote n kan dan ook, analoog aan (12;70) het volgende <u>benaderde betrouwbaarheidsinterval</u> voor de onbekende p aangegeven worden:

$$(12;77)\quad P\left[\frac{1}{n}\left(\underline{x} - u_{\frac{1}{2}\alpha}\sqrt{\frac{\underline{x}(n-\underline{x})}{n-1}}\right) \leqq p \leqq \frac{1}{n}\left(\underline{x} + u_{\frac{1}{2}\alpha}\sqrt{\frac{\underline{x}(n-\underline{x})}{n-1}}\right)\right] \approx 1-\alpha .$$

In de literatuur vindt men gewoonlijk onder de wortel de factor n in de noemer in plaats van $n-1$. Daar de formule alleen voor grote waarden van n nauwkeurig is, maakt dit niet veel verschil. De schatting (12;76) is trouwens wel zuiver als schatting van $\sigma^2(\underline{x}/n)$, maar de wortel eruit is weer (vergelijk opmerking 12;1 op blz. 90) een onzuivere schatting van $\sigma(\underline{x}/n)$. In plaats van (12;77) kan een betere benadering opgesteld worden, hetgeen wij echter tot een later hoofdstuk uitstellen evenals de afleiding van exacte betrouwbaarheidsintervallen voor een onbekende kans.

<u>Voorbeeld 12;6. Lineaire regressie</u>

Als laatste voorbeeld berekenen wij nog de varianties van de <u>regressiecoëfficiënten</u> $\underline{a}'$ en $\underline{b}$ uit (12;56) en (12;57). Daar alle $\underline{y}_i$ dezelfde variantie σ^2 bezitten en onafhankelijk verdeeld zijn, volgt uit (12;56) onmiddellijk

$$(12;78)\qquad \sigma^2(\underline{a}') = \frac{\sigma^2}{n}.$$

Volgens (12;57) is $\underline{b}$ een lineaire combinatie der $\underline{y}_i$ van de vorm

$$\underline{b} = \sum_{i=1}^{n} c_i\,\underline{y}_i ,$$

met

$$c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

waarin de x_i bekende (niet stochastische) getallen zijn. Derhalve geldt

$$\sigma^2(\underline{b}) = \sum_{i=1}^{n} c_i^2 \sigma^2(\underline{y}_i) = \sigma^2 \sum_{i=1}^{n} c_i^2,$$

dus

(12;79)
$$\sigma^2(\underline{b}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Daar verder zowel $\underline{a}'$ als $\underline{b}$ lineaire combinaties van de normaal verdeelde grootheden $\underline{y}_i$ zijn, zijn zij zelf ook normaal verdeeld. Met s^2 (zie (12;62)) als schatting voor σ^2 kan men op grond daarvan weer benaderde betrouwbaarheidsintervallen voor α' en β opstellen. De afleiding van exacte betrouwbaarheidsintervallen geven wij hier niet. Deze berust weer op de hier nog niet behandelde verdeling van STUDENT.

Verder kan men bewijzen, dat $\underline{a}'$ en $\underline{b}$ onafhankelijk verdeeld zijn (bij berekening blijkt de convariantie gelijk aan 0 te zijn), hetgeen bij verdere uitwerking - bijv. ter bepaling van een betrouwbaarheidsstrook voor de gehele regressielijn - aanzienlijke voordelen geeft. Ook voor deze schattingen geldt, dat zij - althans in het normale geval - de kleinst mogelijke spreiding bezitten.

Tenslotte valt op te merken, dat de optimale eigenschap van kleinst mogelijke spreiding, die in het bovenstaande herhaaldelijk is vermeld, niet toevallig ontstaat. Onder vrij algemene voorwaarden bezitten nl. de meest aannemelijke schattingen, althans asymptotisch voor $n \to \infty$, de eigenschap minimale spreidingen te bezitten en normaal verdeeld te zijn. In vele gevallen is de eerstgenoemde eigenschap ook reeds voor eindige n vervuld, terwijl de laatstgenoemde vaak voorkomt als de waarnemingen normaal verdeeld zijn. Het zijn o.a. deze eigenschappen, die de theorie der meest aannemelijke schattingen haar aantrekkelijkheid verlenen.

Hoofdstuk 13. Toetsingstheorie; binomiale toetsen

Men kan de meeste elementaire toepassingen van de statistie globaal in twee groepen verdelen: toepassingen van de schattingstheorie, die in hoofdstuk 12 behandeld is, en van de toetsingstheorie, die in dit hoofdstuk ter sprake komt. Het doel van de schattingstheorie is: op grond van steekproeven van waarnemingen tot quantitatieve schattingen van onbekende parameters te komen; de waarden van deze schattingen staan daarbij dus niet van tevore vast. De toetsingstheorie heeft ten doel op grond van een steekproef van waarnemingen te onderzoeken of één of meer onbekende parameters bepaalde gegeven waarden bezitten of niet.

In symbolen: is θ een onbekende parameter van een kansverdeling (bijv. de gemiddelde levensduur van een bepaald soort gloeilamp of de overlevingskans van een mug, die op een bepaalde wijze met een insecticide in aanraking wordt gebracht), dan geeft de schattingstheorie een schatting $\underline{t}$ voor θ met een kansverdeling, die min of meer om θ gecentreerd ligt, terwijl de toetsingstheorie erop gericht is te onderzoeken of θ al of niet gelijk is aan een bepaalde gegeven waarde θ_o.

Er bestaat een nauw verband tussen deze beide theorieën. Zo is het intuïtief reeds duidelijk, dat men, op grond van een gegeven waarnemingsreeks, tot verwerping van de getoetste waarde θ_o zal dienen over te gaan, indien de voor $\underline{t}$ uit deze waarnemingsreeks gevonden waarde sterk van θ_o afwijkt. Vindt men bijv. voor een steekproef van 100 lampen een gemiddelde levensduur van 700 uur (met een niet te grote spreiding daaromheen), terwijl men de waarde $\theta_o = 2000$ wil onderzoeken, dan is deze waarde voor θ niet acceptabel. Dit verband tussen de beide theorieën zal bij de behandeling van de toetsingstheorie duidelijk naar voren komen.

Voor zover het de toepassingen betreft is er echter een duidelijk verschil tussen de twee methoden. Bij vele wetenschappelijke problemen is in de eerste plaats de toetsing van bepaalde hypothesen van belang en pas in de tweede plaats de schattingstheorie. Indien men bijv. (een klassiek probleem) twee geneesmiddelen op hun merites ter bestrijding van een bepaalde ziekte wil vergelijken, is het in hoofdzaak van belang te onderzoeken of één van de twee beter is dan het andere. Is men er eenmaal van overtuigd, dat middel A beter is dan B, dan zal men, ceteris paribus, middel A voortaan preferen boven B; de grootte van het verschil in geneeskracht is van secundair belang. Wenst men daarentegen bijv. te berekenen welke premie men moet vragen voor een bepaalde verzekering, dan heeft men een quantitatieve

schatting van het risico nodig en het is dan niet genoeg (of zelfs niet van belang) om te weten, dat dit risico groter is dan een bepaald bedrag of groter dan het risico, dat aan een andere situatie verbonden is. De keuze tussen verschillende toepasbare statistische methoden hangt dus, zoals steeds het geval is, af van het doel van het onderzoek. Een duidelijke formulering van dit doel leidt gewoonlijk gemakkelijk tot de keuze van een bepaalde methode, terwijl het achterwege laten van deze formulering - een vaak voorkomend euvel, ook bij wetenschappelijke onderzoekingen - kan leiden tot een verkeerde keuze.

Wij zullen de toetsingstheorie in algemene bewoordingen beschrijven met daarnaast, ter illustratie, het speciale geval van binomiale toetsen. In het volgende hoofdstuk bespreken wij dan nog enkele vaak gebruikte toetsen.

Terminologie en notatie

Het waarnemingsmateriaal bestaat - naar wij zullen veronderstellen - uit één of meer reeksen van onafhankelijke waarnemingen. Wij geven het aan met

(13;1) $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$.

Deze stochastische grootheden behoeven niet alle dezelfde verdeling te bezitten; zij kunnen ook meerdimensionaal zijn: ieder ervan kan bijv. een paar van waarnemingen $(\underline{x}_i, \underline{y}_i)$ voorstellen.

Voorbeeld 13;1

Fraaie voorbeelden van het optreden van kansen met bekende waarden vindt men in de erfelijkheidsleer. Het mag wel bekend verondersteld worden, dat aan de proeven van MENDEL een erfelijkheidstheoretisch model ten grondslag ligt, dat bijv. leidt tot de uitkomst, dat bij het kruisen van bepaalde planten één van twee verschijnselen A en B zich voordoet en dat de kans op A bij iedere proef een bekende waarde (bijv. 1/4) bezit. De verschijnselen A en B kunnen twee kleuren (van erwten, of bloemen) zijn of iets dergelijks. De n waarnemingen, verkregen bij n onafhankelijke proeven van deze aard, nemen in dit geval de "waarden" A of B aan; men kan dit in cijfers omzetten door bij iedere proef het aantal malen A (dus 1 of 0) als waarneming (w) te beschouwen; (13;1) is dan dus een rij van n énen en nullen.

De toegelaten hypothesen; de parameterruimte

Als de kansverdelingen der $\underline{w}_i$ volledig bekend waren, zou er geen toetsingsprobleem bestaan. In de regel is er over deze kansverdelingen wel iets bekend, d.w.z. er worden bepaalde veronderstellingen over gemaakt, die als gegeven worden beschouwd. Deze hebben veelal betrekking op stochastische onafhankelijkheid der $\underline{w}_i$ en vaak ook op de vorm der verdelingen. Zo berusten veel toetsen op de veronderstelling van normaliteit der verdelingen. Wij zullen hier het geval beschouwen, dat de verdelingen der $\underline{w}_i$ van één of meer onbekende parameters afhangen, maar overigens volledig bekend zijn. Geven wij de onbekende parameter(s) aan met θ , dan is dus de simultane verdelingsfunctie

$$F(w_1,\ldots,w_n|\theta) \qquad (13;2)$$

gegeven, op de waarde(n) van θ na. Houden de veronderstellingen in, dat de $\underline{w}_i$ onafhankelijk zijn, dan kan (13;2) als een product geschreven worden:

$$F(w_1,\ldots,w_n|\theta) = \prod_{i=1}^{n} F_i(w_i|\theta). \qquad (13;3)$$

Iedere waarde van θ komt nu overeen met een enkelvoudige hypothese omtrent F . Door aan θ een bepaalde waarde te geven is F immers geheel vastgelegd en bekend.*) Alle waarden, die θ aan kan nemen, vormen tezamen de parameterruimte. Ieder punt daarvan komt overeen met een hypothese omtrent F en de verzameling van al deze hypothesen wordt de verzameling der toegelaten hypothesen genoemd. Deze verzameling wordt dus enerzijds bepaald door de omtrent F gemaakte veronderstellingen, anderzijds door de voor θ toegelaten (of: mogelijke) waarden.

Voorbeeld 13;1

Bij n erfelijkheidsproeven van het type van MENDEL is de verdeling van het volgende type:

$$P[\underline{w}_i = w_i | i=1,\ldots,n;p] = \prod_{i=1}^{n} \{w_i p + (1-w_i) q\}, \qquad (13;4)$$

met $w_i = 0$ of 1 $(i=1,\ldots,n)$, als $p = P[A]$ de kans op verschijnsel A $(w_i=1)$ voorstelt en $q = 1-p = P[B]$ de kans op verschijnsel B.

---

*) Wij zullen later gevallen van algemenere aard ontmoeten, waarbij niet verondersteld wordt dat F een bekende vorm bezit. De formulering wordt echter nodeloos ingewikkeld indien wij deze z.g. verdelingsvrije gevallen er direct in betrekken.

Deze schijnbaar ingewikkelde vorm wordt later teruggebracht tot de binomiale verdeling. Volgens de hypothese van MENDEL bezit de kans p een bepaalde waarde (waarvoor wij, om de gedachten te bepalen, de waarde 1/4 aannemen), maar het doel van het statistische onderzoek zal juist zijn na te gaan, of die hypothese juist is. Worden, in principe, alle waarden van p mogelijk geacht, dan is de parameterruimte hier van de vorm

(13;5) $$0 \leq p \leq 1.$$

<u>De te toetsen hypothese</u>

Veelal gaat het er nu om te onderzoeken of een enkelvoudige hypothese H_o van de vorm

(13;6) $$H_o : \theta = \theta_o,$$

waarin θ_o een gegeven getal is, op grond van de waarnemingen (13;1) al of niet verworpen moet worden. Deze hypothese wordt dan de te toetsen hypothese (ook wel: getoetste hypothese of nulhypothese) genoemd.

De te toetsen hypothese kan ook een samengestelde hypothese zijn, bijv. van de vorm

(13;7) $$H_o : \theta \leq \theta_o \quad \text{of} \quad \theta \geq \theta_o.$$

<u>Voorbeeld 13;1</u>

In het bovengenoemde geval zou de te toetsen hypothese zijn:

(13;8) $$H_o : p = 1/4$$

of in het algemeen $p = p_o$.

<u>De alternatieve hypothesen</u>

De uitkomst van het onderzoek is, dat de getoetste hypothese H_o al of niet <u>verworpen</u> wordt. Bij verwerping van (13;6) luidt de conclusie uiteraard: $\theta \neq \theta_o$; bij verwerping van $H_o : \theta \leq \theta_o$ luidt de conclusie: $\theta > \theta_o$.

De hypothesen, die aanvaard worden als H_o verworpen wordt, worden de <u>alternatieve hypothesen</u> genoemd. In de regel wordt, bij verwerping van H_o, een samengestelde (alternatieve) hypothese aanvaard.

Bij niet-verwerpen van H_o wordt, zoals wij later zullen zien, niet tot aanvaarding van H_o besloten. Althans dat is, strikt genomen, niet verantwoord. Niet-verwerpen betekent nl. alleen, dat er geen voldoende aanwijzingen <u>tegen</u> H_o zijn, doch dat houdt nog niet in, dat de aanwijzingen <u>voor</u> H_o voldoende sterk zijn om tot aanvaarding van H_o over te gaan. Dit punt komt nog uitvoeriger ter sprake.

Voorbeeld 13;1

Wordt (13;8) verworpen, dan luidt de conclusie:

(13;9) $$p \neq 1/4.$$

In practische bewoordingen: de erfelijkheidstheoretische hypothese, die tot de waarde $p = 1/4$ leidde, deugt niet en zal dus door een andere (die dan gewoonlijk ingewikkelder is) vervangen moeten worden. Deze nieuwe hypothese moet dan opnieuw getoetst worden, op grond van hernieuwde waarneming.

De toetsingsgrootheid

De beoordeling van de te toetsen hypothese geschiedt met behulp van een toetsingsgrootheid, die uit de waarnemingen wordt berekend:

(13;10) $$\underline{t} = t(\underline{w}_1, \ldots, \underline{w}_n).$$

Deze grootheid kan meerdimensionaal zijn, maar wij zullen voor het merendeel ééndimensionale toetsingsgrootheden ontmoeten.

Gewoonlijk is $\underline{t}$ een schatting van de onbekende parameter θ, of - wat daarmee aequivalent zal blijken te zijn - een monotone functie daarvan (wijkt deze schatting "sterk" van de te toetsen waarde θ_o af, dan zal men H_o verwerpen).

Voorbeeld 13;1

Volgens voorbeeld 12;3 (blz. 98) is

(13;11) $$\hat{\underline{p}} = \frac{\sum \underline{w}_i}{n},$$

de meest aannemelijke schatting voor de onbekende parameter p. In plaats van $\hat{\underline{p}}$ wordt $\sum \underline{w}_i = n\,\hat{\underline{p}}$ in dit geval als toetsingsgrootheid gebruikt. Dit is dus: het aantal malen, dat verschijnsel A bij de n proeven waargenomen wordt.

De verdeling van $\underline{t}$ onder de getoetste hypothese

De gedachtengang van een toets komt nu ongeveer overeen met die van een bewijs uit het ongerijmde. Onder H_o is de verdelingsfunctie F van $(\underline{w}_1, \ldots, \underline{w}_n)$ volledig bekend en daaruit valt die van $\underline{t}$ dan af te leiden. Notatie:

(13;12) $$F(t \mid H_o) \quad \text{of} \quad F(t \mid \theta_o).$$

De verdere redenering berust grotendeels op deze verdeling.

Voorbeeld 13;1

Stellen wij het aantal malen A voor door:

(13;13) $$\underline{x} = \sum \underline{w}_i ,$$

dan bezit $\underline{x}$, onder H_o , een binomiale verdeling:

(13;14) $$P[\underline{x} = x \mid H_o] = \binom{n}{x} \left(\tfrac{1}{4}\right)^x \left(\tfrac{3}{4}\right)^{n-x} \qquad (x = 0, \ldots, n).$$

De kritieke zone

Onder H_o kan $\underline{t}$ allerlei waarden aannemen. Meestal zijn dit dezelfde waarden als die onder alternatieve hypothesen aangenomen kunnen worden (alleen met andere kansen). Het gebied, dat $\underline{t}$ kan doorlopen, geven wij aan met T :

(13;15) $$P[\underline{t} \in T \mid H_o] = 1.$$

Een toets komt nu daarop neer, dat T in twee of meer gebieden verdeeld wordt en dat de te trekken conclusie afhangt van het gebied, waarin de door $\underline{t}$ aangenomen waarde t terecht komt.

Deze gebieden hebben gewoonlijk de volgende vorm:

(13;16) $t \leqq t_\ell$ leidt tot de conclusie: $\theta < \theta_o$,

(13;17) $t \geqq t_r$ " " " " : $\theta > \theta_o$,

(13;18) $t_\ell < t < t_r$ " " " " : $\theta = \theta_o$ wordt niet verworpen.

Hierbij is verondersteld, dat $\underline{t}$ een schatting van θ is, zodat een grote (kleine) waarde van t een aanwijzing vormt dat θ een grote (kleine) waarde bezit; is de toetsingsgrootheid een monotoon stijgende functie van een schatting van θ , dan blijft de situatie onveranderd, is hij een monotoon dalende functie daarvan, dan verwisselen de conclusies van (13;16) en (13;17). De keuze van t_ℓ en t_r komt later ter sprake.

Het schema (13;16-18) heeft betrekking op een tweezijdige toets, waarbij een enkelvoudige hypothese van de vorm $\theta = \theta_o$ (of een samengestelde van de vorm $\theta_o' \leqq \theta \leqq \theta_o$) getoetst wordt tegen de alternatieve hypothese $\theta < \theta_o$ (resp. $\theta < \theta_o'$) en $\theta > \theta_o$. De gebieden $t \leqq t_\ell$ resp. $t \geqq t_r$ worden ook met Z_ℓ resp. Z_r aangegeven en heten de linker- resp. rechterkritieke zone; zij worden ook tezamen de (tweezijdige) kritieke zone genoemd.

Is de getoetste hypothese van de vorm $\theta \geqq \theta_o$ (resp. $\theta \leqq \theta_o$) dan is de alternatieve hypothese <u>éénzijdig</u>, nl. $\theta < \theta_o$ (resp. $\theta > \theta_o$) en dan wordt het schema als volgt:

(13;19) $t \leqq t_\ell$ leidt tot de conclusie: $\theta < \theta_o$,

(13;20) $t > t_\ell$ " " " " : $\theta \geqq \theta_o$ wordt niet verworpen,

(<u>linkséénzijdige toets</u>), resp.

(13;21) $t \geqq t_r$ leidt tot de conclusie: $\theta > \theta_o$,

(13;22) $t < t_r$ " " " " : $\theta \leqq \theta_o$ wordt niet verworpen,

(<u>rechtséénzijdige toets</u>).

De getallen t_ℓ en t_r heten de <u>linker</u>- resp. <u>rechter kritieke waarden</u>.

<u>Voorbeeld 13;1</u>

De getoetste hypothese is: $p = 1/4$. Worden nu zeer weinig proeven met A als uitkomst verkregen (dus kleine x), dan wordt $p = 1/4$ verworpen ten gunste van $p < 1/4$; worden er echter zeer veel gevonden, dan concludeert men $p > 1/4$. Wij hebben hier dus een geval van een tweezijdige toets. De keuze der kritieke waarden komt later. Het is echter wel reeds intuïtief duidelijk, dat het dwaasheid zou zijn de hypothese $p = 1/4$, bij bijv. $n = 1000$, wel te verwerpen als $x = 780$ is, maar niet als $x > 780$ is. Dit leidt dus inderdaad tot kritieke zones "in de staarten" van het gebied, dat x kan doorlopen.

<u>De onbetrouwbaarheid van een toets</u>

Indien men op deze wijze te werk gaat is het klaarblijkelijk niet uitgesloten, dat H_o verworpen wordt, als deze hypothese juist is. Deze foute uitkomst (een juiste hypothese verwerpen) wordt een <u>fout van de eerste soort</u> genoemd. Er is slechts één methode om deze fout met zekerheid uit te sluiten en dat is: H_o nooit verwerpen. Immers, zodra men een niet-lege kritieke zone gebruikt is er, ook als H_o juist is, een positieve kans H_o te verwerpen.

Deze <u>kans op een fout van de eerste soort</u> wordt kortweg de <u>onbetrouwbaarheid</u> α_o van de toets genoemd. Deze α_o mag uiteraard niet groot zijn en men eist daarom, dat de kritieke zone zo gekozen wordt, dat voldaan is aan

(13;23) $$\alpha_o \leqq \alpha,$$

waarin α een door de onderzoeker gekozen waarde bezit.

De grootheid α wordt de onbetrouwbaarheidsdrempel genoemd. Hiervoor wordt veelal de waarde 0,05 genomen, zonder dat daarvoor een andere fundering bestaat dan de traditie. Indien een fout van de eerste soort ernstige gevolgen heeft kiest men ook wel kleinere waarden, bijv. 0,01 of 0,001.

Het onderscheid tussen α_0 en α lijkt wellicht overbodig. Waarom niet $\alpha_0 = \alpha$ gemaakt? Inderdaad; dit doet men steeds als dat mogelijk is. Maar als de toetsingsgrootheid t onder H_0 een discrete verdeling bezit is dit in de regel niet mogelijk. Men neemt dan α_0 zo dicht bij α als deze verdeling en (13;23) toestaan.

Voorbeeld 13;1

Als $n = 50$ is en $H_0: p = 1/4$, dan ziet de verdelingsfunctie $F(x) = P[X \leqq x]$ er - in 3 decimalen berekend - volgens (13;14) uit als in tabel 13;1. In deze tabel is, behalve $F(x)$, ook opgenomen $R(x) = P[X \geqq x] = 1 - F(x-1)$.

Tabel 13;1

$F(x)$ en $R(x)$ van de binomiale verdeling met $n = 50$, $p = 1/4$

x	$F(x)$	$R(x)$
0	0,000001	1
1	0,00001	≈ 1
2	0,0001	≈ 1
3	0,0005	0,9999
4	0,002	0,9995
5	0,007	0,998
6	0,019	0,993
7	0,045	0,981
8	0,092	0,955
9	0,164	0,918
10	0,262	0,836
11	0,382	0,738
12	0,410	0,618
13	0,637	0,490
14	0,748	0,363
15	0,837	0,252
16	0,902	0,163
17	0,945	0,198
18	0,971	0,055
19	0,986	0,029
20	0,994	0,014
21	0,997	0,006
22	0,999	0,003
23	0,9996	0,001
24	0,9998	0,0004
25	0,9999	0,0002
26,...,50	≈ 1	≈ 0

(Deze tabel is ontleend aan H.G. ROMIG, 50-100 Binomial Tables, Wiley, N.Y., Chapman and Hall, London, 1953).

Nemen wij nu voor Z_ℓ, de linkerkritieke zone, $x \leqq 6$ en voor Z_r, de rechterkritieke zone, $x \geqq 19$, dan is volgens deze tabel:

$$\alpha_o = \alpha_\ell + \alpha_r = P[x \leqq 6 \mid H_o] + P[x \geqq 19 \mid H_o] =$$

$$= 0{,}019 + 0{,}029 = 0{,}048.$$

Is de onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ en wenst men de beide delen van de kritieke zone ongeveer symmetrisch te behandelen - d.w.z. allebei ongeveer $\frac{1}{2}\alpha = 0{,}025$ tot de onbetrouwbaarheid te laten bijdragen, dan is dit de meest geschikte kritieke zone. Wij zien, dat $\alpha_o < \alpha$ is.

Voor $\alpha = 0{,}10$ wordt $Z_\ell : x \leqq 7$ en $Z_r : x \geqq 18$.
In dit geval is (toevalligerwijze) $\alpha_o = \alpha$.

Voor $\alpha = 0{,}01$ is: $Z_\ell : x \leqq 5$ en $Z_r : x \geqq 22$ met $\alpha_o = \alpha = 0{,}01$; een andere mogelijkheid is:

$$Z_\ell : x \leqq 4 \quad \text{en} \quad Z_r : x \geqq 21 \quad , \text{ met } \alpha_o = 0{,}008.$$

Een algemeen aanvaarde regel, om in dergelijke gevallen tussen de verschillende mogelijkheden te kiezen, is niet beschikbaar.

<u>Linker- en rechter-onbetrouwbaarheid</u>

Zoals uit de behandeling van het voorbeeld al blijkt, bestaat de onbetrouwbaarheid van een tweezijdige toets uit twee delen, behorende bij het linker- en rechter- deel van de kritieke zone afzonderlijk. Wij hebben:

(13;24) $$\alpha_o = \alpha_\ell + \alpha_r,$$

met

(13;25) $$\alpha_\ell = P[t \leqq t_\ell \mid H_o]$$ (<u>linkeronbetrouwbaarheid</u>)

en

(13;26) $$\alpha_r = P[t \geqq t_r \mid H_o]$$ (<u>rechteronbetrouwbaarheid</u>).

Men noemt een toets <u>symmetrisch tweezijdig</u> als voor t_ℓ en t_r de grootste resp. kleinste waarde van t genomen wordt, waarvoor α_ℓ en α_r beide $\leqq \frac{1}{2}\alpha$ zijn.

De in het voorbeeld genoemde tweezijdige kritieke zones zijn dus niet geheel symmetrisch, maar komen daar wel bij in de buurt.

Als t een continue verdeling bezit (hetgeen bij het beschouwde voorbeeld niet het geval is), werkt men in het tweezijdige geval gewoonlijk met symmetrische kritieke zones. Ook als men benaderingsmethoden gebruikt wordt dit gewoonlijk gedaan, zoals verderop in dit hoofdstuk ook zal geschieden.

Het onderscheidingsvermogen (Eng.: powerfunction)

Het ten onrechte verwerpen van een (juiste) hypothese - een fout van de eerste soort - vindt zijn tegenhanger in het ten onrechte niet verwerpen van een (onjuiste) hypothese - een fout van de tweede soort -.

Een fout van de eerste soort kan alleen gemaakt worden als H_o juist is en de kans daarop is in dat geval gelijk aan α_o .

Een fout van de tweede soort is alleen mogelijk als H_o onjuist is. In dat geval is een alternatieve hypothese $H \neq H_o$ juist en de kans op het maken van een fout van de tweede soort is afhankelijk van deze H . Deze kans is:

$$1 - \pi(H) = 1 - P[t \in Z \mid H], \tag{13;27}$$

waarin Z de kritieke zone voorstelt. Immers als t in Z ligt, wordt H_o (niet H) verworpen en dan wordt dus geen fout van de tweede soort gemaakt. De grootheid $\pi(H)$ wordt het onderscheidingsvermogen van de toets genoemd. Dit is een functie van H waarvoor geldt:

$$\pi(H_o) = \alpha_o. \tag{13;28}$$

Voor $H \neq H_o$ is $\pi(H)$ de kans op verwerping van H_o als H juist is en H_o dus ook verworpen dient te worden. Dit dus, om zo te zeggen, de kans, dat de toets H van H_o zal weten te onderscheiden.

Bij een tweezijdige toets zijn drie uitkomsten mogelijk, nl. - als θ de onbekende parameter is en θ_o de getoetste waarde - de drie uitspraken:

$\theta < \theta_o$,

$\theta > \theta_o$,

en: $\theta = \theta_o$ wordt niet verworpen.

De kansen op deze drie uitspraken geven wij aan met $\pi_\ell(H)$, $\pi_r(H)$ en $1-\pi(H)$, zodat dus

(13;29) $$\pi_\ell(H) + \pi_r(H) = \pi(H)$$

is. De beide termen heten het linker- resp. rechteronderscheidingsvermogen.

Voorbeeld 13;1

Met behulp van een tabel van de binomiale verdeling kan men nu gemakkelijk het onderscheidingsvermogen van een binomiale toets - als functie van de onbekende kans p - bepalen. Wij doen dit voor de kritieke zones

en $$\left.\begin{array}{ll} Z_\ell: & x \leqq 6 \quad \text{met} \quad \alpha_\ell = 0{,}019 \\ Z_r: & x \geqq 19 \quad \text{met} \quad \alpha_r = 0{,}029 \end{array}\right\} \; n = 50 \quad ; p_0 = 1/4.$$

In de reeds genoemde tabel van ROMIG kunnen wij nu voor verschillende waarden van p de waarde van

(13;30) $$\pi_\ell(p) = P[x \leqq 6 \mid p]$$

en van

(13;31) $$\pi_r(p) = P[x \geqq 19 \mid p] = 1 - P[x \leqq 18 \mid p]$$

opzoeken. Deze waarden staan in tabel 13;2 vermeld, evenals hun som: $\pi(p)$. In fig. 13;1 is deze tabel grafisch in beeld gebracht.

Tabel 13;2

Het onderscheidingsvermogen van de binomiale toets met $n=50$ en $Z_\ell: x \leqq 6$, $Z_r: x \geqq 19$.

p	$\pi_\ell(p)$	$\pi_r(p)$	$\pi(p)$
0,05	0,988	0,000	0,988
0,10	0,770	0,000	0,770
0,15	0,361	0,000	0,362
0,20	0,103	0,003	0,106
0,21	0,076	0,004	0,080
0,22	0,056	0,001	0,064
0,23	0,040	0,012	0,032
0,24	0,028	0,019	0,047
0,25	0,019	0,029	0,048
0,26	0,013	0,042	0,055
0,30	0,002	0,141	0,143
0,40	0,000	0,664	0,664
0,50	0,000	0,966	0,966

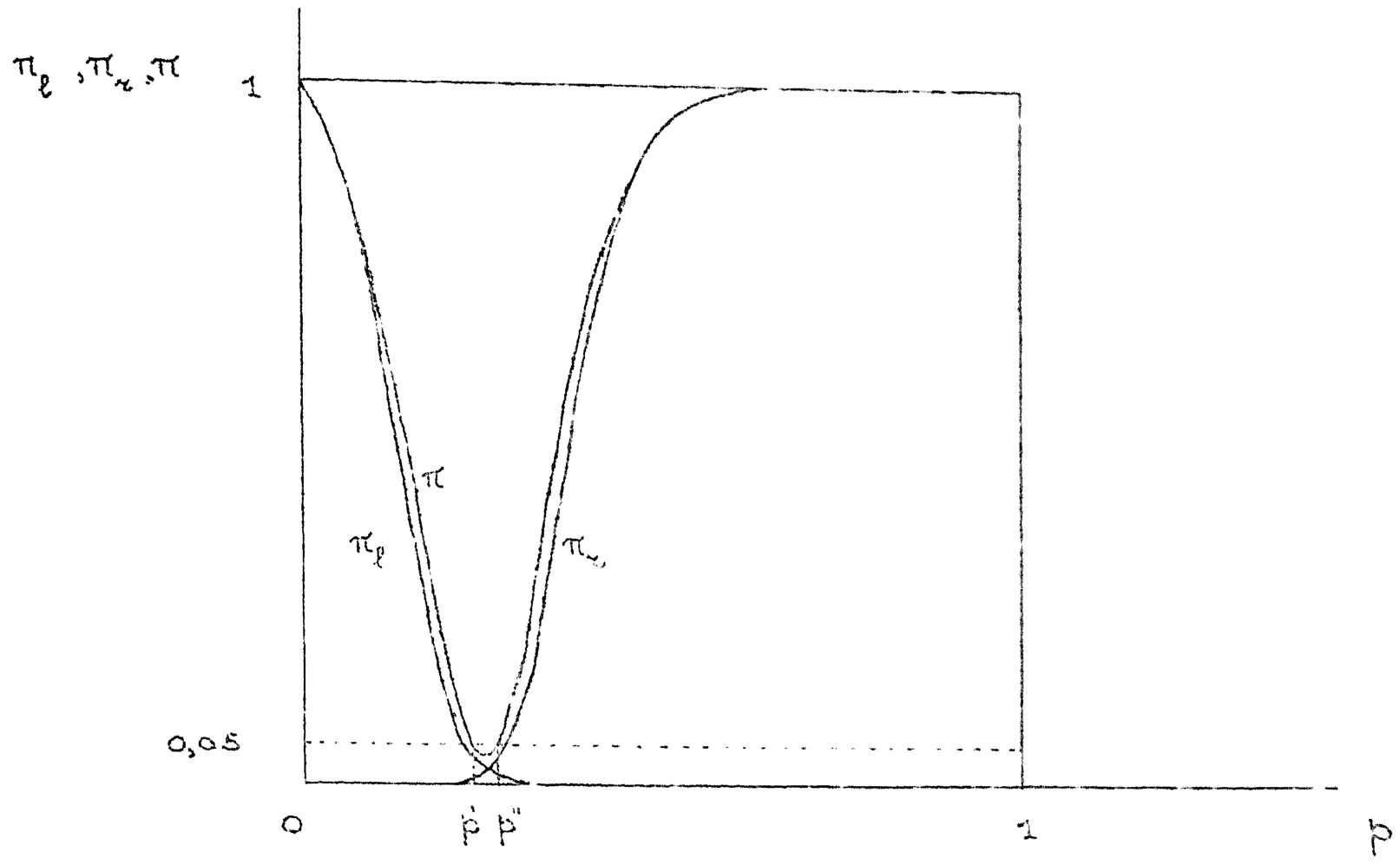


Fig. 13;1. Het onderscheidingsvermogen van de binomiale toets met $n = 50$, $Z_\ell : x \leqq 6$ en $Z_r : x \geqq 19$.

Bij tabel 13;2 en fig. 13;1 zijn nog verschillende opmerkingen te maken.

1e. Hoe steiler het onderscheidingsvermogen verloopt, hoe beter de toets is. Dit volgt direct uit de betekenis van deze functie: de kans op verwerping van H_o, indien H_o onjuist is. Men kan nu bewijzen - hetgeen wij hier nalaten - dat voor toenemende n, bij constante α, de beide krommen π_ℓ en π_r tot limiet een verticale lijn door $p = 1/4$ als abscis hebben. Wordt een andere waarde, $p = p_o$, getoetst, dan worden daarvoor uiteraard andere kritieke zones gebruikt en dan geldt deze stelling analoog, nu met een verticale lijn door $p = p_o$ als limiet.

2e. De toets, die hier geschreven is als een toets voor de hypothese $p = 1/4$ kan evengoed (of zelfs beter) beschouwd worden als een toets voor $p = 0,24$, met $\alpha_o = 0,047$. In feite kan men, met $\alpha = 0,05$, zeggen dat het een toets is voor de samengestelde hypothese $p' \leqq p \leqq p''$, waarin p' en p'' de abscis voorstellen van de snijpunten van de horizontale rechte $\pi = 0,05$ met de kromme $\pi(p)$. Want voor alle tussenliggende waarden van p is de kans dat $\leq$ in Z_ℓ of Z_r valt $\leqq \alpha$. Deze fijne onderscheidingen worden gewoonlijk niet gemaakt.

3e. De kromme π_ℓ stelt de kans voor, dat tot $p < 0,25$ besloten wordt. Deze kans is ook > 0 als p in werkelijkheid $> 0,25$ is, maar hij neemt (gelukkig) met toenemende p sterk af. Voor $p = 0,30$ stelt nu $\pi_\ell (0,30) = 0,002$ de kans voor om te besluiten tot $p < 0,25$ een faliekant verkeerde conclusie; $\pi_r (0,30) = 0,141$ is daarentegen juist de kans op de conclusie $p > 0,25$, die dan juist is.

De som $\pi(0,30)$ van $\pi_\ell(0,30)$ en $\pi_r(0,30)$ is dus eigenlijk de som van de kansen op twee volkomen ongelijksoortige conclusies, nl. een falikant onjuiste en een juiste. Vanuit het standpunt van de toetsingstheorie bezien is het daarom beter de merites van een toets te beoordelen op grond van π_ℓ en π_r afzonderlijk dan op grond van hun som.

4e. Is de toets éénzijdig, dan ontbreekt één van de twee krommen π_ℓ of π_r .

5e. Men kan bewijzen, dat bij gegeven n, p_0 en α het grootste onderscheidingsvermogen verkregen wordt als de kritieke zones "in de staarten" genomen worden, zoals wij hier gedaan hebben. Zou men een kritieke zone nemen van het type:

$$Z : 0 < x_1 \leqq x \leqq x_2 < n$$

dan verkrijgt men een toets met een zeer ongunstig onderscheidingsvermogen, zoals wij later met een voorbeeld zullen laten zien.

Zuiverheid en asymptotische onderscheidendheid

Definitie 13;1

Een toets wordt <u>zuiver</u> genoemd, als het onderscheidingsvermogen voor alle alternatieve hypothesen groter is dan de onbetrouwbaarheid.

De practische betekenis is, dat bij een zuivere toets de kans om de getoetste hypothese te verwerpen als deze juist is kleiner is dan wanneer deze onjuist is.

Definitie 13;2

Een toets voor de hypothese H_0 heet <u>asymptotisch onderscheidend</u> met betrekking tot een alternatieve hypotnese H_1 , als geldt:

$$\lim_{n\to\infty} \pi(H_1) = 1. \qquad (13;32)$$

De practische betekenis hiervan is, dat men, als H_1 juist is, bij een voldoend groot aantal waarnemingen vrijwel zeker is dat de toets (terecht) tot verwerping van H_0 zal leiden.

Eénzijdige toetsen zijn vrijwel altijd zuiver. Het onderscheidend vermogen is dan nl. meestal een monotone kromme, zoals in fig. 13;1 (π_ℓ en π_r afzonderlijk). Tweezijdige toetsen zijn lang niet altijd zuiver, ook niet als dit voor de beide éénzijdige wel geldt.

Toetsen, die voor relevante alternatieve hypothesen niet asymptotisch onderscheidend zijn, zijn voor de practijk in de regel onbruikbaar.

Voorbeeld 13:1

Uit fig. 13;1 blijkt, dat in dit geval de tweezijdige toets, beschouwd als toets voor de hypothese $p = 1/4$, niet zuiver is: immers voor $p = 0,24$ is het onderscheidingsvermogen kleiner dan de onbetrouwbaarheid (vgl. tabel 13;2). Beschouwen wij echter $H_o: p' \leqq p \leqq p''$ als de getoetste hypothese, dan is de toets wel zuiver.

De onzuiverheid ontstaat door 2 oorzaken:

1e: het discrete karakter van de verdeling van de toetsingsgrootheid x, waardoor men de beide delen van de kritieke zone niet volledig "tegen elkaar afwegen" kan,

2e: doordat de verdeling van x, onder H_o, asymmetrisch is, hetgeen het zuiver maken van de toets weliswaar niet onmogelijk maakt, maar wel bemoeilijkt.

Heeft de toetsingsgrootheid onder H_o een symmetrische verdeling, dan behoeft men de tweezijdige kritieke zone slechts symmetrisch te nemen om zuiverheid te bereiken ook voor de tweezijdige toets.

De toets is asymptotisch onderscheidend voor alle van 1/4 verschillende waarden van p; dit is boven (onder opmerkingen, 1e) reeds - zonder bewijs - vermeld.

Overschrijdingskans

De bij een waarde t van de toetsingsgrootheid behorende linker resp. rechteroverschrijdingskans is:

$$(13;33)\quad k_\ell(t) \overset{def}{=} P[t \leqq t_\ell \mid H_o] \quad \text{resp.} \quad k_r(t) \overset{def}{=} P[t \geqq t_r \mid H_o].$$

Is

$$(13;34)\quad k_\ell(t) \leqq \alpha_\ell \quad \text{resp.} \quad k_r(t) \geqq \alpha_r$$

dan ligt t in Z_ℓ resp. Z_r.

De tweezijdige overschrijdingskans $k(t)$, behorende bij een symmetrische tweezijdige toets, wordt gedefinieerd als het dubbele van de kleinste der beide eenzijdige:

$$(13;35)\quad k(t) \overset{def}{=} 2 \min\{k_\ell(t), k_r(t)\}.$$

Dan geldt: is

(13;36) $$k(t) \leqq \alpha,$$

dan is $t \in Z$. Voor een niet-symmetrische wordt de tweezijdige overschrijdingskans op iets andere wijze gedefinieerd.

Voorbeeld 13;1

Volgens tabel 13;1 is $k_\ell(3)=0,0001$, $k_r(21)=0,006$, $k(4)=0,004$.

De betekenis van het begrip overschrijdingskans is, dat men aan de grootte ervan niet alleen kan zien of de beschouwde waarde van t in de kritieke zone van de toets ligt, maar ook of dit maar juist of ruim het geval is. Hoe kleiner de gevonden overschrijdingskans is, des te stelliger kan men de getoetste hypothese verwerpen.

De belangrijkste elementaire begrippen van de toetsingstheorie zijn hiermede in het kort behandeld. Zij vormen tezamen weliswaar slechts een begin van de toetsingstheorie, maar zij zijn voldoende om een aantal toetsen op te baseren. Wij behandelen verder in dit hoofdstuk alleen de in de practijk veelvuldig gebruikte techniek van de binomiale toetsen met behulp van de normale verdeling als benadering van de binomiale.

Binomiale toetsen met normale benadering

In hoofdstuk 11 (Stelling 11;8 en figuur 11;1) hebben wij gezien, dat de binomiale verdeling met behulp van de aangepaste normale verdeling benaderd kan worden. Wij brengen nu op deze benadering nog een verfijning aan, die bekend staat onder de naam continuïteitscorrectie. Deze correctie heeft betrekking op het feit, dat de continue normale verdeling gebruikt wordt om de discrete binomiale te benaderen.

Indien n en p de parameters van de binomiale verdeling zijn, zodat

(13;37) $$P[\underline{x} = x] = \binom{n}{x} p^x q^{n-x}$$

is, met $\mathcal{E}\,\underline{x} = np$ en $\sigma^2(\underline{x}) = npq$, dan wordt de normale verdeling met dezelfde verwachting en variantie voor de benadering gebruikt.

Een discrete kans $P[\underline{x} = x]$ moet nu benaderd worden door een oppervlak onder de normale verdelingsdichtheid en het ligt voor de hand hiervoor het oppervlak boven het interval $(x - \frac{1}{2}, x + \frac{1}{2})$ te nemen, dus symmetrisch om de waarde x gelegen (vgl. fig. 13;2). Alleen voor $P[\underline{x} = o]$ en voor $P[\underline{x} = n]$ neemt men de staart er volledig bij, dus dan neemt men het "normale" oppervlak boven het interval $(-\infty, \frac{1}{2})$ resp. $(n - \frac{1}{2}, \infty)$, omdat anders de som der benaderende kansen kleiner dan 1 zou zijn.

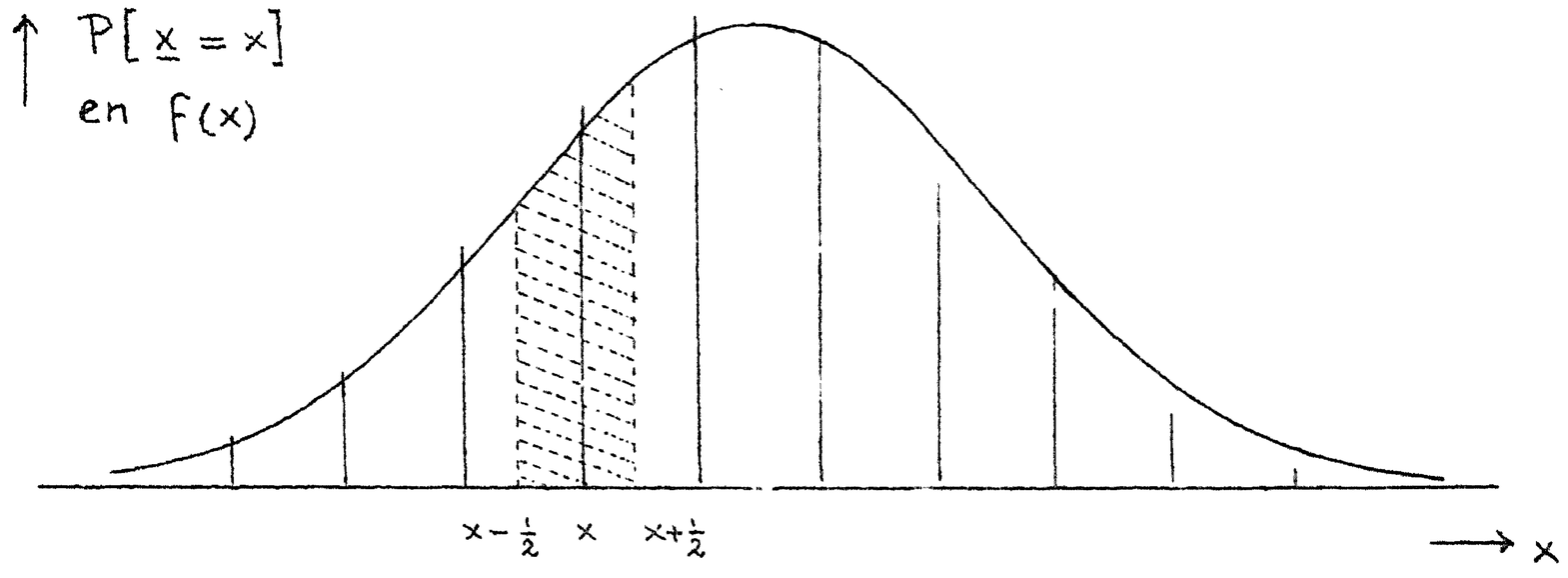


Fig. 13;2. Normale benadering van een binomiale verdeling met continuïteitscorrectie.

Wenst men nu de <u>linkeroverschrijdingskans</u> $P[\underline{x} \leqq x]$ te benaderen, dan neemt men daarvoor dus het "normale" oppervlak boven het interval $(-\infty, x + \frac{1}{2})$, voor de <u>rechteroverschrijdingskans</u> $P[\underline{x} \geqq x]$ dat boven het interval $(x - \frac{1}{2}, \infty)$. Deze benaderende oppervlakken kunnen, na standaardisering, uit tabel 1 van de normale verdeling afgelezen worden. De term $\frac{1}{2}$ is de continuiteitscorrectie.

In formule wordt dit:

$$(13;38) \qquad k_\ell(x) = P[\underline{x} \leqq x \mid n,p] \approx P\left[\underline{u} \leqq \frac{x - np + \frac{1}{2}}{\sqrt{npq}}\right]$$

resp.

$$(13;39) \qquad k_r(x) = P[\underline{x} \geqq x \mid n,p] \approx P\left[\underline{u} \geqq \frac{x - np - \frac{1}{2}}{\sqrt{npq}}\right]$$

en voor de <u>tweezijdige overschrijdingskans</u>:

$$(13;40) \qquad k(x) \approx 2\,P\left[\underline{u} \geqq \frac{|x - np| - \frac{1}{2}}{\sqrt{npq}}\right].$$

Voorbeeld 13;1

Voor $n = 50$ en $p = 1/4$ geldt volgens tabel 13;1

$$k_x(18) = P[\underline{x} \geqq 18 \mid n = 50, p = 1/4] = 0{,}055.$$

Wij passen nu de normale benadering met continuïteits-correctie toe om deze zelfde overschrijdingskans te benaderen:

$\mathcal{E}\underline{x} = np = 12{,}5$, $\sigma^2(\underline{x}) = npq = 9{,}375$, dus $\sigma(\underline{x}) = 3{,}06$.

Derhalve:

$$P[\underline{x} \geqq 18] = P\left[\frac{\underline{x} - 12{,}5}{3{,}06} \geqq \frac{18 - 12{,}5}{3{,}06}\right] \approx P\left[\underline{u} \geqq \frac{17{,}5 - 12{,}5}{3{,}06}\right] =$$

$$= P[\underline{u} \geqq 1{,}63] = 0{,}0516.$$

Zonder continuïteitscorrectie zouden wij vinden

$$P[\underline{x} \geqq 18] \approx P\left[\underline{u} \geqq \frac{18 - 12{,}5}{3{,}06}\right] = P[\underline{u} \geqq 1{,}79] = 0{,}0367.$$

De benadering met continuïteitscorrectie is beter, hetgeen echter niet altijd het geval is. Wel geeft de benadering met continuïteitscorrectie altijd een grotere uitkomst voor een overschrijdingskans dan zonder, zodat minder gauw tot verwerping van de getoetste hypothese wordt overgegaan. In het onderhavige geval zou bij rechtséénzijdige toetsing met onbetrouwbaarheids-drempel 0,05 bij gebruik van de exacte overschrijdingskans of de benadering met continuïteitscorrectie op grond van $x = 18$ niet tot verwerping van $H_0: p = 1/4$ worden overgegaan, maar zonder continuïteitscorrectie wel. Het is daarom "voorzichtiger" de continuïteitscorrectie te gebruiken dan deze weg te laten. Voor grote n is het effect echter gering, zodat weglating dan wel gerechtvaardigd geacht mag worden.

Opgave 13;1

Bereken voor $n = 50$, $p = 1/4$ de linkeroverschrijdingskans van $x = 6$; exact, met en zonder continuïteitscorrectie

(Antwoord: 0,019; 0,025; 0,017.)

Opgave 13;2

Bereken voor $n = 900$ en $p = 0{,}5$ de tweezijdige overschrijdingskans van de waarde $x = 427$ met en zonder continuïteits-correctie.

Opgave 13;3

Bij 225 onafhankelijke experimenten, die alle dezelfde kans p op succes hebben, worden 85 successen gevonden.

a) Toets de hypothese, dat $p=\frac{1}{2}$ is, met onbetrouwbaarheidsdrempel 0,05 .

b) Toets, met dezelfde onbetrouwbaarheidsdrempel, de hypothese dat $p=1/3$.

Voer beide toetsen tweezijdig en linkséénzijdig uit.
(Opgave 5 van het sommenboekje; oplossing blz. 50-51.)

Opgave 13;4

Van een binomiale verdeling is $n=10$ en $p=0,3$.

a) Bereken met behulp van de normale benadering van de binomiale verdeling de linkeroverschrijdingskans van $x=1$, met en zonder continuïteitscorrectie.

b) Bereken deze kans ook exact en bereken vervolgens hoe groot de continuïteitscorrectie in dit geval genomen zou moeten worden, om te bereiken dat de benaderende en de exacte overschrijdingskans precies gelijk zijn.
(Opgave 26 van het sommenboekje; oplossing blz 71.)

Opgave 13;5

Een organisator van een vuurwerk weet uit ervaring, dat van de vuurpijlen, die hij gewoonlijk gebruikt, ongeveer één op de 5 weigert. In verband daarmee maakt hij een aantal reserve pijlen gereed. Hij wenst het publiek minstens 20 goede vuurpijlen te laten zien en hij aanvaardt hoogstens een kans van 1/100, dat hem dit niet zal lukken. Hoeveel reserve pijlen moet hij opstellen?
(Opgave 40 van het sommenboekje; oplossing blz. 83.)

Opgave 13;6

Wat is, voor $n=50$ en $H_0: p=1/4$, de onbetrouwbaarheid van de kritieke zone, die uit de waarden 18, 19 en 20 bestaat?

Bereken het onderscheidingsvermogen van deze kritieke zone voor $p=0,2$; 0,4; 0,6; 0,8 en 1 met behulp van de normale benadering en vergelijk dit met π_x in fig. 13;1 en tabel 13;2.
(Antwoord: $\alpha_0=0,049$,

p	0,2	0,25	0,4	0,6	0,8	1
$\pi \approx$	0,007	0,049	0,37	0,001	0	0.)

Ook voor de kritieke waarden zijn nu eenvoudige benaderingsformules af te leiden. Is u_α de waarde van een $N(0,1)$ -grootheid, waarvoor

(13;41) $$P[\underline{u} \geq u_\alpha] = \alpha$$

is - u_α is dus in tabel 1 te vinden voor $k_\alpha = \alpha$ - dan zijn de kritieke waarden van de linker- en rechtereenzijdige toets

(13;42) $$x_l \approx np_0 - \left\{ u_{\alpha_l}\sqrt{np_0q_0} + \tfrac{1}{2} \right\}$$

resp.

(13;43) $$x_r \approx np_0 + \left\{ u_{\alpha_r}\sqrt{np_0q_0} + \tfrac{1}{2} \right\},$$

terwijl voor de symmetrische tweezijdige toets geldt:

(13;44) $$x_{r,l} \approx np_0 \pm \left\{ u_{\frac{1}{2}\alpha}\sqrt{np_0q_0} + \tfrac{1}{2} \right\}.$$

De termen $\frac{1}{2}$ stellen de continuïteitscorrectie voor.

Opgave 13;7

Geef de afleiding van deze formules.

Opgave 13;8

Hoe groot moet men het aantal experimenten minstens nemen om bij rechtséénzijdige toetsing van de hypothese $p = 0,3$ met $\alpha_r = 0,05$ voor $p=0,6$ een onderscheidingsvermogen van minstens 0,9 te bereiken? (De continuïteitscorrectie mag verwaarloosd worden. Houd bij de oplossing rekening met het feit, dat de toetsingsgrootheid alleen gehele waarden aan kan nemen.)
(Opgave 58 van het sommenboekje; oplossing blz. 97-98.)

De keuze tussen één- of tweezijdige toetsen hangt af van het doel van het onderzoek en van de toegelaten hypothesen. Wij zullen hierover kort zijn.

Indien men een beslissing moet nemen tussen twee handelwijzen, bijv. het al of niet aanschaffen van een nieuwe machine, dan zal men daartoe alleen wensen te besluiten als de nieuwe machine beter is dan de oude. Is nu van de oude machine bekend, dat deze een fractie p_0 aan uitval geeft, dan toetst men de hypothese $H_0: p \geq p_0$ (waarin p de fractie uitval van de nieuwe machine voorstelt) tegen de alternatieve hypothese $p < p_0$.

Alleen bij verwerping van H_o gaat men tot aanschaf van de nieuwe machine over. De onbetrouwbaarheid α_x van de toets stelt dan de kans voor, dat men de nieuwe machine toch zal aanschaffen indien de oude even goed is. Is de oude beter, dan is de kans op deze foute beslissing kleiner dan α_x. Een tweezijdige toets zou hier niet op zijn plaats zijn, daar verwerping van $H_o: p = p_o$ ten gunste van $p > p_o$ evenmin tot aanschaffing van de nieuwe machine leidt als niet-verwerpen van H_o. En het onderscheidingsvermogen van de rechtséénzijdige toets is, bij dezelfde onbetrouwbaarheidsdrempel, voor $p < p_o$ het grootst voor de rechtséénzijdige toets, zodat deze een kleinere kans dan de tweezijdige geeft om een fout van de tweede soort te maken (hetgeen hier zou zijn: de nieuwe machine niet aanschaffen, hoewel hij beter is dan de oude).

Het kan ook voorkomen, dat hypothesen $p > p_o$ onmogelijk geacht worden. Deze situatie doet zich bijv. voor bij het toedienen van middelen, waarvan men met recht kan zeggen: "baat het niet, dan schaadt het toch in ieder geval niet". In zo'n geval is het gebruik van een tweezijdige kritieke zone zinloos, daar het rechterdeel hiervan behoort bij de dan zeker onjuist conclusie, dat $p > p_o$ is.

Bij de meeste wetenschappelijke onderzoekingen zal men echter met beide mogelijkheden ($p < p_o$ en $p > p_o$) rekening moeten houden en dus tweezijdig moeten toetsen.

Hoofdstuk 14. Enige gangbare toetsen.

14.1;1. De 2x2-tabel (dubbele dichotomie).

In het vorige hoofdstuk zijn toetsen behandeld, die op de binomiale verdeling berusten en waarmee de hypothese H_o: $p = p_o$ getoetst kan worden. Wij behandelen nu een toets voor de hypothese

(14;1;1) H_o: $p_1 = p_2$,

inhoudende, dat twee onbekende kansen gelijk zijn.

Dit is een samengestelde hypothese. De parameterruimte (zie fig.14.1;1) is een deel van het (p_1, p_2)-vlak begrensd door de ongelijkheden $0 \leqq p_1 \leqq 1$; $0 \leqq p_2 \leqq 1$ en H_o wordt voorgesteld door de diagonaal van dat gebied.

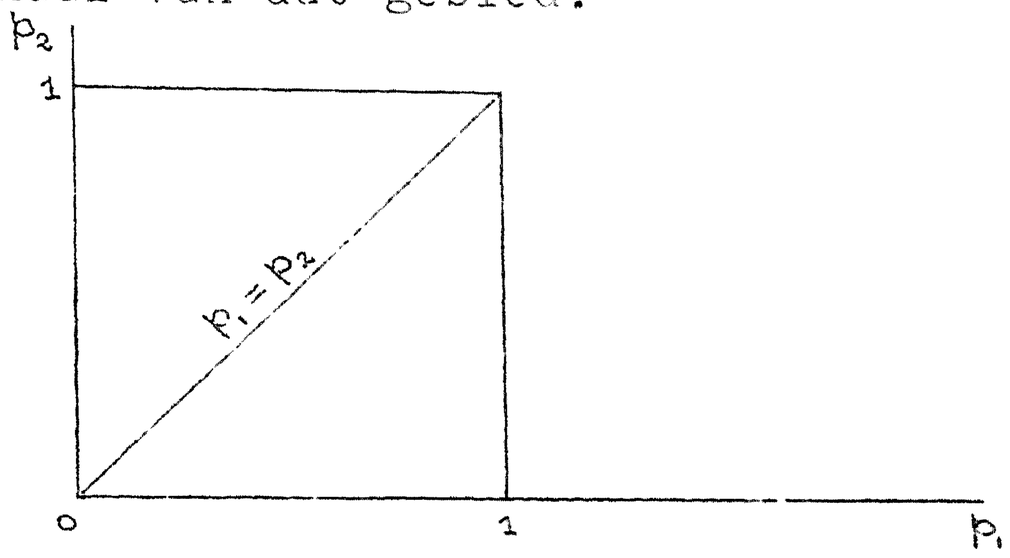


Fig.14.1;1. De parameterruimte en H_o van de toets voor gelijkheid van twee onbekende kansen.

De toets, die wij gaan beschrijven, kan toegepast worden indien het waarnemingsmateriaal bestaat uit twee onafhankelijke reeksen van onafhankelijke experimenten, waarbij p_1 resp. p_2 de kansen op "succes" bij de experimenten van de eerste resp. tweede reeks voorstellen. Zijn de aantallen experimenten n resp. m en de aantallen successen a resp. b, dan kunnen deze resultaten samengevat worden in een tabel van de vorm van tabel 14.1;1.

Tabel 14:1:1

2x2-tabel voor $H_o: p_1 = p_2$

	Aantal malen		
	succes	mislukking	totaal
eerste reeks	a	c	n
tweede reeks	b	d	m
totaal	r	s	N

Hierin is:

(14;1;2) $a + c = n$; $b + d = m$; $r = a + b$; $s = c + d$; $N = n + m = r + s$.

Een dergelijke tabel wordt een 2x2-tabel of ook een dubbele dichotomie genoemd.

Als toetsingsgrootheid wordt de stochastische grootheid

$\underline{a}$ genomen. De verdeling van $\underline{a}$, onder H_o, is echter niet ondubbelzinnig bepaald, maar hangt nog af van de onbekende waarde $p_1 (= p_2)$. Deze onbekende parameter kan echter in dit geval geëlimineerd worden door niet met de onvoorwaardelijke verdeling van $\underline{a}$ (onder H_o) te werken, maar met de <u>voorwaardelijke verdeling</u> van $\underline{a}$ onder de voorwaarde $\underline{r} = r$ (hetgeen impliceert $\underline{s} = N - r$). Deze voorwaardelijke verdeling is de hypergeometrische:

$$(14.1;3) \qquad P[\underline{a} = a \mid \underline{r} = r; H_o] = \frac{\binom{n}{a}\binom{m}{b}}{\binom{N}{r}},$$

of ook

$$(14.1;4) \qquad P[\underline{a} = a \mid \underline{r} = r; H_o] = \frac{\binom{r}{a}\binom{s}{c}}{\binom{N}{n}}.$$

<u>Opgave 14.1;1</u>. Bewijs de formules (14.1;3) en (14.1;4).

Zijn nu r en s de bij het beschouwde experiment door $\underline{r}$ en $\underline{s}$ aangenomen waarden, dan maken wij gebruik van de voorwaardelijke verdeling van $\underline{a}$ met deze waarden van $\underline{r}$ en $\underline{s}$ als voorwaarden. De formules (14.1;3) en (14.1;4) stellen ons dan in staat tot exacte berekening van de linker en rechter overschrijdingskans van de gevonden waarde a van a:

$$(14.1;5) \qquad k_\ell(a) = \sum_{\nu=0}^{a} \frac{\binom{n}{\nu}\binom{m}{r-\nu}}{\binom{N}{r}} = \sum_{\nu=0}^{a} \frac{\binom{r}{\nu}\binom{s}{n-\nu}}{\binom{N}{n}}$$

resp.

$$(14.1;6) \qquad k_r(a) = \sum_{\nu=a}^{n} \frac{\binom{n}{\nu}\binom{m}{r-\nu}}{\binom{N}{r}} = \sum_{\nu=a}^{r} \frac{\binom{r}{\nu}\binom{s}{n-\nu}}{\binom{N}{n}}.$$

<u>Opmerking 14.1;1</u>

De grenzen, waartussen a zich kan bewegen, zijn

$$(14.1;7) \qquad \max(n-s, 0) \leq a \leq \min(n, r).$$

Buiten deze grenzen worden de termen van de in (14.1;5 en 6) vermelde sommen gelijk aan 0.

<u>Opgave 14.1;2</u>

Bereken de exacte linkeroverschrijdingskans, behorende bij de volgende 2x2-tabel:

	succes	mislukking
eerste reeks	2	5
tweede reeks	3	2

(Oplossing; 0,001+0,044+0,265=0,310).

Deze exacte berekeningswijze is alleen dan zonder veel

rekenwerk uitvoerbaar, als de aantallen klein zijn. Bij grotere aantallen wordt een benaderingsmethode gebruikt.

Men kan bewijzen, dat $\underline{a}$ voor m, n, r en s $\to \infty$ <u>asymptotisch normaal</u> verdeeld is als H_o juist is.

Volgens (9;26) en (9;33) geldt

(14;9;8) $$\mathcal{E}(\underline{a} \mid H_o) = \frac{nr}{N}$$

en

(14;9;9) $$\sigma^2(\underline{a} \mid H_o) = \frac{mnrs}{N^2(N-1)} .$$

De asymptotische normaliteit houdt in, dat de grootheid

(14;9;10) $$\underline{a}^* = \frac{\underline{a} - \frac{nr}{N}}{\sqrt{\frac{mnrs}{N^2(N-1)}}}$$

bij benadering N(0,1) verdeeld is.

Bij de berekening van overschrijdingskansen past men dan bovendien nog een continuïteitscorrectie $+\frac{1}{2}$ of $-\frac{1}{2}$ toe.

<u>Voorbeeld 14.1;1</u>

Men vermoedt, dat bespuiting met een bepaald chemisch preparaat het vóórkomen van een bepaalde ziekte onder selderieplanten zal verminderen. Om te onderzoeken of dit waar is neemt men twee groepen van 100 planten en behandelt één van deze beide groepen met het middel, de andere niet. Van de behandelde planten vertonen 20 de ziekte, van de onbehandelde 46. Betekent dit dat het middel werkt?

<u>Oplossing</u>:

De 2x2-tabel is:

	ziek	niet ziek	totaal
behandeld	20	80	100
onbehandeld	46	54	100
totaal	66	134	200

$$a = 20 \quad ; \quad \mathcal{E}(\underline{a} \mid H_o) = \frac{100 \cdot 66}{200} = 33;$$

$$\sigma^2(\underline{a} \mid H_o) = \frac{10^4 \cdot 66 \cdot 134}{\$4 \cdot 10^4 \cdot 199} = 11,11 \qquad \sigma(\underline{a} \mid H) = 3,33.$$

Voor de tweezijdige overschrijdingskans berekenen wij dus

$$\frac{\left|a - \frac{nr}{N}\right| - \frac{1}{2}}{\sqrt{\frac{mnrs}{N^2(N-1)}}} = \frac{12,5}{3,33} \approx 3,8.$$

De tweezijdige overschrijdingskans is zeer klein, zodat H_o verworpen wordt ten gunste van de hypothese, dat de behandeling het vóórkomen van de ziekte vermindert.

Opmerkingen

14.1;2. In een geval als het onderhavige kan men van mening verschillen over de vraag of men linkseenzijdig of tweezijdig moet toetsen. Gaat het er alleen om of men tot uitvoering van het middel wil besluiten of niet, dan kan men met linkseenzijdige toetsing volstaan. Indien men echter ook de mogelijkheid onder het oog ziet, dat de planten door de behandeling een verminderde weerstand tegen de ziekte gaan vertonen en stelt men, uit wetenschappelijke interesse, ook belang in die mogelijkheid, dan is tweezijdige toetsing aan te bevelen. In dit geval leiden trouwens beide methoden tot dezelfde conclusie, omdat het resultaat zeer duidelijk is.

14.1;3. De uit het experiment getrokken conclusie is alleen betrouwbaar indien de proef met de nodige voorzorg is opgezet. Is de ziekte b.v. besmettelijk en is het preparaat niet een geneesmiddel voor zieke planten maar een afweermiddel tegen besmetting, dan dienen de 200 planten alle gelijke besmettingskans te hebben. Dit houdt in, dat de twee groepen niet in twee bakken of velden van elkaar gescheiden geplaatst mogen zijn. Immers dan zou een toevallig iets zwaardere besmetting op het niet behandelde veld het grotere aantal zieke planten aldaar veroorzaakt kunnen hebben. In statistisch idioom: dan zijn de proeven binnen ieder der twee waarnemingsreeksen niet onafhankelijk, zodat de toets zijn geldigheid verliest. Hoe kan men dan in zo'n geval de proef zo opzetten, dat de voorwaarden wel vervuld zijn? Dit zou - met excuses voor de bewerkelijkheid van de proef - b.v. als volgt kunnen geschieden. De 200 planten worden ieder in een bakje geplant ((of gezaaid), waarin nog een tweede plant geplaatst kan worden. Deze 200 bakjes, alle met de zelfde aarde gevuld, worden zo opgesteld, dat ze zoveel mogelijk aan de zelfde weersomstandigheden zijn blootgesteld. Door loting worden 100 bakken voor behandeling aangewezen en deze behandeling wordt uitgevoerd. Daarna wordt in ieder der bakken, op gelijke afstand van de proefplant, een zieke plant geplaatst. Alle zieke planten worden zoveel mogelijk van gelijk formaat en van gelijke ziektegraad genomen en zij worden ofwel vóór de bovengenoemde loting ieder aan een bepaalde bak toegewezen ofwel door een nieuwe loting over de bakken verdeeld.

De aselecte verdeling van de zieke planten en van de behandeling over de bakken, tezamen met het vóorkómen van besmetting van één bak naar een andere ((een boven nog niet vermelde, maar essentiële voorwaarde), garanderen dat een eventueel optredend systematisch verschil in ziektefrequentie($p_1 \neq p_2$) inderdaad al-

leen een gevolg kan zijn van het enige systematische verschil dat nu nog tussen de beide groepen bestaat: het al of niet behandelen. De kans op een verkeerde conclusie van de vorm $p_1 < p_2$ of $p_1 > p_2$ is dan dus inderdaad gelijk aan de onbetrouwbaarheid van de toets. Worden dergelijke voorzorgen niet genomen, dan is deze kans echter volkomen onbekend.

Wij vermelden de volgende eigenschappen van deze toets zonder bewijs.

a) De toets bezit maximaal onderscheidingsvermogen, d.w.z. men kan geen toets ontwerpen met een grotere kans op verwerping van H_o indien $p_1 \neq p_2$ is. Het onderscheidingsvermogen is een functie van p_1, p_2, m en n van een vrij ingewikkeld karakter. Bij gegeven p_1 en p_2 (met p_1 / p_2) en N is het maximaal als $m=n=\frac{1}{2}N$ is. Bij het opzetten van proeven kan men daar rekening mee houden.

b) De eenzijdige toetsen zijn zuiver.

c) De toets is asymptotisch (voor m en $n \to \infty$) onderscheidend voor alle alternatieven (p_1, p_2) met $p_1 \neq p_2$.

De zelfde toets kan ook gebruikt worden als toets voor de onafhankelijkheid van twee kenmerken. Beschouwen wij een populatie, waarvan de elementen de kenmerken A en B al of niet bezitten, en nemen wij daaruit een steekproef van N elementen, dan kunnen wij de resultaten rangschikken in de volgende 2x2-tabel.

Tabel 14.1;2

2x2-tabel voor onafhankelijkheidstoets

	A	niet A	totaal
B	a	b	m
niet B	c	d	n
totaal	r	s	N

Hierin zijn nu alle randtotalen stochastisch behalve N $(= m+n = r+s)$.

De te toetsen hypothese is nu:

(14.1;11) H_o: A en B stochastisch onafhankelijk,
of: $P[A|B] = P[A|\bar{B}] = P[A]$.

Onder de voorwaarde m=m (dus ook n=N-m) hebben wij nu twee reeksen waarnemingen, waarin als H_o juist is, $P[A]$ dezelfde waarde bezit. Onder de voorwaarde m=m en r=r bezit a dus weer de hypergeometrische verdeling (14.1;3). De toets kan dus onveranderd toegepast worden.

Opgave 14.1;3

In een steekproef van gezinnen werd de oogkleur van man en vrouw volgens een bepaald criterium geclassificeerd als "licht" of "niet-licht". De volgende gegevens werden verkregen:

man en vrouw beide lichte oogkleur	: 309
man licht, vrouw niet licht	: 214
man niet licht, vrouw licht	: 133
beide niet licht	: 119

Onderzoek of de oogkleuren van man en vrouw - voor zoverre het de splitsing in "licht" en "niet licht" betreft - stochastisch onafhankelijk zijn of niet.

Een derde mogelijkheid voor toepassing van de 2x2-tabel is als mediaantoets. Laat

(14.1;12) $\underline{x}_1, \ldots, \underline{x}_m$ en $\underline{y}_1, \ldots, \underline{y}_m$

twee onafhankelijke steekproeven zijn uit twee verdelingen, terwijl wij de hypothese H_o willen toetsen dat deze verdelingen identiek zijn. Wij kunnen dan als volgt te werk gaan. Laat $x_1, \ldots, x_m$ en $y_1, \ldots, y_m$ de waargenomen waarden voorstellen. Deze vormen dan, onder H_o, eigenlijk tezamen één steekproef, daar de beide verdelingen onder H_o dezelfde zijn. Noem de mediaan (de middelste waarneming naar grootte, of als m+n even is het gemiddelde van de twee middelste waarnemingen) : M. Dan kan de volgende 2x2-tabel opgesteld worden.

Tabel 14.1;3.

2x2-tabel als mediaantoets (N even)

	aantal>M	aantal<M	totaal
eerste steekproef	a	b	m
tweede steekproef	c	d	n
totaal	$\frac{1}{2}$N	$\frac{1}{2}$N	N

Ook nu bezit a, onder H_o, de hypergeometrische verdeling (14.1;3). Immers wij hebben aan ieder van de N waarnemingen twee kenmerken toegekend, n.l. eerste (resp. tweede = niet eerste) steekproef en > M (resp. niet > M = < M). Deze beide kenmerken zijn echter, onder H_o, stochastisch onafhankelijk; waren zij dit niet, dan zou het kenmerk > M (resp. < M) van de steekproef afhankelijk zijn, hetgeen alleen mogelijk is als de twee steekproeven niet uit dezelfde verdeling afkomstig zijn. (Dit is geen streng bewijs, doch slechts een intuïtieve redenering.) De toets kan nu onveranderd worden toegepast.

Voorbeeld 14.1;2.

Bij twee groepen studenten (van studierichtingen A en B) is de studieduur als volgt verdeeld.

aantal jaren	5	6	7	8	9	10	>10	totaal
richting A	4	10	25	24	12	9	13	97
richting B	1	7	21	30	15	8	20	102
totaal	5	17	46	54	27	17	33	199

In dit geval is de mediaan 8 jaar, en deze valt middenin een groep van gelijke waarnemingen. Wij brengen daarom de scheiding naast deze groep aan en wel zodanig, dat de totalen links en rechts van deze scheiding zo weinig mogelijk van elkaar verschillen. Dit geeft ons de volgende 2x2-tabel.

aantal jaren	$\leqq 8$	> 8	totaal
richting A	63	34	97
richting B	59	43	102
totaal	122	77	199

(Deze modificatie van de toets schaadt de geldigheid niet; in feite mag de scheiding tussen twee tweetal klassen aangebracht worden).

Nu is: $\mathcal{E}(a \mid H_o) = \frac{97.122}{199} \approx 59,5$

en $\sigma^2(a \mid H_o) = \frac{97.102.122.77}{199^2.198} \approx 11,85$ dus $\sigma(a \mid H_o) \approx 3,44$.

Dus, voor tweezijdige toetsing:

$$\frac{|63 - 59,5| - 0,5}{3,44} = \frac{3}{3,44} < 1,$$

Zodat H_o niet verworpen wordt.

Opmerkingen

14.1;4. Uit deze toepassing zien wij, dat de toets toegepast kan worden ondanks het feit, dat de precieze studieduur van hen, die meer dan 10 jaar nodig hadden, niet opgegeven is. In feite is de methode ook toepasbaar indien de gegevens niet kwantitatief, maar kwalitatief zijn: alleen van de volgorde naar grootte wordt n.l. gebruik gemaakt, niet van de grootte der waarnemingen zelf.

14.1;5. Over de vorm van de verdeling van het aantal studiejaren is geen veronderstelling gemaakt. Wat deze ook is, de toets kan steeds worden toegepast. Een dergelijke toets heet verdelingsvrij (beter ware: verdelingsvormveronderstellingsvrij, maar dat

is wat lang) of ook (minder juiste benaming): parametervrij.

14.1;6. De toets is asymptotisch onderscheidend voor alle alternatieven, waarbij de medianen der twee verdelingen gelijk zijn, ook al zijn de verdelingen in andere opzichten verschillend van elkaar. In dat geval is echter de verdeling van $\underline{a}$ niet meer de hypergeometrische en is de onbetrouwbaarheid van de toets onbekend. Indien de verdelingen echter in een omgeving van de (voor beide gelijke), mediaan dezelfde is blijft de toets asymptotisch geldig.

Tenslotte zij nog vermeld, dat de toets vaak in iets andere vorm gegeven wordt. Als toetsingsgrootheid wordt vaak opgegeven:

$$\underline{\chi}^2 = \frac{N\left(|\underline{a}\underline{d}-\underline{b}\underline{c}| - \frac{1}{2}N\right)^2}{mnrs}, \tag{14.1;13}$$

die dan, onder H_o, een z.g. χ^2-verdeling bezit (χ^2=chi-kwadraat). Deze toets is geheel equivalent met de boven beschreven normale benadering (met continuïteitscorrectie).

Ook wordt vaak gewerkt met het verschil der fracties successen

$$\frac{\underline{a}}{n} - \frac{\underline{b}}{m}, \tag{14.1;14}$$

waarbij dan eveneens een normale verdeling als benadering gebruikt wordt. De verwachting, onder H_o, van dit verschil is 0, maar de spreiding is nog afhankelijk van $p_1(=p_2)$. Voor deze spreiding wordt dan, op de in hoofdstuk 12 beschreven wijze, een schatting genomen. Deze methode is daarom minder betrouwbaar dan de hier beschrevene, waarbij $p_1(=p_2)$ geëlimineerd wordt.

14.2. De toetsen van Student

W.S.GOSSET (1876-1937), als brouwer in dienst bij Guinness in Dublin, heeft - schrijvende onder de schuilnaam STUDENT - een aantal voor de moderne statistiek zeer belangrijke artikelen gepubliceerd. Zijn meest bekende resultaat is de afleiding van de "verdeling van Student", die ten grondslag ligt aan de naar hem genoemde toetsen.

Deze verdeling van STUDENT is in feite een familie van verdelingen, die beschreven kan worden door de volgende formule voor de verdelingsdichtheid.

Definitie 14.2;1

De grootheid $\underline{t}_\nu$ heeft een STUDENT-verdeling "met ν vrijheidsgraden" als de verdelingsdichtheid is:

$$(14.2;1) \qquad f(t) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad (\nu = 1,2,\ldots),$$

waarin Γ de (volledige) Γ-functie voorstelt.

De verdelingen van deze familie zijn alle symmetrisch t.o.v. nul en voor $\nu \to \infty$ treedt de N(0,1)-verdeling als limiet op. Voor eindige ν is de verdeling vlakker dan de N(0,1)-verdeling, zoals in fig. 14.2;1 geschetst is.

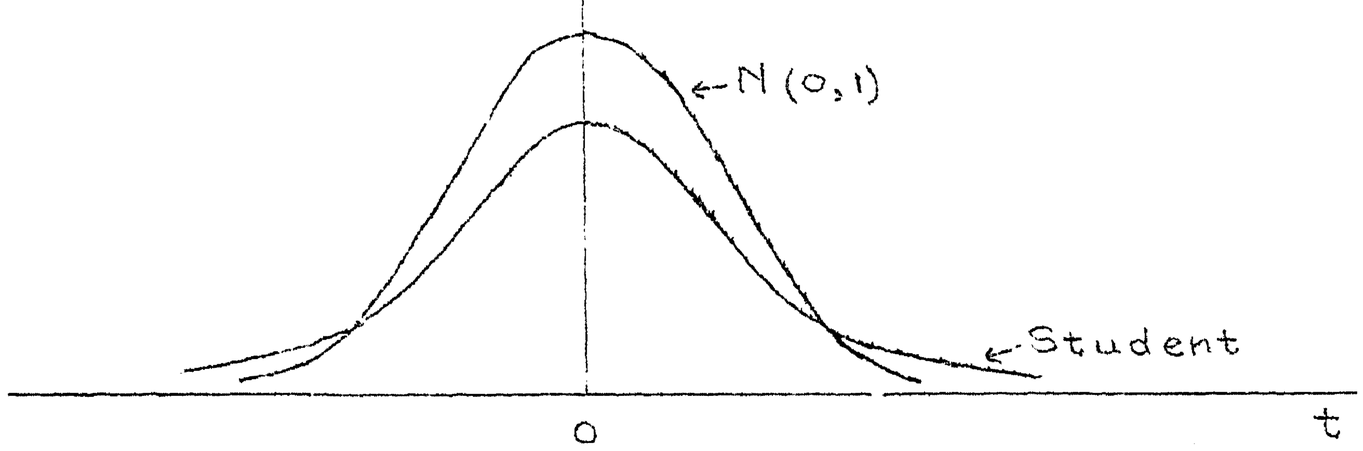


Fig.14.2;1. STUDENT-verdeling in vergelijking met N(0,1).

In tabel 2 vindt men voor $\nu = 1(1)30$ de bij een aantal tweezijdige overschrijdingskansen behorende waarden van t_ν. In de laatste regel staan de overeenkomstige waarden voor de N(0,1)-verdeling vermeld. Met behulp daarvan kan men verifiëren, dat het in fig. 14.2;1 geschetste verschijnsel optreedt.

De toepassingen van deze verdeling berusten nu op de volgende van GOSSET afkomstige stellingen.

Stelling 14.2;1

Indien $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onafhankelijk $N(\mu,\sigma)$ verdeeld zijn en

---

*) "Student's" collected papers, edited by E.S.PEARSON and J.WISHART, University College, London, 1943.

(14.2;2) $$\underline{\bar{x}} \overset{def}{=} \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i \quad , \quad \underline{s}^2 \overset{def}{=} \frac{1}{n-1}\sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})^2,$$

dan bezit de grootheid

(14.2;3) $$\underline{t}_{n-1} \overset{def}{=} \frac{\underline{\bar{x}} - \mu}{\underline{s}}\sqrt{n}$$

een STUDENT-verdeling met n-1 vrijheidsgraden.

Stelling 14.2;2

Indien $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ en $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ onafhankelijk $N(\mu_1, \sigma)$ resp. $N(\mu_2, \sigma)$ verdeeld zijn en

(14.2;4) $$\underline{\bar{x}} \overset{def}{=} \frac{1}{m}\sum_{i=1}^{m} x_i \quad , \quad \underline{\bar{y}} \overset{def}{=} \frac{1}{n}\sum_{i=1}^{n} y_i,$$

(14.2;5) $$\underline{s}_1^2 \overset{def}{=} \sum_{i=1}^{m} (\underline{x}_i - \underline{\bar{x}})^2 \quad , \quad \underline{s}_2^2 \overset{def}{=} \sum_{i=1}^{n} (\underline{y}_i - \underline{\bar{y}})^2,$$

dan bezit de grootheid

(14.2;6) $$\underline{t}_{m+n-2} \overset{def}{=} \frac{\underline{\bar{x}} - \underline{\bar{y}} - (\mu_1 - \mu_2)}{\sqrt{\underline{s}_1^2 + \underline{s}_2^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

een STUDENT-verdeling met m+n-2 vrijheidsgraden.

Opmerkingen

14.2;1 Het belang van deze stellingen is, dat de parameter σ noch in de formule voor de $\underline{t}$'s (de indices worden vaak weggelaten), noch in de verdeling (14.2;1) optreedt. Deze parameter is dus geëlimineerd, waardoor het mogelijk wordt de aandacht volledig op μ resp. $\mu_1 - \mu_2$ te richten

14.2;2 De veronderstelling van normaliteit is essentieel evenals - bij stelling 14.2;2 - de gelijkheid van σ voor beide reeksen stochastische grootheden. Niet al te grote afwijkingen van de normaliteit, in het bijzonder indien de symmetrie der verdelingen bewaard blijft, hebben een vrij geringe invloed op de verdeling der $\underline{t}$'s. Ongelijkheid der spreidingen van de $\underline{x}_i$ en $\underline{y}_j$ kan daarentegen de verdeling van $\underline{t}_{m+n-2}$ sterk beïnvloeden.

Toepassingen

Het is duidelijk, dat met behulp van deze stellingen hypothesen omtrent μ resp. $\mu_1 - \mu_2$ getoetst kunnen worden. Immers indien men stelt

(14.2;7) $$H_o : \mu = \mu_o,$$

waarin μ_o een bekend getal is, dan kan men op grond van waarnemingen $x_1, x_2, \ldots, x_n$ de bijbehorende waarde t_{n-1} van $\underline{t}_{n-1}$ berekenen en de daarbij behorende één- of tweezijdige overschrijdingskans in tabel 2 opzoeken. Deze toets wordt de toets van STUDENT voor

één steekproef genoemd.

Beschikt men over twee onafhankelijke steekproeven $x_1, x_2, \ldots, x_m$ en $y_1, y_2, \ldots, y_n$ dan kan men de hypothese

(14.2;8) $$H_0 : \mu_1 - \mu_2 = \Delta,$$

waarin Δ een bekend getal is, toetsen door t_{m+n-2} te berekenen en de bijbehorende overschrijdingskans op te zoeken. Dit is de toets van STUDENT voor twee steekproeven.

Voorbeeld 14.2;1

Volgens opgave van een leverancier, die en gros vuursteentjes levert aan een afnemer, is het gemiddelde aantal steentjes per doosje gelijk aan 1000. De afnemer telt, ter controle, de inhoud van 10 aselect gekozen doosjes en vindt de in tabel 14.2;1 in de eerste kolom vermelde aantallen.

Tabel 14.2;1

Aantallen vuursteentjes in 10 aselect gekozen doosjes

x	x-995	$(x-995)^2$
983	-12	144
1002	7	49
998	3	9
996	1	1
1002	7	49
983	-12	144
994	-1	1
991	-4	16
1005	10	100
986	-9	81
totaal	-10	594

Op grond van deze steekproef van waarnemingen kan men - in de veronderstelling, dat het aantal vuursteentjes in een doos een normale verdeling bezit - de toets van STUDENT voor één steekproef toepassen.

Daartoe moeten $\bar{x}$ en s berekend worden. Dit gaat het gemakkelijkst door een "voorlopig gemiddelde" a te kiezen (een rond getal, dat in de buurt van $\bar{x}$ ligt, op het oog gekozen) en gebruik te maken van de volgende formules (vgl. voorbeeld 12;1):

(14.2;9) $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i - a) + a,$$

(14.2;10) $$s^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n}(x_i - a)^2 - n(\bar{x} - a)^2\right\}.$$

Wij kiezen a=995; in tabel 14.2;1 zijn de waarden van $x_i - a$ en $(x_i - a)^2$ en de sommen daarvan vermeld. Wij vinden dus

$\bar{x} = -1 + 995 = 994$ en $s^2 = \frac{1}{9}(594 - 10) = 64,89.$

Vullen wij deze waarden in in (14.2;3), tezamen met $\mu = 1000$, omdat wij die waarde willen toetsen, dan vinden wij

$$t_9 = \frac{994 - 1000}{\sqrt{64,89}} = -2,35.$$

Voor de afnemer zijn in de regel alleen de alternatieven $\mu < 1000$ interessant, daar hij in dat geval bij de leverancier zal reclameren, doch niet als $\mu \geqq 1000$ is. Hier is dus de links-éénzijdige toetsing op zijn plaats. Kiest men $\alpha_\ell = 0,05$ (dus $\alpha = 0,10$) dan vindt men in tabel 2 als kritieke waarde -1,83 (in de tabel is alleen de absolute waarde vermeld). Daar de gevonden waarde, -2,35, kleiner is dan deze kritieke waarde, wordt de getoetste hypothese $\mu = 1000$ verworpen ten gunste van $\mu < 1000$. De conclusie luidt dus: de leverancier verpakt te weinig vuurstreentjes in zijn doosjes.

Opmerking 14.2;3

De veronderstelling van normaliteit, die essentieel is bij de toepassing van de toets van STUDENT, wordt in de practijk vaak stilzwijgend en zonder controle op juistheid gemaakt. Bij een gering aantal waarnemingen is deze veronderstelling ook slechts gebrekkig verifieerbaar. Wij zullen later zien, hoe hij vermeden kan worden door toepassing van een verdelingsvrije toets.

Opgave 14.2;1

Het opgegeven gemiddelde gewicht van zakjes suiker is 10 gram. Weging van een steekproef van 10 zakjes geeft de volgende uitkomsten: 10,04; 10,08; 10,01; 9,97; 10,02; 11,99; 10,02; 10,03; 10,00; 10,01. Toets tweezijdig, met behulp van STUDENT's toets en met $\alpha = 0,05$, of de opgegeven gemiddelde waarde juist is.
(Antwoord: $\bar{x} = 10,017$; $s^2 = 0,00089$; $t_9 = 1,8$. De getoetste hypothese wordt niet verworpen.)

Toepassing op verschillen van paren waarnemingen

De meeste toepassingen van de toets van STUDENT voor één steekproef hebben betrekking op een andere situatie dan de boven beschrevene.

Laat $(\underline{x}_i, \underline{y}_i)$ $(i = 1, \ldots, n)$ n onafhankelijke paren van onafhankelijke normaal verdeelde waarnemingen zijn, met

(14.2;11) $$\mathcal{E}\,\underline{x}_i = \mu_i \quad , \quad \mathcal{E}\,\underline{y}_i = \nu_i \, ,$$

dan zijn de verschillen

(14.2;12) $$\underline{v}_i \overset{\text{def}}{=} \underline{x}_i - \underline{y}_i$$

ook normaal verdeeld, met

(14.2;13) $$\mathcal{E}\, \underline{v}_i = \mu_i - \nu_i .$$

Indien de $\underline{v}_i$ alle dezelfde variantie hebben (b.v. omdat alle $\underline{x}_i$ dezelfde variantie hebben, en alle $\underline{y}_i$ ook), kan nu met behulp van de toets van STUDENT de hypothese:

(14.2;14) $$H_o : \quad \mu_i - \nu_i = \Delta_i,$$

met gegeven Δ_i $(i = 1, \ldots, n)$ getoetst worden. De grootheden $\underline{v}_i - \Delta_i$ vormen dan nl. onder H_o een steekproef uit een normale verdeling met verwachting 0, zodat stelling 14.2;1 toegepast kan worden.

Gewoonlijk zijn de paren $(\underline{x}_i, \underline{y}_i)$ bepalingen in duplo van de één of andere grootheid en is $\Delta_i = 0$ $(i = 1, \ldots, n)$ in H_o. De toepassingen behoeven echter niet tot dat eenvoudige geval beperkt te blijven.

Een belangrijk aspect van deze toets is, dat de μ_i onderling mogen verschillen; door niveauverschillen tussen de paren wordt de toets niet verstoord.

Voorbeeld 14.2;2

De bepaling van de isolerende eigenschappen van olie kan geschieden door een glazen buisje, waarin zich twee polen bevinden, met olie te vullen en vervolgens op de polen een spanningsverschil aan te brengen, dat men laat stijgen tot de isolatie doorbroken wordt. Men kan deze bepaling van de doorslagspanning met dezelfde olievulling zovaak herhalen als men wil. Wij beschouwen speciaal de eerste twee bepalingen (meting in duplo). Geven wij de doorslagspanning bij de eerste bepaling aan met $\underline{x}$ en die bij de tweede met $\underline{y}$, dan vindt men - hetgeen niet te verbazen is - in de regel verschillende uitkomsten. Het ligt echter voor de hand te veronderstellen, dat $\underline{x}$ en $\underline{y}$ wel dezelfde kansverdeling zullen bezitten, ook al is deze in de regel voor verschillende oliesoorten verschillend. Indien $\underline{x}$ en $\underline{y}$ niet dezelfde kansverdeling bezitten kan men strict genomen niet van duplo-bepalingen spreken; de volgorde van bepaling heeft dan invloed op de uitkomst. Dit is echter geenszins uitgesloten te achten, daar een door de olie slaande vonk ionen daarin kan achterlaten.

Bij een door J.W.YOUDEN en J.M.CAMERON beschreven experiment *) werden 14 verschillende oliemomsters (één van ieder van 14 soorten olie) gebruikt; van ieder monster werden twee vullingen onderzocht, bij iedere vulling twee bepalingen verricht. In tabel 14.2;2 staan de verschillen v_1 resp. v_2 van de duplobepalingen bij de eerste resp.

*) W.J. YOUDEN en J.M. CAMERON, Use of statistics to determine precision of test methods, Symposium on Application of Statistics, Am.Soc.for testing Materials, Philadelphia 1950 pp. 27-34.

tweede vulling vermeld.

Tabel 14.2;2

Verschillen (in K.V.) van duplobepalingen van de doorslagspanning van oliemonsters.

olie-monster no.	bij eerste vulling v_1	bij tweede vulling v_2	$v_1 - v_2$
1	3	-1	4
2	2	1	1
3	4	2	2
4	4	5	-1
5	1	-2	3
6	5	1	4
7	6	3	3
8	0	0	0
9	-1	7	-8
10	1	5	-4
11	7	1	6
12	-1	-1	0
13	1	0	1
14	1	8	-7

De veronderstelling, dat de volgorde geen invloed heeft op de uitkomsten, komt overeen met H_o van (14.2;14), met $\Delta_i = 0$.

Daar er geen reden is aan te nemen, dat er systematische verschillen optreden tussen de eerste en de tweede vullingen met dezelfde soort olie, laten wij dit onderscheid voorlopig vallen. Wij beschikken dan over 28 onafhankelijke waarnemingen, nl. de in de kolommen met v_1 en v_2 aangegeven verschillen.

Bepalen wij daarvan gemiddelde en variantie, dan vinden wij:

$$\bar{v} = \frac{62}{14} = 2{,}215 ,$$

$$\bar{v}^2 = 4{,}906 ,$$

$$\Sigma v_i^2 = 346 ,$$

$$s^2 = \frac{1}{n-1}\left(\Sigma v_i^2 - n\bar{v}^2\right) = \frac{1}{27}\left(346 - 28.4{,}906\right) = 7{,}73 ,$$

$$s = \sqrt{7{,}73} = 2{,}78 ,$$

$$t_{27} = \frac{\bar{v}}{s}\sqrt{n} = \frac{2{,}215}{2{,}78}\sqrt{28} = 4{,}2 .$$

Deze waarde is aanzienlijk groter dan de grootste voor $\nu = 27$ in tabel 2 voorkomende waarde. De overschrijdingskans (tweezijdig in dit geval) is dus veel kleiner dan 0,01 en de getoetste hypothese moet met stelligheid worden verworpen. Daar $t > 0$ is luidt de conclusie, dat er een systematisch verschil tussen de beide bepalingen is in die zin, dat de tweede doorslagspanning de neiging heeft kleiner uit te vallen dan de eerste. Dit betekent niet, dat nu voor alle oliemonsters de waarden van $\Delta_i = \mu_i - \nu_i$ dezelfde behoeven te zijn. Zelfs behoeven zij nog niet alle > 0 te zijn, maar dit toch wel in

overwegend aantal. Dit effect kan zeer wel het gevolg zijn van ioni sering door de eerste doorslag.

Opgave 14.2;2

Indien men nu nog wil nagaan, of dit effect verschillend is voor de eerste en de tweede vulling, dan kan men de "dubbelverschillen" $v_1 - v_2$ bepalen en daarop dezelfde toets toepassen. Indiën de daling van μ_i tot ν_i door de eerste doorslag bij de eerste en tweede vulling dezelfde is, vormen deze verschillen opnieuw een steekproef (nu van 14 waarnemingen) uit een verdeling met verwachting 0.Voer deze berekeningen uit.
(Oplossing: $t_{13} = 0,26$. De hypothese wordt niet verworpen.)

Opmerkingen

(14.2;4) De boven beschreven toets is asymptotisch onderscheidend als

(14.2;15) $$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mu_i - \nu_i) \longrightarrow \pm \infty \quad \text{voor } n \longrightarrow \infty.$$

Immers voor $n \longrightarrow \infty$ zal s tot σ naderen, terwijl $\bar{v}$ nadert tot $\frac{1}{n} \Sigma \Delta_i$, zodat $|t| \longrightarrow \infty$ als aan (14.2;15) voldaan is.

(14.2;5) De hier gevolgde methode van het gebruiken van beide kolommen (v_1 en v_2) tegelijk in één toets kan tot moeilijkheden leiden indien de aantallen vullingen voor verschillende oliesoorten verschillend zijn. Indien b.v. slechts bij één olie de doorslagspanning daalt, maar bij andere stijgt, kan men toch als conclusie een daling krijgen, indien men van die ene oliesoort meer proeven (vullingen) in de toetsing betrekt dan van de andere samen. Wanneer men echter met alle oliesoorten evenveel proeven doet, bestaat dit bezwaar niet. Proeven als deze berusten trouwens steeds op de gedachte, dat het (eventuele) verschil tussen de beide doorslagspanningen voor alle of nagenoeg alle oliemonsters wel hetzelfde teken zal bezitten.

Voorbeeld van de toets van STUDENT voor twee steekproeven

Voorbeeld 14.2;3

Van twee soorten varkensvoer, A en B, wenst men na te gaan of één van beide meer gewicht aanzet dan het andere. Men verdeelt daartoe een groep van 24 varkens aselect in twee groepen van 12 en geeft gedurende enige tijd aan de ene groep voedsel A en aan de andere voedsel B. De gewichtstoenamen,die hiervan het gevolg zijn, staan in tabel 1 14.2;3 vermeld.

Tabel 14.2;3

Gewichtstoenamen (in K.G.) van varkens gevoed met twee typen varkensvoer

A	B
31	26
34	24
29	28
26	29
32	30
35	29
38	32
34	26
30	31
29	29
32	32
31	28

Aannemende, dat de gewichtstoename een normale verdeling bezit met gelijke spreiding voor A en B, kan de toets van STUDENT voor twee steekproeven toegepast worden om de hypothese te toetsen, dat de verwachting der gewichtstoename voor beide voedsels gelijk is. Daartoe wordt formule (14.2;6) gebruikt, met $\mu_1 - \mu_2 = 0$. Berekening geeft: $t_{22} = 2{,}65$, hetgeen overeenkomt met een overschrijdingskans (tweezijdig) tussen 0,02 en 0,01. De hypothese wordt dus verworpen (als $\alpha \geqq 0{,}02$ gekozen was) ten gunste van voedsel A (waarvoor de gemiddelde gewichtstoename in de steekproef groter is dan bij B).

Opmerkingen

14.2;6 Bij deze toets is niet alleen de normaliteit van belang, maar vooral ook de gelijkheid der beide spreidingen. In de beide steekproeven verschillen deze niet sterk, zoals grofweg te zien is door de spreidingsbreedte (de grootste min de kleinste waarneming) te bepalen. Deze bedraagt voor de eerste steekproef 38 - 26 = 12 en voor de tweede 32 - 24 = 8. Daar beide steekproeven van de zelfde omvang zijn, zijn deze twee uitkomsten vergelijkbaar (bij ongelijke omvang is dat niet zo: hoe groter de steekproef, hoe groter ook in de regel de spreidingsbreedte is. Een verschil van 30% is voor kleine steekproeven niet verontrustend. Men kan de gelijkheid der spreidingen ook rigoreus toetsen (uitgaande van de veronderstelling van normaliteit), met behulp van het quotiënt der steekproefvarianties als toetsingsgrootheid; deze toets behandelen wij echter niet.

14.2;7 Bij het opzetten van een proef als deze zijn er twee belangrijke richtlijnen ter vergroting van het onderscheidingsvermogen van de toets. Ten eerste wordt dit, onder overigens gelijkblijvende omstandigheden, maximaal, indien men de beide steekproeven van gelijke omvang neemt. Ten tweede doet men er goed aan het proef-

materiaal - hier de 24 varkens - zoveel mogelijk "homogeen" te maken, d.w.z. van gelijke leeftijd, geslacht en gewicht. Alle verschillen tussen de varkens verkleinen het onderscheidingsvermogen, al wordt de onbetrouwbaarheid van de toets - door het aselect verdelen in twee groepen - er niet door beïnvloed. Zware varkens zijn niet voor niets zwaar; zij zijn in het verleden snel gegroeid en zullen dit tijdens de proef dus waarschijnlijk ook doen. Grote verschillen in groeisnelheid vergroten echter σ, dus ook de noemer van t, waardoor deze zelf kleiner neigt te worden dan bij varkens met kleinere verschillen in groeisnelheid. Zorgvuldige selectie van proefdieren - gevolgd door aselecte verdeling in twee groepen, voor alle zekerheid - loont daarom de moeite.

14.2;8 Tabel 14.2;3 beschouwende zou men wellicht geneigd zijn de verschillen in de regels te vormen en op deze verschillen de toets van STUDENT voor één steekproef toe te passen. Hoewel deze toets, mathematisch gezien, niet onjuist is - de onbetrouwbaarheid is immers bekend - bestaat het bezwaar, dat het onderscheidingsvermogen ervan veel kleiner is dan dat van de toets voor twee steekproeven. De gewichtstoenamen, die in één regel staan, hebben in feite niets met elkaar te maken; vermoedelijk staan de beide kolommen opgeschreven in de volgorde van weging, zodat de samenvoeging in paren op aselecte wijze tot stand is gekomen. Het is intuïtief reeds wel duidelijk, dat dit een gebrekkige, weinig onderscheidende, wijze van toetsen is.

Indien echter geen "homogene" groep varkens beschikbaar is, kan men met vrucht terugvallen op de toets voor één steekproef van verschillen. Men kan dan nl. paren van vrijwel gelijke varkens vormen, waarbij het ene paar aanzienlijk kan verschillen van het andere. Van ieder paar wordt dan aselect één varken aan A, één aan B toegewezen. Deze paren verschijnen dan ook - althans hun gewichtstoenamen - ieder in één regel. Nu is de situatie juist andersom: de toets voor één steekproef, toegepast op de verschillen der gewichtstoenamen, is nu de aangewezene.

Opgave 14.2;3

Onderstaande twee steekproeven zijn afkomstig uit twee normale verdelingen met gelijke spreiding:

0,87	0,52
0,75	1,94
0,03	0,53
0,17	1,02
0,67	0,94
1,92	0,06
0,23	0,25
2,14	2,05
2,69	1,93
1,31	1,17
	1,38
	1,43
	1,74

Toets, met behulp van de toets van STUDENT voor twee steekproeven, de hypothese, dat de verwachting der beide verdelingen dezelfde is.

(Opgave 80 van het opgavenboekje, onderdeel a; oplossing blz.118-119.)

14.3 Verdelingsvrije toetsen voor dezelfde of analoge hypothesen

Zoals reeds is opgemerkt kan de onderstelling van normaliteit der verdelingen vermeden worden zonder dat daarvoor een andere veronderstelling omtrent de vorm der verdelingen in de plaats gesteld behoeft te worden. Er zijn vele dergelijke "verdelingsvrije" toetsen en wij behandelen er slechts enkele van.

Toetsen voor één steekproef

Als toets voor één steekproef kan in plaats van de toets van STUDENT voor de hypothese (14.2;7), als bekend is dat de verdeling symmetrisch is, gewerkt worden met het aantal der waarnemingen, die $< \mu_o$ zijn, als toetsingsgrootheid. Is nl. de verdeling symmetrisch dan impliceert $\mu = \mu_o$:

$$(14.3;1) \qquad P[x < \mu_o] = P[x > \mu_o].$$

Is $P[x = \mu_o] = o$ (b.v. omdat de verdeling continu is), dan zijn de beide kansen van (14.3;1) dus gelijk aan $\frac{1}{2}$ en dit kan getoetst worden met behulp van een binomiale toets - zie hoofdstuk 13 - met $p = \frac{1}{2}$.

Is $P[x = \mu_o] > o$, dan volgt uit de symmetrie van de verdeling, als $\mu = \mu_o$ is:

$$(14.3;2) \qquad P[x < \mu_o \mid x \neq \mu_o] = P[x > \mu_o \mid x \neq \mu_o] = \tfrac{1}{2}$$

(de lezer bewijze dit). Voor de toepassing van de toets betekent dit, dat waarnemingen, die $= \mu_o$ zijn, buiten beschouwing gelaten worden en dat met behulp van de overige waarnemingen de relatie (14.3;2) wordt getoetst.

De toepassing van deze toets wordt, voor $n \leqq 100$, vergemakkelijkt door tabel 3, waarin de linker kritieke waarden voor deze toets (de z.g. tekentoets; deze naam wordt later duidelijk) zijn vermeld. De rechter kritieke waarden kunnen gemakkelijk berekend worden, daar zij (wegens $p = \frac{1}{2}$) opgeteld bij de linker kritieke waarden het aantal waarnemingen $\neq \mu_o$ geven.

Voorbeeld 14.3;1

Wij kunnen deze toets toepassen op de gegevens van tabel 14.2;1 (voorbeeld 14.2;1). In dat geval is $\mu_o = 1000$ en er zijn 10 waarnemingen, waarvan er geen = 1000 is. Het aantal waarnemingen > 1000 (dit kan men evengoed gebruiken als het aantal<1000; het is het gemakkelijkste het kleinste aantal van deze twee te nemen, daar dit in tabel 3 direct met de daarin vermelde kritieke waarden vergeleken kan worden) is gelijk aan 3. Volgens tabel 3 leidt dit resultaat, ook bij éénzijdige toetsing met onbetrouwbaarheidsdrempel 0,05 (dus bij gebruik van de kolom waarboven 0,10 staat) niet tot verwerping van de getoetste hypothese.

Opmerking 14.3;1

Wij vinden dus bij toepassing van de toets van STUDENT in dit geval een andere uitkomst dan bij de tekentoets. De oorzaak is, dat in geval van normale verdelingen - en verdelingen, die daarvan slechts weinig afwijken - de eerstgenoemde toets een groter onderscheidingsvermogen heeft dan de tekentoets. Hierover later meer.

Indien de verdeling, waaruit de steekproef afkomstig is, niet symmetrisch is, gelden (14.3;1 en 2) niet, ook niet onder de hypothese $\mu = \mu_0$. Deze hypothese kan dan noch met de toets van STUDENT - wegens de afwijking van normaliteit - noch met de tekentoets onderzocht worden. In feite is er geen enkele toets, behalve voor grote steekproeven, die zonder verdere onderstellingen over de vorm van de scheve verdeling van toepassing is. Men kan dan echter met de tekentoets wel een andere hypothese toetsen, nl. dat de mediaan een voorgeschreven waarde bezit; althans indien de verdeling continu is, want dan is de kans op een waarde onder de mediaan gelijk aan $\frac{1}{2}$, evenals de kans op een waarde daarboven. Zoals wij in hoofdstuk 12 (blz 111 e.v.) gezien hebben is, juist voor scheve verdelingen, de mediaan vaak een geschiktere parameter dan de verwachting (of het gemiddelde).

Opgave 14.3;1

In een steekproef van 225 waarnemingen uit een verdeling met onbekende vorm worden 150 waarnemingen aangetroffen, die kleiner dan een zeker getal A zijn, terwijl de overige groter dan A zijn.

Toets de volgende twee hypothesen ($\alpha = 0{,}05$):

a) A is de mediaan van de verdeling,

a) A is het derde quartiel, d.w.z. voor A geldt:

$$P[x < A] = \tfrac{3}{4}.$$

Ga verder na welke waarden voor p wel en welke niet verworpen worden als de getoetste hypothese luidt:

(14.3;3) $$P[x < A] = p.$$

De toepassing op verschillen van paren waarnemingen, die in het voorgaande voor de toets van STUDENT beschreven is, geldt analoog voor de tekentoets. Zijn x en y onafhankelijk verdeeld volgens dezelfde kansverdeling en is $v = x - y$, dan geldt voor de verdeling van v, als $P[v=0] = 0$ is:

(14.3;4) $$P[v < 0] = P[v > 0] = \tfrac{1}{2}$$

en als $P[v=0] > 0$ is:

(14.3;5) $$P[\underline{v} < 0 \mid \underline{v} \neq 0] = P[\underline{v} > 0 \mid \underline{v} \neq 0] = \tfrac{1}{2}.$$

(de lezer bewijze dit.)

De hypothese, dat voor iedere i $(=1,\ldots,n)$ twee grootheden (b.v. duplobepalingen) $\underline{x}_i$ en $\underline{y}_i$ dezelfde verdeling hebben, kan nu dus getoetst worden met behulp van het aantal positieve (of negatieve) tekens der verschillen $\underline{v}_i = \underline{x}_i - \underline{y}_i$. Hieraan ontleent de toets zijn naam.

De vorm der verdelingen is nu van geen enkel belang meer; ook de spreidingen der $\underline{v}_i$ behoeven niet gelijk te zijn.

Voorbeeld 14.3;2

Beschouw de gegevens van tabel 14.2;2 (doorslagspanningen van oliesoorten).

Het aantal waarnemingen $\neq 0$ in de eerste twee kolommen (v_1 en v_2) bedraagt 25 (de nullen worden buiten beschouwing gelaten). Het aantal negatieve hieronder is 5. Volgens tabel 3 is de bijbehorende tweezijdige overschrijdingskans $\leq 0{,}01$, zodat de getoetste hypothese - dat de duplobepalingen aequivalent zijn - verworpen moet worden. (Hier vinden wij dus wel dezelfde uitkomst als met de toets van STUDENT).

Onder de 12 "dubbelverschillen" $v_1 - v_2$, die $\neq 0$ zijn, zijn er 4 negatief. Volgens tabel 3 is dit niets bijzonders; de tweezijdige overschrijdingskans is $> 0{,}10$, daar de dáárbij behorende kritieke waarde 2 bedraagt.

Opmerking 14.3;2

Eén der meest opvallende verschillen tussen de tekentoets en de toets van STUDENT is wel, dat toepassing van de eerstgenoemde zoveel gemakkelijker is en vlugger gaat dan van de laatstgenoemde. Vooral bij het verkrijgen van een snel overzicht over waarnemingsmateriaal van uitgebreide en complexe aard kunnen toetsen zoals de tekentoets zeer nuttig zijn.

Opgave 14.3;2

Een chemisch analyste verricht een aantal titraties in duplo. Iedere te titreren oplossing wordt daartoe, direct na de bereiding, in twee kolfjes gedaan en terwijl de eerste titratie plaats vindt staat het tweede kolfje te wachten. Zij verkrijgt de volgende uitkomsten:

Tabel 14.3;1

Uitkomsten van 12 in duplo uitgevoerde titraties

Oplossing nr	1e titratie	2e titratie
1	21,24	25,83
2	16,84	17,35
3	15,52	16,12
4	25,68	23,54
5	24,04	24,58
6	19,77	27,42
7	11,92	14,73
8	28,83	27,52
9	17,38	14,91
10	11,01	19,87
11	23,43	24,38
12	17,16	20,55

Ga na (met onbetrouwbaarheidsdrempel 0,05) of het wachten invloed heeft uitgeoefend op de uitkomst van de titraties.
(Opgave 3 van het sommenboekje; oplossing op bl. 48-49.)

Het verschil in onderscheidingsvermogen tussen de tekentoets en de toets van STUDENT is, als aan de voor de toets van STUDENT nodige veronderstellingen voldaan is, nogal groot, vooral bij kleine steekproeven. Dit is begrijpelijk, daar alleen het teken van ieder der waarnemingen in de toetsing betrokken wordt, maar niet de grootte. Hierdoor gaat informatie verloren, die in de steekproef aanwezig is. Aan dit bezwaar wordt b.v. tegemoet gekomen door de symmetriestoets van WILCOXON.

De te toetsen hypothese van deze toets luidt als volgt:

(14.3;6) H_o: ieder der onafhankelijke stochastische grootheden $\underline{x}_1, \ldots, \underline{x}_m$ heeft een verdeling, die symmetrisch is met betrekking tot een bekende waarde a_i $(i = 1, \ldots, m)$.

De waarden a_i behoeven niet gelijk te zijn, de verdelingen ook overigens evenmin. De toets wordt toegepast met behulp van de verschillen $\underline{x}_i - a_i$, die onder H_o symmetrisch verdeeld zijn om 0. Het is duidelijk, dat deze toets gebruikt kan worden om, bij symmetrische verdelingen, de hypothese $\mu = \mu_o$ te toetsen, terwijl hij anderzijds, evenals de tekentoets, op verschillen van duplo-waarnemingen toegepast kan worden. Immers het verschil $\underline{v}$ van twee onafhankelijke stochastische grootheden $\underline{x}$ en $\underline{y}$ is, indien $\underline{x}$ en $\underline{y}$ dezelfde verdeling bezitten, symmetrisch om 0 verdeeld.

Wij geven hier een korte beschrijving van deze toets. Een uitvoerige handleiding met tabellen is gegeven door A. BENARD en

CONSTANCE VAN EEDEN *).

De toetsingsgrootheid $\underline{T}$ is als volgt gedefinieerd. De waarnemingen ($\underline{x}_i - a_i$; $i = 1,\ldots,n$; of: $\underline{x}_i - \underline{y}_i$ als het over verschillen gaat), worden, evenals bij de tekentoets, voor zoverre zij = 0 zijn, buiten beschouwing gelaten. Noem het aantal, dat dan overblijft, n. Aan ieder van deze waarnemingen wordt nu een rangnummer toegekend naar opklimmende grootte van de absolute waarden der waarnemingen. Zijn er gelijke absolute waarden, dan delen de desbetreffende waarnemingen de hun toekomende rangnummers gelijkelijk op. Deze rangnummers worden nu voorzien van de tekens der bijbehorende waarnemingen en $\underline{T}$ is de som van de op deze wijze verkregen getallen.
Men kan nu de volgende stelling bewijzen:

Stelling 14.3;1

Indien H_o juist is, geldt voor $\underline{T}$:

(14.3;7) $$\mathcal{E}(\underline{T} \mid H_o) = 0$$

(14.3;8) $$\sigma^2(\underline{T} \mid H_o) = \frac{3n(n+1)^2 + n^3 - D}{12} = \frac{1}{6}n(n+1)(2n+1) - \frac{D-n}{12};$$

waarin D de volgende betekenis heeft: verdeel alle waarnemingen, die $\neq 0$ zijn, in groepen met gelijke absolute waarden **). Laat t_j ($j = 1,\ldots,k$) de omvang van de j^e knoop voorstellen, dan is

(14.3;9) $$D \overset{\text{def}}{=} \sum_{j=1}^{k} t_j^3 \quad ***)$$

Voor $n \to \infty$ is verder $\underline{T}$, onder H_o, asymptotisch normaal verdeeld.

Wij bewijzen deze stelling niet volledig, maar geven slechts het principe aan.

Onder H_o is ieder der waarnemingen symmetrisch t.o.v. 0 verdeeld, zodat iedere absolute waarde > 0, die in de steekproef gevonden wordt, gelijke kans (nl. $\frac{1}{2}$) heeft om bij een + resp. bij een - teken te behoren. Ieder der rangnummers heeft dus gelijke kans om positief of negatief in $\underline{T}$ voor te komen. De verdeling van $\underline{T}$ kan dus verkregen worden door voor ieder der rangnummers met kans $\frac{1}{2}$ een + en met kans $\frac{1}{2}$ een - te zetten en ze dan op te tellen. De lotingen voor + of -

---

*) A. BENARD en CONSTANCE VAN EEDEN, Handleiding voor de symmetrietoets van WILCOXON, Rapport S 206 (M 76) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam 1956.

**) "knopen" (Engels: "ties").

***) Zijn er geen gelijken onder de absolute waarden, dan zijn dus alle $t_j = 1$, $k = D = n$, en de laatste term van (14.3;8) valt weg. Tabellen van $\frac{1}{6}n(n+1)(2n+1)$ en $\sqrt{\frac{1}{6}n(n+1)(2n+1)}$ zijn, voor $n = 21(1)100$, te vinden in de bovengenoemde handleiding voor deze toets.

zijn bovendien onafhankelijk.

Daar deze procedure symmetrisch is met betrekking tot + en - (hij verandert niet als + en - verwisseld worden), is $\underline{T}$, onder H_o, symmetrisch verdeeld met betrekking tot 0, waaruit (14.3;7) direct volgt.

Het 1 wijs van (14.3;8) is iets ingewikkelder. Zijn de waarden der t_j, gerangschikt naar opklimmende grootte van de rangnummers, waarbij zij behoren: $t_1, \ldots, t_k$, en geven wij de bijbehorende rangnummers aan met $r_1, \ldots, r_k$, dan geldt dus voor de eerste rangnummers, die alle gelijk aan r_1 zijn;

$$r_1 = \frac{1}{t_1}(1 + 2 + \ldots + t_1) = \tfrac{1}{2}(t_1 + 1).$$

Verder voor r_2 (de volgende t_2 rangnummers hebben deze waarde):

$$r_2 = t_1 + \tfrac{1}{2}(t_2 + 1),$$

enz. Algemeen;

$$(14.3;10) \qquad r_j = \sum_{h=1}^{j-1} t_h + \tfrac{1}{2}(t_j + 1).$$

Daar verder de verwachting van $\underline{T}$, onder H_o, gelijk aan 0 is, geldt:

$$\sigma^2(\underline{T} \mid H_o) = \sum_{j=1}^{k} t_j r_j^2 = \sum_{j=1}^{k} t_j \left\{ \sum_{h=1}^{j-1} t_h + \tfrac{1}{2}(t_j + 1) \right\}^2 =$$

$$= \sum_{j=1}^{k} \left\{ t_j \left(\sum_{h=1}^{j-1} t_h \right)^2 + t_j (t_j + 1) \sum_{h=1}^{j-1} t_h + \tfrac{1}{4} t_j (t_j + 1)^2 \right\}.$$

Deze nogal ingewikkelde vorm laat zich als volgt reduceren:

$$\sum_{\nu=1}^{n} \nu^2 = \sum_{\nu=1}^{t_1} \nu^2 + \sum_{\nu=1}^{t_2} (t_1 + \nu)^2 + \sum_{\nu=1}^{t_3} (t_1 + t_2 + \nu)^2 + \ldots =$$

$$= \sum_{j=1}^{k} \left\{ \sum_{\nu=1}^{t_j} \left(\sum_{h=1}^{j-1} t_h + \nu \right)^2 \right\} = \sum_{j=1}^{k} \left\{ t_j \left(\sum_{h=1}^{j-1} t_h \right)^2 + 2 \sum_{h=1}^{j-1} t_h \sum_{\nu=1}^{t_j} \nu + \sum_{\nu=1}^{t_j} \nu^2 \right\} =$$

$$= \sum_{j=1}^{k} \left\{ t_j \left(\sum_{h=1}^{j-1} t_h \right)^2 + t_j (t_j + 1) \sum_{h=1}^{j-1} t_h + \tfrac{1}{6} t_j (t_j + 1)(2t_j + 1) \right\},$$

daar $\sum_{\nu=1}^{t_j} \nu^2 = \frac{1}{6} t_j (t_j + 1)(2 t_j + 1)$.

Het linkerlid, $\sum \nu^2$, is gelijk aan $\frac{1}{6} n(n+1)(2n+1)$, zodat wij voor σ^2 krijgen:

$$\sigma^2(\underline{T} \mid H_o) = \tfrac{1}{6} n(n+1)(2n+1) - \tfrac{1}{6} \sum_{j=1}^{k} t_j (t_j + 1)(2 t_j + 1) + \tfrac{1}{4} \sum_{j=1}^{k} t_j (t_j + 1)^2,$$

hetgeen door een korte herleiding in (14.3;8) overgaat.

Opmerking 14.3;3

Bij de berekening van de variantie van $\underline{T}$ zijn de t_j als niet-stochastisch grootheden behandeld, terwijl zij toch eigenlijk ook

stochastisch zijn. De in (14.3;8) gebruikte notatie is dan ook niet juist; de berekende variantie is die van de voorwaardelijke verdeling van T, onder H_o en onder de voorwaarde: $t_1 = t_1, t_2 = t_2, \ldots, t_k = t_k$. Wij zouden dus, strikt genomen, de volgende notatie moeten gebruiken.

(14.3;11) $$\sigma^2(T \mid H_o; t_1, t_2, \ldots, t_k).$$

Opmerkenswaard is overigens, dat de volgorde der t_j er in (14.3;8) niet toe doet; alleen $D = \Sigma t_j^3$ komt in de formule voor. Bij de berekening van D behoeft dus op deze volgorde niet gelet te worden.

De asymptotische normaliteit van T bewijzen wij hier niet. Dit bewijs is te vinden in:
CONSTANCE VAN EEDEN and A.BENARD, A general class of distribution-free tests for symmetry containing the tests of Wilcoxon and Fisher, Proc. Kon. Ned. Akad. v. Wet. A 60 (1957) and Indagationes Mathematicae 19 (1957) 381-391, 392-400, 401-408.

De exacte verdeling van T, onder H_o en onder de voorwaarde van gegeven t's (nu is de volgorde wel van belang) is, voor kleine n, gemakkelijk af te leiden. Voor $n = 3$ b.v. en zonder gelijken onder de absolute waarden, is deze verdeling te verkrijgen door op de $2^3 = 8$ mogelijke manieren + of - tekens toe te kennen aan de rangnummers 1, 2 en 3. Dit is in tabel 14.3;2 uitgevoerd.

Tabel 14.3;2

Afleiding van de verdeling van T, onder H_o, voor n=3, zonder gelijken

1	+	-	+	+	-	-	+	-
2	+	+	-	+	-	+	-	-
3	+	+	+	-	+	-	-	-
T	6	4	2	0	0	-2	-4	-6

Deze 8 mogelijke uitkomsten hebben alle kans $\frac{1}{8}$, dus heeft de waarde 0 de kans $\frac{1}{4}$, terwijl 6,4,2,-2,-4,-6, alle kans $\frac{1}{8}$ hebben.

Voor de berekening van deze exacte verdeling voor grotere n zijn recursieformules opgesteld, die te vinden zijn in het bovenaangehaalde artikel van CONSTANCE VAN EEDEN en A. BENARD.

Met behulp van deze methode zijn, voor $n = 4(1)20$, de exacte kritieke waarden van de toets berekend, voor $\alpha = 0{,}01$; 0,02; 0,05 en

0,10 (tweezijdig); deze zijn in tabel 4 te vinden. De overige in tabel 4 vermelde kritieke waarden zijn gevonden met behulp van de normale benadering voor de verdeling van $\underline{T}$ onder H_o. Zij maken de toepassing van de toets uitermate eenvoudig, althans als er geen gelijke absolute waarden zijn. Uitgebreidere tabellen van kritieke waarden en tabellen van overschrijdingskansen zijn te vinden in de reeds eerder genoemde handleiding.

Opmerkingen

14.3;4. Zijn er wel gelijke absolute waarden, dan wordt hierdoor D vergroot, dus de variantie verkleind. Dit betekent, dat men, toch werkende met de variantie zonder gelijken, tot iets te grote overschrijdingskansen komt. Het verschil is echter, als er geen grote groepen gelijken zijn, klein.

Zijn er gelijken, dan geldt tabel 4 niet meer exact, maar als het er niet te veel zijn, nog wel bij goede benadering. Twijfelt men, dan kan men naast het gebruik van de tabel de normale benadering gebruiken. Bij kleine n en veel gelijken kan men de exacte verdeling bepalen langs de boven aangegeven weg, en daarmede dan de exacte overschrijdingskans van de gevonden waarde van $\underline{T}$.

14.3;5. Uit tabel 14.3;1 blijkt, dat de intervallen tussen twee op elkaar volgende waarden, die door $\underline{T}$ aangenomen kunnen worden, de lengte 2 bezitten, althans als er geen gelijken zijn. De continuïteitscorrectie, bij gebruik van de normale benadering, bedraagt daarom $\pm$ 1. Zijn er wel gelijken, dan is de zaak veel ingewikkelder. Men handhaaft dan gewoonlijk toch deze continuïteitscorrectie.

Voorbeeld 14.3;1(vervolg)

Wij passen de toets nu toe op de gegevens van voorbeeld 14.2;1, waarop wij reeds eerder de toets van STUDENT en de teken-toets hebben toegepast. De waarde van T kan berekend worden als in tabel 14.3;3.

Tabel 14.3;3

Berekening van T voor de symmetrietoets van WILCOXON voor de gegevens van tabel 14.2;1

verschillen x_i-1000	in volgorde van absolute grootte	rangnummer met teken
-17	+ 2	+ 2
+ 2	+ 2	+ 2
- 2	- 2	- 2
- 4	- 4	- 4
+ 2	+ 5	+ 5
-17	- 6	- 6
- 6	- 9	- 7
- 9	-14	- 8
+ 5	-17	- 9½
-14	-17	- 9½
n=10		T =9-46=-37 +

Volgens tabel 4, geen rekening houdende met de gelijken, geldt

(14.3;11) $\qquad 0{,}025 < k_\ell < 0{,}05,$

hetgeen dus bij linkseenzijdige toetsing met $\alpha_\ell = 0{,}05$ tot verwerping van de getoetste hypothese leidt.

Voor toepassing van de normale benadering berekenen wij eerst D:

$$D = 3^3+1+1+1+1+1+2^3 = 40$$

en vervolgens (met (14.3;8)):

$$\sigma^2 = \frac{30.11^2 + 1000 - 40}{12} = 382{,}5 \; ; \; \sigma = 19{,}56.$$

Derhalve wordt

$$\frac{T+1}{\sigma} = \frac{-36}{19{,}56} = -1{,}84$$

hetgeen, volgens tabel 1 geeft: $k_\ell \approx 0{,}033$.

Dit klopt goed met de vorige uitkomst.

Opmerking 14.3;6

Dit voorbeeld is een illustratie van het grotere onderscheidingsvermogen van deze toets in vergelijking met de tekentoets. Immers de laatstgenoemde leidde niet tot verwerping van de getoetste hypothese, de eerstgenoemde wél, evenals de toets van STUDENT (voorbeeld 14.2;1).

Opgave 14.3;3

Los opgave 14.2,1 (bl 147) op met behulp van de symmetrie-

toets van WILCOXON.
(Antwoord: $T = 34, n = 9$; volgens tabel 4 is bij $\alpha = 0{,}05$ de rechter kritieke waarde 35, zodat H_o nog net niet verworpen wordt. Met de normale benadering: $\sigma = 16{,}8$; $33/16{,}8 = 1{,}96$; $k = 0{,}05$, dus net op het randje. Zou men in dit geval een exacte beslissing willen nemen, dan kan dit alleen op grond van de exacte verdeling, rekening houdende met de gelijken.)

Opgave 14.3;4

Pas de symmetrietoets van WILCOXON toe op voorbeeld 14.2;2 en opgave 14.2;2.
(Antwoord: voor de 25 verschillen $\neq 0$ van tabel 14.2;2 is $T = 251$, $D = 1411$, $\sigma = 73{,}5$, $T_\alpha = 193$ voor $\alpha = 0{,}01$; voor de 12 dubbelverschillen is $T = 12$, $T_\alpha = 44$ voor $\alpha = 0{,}10$.)

Toetsen voor twee steekproeven

Een verdelingsvrije toets voor twee steekproeven, de mediaantoets, is reeds in § 14.1 behandeld. Een tweede, de twee-steekproeven-toets van WILCOXON, zullen wij nu bespreken.

(14.312) Zijn $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ en $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ twee onafhankelijke steekproeven, dan houdt de getoetste hypothese H_o in, dat al deze m+n stochastische grootheden dezelfde kansverdeling bezitten. Aan de vorm van deze kansverdeling wordt geen enkele voorwaarde opgelegd.

Een uitvoerige handleiding (met tabellen) voor deze toets is: DORALINE WABEKE en CONSTANCE VAN EEDEN, Handleiding voor de toets van WILCOXON, Rapport S 176 (M 65) van het Mathematisch Centrum, Amsterdam 1955, waarin ook verdere litteratuur is vermeld.

Een korte beschrijving van de toets volgt hieronder.

De toetsingsgrootheid W is gelijk aan de som van

a) twee maal het aantal paren (x_i, y_j) met $x_i > y_j$, en

b) het aantal paren (x_h, y_k) met $x_h = y_k$,

met $i, h = 1, \ldots, m$; $j, k = 1, \ldots, n$*)

De toets berust verder op de volgende stelling.

Stelling 14.3;2

Indien H_o juist is, geldt voor $\underline{W}$:

(14.3;13) $$\mathcal{E}(\underline{W} \mid H_o) = mn,$$

---

*) Vaak wordt ook de grootheid $U = \frac{1}{2}W$ als toetsingsgrootheid gebruikt.

(14.3;14) $$\sigma^2(\underline{W}\mid H_o) = \frac{mn(N^3 - D)}{3N(N-1)}\ ,$$ *)

met

(14.3;15) $$N = m + n\ ,\quad D = \sum_{j=1}^{k} t_j^3\ .$$

De t_j geven weer de omvangen der knopen in n; nu echter niet van de absolute waarden der waarnemingen, maar van de waarnemingen zelf.

Voor m en $n \to \infty$ is $\underline{W}$ verder, onder H_o, asymptotisch normaal verdeeld.

De formule voor σ^2 geldt, evenals bij de symmetrietoets van WILCOXON, voorwaardelijk voor gegeven waarden $t_1, t_2, \ldots, t_k$ (als de N waarnemingen in totaal uit k knopen bestaan).

Het bewijs van (14.3;13) is zeer eenvoudig. Is nl. H_o vervuld, dan geldt voor ieder paar (i,j) (vgl. (14.3;5)):

$$P[\underline{x}_i > \underline{y}_j \mid \underline{x}_i \neq \underline{y}_j] = P[\underline{x}_i < \underline{y}_j \mid \underline{x}_i \neq \underline{y}_j] = \tfrac{1}{2}\ ,$$

zodat, als $\underline{x}_i \neq \underline{y}_j$, de verwachte bijdrage tot $\underline{W}$ gelijk is aan 1. Is echter $x_i = y_j$, d.w.z. nemen $\underline{x}_i$ en $\underline{y}_j$ dezelfde waarde aan, dan is de bijdrage tot $\underline{W}$ ook 1. Daar er mn paren (i,j) zijn, volgt (14.3;13) hieruit onmiddellijk.

De afleiding van(14.3;14) is ingewikkelder, maar eveneens geheel elementair. Het kan gevoerd worden met behulp van de reeds eerder afgeleide relaties (14.3;10) en (14.3;8). Daartoe herleiden wij eerst W tot een som van rangnummers.

Wij nummeren alle $N = m + n$ waarnemingen naar opklimmende grootte en geven deze aan met

(14.3;16)
$$\begin{aligned} &\text{aantal} &&: t_1, t_2, \ldots, t_k\ ,\\ &\text{rangnummer} &&: r_1, r_2, \ldots, r_k\ , \end{aligned}$$

waarbij dan weer geldt:

(14.3;10) $$r_j = \sum_{h=1}^{j-1} t_h + \tfrac{1}{2}(t_j + 1)\ .$$

Laat nu de rangnummers der x_i zijn: $R_1, R_2, \ldots, R_m$ dan definiëren wij

---

*) Zijn er geen gelijken ($D = N$), dan wordt deze formule:

(14.3;14a) $$\sigma^2(\underline{W}\mid H_o) = \tfrac{1}{3}mn(N+1).$$

Tabellen van

$$\left.\begin{aligned} \mathcal{E}(\underline{U}\mid H_o) &= \tfrac{1}{2}\mathcal{E}(\underline{W}\mid H_o) = \tfrac{1}{2}mn,\\ \sigma(\underline{U}\mid H_o) &= \tfrac{1}{2}\sigma(\underline{W}\mid H_o) = \sqrt{\tfrac{1}{12}mn(N+1)} \end{aligned}\right\}\ U = \tfrac{1}{2}W$$

en

zijn, voor $m = 1(1)100$, $n = 11(1)100$ en $m \leq n$, te vinden in: Auxiliary table for WILCOXON's two sample test, R 132/S 86 of the Mathematical Centre, Amsterdam.

(14.3;17) $$R \overset{def}{=} \sum_{i=1}^{m} R_i .$$

De volgende relatie bestaat dan tussen W en R :

(14.3;18) $$W = 2R - m(m+1).$$

Laat nl. onder $R_1, R_2, \ldots, R_m$ voorkomen:

(14.3;19) rangnummer: $r_1, r_2, \ldots, r_k$,
aantal : $\nu_1, \nu_2, \ldots, \nu_k$,

dan geldt

(14.3;19^a) $$\sum_j t_j = N, \quad \sum_j t_j r_j = \tfrac{1}{2}N(N+1), \quad \sum_j \nu_j = m.$$

Verder is dan

(14.3;20) $$R = \sum_j \nu_j r_j$$

en

$$W = \sum_j \nu_j (t_j - \nu_j) + 2 \sum_{j=2}^{k} \sum_{h=1}^{j-1} \nu_j (t_h - \nu_h) =$$

$$= \Big(\sum_j \nu_j t_j + 2 \sum_{j=2}^{k} \sum_{h=1}^{j-1} \nu_j t_h\Big) - \Big(\sum_j \nu_j^2 + 2 \sum_{j=2}^{k} \sum_{h=1}^{j-1} \nu_j \nu_h\Big).$$

De laatste van deze twee termen is gelijk aan $(\sum_j \nu_j)^2$, dus $= m^2$.

Verder volgt, door invullen van (14.3;10) in (14.3;20):

$$2R = 2 \sum_j \nu_j \Big\{ \sum_{h=1}^{j-1} t_h + \tfrac{1}{2}(t_j + 1) \Big\} = \sum_j \nu_j t_j + 2 \sum_{j=2}^{k} \sum_{h=1}^{j-1} \nu_j t_h + \sum_j \nu_j =$$

$$= W + m^2 + m = W + m(m+1).$$

Wij kunnen dus i.p.v. met W ook met R werken, terwijl krachtens (14.3;18) geldt:

(14.3;21) $$\mathcal{E}(\underline{W} \mid H_0) = 2\,\mathcal{E}(\underline{R} \mid H_0) - m(m+1)$$

en

(14.3;22) $$\sigma^2(\underline{W} \mid H_0) = 4\,\sigma^2(\underline{R} \mid H_0).$$

Onder H_0 bezitten alle $\underline{x}_i$ en $\underline{y}_j$ dezelfde verdeling en zij vormen dus tezamen een steekproef $\underline{z}_1, \ldots, \underline{z}_N$ uit deze verdeling, die als het ware door een aselecte verdeling in twee groepen van m en n waarnemingen is gesplitst. De voorwaarde $\underline{t}_1 = t_1, \underline{t}_2 = t_2, \ldots, \underline{t}_k = t_k$, waardoor de bijbehorende verzameling (14.3;16) van rangnummers bepaald wordt, voegt aan de steekproef $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_N$ deze verzameling rangnummers op ondubbelzinnige wijze toe. De stochastische grootheden $\underline{R}_1, \underline{R}_2, \ldots, \underline{R}_m$ kunnen dus, onder H_0 <u>en</u> de voorwaarden $\underline{t}_1 = t_1, \underline{t}_2 = t_2, \ldots, \underline{t}_k = t_k$ beschouwd worden als een aselecte steekproef (zonder teruglegging) uit deze verzameling van rangnummers.

Voor iedere $i = 1,2,\ldots,m$ geldt dan echter:

$$\mathcal{E}\,\underline{R}_i = \frac{1}{N}\sum_j t_j r_j = \tfrac{1}{2}(N+1) \qquad \text{(vgl. (14.3;19))}$$

en dus

$$\mathcal{E}\,\underline{R} = m\,\mathcal{E}\,\underline{R}_i = \tfrac{1}{2}m(N+1). \tag{14.3;23}$$

Tezamen met (14.3;21) wordt hieruit (14.3;13) gemakkelijk opnieuw verkregen.

Verder is

$$\mathcal{E}\,\underline{R}_i^2 = \frac{1}{N}\sum_j t_j r_j^2. \tag{14.3;24}$$

Beschouwen wij vervolgens de simultane verdeling van een willekeurig paar $(\underline{R}_j, \underline{R}_\ell)$, dan is dit dezelfde als van een steekproef van 2 zonder teruglegging uit de verzameling (14.3;16) van rangnummers. Dit paar kan dus $\binom{N}{2}$ verschillende paren van waarden met gelijke kansen aannemen (als de volgorde binnen het paar buiten beschouwing gelaten wordt), zodat geldt:

$$\mathcal{E}\,\underline{R}_i\underline{R}_\ell = \binom{N}{2}^{-1}\left\{\sum_j \binom{t_j}{2} r_j^2 + \sum_{j=2}^{k}\sum_{h=1}^{j-1} t_j t_h r_j r_h\right\} =$$

$$= \tfrac{1}{2}\binom{N}{2}^{-1}\left\{\sum_j (t_j^2 - t_j) r_j^2 + 2\sum_{j=2}^{k}\sum_{h=1}^{j-1} t_j t_h r_j r_h\right\}.$$

Nu is, weer volgens (14.3;19):

$$\tfrac{1}{4}N^2(N+1)^2 = \left(\sum_j t_j r_j\right)^2 = \sum_j t_j^2 r_j^2 + 2\sum_{j=2}^{k}\sum_{h=1}^{j-1} t_j t_h r_j r_h,$$

zodat wij, ook van (14.3;24) gebruik makende, krijgen:

$$\mathcal{E}\,\underline{R}_i\underline{R}_\ell = \left\{N(N-1)\right\}^{-1}\left\{\tfrac{1}{4}N^2(N+1)^2 - N\,\mathcal{E}\,\underline{R}_i^2\right\}.$$

Verder is dan

$$\mathcal{E}\,\underline{R}^2 = \mathcal{E}\left(\sum_i \underline{R}_i\right)^2 = m\,\mathcal{E}\,\underline{R}_i^2 + m(m-1)\,\mathcal{E}\,\underline{R}_i\underline{R}_\ell =$$

$$= m\left\{\left(1 - \frac{m-1}{N-1}\right)\mathcal{E}\,\underline{R}_i^2 + \frac{(m-1)N(N+1)^2}{4(N-1)}\right\} =$$

$$= m\left\{\frac{m}{N-1}\,\mathcal{E}\,\underline{R}_i^2 + \frac{(m-1)N(N+1)^2}{4(N-1)}\right\}.$$

Hierin kan nu $\mathcal{E}\,\underline{R}_i^2 = \frac{1}{N}\sum_j t_j r_j^2$ rechtstreeks ontleend worden aan (14.3;8), indien daarin voor m gesubstitueerd wordt: N. Immers volgens het bewijs van deze formule is $\sigma^2(T\,|\,H_0) = \sum_j t_j r_j^2$, zodat

$$\mathcal{E}\,\underline{R}_i^2 = \frac{1}{N}\left\{\tfrac{1}{6}N(N+1)(2N+1) - \tfrac{1}{12}(D-N)\right\}.$$

Substitutie hiervan in de vorige vergelijking en herleiding geeft:

$$\mathcal{E}\,\underline{R}^2 = \frac{mn(N^3-D)}{12N(N-1)} + \frac{1}{4}m^2(N+1)^2.$$

zodat volgt

$$\sigma^2(\underline{R}) = \mathcal{E}\,\underline{R}^2 - (\mathcal{E}\,\underline{R})^2 = \frac{mn(N^3-D)}{12N(N-1)},$$

waarin $D = \sum_j t_j^3$ is.

Het bewijs van de asymptotische normaliteit van $\underline{W}$, onder H_o, geven wij niet. Dit is, voor het geval zonder gelijke waarnemingen, te vinden in

H.B. MANN and D.R. WHITNEY, On a test of whether one of two random variables is stochastically larger than the other, Ann.Math.Stat. 18 (1947) 50-60

en in

D. VAN DANTZIG, Hoofdstuk 6 van de Kadercursus Mathematische Statistiek, Mathematisch Centrum, Amsterdam, 1947-1950.

Het bewijs van de asymptotische normaliteit als er wel gelijken optreden is impliciet te vinden in:

W.H. KRUSKAL, A non-parametric test for the several sample problem, Ann. Math. Stat. 23 (1952) 525-540

en in

D.J. STOKER, Oor'n klas van toetsingsgroothede vir die probleem van twee steekproewe, dissertatie Amsterdam, 1955.

De exacte verdeling van $\underline{W}$, onder H_o en onder voorwaarde van gegeven t's (met hun volgorde) is ook nu weer op eenvoudige wijze af te leiden, indien m en n klein zijn. Daartoe schrijft men alle mogelijke steekproeven van m uit de $N = m+n$ door de t_j bepaalde rangnummers op en berekent voor ieder daarvan de door $\underline{W}$ aangenomen waarde. Iedere steekproef heeft de kans $\binom{N}{m}^{-1}$ en daaruit volgt direct de gezochte kansverdeling.

Opgave 14.3;5

Bereken de exacte verdeling van $\underline{W}$ voor het geval $m = 2, n = 3$ met $t_1 = 1$, $t_2 = 2$, $t_3 = 2$.

Ook voor deze verdeling zijn recursieformules afgeleid, nl. in het boven aangehaalde artikel van MANN en WHITNEY voor het geval zonder gelijke waarnemingen en in

L.J. SMID, On the distribution of the test Statistics of KENDALL and WILCOXON's two sample test when ties are present, Statistica neerlandica 10 (1956) 205-214, voor het geval met gelijken.

In tabel 5 (a,b en c) zijn voor α = 0,02; 0,05 en 0,10 (tweezijdig) de kritieke waarden van W voor m, n = 1(1)15 en $m \leq n$ getabelleerd. Tabellen van exacte overschrijdingskansen, voor $m \leq n \leq 10$ zijn te vinden in de bovengenoemde handleiding voor deze toets. Een nomogram, waaruit men de overschrijdingskansen kan aflezen, samengesteld door P.J. RIJKOORT, is te vinden in: Statistische tabellen en nomogrammen, uitgegeven onder redactie van de Vereniging voor Statistiek door H.E. Stenfert Kroese N.V., Leiden.

Voor andere waarden van m en n gebruike men de normale benadering met continuïteitscorrectie $\pm$ 1. Is één van beide, m of n, zeer klein, b.v. $m < 5$, terwijl $n > 15$ is, dan is de tabel niet te gebruiken terwijl de normale benadering nog niet erg goed is. Voor dat geval kan een andere benadering gebruikt worden (een n maal geïtereerde homogene verdeling), die wij hier echter niet bespreken. In belangrijke gevallen kan men verder altijd de exacte verdeling afleiden, hoewel dit soms zeer bewerkelijk kan zijn.

Voorbeeld 14.3;2

Wij behandelen nu het probleem van voorbeeld 14.2;3 (bl.151) met behulp van de twee-steekproeven-toets van WILCOXON.

Dit geschiedt gemakkelijk in de vorm van tabel 14.3;4.

Tabel 14.3;4

Schema ter berekening van W

waarneming	aantal malen bij A	aantal malen bij B	bijdrage tot W	t_i	t_i^3
24	0	1	0	1	1
26	1	2	4	3	27
28	0	2	0	2	8
29	2	3	26	5	125
30	1	1	17	2	8
31	2	1	38	3	27
32	2	2	44	4	64
34	2	0	48	2	8
35	1	0	24	1	1
38	1	0	24	1	1
totaal	m = 12	n = 12	W = 225	N = 24	D = 270

Het is in dit geval iets eenvoudiger x en y in gedachten te verwisselen en dan W te berekenen. De uitkomst is dan 63 (de lezer controlere dit), hetgeen tezamen met de gevonden waarde, 225, gelijk is aan $2mn$ = 288. Deze controle is altijd nuttig voor het ontdekken van fouten bij de berekening van W.

Storen wij ons niet aan de knopen en gebruiken wij tabel 5, dan vinden wij voor de tweezijdige overschrijdingskans

(14.3;25) $0,02 < k < 0,05$ (maar zeer dicht bij 0,02), hetgeen iets groter is dan op bl. 151 met de toets van STUDENT gevonden werd.

Berekenen wij de variantie met behulp van (14.3;14), dan vinden wij

$$\sigma^2 = \frac{144\ (24^3 - 270)}{3.24.23} = 1178,6 \ ; \quad \sigma = 34,3$$

(de invloed der gelijken is gering; past men formule (14.3;14a) toe, dan krijgt men $\sigma = 34,6$).

Nu is

$$\frac{|W - mn| \ -1}{\sigma} = \frac{225-144-1}{34,3} = 2,33 \ ,$$

dus $k \approx 0,02$, in goede overeenstemming met (14.3;25).

Opgave 14.3;6 Pas op de gegevens van opgave 14.2;3 de twee-steekproeven-toets van WILCOXON toe.
(Opgave 80 van het sommenboekje, onderdeel c; oplossing op blz. 119-120.)

Opgave 14.3;7 Behandel voorbeeld 14.1;2 met behulp van de toets van WILCOXON. Beschouw daarbij alle waarnemingen > 10 als één knoop.
(Antwoord: $W = 11.265$; $mn = 9894$; $D = 320.371$; $\sigma^2 = 632.801$; $\sigma = 795,5$; $\frac{W-\mu-1}{\sigma} = 1,72$; $k = 0,085$).

Opgave 14.3;8

Over het aantal verkeersongevallen met dodelijke afloop in twee steden A en B zijn de gegevens van tabel 14.3;5 beschikbaar.

Tabel 14.3;5

Aantal ongevallen met dodelijke afloop in twee steden A en B

Jaar	Aantal ongevallen	
	in A	in B
1920	99	75
1921	87	83
1922	105	81
1923	97	101
1924	109	93
1925	105	97
1926	116	87
1927	110	113
1928	125	103
1929	117	93
1930	130	105
1931	137	107
1932	121	124

vervolg tabel 14.3;5

1933	126	112
1934	131	121
1935	145	130
1936	142	121

Onderzoek of in één van beide steden systematisch meer ongevallen met dodelijke afloop optreden dan in de andere (onbetrouwbaarheidsdrempel 0,05). Ga hierbij uit van de onderstelling, dat een eventueel verschil tussen A en B in ieder der onderzochte jaren in de zelfde richting ligt en licht de keuze van de door U gekozen toets toe aan de hand van een grafiek van de gegevens. Wat is UW conclusie. Welke verklaringen van het door U gevonden verschijnsel acht U aannemelijk?
(Opgave 13 uit het sommenboekje; oplossing (gedeeltelijk) op blz. 56-58.)

Opmerkingen

14.3;7. Bestaan de waarnemingen uit twee knopen, dan gaat de twee-steekproeven-toets van WILCOXON over in de 2x2-tabel toets van §14.1. Men vergelijke:
CONSTANCE VAN EEDEN, Verdelingsvrije toetsen voor twee steekproeven en de methode der 2x2-tabel, Statistica Neerlandica 10 (1956) 157-162.

14.3;8. Zoals in opgave 14.3;7 blijkt, is de toets van WILCOXON voor twee steekproeven geschikt voor toepassing op afgeknot waarnemingsmateriaal. Dit is b.v. van belang bij levensduurproeven, waarbij het grootste deel der waarnemingen in betrekkelijk korte tijd verkregen wordt, terwijl men op enkele waarnemingen zeer lang moet wachten. Men kan dan beter de proef enigzins uitbreiden (m en n groter) en na een van tevoren vastgestelde tijd afbreken (waarom "- "van tevoren vast gesteld"?).

Opgave 14.3;9

a) Twee soorten gloeilampen moeten met elkaar vergeleken worden wat hun levensduur betreft. Daartoe worden van de eerste soort 10 exemplaren en van de tweede 12 exemplaren genomen en deze worden gebrand tot ze doorbranden. De uitkomsten zijn (in uren):
x: 625,637,710,770,820,843,856,920,1070,1225.
y: 630,683,780,830,889,970,1028,1150,1210,1470,1520,2090.
Onderzoek of één van beide soorten een systematisch langere levensduur heeft dan de andere soort (onbetrouwbaarheidsdrempel 0,05).

b) Indien men de bovenbeschreven proef wil bekorten tot maximaal 1000 uren nemen de gegevens de volgende vorm aan:

x : 625,637,710,770,820,843,856,920, 2 waarnemingen >1000.
y : 630,683,780,830,889,970, 6 waarnemingen >1000.

Toets ook voor deze gegevens de hypothese, die U in deel a) van deze opgave getoetst hebt. (Aanwijzing: beschouw de waarnemingen die groter dan 1000 zijn; alle als gelijk uitgevallen waarnemingen.)
(Opgave 24 uit het sommenboekje; oplossing op blz. 68-70.)

14.4 Onderscheidingsvermogen van de toetsen van STUDENT, de tekentoets en de toetsen van WILCOXON en asymptotische eigenschappen.

De berekening van het onderscheidingsvermogen hebben wij in hoofdstuk 13 voor binomiale toetsen uitgevoerd. Voor de toetsen van STUDENT kan een soortgelijke berekening uitgevoerd worden, die echter niet tot een integraal voert, die met de normale verdeling te benaderen is, maar tot de z.g. niet-centrale t-verdeling, die evenals de normale verdeling (maar pas veel later) uitvoerig getabelleerd is. Het onderscheidingsvermogen heeft dezelfde soort vorm als bij de tekentoets, indien men op de horizontale as uitzet

(14.4;1) $$\delta_1 = \frac{\mu - \mu_0}{\sigma}$$

voor de toets voor één steekproef (μ is de werkelijke verwachting, μ_0 de getoetste en σ de spreiding van x) en

(14.4;2) $$\delta_2 = \frac{\mu_1 - \mu_2 - \Delta}{\sigma}$$

voor de toets voor twee steekproeven ($\mu_1 - \mu_2$ is het werkelijke verschil tussen de verwachtingen van x en y en Δ de getoetste waarde daarvan; σ de gemeenschappelijke spreiding van x en van y).

Het onderscheidingsvermogen hangt dan af van het aantal vrijheidsgraden $\nu = n-1$ bij één steekproef en $= m+n-2$ bij twee steekproeven.

In figuur 14.4;1 is het onderscheidingsvermogen geschetst in de vorm van enkele curven voor verschillende waarden van ν, behorende bij $\alpha = 0,05$.

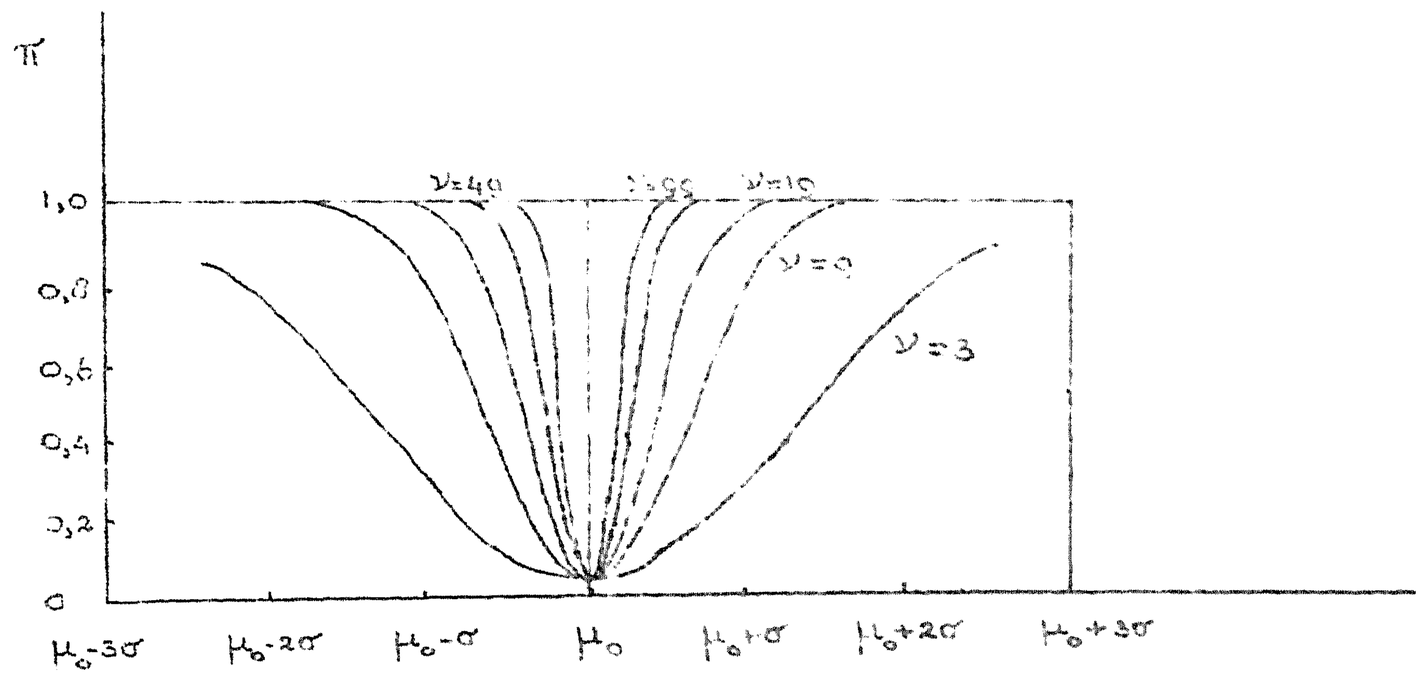


Fig. 14.4;1. Het onderscheidingsvermogen van STUDENT's tweezijdige toetsen met $\alpha = 0,05$.

Men kan bewijzen dat de toetsen van STUDENT, indien de bijbehorende voorwaarden van normaliteit en gelijke spreidingen vervuld zijn, het grootst mogelijke onderscheidingsvermogen bezitten. Dit

houdt dus in, dat zij in dat geval onderscheidender zijn dan verdelingsvrije toetsen. Het omgekeerde kan voorkomen als aan de genoemde veronderstellingen niet is voldaan, terwijl dan bovendien de onbetrouwbaarheidsdrempel van de toetsen van STUDENT niet meer de opgegeven waarde bezit.

De minst onderscheidende toets voor één steekproef die wij hebben behandeld, is de tekentoets. Ter illustratie zullen wij, voor $n = 20$, deze toets vergelijken met die van STUDENT.

Wij stellen daarbij, voor de eenvoud, $\sigma = 1$ en $\mu_0 = 0$. Het onderscheidingsvermogen van de toets van STUDENT kan dan uit fig. 14.4;1 afgelezen worden. Voor de tekentoets kan het als volgt berekend worden.

Voor $n = 20$ zijn de kritieke waarden bij $\alpha = 0{,}05$ volgens tabel 3, gelijk aan 5 en 15. Wij berekenen nu eerst, b.v. voor $\mu = 0{,}5$, de kans $p_{0,5}$ op een positieve uitkomst. De beschouwde verdeling is dus $N(0{,}5;1)$, zodat tabel 1 ons geeft: $p_{0,5} = 1 - 0{,}31 = 0{,}69$. Vervolgens berekenen wij, met behulp van de normale benadering voor de binomiale verdeling, de kans op 15 of meer positieve uitkomsten onder de 20 waarnemingen. Noemen wij het aantal positieve uitkomsten $\underline{a}$, dan is

$$\mathcal{E}\,\underline{a} = 20.0{,}69 = 13{,}8; \quad \sigma^2(\underline{a}) = 20.0{,}69.0{,}31 = 4{,}28; \quad \sigma = 2{,}07;$$

$$\frac{14{,}5 - 13{,}8}{2{,}07} = 0{,}34,$$

zodat de gezochte kans, volgens tabel 1, $\approx 0{,}367$ is. De kans om in het andere deel van de kritieke zone ($a \leqq 5$) terecht te komen is gering en wordt buiten beschouwing gelaten. Uit fig. 14.4;1 volgt, dat de overeenkomstige kans, voor de toets van STUDENT, $\approx 0{,}55$ is.

Analoge berekeningen voor $\mu = 1$; 1,5 en 2 leiden tot de in fig. 14.4;2 geschetste curve. In deze figuur is, ter vergelijking, de desbetreffende curve voor de toets van STUDENT eveneens opgenomen. Voor $\mu = 0$ is $p_0 = \frac{1}{2}$ en de onbetrouwbaarheid blijkt dan, bij opzoeken in een tabel van de binomiale verdeling, 0,0414 te bedragen, dus vrij aanzienlijk minder dan die van de toets van STUDENT, waarvoor deze exact 0,05 is. Een gedeelte van het geringere onderscheidingsvermogen is aan deze kleinere onbetrouwbaarheid te wijten.

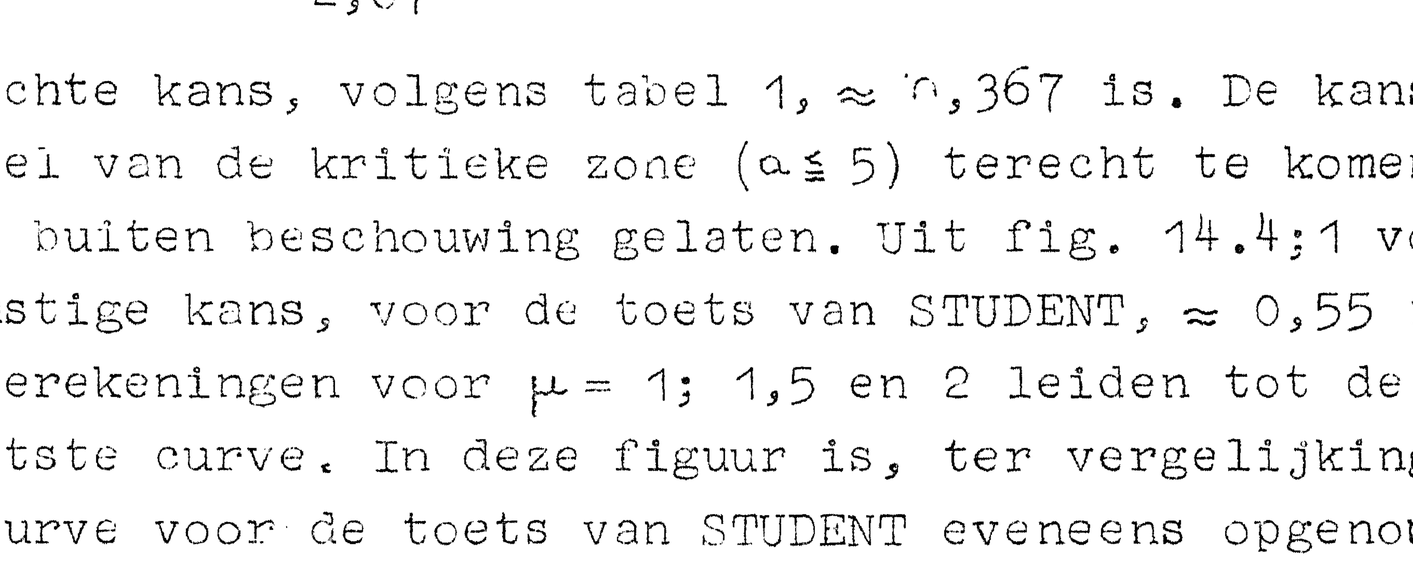


Fig. 14.4;2. Onderscheidingsvermogen van de tekentoets in vergelijking met de toets van STUDENT (één steekproef, $\alpha = 0{,}05$ tweezijdig, $n = 20$).

Opgave 14.4;1
Voer de berekeningen voor fig. 14.4;2 uit.

Voor de toetsen van WILCOXON is veel minder over het onderscheidingsvermogen bekend dan voor de andere behandelde toetsen. De belangrijkste reden hiervan is, dat dit onderscheidingsvermogen niet verdelingsvrij is. Ook bij verschuiving der verdelingen ten opzichte van de getoetste ligging, met behoud van de vorm, is het onderscheidingsvermogen van deze vorm afhankelijk. Verder blijkt, dat, in tegenstelling met het eenvoudige karakter van de verdeling van de toetsingsgrootheden T en W onder H_o, deze grootheden onder alternatieve hypothesen juist zeer ingewikkelde verdelingen bezitten. Curven zoals de boven in fig. 14.4;1 en 2 geschetst, zijn dan ook voor deze toetsen nog niet berekend.

Wel bestaat er practische ervaring met deze toetsen in vergelijking met die van STUDENT, die doet vermoeden dat :a, het onderscheidingsvermogen slechts weinig geringer is in het geval van normale verdelingen en :b , groter als de basisverdelingen scheef zijn. De in dit hoofdstuk behandelde voorbeelden wijzen ook in de richting van bewering a. Beide beweringen worden bovendien bevestigd door G. E. NOETHER in een nog niet gepubliceerd artikel: "The efficiency of some distributionfree tests", dat in de loop van 1958 in Statistica Neerlandica zal verschijnen.

In dit artikel wordt een overzicht gegeven van verschillende onderzoekingen, die betrekking hebben op de verhouding waarin men het aantal waarnemingen moet verkleinen om de toets van STUDENT een even groot onderscheidingsvermogen te geven als een daarmee vergeleken toets. De uitkomsten gelden voor grote steekproeven en dicht bij de getoetste hypothese. Deze verhouding, die de asymptotische doeltreffendheid van de toets heet, is in geval van normale verdelingen voor de tekentoets $2/\pi = 0{,}64$ (zowel voor het geval van één als voor dat van twee steekproeven) en voor de beide toetsen van WILCOXON $3/\pi = 0{,}95$.

Voor bepaalde, in de practijk zeker voorkomende, scheve basisverdelingen, werden in het laatste geval waarden variërend van 1,5 tot 3 gevonden.

De conclusie, die men hieruit kan trekken is, dat de tekentoets wel degelijk als een zeer zwakke toets beschouwd moet worden, die in de regel alleen geschikt is voor oriëntatie (nl. tenzij het aantal waarnemingen zo groot is, dat het onderscheidingsvermogen voldoende wordt geacht). De toetsen van WILCOXON zijn echter practisch gelijkwaardig met die van STUDENT in het normale geval en in andere gevallen vaak verre superieur.

Bovendien kan men desgewenst, ten koste van enige complicaties

van de uit te voeren berekeningen, modificaties van deze toetsen gebruiken, waarvoor de asymptotische doeltreffendheid in het normale geval 1 is, terwijl zij toch verdelingsvrij zijn. *)

Van alle besproken toetsen kan, grofweg, gezegd worden, dat zij asymptotisch onderscheidend zijn, zodra de getoetste hypothese niet is vervuld. Zo is dit b.v. het geval voor de tweezijdige toetsen van STUDENT als $\mu \neq \mu_o$ resp. $\mu_1 - \mu_2 \neq \Delta$ is, maar niet als $\mu = \mu_o$ resp. $\mu_1 - \mu_2 = \Delta$ is, terwijl de verdeling niet normaal is of de spreidingen niet gelijk zijn. Het onderscheidingsvermogen stijgt dan soms wel boven de onbetrouwbaarheidsdrempel (zoals reeds eerder gesignaleerd), maar het behoeft niet tot 1 te naderen.

Bij de toets van WILCOXON voor twee steekproeven is de grootheid, waar het om gaat:

$$P[\underline{x} > \underline{y}] - P[\underline{x} < \underline{y}]. \qquad (14.4;3)$$

Onder H_o is dit verschil = 0 en als het $\neq$ is, is de tweezijdige toets asymptotisch onderscheidend. Voor de symmetrietoets is de zaak iets ingewikkelder. Men zie voor dat geval de eerder geciteerde litteratuur.

---

*) Men zie voor deze toetsen:
M.E.TERRY, Some rank order tests, which are most powerful against specific parametric alternatives, Ann.Math.Stat. (1952) 346-366,
en
B.L. VAN DER WAERDEN, Ein neuer Test für das Problem der zwei Stichproben, Mathem. Annalen 126 (1953) 93-107.

Hoofdstuk 15. Betrouwbaarheidsgrenzen

De theorie der betrouwbaarheidsgrenzen vormt de verbinding tussen de schattings- en de toetsingstheorie. Ook hier wordt het onderwerp van onderzoek gevormd door één (of meer) onbekende parameter(s) van een kansverdeling, waarover door een reeks, gewoonlijk onafhankelijke, waarnemingen

(15;1) $$\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$$

informatie wordt verschaft.

Definitie 15;1

Een functie $\underline{t}_\ell = t(\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n)$ van de waarnemingen is een betrouwbaarheidsondergrens voor een onbekende parameter θ, met onbetrouwbaarheidsdrempel α_ℓ, indien geldt:

(15;2) $$P[\underline{t}_\ell < \theta] \geqq 1-\alpha_\ell ,$$

waarin θ de (onbekende) werkelijke waarde der parameter voorstelt.

Analoog geldt, voor een betrouwbaarheidsbovengrens met onbetrouwbaarheidsdrempel α_r:

(15;3) $$P[\underline{t}_r > \theta] \geqq 1-\alpha_r .$$

Tezamen vormen $\underline{t}_\ell$ en $\underline{t}_r$ een betrouwbaarheidsinterval voor θ, met onbetrouwbaarheidsdrempel α, indien geldt:

(15;4) $$P[\underline{t}_\ell < \theta < \underline{t}_r] \geqq 1-\alpha .$$

Opmerkingen

15;1. In alle praktisch voorkomende gevallen, waarbij voor één parameter zowel onder- als bovengrenzen zijn afgeleid, geldt:

(15;5) $$P[\underline{t}_\ell < \underline{t}_r] = 1 .$$

Indien hieraan voldaan is, geldt

(15;6) $$\alpha = \alpha_r + \alpha_\ell .$$

(De lezer bewijze dit.)

15;2. Evenals bij toetsen kan de toevoeging "drempel" in "onbetrouwbaarheidsdrempel" vervallen, indien $\underline{t}_\ell$ en $\underline{t}_r$ een continue kansverdeling bezitten; de ongelijkheden (15;2, 3, en 4) gaan dan in gelijkheden over - niet binnen, maar buiten de vierkante haken - . Zijn $\underline{t}_\ell$ en $\underline{t}_r$ discreet verdeeld, dan is de onbetrouwbaarheid hoog-

stens gelijk aan en in de regel kleiner dan de onbetrouwbaarheidsdrempel; de eerstgenoemde hangt van de onbekende waarde van θ af, zodat hij zelf ook onbekend blijft. Voor sommige (mogelijke) waarden van θ worden onbetrouwbaarheid en onbetrouwbaarheidsdrempel gelijk, zodat de ongelijkheden (15;2,3, en 4) niet "verbeterd" kunnen worden.

15;3. $1-\alpha_l$, $1-\alpha_r$ resp. $1-\alpha$ worden <u>betrouwbaarheidscoëfficiënten</u> genoemd (Engels: confidence coefficient, of: confidence level).

15;4. Aan het einde van hoofdstuk 12 hebben wij reeds enkele voorbeelden van (benaderde) betrouwbaarheidsgrenzen ontmoet. Wij zullen in dit hoofdstuk de exacte grenzen, behorende bij die voorbeelden afleiden.

<u>Afleiding met behulp van de toetsingstheorie</u>

Wij bespreken eerst het verband van de theorie der betrouwbaarheidsgrenzen met de toetsingstheorie, daarna met de schattingstheorie.

Indien wij beschikken over een toetsingsmethode T, waarmee wij, op grond van de waarnemingen (15;1), iedere toegelaten (mogelijk geachte) waarde van de onbekende parameter kunnen toetsen, dan geldt de volgende stelling.

<u>Stelling 15;1</u>

De verzameling der voor θ toegelaten waarden (de parameterruimte Ω) wordt door toepassing van de toetsingsmethode met onbetrouwbaarheidsdrempel α, in twee deelverzamelingen gesplitst.

Dit zijn:

(15;7) $$\underline{G} = G(\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n)$$

de verzameling van alle <u>niet</u> verworpen waarden, en

(15;8) $$\underline{\bar{G}} = \bar{G}(\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n) = \Omega - \underline{G}$$

de verzameling van alle <u>wel</u> verworpen waarden, het complement van $\underline{G}$ met betrekking tot Ω.

Voor $\underline{G}$ geldt:

(15;9) $$P[\theta \in \underline{G}] \geqq 1-\alpha,$$

waarin θ de werkelijke waarde van de onbekende parameter voorstelt (en "$\in$" de relatie: "is een element van").

$\underline{G}$ is dus een betrouwbaarheidsgebied voor θ, met onbetrouwbaarheidsdrempel α.

Bewijs

Indien de werkelijke waarde θ getoetst wordt, is , per definitie van α, de kans op verwerping $\leqq \alpha$.
Hieruit volgt (15;9) direct.

Opmerking 15;5

Bij een toets T behoort dus een betrouwbaarheidsgebied $\underline{G}$. Omgekeerd wordt T weer uit $\underline{G}$ verkregen door een getoetste waarde θ_0 te verwerpen indien θ_0 niet in G ligt en niet, indien $\theta_0 \in G$.

Voorbeeld 15;1. Betrouwbaarheidsinterval voor een onbekende kans.

Wij leiden nu, volgens het in stelling 15;1 beschreven principe, een methode af om exacte betrouwbaarheidsgrenzen voor een onbekende kans te vinden.

Daarvoor is de volgende hulpstelling nodig.

Lemma 15;1

De rechteroverschrijdingskans

$$P[x \geqq x \mid p] = \sum_{i=x}^{n} \binom{n}{i} p^i q^{n-i} \tag{15;10}$$

van een binomiale verdeling, behorende bij een gegeven waarde van x, is een monotoon stijgende functie van p. (De linkeroverschrijdingskans is dus een monotoon dalende functie van p).

Bewijs

Beschouw één term van (15;10):

$$\varphi(p) \overset{def}{=} \binom{n}{i} p^i q^{n-i}. \tag{15;11}$$

Voor de afgeleide hiervan geldt

$$\varphi'(p) = \binom{n}{i} p^{i-1} q^{n-i-1} (i - np) \tag{15;12}$$

een relatie, die voor $i = 0$ en $i = n$ zijn geldigheid behoudt.

Uit (15;12) volgt

$$\varphi'(p) \begin{cases} > 0 & \text{voor } i > np, \\ = 0 & \text{voor } i = np \\ < 0 & \text{voor } i < np. \end{cases} \tag{15;13}$$

Is nu in (15;10) $x > np$, dan bezitten alle termen een positieve

afgeleide, zodat dit voor de som ook geldt. Is $x \leqq np$, dan is $x-1 < np$ en dan bezitten alle termen van de complementaire som (de linkeroverschrijdingskans van x) negatieve afgeleiden, dus hun som ook. Trekken wij deze som van 1 af, dan wordt weer (15;10) verkregen, die dus ook in dit geval een positieve afgeleide bezit, q.e.d.

Wij beschouwen nu eerst de rechtseenzijdige toets, die, bij gegeven n en α_x, berust op een kritieke waarde x_x, die echter nog afhangt van de getoetste waarde p_o. Deze waarde x_x is de kleinste, waarvoor geldt

(15;14) $$P[\underline{x} \geqq x_x \mid p_o] \leqq \alpha_x,$$

zodat

(15;15) $$P[\underline{x} \geqq x_x - 1 \mid p_o] > \alpha_x$$

is. Vergroten wij nu p_o, dan wordt volgens lemma 15;1 het linkerlid van (15;15) bij vasthouden van x_x, eveneens vergroot, zodat hij $> \alpha_x$ blijft. Dit betekent dat x_x door vergroting van p_o niet kan dalen,, maar - uiteraard - wel kan stijgen.

Bij een gegeven waargenomen waarde x van $\underline{x}$ (het aantal "successen" in n onafhankelijke proeven; dit is in dit geval het waarnemingsmateriaal), worden nu die waarden p_o verworpen, waarvoor $x > x_x$ is. De waarde $p_o = 1$ wordt zeker niet verworpen, want daarvoor is de rechter kritieke zone leeg. Laten wij p_o nu dalen vanaf de waarde 1, dan wordt voor een zekere waarde van p_o : $x_x = n$, en x_x daalt vervolgens trapsgewijze.

Zolang echter $x_x > x$ is, wordt p_o nog niet verworpen. Bij een zekere waarde van p_o gaat x_x echter van de waarde $x+1$ over in de waarde x. Deze waarde van p_o geven wij aan met p_ℓ, het is de eerste, die verworpen wordt. Alle waarden $p_o > p_\ell$ worden niet, alle waarden $\leqq p_\ell$ worden wel verworpen.

Volgens stelling 15;1 is dus p_ℓ, die een functie van het aantal successen x is, de betrouwbaarheidsondergrens voor p, zodat, als wij x weer als stochastische beschouwen, geldt:

(15;16) $$P[\underline{p}_\ell < p] \geqq 1 - \alpha_x,$$

waarin p de (onbekende) werkelijke waarde van de kans op succes voorstelt. De onbetrouwbaarheid α_x van de rechtseenzijdige toets is dus gelijk aan die van de betrouwbaarheidsondergrens. In deze laatste betekenis zou hij juist met α_ℓ aangegeven worden.

Uit het bovenstaande blijkt:

Stelling 15;2

Een betrouwbaarheidsondergrens p_ℓ voor een onbekende kans wordt verkregen door de grootste waarde van p_o op te zoeken, waarvoor geldt:

(15;17) $$x = x_\alpha,$$

met x = aantal successen,
en x_α = de bij de voor deze ondergrens gekozen onbetrouwbaarheidsdrempel behorende rechter kritieke waarde der binomiale toets.

Op analoge wijze wordt een betrouwbaarheidsbovengrens p_α voor p gevonden als de kleinste waarde van p_o, waarvoor

(15;18) $$x = x_\ell,$$

met x_ℓ = de linker kritieke waarde van de binomiale toets met de voor de bovengrens gekozen betrouwbaarheidsdrempel.

Beschikt men over exacte tabellen van de binomiale verdeling of van x_α, voor vele waarden van n en p, dan kan men dus bij gegeven n en x, de bijbehorende waarden van p_ℓ en p_α door opzoeken bepalen.

Voor $n=10$ en $x=0,1,\ldots,10$ is het resultaat daarvan voor $\alpha_\ell = \alpha_\alpha = 0,025$ in fig. 15;1 in beeld gebracht. De dik getrokken stukken der verticale lijnen stellen de betrouwbaarheidsintervallen voor, de gestippelde gedeelten de verworpen hypothesen. De gestippelde verbindingslijnen der grenspunten hebben slechts aanschouwelijke betekenis. De horizontale dienen voor het gemak van de aflezing *).

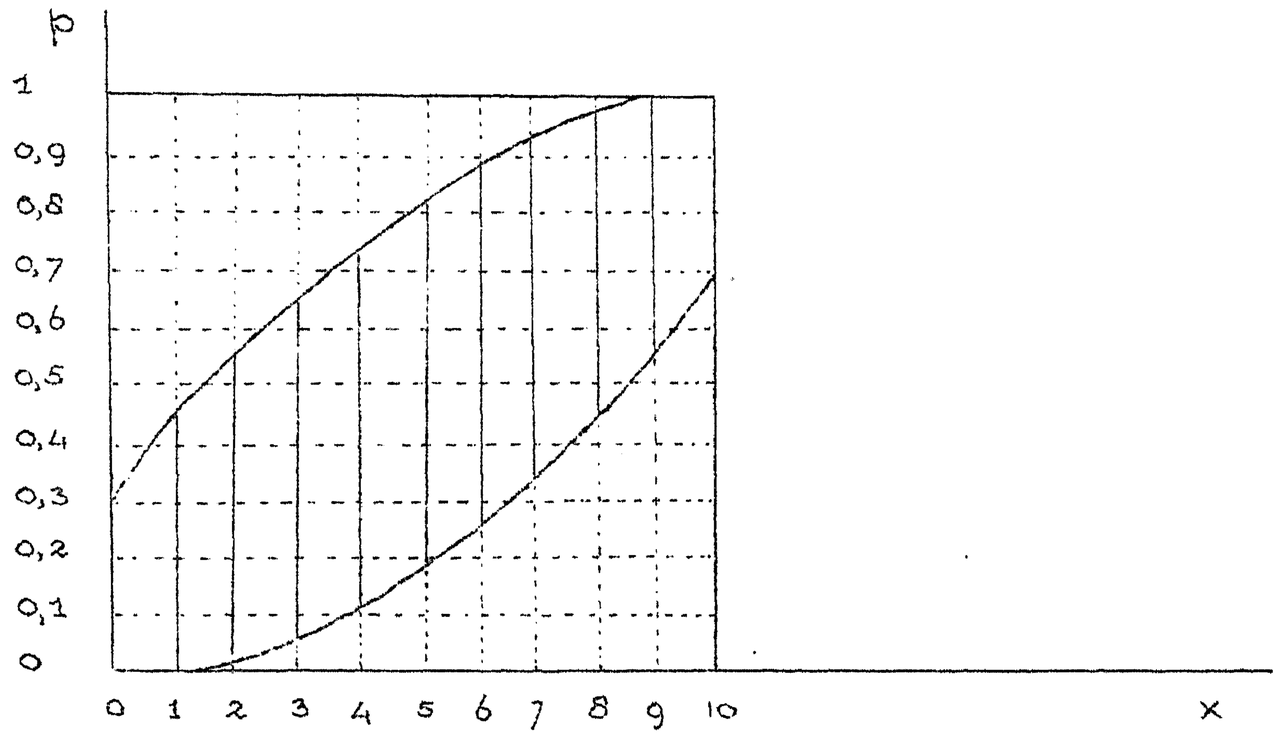


Fig. 15;1. Betrouwbaarheidsintervallen voor p bij $n=10$ en $\alpha_\ell = \alpha_\alpha = 0,025$

---

*) Deze figuur vindt men ook in het leerboek van Dr L.N.H. BUNT, "Statistiek voor het voorbereidend hoger en middelbaar onderwijs", Wolters, Groningen 1956, blz 150. Daar wordt de hier kort besproken theorie uitvoerig en elementair uiteengezet.

Voor grotere n kan men gebruik maken van de normale benadering met continuïteitscorrectie. Dit leidt tot expliciete formules voor p_ℓ en p_r, als functie van n, x, α_ℓ en α_r, die echter een benaderend karakter bezitten. Zij zijn echter veel nauwkeuriger dan de benaderingsformule (12;77), die op slordigere wijze is afgeleid. De beide formules zijn wel asymptotisch, voor $n \to \infty$, aequivalent.

Voor de afleiding van de benaderingsformules maken wij gebruik van formules (13;42) en (13;43) voor x_ℓ en x_r, die op de normale benadering met continuïteitscorrectie berusten. De tweede hiervan luidt:

$$x_r \approx n p_0 + u_{\alpha_r} \sqrt{n p_0 (1-p_0)} + \tfrac{1}{2}, \tag{15;19}$$

waarin u_{α_r} de bij rechteroverschrijdingskans α_r behorende waarde van de $N(0,1)$-verdeling is (tabel 1) en p_0 de getoetste waarde voorstelt. Deze formule geldt, zolang de uitkomst van het rechterlid $\leqq n$ is; is deze $> n$, dan is de benaderde rechter kritieke zone leeg.

Lemma 15;1 doet vermoeden, dat x_r (eigenlijk wordt bedoeld: het rechterlid van (15;19)), bij gegeven n en α_r, een monotoon stijgende functie, van p_0 zal zijn. Daar het hier een benaderingsformule betreft is het wenselijk dit te controleren.

Daartoe beschouwen wij de afgeleide naar p (voor het gemak laten wij de indices 0 en α_r in deze berekening weg).

Deze is

$$x_r' = n + \tfrac{1}{2} u n (1-2p) / \sqrt{n p (1-p)}.$$

Voor kleine p is deze inderdaad positief, maar voor p dicht bij 1 wordt hij negatief. Wij tonen nu aan dat dit pas het geval is voor zo grote p, dat daarvoor $x_r > n$ is. Daartoe berekenen wij het punt, waar $x_r' = 0$ is. Dit is het geval voor

$$\tfrac{1}{2} u (2p-1) = \sqrt{n p (1-p)}.$$

Het rechterlid van deze vergelijking is, voor $p > 0$, positief, het linkerlid dus ook. Dit betekent, dat de oplossing een waarde van p is, die $> \frac{1}{2}$ is.

Lossen wij de vergelijking nu op, dan verkrijgen wij als oplossing

$$p = \tfrac{1}{2}\left\{1 + \sqrt{\frac{n}{n+u^2}}\right\} \quad ; \quad 1-p = \tfrac{1}{2}\left\{1 - \sqrt{\frac{n}{n+u^2}}\right\}.$$

Invullen van deze waarden in (15;19) geeft na herleiding

$$x_r = \tfrac{1}{2} n \left\{1 + \sqrt{\frac{n+u^2}{n}}\right\} > n,$$

waarmee het gestelde bewezen is.

Wij kunnen nu dus (15;17) toepassen om de bij een waarde x behorende p_ℓ te berekenen. Dit geeft

$$(15;20) \qquad x \approx n p_\ell + u_{\alpha_\ell}\sqrt{n p_\ell (1-p_\ell)} + \tfrac{1}{2},$$

waarin nu α_ℓ ($= \alpha_n$ uit (15;19)) de onbetrouwbaarheidsdrempel van p_ℓ voorstelt. Schrijven wij deze vergelijking als

$$x - \tfrac{1}{2} - n p_\ell \approx u_{\alpha_\ell}\sqrt{n p_\ell (1-p_\ell)}$$

dan zien wij, dat het rechterlid $\geqq 0$ is, zodat dus moet gelden

$$p_\ell \leqq \tfrac{1}{n}(x - \tfrac{1}{2}).$$

Deze ongelijkheid geeft aan welke der beide oplossingen van de vierkantsvergelijking, die door kwadrateren ontstaat, de juiste is. Deze blijkt te zijn

$$(15;21) \qquad p_\ell \approx \frac{(x-\tfrac{1}{2}) + \tfrac{1}{2}u^2_{\alpha_\ell} - \sqrt{u^2_{\alpha_\ell}\left\{\frac{(x-\frac{1}{2})(n-x+\frac{1}{2})}{n} + \frac{1}{4}u^2_{\alpha_\ell}\right\}}}{n + u^2_{\alpha_\ell}},$$

Op analoge wijze verkrijgt men voor de bovengrens

$$(15;22) \qquad p_n \approx \frac{(x+\tfrac{1}{2}) + \tfrac{1}{2}u^2_{\alpha_n} + \sqrt{u^2_{\alpha_n}\left\{\frac{(x+\frac{1}{2})(n-x-\frac{1}{2})}{n} + \frac{1}{4}u^2_{\alpha_n}\right\}}}{n + u^2_{\alpha_n}},$$

terwijl beide grenzen tezamen een betrouwbaarheidsinterval vormen met $\alpha = \alpha_n + \alpha_\ell$.

Opmerking 15;5

Voor grote n en x wordt, door verwaarlozing van kleine termen in teller en noemer, formule (12;77) teruggekregen

Zet men de zo verkregen grenzen voor $n=100$ uit tegen x/n als abcis , dan wordt fig. 15;2 verkregen. Doet men dit voor verschillende waarden van n, dan verkrijgt men een bundel van dergelijke lijnen *).

---

*) Zie b.v. W.J. DIXON and F.J. MASSEY Jr., "Introduction to statistical analysis", 2nd ed., McGraw-Hill, N.Y.- Toronto-London 1957, table A-9 of L.N.H. BUNT, l.c. fig. 23.

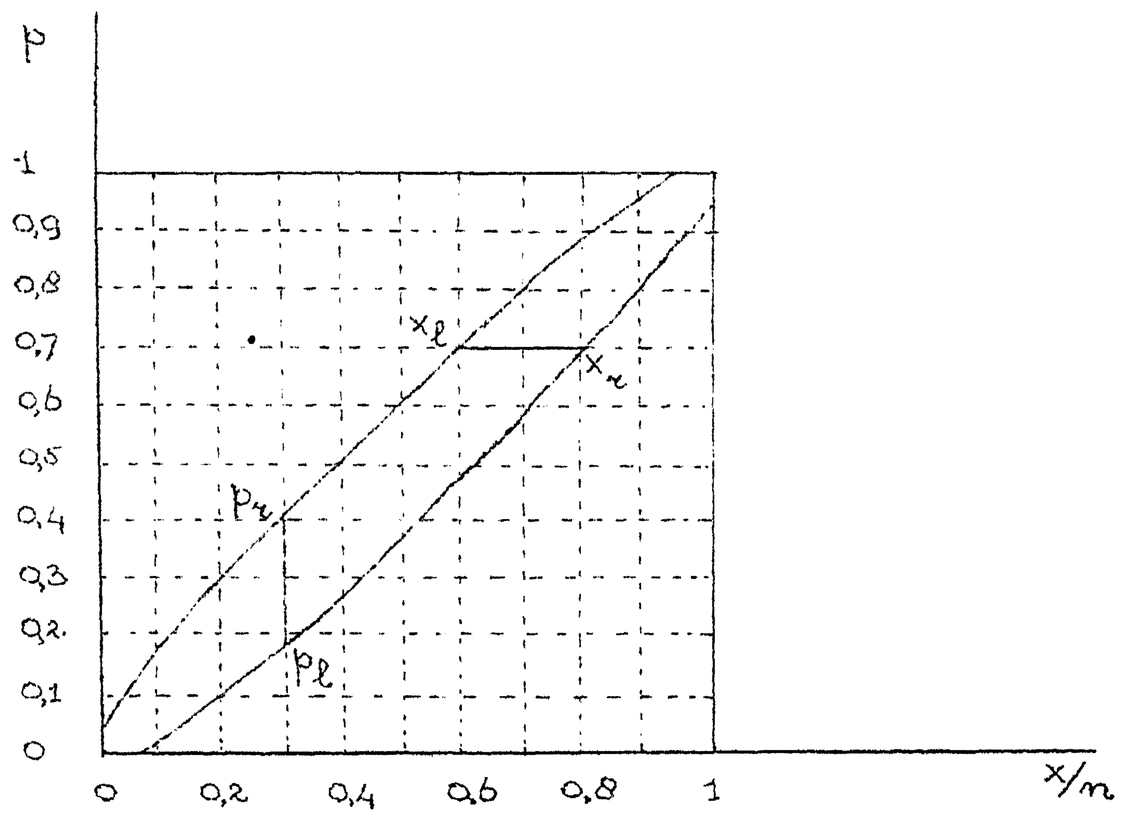


Fig. 15;2. Benaderde betrouwbaarheidsintervallen voor p' ($n = 100$, $\alpha_l = \alpha_r = 0{,}025$)

Geven dus de verticale lijnen in deze figuur betrouwbaarheidsintervallen voor p aan (nl. p_l en p_r als functie van x), men kan zich ook afvragen, wat de interpretatie is van de snijpunten der twee krommen met een horizontale lijn. De vergelijkingen der beide krommen worden echter gegeven door (15;20), waarvan het rechterlid (dus ook het linkerlid) x_r voorstelt, en een daaraan analoge uitdrukking voor x_l. De snijpunten, bij constante p, zijn dus niet anders dan de kritieke waarden x_l en x_r, behorende bij deze waarden van p als getoetste waarde.

Behalve uit de formule valt dit ook direct in te zien. Immers ieder punt buiten de "vis" komt overeen met een x en een p, die niet met elkaar in overeenstemming zijn: de waarde van p wordt, bij toetsing op grond van x, verworpen. De buiten de vis gelegen punten van een horizontale lijn (dus bij één p behorend) vormen dus tezamen alle waarden van x, die tot verwerping van deze waarde p leiden, dus juist de kritieke zones.

In de figuur komt duidelijk tot uiting dat x_l met p_r en x_r met p_l correspondeert.

Het oorspronkelijk door J. NEYMAN bij de afleiding van deze methode gegeven bewijs verliep juist omgekeerd *). Zijn redenatie was: teken de kromme x_r en x_l als functie van p_o, de getoetste waarde. Er is één horizontale lijn (onbekend welke), die bij de werkelijke p behoort. Hierop komt dus het punt (p, x) terecht. De kans, dat het links van (of in) x_l valt is $\leqq \alpha_l$ (als toetsings-onbetrouwbaarheid) en de kans, dat het rechts van x_r komt is $\leqq \alpha_r$. Realiseert

---

*) In zijn allereerste publicatie hierover gaf NEYMAN een veel ingewikkelder bewijs, dat in een later artikel tot het hieronder gegevene werd herleid.

zich geen van deze beide mogelijkheden, dan komt het punt $(\mu, \underline{x})$ dus in de vis terecht (al weten wij niet waar) en deze kans is dus $\geqq 1-\alpha$. Wij voorspellen nu:
"Het punt $(\mu, \underline{x})$ komt in de vis terecht",
een voorspelling met kans $\geqq 1-\alpha$ op juistheid. Vervolgens wordt $\underline{x}$ waargenomen; dit geeft een waarde x. Het punt (μ, x) ligt nu, volgens de voorspelling, in de vis, dus de bewering:
"Het punt (μ, x) ligt in de vis"
heeft een betrouwbaarheid $\geqq 1-\alpha$. Maar deze bewering is aequivalent met: μ ligt tussen de snijpunten, die de verticale lijn door x met de beide krommen heeft. Dit zijn dus betrouwbaarheidsgrenzen.

Opgave 15;1

In verband met de bij boekingen voor bungalows te volgen politiek wil een reisbureau inzicht hebben in de kans op afzeggen van een vroeg gemaakte reservering. Uit de gegevens van het vorige jaar blijkt, dat van 221 vroege boekingen 33 tenslotte niet zijn doorgegaan. Overwogen wordt in de winter en lente een zekere overboeking toe te laten. Daartoe wenst men te beschikken over een boven- en een ondergrens voor de kans op afzeggen van een reeds geboekte bungalow.

Bereken een betrouwbaarheidsinterval voor deze kans met tweezijdige onbetrouwbaarheid 0,01.

Oplossing:

$\alpha_\ell = \alpha_r = 0{,}005$, dus $u_{\alpha_\ell} = u_{\alpha_r} = 2{,}575$ (tabel 1). Invullen in (15;21) en (15;22) geeft

$$p_\ell = 0{,}096 \qquad \text{en} \qquad p_r = 0{,}224.$$

Eigenschappen van betrouwbaarheidsintervallen

Definitie 15;2. Een betrouwbaarheidsgebied G heet zuiver met betrekking tot de parameter θ, indien voor iedere van de werkelijke waarde θ verschillende waarde θ' geldt:

(15;23) $$P[\theta \in \underline{G}] \geqq P[\theta' \in \underline{G}].$$

Stelling 15;2

Indien een toets T, voor de hypothese $\theta = \theta_o$, berust op een continu verdeelde toetsingsgrootheid en voor iedere toegelaten waarde van θ_o zuiver is, dan is het bijbehorende betrouwbaarheidsgebied voor θ ook zuiver.

Bewijs

De continuïteit van de toetsingsgroctheid heeft tot gevolg, dat onbetrouwbaarheid en onbetrouwbaarheidsdrempel gelijk zijn. Geef beide aan met α .

Wordt nu een onjuiste waarde θ' getoetst, dan is, wegens de zuiverheid van de toets, de kans op verwerping van θ' minstens gelijk aan α . De kans, dat θ' niet in $\underline{G}$ ligt, is dus $\geqq \alpha$ (en in het algemeen uiteraard $> \alpha$).

Wordt echter de werkelijke waarde θ getoetst, dan is de kans op verwerping daarvan $\leqq \alpha$.

Opmerkingen

15;6 Voor tweezijdige toetsen is het begrip zuiverheid niet erg zinvol, omdat (vgl. bl. 128 bovenaan) het onderscheidingsvermogen in dat geval verkregen wordt door optelling van de kansen op twee geheel ongelijksoortige gebeurtenissen, nl. het trekken van de juiste conclusie enerzijds en het trekken van een falikant verkeerde anderzijds. Voor tweezijdige betrouwbaarheidsintervallen is het begrip zuiverheid echter wel zinvol.

15;7 Eénzijdige betrouwbaarheidsintervallen zijn nooit zuiver in bovengenoemde zin. Immers is, voor een ondergrens t_ℓ , de relatie $t_\ell < \theta$ juist, dan geldt ook $t_\ell < \theta'$ voor iedere $\theta' > \theta$. Dit betekent, dat in plaats van (15;23) geldt:

$$(15;24) \qquad P[\underline{t}_\ell < \theta] \leqq P[\underline{t}_\ell < \theta'] \qquad \text{voor iedere } \theta' > \theta,$$

hetgeen juist overeenkomt met onzuiverheid.

Naar de andere kant, nl. voor $\theta'' < \theta$, is een ondergrens echter altijd zuiver. Immers is $t_\ell < \theta''$, dan ook $t_\ell < \theta$ (daar $\theta'' < \theta$), dus geldt

$$(15;25) \qquad P[\underline{t}_\ell < \theta] \geqq P[\underline{t}_\ell < \theta''] \qquad \text{voor iedere } \theta'' < \theta.$$

Analoge relaties gelden voor een bovengrens t_ν . De al of niet zuiverheid van de bijbehorende toets is hierbij niet ter sprake gebracht, hetgeen misschien de indruk zou kunnen wekken, dat deze voor dit resultaat niet van belang is. Dit is echter slechts schijn. Immers het is gemakkelijk uit (15;24 en 25) de (éénzijdige) zuiverheid van de bijbehorende toets af te leiden. Dit betekent dus, dat de bij een éénzijdig betrouwbaarheidsinterval behorende toets altijd zuiver is; een onzuivere toets kan nooit tot een éénzijdig begrensd betrouwbaarheidsgebied leiden.

Definitie 15;3

Een betrouwbaarheidsinterval voor θ heet asymptotisch verdwijnend als voor de (stochastische) lengte $\underline{L} = \underline{t}_r - \underline{t}_l$ geldt:

(15;26) $$\lim_{n\to\infty} P[\underline{L} > \varepsilon] = 0 \qquad \text{voor iedere } \varepsilon > 0.$$

Stelling 15;3

Wordt een betrouwbaarheidsondergrens $\underline{t}_l$ afgeleid uit een éénzijdige toets T, voor de hypothese $\theta \leqq \theta_o$, die voor iedere θ_o asymptotisch onderscheidend is met betrekking tot iedere $\theta' > \theta_o$, dan geldt:

(15;27) $$\lim_{n\to\infty} P[\underline{t}_l < \theta - \tfrac{1}{2}\varepsilon] = 0 \qquad \text{voor iedere } \varepsilon > 0.$$

Hetzelfde geldt, mutatis mutandis voor $\underline{t}_r$.

Bewijs

$t_l < \theta - \frac{1}{2}\varepsilon$ betekent, dat de waarde $\theta - \frac{1}{2}\varepsilon$ niet verworpen wordt. De kans hierop nadert echter voor $n \to \infty$ tot 0.

Stelling 15;4

Wordt een tweezijdige betrouwbaarheidsinterval afgeleid uit een tweezijdige toets, samengesteld uit twee eenzijdige asymptotisch onderscheidende toetsen, dan is het betrouwbaarheidsinterval asymptotisch verdwijnend.

Bewijs

Volgens de vorige stelling geldt (15;27) en

(15;28) $$\lim_{n\to\infty} P[\underline{t}_r > \theta + \tfrac{1}{2}\varepsilon] = 0 \qquad \text{voor iedere } \varepsilon > 0.$$

Daar voor ieder tweetal gebeurtenissen geldt:

$$P[A \vee B] \leqq P[A] + P[B]$$

volgt hieruit, dat ook geldt:

$$\lim_{n\to\infty} P[\underline{t}_l < \theta - \tfrac{1}{2}\varepsilon \vee \underline{t}_r > \theta + \tfrac{1}{2}\varepsilon] = 0.$$

dus

$$\lim_{n\to\infty} P[\theta - \tfrac{1}{2}\varepsilon \leqq \underline{t}_l < \underline{t}_r \leqq \theta + \tfrac{1}{2}\varepsilon] = 1.$$

waaruit (15;26) volgt.

Heeft van twee toetsen T_1 en T_2 voor eenzelfde hypothese T_1 overal een groter onderscheidingsvermogen dan T_2, dan verdient het bij T_1 behorende betrouwbaarheidsgebied $\underline{G}_1$ de voorkeur boven het bij T_2 behorende $\underline{G}_2$. Immers voor iedere $\theta' \neq \theta$ geldt dan

(15;30) $$P[\theta' \in \underline{G}_1] < P[\theta' \in \underline{G}_2],$$

hetgeen betekent, dat $\underline{G}_1$ in de regel korter zal zijn dan $\underline{G}_2$.

Voorbeeld 15;2 Betrouwbaarheidsgrenzen voor de mediaan

Laat van een grootheid $\underline{x}$ n onafhankelijke waarnemingen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ beschikbaar zijn, terwijl gevraagd wordt een (éénzijdige of tweezijdige) betrouwbaarheidsinterval voor de mediaan x_M te bepalen. Wij veronderstellen - om complicaties te vermijden - dat $\underline{x}$ continu verdeeld is, zodat geen gelijke waarden onder de waarnemingen voorkomen (althans de kans daarop is 0). Voor x_M geldt dan:

(15;31) $$P[\underline{x} < x_M] = P[\underline{x} > x_M] = \tfrac{1}{2}.$$

Wij rangschikken nu de waarnemingen naar opklimmende grootte en geven ze dan aan met $\underline{x}_{(1)}, \underline{x}_{(2)}, \ldots, \underline{x}_{(n)}$, waarvoor dus geldt

(15;32) $$P[\underline{x}_{(1)} < \underline{x}_{(2)} < \ldots < \underline{x}_{(n)}] = 1.$$

Noem nu $\underline{a}$ het aantal der waarnemingen, die kleiner zijn dan een te toetsen waarde $x_{M,o}$ voor de mediaan, dan bezit $\underline{a}$, onder de hypothese $x_M = x_{M,o}$, een binomiale verdeling met $p = \frac{1}{2}$, zodat voor toetsing van tabel 2, de tekentoets, gebruik gemaakt kan worden.

Laat a_ℓ de in die tabel gevonden (of eventueel de met behulp van de desbetreffende formule berekende) linker kritieke waarde zijn, behorende bij een onbetrouwbaarheidsdrempel α_ℓ. Dan wordt dus $x_{M,o}$ verworpen indien $\underline{a}$ een waarde $\leqq a_\ell$ aanneemt. Nu is echter $a \leqq a_\ell$ equivalent met: onder $x_{M,o}$ liggen a_ℓ of minder waarnemingen. Of ook: $x_{M,o} \leqq x_{(a_\ell+1)}$.

Alle waarden $\leqq x_{(a_\ell+1)}$ worden dus verworpen, zodat volgens stelling 15;1 geldt:

(15;33) $$P[\underline{x}_{(a_\ell+1)} < x_M] \geqq 1 - \alpha_\ell.$$

De betrouwbaarheidsondergrens is dus de $(a_\ell+1)^e$ waarneming bij rangschikking naar grootte.

Om redenen van symmetrie moet de betrouwbaarheidsbovengrens met $\alpha_r = \alpha_\ell$, op dezelfde wijze gevonden worden, maar nu van "rechts" naar "links" werkende. In plaats van de $(a_\ell+1)^e$ van onderen af moet nu dus de $(a_\ell+1)^e$ van boven af genomen worden. Deze draagt het nummer $n-a_\ell$ (N.B.: niet $n-a_\ell-1$!), zodat wij verkrijgen:

(15;34) $$P[x_M < \underline{x}_{(n-a_\ell)}] \geq 1-\alpha_\ell.$$

Een tweezijdige betrouwbaarheidsinterval wordt verkregen door (15;33) en (15;34) te combineren, De notatie kan vereenvoudigd worden als wij $\alpha_\ell = \alpha_r = \frac{1}{2}\alpha$ stellen. Geven wij de linker-kritieke waarde van de tekentoets voor dat geval aan met r, zodat de rechter-kritieke waarde $n-r$ wordt, dan geldt:

(15;35) $$P[\underline{x}_{(r+1)} < x_M < \underline{x}_{(n-r)}] \geq 1-\alpha,$$

terwijl de beide grenzen ieder afzonderlijk $\frac{1}{2}\alpha$ als onbetrouwbaarheidsdrempel hebben.

Opmerking 15;8

Hoewel wij verondersteld hebben, dat $\underline{x}$ continu verdeeld is, is toch de onbetrouwbaarheid hier niet gelijk aan de onbetrouwbaarheidsdrempel, omdat de toetsingsgrootheid discreet verdeeld is.

Is $\underline{x}$ discreet verdeeld, dan treden complicaties op, die wij hier buiten beschouwing laten.

Opgave 15;2

Het tingehalte van 20 aselect uit een grondlaag genomen monsters bedraagt, uitgedrukt in de een of andere eenheid:
21,4; 13,7; 14,0; 18,2; 7,7; 16,3; 27,4; 17,4; 23,1; 5,6; 22,0; 16,9; 27,9; 22,8; 10,8; 23,5; 13,5; 19,0; 17,8; 18,0.

Bepaal een tweezijdig betrouwbaarheidsinterval voor de mediaan van het tingehalte, met onbetrouwbaarheidsdrempel 0,05.

Oplossing: Volgens tabel 2 is r=5. Bij rangschikking naar grootte blijken de grenzen te zijn: 14,0 en 22,0.

Voorbeeld 15;3 Betrouwbaarheidsgrenzen voor

Indien van een reeks van n onafhankelijke experimenten, met voor iedere kans p op succes, de waarde van p bekend is, maar n niet, kan men door directe toepassing van stelling 15;1 betrouwbaarheidsgrenzen voor deze onbekende n bepalen. De binomiale toets wordt daar-

bij dan niet gebruikt als toets voor hypothesen van de vorm $p = p_0$, maar, bij de bekende p, voor hypothesen van de vorm $n = n_0$.

Opgave 15;3

Bij een reeks van n onafhankelijke experimenten, ieder met kans $\frac{1}{2}$ op succes, worden 20 sucessen verkregen. Het aantal (n) is onbekend.

a. Gevraagd wordt een tweezijdig betrouwbaarheidsinterval voor n te bepalen met $\alpha = 0{,}05$,
b. Los de opgave ook op als $p = \frac{2}{3}$ is,
c. Leid (met behulp van de normale benadering zonder continuïteitscorrectie) eem algemene formule voor dit betrouwbaarheidsintervan af.

(Opgave 21 uit het sommenboekje; oplossing op blz. 63-66.)

Voorbeeld 15;4 Betrouwbaarheidsgrenzen voor een symmetriepunt

Indien de grootheid x, waarvan n onafhankelijke waarnemingen beschikbaar zijn, symmetrisch verdeeld is om een onbekend punt ξ (dat dan tevens $= x_m$ en $= \mu$ is), kan de symmetrietoets van WILCOXON gebruikt worden om een betrouwbaarheidsinterval voor ξ te bepalen.

De procedure is eenvoudig, maar eist, indien men geen vaardigheid in de toepassing bezit, nogal wat rekenwerk. Het komt eenvoudig hierop neer, dat men in plaats van de waarnemingen zelf de getallen

(15;36) $$x_1 - \Delta, x_2 - \Delta, \ldots, x_n - \Delta$$

beschouwt en daarop, voor verschillende waarden van Δ, de tweezijdige symmetrietoets toepast. Er zijn twee grenswaarden voor Δ, Δ_k en Δ_ℓ, waarbij juist verworpen moet worden, terwijl dit voor intermediaire waarden niet het geval is. Dit zijn dan de betrouwbaarheidsgrenzen voor ξ.

De procedure kan ook éénzijdig worden uitgevoerd. De kritieke waarden van de toets kunnen aan tabel 4 of aan de desbetreffende formules worden ontleend.

Opgave 15;4

De gegevens van opgave 15;2 zijn ontleend aan een symmetrische verdeling. Bepaal een tweezijdige betrouwbaarheidsinterval voor het symmetriepunt, met $\alpha = 0{,}05$.

Oplossing

Δ	T	
15,0	109	$n = 20$, rechterkritieke waarde = 106
15,1	108	
15,2	105	
21	-122	$n = 20$, linker kritieke waarde = -106
20,7	-110	
20,6	-104	

Het betrouwbaarheidsinterval is dus

$$15,1 < \xi < 20,7.$$

Voorbeeld 15;5 Betrouwbaarheidsinterval voor de verschuiving van één verdeling ten opzichte van een andere

Indien twee stochastische grootheden $\underline{x}$ en $\underline{y}$ op een verschuiving na dezelfde verdeling bezitten, d.w.z. als

(15;37) $\underline{x} + \Delta$ en $\underline{y}$

voor een onbekende waarde van Δ dezelfde verdeling bezitten, dan kan, op grond van twee onafhankelijke steekproeven $\underline{x}_1, \ldots, \underline{x}_m$ en $y_1, \ldots, y_n$ een betrouwbaarheidsinterval voor de "verschuiving" Δ bepaald worden.

Dit geschiedt geheel analoog aan het vorige voorbeeld, maar nu met behulp van de toets van WILCOXON voor twee steekproeven. De hypothese $\Delta = \Delta_0$ kan, met behulp van die toets, voor iedere Δ_0 getoetst worden en de niet verworpen waarden vormen tezamen het betrouwbaarheidsgebied.

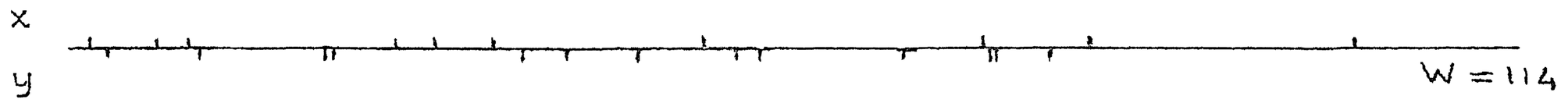
De uitvoering is in dit geval eenvoudiger dan in het geval van de symmetrietoets. Door de beide steekproeven ieder op een strookje millimeterpapier uit de zetten en deze langs elkaar te schuiven, zijn de kritieke verschuivingen (waarbij nog juist niet of net wel verworpen wordt) gemakkelijk te bepalen. Bij verschuiving van het ene strookje ten opzichte van het andere verandert de waarde van de toetsingsgrootheid met 2 wanneer een x-streepje een y-streepje passeert.

Opgave 15;5

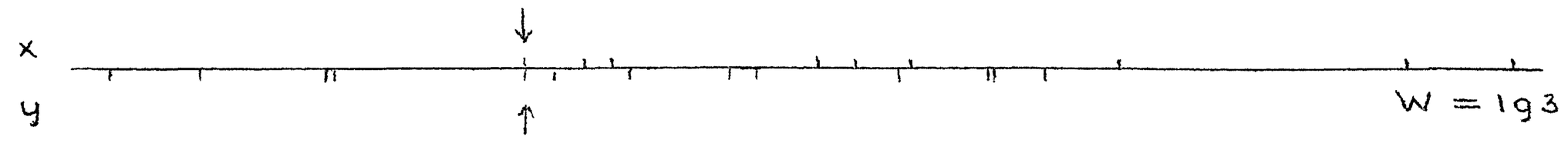
Bepaal uit de gegevens van opgave 14.2;3 een tweezijdig betrouwbaarheidsinterval, met $\alpha = 0,05$, voor de verschuiving van de verdelingen ten opzichte van elkaar.

Oplossing: $m = 10$, $n = 13$, $\alpha = 0,05$; de kritieke waarden van W zijn dus 66 en 194.

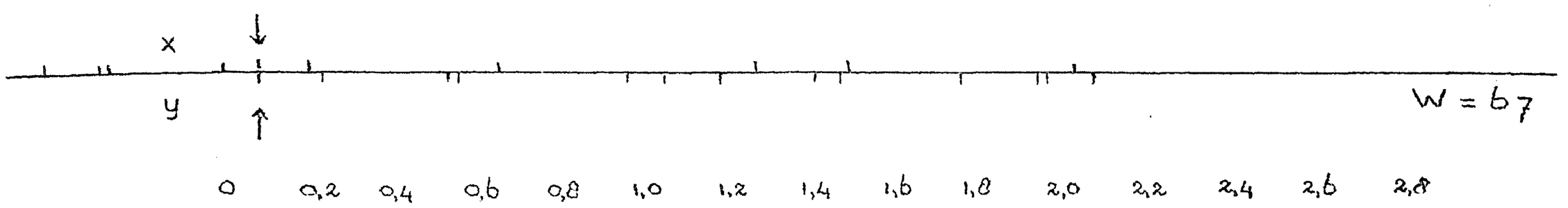
Zonder verschuiving ($\Delta = 0$):



Met verschuiving van x over 0,93 naar rechts:



Met verschuiving van x over 0,69 naar links:



Het interval is:

$$-0{,}93 \leqq \Delta \leqq 0{,}69.$$

<u>Afleiding van betrouwbaarheidsgrenzen met behulp van de schattingstheorie</u>

Een andere methode om betrouwbaarheidsintervallen af te leiden berust op de schattingsthecrie. Het principe van deze methode kan als volgt uiteengezet worden.

Laat $\underline{t}$ een schatting van de onbekende parameter θ zijn met $\mathcal{E}_{\underline{t}} = \theta$ en $\sigma^2(\underline{t}) = \sigma^2$ onafhankelijk van θ. Laat verder de verdeling van $\underline{t}$, op de onbekendheid van θ na, bekend zijn, b.v. normaal. Dan geldt:

(15;38) $$P[\theta - u_{\frac{1}{2}\alpha}\sigma < \underline{t} < \theta + u_{\frac{1}{2}\alpha}\sigma] = 1-\alpha.$$

Het interval $(\theta - u_{\frac{1}{2}\alpha}\sigma\ ,\ \theta + u_{\frac{1}{2}\alpha}\sigma\ )$ heet een <u>voorspellingsinterval</u> voor t; de voorspelling, dat $\underline{t}$ in dit interval zal liggen heeft een kans $1-\alpha$ juist te zijn.

Dit interval laat zich echter gemakkelijk omrekenen tot een betrouwbaarheidsinterval voor θ. Immers (15;38) is equivalent met

(15;39) $$P[\underline{t} - u_{\frac{1}{2}\alpha}\sigma < \theta < \underline{t} + u_{\frac{1}{2}\alpha}\sigma] = 1-\alpha$$

en het interval $(\underline{t} - u_{\frac{1}{2}\alpha}\sigma\ ,\ \underline{t} + u_{\frac{1}{2}\alpha}\sigma\ )$ is nu een betrouwbaarheidsinterval met onbetrouwbaarheid α.

Op dit principe, zij het soms in gewijzigde vorm, berust de afleiding van vele betrouwbaarheidsintervallen.

Volgens opmerking 15;5 behoort bij zo'n interval ook weer een toets, zodat er geen essentieel verschil met de in het voorafgaande behandelde methode is.

Een moeilijkheid is echter, dat de verdeling van $\underline{t}$ vaak niet zo eenvoudig van θ afhangt, dat een vergelijking van de vorm (15;38) opgeschreven kan worden.

Beschouwen wij b.v. het geval van een binomiale grootheid $\underline{x}$ met parameters n en p, waarvoor dus geldt: $\mathcal{E}\underline{x} = np$ en $\sigma^2(\underline{x}) = npq$, dan hangen verwachting en spreiding beide van n <u>en</u> van p af. Zowel indien bij bekende n een betrouwbaarheidsinterval voor p bepaald moet worden (voorbeeld15;1) als wanneer p bekend is en een betrouwbaarheidsinterval voor n gezocht wordt (voorbeeld 15;3), leidt dit tot complicaties, die gewoonlijk omzeild worden op de in hoofdstuk 12 beschreven wijze: vervanging van σ door een schatting daarvan zonder rekening te houden met het stochastische karakter daarvan (resultaat: formule (12;77)). Deze moeilijkheid treedt echter niet op indien men van de toetsingstheorie uitgaat.

In nog sterkere mate geldt dit voor de voorbeelden 15;2, 4 en 5. In die gevallen berust de toets op een toetsingsgrootheid, die geen schatting van de onbekende parameter is, zodat directe toepassing van de schattingstheorie niet mogelijk is. En hoewel men in alle drie genoemde gevallen wel een schatting voor de onbekende parameter kan aangeven, is de verdeling van deze schatting of wel onbekend, ofwel afhankelijk van de vorm van de verdeling der waarnemingen ofwel beide, zodat het met behulp van de toetsingstheorie opgeloste probleem uitgaande van de schattingstheorie niet opgelost is of zelfs niet kan worden.

Anderzijds zijn er echter gevallen, waarin de schattingstheorie even snel als de toetsingstheorie tot een oplossing (en dan ook tot dezelfde oplossing) voert.

<u>Voorbeeld 15;6 Betrouwbaarheidsgrenzen voor de verwachting van een normale verdeling met onbekende spreiding</u>

Indien $\underline{x}$ $N(\mu,\sigma)$-verdeeld is, en $x_1, x_2, \ldots, x_n$ vormen een steekproef uit deze verdeling, dan bezit volgens stelling 14.2;1 de grootheid

$$\underline{t} = \frac{\bar{\underline{x}} - \mu}{\underline{s}} \sqrt{n} \qquad (15;38)$$

een STUDENT-verdeling met $n-1$ vrijheidsgraden.

In tabel 2 kan men nu de waarde $t_{\frac{1}{2}\alpha}$ opzoeken, die een tweezijdige overschrijdingskans α bezit en daarvoor geldt dus

(15;39) $$P\left[\,|\underline{t}| < t_{\frac{1}{2}\alpha}\right] = 1-\alpha$$

of, na enige herleiding

(15;40) $$P\left[\,\overline{\underline{x}} - t_{\frac{1}{2}\alpha}\frac{\underline{s}}{\sqrt{n}} < \mu < \overline{\underline{x}} + t_{\frac{1}{2}\alpha}\frac{\underline{s}}{\sqrt{n}}\,\right] = 1-\alpha.$$

Dit is een tweezijdig betrouwbaarheidsinterval met onbetrouwbaarheid α, ieder der betrouwbaarheidsgrenzen, die de eindpunten van dit interval zijn, bezit afzonderlijk onbetrouwbaarheid $\frac{1}{2}\alpha$.

Opgave 15;6

Bereken, in de veronderstelling van normaliteit, een tweezijdig betrouwbaarheidsinterval voor μ, met onbetrouwbaarheid 0,05, uit de gegevens van opgave 15;2.

Oplossing: $\bar{x} = \frac{1}{n}\sum x_i = 17{,}85$; $s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 = 34{,}99$; $s = 5{,}8$; $t_{\frac{1}{2}\alpha} = t_{0{,}025} = 2{,}093$ $(\nu = 19)$. De grenzen worden: 15,1 en 20,6.

Voorbeeld 15;7 Betrouwbaarheidsgrenzen voor het verschil van de verwachtingen van twee normale verdelingen met gelijke spreiding

Indien $\underline{x}$ en $\underline{y}$ beide normaal verdeeld zijn, met verwachtingen μ_1 resp. μ_2 en met dezelfde spreiding σ, terwijl $x_1, x_2, \ldots, x_m$ en $y_1, y_2, \ldots, y_n$ twee onafhankelijke steekproeven uit deze verdelingen zijn, dan heeft volgens stelling 14.2;2 de grootheid

(15;41) $$\underline{t} = \frac{\overline{\underline{x}} - \overline{\underline{y}} - (\mu_1 - \mu_2)}{\sqrt{\underline{S}_1^2 + \underline{S}_2^2}}\sqrt{\frac{mn(m+n-2)}{m+n}}$$

een STUDENT-verdeling met $m+n-2$ vrijheidsgraden.

Derhalve geldt dan weer (15;39), waarbij $t_{\frac{1}{2}\alpha}$ in de juiste regel van tabel 2 moet worden opgezocht.

Nu valt deze vergelijking te herleiden tot

(15;42) $$P\Big[\overline{\underline{x}} - \overline{\underline{y}} - t_{\frac{1}{2}\alpha}\sqrt{(\underline{S}_1^2 + \underline{S}_2^2)\frac{m+n}{mn(m+n-2)}} < \mu_1 - \mu_2 < \overline{\underline{x}} - \overline{\underline{y}} + t_{\frac{1}{2}\alpha}\sqrt{(\underline{S}_1^2 + \underline{S}_2^2)\frac{m+n}{mn(m+n-2)}}\,\Big] = 1-\alpha.$$

Opgave 15;7

Bepaal, in de veronderstelling van normaliteit, uit de gegevens van opgave 14.2;3 een tweezijdig betrouwbaarheidsinterval, met $\alpha = 0{,}05$, voor $\mu_1 - \mu_2$.

(Opgave 80 b) van het sommenboekje; oplossing op bl. 119: $-0{,}755 < \mu_1 - \mu_2 < 0{,}609$.)

Voorbeeld 15;8 Benaderde betrouwbaarheidsgrenzen voor het verschil van twee onbekende kansen

Een toets voor de gelijkheid van 2 kansen ($H_0: p_1 = p_2$) is in het begin van hoofdstuk 14 behandeld. Deze toets leent zich echter niet voor het toetsen van hypothesen van de vorm $p_1 - \Delta = p_2$, zodat hieraan geen betrouwbaarheidsinterval voor $\Delta = p_1 - p_2$ te ontlenen valt. Een dergelijk interval moet dus op andere wijze afgeleid worden en hier voor is geen andere methode bekend dan de aan het eind van hoofdstuk 12 reeds beschreven methode "voor grote steekproeven".

Beschikken wij over 2 onafhankelijke reeksen van n_1 resp. n_2 proeven, waarvan de ene betrekking heeft op p_1 en de andere op p_2, en zijn de aantallen successen $\underline{x}_1$ resp. $\underline{x}_2$, dan zijn deze grootheden binomiaal verdeeld en onafhankelijk van elkaar.

Als schatting voor p_1 resp. p_2 kunnen $\underline{x}_1/n_1$ resp. $\underline{x}_2/n_2$ dienen (vgl. (12;31)) en als schatting voor $p_1 - p_2$ gebruiken wij daarom de grootheid

$$\underline{d} = \frac{\underline{x}_1}{n_1} - \frac{\underline{x}_2}{n_2}. \tag{15;43}$$

Voor de verwachting en de variantie van $\underline{d}$ geldt:

$$\mathcal{E}\,\underline{d} = p_1 - p_2 \; ; \quad \sigma^2(\underline{d}) = \frac{p_1 q_1}{n_1} - \frac{p_2 q_2}{n_2}. \tag{15;44}$$

(De lezer bewijze dit.)

Zijn n_1 en n_2 niet te klein, dan zijn $\underline{x}_1$ en $\underline{x}_2$ en derhalve ook $\underline{d}$ bij benadering normaal verdeeld, zodat geldt:

$$P\left[\,|\,\underline{d} - (p_1 - p_2)\,| < u_{\frac{1}{2}\alpha}\,\sigma(\underline{d})\right] \approx 1 - \alpha. \tag{15;45}$$

Hieruit volgt een benaderd betrouwbaarheidsinterval, als men $\sigma(\underline{d})$ door een schatting van die grootheid vervangt. Daartoe vult men gewoonlijk in de formule voor $\sigma^2(\underline{d})$ (15;44) de schattingen $\underline{x}_1/n_1$ resp. $\underline{x}_2/n_2$ in voor p_1 resp. p_2, hetgeen leidt tot de schatting:

$$\underline{s}^2(\underline{d}) = \frac{\underline{x}_1(n_1 - \underline{x}_1)}{n_1^3} + \frac{\underline{x}_2(n_2 - \underline{x}_2)}{n_2^3}. \tag{15;46}$$

Soms gebruikt men ook noemers $n_1^2(n_1 - 1)$ en $n_2^2(n_2 - 1)$, om de schatting zuiver te krijgen (vgl. (12;76)), maar de betekenis van deze correctie is beperkt, omdat dan wel voor σ^2, maar niet voor σ, een zuivere schatting verkregen wordt, terwijl bovendien voor grote n de invloed van de correctie gering is.

Het (tweezijdige) betrouwbaarheidsinterval, al of niet met deze correctie, wordt dus:

(15;47) $P\left[\frac{x_1}{n_1} - \frac{x_2}{n_2} - u_{\frac{1}{2}\alpha}\,\underline{\leq}(\underline{d}) < p_1 - p_2 < \frac{x_1}{n_1} - \frac{x_2}{n_2} + u_{\frac{1}{2}\alpha}\,\underline{\leq}(\underline{d})\right] \approx 1-\alpha.$

Beide grenzen afzonderlijk bezitten als (benaderde) onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$.

Opgave 15;8

Bereken een tweezijdig betrouwbaarheidsinterval met $\alpha = 0{,}01$, voor het verschil $p_1 - p_2$ van de ziektekans van selderieplanten op grond van de gegevens van voorbeeld 14.1;1.

Oplossing: $\frac{x_1}{n_1} = 0{,}2$; $\frac{x_2}{n_2} = 0{,}46$; $s^2(\underline{d}) = \frac{20.80}{10^6} + \frac{46.54}{10^6} = 0{,}004084$;
$s(\underline{d}) = 0{,}0639$.

Het betrouwbaarheidsinterval wordt (met $u_{\frac{1}{2}\alpha} = 2{,}575$):

$$-0{,}42 < p_1 - p_2 < -0{,}10$$

Opmerking 15;9

De bij dit betrouwbaarheidsinterval behorende toets stelt ons in staat de hypothese $p_1 - p_2 = d_0$, met gegeven waarde van d_0, bij benadering te toetsen door na te gaan of d_0 in het betrouwbaarheidsinterval ligt of niet. De methode der 2x2-tabel (par.14;1) geldt alleen voor $d_0 = 0$ maar is voor dat geval exact uitvoerbaar voor kleine n_1 en n_2, terwijl voor grote n_1 en n_2 de bijbehorende benadering (zie blz. 38) beter is dan de hier beschrevene. Voor $d_0 = 0$ is dus de laatstgenoemde methode te prefereren.

Hiermede zijn een aantal der gangbare schattings-en toetsingsmethoden beschreven. Een veel groter aantal is buiten beschouwing gelaten. Het in deze cursus bestreken gebied bevat bovendien slechts een zeer klein gedeelte van de momenteel beschikbare statistische technieken. Voor lezers, die zich verder in de mathematische statistiek wensen te verdiepen volgt hieronder een korte lijst van enkele goede leer- en handboeken.

M.G. Kendall, The advanced theory of statistics, I and II,
Ch. Griffin, London, 1947

M.G. Kendall, Rank correlation methods,
Ch. Griffin 1955

A.M. Mood Introduction to the theory of statistics,
McGraw Hill Book Comp., 1950

A. Hald Statistical theory with engineering applications,
John Wiley and Sons, New York, 1952

H. Cramér, Mathematical methods of statistics,
Princeton University Press, 1946

Appendix

§ 1. Normale verdeling. Bewijs van de formule

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

Geven wij het linkerlid aan door L en substitueren wij $u = \frac{x-\mu}{\sigma}$, dan komt er, daar $dx = \sigma\,du$ is,

$$L = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2}\,du.$$

Derhalve is

$$L^2 = \frac{1}{2\pi}\iint_{-\infty}^{+\infty} e^{-\frac{1}{2}(u^2+v^2)}\,du\,dv$$

en overgang op poolcoördinaten r en φ geeft, daar $r^2 = u^2+v^2$ en

$du\,dv = r\,dr\,d\varphi$

$$L^2 = \frac{1}{2\pi}\int_0^{2\pi} d\varphi \int_0^{\infty} r\,e^{-\frac{1}{2}r^2}\,dr = \int_0^{\infty} e^{-\frac{1}{2}r^2}\,d\left(\tfrac{1}{2}r^2\right) = 1.$$

Afleiding van de momenten (formule (9;9))

Volgens (9;4) is

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

Substitueer $u = \frac{x-\mu}{\sigma}$; $x = u\sigma+\mu$, $dx = \sigma\,du$:

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

$$= \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} u\,e^{-\frac{1}{2}u^2}\,du + \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2}\,du.$$

De eerste term valt weg wegens symmetrie van het argument onder de integraal en de tweede term is volgens (A.1;1) gelijk aan μ. Dus is

$$\mu_1 = \mathcal{E}\,x = \mu. \tag{A.1;2}$$

Verder is

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

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

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

daar nu de middelste term wegvalt. Nu is, met partieel integreren,

$$\int_{-\infty}^{+\infty} u^2 e^{-\frac{1}{2}u^2} du = \left[-u e^{-\frac{1}{2}u^2}\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du = \sqrt{2\pi},$$

hetgeen na invullen geeft

(A.1;3) $$\mu_2 = \mathcal{E}\, x^2 = \sigma^2 + \mu^2.$$

§ 2. Exponentiële verdeling

Als $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onderling onafhankelijk verdeeld zijn volgens dezelfde exponentiële verdeling

$$P[\underline{x}_i \leqq x_i] = 1 - e^{-\lambda x_i} \qquad x_i \geqq 0, \tag{A.2;1}$$

dan is

$$\ell\, \frac{1}{\bar{\underline{x}}} = \frac{n}{n-1}\lambda \qquad \left(\bar{\underline{x}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i\right). \tag{A.2;2}$$

Bewijs

Als

$$\underline{y} \overset{\text{def}}{=\!=} \sum_{i=1}^{n} \underline{x}_i \tag{A.2;3}$$

dan is de verdelingsdichtheid $f(y)$ van $\underline{y}$

$$f(y) = \frac{\lambda^n}{(n-1)!}\, y^{n-1}\, e^{-\lambda y}. \tag{A.2;4}$$

Dit kan als volgt met behulp van volledige inductie bewezen worden. Uit (A.2;1) volgt voor de verdelingsdichtheid $g(x_i)$ van $\underline{x}_i$

$$g(x_i) = \lambda\, e^{-\lambda x_i}, \tag{A.2;5}$$

dus voor $n=2$ is (zie (11;5))

$$f(y) = \int_0^y g(x)\, g(y-x)\, dx = \lambda^2 \int_0^y e^{-\lambda x}\, e^{-\lambda(y-x)}\, dx = \tag{A.2;6}$$

$$= \lambda^2 \int_0^y e^{-\lambda y}\, dx = \lambda^2 y\, e^{-\lambda y}.$$

Stel nu dat (A.2;4) juist is voor $n-1$, dan wordt de verdelingsdichtheid voor de som van n grootheden $\underline{x}_i$

$$\int_0^y \frac{\lambda^{n-1}}{(n-2)!}\, x^{n-2} e^{-\lambda x} \cdot \lambda\, e^{-\lambda(y-x)}\, dx = \tag{A.2;7}$$

$$= \frac{\lambda^n}{(n-2)!}\, e^{-\lambda y} \int_0^y x^{n-2}\, dx = \frac{\lambda^n}{(n-1)!}\, y^{n-1}\, e^{-\lambda y}.$$

Uit (A.2;4) volgt nu

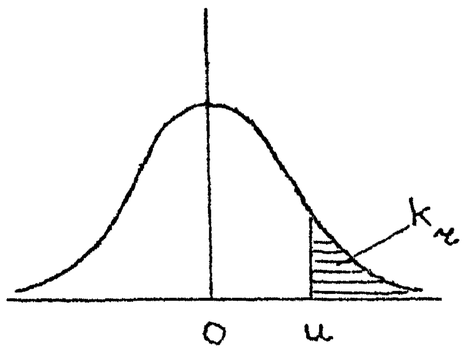
$$(A.2;8)\qquad \mathcal{E}\,\frac{1}{\underline{\underline{y}}} = \frac{\lambda^n}{(n-1)!}\int_0^\infty y^{n-2}\,e^{-\lambda y}\,dy = \frac{\lambda^n}{(n-1)!}\,\frac{(n-2)!}{\lambda^{n-1}} = \frac{\lambda}{n-1}\,,$$

daar (zie (A.2;7))

$$\frac{\lambda^n}{(n-1)!}\int_0^\infty y^{n-1}\,e^{-\lambda y}\,dy = 1.$$

Dus is

$$(A.2;9)\qquad \mathcal{E}\,\frac{1}{\underline{\underline{x}}} = n\,\mathcal{E}\,\frac{1}{\underline{\underline{y}}} = \frac{n}{n-1}\,\lambda.$$



Tabel 1

N(0,1)-verdeling

Waarden van $k_u \cdot 10^4$ voor $u = 0,00(0,01)3,49$.

$$k_u = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{1}{2}v^2} dv.$$

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

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

Tabel 2

Waarden van t_ν bij tweezijdige overschrijdingskansen k van de STUDENT-verdeling

ν \ k	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,12566	0,25335	0,38532	0,52440	0,67449	0,84162	1,03643	1,28155	1,64485	1,95996	2,32634	2,57582

Tabel 3

Linker Kritieke waarden van de tweezijdige tekentoets *)

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

*) "-" betekent dat voor de betreffende waarden van n en α voor geen enkele waarde van de toetsingsgrootheid de tweezijdige overschrijdingskans $\leq \alpha$ is.

Tabel 4

Rechter kritieke waarden van de tweezijdige symmetrietoets van WILCOXON *)

n	Onbetrouwbaarheidsdrempel (α)			
	0,01	0,02	0,05	0,10
4	-	-	-	-
5	-	-	-	15
6	-	-	21	17
7	-	28	24	22
8	36	34	30	26
9	43	39	35	29
10	49	45	39	35
11	56	52	46	40
12	64	60	52	44
13	73	67	57	49
14	81	75	63	55
15	90	82	70	60
16	98	90	78	66
17	107	99	85	71
18	117	107	91	77
19	126	116	98	84
20	136	124	106	90
21	151	135	115	97
22	161	145	123	103
23	172	154	130	110
24	182	164	140	118
25	193	175	147	125
26	205	185	157	131
27	216	196	166	138
28	228	206	174	146
29	241	217	183	155
30	253	229	193	161

---

*) "-" betekent dat voor de betreffende waarden van n en α voor geen enkele waarde van T de tweezijdige overschrijdingskans $\leqq \alpha$ is. Voor $n \leqq 20$ zijn de kritieke waarden exact berekend, daarboven met behulp van de normale benadering.

Tabel 5a

Kritieke waarden van de tweezijdige toets voor twee steekproeven van Wilcoxon voor $m \leqq n \leqq 15$ en $\alpha = 0{,}02$ *)

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	52	56	60
3	-	-	-	-	-	-	42	48	52	58	64	68	74	80	84
4	-	-	-	-	40	46	54	60	66	74	80	86	94	100	106
5	-	-	-	0	48 / 2	56	64	72	80	88	96	104	112	120	128
6	-	-	-	2	4	66 / 6	76	84	94	104	114	122	132	142	150
7	-	-	0	2	6	8	86 / 12	98	108	118	130	140	154	162	172
8	-	-	0	4	8	12	14	110 / 18	122	134	146	158	168	180	192
9	-	-	2	6	10	14	18	22	134 / 28	148	162	174	188	200	214
10	-	-	2	6	12	16	22	26	32	162 / 38	176	192	206	220	234
11	-	-	2	8	14	18	24	30	36	44	192 / 50	208	224	240	256
12	-	-	4	10	16	22	28	34	42	48	56	220 / 62	242	260	276
13	-	0	4	10	18	24	32	40	46	54	62	70	260 / 78	278	296
14	-	0	4	12	20	26	34	44	52	60	68	76	86	298 / 94	318
15	-	0	6	14	22	30	38	48	56	66	74	84	94	102	338 / 112

*) "-" betekent dat voor de betreffende waarden van m en n voor geen enkele waarde van W de tweezijdige overschrijdingskans $\leqq \alpha$ is.

Tabel 5b

Kritieke waarden van de tweezijdige toets voor twee steekproeven van Wilcoxon voor $m \leq n \leq 15$ en $\alpha = 0{,}05$ *)

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	32	36	40	44	46	50	54	58
3	-	-	-	-	30	34	40	44	50	54	60	64	70	74	80
4	-	-	-	0 / 32	38	44	50	56	64	70	76	82	88	94	100
5	-	-	0	2	4 / 46	54	60	68	76	84	92	98	106	114	122
6	-	-	2	4	6	10 / 62	72	80	88	98	106	116	124	134	142
7	-	-	2	6	10	12	16 / 82	92	102	112	122	132	142	152	162
8	-	0	4	8	12	16	20	26 / 102	114	126	138	148	160	172	182
9	-	0	4	8	14	20	24	30	34 / 128	140	152	164	178	190	202
10	-	0	6	10	16	22	28	34	40	46 / 154	168	182	194	208	222
11	-	0	6	12	18	26	32	38	46	52	60 / 182	198	212	228	242
12	-	2	8	14	22	28	36	44	52	58	66	74 / 214	230	246	262
13	-	2	8	16	24	32	40	48	56	66	74	82	90 / 248	264	282
14	-	2	10	18	26	34	44	52	62	72	80	90	100	110 / 282	302
15	-	2	10	20	28	38	48	58	68	78	88	98	108	118	128 / 322

*) Zie voetnoot bij tabel 5a.

Tabel 5c

Kritieke waarden van de tweezijdige toets voor twee steekproeven van Wilcoxon voor $m \leq n \leq 15$ en $\alpha = 0{,}10$ *)

n \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	20	24	28	30	34	38	42	44	48	50	54
3	-	-	0	24	28	32	38	42	46	52	56	62	66	70	76
4	-	-	0	2	36	42	48	54	60	66	72	78	84	90	96
5	-	0	2	4	8	50	58	64	72	78	86	94	100	108	114
6	-	0	4	6	10	14	68	76	84	92	100	110	118	126	134
7	-	0	4	8	12	16	22	86	96	106	116	126	134	144	154
8	-	2	6	10	16	20	26	30	108	120	130	140	152	162	174
9	-	2	8	12	18	24	30	36	42	132	144	156	168	180	192
10	-	2	8	14	22	28	34	40	48	54	158	172	186	198	212
11	-	2	10	16	24	32	38	46	54	62	68	188	202	216	230
12	-	4	10	18	26	34	42	52	60	68	76	84	218	234	250
13	-	4	12	20	30	38	48	56	66	74	84	94	102	252	268
14	-	6	14	22	32	42	52	62	72	82	92	102	112	122	288
15	-	6	14	24	36	46	56	66	78	88	100	110	122	132	144

*) Zie voetnoot bij tabel 5a.

Inhoud

			blz.
Hoofdstuk	1	Inleiding	1
	2	Het experimentele fundament van de statistiek	5
	3	Eigenschappen van frequentiequotiënten	9
	4	Het kansbegrip	15
	5	Voorbeelden van het berekenen van kansen	26
	6	Het meten van kansen	34
	7	Stochastische grootheden, kansverdelingen	36
	8	Onafhankelijkheid van stochastische grootheden	47
	9	Mathematische verwachting en momenten	54
	10	De ongelijkheid van Bienaymé-Tchebycheff en de theoretische wet van de grote aantallen	68
	11	Functies van stochastische grootheden; centrale limietstelling	73
	12	Schattingstheorie; meest aannemelijke schattingen	87
	13	Toetsingstheorie; binomiale toetsen	116
	14	Enige gangbare toetsen	
		14.1 De 2x2-tabel (dubbele dichotomie)	136
		14.2 De toetsen van Student	144
		14.3 Verdelingsvrije toetsen voor dezelfde of analoge hypothesen	154
		14.4 Onderscheidingsvermogen van de toetsen van Student, de tekentoets en de toetsen van Wilcoxon en asymptotische eigenschappen	172
	15	Betrouwbaarheidsgrenzen	176

Opgaven

	blz.		blz.		blz.
11;1	74	13;6	133	14.3;6	169
11;2	74	13;7	134	14.3;7	169
11;3	76	13;8	134	14.3;8	169
11;4	77	14.1;1	137	14.3;9	170
11;5	84	14.1;2	137	14.4;1	174
11;6	84	14.1;3	141	15;1	184
11;7	85	14.2;1	147	15;2	188
11;8	85	14.2;2	150	15;3	189
11;9	85	14.2;3	153	15;4	189
13;1	132	14.3;1	155	15;5	190
13;2	132	14.3;2	156	15;6	193
13;3	133	14.3;3	162	15;7	193
13;4	133	14.3;4	163	15;8	195
13;5	133	14.3;5	167		

Stellingen

3;1	10	9;4	60	11;5	76
3;2	11	9;5	62	11;6	78
3;3	11	9;6	63	11;7	79
3;4	11	9;7	63	11;8	79
3;5	12	9;8	65	11;9	82
3;6	12	9;9	66	14.2;1	144
3;7	12	9;10	66	14.2;2	145
3;8	13	10;1	68	14.3;1	158
4;1	21	10;2	69	14.3;2	163
8;1	47	10;3	70	15;1	177
8;2	48	11;1	73	15;2	180
8;3	49	11;2	73	15;3 }	184
9;1	58	11;3	74	15;4 } 1)	186
9;2	58	11;4	75	15;5 }	186
9;3	59				

Voorbeelden

5;1	26	12;2	92	14.2;3	150
5;2	27	12;3	93, 98, 113	14.3;1	154, 161
5;3	28	12;4	94, 98	14.3;2 2)	156
5;4	28	12;5	99	14.3;3	168
5;5	29	12;6	102, 114	15;1	178
5;6	30	12;7	110	15;2	187
5;7	31	13;1	117 ev, 126	15;3	188
5;8	32		129, 130, 132	15;4	189
5;9	33	14.1;1	138	15;5	190
11;1	82	14.1;2	142	15;6	192
11;2	83	14.2;1	146	15;7	193
12;1	87, 91 96, 108	14.2;2	148	15;8	194

1) In de tekst abusievelijk aangegeven als 15;2, 15;3 en 15;4

2) In de tekst abusievelijk aangegeven als 14.3;2

<u>Errata</u> bij

Rapport S 200 (C8)

Syllabus van een oriënterende cursus

mathematische statistiek

pag.	regel (resp. formule-nr)	staat	moet staan
7	21 v.o.	0,50	0,5
26	13 v.b.	(3;11)	(3;10)
	17 v.b.	(3;9)	(3;8)
	12 v.o.-1 v.o. vervalt		
27	18 v.b. } 22 v.b. }	stelling 3;2	stelling 3;6
28	9 v.b.	$\frac{1}{n-k}$	$\frac{1}{n-k+1}$
	12 v.b.	stelling 3;3	stelling 3;2
	17 v.o.	stelling 3;5	stelling 3;4
30	11 v.o.	(4;16)	(4;14)
	8 v.o.	W W M W M M ... M	W W Z W Z Z ... Z
31	10 v.b.	katagorisch	kategorisch
	11 v.b.	stelling 3;4	stelling 3;3
32	(5;10)	$(\frac{N}{N+M})^x (\frac{M}{N+M})^{n-x}$	$(\frac{M}{N+M})^x (\frac{N}{N+M})^{n-x}$
	12 v.b.	$x << N,\ n-x << M$	$x << M,\ n-x << N$

33 Opmerking bovenaan de pagina vervalt en wordt vervangen door:

Verder geldt voor $k \geqq 1$

(5;12) $$\sum_{n=k}^{\infty} Q_n = \sum_{n=k}^{\infty} \binom{n-1}{k-1} p^k q^{n-k} = \begin{cases} 1 \text{ als } p \neq 0 \\ 0 \text{ als } p = 0 \end{cases}$$

<u>Bewijs</u>

Iedere Q_n bevat een factor p^k met $k \geqq 1$, dus

$$\sum_{n=k}^{\infty} Q_n = 0 \qquad \text{als } p=0.$$

Stel nu

$$S_k = \sum_{n=k}^{\infty} Q_n = \sum_{n=k}^{\infty} \binom{n-1}{k-1} p^k q^{n-k}$$

dan is

$$qS_k = \sum_{n=k}^{\infty} \binom{n-1}{k-1} p^k q^{n-k+1} = \sum_{n=k+1}^{\infty} \binom{n-2}{k-1} p^k q^{n-k}$$

dus

$$pS_k = p^k + \sum_{n=k+1}^{\infty} \left[\binom{n-1}{k-1} - \binom{n-2}{k-1} \right] p^k q^{n-k} =$$

$$= p^k + \sum_{n=k+1}^{\infty} \binom{n-2}{k-2} p^k q^{n-k} = \sum_{n=k}^{\infty} \binom{n-2}{k-2} p^k q^{n-k} =$$

$$= pS_{k-1}.$$

Voor $p \neq 0$ geldt dus $S_k = S_{k-1} (k=1,2,\ldots)$ en dus is

$$S_k = S_1 = \sum_{n=1}^{\infty} pq^{n-1} = p \cdot \frac{1}{1-q} = 1.$$

pag.	regel (resp. formule-nr)	staat	moet staan
34	16 v.b.	uitgevoerd	ingevoerd
36	2 v.o.	στοχαζειν	στοχαξεσθαι
37	4 v.o.	F(x) voor iedere x	F(x) voor iedere x continue is en voor iedere x
44	(7;35)	$f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} \ldots$	$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \ldots$
45	5 v.o.	$\int_{-\infty}^{+\infty} f(x_1 \mid x_2) = 1$	$\int_{-\infty}^{+\infty} f(x_1 \mid x_2) dx_1 = 1$
47	4 v.b.	(4;15)	(4;13)
49	19 v.o.	(4;16)	(4;14)
	18 v.o. / 17 v.o.	(4;15)	(4;13)
66	(9;37)	$\mathrm{cov}(\tilde{x},\tilde{y})$	$\mathrm{cov}(\underline{x},\underline{y})$
68	1 v.o.	(10;2)	(10;1)
76	9 v.b.	gereduceerde	gereduceerden
	17 v.b.	$\sigma(\underline{y})$	$\sigma(\underline{z})$
77	15 v.b.	300	350
89	13 v.o.	$-2(\underline{x}-\mu)$	$-2(\bar{\underline{x}}-\mu)$
	9 v.o.	$-n(\underline{x}-\mu)$	$-n(\bar{\underline{x}}-\mu)$
94	17 v.b.	ziet	zien

pag.	regel (resp. formule-nr)	staat	moet staan
96	16 v.b.	aannemlijkheid	aannemelijkheid
	6 v.o.	$\frac{1}{\sigma\sqrt{2\pi}}$	$(\frac{1}{\sigma\sqrt{2\pi}})^n$
98	1 v.b.	12;2	12;3
103	7 v.b.	uirgezet	uitgezet
104	7 v.o.	twee laatste	laatste twee
	1 v.o.	Y	Y_i
109	3 v.b.	$\lvert \underline{x}^* \rvert$	$\lvert \underline{\bar{x}}^* \rvert$
118	(13;4)	$P[\underline{w}_i = w_i \mid i=1,\ldots,n;p]$	$P[\underline{w}_1 = w_1 \wedge \ldots \wedge \underline{w}_n = w_n \mid p]$
123	17 v.o.	0,410	0,510
126	8 v.o.	0,001	0,008
	7 v.o.	0,032	0,052
	1 v.o.	0,966 (2x)	0,968
129	(13;33)	t_l, t_r	t
	(13;34)	$k_r(t) \geqq \alpha_r$	$k_r(t) \leqq \alpha_r$
130	6 v.b.	$k_1(3) = 0,0001$	$k_1(3) = 0,0005$
135	2 v.b. } 5 v.b. }	α_r	α_1
136	2 v.b.	14.1;1	14.1
140	13 v.b.	p_1/p_2	$p_1 \neq p_2$
143	3 v.b.	gelijk	verschillend zijn en niet asymptotisch onderscheidend als de medianen gelijk
147	4 v.b.	$\frac{994-1000}{\sqrt{64,89}}$	$\frac{994-1000}{\sqrt{64,89}}\sqrt{10}$
148	8 v.o.	oliemomsters	oliemonsters
175	15 v.b.	$\neq$	$\neq 0$
179	7 v.o.	stochastische	stochastisch
180	9 v.o.	De gestippelde ver-	De ver-
183	fig. 15;2	x_l, x_r	$\frac{x_l}{n}, \frac{x_r}{n}$

pag.	regel (resp. formule-nr)	staat	moet staan
188	5 v.o.	Betrouwbaarheids-grenzen voor	Betrouwbaar-heidsgrenzen voor n
189	11 v.b.	eem	een

REGISTER

Aannemelijkheidsfunctie 96
Afhankelijk 24, 25, 26, 31, 32, 33, 41, 47, 48, 49, 50, 51, 52, 58, 59, 60, 61, 65
Aselect 27, 28, 29, 30, 31, 33, 40, 42
Asymptotisch:
  doeltreffend 113, 174
  onderscheidend 128, 143, 150, 175
  raak 88, 90, 94, 112
  verdwijnend 186

BAYES 12
BENARD, A. 158, 160
Betrouwbaarheids:
  coëfficiënt 109, 177
  gebied 178
  grenzen (interval) 109, 176, 178 e.v., 191 (afleiding van), 114 (benaderd), 194 (benaderd - voor het verschil van twee onbekende kansen), 184 (eigenschappen van), 187 (voor de mediaan), 188 (voor n), 189 (voor een symmetriepunt), 193 (voor het verschil van de verwachtingen van twee normale verdelingen met gelijke spreiding), 190 (voor de verschuiving van één verdeling t.o.v. een andere), 192 (voor de verwachting van een normale verdeling met onbekende spreiding)
BIENAYMÉ - TCHEBYCHEFF (ongelijkheid van) 68
BRUIJN, N.G. de 33
BUNT, L.N.H. 182

CAMERON, J.M. 148
Continuïteitscorrectie 130, 161
Correlatiecoëfficiënt 67, 76
Covariantie 55 e.v., 76
CRAMÉR, H. 4, 195

DANTZIG, D. van 20, 27, 167
Dichotomie 13, 14, 55
  dubbele - 136, 170
DIXON, W.J. 4, 182

EEDEN, C. van 158, 160, 163, 170
ETTINGER, J van 92
Eventualiteit 20 e.v.

FERMAT 5
FISHER, R.A. 95
Fout 122 (van de eerste soort), 125 (van de tweede soort)
Frequentie (quotiënt) 5 e.v., 9 e.v , 17 e.v., 25
  verdeling 99
  voorwaardelijk - 9 e.v., 21

Gemiddelde 69 e.v., 88 (schatting voor)
  voorlopig - 91, 146
Gestandaardiseerde grootheid 63, 68
GOSSET, W.S. zie STUDENT

HALD, A. 195
HEMELRIJK, J. 52
HUYGENS 15
Hypothesen 118 (toegelaten), 119 (te toetsen), 119 (alternatieve)

Kans:
  rekening 2, 5, 6, 9, 15 e.v., 25 e.v., 34, 35
  verdeling 36 e.v., 120
  voorwaardelijke - 21 e.v , 26, 30, 42, 44, 45, 47
KENDALL, M.G. 4, 167, 195
Kleinste kwadraten (theorie der) 102 e.v.
KOLMOGOROV, A. 25

Kritieke:
waarde 134
zone 121
KRUSKAL, W.H. 167

LAPLACE 15 e.v., 25, 27, 29, 31, 51, 79, 80
Limietstelling (centrale) 73, 78, 79, 85

MANN, H.B. 167
MASSEY Jr, F.J. 4, 182
Mediaan 111
MENDEL 117 e.v.
MÉRÉ, de 5, 16
MISES, von 16
Model 1, 2, 8, 16, 19, 24, 27, 35, 36, 42, 50, 51
MOIVRE, de 79, 80
Moment 54 e.v., 59, 61
MOOD, A.M. 195

NEYMAN, J. 4, 183
NOETHER, G.E. 174
Normale benadering 181 (van de betrouwbaarheidsgrenzen), 130 (van de binomiale toets)

Onafhankelijk 22 e.v.
Onbetrouwbaarheid 109, 176, 179, 123 (drempel), 124 (linker- en rechter-), 122 (van een toets)
Onderscheidingsvermogen 125, 126 (linker- en rechter-), 172 (voor STUDENT-toetsen), 174 (voor WILCOXON-toetsen)
Overschrijdingskans 131, 137, 129 (linker- en rechter-), 129 (tweezijdige)

Parameterruimte 118
PASCAL 5
PEARSON, E.S. 144
Populatie 93

QUÉTELET 5

Regressie:
coëfficiënt 114
lineaire - 102 e.v., 114
lijn 105
Residu 106
ROMIG, H.G. 123, 126
RIJKOORT, T.J. 168

Schatting:
meest aannemelijke - 95 e.v., 101, 103
meest doeltreffende - 108
nauwkeurigheid van een - 108
-stheorie 87, 191
SMID, L.J. 167
Spreiding zie: Variatie
-sbreedte 151
Stochastisch (zie ook: Afhankelijk) 8, 22, 23, 30, 36, 37, 40, 47, 49 e.v., 54, 58 e.v., 62, 63, 65, 67, 73
convergentie 59
STOKER, D.J. 167
STUDENT 144 e.v., 150 e.v., 155 e.v., 161 e.v., 172 e.v., 192
SUSSKIND 5
Systeem:
exclusief - 11, 21
kategorisch - 11 e.v., 31

TERRY, M.E. 175
Toets:
binomiale - 116 e.v., 126
mediaan- 141
onafhankelijkheids- 140
van STUDENT voor één steekproef 144 e.v., 146 e.v., 152 e.v., 155 e.v., 161 e.v.

van STUDENT voor twee steekproeven 146, 150 e.v.
teken- 154 e.v.
verdelingsvrije - 142, 154, 174 e.v.
WILCOXON, symmetrie- 157 e.v., 189
WILCOXON, twee steekproeven- 163 e.v., 168, 170, 19●
Toetsings:
grootheid 120, 136
theorie 116

Variantie 56, 59 e.v., 108 (van het gemiddelde van een steekproef), 88 (schatting voor), 158 e.v. (van symmetrietoets van WILC●XON)
Verdeling:
asymptotisch normale - 138, 158
binomiale - 37, 41, 61, 94, 98, 113, 121, 178, 18●
chi-kwadraat - 143
continue - 37 e.v., 42 e.v., 4●, 54
n-dimensionale - 42 e.v., 47 e.v., 57 e.v.
discrete - 37, 40, 43, 48, 52, 54
exacte - 160, 167
exponentiële - 99
homogene - 39, 41, 43, 56, 59
hypergeometrische - 37, 62, 93, 137
marginale - 43 e.v., 49, 50, 57, 58
negatief binomiale - 37, 52, 55, 61, 94, 98
normale - 38, 39, 42, 44, 46, 50, 51, 57, 58, 62, 63, 73, 76, 78 e.v., 82, 83, 85, 87, 96
simultane - 40, 42, 44, 49 e.v., 57, 59, 63, 66, 73, 96
van STUDENT 144 e.v., 192
voorwaardelijke - 44 e.v., 50, 57, 137, 160
Verdelingsdichtheid, -functie zie: kansverdeling
Verwachting 54, 57 e.v.
marginale - 57 e.v.
voorwaardelijke - 57, 64
Voorspellingsinterval 191

Waarschijnlijkheidsrekening zie: kansrekening
WABEKE, D. 52, 163
WAERDEN, B.L. van der 175
WEMELSFELDER, P.J. 99
Wet der grote aantallen 5, 70 e.v.
WHITNEY, D.R. 167
WILCOXON 157 e.v., 163 e.v., 167, 168, 170, 189
WISHART, J. 144

YOUDEN, W.J. 148

Zuiverheid 184 (van een betrouwbaarheidsgebied), 128 (van een toets)
van schattingen 88 e.v., 94, 97, 105, 112