MC SYLLABUS 1.2

J. HEMELRIJK
J. KRIENS

LEERGANG BESLISKUNDE

DEEL 2 KANSREKENING

MATHEMATISCH CENTRUM AMSTERDAM 1976

AMS(MOS) subject classification scheme (1970): 60-01

ISBN 90 6196 015 0

1e druk 1965 3e druk 1972
2e druk 1969 4e druk 1976

Inhoud

Voorwoord

Het zou misschien prettig zijn als iedereen die beslissingen moet nemen, beschikte over volmaakt betrouwbare gegevens en een volmaakt inzicht in alle gevolgen van de diverse mogelijke besluiten. Omdat de werkelijkheid wel heel anders is, spelen onvolledige waarnemingen, onzekere faktoren en onbekende gevolgen bij onze beslissingen een grote rol, zelfs als we het beslissen tot een kunde verheffen.
De statistiek en de waarschijnlijkheidsrekening kunnen dit bezwaar enigszins opheffen, doordat zij in bepaalde omstandigheden toestaan de onzekerheden te kwantificeren en ondanks de onzekerheden bepaalde uitspraken te doen.

Vooral als voorbereiding van de delen 3 (Statistiek) en 4 (Markovketens en Wachttijden), maar ook van de overige delen van de Leergang Besliskunde, worden in dit tweede deel de grondbeginselen van het rekenen met kansen besproken. Na een inleiding over statistische verschijnselen wordt het begrip "kans" axiomatisch ingevoerd als een soort idealisering van het frequentiequotiënt. Het berekenen en meten van kansen wordt besproken, waarna discrete en continue kansverdelingen in één en meer dimensies worden behandeld. Nadat de eigenschappen van de verwachting en van de overige momenten van stochastische variabelen zijn genoemd, besluit dit deel met een korte bespreking van enige belangrijke limietstellingen. Talrijke voorbeelden kunnen het inzicht van de lezer verdiepen en zijn vaardigheid in het werken met kansrekening vergroten.

Dit deel is een bewerking door Prof. J. KRIENS van de eerste helft van Rapport S 200 (C 8): Syllabus van een oriënterende cursus Mathematische Statistiek, door Prof.Dr. J. HEMELRIJK. Enige nadere correcties werden aangebracht door Dr. W.R. VAN ZWET en Drs. W. MOLENAAR.

Amsterdam, 1965.

1. Inleiding.

Gaandeweg is de wiskunde als hulpwetenschap doorgedrongen in velerlei gebieden van wetenschap en praktijk; gedeeltelijk is het ontstaan ervan zelfs direct aan bepaalde praktische problemen verbonden. Meetkunde, algebra, analyse, enz. worden als routinemiddelen gehanteerd in de natuurkunde, sterrekunde, scheepvaartkunde, en waar al niet meer. In de regel - en tot voor kort vrijwel steeds - wordt de wiskunde gebruikt om problemen van deterministische aard op te lossen. De elementaire natuurkunde bijv. 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 bijv. 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 praktijk 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", enz. 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. de levensduur van een machine-onderdeel;
5. de vraag naar een artikel in de volgende maand;
6. het jaarlijkse nationale inkomen;
7. de werkzaamheid van een geneesmiddel;
8. 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 wanneer men voorspelt, dat het weer morgen "hetzelfde" zal zijn als vandaag. Indien men beschikt over de ongevallen-statistieken 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 kan worden beantwoord. Zo kan men bijv. 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 waar de statistiek een belangrijk hulpmiddel is gebleken: biologie, medicijnen, landbouw, industrie, geodesie, astronomie, verzekeringswezen, economie, sociologie, bevolkingsleer, strategie en kernfysica.

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 essentieel element is, nl. van dat deel van de zuivere wiskunde dat de kansrekening of waarschijnlijkheidsrekening genoemd wordt en dat een onderdeel van de maattheorie is. Het essentiële praktische 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, bijv. 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 de volgende paragrafen en in deel 3. 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 (meestal) 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 vraagstukken uit de industrie, zoals het onderzoek naar de factoren die een produktieproces of de verkoop van een artikel beïnvloeden.

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

Tot slot van deze inleiding merken wij nog 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 bijv. vormt dit beschrijvende statistische werk het grootste deel van de statistische werkzaamheden. De techniek van het verzamelen van gegevens door middel van enquêtes neemt daarbij de plaats in van de techniek van het opzetten van experimenten op gebieden als landbouw, biologie en industrie. Voor beide technieken zijn aparte onderdelen van de statistische theorie ontwikkeld.

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 procentsgewijs 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 de eigenschappen van allerlei dobbelspelen berekenen; bijv.: zal het bij een groot aantal pogingen 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, enz. Men noemt het beschreven verschijnsel de experimentele wet der grote aantallen. Een iets nauwkeuriger 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*) 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 fq(S) gevonden worden. De experimentele wet van de grote aantallen zegt nu, dat voor een grote klasse van reeksen experimenten geldt dat deze verschillen tussen de voor fq(S) gevonden waarden voor grote waarden van N zeer klein worden.

De term "experiment" vatte men hierbij zeer ruim op. Bij de

---

*) Afkorting: fq; meervoud: fqn; notatie: fq(S).

voorgaande voorbeelden valt daar niet alleen een worp met een dobbelsteen onder, maar ook het verrichten van een waarneming, bijv. 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 praktische problemen en deze theorie.

Een praktisch 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.

De fqn 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 wijken bij elkaar te nemen, dan worden de verschillen geringer. Dit is in tabel II gedaan.

wijk nr.	jongens	meisjes	totaal (N)	fq (jongens)
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

Tabel I

Aantallen jongens- en meisjesgeboorten
in 6 wijken van Amsterdam in 1954

wijken	jongens	meisjes	totaal (N)	fq (jongens)
1 en 2	323	304	627	0,52
3 en 4	302	299	601	0,50
5 en 6	349	350	699	0,50
1,2,3	600	576	1176	0,51
4,5,6	374	377	751	0,50
1 t/m 6	974	953	1927	0,51

Tabel II

De gegevens van Tabel I bij samenvoegen van wijken

Hierbij valt nog op te merken, dat de fqn, 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 fq, verkregen uit alle gegevens tezamen, eigenlijk niet op 0,5 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 systematische 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 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.

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. De verzameling Λ kan bijv. bestaan uit 500 op een enquête binnengekomen formulieren, waarop het kenmerk A kan zijn het antwoord dat men een artikel regelmatig gebruikt, het kenmerk B dat men het artikel niet gebruikt, het kenmerk C dat men niet geheel tevreden is over de verpakking, enz. In een ander voorbeeld bestaat Λ uit 928 Nederlandse mannen van wie gegevens verzameld zijn, zoals de kleur van de ogen (bruin is kenmerk

A, blauw is kenmerk B, enz.), de kleur van het haar (rood is kenmerk K, blond is kenmerk L), enz.

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. Zo stelt in het tweede voorbeeld $\bar{A}$ voor: geen bruine ogen, $B \wedge L$ betekent zowel blauwe ogen als blond haar en $A \vee L$ bruine ogen of blond haar (of beide).

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

$$\text{(3.1)} \qquad fq(S) \overset{\text{def}}{=} \frac{N(S)}{N}$$

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

$$\text{(3.2)} \qquad fq(S \mid T) \overset{\text{def}}{=} \frac{N(S \wedge T)}{N(T)} \quad .$$

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

De volgende eigenschappen van fqn zijn nu evident:

(3.3)

I	$0 \leq fq(S) \leq 1$	voor ieder der beschouwde kenmerken.
II	$fq(S) = 0$	dan en slechts dan als geen $\lambda \in \Lambda$ het kenmerk S bezit.
III	$fq(S) = 1$	dan en slechts dan als iedere $\lambda \in \Lambda$ het kenmerk S bezit.
IV	$fq(S \vee U) = fq(S) + fq(U) - fq(S \wedge U)$	voor ieder paar S,U van beschouwde kenmerken.

Eigenschap IV volgt direct uit $N(S \vee U) = N(S) + N(U) - N(S \wedge U)$, na deling door N.

Een meetkundige illustratie van deze eigenschap vindt men in figuur 3.1; de verzameling Λ bestaat uit een rechthoek, waarvan ieder punt een element λ voorstelt. Er zijn twee kenmerken S en U en een deel van de elementen bezit beide kenmerken.

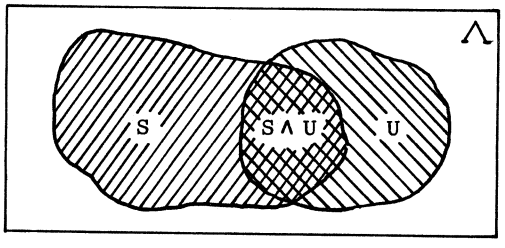


fig. 3.1

Volgens definitie (3.2) en het daaronder vermelde gelden de eigenschappen (3.3) ook voor voorwaardelijke fqn. Indien $fq(T) \neq 0$ gaat bijv. eigenschap IV over in

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

Verder volgt uit (3.1) en (3.2)

(3.4) $$fq(S \mid T) = \frac{fq(S \wedge T)}{fq(T)}, \quad \text{of} \quad fq(S \wedge T) = fq(S \mid T)\, fq(T).$$

Nu kunnen op grond van (3.3) en (3.4) alleen (en onder gebruikmaking van stellingen uit de logica met betrekking tot conjunctie en disjunctie), dus zonder terug te grijpen op de definities (3.1) en (3.2), 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') $$\begin{cases} \text{II} & fq(S) = 0 & \text{als geen } \lambda \in \Lambda \text{ het kenmerk S bezit,} \\ \text{III} & fq(S) = 1 & \text{als iedere } \lambda \in \Lambda \text{ het kenmerk S bezit,} \end{cases}$$

in plaats van (3.3;II) en (3.3;III).

De te bewijzen stellingen gelden, zoals gezegd, ook voor voorwaardelijke frequentiequotiënten; 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.

De eerste stelling die wij noemen is een generalisatie van (3.3;IV). Voor drie kenmerken S_1, S_2 en S_3 kan uit figuur 3.2 gemakkelijk afgeleid worden, dat geldt

$$fq(S_1 \vee S_2 \vee S_3) = fq(S_1) + fq(S_2) + fq(S_3) - fq(S_1 \wedge S_2) +$$
$$- fq(S_1 \wedge S_3) - fq(S_2 \wedge S_3) + fq(S_1 \wedge S_2 \wedge S_3) .$$

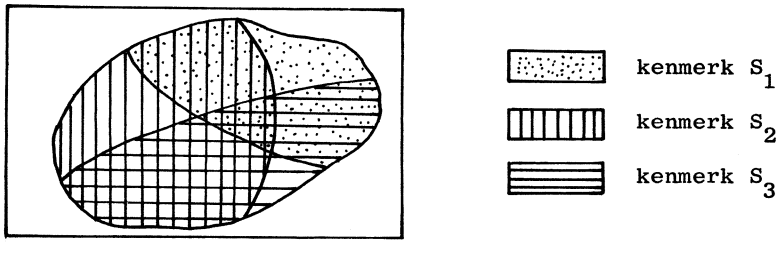


fig. 3.2

Een dergelijke stelling geldt niet alleen voor twee en voor drie kenmerken, doch algemeen voor een willekeurig aantal van k kenmerken.

Stelling 3.1 (algemene optelregel)

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

$$\text{(3.5)} \qquad \bigvee_{i=1}^{k} S_i \overset{\text{def}}{=} S_1 \vee S_2 \vee \ldots \vee S_k$$

en

$$\text{(3.6)} \qquad \bigwedge_{i=1}^{k} S_i \overset{\text{def}}{=} S_1 \wedge S_2 \wedge \ldots \wedge S_k$$

dan is

$$\text{(3.7)} \qquad fq(\bigvee_{i=1}^{k} S_i) = \sum_{1 \leq i \leq k} fq(S_i) - \sum_{1 \leq i < j \leq k} fq(S_i \wedge S_j) +$$
$$+ \sum_{1 \leq i < j < h \leq k} fq(S_i \wedge S_j \wedge S_h) - \ldots + (-1)^{k-1} fq(\bigwedge_{i=1}^{k} S_i).$$

Bewijs:

D.m.v. 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

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

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

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

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

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 $h \geq 2$, 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 $h \geq 2$ geldt

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

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) $\quad fq(S \wedge U) \leq fq(S)$

en

(3.14) $\quad fq(S \vee U) \geq fq(S)$.

Bewijs:

Uit $\quad S \equiv (S \wedge U) \vee (S \wedge \overline{U})$

en (3.8) volgt

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

daar $S \wedge U$ en $S \wedge \overline{U}$ op Λ een exclusief systeem vormen. Verder is volgens (3.3;I)

$$fq(S \wedge \overline{U}) \geq 0 ,$$

dus

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

Verder is volgens (3.3;IV)

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

Daar volgens (3.13) geldt $fq(S \wedge U) \leq fq(U)$ is dus

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

De tot dusverre afgeleide stellingen kunnen alle gezien worden als generalisaties van (3.3;IV), respectievelijk specialisaties van (3.7).

Wij geven nu nog drie stellingen, waarin voorwaardelijke fqn een belangrijke rol vervullen. De eerste hiervan is een generalisatie van (3.4) en luidt als volgt:

Stelling 3.6 (algemene vermenigvuldigingsregel)

Als

$$fq(\bigwedge_{i=1}^{k-1} S_i) > 0 \quad , \tag{3.15}$$

dan geldt

$$fq(\bigwedge_{i=1}^{k} S_i) = fq(S_1).fq(S_2 \mid S_1).fq(S_3 \mid S_1 \wedge S_2)\ldots fq(S_k \mid \bigwedge_{i=1}^{k-1} S_i). \tag{3.16}$$

Bewijs:

Uit (3.13) volgt

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

Wegens (3.15) zijn dus alle voorwaardelijke fqn in (3.16) gedefinieerd. De stelling wordt nu bewezen met volledige inductie.

Voor k = 2 gaat de stelling over in (3.4)

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

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

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

$$= fq(S_1).fq(S_2 \mid S_1) \ldots fq(S_{k-1} \mid \bigwedge_{i=1}^{k-2} S_i).fq(S_k \mid \bigwedge_{i=1}^{k-1} S_i) \quad .$$

Stelling 3.7

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

$$fq(S) = \sum_{i=1}^{k} fq(S \mid T_i).fq(T_i) \quad . \tag{3.17}$$

Bewijs:

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

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

geldt volgens (3.8)

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

De stelling volgt dan uit

$$fq(S \wedge T_i) = fq(S \mid T_i).fq(T_i) \qquad \text{als } fq(T_i) > 0 .$$

Stelling 3.8 (stelling van Bayes)

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

$$(3.18) \qquad fq(T_i \mid S) = \frac{fq(S \mid T_i).fq(T_i)}{\sum_{j=1}^{k} fq(S \mid T_j).fq(T_j)} \qquad \text{voor } i=1,\ldots,k .$$

Bewijs:

Volgens (3.4) is

$$fq(S \wedge T_i) = fq(S).fq(T_i \mid S) = fq(T_i).fq(S \mid T_i) \qquad \text{voor } i=1,\ldots,k ,$$

dus

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

De stelling volgt dan uit (3.17).

Opmerking

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. de kenmerken $S \wedge T_i$ voor $i=1,2,\ldots,k$ een exclusief systeem op Λ

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

Ter illustratie van de in deze paragraaf behandelde begrippen en stellingen geven wij nog een voorbeeld.

Gegevens betreffende 928 Nederlandse mannen:

kleur ogen \ kleur haar	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:

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

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

Voorwaardelijke frequenties:

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

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

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

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

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

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

Toepassing algemene optelregel (st. 3.1)

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

Algemene vermenigvuldigingsregel (st. 3.6)

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

Bijzondere optelregel (st. 3.2)

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

$$fq(\text{r.h. of bl.h.}|\text{bl.o.}) = fq(\text{r.h.}|\text{bl.o.}) + fq(\text{bl.h.}|\text{bl.o.}) = \frac{9}{508} + \frac{184}{508} = 0{,}380.$$

St. 3.7

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

$$fq(\text{bl.o.}) = fq(\text{bl.o.}|\text{r.h.}).fq(\text{r.h.}) + \ldots + fq(\text{bl.o.}|\text{zw.h.}).fq(\text{zw.h.}) = 0{,}547.$$

St. 3.8

$T_1,\ldots,T_k$ haarkleur, S blauwe ogen:

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

Een tweede toepassing van de bovenvermelde stellingen wordt in het begin van de volgende paragraaf gegeven.

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 en gelijkelijk mogelijke) mogelijkheden, waarbij S gerealiseerd wordt en het totale aantal mogelijkheden.

Zo is bijv. 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 fysische eisen van symmetrie zijn niet voldoende. Fabriceert men bijv. een zo zuiver mogelijk symmetrische munt van zeer homogeen materiaal, doch legt men deze telkens met "kruis" boven op tafel, 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 met 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 de vorige paragraaf gegeven zijn, duidelijk. Een verzameling Λ van n elementen λ_i (i=1,2,...,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 gedefinieerde 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 paragraaf 2 gestelde vraag trachten te beantwoorden met behulp van een stelling uit paragraaf 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?" *)
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.

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 de eventualiteit "i^e worp levert 6 op" kortweg aan met 6_i. Dan is

$$S_1 \equiv 6_1 \vee 6_2 \vee 6_3 \vee 6_4$$

en

$$\overline{S}_1 \equiv \overline{6}_1 \wedge \overline{6}_2 \wedge \overline{6}_3 \wedge \overline{6}_4 \ .$$

---

*) 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 praktijk won hij echter met het eerste, maar verloor met het tweede spel. Hij concludeerde hieruit, dat de wiskunde niet deugde.

Volgens (3.10) is

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

zodat het voldoende is $fq(\overline{S}_1)$ te berekenen. Volgens (3.16) is

$$fq(\overline{S}_1) = fq(\overline{6}_1).fq(\overline{6}_2 \mid \overline{6}_1).fq(\overline{6}_3 \mid \overline{6}_1 \wedge \overline{6}_2).fq(\overline{6}_4 \mid \overline{6}_1 \wedge \overline{6}_2 \wedge \overline{6}_3) .$$

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

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

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 procentsgewijs (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 in meer dan 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. Wij beperken ons tot één van deze wegen, namelijk de axiomatische opzet van de theorie, afkomstig van KOLMOGOROF. Deze is tegenwoordig de meest gebruikelijke, mede vanwege de eenvoudige wijze, waarop langs deze weg exactheid kan worden bereikt.

Bij de axiomatische opzet van de kansrekening laat men de gelijkwaardigheid van mogelijkheden - die de principiële moeilijkheid vormt bij de definitie van LAPLACE - vallen. Overigens is de opzet 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 ($10^6 \approx 10^6-1 \approx .. .. \approx 1$), 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.

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 gedefiniëerd waren. Een dergelijke verzameling hebben wij nu ook nodig en deze verzameling, die de basis van het wiskundige model vormt, is voor elk praktisch probleem verschillend. Een praktisch 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 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, $\overline{6}$, enz. Deze kenmerken

worden ook "eventualiteiten" genoemd, een term (ingevoerd door Prof. 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 een 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 inf. aannemen en Γ moet dus oneindig veel elementen bevatten. Dit maakt een kleine wijziging in de eigenschappen II en III wenselijk, terwijl een vijfde 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 gedefinieerde eventualiteiten wordt nu een constant getal als kans toegevoegd en deze kansen (die in de regel van onbekende grootte zijn) voldoen aan de volgende axioma's:

(4.1)

I	$0 \leq P[S] \leq 1$	voor iedere beschouwde eventualiteit S.
II	$P[S] = 0$	als geen element van Γ het kenmerk S bezit.
III	$P[S] = 1$	als ieder element van Γ het kenmerk S bezit.
IV	$P[S \vee U] = P[S] + P[U] - P[S \wedge U]$	voor ieder paar S,U van beschouwde eventualiteiten.

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

Het hier gegeven axiomastelsel kan door het volgende, eenvoudigere stelsel vervangen worden:

(4.1')
$$\begin{cases} \text{I'} & P[S] \geq 0 \quad \text{voor iedere beschouwde eventualiteit S.} \\ \text{III'} & P[S] = 1 \quad \text{als ieder element van } \Gamma \text{ het kenmerk S bezit.} \\ \text{IV'} & P[S \vee U] = P[S] + P[U] \quad \text{voor ieder paar eventualiteiten, dat elkaar uitsluit.} \end{cases}$$

Uit de axioma's I', III' en IV' kunnen de in de axioma's I t/m IV genoemde eigenschappen worden afgeleid, zodat het voor de kansrekening geen verschil maakt, of men uitgaat van het stelsel (4.1), dan wel van het stelsel (4.1'). Op grond van symmetrie-overwegingen en het feit dat het stelsel (4.1) iets meer aanspreekt, wordt hier van dit stelsel uitgegaan.

De op grond van de eigenschappen (3.3) voor onvoorwaardelijke fqn afgeleide stellingen gelden nu ook voor kansen, daar bij de bewijzen in paragraaf 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 dan:

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

(4.2)
$$P\left[\bigvee_{i=1}^{k} S_i\right] = \sum_{i=1}^{k} P[S_i] \quad .$$

Deze eigenschap geldt voor iedere eindige k. Als vijfde 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)
$$P\left[\bigvee_{i=1}^{\infty} S_i\right] = \sum_{i=1}^{\infty} P[S_i] \quad .$$

Het axiomasysteem is nu gereed. Een verzameling Γ van eventualiteiten met bijbehorende kansen, die aan deze rij 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, $P[S \mid T]$, op de eventualiteit S onder voorwaarde van het optreden van de eventualiteit T, die een kans $P[T] > 0$ bezit, wordt gedefinieerd door

$$P[S \mid T] \overset{\text{def}}{=} \frac{P[S \wedge T]}{P[T]} \quad . \tag{4.4}$$

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 paragraaf 3 afgeleide stellingen voor voorwaardelijke kansen.

Stelling 4.1

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

(4.5)

I	$0 \leq P[S \mid T] \leq 1$	voor iedere eventualiteit S.
II	$P[S \mid T] = 0$	als geen der elementen van Γ, die het kenmerk T bezitten, ook S bezit.
III	$P[S \mid T] = 1$	als ieder element, dat het kenmerk T bezit, ook S bezit.
IV	$P[S \vee U \mid T] = P[S \mid T] + P[U \mid T] - P[S \wedge U \mid T]$	voor ieder paar kenmerken S en U.
V	$P\left[\bigvee_{i=1}^{\infty} S_i \mid T\right] = \sum_{i=1}^{\infty} P[S_i \mid T]$	als $S_1, S_2, \ldots$ een oneindig exclusief systeem vormen voor de elementen, waarvoor eventualiteit T optreedt.

Bewijs:

I Volgens stelling 3.5 en axioma (4.1;I) geldt

(4.6) $$0 \leq P[S \wedge T] \leq P[T] ,$$

dus $$0 \leq P[S \mid T] = \frac{P[S \wedge T]}{P[T]} \leq 1 .$$

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 $P[T] = P[T \wedge S] + P[T \wedge \bar{S}]$ volgt dan: $P[T] = P[T \wedge S]$.

IV $$P[(S \vee U) \wedge T] = P[(S \wedge T) \vee (U \wedge T)] = P[S \wedge T] + P[U \wedge T] - P[S \wedge U \wedge T].$$

$$P[S \vee U \mid T] = \frac{P[(S \vee U) \wedge T]}{P[T]} = \frac{P[S \wedge T]}{P[T]} + \frac{P[U \wedge T]}{P[T]} - \frac{P[S \wedge U \wedge T]}{P[T]} =$$

$$= P[S \mid T] + P[U \mid T] - P[S \wedge U \mid T].$$

V Als $S_1, S_2, \ldots$ een exclusief systeem vormen voor de elementen waarbij T optreedt, dan is

$$P\Big[\bigvee_{i=1}^{\infty} S_i \mid T\Big] = \frac{P\Big[\bigvee_{i=1}^{\infty} S_i \wedge T\Big]}{P[T]} = \frac{P\Big[\bigvee_{i=1}^{\infty} (S_i \wedge T)\Big]}{P[T]} =$$

$$= \frac{\sum_{i=1}^{\infty} P[S_i \wedge T]}{P[T]} = \sum_{i=1}^{\infty} P[S_i \mid T] .$$

Tenslotte voeren wij nog het begrip stochastische onafhankelijkheid in. De eventualiteit S en de eventualiteit T zijn per definitie stochastisch onafhankelijk als

(4.7) $$P[S \wedge T] = P[S].P[T] .$$

Als S en T stochastisch onafhankelijk zijn en $0 < P[T] < 1$ is, geldt

$$(4.8) \qquad \begin{cases} P[S \mid T] = P[S \mid \bar{T}] = P[S] \ , \\ P[\bar{S} \mid T] = P[\bar{S} \mid \bar{T}] = P[\bar{S}] \end{cases}$$

en analoog met verwisseling van S en T. Omgekeerd volgt uit elk van de regels (4.8) de relatie (4.7). Het bewijs van deze eigenschappen en van de hieronder volgende laten wij aan de lezer over.

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

(4.10) Is $P[S] = 0$ of $P[S] = 1$, 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 ($h \leq k$) $S_{i_1}, S_{i_2}, \ldots, S_{i_h}$ daaruit, geldt

$$(4.11) \qquad P\left[\bigwedge_{j=1}^{h} S_{i_j}\right] = \prod_{j=1}^{h} P\left[S_{i_j}\right] .$$

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

$$(4.12) \qquad P\left[\bigwedge_{i=1}^{k} S_i\right] = \prod_{i=1}^{k} P[S_i] \ ;$$

(d.i. de <u>bijzondere productregel</u> voor onafhankelijke eventualiteiten).

Paarsgewijze onafhankelijkheid is echter niet voldoende om (4.11) te garanderen, en het vervuld zijn van (4.12) ook niet. Een tegenvoorbeeld voor het tweede is als volgt te construeren:

Laat Γ uit 30 elementen bestaan, aangegeven door de nummers $1, \ldots, 30$ 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 produktregel niet opgaat, is een Γ met 12 elementen $(1,\ldots,12)$, die ieder een kans 1/12 bezitten en

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

B: $1 \vee 2 \vee 5 \vee 6 \vee 7 \vee 8$

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

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

Een minder kunstmatig voorbeeld wordt gevonden door bij twee worpen met een zuivere dobbelsteen de eventualiteiten

A: eerste worp even

B: tweede worp even

C: som der ogen bij de twee worpen even

te beschouwen.

Het is van belang na te gaan, wat de praktische 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 (4.8) 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 praktische interpretatie en indien het praktisch 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 kan men in het model een eventuele afhankelijkheid toelaten. Er kan dan statistisch onderzocht worden of er inderdaad afhankelijkheid aanwezig is.

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 elkaars 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 bijv. voor de Nederlandse gezinshoofden anno 1969 de kenmerken "bezit een centraal verwarmd huis" en "bezit een auto", dan is er een duidelijke afhankelijkheid; de kans op autobezit is groter bij wie centrale verwarming bezit en omgekeerd.Toch koopt men geen auto omdat het hele huis warm is en laat men geen centrale verwarming aanleggen omdat men een auto heeft. Het is de parallelle invloed van het welvaartspeil, die zorgt voor een statistisch verband dat geen causaal verband is. Over een jaar of wat kan het er trouwens 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 (bijv. één worp met een zuivere dobbelsteen) wordt nl. uiteraard een basisverzameling genomen, waarvan alle elementen (de elementaire mogelijke uitkomsten van het spel) gelijke kans bezitten. Elk der elementen heeft dus 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 praktische 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 verifieerbaar karakter een hechtere basis voor de toepassingen geeft dan de andere interpretaties.

5. Voorbeelden van het berekenen van kansen.

Ter illustratie van het gebruik van de axioma's geven wij in deze paragraaf een aantal elementaire voorbeelden van berekeningen van kansen.

Voorbeeld 5.1

Van een produkt dat aan de lopende band gemaakt wordt, bezit 10% een afwerkingsfout. Men kiest onafhankelijk van elkaar twee exemplaren van het produkt als deze van de band komen. Wat is de kans dat beide goed zijn, dat beide fout zijn en dat er één goed en één fout is? Wanneer men weet dat er één goed is en één fout, hoe groot is dan de kans dat eerst een goed en daarna een fout exemplaar werd aangetroffen?

Oplossing

Geven wij een goed exemplaar aan door het kenmerk S ("succes") en een fout door het kenmerk M ("mislukking"), dan is $P[M] = 0,1$ en volgens (3.10) $P[S] = 0,9$. Daar de beide uitvoeringen van het experiment onafhankelijk zijn, is (4.7) van toepassing, zodat geldt

$$P[SS] = 0,81,\ P[SM] = P[MS] = 0,09,\ P[MM] = 0,01\ .$$

Verder is volgens (3.8)

$$P[SM \vee MS] = P[SM] + P[MS] = 0,18\ ,$$

zodat definitie (4.4) geeft

$$P[SM \mid SM \vee MS] = \frac{P[SM \wedge (SM \vee MS)]}{P[SM \vee MS]} = \frac{P[SM]}{P[SM \vee MS]} = \frac{0,09}{0,18} = \frac{1}{2} .$$

Hetzelfde geldt voor de voorwaardelijke kans $P[MS \mid SM \vee MS]$ en dus is

$$P[SM \mid SM \vee MS] = P[MS \mid SM \vee MS] = \frac{1}{2} . \qquad (5.1)$$

Men kan op dezelfde wijze als hier is aangegeven, nagaan dat het resultaat (5.1) ook verkregen wordt, wanneer $P[S] = p$ en $P[M] = q = 1-p$ is.

Voorbeeld 5.2

Een doos bevat 100 briefjes waarop de getallen 00 t/m 99 geschreven zijn. 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 *). Door dit model is het kansveld volledig beschreven.

De kans, dat de briefjes bijv. in de volgorde 00, 01, 02, ... , 99 getrokken zullen worden is nu gemakkelijk te berekenen met behulp van stelling 3.6. Geven wij het nummer van de trekking aan met een index,

---

*) 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.

dan is, daar in dit geval de definitie van LAPLACE toepasbaar is,

$$P[00_1] = \frac{1}{100} \ , \ P[01_2 \mid 00_1] = \frac{1}{99} \ , \ P[02_3 \mid 00_1 \wedge 01_2] = \frac{1}{98} \ , \text{ enz.;}$$

dus volgens stelling 3.6 is

$$(5.2) \qquad P[00_1 \wedge 01_2 \wedge \ldots \wedge 99_{100}] = \frac{1}{100} \cdot \frac{1}{99} \cdot \ldots \cdot \frac{1}{2} \cdot 1 = \frac{1}{100!} \ .$$

De kans dat de briefjes in de volgorde 00 t/m 99 getrokken worden is dus $\frac{1}{100!}$. Ditzelfde geldt ook voor iedere andere volgorde; m.a.w. alle 100! permutaties hebben kans $\frac{1}{100!}$.

Bevatte de doos n briefjes met de getallen 1 t/m n, dan hebben alle n! permutaties de 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 *) genoemd. De verzameling van elementen, waaruit deze trekkingen worden verricht, wordt veelal met het woord populatie aangegeven. 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 teruglegging; maar dan is schudden voor iedere trekking nodig.

Voorbeeld 5.3

Uit een doos met n briefjes, genummerd 1,2,...,n, worden aselect en zonder teruglegging k ($k \leq n$) briefjes getrokken. Hoe groot is de kans dat juist de nummers 1,...,k in één of andere volgorde worden getrokken?

Oplossing

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

---

*) Deze term is geïntroduceerd door Prof. VAN DANTZIG.

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

Voor iedere andere gegeven volgorde van deze k nummers geldt het zelfde en daar er slechts één volgorde uit kan komen, is volgens stelling 3.2 (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

$$\frac{k!(n-k)!}{n!} = \binom{n}{k}^{-1} . \tag{5.3}$$

Opmerking

Deze uitkomst geldt uiteraard niet alleen voor de nummers 1,...,k, doch voor ieder 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 3.4. Een aselect zonder teruglegging getrokken k-tal wordt, indien de volgorde van trekking buiten beschouwing wordt gelaten, een steekproef van de omvang k genoemd.

Wanneer men in de praktijk een dergelijke steekproef moet nemen, dan kan men deze uit de populatie kiezen met behulp van een lijst met aselecte cijfers (random digits). Een dergelijke lijst bestaat uit rijen cijfers, die verkregen zijn door onderling onafhankelijke trekkingen met teruglegging uit de getallen 0 t/m 9. Bij iedere trekking hebben dus alle waarden 0 t/m 9 een gelijke kans 1/10 om getrokken te worden. In een lijst met aselecte cijfers kan men ook steeds een gelijk aantal cijfers combineren; er ontstaan dan aselecte getallen. Combineert men bijv. steeds drie cijfers, dan verkrijgt men aselecte getallen tussen 000 en 999. Een steekproef van 25 stuks wordt nu uit een monster van 1000 genummerde exemplaren genomen door 25 aselecte getallen van drie cijfers te zoeken en de met deze getallen corresponderende exemplaren te controleren.

Over het opstellen van lijsten met aselecte cijfers/getallen en over hun eigenschappen zal bij de behandeling van Monte Carlo-methoden uitvoerig gesproken worden (vgl. deel 8 van deze syllabusserie).

Voorbeeld 5.4

Men zoekt in een tabel met aselecte cijfers 10 getallen op tussen 00 en 24. Hoe groot is de kans dat hierbij twee of meer gelijke getallen gevonden worden?

Oplossing

De gebeurtenis "het getal ligt tussen 00 en 24" geven wij aan met het kenmerk A. Alleen getallen met het kenmerk A kunnen gebruikt worden. Onder de voorwaarde dat A heeft plaatsgevonden, is de kans dat men een getal trekt met als eerste cijfer y en als tweede cijfer z, volgens (4.4), (4.7) en (4.9), gelijk aan

$$(5.4)\qquad P[yz|A] = \frac{P[yz \wedge A]}{P[A]} = \frac{P[yz]}{P[A]} = \frac{P[y].P[z]}{P[A]} = \frac{(\frac{1}{10})^2}{\frac{25}{100}} = \frac{1}{25}\ .$$

Onder de voorwaarde dat het aselecte getal ligt tussen 00 en 24 is de kans voor ieder getal dus even groot en wel 1/25. De gevraagde kans is één min de kans dat de aselecte getallen alle ongelijk zijn. Geven wij het eerst gevonden aselecte getal aan met a_1, het tweede met a_2, enz., dan is

$$(5.5)\quad \begin{cases} P[\text{alle } a_i \text{ ongelijk voor } i \leq 10] = \\ = P[a_2 \neq a_1] \,.\, P[\text{alle } a_i \text{ ongelijk voor } i \leq 3 \mid a_2 \neq a_1] \ \ldots \\ .\ P[\text{alle } a_i \text{ ongelijk voor } i \leq 10 \mid \text{alle } a_i \text{ ongelijk voor } i \leq 9] = \\ = \frac{24}{25} \cdot \frac{23}{25} \ \ldots \ \frac{16}{25} = \frac{24.23\ldots16}{25^9} = \frac{24!}{25^9.15!} = 0{,}12\ . \end{cases}$$

De gezochte kans is dus 1 - 0,12 = 0,88.

Voorbeeld 5.5

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

A) de trekkingen met teruglegging geschieden,

B) de trekkingen zonder teruglegging geschieden?

Oplossing

A) Bij trekkingen met teruglegging is de k^{de} trekking een aselecte trekking uit dezelfde verzameling knikkers als de eerste, zodat volgens de definitie van LAPLACE de gevraagde kans gelijk is aan

(5.6) $$\frac{M}{M + N} .$$

B) Bij trekkingen zonder teruglegging is de samenstelling van de knikker-verzameling bij de k^{de} trekking niet dezelfde als bij de eerste, maar hangt af van de resultaten der eerste k-1 trekkingen. Toch is de uitkomst dezelfde als bij A, zoals op de volgende wijze in te zien is. Als wij na de k^{de} trekking doorgaan tot alle M + N knikkers getrokken zijn, beïnvloeden wij het resultaat van de k^{de} trekking niet. Denken wij de knikkers genummerd van 1,...,M + N, dan zijn alle (M + N)! permutaties der trekkingsvolgorde even waarschijnlijk, zodat alle knikkers dezelfde kans bezitten om de k^{de} getrokkene te zijn. Er zijn dus voor de k^{de} trekking weer M + N mogelijkheden met gelijke kansen, waaruit volgt dat de kans op een witte k^{de} knikker weer gelijk is aan

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

Opmerking

De berekende kans is de onvoorwaardelijke kans dat de k^{de} 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 op een bepaald resultaat 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 5.6

Iemand passeert elke ochtend op weg van zijn huis naar kantoor een brug, die bij nadering van een schip wordt geopend. Als hij heeft opgemerkt dat de brug één op de tien keer open is, hoe groot is dan de kans P_4 dat hij precies vier keer in een maand van 25 werkdagen moet wachten? En hoe groot is de kans dat hij minstens vier maal moet wachten?

Oplossing

Wij noemen een dichte brug een succes (S), een geopende brug een mislukking (M) en geven de kans op succes aan met p, de kans op mislukking met $q = 1-p$ en het aantal experimenten (i.e. ritjes van huis naar kantoor) met n (=25).

In ons model nemen wij aan dat $q = 0,1$ en dus $p = 0,9$. Bovendien maken wij de redelijk lijkende veronderstelling dat de experimenten onafhankelijk zijn (zie de opmerking bij voorbeeld 5.5), zodat wij de bijzondere produktregel (4.12) kunnen toepassen. Daaruit volgt dat de kans op de rij uitkomsten

$$S\ S\ M\ S\ M\ M\ \ldots\ M$$

gelijk is aan

$$p.p.q.p.q.q\ \ldots\ q\ .$$

Als er dus in de rij k (=21) successen en $n-k$ (=4) mislukkingen voorkomen, is deze kans gelijk aan

$$p^k . q^{n-k} .$$

Het aantal verschillende rijen met k successen en $n-k$ mislukkingen is gelijk aan het aantal manieren waarop we de k successen in de rij van n uitkomsten kunnen plaatsen. Dit aantal is $\binom{n}{k}$. Volgens de bijzondere optelregel geldt dus

(5.7) $$P_k = \binom{n}{k} p^k q^{n-k} \quad (\text{met } \binom{n}{0} \overset{\text{def}}{=} 1).$$

Vullen wij in (5.7) de waarden $n = 25$, $k = 21$ en $p = 0,9$ in, dan vinden wij dat het antwoord op de eerste vraag luidt

$$P_{21} = \binom{25}{21} (0,9)^{21} (0,1)^{4} = 0,14 \ . \tag{5.8}$$

Het antwoord op de tweede vraag wordt gevonden door de in (5.7) opgegeven kansen P_k na invullen van $n = 25$ en $p = 0,9$ te sommeren van $k = 0$ t/m $k = 21$. Het resultaat is

$$\sum_{k=0}^{21} \binom{25}{k} (0,9)^{k} (0,1)^{25-k} = 0,24 \ . \tag{5.9}$$

Opmerking

De grootheid k kan de waarde $0,1,\ldots,n$ aannemen. Daar deze $n+1$ mogelijkheden een kategorisch systeem vormen, geldt volgens stelling 3.3

$$\sum_{k=0}^{n} P_k = \sum_{k=0}^{n} \binom{n}{k} p^{k} q^{n-k} = 1 \ , \tag{5.10}$$

hetgeen overeenkomt met de binomiale ontwikkeling van $(p+q)^n$.

De hier gevonden kansen P_k behoren bij de binomiale verdeling, waarop wij o.a. in paragraaf 7 terugkomen.

Voorbeeld 5.7

Een rij van N posten in een kasboek moet nagekeken worden om te controleren of er veel fouten gemaakt zijn. Men neemt een steekproef zonder teruglegging van n exemplaren. Hoe groot is de kans dat er k fouten aangetroffen worden, wanneer er in totaal F gemaakt zijn?

Oplossing

Volgens voorbeeld 5.3 hebben alle $\binom{N}{n}$ n-tallen gelijke kans om getrokken te worden. De steekproef bezit het gezochte kenmerk wanneer er k posten zijn met een fout en n-k zonder fout. Uit de F posten met een fout kunnen $\binom{F}{k}$ verschillende k-tallen gekozen worden en uit de

N-F posten zonder fout $\binom{N-F}{n-k}$ verschillende (n-k)-tallen. Onder de $\binom{N}{n}$ n-tallen zijn er dus

$$\binom{F}{k}\cdot\binom{N-F}{n-k}$$

met het gevraagde kenmerk, zodat volgens de definitie van LAPLACE geldt

(5.11) $$P'_k = \frac{\binom{F}{k}\cdot\binom{N-F}{n-k}}{\binom{N}{n}} .$$

Opmerking

De grootheid k neemt nu gehele waarden aan tussen max(n-N+F,0) en min(F,n), deze grenzen inbegrepen. Buiten deze grenzen is $P'_k = 0$ en het rechterlid van (5.11) ook, volgens de afspraak dat $\binom{a}{b} = 0$ als $b < 0$ of $> a$. Derhalve geldt volgens stelling 3.3

(5.12) $$\sum_k \binom{F}{k}\cdot\binom{N-F}{n-k} = \binom{N}{n} .$$

Als n, dus ook k, klein is t.o.v. F en N-F, dan is

(5.13) $$P'_k \approx P_k \quad \text{met} \quad p = \frac{F}{N} .$$

Dit is als volgt in te zien:

$$P'_k = \frac{F!}{k!(F-k)!}\cdot\frac{(N-F)!}{(n-k)!(N-F-n+k)!}\cdot\frac{n!(N-n)!}{N!} .$$

Hierin is, als $k \ll F$,

$$\frac{F!}{(F-k)!} = F(F-1)\ldots(F-k+1) \approx F^k$$

en, als $n-k \ll N-F$,

$$\frac{(N-F)!}{(N-F-n+k)!} = (N-F)(N-F-1)\ldots(N-F-n+k+1) \approx (N-F)^{n-k} .$$

Voor $n << N$ is verder

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

Vult men deze drie benaderingen, samen met $\frac{F}{N} = p$ in (5.11) in, dan vindt men

$$P'_k \approx \binom{n}{k} p^k (1-p)^{n-k} . \qquad (5.14)$$

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

Voorbeeld 5.8

Wanneer in voorbeeld 5.7 het aantal posten 150 bedraagt, het aantal fouten 10 en de steekproef 50 posten omvat, hoe groot is dan de kans dat geen enkele fout ontdekt wordt?

Oplossing

In dit voorbeeld is $N = 150$, $F = 10$ en $n = 50$, terwijl het vinden van geen enkele fout betekent dat $k = 0$ is. Substitutie van deze waarden in (5.11) geeft dan voor de gezochte kans

$$P'_0 = \frac{\binom{10}{0} \cdot \binom{140}{50}}{\binom{150}{50}} = 0{,}015 . \qquad (5.15)$$

Voorbeeld 5.9

Een bureau voor marktanalyse heeft voor zijn consumenten-panel 10 arbeidersgezinnen uit een provincieplaats nodig. Uit ervaring weet men dat het bij huisbezoek gelukt 40% van de bezochte gezinnen te overreden aan het panel deel te nemen. Hoe groot is de kans dat men bij het 20^{ste} bezoek de 10^{de} deelnemer vindt?

Oplossing

De gezochte kans is gelijk aan de kans, dat bij de eerste 19 bezoeken 9 successen verkregen worden en bij het 20^{ste} ook een succes. In ons model stellen wij de kans op het laatste op 2/5, terwijl wij bovendien aannemen dat de resultaten van verschillende bezoeken onafhankelijk zijn. De kans op het eerste wordt dan gegeven door (5.7) met $n = 19$, $k = 9$, $p = 2/5$ en $q = 3/5$. Wederom wegens de onafhankelijkheid der experimenten moeten deze beide kansen met elkaar vermenigvuldigd worden, hetgeen leidt tot

$$Q_{20} = \binom{19}{9} \left(\frac{2}{5}\right)^{10} \left(\frac{3}{5}\right)^{10} = 0{,}059 \; . \tag{5.16}$$

Algemeen geldt voor een reeks van onafhankelijke experimenten, ieder met een kans p op succes, die wordt voortgezet tot precies k successen verkregen zijn, dat de kans Q_n, dat hiervoor n experimenten nodig zijn, gelijk is aan

$$Q_n = \binom{n-1}{k-1} p^k q^{n-k} \qquad (q = 1-p) \; . \tag{5.17}$$

Voorbeeld 5.10 *)

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?

---

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

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 meisjesgeboorten 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-samenstelling 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 laatstgenoemde.

Voorbeeld 5.11

Op een cursus is afgesproken dat alleen kan worden begonnen, wanneer de docent en alle cursisten aanwezig zijn. In het verleden is gebleken dat de docent één op de tien keer te laat komt en dat er één op de vijf keer één of meer cursisten te laat komen. Verder heeft de concierge opgemerkt dat de cursus één op de vijf keer te laat begint. Welke conclusie kunt U hieruit trekken over het al dan niet stochastisch onafhankelijk zijn van het te laat komen van cursisten en docent?

Oplossing

Noemen wij het te laat komen van de docent gebeurtenis A en het te laat komen van minstens één cursist gebeurtenis B, dan kunnen wij de voorgaande observaties in ons model voorstellen door $P[A] = 1/10$, $P[B] = 1/5$ en $P[A \vee B] = 1/5$. De cursus begint immers te laat wanneer de gebeurtenis $A \vee B$ optreedt. Uit

$$P[A \vee B] = P[A] + P[B] - P[A \wedge B]$$

samen met de gegevens, volgt $P[A \wedge B] = 1/10$. Wegens $P[A].P[B] = 1/50$ is $P[A \wedge B] \neq P[A].P[B]$ en dus zijn de gebeurtenissen A en B niet stochastisch onafhankelijk.

De voorwaardelijke kansen $P[A \mid B]$ en $P[B \mid A]$ zijn resp.

$$P[A \mid B] = \frac{P[A \wedge B]}{P[B]} = \frac{\frac{1}{10}}{\frac{1}{5}} = \frac{1}{2}$$

en

$$P[B \mid A] = \frac{P[A \wedge B]}{P[A]} = \frac{\frac{1}{10}}{\frac{1}{10}} = 1 \ .$$

Voorbeeld 5.12

Gegeven zijn N+1 urnen $U_0, U_1, \ldots, U_N$. In elk der urnen bevinden zich N knikkers; U_k bevat k rode en N-k witte knikkers. Uit deze verzameling urnen wordt aselect een urn gekozen, waaruit aselect met teruglegging n knikkers worden genomen. Bereken de kans dat alle n knikkers rood zijn.

Oplossing

Wanneer kenmerk A betekent "alle n knikkers zijn rood", dan is volgens (3.17)

$$P[A] = \sum_{i=0}^{N} P[A \mid U_i].P[U_i] =$$

$$= \sum_{i=0}^{N} \left(\frac{i}{N}\right)^n . \frac{1}{N+1} = \frac{1}{N^n(N+1)} \sum_{i=1}^{N} i^n \ .$$

6. Het meten van kansen.

In de vorige paragraaf zijn een aantal voorbeelden gegeven van het berekenen van kansen. Meestal geschiedde dit op grond van numeriek gegeven kansen, maar soms, zoals in voorbeeld 5.6, door de berekening te baseren op een onbekende kans p, die dan ook in het resultaat (5.7) voorkomt. In alle voorbeelden, op het tiende na, konden de gevraagde kansen expliciet berekend worden. In feite werd in al die gevallen waarin numerieke uitkomsten verkregen zijn, uitgegaan van de gelijkheid van een aantal kansen.

Voor de praktische 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 ingevoerd maatbegrip nodig is dit ook praktisch meetbaar te maken, d.w.z. experimentele middelen aan te geven om er in praktijkproblemen een (min of meer) bepaalde waarde aan toe te kennen. Denkt men bijv. aan het begrip "lengte", dan zal men een praktisch uitvoerbare procedure aan moeten geven, die aan de volgende eisen voldoet:

a) Men moet kunnen nagaan 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 worden gebracht 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.

Punt b) is daarbij het eenvoudigste, daar een absolute maat voor het kansbegrip reeds in de axioma's II en III vervat is. Men kan dan ook langs eenvoudige weg en met simpele middelen een kans van willekeurige, gegeven grootte "construeren", althans als die grootte door een rationaal getal wordt gegeven.

De punten a)en c)eisen meer voorbereiding. Men ontmoet in de statistiek de oplossing van a)resp. c)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 verkrijgen 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.

Wij zullen deze methoden, die overigens slechts een klein gedeelte van de statistiek vormen, in deze cursus niet volledig behandelen, maar wel voor zover wij ze later nodig hebben.

7. Stochastische grootheden en één-dimensionale kansverdelingen.

Definitie 7.1

Een grootheid x, die bij een experiment verschillende reële waarden aan kan nemen, wordt een stochastische grootheid genoemd, indien voor ieder gegeven reëel getal x, de kans dat 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 bijv.: 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 levertijd van een uitgegeven order; het aantal fouten

dat in een week in een administratie gemaakt wordt en het aantal telefoongesprekken dat gedurende een bepaald tijdsinterval over een bepaalde lijn gevoerd wordt, enz., enz. .

Notatie en terminologie

Stochastische grootheden geven wij aan met onderstreepte letters: $\underline{x}$, $\underline{y}$, $\underline{z}$, $\underline{x}_1$, $\underline{x}_2$, $\underline{x}_i$, enz. De getalwaarden die zij kunnen aannemen, vaak beschouwd als algebraïsche variabelen, worden door dezelfde letters zonder onderstreping aangegeven. In plaats van onderstreepte letters worden in boeken vaak vet gedrukte letters of hoofdletters gebruikt.

De kansverdeling van een stochastische grootheid $\underline{x}$ is de verzameling van alle waarden x, die $\underline{x}$ kan aannemen, met de bijbehorende kansen $P\left[\underline{x} \leq x\right]$.

De verdelingsfunctie van $\underline{x}$ is de functie

$$F(x) \stackrel{\text{def}}{=} P\left[\underline{x} \leq x\right] = \tag{7.1}$$

= de kans dat $\underline{x}$ een waarde $\leq$ x aanneemt, opgevat als functie van de algebraïsche variabele x.

Uit deze definitie en de axioma's I t/m V uit paragraaf 4 volgt, dat iedere verdelingsfunctie monotoon niet dalend is en uitsluitend waarden in het gesloten interval $[0,1]$ aanneemt, terwijl

$$\begin{cases} F(-\infty) \stackrel{\text{def}}{=} \lim\limits_{x \to -\infty} F(x) = 0 , \\ F(+\infty) \stackrel{\text{def}}{=} \lim\limits_{x \to +\infty} F(x) = 1 . \end{cases} \tag{7.2}$$

Definitie 7.2

Discrete kansverdelingen zijn kansverdelingen waarbij de beschouwde stochastische grootheid een eindig of aftelbaar aantal waarden $x_1, x_2, \ldots$ kan aannemen, terwijl

$$P_i \stackrel{\text{def}}{=} P\left[\underline{x} = x_i\right] > 0 , \quad \sum_i P_i = 1 . \tag{7.3}$$

De bijbehorende verdelingsfuncties zijn trapfuncties

(7.4) $$F(x) = \sum_{x_i \leq x} P_i \ .$$

Ter illustratie dienen de figuren 7.1 en 7.2.

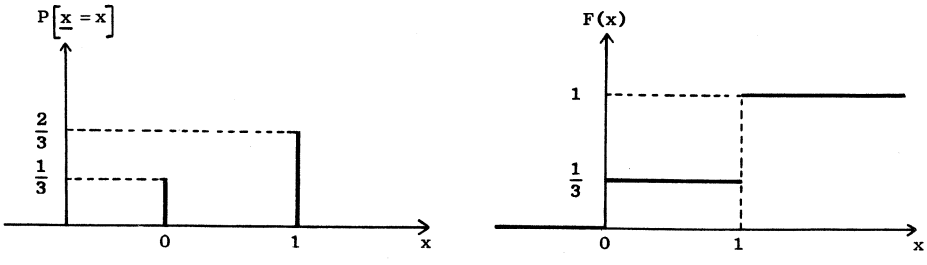


fig. 7.1

Afzonderlijke kansen en verdelingsfunctie
van een alternatief met p = 2/3 (zie 7.5)

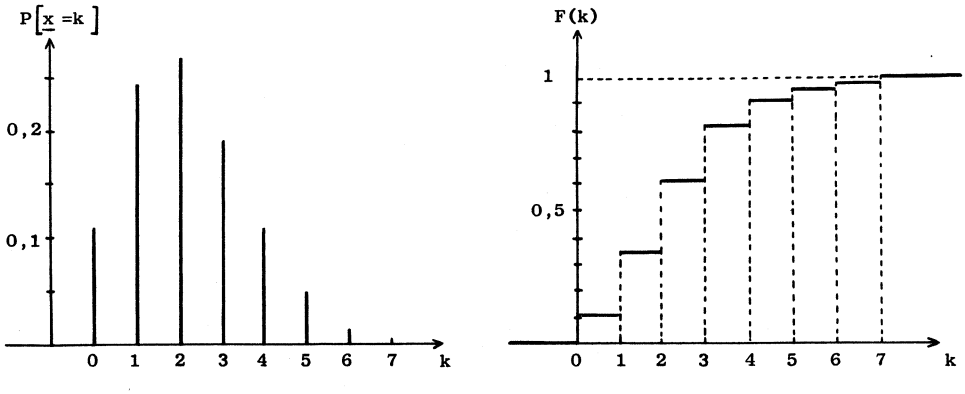


fig. 7.2

Afzonderlijke kansen en verdelingsfunctie
van een Poissonverdeling met λ = 2,2 (zie 7.9)

Wij geven nu een opsomming van de belangrijkste discrete verdelingen. Daarbij zullen wij zien dat in de algebraïsche vorm van de afzonderlijke kansen en van de verdelingsfunctie vaak één of meer constanten optreden, die men parameters noemt. Bij iedere keuze van de parameterwaarden behoort dus één kansverdeling. Wanneer de berekening van de afzonderlijke kansen of van de verdelingsfunctie bewerkelijk is, kan men dikwijls gebruik maken van tabellen. Wij zullen er in het volgende een aantal noemen. Een algemeen overzicht van statistische tabellen geeft: J.A. GREENWOOD & H.O. HARTLEY, Guide to Tables in Mathematical Statistics, Princeton University Press, (1962).

1. Het alternatief of de dichotomie (voorbeeld 5.1) met parameter p $(0 < p < 1)$:

$$P[\underline{x} = 1] = p; \qquad P[\underline{x} = 0] = q = 1-p \ . \tag{7.5}$$

2. De binomiale verdeling (voorbeeld 5.6) met parameters n (positief geheel) en p $(0 < p < 1)$:

$$P[\underline{x} = k] = \binom{n}{k} p^k q^{n-k} \quad (k=0,1,\ldots,n; q \overset{\text{def}}{=} 1-p). \tag{7.6}$$

Van deze verdeling bestaan uitvoerige tabellen, waarvan wij noemen: Tables of the Binomial Probability Distribution, National Bureau of Standards, Applied Mathematics Series 6 (1950) en Tables of the Cumulative Binomial Probabilities, Ordnance Corps Pamphlet ORDP 20-1, U.S. Army Ordnance Corps (1952). Beide tabellen bevatten de waarden van de verdelingsfunctie en de eerstgenoemde ook die van de afzonderlijke termen; p doorloopt de waarden 0,01 (0,01) 0,50 en n de waarden 2 (1) 49 resp. 1 (1) 150.*) Het benaderen van de kansen voor grote waarden van n zullen wij in paragraaf 13 behandelen.

3. De hypergeometrische verdeling (voorbeeld 5.7) met parameters M, N en n (alle positief geheel):

$$P[\underline{x} = k] = \frac{\binom{M}{k}\binom{N}{n-k}}{\binom{M+N}{n}} \quad (k \text{ geheel tussen } \max(n-N,0) \text{ en } \min(M,n)). \tag{7.7}$$

*) De notatie a (c) b betekent "van a met stapgrootte c tot en met b"; dus a, a+c, a+2c, ..., b.

Ook deze verdeling is, voor M+N = 2 (1) 50 (10) 100, getabelleerd en wel in: G.L. LIEBERMAN and D.B. OWEN, Tables of the Hypergeometric Probability Distribution, Stanford University Press, Stanford (1961). Tabellen van de binomiaalcoëfficiënten $\binom{m}{n}$ kan men o.a. vinden in: TH.C. FRY, Probability and its Engineering Uses, D. van Nostrand, New York (1928), waarin m de waarden 1 (1) 100 doorloopt. Verder kan deze verdeling in veel gevallen goed benaderd worden door de binomiale verdeling (vgl. de opmerking bij voorbeeld 5.7).

4. De negatief binomiale verdeling of Pascalverdeling (voorbeeld 5.9), met parameters k (positief geheel) en p (0 < p < 1):

$$(7.8) \qquad P\left[\underline{x} = x\right] = \binom{x-1}{k-1} p^k q^{x-k} \qquad (x=k,k+1,\ldots;q = 1-p).$$

Tussen de kansen (7.8) en (7.6) bestaat een eenvoudig verband; vermenigvuldigt men de kansen (7.6) met

$$\frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n},$$

en vervangt men daarna n door x, dan ontstaan de kansen (7.8). Met behulp van deze relatie kunnen de kansen van de negatief binomiale verdeling altijd uit een tabel van de binomiale verdeling berekend worden.

5. De Poissonverdeling met positieve parameter λ:

$$(7.9) \qquad P\left[\underline{x} = k\right] = \frac{\lambda^k . e^{-\lambda}}{k!} \qquad (k=0,1,\ldots).$$

Deze verdeling is o.a. getabelleerd in: GENERAL ELECTRIC, Tables of the Individual and Cumulative Terms of the Poisson Distribution, D. van Nostrand, Princeton (1962); de tabel geeft $P\left[\underline{x} = k\right]$, $P\left[\underline{x} \geq k\right]$ en $P\left[\underline{x} \leq k\right]$ voor λ-waarden tussen 10^{-7} en 205. Verder bevat het uitgebreide tabellenboek: E.S. PEARSON and H.O. HARTLEY, Biometrika Tables for Statisticians, Vol. I, Cambridge (1954), een tabel van afzonderlijke termen voor λ = 0,1 (0,1) 15 (tabel 39) en een cumulatieve tabel voor λ-waarden tussen 0,0005 en 67 (tabel 7).

6. De (discrete) homogene verdeling met parameter n (positief geheel):

$$P\left[\underline{x} = k\right] = \frac{1}{n+1} \qquad (k=0,1,\ldots,n). \tag{7.10}$$

Definitie 7.3

Continue kansverdelingen zijn kansverdelingen, waarbij de stochastische grootheid een continuum van waarden (bijv. alle waarden van een eindig of oneindig interval) kan aannemen, terwijl de verdelingsfunctie F(x) voor iedere x continu is en voor iedere x, op hoogstens eindig veel na, continu differentieerbaar.

De afgeleide

$$f(x) \overset{\text{def}}{=} \frac{d}{dx} F(x) \tag{7.11}$$

wordt dan de verdelingsdichtheid genoemd. Daar F(x) niet dalend is, geldt $f(x) \geq 0$ voor alle x, waarvoor f(x) gedefinieerd is, terwijl voorts geldt

$$F(x) = \int_{-\infty}^{x} f(v)\, dv \qquad *) \tag{7.12}$$

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

$$P\left[x_1 < \underline{x} \leq x_2\right] = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\, dx \tag{7.13}$$

(vgl. fig. 7.3),

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \,, \tag{7.14}$$

$$P\left[\underline{x} = x\right] = 0 \qquad \text{voor iedere } x. \tag{7.15}$$

---

*) 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 $x \leq a$, 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.

Uit (7.13) en (7.15) volgt nu: $P[x_1 \leq x \leq x_2] = \int_{x_1}^{x_2} f(x)\,dx.$

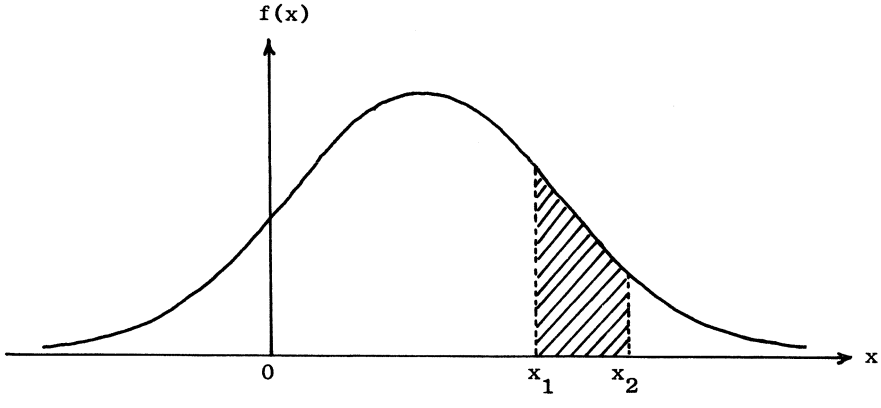


fig. 7.3

Meetkundige illustratie van (7.13)

Deze kansverdelingen treden op als limiet van discrete kansverdelingen (zij worden daarom vaak voor benaderingsdoeleinden gebruikt; zie paragraaf 13) en als wiskundig model bij grootheden, die kwantitatief op een bij benadering continue schaal afgelezen worden. Voorbeelden daarvan zijn: de lengte van een rekruut, de tijdsduur van een treinreis, de snelheid van een vliegtuig en de vraag per week naar een massa-artikel. De belangrijkste continue verdelingen zijn hieronder opgesomd.

7. De (continue) homogene verdeling, met parameters θ_1 (beginpunt) en $\theta_2 > \theta_1$ (eindpunt):

(7.16)
$$f(x) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1} & \text{voor } \theta_1 \leq x \leq \theta_2\,, \\ 0 & \text{elders;} \end{cases}$$

(7.17)
$$F(x) = \begin{cases} 0 & \text{voor } x < \theta_1\,, \\ \dfrac{x-\theta_1}{\theta_2 - \theta_1} & \text{voor } \theta_1 \leq x \leq \theta_2\,, \\ 1 & \text{voor } x > \theta_2\,. \end{cases}$$

8. De normale verdeling, met parameters $\sigma > 0$ en μ:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (-\infty < x < +\infty), \tag{7.18}$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv . \tag{7.19}$$

Notatie

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

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 $\underline{u}$. Tabellen van de verdelingsdichtheid en de verdelingsfunctie van $\underline{u}$ vindt men o.a. in de reeds genoemde verzameling van PEARSON en HARTLEY; $\underline{u}$ doorloopt hierin de waarden 0,00 (0,01) 6,00 (tabel 1). Een beknopte tabel is aan het eind van dit boekje opgenomen. Het is overbodig de normale verdeling voor andere parameterwaarden dan $\mu = 0$, $\sigma = 1$ te tabelleren wegens de volgende belangrijke stelling.

Stelling 7.1

Als $\underline{x}$ een $N(\mu,\sigma)$-verdeelde grootheid is, dan is $\underline{u} \overset{\text{def}}{=} \frac{\underline{x} - \mu}{\sigma}$ $N(0,1)$-verdeeld.

Bewijs:

De gebeurtenis $\frac{\underline{x}-\mu}{\sigma} \leq u$ treedt dan en slechts dan op als $\underline{x} \leq \mu+u\sigma$ optreedt; m.a.w. beide gebeurtenissen hebben dezelfde kans *). De verdelingsfunctie G van $\underline{u}$ voldoet dus aan

$$G(u) = P[\underline{u} \leq u] = P[\underline{x} \leq \mu+u\sigma] = F(\mu+u\sigma).$$

---

*) Dit enigszins intuïtieve argument kan gepreciseerd worden met stelling 11.1.

De verdelingsdichtheid g van $\underline{u}$ wordt hieruit verkregen door naar u te differentiëren (vgl. deel 1, stelling 12.4). Dit levert

$$g(u) = \frac{d}{du} G(u) = \frac{d}{du} F(\mu+u\sigma) = \sigma f(\mu+u\sigma) =$$

$$= \frac{\sigma}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu+u\sigma-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} .$$

Dus $\underline{u}$ heeft juist de verdelingsdichtheid (7.18) met $\mu = 0$ en $\sigma = 1$

Voorbeeld 7.1

$\underline{x}$ is N(4;0,5)-verdeeld. Gevraagd wordt $P\left[\underline{x} \geq 4,75\right]$ te berekenen.

Oplossing

De berekening verloopt als volgt

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

Uit de tabel van de normale verdeling blijkt, dat de gezochte kans 0,067 bedraagt.

9. De exponentiële verdeling, met positieve parameter λ:

(7.20) $$f(x) = \lambda e^{-\lambda x} \qquad (0 \leq x < \infty),$$

(7.21) $$F(x) = \int_0^x \lambda e^{-\lambda v} \, dv = 1 - e^{-\lambda x} .$$

Zowel de waarden van f(x) als van F(x) kunnen gemakkelijk gevonden worden m.b.v. een tabel van natuurlijke logarithmen of van e-machten; een dergelijke tabel is bijv.: H.W. HOLTAPPEL, Tafels van e^x, P. Noordhoff, Groningen (1938); x loopt van -9,999 met sprongen van 0,001 op t/m 9,999.

10. De χ^2 (chi-kwadraat)-verdeling, met parameter n (positief geheel):

$$(7.22) \qquad f(x) = \frac{1}{2^{\frac{1}{2}n}\,\Gamma(\frac{1}{2}n)}\; x^{\frac{1}{2}n-1}\; e^{-\frac{1}{2}x} \qquad (0 \leq x < \infty) ,$$

$$(7.23) \qquad F(x) = \frac{1}{2^{\frac{1}{2}n}\,\Gamma(\frac{1}{2}n)} \int_0^x v^{\frac{1}{2}n-1}\; e^{-\frac{1}{2}v}\; dv .$$

In de formules (7.22) en (7.23) is Γ de gamma-functie; voor p positief, geheel is $\Gamma(p) = (p-1)!$ terwijl (vgl. deel 1, voorbeeld 11.11) algemeen geldt

$$(7.24) \qquad \Gamma(p) = \int_0^\infty e^{-x}\; x^{p-1}\; dx .$$

De parameter n wordt vaak het aantal vrijheidsgraden van de χ^2-verdeling genoemd. Een tabel van de verdelingsfunctie voor n = 1 (1) 30 (2) 70 is opgenomen in de verzameling van PEARSON en HARTLEY (tabel 7).

11. De gamma-verdeling, met positieve parameters α en λ:

$$(7.25) \qquad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\; e^{-\lambda x}\; x^{\alpha-1} \qquad (0 \leq x < \infty) ,$$

$$(7.26) \qquad F(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^x e^{-\lambda v}\; v^{\alpha-1}\; dv .$$

De onder 9 en 10 genoemde verdelingen zijn speciale gevallen van de gamma-verdeling. De exponentiële verdeling verkrijgt men door in (7.25) voor α de waarde 1 in te vullen ($\Gamma(1) = 1$) en de χ^2-verdeling met n vrijheidsgraden door $\lambda = \frac{1}{2}$ en $\alpha = \frac{1}{2}n$ in te vullen.

Een ander speciaal geval van de gamma-verdeling is de zg. Erlang-verdeling, die ontstaat wanneer men voor α alleen gehele positieve waarden kiest.

Voor een nadere beschouwing van de klasse van gamma-verdelingen

en een overzicht van de bestaande tabellen verwijzen wij naar: GERDA KLERK-GROBBEN, De gamma-verdeling: overzicht betreffende eigenschappen, schattings- en toetsingsmethoden, rapport S 250, Afdeling Mathematische Statistiek, Mathematisch Centrum, Amsterdam (1959).

12. De bêta-verdeling, met positieve parameters α en β:

$$(7.27) \qquad f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)} \; x^{\alpha-1} \; (1-x)^{\beta-1} \qquad (0 \leq x \leq 1) \; ,$$

$$(7.28) \qquad F(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha).\Gamma(\beta)} \int_0^x v^{\alpha-1} \; (1-v)^{\beta-1} \; dv \; .$$

Waarden van F(x) kunnen o.a. gevonden worden met behulp van tabel 16 en nomogram 17 uit de verzameling van PEARSON en HARTLEY; in de tabel doorloopt α de waarden 0,5 (0,5) 15 20 30 60 ∞ en β de waarden 0,5 (0,5) 5 6 7,5 10 12 15 20 30 ∞.

13. De logarithmisch normale verdeling, met parameters $\sigma > 0$ en μ:

$$(7.29) \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \; . \; \frac{1}{x} \qquad (0 \leq x < \infty) \; ,$$

$$(7.30) \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x e^{-\frac{(\log v - \mu)^2}{2\sigma^2}} \; . \; \frac{1}{v} \; dv \; .$$

Deze verdeling kan uit de normale verdeling worden afgeleid door voor x te substitueren log y; de grootheid x is dan zelf niet normaal verdeeld, maar de logarithme ervan wel. Deze transformatie kan ook gebruikt worden om met behulp van tabellen van de normale verdeling de waarden van f(x) en F(x) te vinden.

Ter illustratie is in figuur 7.4 de verdelingsdichtheid en de verdelingsfunctie van een normale verdeling met parameters $\mu = 0$ en $\sigma = 1$ getekend. De figuur geeft de verdeling $N(\mu,\sigma)$ weer als men

langs de x-as voor elk getal x_0 leest: $x_0\sigma+\mu$, en uitsluitend in de linker figuur voor elke f_0 langs de vertikale as leest: f_0/σ. De verdelingsdichtheid is dus spitser en de verdelingsfunctie stijgt sneller als σ klein is, terwijl μ alleen een verschuiving bewerkstelligt. Bij elke $N(\mu,\sigma)$-verdeling is de afstand van het punt $x = \mu$ tot de abscis van het buigpunt van f(x) gelijk aan σ, hetgeen men kan verifiëren door de tweede afgeleide van f(x) gelijk aan nul te stellen.

Een tweede illustratie geeft figuur 7.5, waarin de verdelingsdichtheid en de verdelingsfunctie van een exponentiële verdeling met parameter $\lambda = \frac{1}{2}$ is getekend.

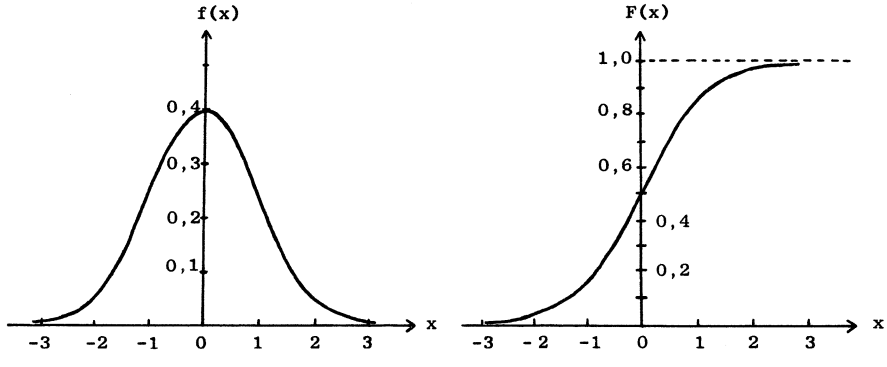


fig. 7.4

Verdelingsdichtheid en verdelingsfunctie voor de normale verdeling met $\mu = 0$ en $\sigma = 1$

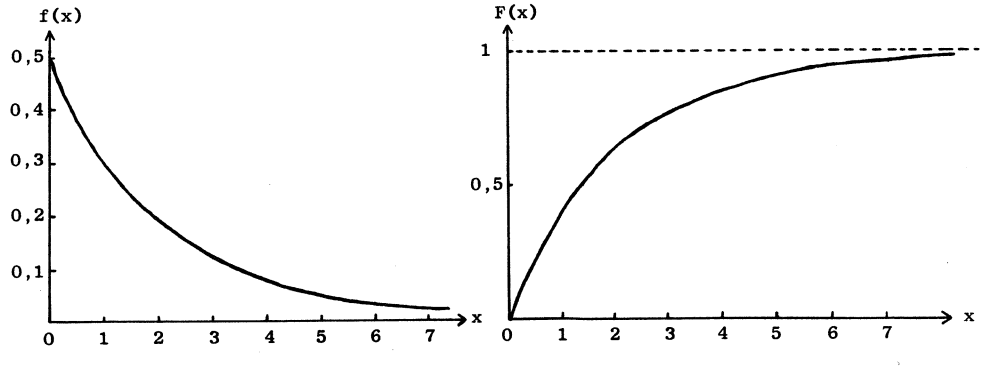


fig. 7.5

Verdelingsdichtheid en verdelingsfunctie voor de exponentiële verdeling met $\lambda = 0,5$

Na deze opsomming van verdelingen willen wij nog verwijzen naar twee Nederlandse publikaties.

1. In een artikel van H.J. PRINS, Statistica Neerlandica, 14 (1960), 1-18, wordt een overzicht gegeven van methoden om kansen, behorende bij bepaalde verdelingen, te berekenen met behulp van tabellen van andere verdelingen.

2. Onder redaktie van de Vereniging voor Statistiek is bij uitgeverij Stenfert Kroese in Leiden een losbladige verzameling "Statistische Tabellen en Nomogrammen" verschenen, die een groot aantal van de hier genoemde verdelingen omvat.

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

$$F(x) = \lambda F_1(x) + (1-\lambda)F_2(x) \qquad \text{voor } 0 < \lambda < 1, \tag{7.31}$$

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

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 differentieerbaar zijn. Deze pathologische gevallen zijn voor de praktijk van weinig of geen belang en blijven hier buiten beschouwing.

8. Twee- en meer-dimensionale kansverdelingen.

Definitie 8.1

Twee stochastische grootheden $\underline{x}$ en $\underline{y}$ bezitten een simultane kansverdeling, als voor ieder paar reële getallen x en y de (simultane) verdelingsfunctie

$$F(x,y) \overset{\text{def}}{=} P\left[\underline{x} \leq x \wedge \underline{y} \leq y\right] \tag{8.1}$$

gedefinieerd is.

F(x,y) is zowel als functie van x als van y niet-dalend, terwijl $F(+\infty,+\infty) = \lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x,y) = 1$ en $F(x,-\infty) = F(-\infty,y) = 0$ is voor alle x resp. y.

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. De levertijd van een artikel en het aantal binnenkomende bestellingen per week voor dit artikel. Het aantal klanten dat zich per dag bij een loket meldt en de bedieningstijd per klant.

Definitie 8.2

Een discrete twee-dimensionale kansverdeling is een verdeling, waarbij in de twee-dimensionale ruimte R_2 met x en y als coördinaten een eindig of aftelbaar aantal punten $P_j = (x_j,y_j)$ (j=1,2,...) aanwezig is, waarvoor geldt

$$P\left[\underline{P} = P_j\right] = p_j > 0, \qquad \sum_j p_j = 1 , \tag{8.2}$$

waarin $\underline{P} = (\underline{x},\underline{y})$ een stochastisch punt in R_2 voorstelt. $\underline{P} = P_j$ is dus een verkorte schrijfwijze voor $\underline{x} = x_j \wedge \underline{y} = y_j$.

De bijbehorende verdelingsfunctie is dan een terrasfunctie:

(8.3) $$F(x,y) = \sum_{\substack{j \\ x_j \leq x \\ y_j \leq y}} p_j \ .$$

Voorbeeld 8.1

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

(8.4) $$P\left[\underline{x} = x \wedge \underline{y} = y\right] = p^2 q^{y-2} \qquad \text{voor} \quad y > x.$$

Deze kans is dus voor iedere $x < y$ dezelfde. Dit doet vermoeden dat bij gegeven y alle mogelijke waarden van x dezelfde kans zullen bezitten. Dat dit zo is, kan als volgt bewezen worden (vgl. 7.8):

(8.5) $$P\left[\underline{x} = x \mid \underline{y} = y\right] = \frac{P\left[\underline{x} = x \wedge \underline{y} = y\right]}{P\left[\underline{y} = y\right]} = \frac{p^2 q^{y-2}}{\binom{y-1}{1} p^2 q^{y-2}} = \frac{1}{y-1} \ .$$

Onder de voorwaarde $\underline{y} = y$ volgt $\underline{x}$ dus de discrete homogene verdeling over de getallen $1,2,\ldots,y-1$.

Definitie 8.3

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

(8.6) $$f(x,y) \stackrel{\text{def}}{=} \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

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 (8.6) is

(8.7) $$F(x,y) = \int_{\substack{v \leq x \\ w \leq y}} \int f(v,w)\, dv\, dw.$$

Is in de ruimte R_2 met coördinaten x en y een gebied G gegeven, en is $\underline{P} \equiv (\underline{x},\underline{y})$, dan geldt

$$(8.8) \qquad P\left[\underline{P} \in G\right] = \int\int_G f(x,y)\, dx\, dy \ .$$

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

Voorbeeld 8.2

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

$$(8.9) \quad f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\,\frac{x-\mu_1}{\sigma_1}\,\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}$$

voor $-\infty < x,y < \infty$, waarbij voor de vijf parameters geldt $-\infty < \mu_1,\mu_2 < \infty$, $0 < \sigma_1,\sigma_2 < \infty$, $-1 < \rho < 1$.

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

Marginale verdelingen

Wanneer de simultane verdelingsfunctie F(x,y) van $\underline{x}$ en $\underline{y}$ is gegeven, is de verdelingsfunctie van bijv. $\underline{x}$ hieruit eenvoudig af te leiden, nl.

$$(8.10) \qquad F_{\underline{x}}(x) \overset{\text{def}}{=} P\left[\underline{x} \leq x\right] = P\left[\underline{x} \leq x \wedge \underline{y} < \infty\right] = F(x,\infty) = \lim_{y \to \infty} F(x,y) \ .$$

$F_{\underline{x}}(x)$ wordt veelal aangeduid als de marginale verdelingsfunctie van $\underline{x}$ behorende bij de simultane verdelingsfunctie F(x,y). Dit is echter niet meer dan een andere naam voor een reeds behandeld begrip;

wij zouden dezelfde verdelingsfunctie voor $\underline{x}$ hebben verkregen, indien wij $\underline{y}$ van het begin af niet in onze beschouwing hadden betrokken. Analoog vinden wij voor de (marginale) verdelingsfunctie van $\underline{y}$

$$F_{\underline{y}}(y) \overset{\text{def}}{=} P\left[\underline{y} \leq y\right] = F(\infty, y) = \lim_{x \to \infty} F(x,y) \quad . \tag{8.11}$$

Bezitten $\underline{x}$ en $\underline{y}$ een twee-dimensionale discrete verdeling, dan worden de (marginale) kansen voor $\underline{x}$ verkregen door optelling der kansen $p_j = P\left[\underline{x} = x_j \wedge \underline{y} = y_j\right]$ voor alle waarden van j waarvoor $x_j = x$, ongeacht de waarde van y_j:

$$P\left[\underline{x} = x\right] = \sum_{\substack{j \\ x_j = x}} p_j \quad , \tag{8.12}$$

en analoog

$$P\left[\underline{y} = y\right] = \sum_{\substack{j \\ y_j = y}} p_j \quad . \tag{8.13}$$

Voorbeeld 8.3

De marginale verdeling van $\underline{x}$ behorende bij de door (8.4) gegeven verdeling is volgens (8.12)

$$P\left[\underline{x} = x\right] = \sum_{y > x} p^2 q^{y-2} = p^2 \sum_{y-x > 0} q^{y-x+x-2} =$$

$$= p^2 q^{x-2} \sum_{m > 0} q^m = p^2 q^{x-2} \frac{q}{1-q} = pq^{x-1} \quad .$$

Deze marginale verdeling is dus dezelfde als (7.8) met k = 1. (Was dit ook direct in te zien? De lezer berekene de marginale verdeling van $\underline{y}$).

Bezitten $\underline{x}$ en $\underline{y}$ een twee-dimensionale continue verdeling, dan geldt volgens (8.10) en (8.7) voor de marginale verdelingsfunctie van $\underline{x}$

$$F_{\underline{x}}(x) = \int_{v=-\infty}^{x} \int_{w=-\infty}^{\infty} f(v,w)\, dv\, dw \,. \tag{8.15}$$

De marginale verdelingsdichtheid van $\underline{x}$ is dus

$$f_{\underline{x}}(x) = \frac{dF_{\underline{x}}(x)}{dx} = \int_{-\infty}^{\infty} f(x,w)\, dw \,. \tag{8.16}$$

Voor de marginale verdelingsfunctie en de marginale verdelingsdichtheid van y gelden analoge formules.

Voorbeeld 8.4

De marginale verdeling van $\underline{x}$ behorende bij (8.9) laat zich eenvoudig als volgt berekenen. Stellen wij

$$t = \frac{x-\mu_1}{\sigma_1} \quad \text{en} \quad u = \frac{y-\mu_2}{\sigma_2} \,, \tag{8.17}$$

dan is $dx = \sigma_1 dt$ en $dy = \sigma_2 du$.
Substitueren wij dit in (8.9) en stelt :: voor "op een constante factor na gelijk aan", dan is dus

$$f(x,y) :: e^{-\frac{1}{2(1-\rho^2)}(t^2-2\rho tu+u^2)} \,. \tag{8.18}$$

Volgens (8.16) geldt dus voor de marginale verdelingsdichtheid $f_{\underline{x}}(x)$ van $\underline{x}$

$$\begin{aligned} f_{\underline{x}}(x) &:: e^{-\frac{t^2}{2(1-\rho^2)}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}\{(u^2-2\rho tu+\rho^2t^2)-\rho^2t^2\}}\, du = \\ &= e^{-\frac{1}{2}t^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}(u-\rho t)^2}\, du \,. \end{aligned} \tag{8.19}$$

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

$$f_{\underline{x}}(x) :: e^{-\frac{1}{2}t^2} = e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}} . \tag{8.20}$$

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 (7.14), die uiteraard ook voor marginale verdelingen geldt. Uit formule (7.18) blijkt dat het resultaat moet zijn

$$f_{\underline{x}}(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}} . \tag{8.21}$$

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

Voorwaardelijke verdelingen

In het voorgaande berekenden wij uit de simultane verdeling van $\underline{x}$ en $\underline{y}$ de (marginale) verdeling van $\underline{x}$, d.w.z. de verdeling van $\underline{x}$ ongeacht de bij $\underline{x}$ optredende waarde van $\underline{y}$. Thans beschouwen wij de voorwaardelijke verdeling van $\underline{x}$ onder de voorwaarde dat $\underline{y}$ een gegeven waarde y aanneemt.

Voor discrete verdelingen levert het vinden van deze voorwaardelijke verdeling geen moeilijkheden op:

$$P\left[\underline{x} = x \mid \underline{y} = y\right] = \frac{P\left[\underline{x} = x \wedge \underline{y} = y\right]}{P\left[\underline{y} = y\right]} \quad \text{mits} \quad P\left[\underline{y} = y\right] > 0 . \tag{8.22}$$

Deze voorwaardelijke verdeling van $\underline{x}$ is in het algemeen een andere dan de onvoorwaardelijke of marginale verdeling. Dit wordt duidelijk indien wij nogmaals de simultane verdeling (8.4) beschouwen.

In figuur 8.1 zijn de mogelijke paren waarden voor $\underline{x}$ en $\underline{y}$ aangegeven.

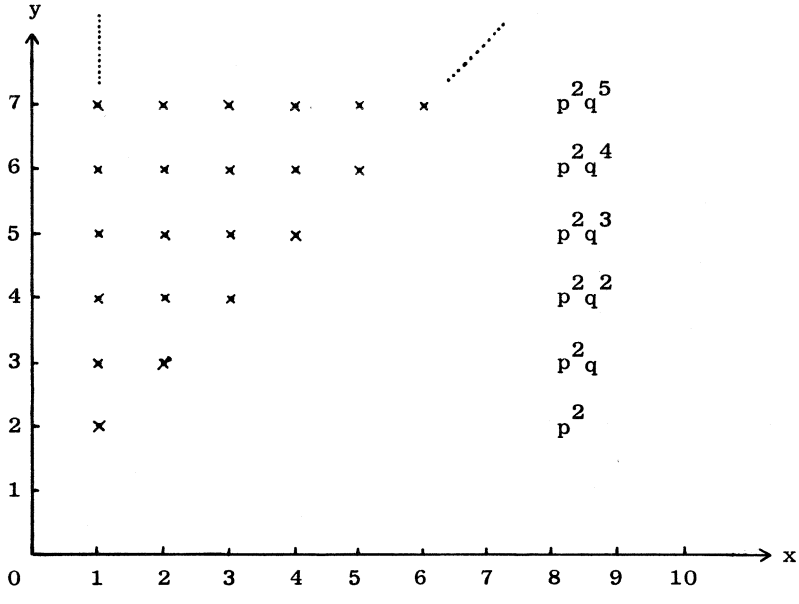


fig. 8.1

Bij iedere waarde van y is volgens formule (8.4) de kans aangegeven op ieder der (x,y)-combinaties, die bij deze waarde van y kunnen voorkomen. De marginale kansen $P[\underline{x} = x]$ worden nu verkregen door steeds alle kansen, behorende bij punten op de vertikale rechte door het punt (x,0), op te tellen. Dit zijn immers juist de punten waar $\underline{x}$ de waarde x aanneemt. Men drukt dit vaak uit door te zeggen, dat de marginale verdeling van $\underline{x}$ uit de simultane verdeling wordt verkregen door de waarschijnlijkheden op de x-as te projekteren. Op deze wijze ontstaat de verdeling uit voorbeeld 8.3.

Voor het bepalen van de voorwaardelijke kansen $P[\underline{x} = x \mid \underline{y} = y]$ behoeven wij slechts de punten op de horizontale rechte door het punt y te beschouwen. Om de voorwaardelijke verdeling te verkrijgen behoeven de bijbehorende kansen slechts door $P[\underline{y} = y]$ (d.w.z. de som der kansen op deze rechte) te worden gedeeld, waardoor hun som gelijk aan 1 wordt. Op deze wijze ontstaat de verdeling gegeven in formule (8.5). Dat de

marginale en de voorwaardelijke verdelingen van $\underline{x}$ in dit geval verschillend zijn, blijkt overigens reeds uit het feit dat $\underline{x}$ in het marginale en het voorwaardelijke geval verschillende waardenverzamelingen doorloopt.

In het continue geval doet zich de complicatie voor dat de voorwaarde, waaronder men werkt, zelf een kans 0 bezit, zodat de definitie (8.22) niet zonder meer bruikbaar is. Bezitten $\underline{x}$ en $\underline{y}$ een continue twee-dimensionale verdeling, dan is $P[\underline{y} = y] = 0$ voor iedere y , zodat een voorwaardelijke kans van de vorm

$$P[\underline{x} \leq x \mid \underline{y} = y] \tag{8.23}$$

vooralsnog niet gedefinieerd is. Deze kans wordt nu gedefinieerd als

$$F(x \mid y) = P[\underline{x} \leq x \mid \underline{y} = y] = \lim_{\Delta y \downarrow 0} P[\underline{x} \leq x \mid y \leq \underline{y} \leq y+\Delta y] \ . \tag{8.24}$$

Volgens de definitie van een voorwaardelijke kans en (8.8) is dit gelijk aan

$$\lim_{\Delta y \downarrow 0} \left\{ \int_{-\infty}^{x} dv \int_{y}^{y+\Delta y} dw\ f(v,w) \Big/ \int_{-\infty}^{\infty} dv \int_{y}^{y+\Delta y} dw\ f(v,w) \right\}, \tag{8.25}$$

en hierin is, na deling van teller en noemer door Δy,

$$\begin{aligned} &\lim_{\Delta y \downarrow 0} \frac{1}{\Delta y} \int_{-\infty}^{x} dv \int_{y}^{y+\Delta y} dw\ f(v,w) = \int_{-\infty}^{x} f(v,y)\ dv = \\ &= \frac{\partial}{\partial y} \int_{-\infty}^{x} \int_{-\infty}^{y} f(v,w)\ dv\ dw = \frac{\partial}{\partial y} F(x,y) \ , \end{aligned} \tag{8.26}$$

terwijl voor de noemer volgens (8.15) geldt

$$\begin{aligned} &\lim_{\Delta y \downarrow 0} \frac{1}{\Delta y} \int_{-\infty}^{\infty} dv \int_{y}^{y+\Delta y} dw\ f(v,w) = \int_{-\infty}^{\infty} f(v,y)\ dv = \\ &= \frac{d}{dy} \int_{-\infty}^{\infty} dv \int_{-\infty}^{y} dw\ f(v,w) = \frac{d}{dy} F_{\underline{y}}(y) = f_{\underline{y}}(y) \ , \end{aligned} \tag{8.27}$$

waar $F_{\underline{y}}$ en $f_{\underline{y}}$ de marginale verdelingsfunctie resp. verdelingsdichtheid van $\underline{y}$ voorstellen. Derhalve is, indien $f_{\underline{y}}(y) \neq 0$,

$$F(x \mid y) = \frac{\frac{\partial}{\partial y} F(x,y)}{f_{\underline{y}}(y)} , \tag{8.28}$$

zodat voor de voorwaardelijke dichtheid $f(x \mid y)$ geldt

$$f(x \mid y) = \frac{\partial}{\partial x} F(x \mid y) = \frac{f(x,y)}{f_{\underline{y}}(y)} . \tag{8.29}$$

De lezer controlere of dit inderdaad een kansdichtheid is, d.w.z. of (8.29) niet-negatief is en bij integratie over x de waarde 1 oplevert.

Voorbeeld 8.5

Wij berekenen nu de voorwaardelijke verdelingsdichtheid van de twee-dimensionale normale verdeling (8.9). Wij substitueren in (8.29) het analogon voor $\underline{y}$ van (8.21) en in de teller (8.9). Wanneer :: weer betekent:"is op een constante factor na gelijk aan", vindt men met de substitutie van (8.17)

$$f(x \mid y) :: e^{-\frac{1}{2(1-\rho^2)}(t^2-2\rho tu+u^2)+\frac{1}{2}u^2} =$$

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

$$= e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1} - \rho\frac{y-\mu_2}{\sigma_2}\right)^2} = \tag{8.30}$$

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

waarin nu y wegens de voorwaarde $\underline{y} = y$ als een gegeven constante

optreedt. Derhalve volgt uit (7.18)

$$(8.31) \qquad f(x \mid y) = \frac{1}{\sigma_1\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{\left[x - \{\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)+\mu_1\}\right]^2}{2(1-\rho^2)\sigma_1^2}} .$$

Onder de voorwaarde $\underline{y} = y$ bezit $\underline{x}$ dus een $N\left(\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)+\mu_1;\ \sigma_1\sqrt{1-\rho^2}\right)$-verdeling. Ook hier is de voorwaardelijke verdeling in het algemeen dus niet gelijk aan de marginale (vgl. (8.21)).

In sommige praktijksituaties zijn van een twee-dimensionale verdeling $F(x,y)$ de marginale verdeling $F_{\underline{x}}(x)$ en de voorwaardelijke verdeling $F(y \mid x)$ gegeven, terwijl de marginale verdeling $F_{\underline{y}}(y)$ berekend moet worden.

Voor een discrete verdeling, waarin $\underline{x}$ de waarden $1,\ldots,m$ en $\underline{y}$ de waarden $1,\ldots,n$ kan aannemen, geschiedt een dergelijke berekening met de formule (vgl. (3.17))

$$(8.32) \qquad P[\underline{y}=j] = \sum_{i=1}^{m} P[\underline{x}=i \wedge \underline{y}=j] = \sum_{i=1}^{m} P[\underline{x}=i].P[\underline{y}=j \mid \underline{x}=i] .$$

Bij continue verdelingen kan de gezochte marginale verdelingsdichtheid gevonden worden via (vgl. (8.16) en (8.29))

$$(8.33) \qquad f_{\underline{y}}(y) = \int_{-\infty}^{\infty} f(v,y)\,dv = \int_{-\infty}^{\infty} f_{\underline{x}}(v) f(y \mid v)\,dv .$$

Een voorbeeld van een dergelijke berekening volgt hier voor het geval de ene marginale verdeling discreet is en de andere continu.

Voorbeeld 8.6

Een leverancier verkoopt een artikel uit voorraad aan zijn klanten, terwijl hij het artikel zelf op order inkoopt bij een fabriek. De frequenties van de levertijden van de fabriek naar de leverancier sluiten goed aan bij een Erlang-verdeling. De verdelingsfunctie van het aantal

klanten in een periode van T maanden is een Poisson-verdeling met parameter μT. Hoe groot is de kans dat er tijdens de levertijd van de volgende bestelling hoogstens n klanten komen?

Oplossing

De verdelingsfunctie van de levertijd is (zie (7.26))

$$F(T) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_0^T e^{-\lambda t}\, t^{\alpha-1}\, dt \qquad (\alpha \text{ geheel positief}) \tag{8.34}$$

en de verdelingsfunctie van het aantal klanten dat tijdens een vaste levertijd van T maanden komt, is (zie (7.9))

$$P\left[\underline{k} \leq n \mid \underline{t} = T\right] = \sum_{k=0}^{n} \frac{(\mu T)^k\, e^{-\mu T}}{k!} \,. \tag{8.35}$$

Om de verdelingsfunctie van het aantal klanten in de komende levertijd te vinden, moeten wij volgens (8.33) de in (8.35) opgegeven kans integreren over T; deze verdelingsfunctie is dus (vgl. (7.26))

$$\begin{aligned} P\left[\underline{k} \leq n\right] &= \int_0^{\infty} \left(\sum_{k=0}^{n} \frac{(\mu t)^k\, e^{-\mu t}}{k!} \right) f(t)\, dt = \\ &= \sum_{k=0}^{n} \int_0^{\infty} \frac{(\mu t)^k\, e^{-\mu t}}{k!} \cdot \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\, e^{-\lambda t}\, t^{\alpha-1}\, dt = \\ &= \sum_{k=0}^{n} \frac{\mu^k \lambda^{\alpha}}{k!\,\Gamma(\alpha)} \int_0^{\infty} e^{-(\lambda+\mu)t}\, t^{\alpha+k-1}\, dt = \\ &= \sum_{k=0}^{n} \frac{\mu^k \lambda^{\alpha}}{k!\,\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+k)}{(\lambda+\mu)^{\alpha+k}} = \\ &= \sum_{k=0}^{n} \binom{\alpha+k-1}{\alpha-1} \left(\frac{\lambda}{\lambda+\mu}\right)^{\alpha} \left(\frac{\mu}{\lambda+\mu}\right)^{k} . \end{aligned} \tag{8.36}$$

Men kan nu verder nagaan (vgl. (7.8) en de voetnoot op bladzijde 92), dat (8.36) een negatief binomiale verdeling voorstelt met parameters α en $\frac{\lambda}{\lambda+\mu}$. In een geval, waarin de levertijden een Erlang-verdeling volgen en het aantal klanten in een periode van T maanden

een Poisson-verdeling volgt met parameter μT, is de onvoorwaardelijke verdeling van het aantal klanten tijdens de levertijd dus een negatief binomiale verdeling.

Alle in deze paragraaf ingevoerde begrippen kunnen uitgebreid worden tot ruimten van n dimensies; d.w.z. bij n stochastische variabelen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ behoort een n-dimensionale verdelingsfunctie

$$F(x_1, x_2, \ldots, x_n) = P\left[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 \leq x_2 \wedge \ldots \wedge \underline{x}_n \leq x_n\right] . \tag{8.37}$$

Wij zullen dit niet verder bespreken en volstaan met het volgende voorbeeld.

Voorbeeld 8.7

Laat $\underline{x}_i$ $(i=1,2,\ldots,n)$ het aantal malen kruis voorstellen bij de i^{de} 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

$$P\left[\underline{x}_i = 1\right] = p, \qquad P\left[\underline{x}_i = 0\right] = q \qquad (i=1,2,\ldots,n) . \tag{8.38}$$

Zoals intuïtief eenvoudig is in te zien en bovendien in stelling 9.2 bewezen zal worden, geldt wegens de onafhankelijkheid

$$P\left[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2 \wedge \ldots \wedge \underline{x}_n = x_n\right] = p^{\sum_i x_i} q^{n-\sum_i x_i} , \tag{8.39}$$

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

$$\underline{x} = \sum_{i=1}^{n} \underline{x}_i , \tag{8.40}$$

dan is $\underline{x}$ het aantal malen kruis bij n worpen, waarvoor dus (7.6), 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 van pas komen.

9. Onafhankelijkheid van stochastische grootheden.

In paragraaf 4 is het begrip stochastische onafhankelijkheid gedefinieerd voor gebeurtenissen door middel van de produktformules (4.7) en (4.11). Aequivalent hiermee is, dat voorwaardelijke en onvoorwaardelijke kansen gelijk zijn (vgl. (4.8)). Wij sluiten hier bij die definities aan, waarbij nu de beschouwde gebeurtenissen zijn:

$$\underline{x}_1 \leq x_1,\ \underline{x}_2 \leq x_2,\ \text{enz.}\ .$$

Definitie 9.1

Een aantal stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ is onafhankelijk verdeeld (stochastisch onafhankelijk), indien zij een simultane kansverdeling bezitten waarvan de verdelingsfunctie F voldoet aan

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1).F_2(x_2)\ldots F_n(x_n)\ , \tag{9.1}$$

waarin F_i (i=1,2,...,n) de marginale verdelingsfunctie van $\underline{x}_i$ is. Vergelijking (9.1) is identiek met

$$P\left[\underline{x}_1 \leq x_1 \wedge \underline{x}_2 \leq x_2 \wedge \ldots \wedge \underline{x}_n \leq x_n\right] = \prod_{i=1}^{n} P\left[\underline{x}_i \leq x_i\right]. \tag{9.2}$$

Voor de eenvoud beschouwen wij verder in het bijzonder het geval van een twee-dimensionale verdeling van stochastische grootheden $\underline{x}$ en $\underline{y}$. De uitbreiding tot meer dan 2 dimensies is gemakkelijk aan te geven en de bewijzen verlopen analoog. De formules (9.1) en (9.2) gaan voor n = 2 over in

$$F(x,y) = F_{\underline{x}}(x).F_{\underline{y}}(y) \tag{9.1'}$$

en

$$P\left[\underline{x} \leq x \wedge \underline{y} \leq y\right] = P\left[\underline{x} \leq x\right].P\left[\underline{y} \leq y\right]\ . \tag{9.2'}$$

Stelling 9.1

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

$$P\left[\underline{x} \leq x \mid \underline{y} \leq y\right] = P\left[\underline{x} \leq x\right], \text{ indien } P\left[\underline{y} \leq y\right] > 0 \text{ is.} \tag{9.3}$$

Bewijs:

Volgens de definitie van een voorwaardelijke kans, tezamen met (9.1'), geldt

$$P\left[\underline{x} \leq x \mid \underline{y} \leq y\right] = \frac{F(x,y)}{F_{\underline{y}}(y)} = \frac{F_{\underline{x}}(x).F_{\underline{y}}(y)}{F_{\underline{y}}(y)} = F_{\underline{x}}(x) = P\left[\underline{x} \leq x\right] .$$

Stelling 9.2

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

$$P\left[\underline{x} \leq x \wedge \underline{y} = y\right] = P\left[\underline{x} \leq x\right].P\left[\underline{y} = y\right] , \tag{9.4}$$

$$P\left[\underline{x} = x \wedge \underline{y} \leq y\right] = P\left[\underline{x} = x\right].P\left[\underline{y} \leq y\right] , \tag{9.5}$$

$$P\left[\underline{x} = x \wedge \underline{y} = y\right] = P\left[\underline{x} = x\right].P\left[\underline{y} = y\right] . \tag{9.6}$$

Bewijs:

Is y' < y de grootste waarde < y, die door $\underline{y}$ aangenomen kan worden, dan zijn de eventualiteiten $\left[\underline{y} < y\right]$ en $\left[\underline{y} \leq y'\right]$ dezelfde en daar (9.2') voor alle mogelijke waarden van x en y geldt is dus tevens

$$P\left[\underline{x} \leq x \wedge \underline{y} < y\right] = P\left[\underline{x} \leq x\right].P\left[\underline{y} < y\right] . \tag{9.7}$$

Is er geen dergelijke waarde y', dan zijn linker- en rechterlid van (9.7) beide = 0, zodat de geldigheid behouden blijft *). Wegens de exclusiviteit van de gebeurtenissen en het gegeven (9.2') is

$$\begin{aligned} &P\left[\underline{x} \leq x \wedge \underline{y} < y\right] + P\left[\underline{x} \leq x \wedge \underline{y} = y\right] = P\left[\underline{x} \leq x \wedge \underline{y} \leq y\right] = \\ &= P\left[\underline{x} \leq x\right].P\left[\underline{y} \leq y\right] = P\left[\underline{x} \leq x\right].P\left[\underline{y} < y\right] + P\left[\underline{x} \leq x\right].P\left[\underline{y} = y\right]. \end{aligned} \tag{9.8}$$

Van de buitenste leden zijn de eerste termen gelijk volgens (9.7). Dus zijn ook de tweede termen gelijk en is (9.4) bewezen. Het bewijs van

---

*) Strikt genomen moet het geval waarin y de bovenste grens is van een oneindige rij waarden kleiner dan y die $\underline{y}$ kan aannemen, afzonderlijk beschouwd worden. Hierbij is axioma V nodig (zie paragraaf 4, blz.27).

(9.5) verloopt net zo. Voor (9.6) past men een analoge redenering toe op $P\left[\underline{x} = x \wedge \underline{y} < y\right] + P\left[\underline{x} = x \wedge \underline{y} = y\right]$.

Stelling 9.3

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

$$f(x,y) = f_{\underline{x}}(x).f_{\underline{y}}(y) \tag{9.9}$$

waarin f de simultane en $f_{\underline{x}}$ en $f_{\underline{y}}$ de marginale verdelingsdichtheden voorstellen.

Bewijs:

Deze stelling volgt onmiddellijk uit definitie 8.3:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 \{F_{\underline{x}}(x).F_{\underline{y}}(y)\}}{\partial x \partial y} = f_{\underline{x}}(x).f_{\underline{y}}(y).$$

Opmerkingen

In paragraaf 4 is opgemerkt dat de bijzondere produktregel (4.12) niet voldoende is om stochastische onafhankelijkheid te garanderen. Daartoe was (4.11) nodig. In definitie 9.1 ligt echter een analoge relatie als (4.11) opgesloten. Immers vult men in (9.1) bijv. $x_n = \infty$ in, dan gaat deze relatie over in dezelfde relatie, maar nu voor de eerste n-1 stochastische grootheden. Wij hebben hier dus aan (9.1) genoeg voor de definitie van stochastische onafhankelijkheid. Anderzijds kan men door een tegenvoorbeeld aantonen, dat voor meer dan twee stochastische grootheden paarsgewijze onafhankelijkheid niet voldoende is voor volledige onafhankelijkheid.

De relaties (9.6) en (9.9) (voor n dimensies opgeschreven) kunnen ook dienen om stochastische onafhankelijkheid te definiëren. Dan kan (9.1) 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 kunnen worden.

Uit de stellingen 9.2 en 9.3 volgt, dat voor onafhankelijke grootheden geldt

(9.10) $P[\underline{x} = x \mid \underline{y} = y] = P[\underline{x} = x]$ (discontinue geval)

en (volgens (8.18))

(9.11) $f(x \mid y) = f_{\underline{x}}(x)$ (continue geval).

Of algemeen gezegd, 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.

Voorbeelden

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 hierboven zijn afgeleid, gelden. Als voorbeeld van de vereenvoudigingen die optreden, beschouwen wij de twee-dimensionale normale verdeling (8.9). De marginale verdelingen van $\underline{x}$ en $\underline{y}$ zijn, volgens (8.21),

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

Zijn $\underline{x}$ en $\underline{y}$ stochastisch onafhankelijk, dan geldt dus volgens (9.9)

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

hetgeen overeenkomt met (8.9) wanneer daarin $\rho = 0$ wordt gesteld. 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 (8.21) en (8.31), 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 worpresultaten 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.

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 (9.5)

$$(9.13) \qquad P\left[\underline{x}_1 = x_1 \wedge \underline{x}_2 = x_2 \wedge \ldots \wedge \underline{x}_n = x_n\right] = \prod_{i=1}^{n} P\left[\underline{x}_i = x_i\right] = 6^{-n}$$

en deze kans is dus voor alle mogelijke n-tallen uitkomsten dezelfde.

De onderstelling dat de experimenten stochastisch onafhankelijk zijn is niet alleen bij worpen met dobbelstenen of munten vaak gerechtvaardigd. Wanneer men steekproeven zorgvuldig kiest (bijv. met behulp van aselecte getallen, vgl. voorbeeld 5.3), dan kan men veelal uitgaan van de onderstelling dat de meetresultaten door stochastisch onafhankelijke grootheden voorgesteld kunnen worden. Terloops merken wij hierbij op dat men een dergelijke onderstelling niet klakkeloos behoeft te aanvaarden, doch experimenteel kan onderzoeken. De statistiek beschikt over de middelen om op grond van waarnemingsresultaten de

gemaakte onderstelling op zijn juistheid te toetsen.

In sommige gevallen is het mogelijk door een geschikte transformatie stochastisch afhankelijke grootheden te vervangen door stochastisch onafhankelijke grootheden. Wij besluiten deze paragraaf met een illustratie van een dergelijke transformatie.

Beschouw 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

$$P\left[\underline{x}_1 < \underline{x}_2 < \ldots\right] = 1 \quad , \tag{9.14}$$

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

$$\underline{y}_1 = \underline{x}_1, \quad \underline{y}_2 = \underline{x}_2 - \underline{x}_1, \quad \underline{y}_3 = \underline{x}_3 - \underline{x}_2, \text{ enz.}, \tag{9.15}$$

dan zijn deze grootheden weer onafhankelijk en zij bezitten alle de dezelfde negatief binomiale verdeling (formule (7.8), 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 (7.8) van toepassing is. Bij verschillende berekeningen omtrent de kansverdeling van $\underline{x}_k$ (met $k > 1$) is het dan ook eenvoudiger om van de $\underline{y}_i$ uit te gaan.

10. Mathematische verwachting en momenten.

Het begrip mathematische verwachting dat in deze paragraaf 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 één-dimensionale geval.

Definitie 10.1

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

$$p_i = P\left[\underline{x} = x_i\right] \qquad \left(\sum_i p_i = 1\right) \tag{10.1}$$

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

$$\mathfrak{E}\,\phi(\underline{x}) \stackrel{\text{def}}{=} \sum_i p_i \phi(x_i) \,. \tag{10.2}$$

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

$$\mathfrak{E}\,\phi(\underline{x}) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} \phi(x) f(x)\, dx \,. \tag{10.3}$$

Opmerking

De functie $\phi(\underline{x})$ van $\underline{x}$ is, daar $\underline{x}$ stochastisch is, zelf in het algemeen ook een stochastische grootheid. De verwachting *) $\mathfrak{E}$ is een "operator", die deze stochastische grootheid in een getal omzet. Immers de rechterleden van (10.2) en (10.3) zijn getallen.

Voor gemengde kansverdelingen (7.31) is $\mathfrak{E}\,\phi(\underline{x})$ een lineaire combinatie met coëfficiënten λ en $(1-\lambda)$ van de rechterleden van (10.2) en (10.3).

---

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

Definitie 10.2

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

(10.4) $$\mu_k \stackrel{\text{def}}{=} \mathcal{E}\,\underline{x}^k \ .$$

Opmerkingen

Voor iedere kansverdeling geldt $\mu_0 = 1$. Stellen wij de kansverdeling van $\underline{x}$ aanschouwelijk voor als een massaverdeling over de 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.

Voorbeeld 10.1

Indien $\underline{x}$ de kansen p resp. $q = 1-p$ bezit op de waarden 1 resp. 0 (vgl. (7.5) en fig. 7.1), dan geldt

(10.5) $$\mu_k = p \qquad \text{voor } k = 1,2,\ldots \ .$$

Immers dan is $\mu_k = p.1^k + q.0^k = p$.

Voorbeeld 10.2

Indien $\underline{x}$ een negatief binomiale verdeling bezit met $k = 1$ (vgl. (7.8)), dan is $P\left[\underline{x} = x\right] = pq^{x-1}$ ($x=1,2,\ldots$) en dus

$$\mathcal{E}\,\underline{x} = \sum_{x=1}^{\infty} xP\left[\underline{x} = x\right] = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty} xq^{x-1} \ .$$

Nu is

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

en nemen wij links en rechts de afgeleide naar q, dan vinden wij

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

Derhalve is

(10.6) $$\mu = \mathcal{E}\underline{x} = \frac{1}{p} \quad .$$

Voorbeeld 10.3

Is $\underline{x}$ N(μ,σ)-verdeeld (vgl. (7.18)), dan is

$$\mathcal{E}\underline{x} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx \quad .$$

Substitueer $u = \frac{x-\mu}{\sigma}$ (dus $x = u\sigma + \mu$ en $dx = \sigma du$), dan geldt

$$\mathcal{E}\underline{x} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (u\sigma+\mu)e^{-\frac{1}{2}u^2} du =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} ue^{-\frac{1}{2}u^2} du + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du \quad .$$

De eerste term valt weg wegens de antisymmetrie van de functie onder het integraalteken en de tweede term is wegens (7.14) gelijk aan μ. Dus is

(10.7) $$\mathcal{E}\underline{x} = \mu \quad .$$

Wij merken op dat de parameter μ van de normale verdeling dus gelijk is aan het eerste moment van die verdeling, of zoals men gemakshalve vaak zegt, gelijk is aan "de verwachting" van de normale verdeling. Bij wijze van illustratie berekenen wij ook nog het tweede moment $\mathcal{E}\underline{x}^2$. Dit kan als volgt gedaan worden:

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

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

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

daar nu de middelste term wegvalt. Nu is met partieel integreren

$$\int_{-\infty}^{+\infty} u^2 e^{-\frac{1}{2}u^2} du = - \int_{-\infty}^{+\infty} u \, d(e^{-\frac{1}{2}u^2}) =$$

$$= \left[- ue^{-\frac{1}{2}u^2} \right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{1}{2}u^2} du = \sqrt{2\pi} ,$$

hetgeen na invullen geeft

$$\mu_2 = \mathcal{E}\, \underline{x}^2 = \sigma^2 + \mu^2 . \tag{10.8}$$

Voorbeeld 10.4

Bezit $\underline{x}$ een gamma-verdeling met parameters α en λ (vgl. (7.25)), dan is

$$\begin{cases} \mathcal{E}\, \underline{x} = \int_0^\infty x \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda x} x^\alpha dx = \\ = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} \int_0^\infty e^{-\lambda x} x^\alpha dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{\alpha}{\lambda} . \end{cases} \tag{10.9}$$

Verder is

$$\begin{cases} \mathcal{E}\, \underline{x}^2 = \int_0^\infty x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda x} x^{\alpha+1} dx = \\ = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} = \frac{\alpha(\alpha+1)}{\lambda^2} . \end{cases} \tag{10.10}$$

Definitie 10.3

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

$$(10.11) \qquad \tilde{\mu}_k = \mu_k(\tilde{\underline{x}}) = \mathcal{E}\,\tilde{\underline{x}}^k = \mathcal{E}(\underline{x}-\mu)^k = \mathcal{E}(\underline{x} - \mathcal{E}\,\underline{x})^k \ .$$

Opmerkingen

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

$$(10.12) \qquad \sigma^2(\underline{x}) = \mathcal{E}(\underline{x}-\mu)^2 = \mathcal{E}\,\tilde{\underline{x}}^2 = \tilde{\mu}_2 \ .$$

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

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. Een tabel van de in paragraaf 7 genoemde verdelingen met hun verwachtingen en varianties vindt men aan het slot van dit deel (zie blz. 116).

Het geval van meer-dimensionale verdelingen behandelen wij in hoofdzaak voor twee dimensies. Voor meer dan twee verloopt het geheel analoog.

Definitie 10.4

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 $\phi(x,y)$ een functie van x en y, dan is de mathematische verwachting van $\phi(x,y)$

$$(10.13) \qquad \mathcal{E}\,\phi(\underline{x},\underline{y}) \overset{\text{def}}{=} \sum_j p_j\,\phi(x_j,y_j) \ .$$

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

$$\mathcal{E}\,\phi(\underline{x},\underline{y}) \overset{\text{def}}{=} \int_{-\infty}^{+\infty}\int \phi(x,y)f(x,y)\,dxdy \quad . \tag{10.14}$$

Definitie 10.5

De voorwaardelijke verwachting van $\phi(\underline{x},\underline{y})$, onder de voorwaarde $\underline{y} = y$, is de verwachting van $\phi(\underline{x},y)$ met betrekking tot de voorwaardelijke verdeling van $\underline{x}$ onder dezelfde voorwaarde.

Notatie: $\mathcal{E}(\phi(\underline{x},\underline{y}) \mid \underline{y}=y)$ of $\mathcal{E}(\phi(\underline{x},\underline{y}) \mid y)$ of kortweg $\mathcal{E}\,\phi(\underline{x},y)$.

Dus volgens (8.29) geldt voor het continue geval

$$\mathcal{E}\,\phi(\underline{x},y) = \int_{-\infty}^{+\infty} \phi(x,y)\,\frac{f(x,y)}{f_{\underline{y}}(y)}\,dx \quad . \tag{10.15}$$

Voorbeeld 10.5

Volgens (8.31) is de voorwaardelijke verdeling van $\underline{x}$, onder de voorwaarde $\underline{y} = y$, als $\underline{x}$ en $\underline{y}$ een simultane normale verdeling (8.9) bezitten, ook normaal. Uit (8.31) en (10.7) volgt nu direct

$$\mathcal{E}(\underline{x} \mid y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2) \quad . \tag{10.16}$$

Voorbeeld 10.6

Hoe groot is de verwachting van het aantal klanten in voorbeeld 8.6 voor het geval de levertijd T maanden bedraagt?

Oplossing

Gevraagd wordt de verwachting van het aantal klanten $\underline{k}$ dat binnenkomt in een periode van T maanden. De voorwaardelijke kansverdeling van $\underline{k}$ is volgens (8.35)

$$P\left[\underline{k} \leq n \mid \underline{t} = T\right] = \sum_{k=0}^{n} \frac{(\mu T)^k\, e^{-\mu T}}{k!} \quad .$$

De voorwaardelijke verwachting is dus

$$\mathcal{E}(\underline{k} \mid T) = \sum_{k=0}^{\infty} k \frac{(\mu T)^k e^{-\mu T}}{k!} = \mu T \sum_{k=1}^{\infty} \frac{(\mu T)^{k-1} e^{-\mu T}}{(k-1)!} =$$

(10.17)

$$= \mu T \sum_{j=0}^{\infty} \frac{(\mu T)^j e^{-\mu T}}{j!} = \mu T \ .$$

De verwachting van het aantal klanten is dus evenredig met de lengte van het tijdsinterval.

Uit (10.17) volgt dat het eerste moment van een Poisson-verdeling met parameter μT gelijk is aan μT. Substitueert men voor μT de gebruikelijke parameter λ, dan is tevens bewezen dat voor de Poisson-verdeling (7.9) geldt

(10.18) $$\mu_1 = \mathcal{E}\underline{x} = \lambda \ .$$

Stelling 10.1

Bezitten $\underline{x}$ en $\underline{y}$ een simultane verdeling, en is ϕ een functie van x alleen: $\phi(x)$, dan is de verwachting van $\phi(\underline{x})$ gelijk aan de verwachting van ϕ over de marginale verdeling van $\underline{x}$.

Bewijs:

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

Volgens (10.14) is

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

$$= \int_{-\infty}^{+\infty} \phi(x)\, dx \int_{-\infty}^{+\infty} f(x,y)\, dy = \int_{-\infty}^{+\infty} \phi(x) f_{\underline{x}}(x)\, dx \ ,$$

waarin $f_{\underline{x}}(x)$ (vgl. (8.16)) de marginale verdelingsdichtheid van $\underline{x}$ is.

Voorbeeld 10.7

Voor de twee-dimensionale normale verdeling, die gegeven wordt door (8.9), is de marginale verdeling van $\underline{x}$ volgens (8.21) een normale met parameters μ_1 en σ_1. Volgens (10.7) en stelling 10.1 is dus voor deze verdeling

(10.19) $$\mathcal{E}\,\underline{x} = \mu_1 \,, \qquad \mathcal{E}\,\underline{x}^2 = \sigma_1^2 + \mu_1^2$$

en analoog voor $\underline{y}$.

Stelling 10.2

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

(10.20) $$\mathcal{E}\,\phi(\underline{x})\psi(\underline{y}) = \mathcal{E}\,\phi(\underline{x})\ \mathcal{E}\,\psi(\underline{y}) \,.$$

Bewijs (voor het continue geval):

Met behulp van (9.9) en stelling 10.1 verkrijgen wij

$$\mathcal{E}\,\phi(\underline{x})\psi(\underline{y}) = \iint_{-\infty}^{+\infty} \phi(x)\psi(y) f(x,y)\ dxdy =$$

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

Voorbeeld 10.8

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

(10.21) $$\mathcal{E}\,\underline{xy} = \mathcal{E}\underline{x}\ \mathcal{E}\underline{y} \,.$$

Door middel van een tegenvoorbeeld kan men aantonen dat het omgekeerde niet het geval is.

Stelling 10.3

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

(10.22) $$\mathcal{E}\{a\phi(\underline{x},\underline{y}) + b\psi(\underline{x},\underline{y}) + c\} = a\,\mathcal{E}\phi(\underline{x},\underline{y}) + b\,\mathcal{E}\psi(\underline{x},\underline{y}) + c \,,$$

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 (10.22) volgt onmiddellijk uit definitie 10.1.

Toepassing

Deze stelling maakt het mogelijk de gereduceerde momenten in de gewone uit te drukken. Bijv. (vgl. (10.12))

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

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

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

$$\sigma^2 = \mu_2 - \mu^2 \quad \text{of} \quad \sigma^2(\underline{x}) = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 . \tag{10.23}$$

Passen wij deze formule op de variantie van de normale verdeling toe, dan vinden wij met behulp van (10.7) en (10.8)

$$\sigma^2(\underline{x}) = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2 . \tag{10.24}$$

De parameter σ is dus gelijk aan de wortel van de variantie, dat is de standaardafwijking. Daar bovendien de parameter μ, zoals wij zagen, gelijk is aan de verwachting $\mathcal{E}\underline{x}$, wordt een normale verdeling door verwachting en variantie dus volledig bepaald.

Als tweede illustratie passen wij formule (10.23) toe om de variantie van een gamma-verdeling te berekenen; wij krijgen met behulp van (10.9) en (10.10)

$$\sigma^2(\underline{x}) = \mathcal{E}\underline{x}^2 - (\mathcal{E}\underline{x})^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2} . \tag{10.25}$$

Opmerking

Uit (10.23) volgt direct de ongelijkheid

$$\mu_2 \geq \mu^2 \quad \text{of} \quad \mathcal{E}\underline{x}^2 \geq (\mathcal{E}\underline{x})^2 , \tag{10.26}$$

waarbij het gelijkteken, dus de eigenschap $\sigma(\underline{x}) = 0$, dan en alleen dan geldt als $\underline{x}$ met kans 1 een bepaalde waarde aanneemt.

Stelling 10.4 (Optelregel voor varianties)

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

$$\sigma^2(\underline{x}+\underline{y}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y}) . \tag{10.27}$$

De stelling geldt ook voor meer dan twee grootheden.

Bewijs:

Volgens (10.22) 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 (10.22),

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

daar, wegens de onafhankelijkheid, (10.21) toegepast kan worden.
Volgens (10.23) is

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

Substitueren wij de bovenstaande resultaten hierin, dan geldt

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

weer volgens (10.23).

Opmerking

De voorwaarde van onafhankelijkheid is voldoende, doch niet noodzakelijk. Nodig en voldoende is blijkens het bewijs, dat (10.21) geldt.

---

*) Wanneer wij deze onderstelling maken, dan nemen wij steeds stilzwijgend a dat $\underline{x}$ en $\underline{y}$ een simultane verdeling bezitten.

Deze stelling geeft tevens een belangrijk verschil aan tussen verwachtingen en varianties. Voor optelbaarheid van verwachtingen behoeft (10.21) niet te gelden.

De stellingen 10.3 en 10.4 zijn zowel voor de theoretische statistiek als voor de toepassingen van groot belang. Wij geven twee toepassingen.

Voorbeeld 10.9

Bereken de verwachting en de variantie van de binomiale verdeling.

Oplossing

Volgens voorbeeld 8.7 kan een binomiale grootheid $\underline{x}$ (vgl. (7.6)) 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. Voor ieder daarvan geldt $\mathcal{E}\underline{x}_i = p$ en $\mathcal{E}\underline{x}_i^2 = p$, dus (vgl. (10.23)) $\sigma^2(\underline{x}_i) = p - p^2 = pq$.

Voor

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

geldt dus volgens stelling 10.3

(10.28) $$\mu = \mathcal{E}\underline{x} = np$$

en volgens (10.27)

(10.29) $$\sigma^2 = \sigma^2(\underline{x}) = npq \ .$$

Voorbeeld 10.10

Het bureau voor marktanalyse genoemd in voorbeeld 5.9 vraagt zich af hoeveel bezoeken men gemiddeld zal moeten afleggen om k deelnemers aan het consumentenpanel te vinden. Bovendien wil men graag weten of het aantal bezoeken, nodig om k deelnemers te krijgen, sterk zal variëren en daarom vraagt men ook hoe groot de spreiding van dit aantal bezoeken is.

Oplossing

Het aantal bezoeken $\underline{n}$, nodig om precies k deelnemers te vinden, heeft volgens voorbeeld 5.9 een negatief binomiale verdeling. Volgens (9.15) kan de stochastische grootheid $\underline{n}$ beschouwd worden als de som van k onafhankelijke grootheden $\underline{n}_1, \underline{n}_2, \ldots, \underline{n}_k$, waarvan $\underline{n}_1$ het aantal experimenten (= bezoeken) voorstelt tot en met het eerste succes, $\underline{n}_2$ het daarop volgende aantal experimenten tot en met het tweede succes, enz. Daar alle experimenten onafhankelijk zijn, geldt dit ook voor de grootheden $\underline{n}_i$ (i=1,2,...,k) en deze bezitten bovendien alle dezelfde verdeling als $\underline{n}_1$. Het eerste moment daarvan is reeds berekend (vgl. (10.6)); het tweede kan op analoge wijze afgeleid worden. Differentieert men nl. de betrekking

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

tweemaal naar q, dan vindt men

$$\sum_{n=0}^{\infty} n(n-1)q^{n-2} = \frac{2}{(1-q)^3} \quad ,$$

waaruit, samen met (10.6), volgt

(10.30) $$\mathcal{E}\,\underline{n}_i^2 = \sum_{n_i=0}^{\infty} n_i^2 p q^{n_i - 1} = \frac{1+q}{p^2} \, .$$

Passen wij nu (10.23) toe, dan blijkt

(10.31) $$\sigma^2(\underline{n}_i) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} \, ,$$

en dus leiden stelling 10.3 resp. stelling 10.4 tot

(10.32) $$\mu = \mathcal{E}\,\underline{n} = \frac{k}{p} \quad \text{en} \quad \sigma^2 = \sigma^2(\underline{n}) = \frac{kq}{p^2} \, .$$

In het geciteerde voorbeeld, met $p = \frac{2}{5}$ en k = 10, is het verwachte aantal bezoeken dus 25 en de spreiding

$$\sqrt{\frac{kq}{p^2}} = \sqrt{\frac{10 \cdot \frac{3}{5}}{\left(\frac{2}{5}\right)^2}} = 6{,}12 \, .$$

Stelling 10.5

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

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

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:

Stelling 10.6

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

(10.34) $$\underline{x}^* \stackrel{\text{def}}{=} \frac{\underline{x}-\mu}{\sigma} = \frac{\tilde{\underline{x}}}{\sigma}$$

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. Omdat de $N(\mu,\sigma)$-verdeling verwachting μ en spreiding σ heeft, volgt de gestandaardiseerde grootheid $\underline{x}^*$ van elke $\underline{x}$ met een normale verdeling, de gestandaardiseerde normale verdeling (stelling 7.1).

Voorbeeld 10.11 *)

Een verpakkingsmachine, die pakjes thee van nominaal 100 gram maakt, verpakt gemiddeld 101,0 gram in ieder pakje. De verdeling van het gewicht aan thee in een pakje is normaal en bezit een spreiding van 0,60 gram.

Beantwoordt de volgende vragen:

a) Hoeveel procent van de pakjes bevat minder dan het nominale gewicht?

b) Tot welk bedrag zou men het gemiddelde gewicht, bij gelijkblijvende spreiding, op moeten voeren om dit percentage tot 2% terug te brengen?

c) Hoeveel procent van de pakjes bevat dan meer dan 103,0 gram?

Oplossing

Het gewicht van de inhoud is N(101,0;0,60) verdeeld.

a) De kans op pakjes met minder dan het nominale gewicht bedraagt

$$P\left[\underline{x} < 100\right] = P\left[\underline{x} \leq 100\right] = P\left[\frac{\underline{x}-101,0}{0,60} \leq \frac{100-101,0}{0,60}\right] = P\left[\underline{u} \leq -1,67\right] = 0,0475$$

en het gezochte percentage is dus 4,75%.

b) Het percentage van 4,75% kan bij gelijkblijvende spreiding alleen verminderd worden door het gemiddelde te verhogen, bijv. tot μ'. Gegeven is dan dat moet gelden

$$P\left[\underline{x} \leq 100\right] = P\left[\frac{\underline{x}-\mu'}{0,60} \leq \frac{100-\mu'}{0,60}\right] = P\left[\underline{u} \leq \frac{100-\mu'}{0,60}\right] = 0,02.$$

Uit de tabel van de normale verdeling lezen wij af, dat dit percentage ongeveer bereikt wordt bij $u = -2,06$ en dus is $\frac{100-\mu'}{0,60} = -2,06$ of $\mu' = 101,24$. Het gemiddelde gewicht moet dus op de waarde van 101,24 ingesteld worden.

---

*) Dit voorbeeld is ontleend aan J. HEMELRIJK en DORALINE WABEKE, Elementaire statistische opgaven met uitgewerkte oplossingen, Noorduijn en Zoon N.V., Gorinchem 1957. Dit boekje bevat 100 opgaven, waarvan een groot deel op praktijkgevallen is gebaseerd. Voorbeeld 10.11 is opgave 42.

c) Voor de laatste vraag berekenen wij

$$P\left[\underline{x} \geq 103\right] = P\left[\underline{u} \geq \frac{103-101,24}{0,60}\right] = P\left[\underline{u} \geq 2,93\right] = 0\ 0017\ ;$$

m.a.w. stellen wij de verpakkingsmachine in op 101,24 gram, dan bevat 0,17% van de pakjes meer dan 103 gram thee.

Stelling 10.7

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

(10.35) $$g(y) \overset{\text{def}}{=} \mathcal{E}\left(\phi(\underline{x},\underline{y}) \mid \underline{y} = y\right)$$

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

(10.36) $$\mathcal{E}\,\phi(\underline{x},\underline{y}) = \mathcal{E}\,g(\underline{y})\ .$$

Bewijs (voor het continue geval):

Volgens de definities 10.4 en 10.5 geldt

$$\mathcal{E}\,\phi(\underline{x},\underline{y}) = \int_{-\infty}^{+\infty}\!\!\int \phi(x,y) f(x,y)\ dxdy =$$

$$= \int_{-\infty}^{+\infty} f_{\underline{y}}(y)\ dy \int_{-\infty}^{+\infty} \phi(x,y)\ \frac{f(x.y)}{f_{\underline{y}}(y)}\ dx =$$

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

Deze stelling houdt dus in dat de verwachting van een functie t.o.v. een twee-dimensionale kansverdeling berekend kan worden door eerst de voorwaardelijke verwachting t.o.v. één van de variabelen te bepalen en vervolgens van de gevonden functie de verwachting t.o.v. de andere variabele te berekenen. Wij kunnen deze stelling bijv. gebruiken in een geval, zoals in voorbeeld 8.6 geschetst is. Men berekent dan eerst de voorwaardelijke verwachting bij gegeven waarde T van $\underline{t}$; deze is volgens (10.17) gelijk aan μT. Neemt men van $\mu\underline{T}$ de verwachting t.o.v.

de marginale verdeling van $\underline{T}$, dan vindt men (vgl. (10.9)) voor de verwachting van het aantal klanten gedurende de levertijd $\frac{\alpha\mu}{\lambda}$. In dit geval hebben wij de verdeling van het aantal klanten reeds berekend bij voorbeeld 8.6 en dan vormt deze weg geen vereenvoudiging *). Anders is het echter gesteld in voorbeeld 10.12.

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 gedefinieerd door

$$\mathrm{cov}(\underline{x},\underline{y}) \stackrel{\text{def}}{=} \mathcal{E}\,\tilde{\underline{x}}\,\tilde{\underline{y}} = \mathcal{E}(\underline{x}-\mathcal{E}\underline{x})(\underline{y}-\mathcal{E}\underline{y}) \quad . \tag{10.37}$$

<u>Voorbeeld 10.12</u>

Wij berekenen met behulp van stelling 10.7 de covariantie van $\underline{x}$ en $\underline{y}$ voor het geval van de <u>twee-dimensionale normale verdeling</u> (vgl. (8.9)).

<u>Oplossing</u>

Volgens (10.16) is

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

dus

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

Volgens (10.36) is dus

---

*) Past men de formule (10.32) toe op de verdeling (8.36), dan vindt men een andere waarde dan $\frac{\alpha\mu}{\lambda}$. De in (10.32) gegeven verwachting behoort bij de negatief binomiale verdeling, zoals deze in (7.8) gegeven is. Hierin is $\underline{x}$ het aantal worpen tot en met het k^{de} succes. Men kan echter ook de kans opgeven dat er $\underline{x}$ mislukkingen hebben plaatsgevonden voor het k^{de} succes, waarbij $\underline{x}$ dus waarden vanaf 0 en niet vanaf k kan aannemen. Dit is de vorm waarin de verdelingsfunctie (8.36) opgeschreven staat. Bij deze vorm is de verwachting van de negatief binomiale verdeling met parameters k en p gelijk aan kqp^{-1}, terwijl de variantie kqp^{-2} blijft (vgl. (10.32)). Is $k = \alpha$ en $p = \frac{\lambda}{\lambda+\mu}$, dan vindt men voor de gezochte verwachting wederom $\frac{\alpha\mu}{\lambda}$.

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

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

Dus

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

In dit voorbeeld is stelling 10.7 van groot nut geweest, aangezien men zonder kennis van deze stelling $\mathcal{E}(\underline{x}-\mu_1)(\underline{y}-\mu_2)$ geheel had moeten uitschrijven, hetgeen een ingewikkelde vorm tot resultaat heeft.

Stelling 10.8

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

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

Bewijs:

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

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

Stelling 10.9

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

(10.40) $$|\operatorname{cov}(\underline{x},\underline{y})| \leq \sigma(\underline{x}).\sigma(\underline{y}) \ .$$

Bewijs:

Stel in (10.39) b = 1. Daar een variantie steeds ≥ 0 is, geldt

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

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

$$\frac{1}{4}D = \{\mathrm{cov}(\underline{x},\underline{y})\}^2 - \sigma^2(\underline{x})\sigma^2(\underline{y}) \leq 0 \ .$$

Hieruit volgt (10.40).

De correlatiecoëfficiënt van $\underline{x}$ en $\underline{y}$ wordt gedefinieerd door

$$\rho(\underline{x},\underline{y}) \stackrel{\mathrm{def}}{=} \frac{\mathrm{cov}(\underline{x},\underline{y})}{\sigma(\underline{x})\sigma(\underline{y})} \ . \tag{10.41}$$

Volgens (10.40) geldt

$$-1 \leq \rho(\underline{x},\underline{y}) \leq +1 \ , \tag{10.42}$$

en volgens (10.38) is $\rho(\underline{x},\underline{y})$ voor een twee-dimensionale normale verdeling gelijk aan de parameter ρ uit formule (8.9).

Stelling 10.10

Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan is $\mathrm{cov}(\underline{x},\underline{y}) = 0$. Het omgekeerde is in het algemeen niet juist, maar wel als $\underline{x}$ en $\underline{y}$ een twee-dimensionale normale verdeling (8.9) volgen.

Bewijs:

Men houde in het oog dat $\mathcal{E}\,\underline{x}$ en $\mathcal{E}\,\underline{y}$ constanten zijn. Volgens stelling 10.3 is

$$\begin{aligned} \mathcal{E}(\underline{x}-\mathcal{E}\underline{x})(\underline{y}-\mathcal{E}\underline{y}) &= \mathcal{E}(\underline{x}\underline{y} - \underline{x}\mathcal{E}\underline{y} - \underline{y}\mathcal{E}\underline{x} + \mathcal{E}\underline{x}\,\mathcal{E}\underline{y}) = \\ &= \mathcal{E}\underline{x}\underline{y} - \mathcal{E}\underline{x}\,\mathcal{E}\underline{y} - \mathcal{E}\underline{y}\,\mathcal{E}\underline{x} + \mathcal{E}\underline{x}\,\mathcal{E}\underline{y} = \mathcal{E}\underline{x}\underline{y} - \mathcal{E}\underline{x}\,\mathcal{E}\underline{y} \ . \end{aligned} \tag{10.43}$$

Wegens (10.21) is dit nul voor onafhankelijke $\underline{x}$ en $\underline{y}$. Geldt omgekeerd $\mathrm{cov}(\underline{x},\underline{y}) = 0$ voor $\underline{x}$ en $\underline{y}$ met verdeling (8.9), dan is $\rho = \rho(\underline{x},\underline{y}) = 0$ in die formule. Dat hieruit onafhankelijkheid volgt is reeds na (9.12) opgemerkt.

Als men met meer dan twee stochastische variabelen te maken heeft, die niet onafhankelijk zijn, is het makkelijk om hun eerste twee momenten samen te vatten in een vector (deel 1, paragraaf 1) en een matrix (deel 1, paragraaf 5). Als bijv. $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ stochastische variabelen

zijn, schrijft men $\mathcal{E}\,\underline{X} = (\mathcal{E}\underline{x}_1, \mathcal{E}\underline{x}_2, \ldots, \mathcal{E}\underline{x}_n)$ voor de vector van de verwachtingen. In plaats van de tweede momenten vermeldt men meestal de varianties (op de hoofddiagonaal) en de covarianties in de vorm van de <u>covariantiematrix</u> $\mathcal{C}$ met elementen

$$c_{ik} = \mathcal{E}\,\tilde{\underline{x}}_i\,\tilde{\underline{x}}_k = \mathcal{E}(\underline{x}_i - \mathcal{E}\,\underline{x}_i)(\underline{x}_k - \mathcal{E}\,\underline{x}_k) \quad . \tag{10.44}$$

Men ziet direct dat $\mathcal{C}$ een symmetrische matrix is. Bovendien is $\mathcal{C}$ positief semi-definiet (deel 1, definitie 13.3); d.w.z. voor elke vector $Y \neq 0$ geldt $Y\,\mathcal{C}\,Y \geq 0$. Wij zullen dit laatste niet bewijzen.

11. <u>Functies van stochastische grootheden.</u>

Reeds eerder is opgemerkt, dat een functie van één of meer stochastische grootheden zelf in het algemeen weer een stochastische grootheid is. De volgende stelling geeft ons een middel om de kansverdeling van een dergelijke functie te berekenen. Evenals in paragraaf 8 formuleren wij de stellingen voor twee-dimensionale kansverdelingen. Ze zijn echter algemeen geldig en kunnen voor n-dimensionale kansverdelingen op analoge wijze toegepast worden.

<u>Stelling 11.1</u>

Laten $\underline{x}$ en $\underline{y}$ een simultane kansverdeling bezitten met $f(x,y)$ als kansdichtheid en laat $\phi(x,y)$ een functie van x en y zijn. Laat verder, in de twee-dimensionale ruimte $\mathbb{R}_2$ met x en y als coördinaten, G_a het gebied voorstellen van al dié punten waarvoor geldt

$$\phi(x,y) \leq a \quad ; \tag{11.1}$$

dan is

$$\begin{aligned} F_{\underline{\phi}}(a) &= P\left[\underline{\phi} \leq a\right] = P\left[\phi(\underline{x},\underline{y}) \leq a\right] = \\ &= P\left[\underline{P} \in G_a\right] = \iint\limits_{G_a} f(x,y)\,dxdy \quad . \end{aligned} \tag{11.2}$$

Hierin stelt $\underline{P}$ het stochastische punt $(\underline{x},\underline{y})$ voor. In het geval van een discrete verdeling wordt in het laatste lid van (11.2) de tweevoudige integraal door een dubbele som vervangen.

Bewijs:

De stelling volgt uit (8.8).

Voorbeeld 11.1 *)

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 $\frac{2}{3}$ 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 $\frac{1}{2}$.

Oplossing

a) Geven wij het eindcijfer van candidaat A aan met $\underline{x}$, dan bezit $\underline{x}$ een binomiale verdeling met $p = \frac{2}{3}$ en $n = 5$. De kansen worden dus gegeven door (vgl. (7.6))

$$P[\underline{x} = k] = \binom{5}{k} \left(\frac{2}{3}\right)^{k} \left(\frac{1}{3}\right)^{5-k} , \tag{11.3}$$

terwijl de verwachting volgens (10.28) is $\mathcal{E}\underline{x} = \frac{10}{3}$ en de variantie volgens (10.29) $\sigma^2(\underline{x}) = \frac{10}{9}$; de spreiding is dus $\sqrt{\frac{10}{9}} = 1{,}05$.

b) Voor candidaat B noemen wij $\underline{x}$ het aantal punten dat hij met het oplossen van de vragen 1, 2 of 3 kan bereiken en $\underline{y}$ het aantal punten dat hij met het oplossen van de opgaven 4 en 5 kan bereiken. Het eindcijfer, voorgesteld door $\underline{z}$, is dan dus gelijk aan $\underline{z} = \underline{x}+\underline{y}$, en $\underline{x}$ en $\underline{y}$ zijn onafhankelijk wegens de onafhankelijkheid van de antwoorden.

---

*) Dit is opgave 73 van het reeds eerder genoemde opgavenboekje (vgl. de voetnoot op blz. 90).

Nu is $$P[\underline{x}=k] = \binom{3}{k}\left(\frac{2}{3}\right)^{k}\left(\frac{1}{3}\right)^{3-k}$$

en $$P[\underline{y}=j] = \binom{2}{j}\left(\frac{1}{2}\right)^{j}\left(\frac{1}{2}\right)^{2-j} = \binom{2}{j}\cdot\frac{1}{4}\ .$$

De verzamelingen G_a uit stelling 11.1 kunnen nu aangegeven worden in een twee-dimensionale figuur (zie figuur 11.1).

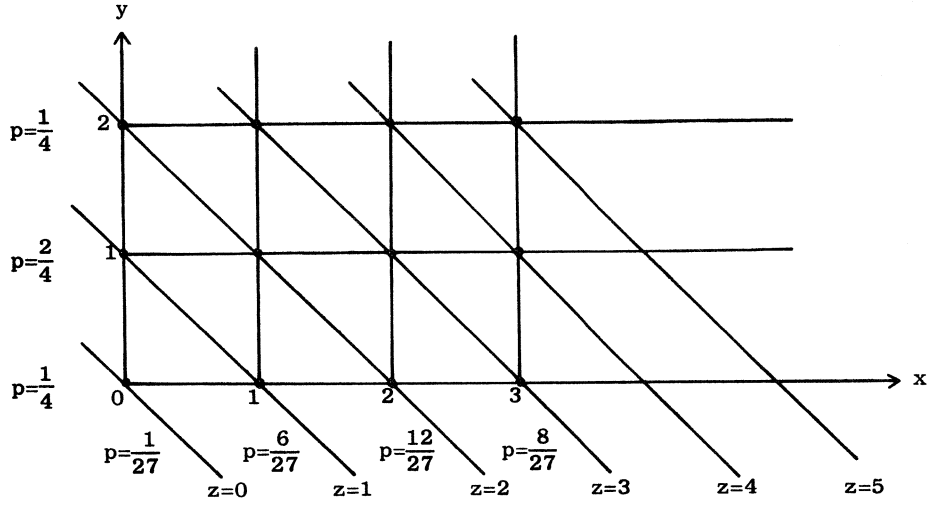


fig. 11.1

De kansverdeling van $\underline{z} = \underline{x}+\underline{y}$

We vinden op deze wijze

$$P[\underline{z}=0] = \frac{1}{4}\cdot\frac{1}{27} = \frac{1}{108}\ ,$$

$$P[\underline{z}=1] = \frac{2}{4}\cdot\frac{1}{27} + \frac{1}{4}\cdot\frac{6}{27} = \frac{8}{108}\ ,$$

$$P[\underline{z}=2] = \frac{1}{4}\cdot\frac{1}{27} + \frac{2}{4}\cdot\frac{6}{27} + \frac{1}{4}\cdot\frac{12}{27} = \frac{25}{108}\ ,$$

$$P[\underline{z}=3] = \frac{1}{4}\cdot\frac{6}{27} + \frac{2}{4}\cdot\frac{12}{27} + \frac{1}{4}\cdot\frac{8}{27} = \frac{38}{108}\ ,$$

$$P[\underline{z}=4] = \frac{1}{4}\cdot\frac{12}{27} + \frac{2}{4}\cdot\frac{8}{27} = \frac{28}{108}\ ,$$

$$P[\underline{z}=5] = \frac{1}{4}\cdot\frac{8}{27} = \frac{8}{108}\ .$$

De verwachting bedraagt $\mathcal{E}_{\underline{z}} = \mathcal{E}_{\underline{x}} + \mathcal{E}_{\underline{y}} = 3.\frac{2}{3} + 2.\frac{1}{2} = 3,$

de variantie $\sigma^2(\underline{z}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y}) = 3.\frac{2}{3}.\frac{1}{3} + 2.\frac{1}{2}.\frac{1}{2} = \frac{7}{6}$

en de spreiding $\sigma(\underline{z}) = \sqrt{\frac{7}{6}} = 1,08.$

Stelling 11.2

Zijn $\underline{x}$ en $\underline{y}$ continu en onafhankelijk verdeelde stochastische grootheden met verdelingsfuncties F(x) en G(y) en verdelingsdichtheden f(x) en g(y), dan geldt voor de verdelingsfunctie H(z) en de verdel ngsdichtheid h(z) van $\underline{z} = \underline{x} + \underline{y}$

(11.4) $$H(z) = \int_{-\infty}^{+\infty} f(x)G(z-x)\,dx$$

resp.

(11.5) $$h(z) = \int_{-\infty}^{+\infty} f(x)g(z-x)\,dx\ .$$

Bewijs:

Volgens (11.2) en stelling 9.3 is

$$H(z) = \iint\limits_{x+y \leq z} f(x)g(y)\,dxdy =$$

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

Hieruit wordt h(z) verkregen door naar z te differentiëren (vgl. stelling 12.4 uit deel 1).

Deze stelling, die men bij berekeningen vaak moet toepassen, illustreren wij aan de hand van twee voorbeelden.

Stelling 11.3

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

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

waarin :: betekent "op een constante factor na gelijk aan".
Het produkt 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, zoals eerder (8.19), 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 vóór 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.

Opmerkingen

Stelling 11.3 is direct (door inductie) uit te breiden tot meer dan twee 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 (8.9) en (11.2), waarin voor G_a het gebied $x + y \leq z$ 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 (vgl. stelling 7.1), evenmin als het optellen van constante termen, zodat de volgende algemene stelling geldt.

Stelling 11.4

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.

Stelling 11.5

Zijn $\underline{x}$ en $\underline{y}$ onderling onafhankelijk verdeeld en bezit $\underline{x}$ een gamma-verdeling met parameters α en λ, en $\underline{y}$ een exponentiële verdeling met parameter λ, dan heeft $\underline{z} = \underline{x} + \underline{y}$ een gamma-verdeling met parameters $\alpha+1$ en λ.

Bewijs:

De verdelingsdichtheden van $\underline{x}$ en $\underline{y}$ zijn volgens (7.25) resp. (7.20)

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \quad \text{resp.} \quad \lambda e^{-\lambda y} .$$

Substitueren wij dit in (11.5), dan vinden wij

$$h(z) = \int_0^z \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \lambda e^{-\lambda(z-x)} dx =$$

$$= \frac{\lambda^{\alpha+1}}{\Gamma(\alpha)} e^{-\lambda z} \int_0^z x^{\alpha-1} dx = \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} e^{-\lambda z} z^\alpha ,$$

hetgeen de verdelingsdichtheid van een gamma-verdeling met parameters $\alpha+1$ en λ voorstelt.

Opmerking

Uit stelling 11.5 volgt voor $\alpha = 1$, dat de som van twee exponentiële verdelingen met dezelfde parameter λ een gamma-verdeling bezit met parameters $\alpha = 2$ en λ. Successieve toepassing van stelling 11.5 voor $\alpha = 2,3,\ldots$ leidt tot de volgende stelling.

Stelling 11.6

De som van k onderling onafhankelijke, exponentieel verdeelde grootheden met parameter λ bezit een gamma-verdeling met parameters $\alpha = k$ en λ.

Ook deze stelling kan nog algemener gemaakt worden, doch daarop zullen wij hier niet ingaan. Wel merken wij op dat een stelling zoals 11.4 voor gamma-verdelingen niet geldt (zelfs niet voor exponentiële verdelingen).

Wij besluiten deze paragraaf met een praktische toepassing van stelling 11.4.

Voorbeeld 11.2 *)

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 350 gram olie bevatten,
c) de covariantie van het gewicht van een gevulde fles en dat van de olie die erin zit.

---

*) Dit is opgave 67 uit het opgavenboekje (vgl. de voetnoot op blz.90).

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?

Oplossing

Geef het gewicht van de olie-inhoud aan met $\underline{x}$, van de flessen met $\underline{y}$ en van de gevulde flessen met $\underline{z}$. Gegeven is dat $\underline{y}$ N(228,3;11,3) verdeeld is en $\underline{z}$ N(585,2;12,8) verdeeld. Nu is $\underline{z} = \underline{x} + \underline{y}$ en dus is, mede wegens de onafhankelijkheid van $\underline{x}$ en $\underline{y}$, $\mathcal{E}\underline{z} = \mathcal{E}\underline{x} + \mathcal{E}\underline{y}$ en $\sigma^2(\underline{z}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y})$.

a) Volgens stelling 11.4 is $\underline{x}$ normaal verdeeld; de verwachting is $\mathcal{E}\underline{x} = 585,2 - 228,3 = 356,9$ en de variantie $\sigma^2(\underline{x}) = (12,8)^2 - (11,3)^2 = 36,15 = (6,0)^2$; $\underline{x}$ is dus N(356,9;6,0) verdeeld.

b) De kans op een fles met minder dan 350 gram olie bedraagt

$$P[\underline{x} < 350] = P[\underline{x} \leq 350] = P\left[\frac{\underline{x} - 356,9}{6,0} \leq \frac{350-356,9}{6,0}\right] = P[\underline{u} \leq -1,15] = 0,1251$$

en dus bevat 12,5% van de flessen minder dan 350 gram olie.

c) Uit (10.39) volgt $\sigma^2(\underline{y}) = \sigma^2(\underline{z}) + \sigma^2(\underline{x}) - 2\mathrm{cov}(\underline{x},\underline{z})$ en dus is $\mathrm{cov}(\underline{x},\underline{z}) = \frac{1}{2}\left[(12,8)^2 + 36,15 - (11,3)^2\right] = 36,15$.

d) Laat de nieuwe spreiding van $\underline{x}$ $\sigma'(\underline{x})$ bedragen; dan is de kans op een fles met minder dan 350 gram olie $P[\underline{x} \leq 350] = P\left[\underline{u} \leq \frac{350-356,9}{\sigma'(\underline{x})}\right]$. Deze kans mag hoogstens 0,025 bedragen en dus moet voldaan zijn aan $\frac{350-356,9}{\sigma'(\underline{x})} = -1,96$, zodat $\sigma'(\underline{x})$ hoogstens 3,5 mag zijn. Hieruit volgt voor de variantie van $\underline{z}$ $\sigma^2(\underline{z}) = (\sigma'(\underline{x}))^2 + \sigma^2(\underline{y}) = (3,5)^2 + (11,3)^2 = (11,8)^2$ en de spreiding der gevulde flessen moet dus tot 11,8 gram teruggebracht worden.

Opmerking

Deze opgave geeft een typerend beeld van enigszins primitieve toepassingen van de statistiek, zoals deze in de praktijk 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.

12. De theoretische wet van de grote aantallen.

Limietstellingen vormen een zeer belangrijk gedeelte van de waarschijnlijkheidsrekening en de mathematische statistiek. Gezien het karakter van deze leergang kunnen wij er niet lang bij blijven stilstaan en zullen wij ons moeten beperken tot het noemen van de voor ons belangrijkste stellingen. Een uitvoerige behandeling en bewijzen kan men vinden in ieder fundamenteel boek over statistiek en waarschijnlijkheidsrekening, bijv. in de aan het eind van dit deel genoemde boeken van M.G. KENDALL en A. STUART, H. CRAMÉR of S.S. WILKS.

In paragraaf 8 van deel 1, hebben wij de limiet van een rij getallen als volgt gedefinieerd:
Een oneindige rij getallen $a_1, a_2, \ldots$ convergeert naar de limiet a, wanneer bij ieder positief getal ε een getal N gevonden kan worden, zodanig dat voor $n > N$ geldt $|a_n - a| < \varepsilon$.
Dit betekent dus dat vanaf de index N de getallen a_n minder dan ε verwijderd zijn van de limiet a.

Een dergelijke definitie kan bij stochastische grootheden niet gegeven worden. Onderzoekt men bijv. een rij stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots$, dan zal niet alleen $\underline{x}_n$ voor iedere n nog geheel verschillende waarden aan kunnen nemen, maar hetzelfde zal in het algemeen gelden voor iedere functie, die men van deze grootheden kan vormen. Men zal dus een ander, zwakker convergentiebegrip moeten invoeren dan dat voor rijen getallen. Dit gebeurt door niet te eisen dat de

stochastische grootheden $\underline{x}_n$ vanaf een bepaalde n alleen waarden dichtbij een bepaald getal, bijv. μ aannemen, maar door slechts te eisen dat de kans op een waarde van $\underline{x}_n$, welke meer dan ε van μ verwijderd is, naar nul convergeert. Deze kans is $P\left[|\underline{x}_n-\mu| > \varepsilon\right]$ en wij zeggen nu dat $\underline{x}_n$ stochastisch naar μ convergeert voor $n \to \infty$, als deze kans voor $n \to \infty$ tot nul nadert; ook spreekt men wel van convergentie in waarschijnlijkheid. Stelling 12.1 bevat een eigenschap van dit type convergentie.

Stelling 12.1 (ongelijkheid van BIENAYMÉ-TCHEBYCHEF)

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

$$P\left[|\underline{x}^*| \geq \delta\right] \leq \frac{1}{\delta^2} \quad , \tag{12.1}$$

waar $\underline{x}^* = (\underline{x}-\mu)/\sigma$ de gestandaardiseerde $\underline{x}$ voorstelt.

Bewijs:

Wij geven het bewijs voor discreet verdeelde $\underline{x}$; voor een continue verdeling verloopt het analoog. Als $\underline{x}$ met kans p_i de waarde x_i aanneemt (i=1,2,...) dan is

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

Dus

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

$$= \sum_{|x_i^*| \geq \delta} p_i x_i^{*2} + \sum_{|x_i^*| < \delta} p_i x_i^{*2} \geq \sum_{|x_i^*| \geq \delta} p_i x_i^{*2} \geq$$

$$\geq \delta^2 \sum_{|x_i^*| \geq \delta} p_i = \delta^2 P\left[|\underline{x}^*| \geq \delta\right]$$

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 (voorbeeld 10.9). Dan volgt uit (12.1) dat

$$(12.2) \qquad P\left[\,|\underline{x}-p| \geq \delta\sqrt{pq}\,\right] \leq \frac{1}{\delta^2} \ .$$

Bij invullen van $p = \frac{1}{2}$ en $\delta = 1$ staat hier $P\left[|\underline{x}-\frac{1}{2}| \geq \frac{1}{2}\right] \leq 1$.
De kans in het linkerlid is 1, dus de ongelijkheid is in dit speciale geval scherp; verkleining van het rechterlid is niet mogelijk zonder tot een onjuist resultaat te komen.

In de regel is de ongelijkheid (12.1) echter zeer grof. Voor een N(0,1) verdeelde grootheid $\underline{u}$ volgt uit (12.1) voor $\delta = 2$

$$(12.3) \qquad P\left[\,|\underline{u}| \geq 2\,\right] \leq \frac{1}{4} \ .$$

Volgens de tabel van de normale verdeling is de kans in het linkerlid gelijk aan 0,046.

Ondanks dit gebrek aan scherpte is de ongelijkheid vooral voor theoretische doeleinden van groot belang, zoals uit de stellingen 12.2 en 12.3 zal blijken.

Stelling 12.2

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

$$(12.4) \qquad \underline{\bar{x}}_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \underline{x}_i$$

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

$$(12.5) \qquad \lim_{n\to\infty} P\left[\,|\underline{\bar{x}}_n - \mu| > \varepsilon\,\right] = 0.$$

Bewijs:

Volgens stelling 10.3 is

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

en volgens de stellingen 10.4 en 10.5 is

$$\sigma^2(\underline{\bar{x}}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2(\underline{x}_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}\ .$$

Volgens (12.1) toegepast op $\underline{\bar{x}}_n$ is voor elke $\delta > 0$

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

Invullen van $\delta = \frac{\varepsilon\sqrt{n}}{\sigma}$ geeft

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

en voor $n \to \infty$ is de limiet van het rechterlid 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 12.3 (theoretische wet der grote aantallen)

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

(12.6) $$\lim_{n\to\infty} P\left[\,\left|\frac{\underline{x}}{n} - p\right| > \varepsilon\,\right] = 0\ .$$

Opmerkingen

Deze stelling heeft een grote praktische betekenis. Het is de theoretische tegenhanger van de experimentele wet van de grote aantallen (zie paragraaf 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 axioma-stelsel 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, adequaat blijken te zijn ter beschrijving van de nogal ingewikkelde experimentele wet van de grote aantallen, die wij fundamenteel voor statistische verschijnselen hebben genoemd. Dit betekent, dat de interpretatie van een kans p als - bij benadering - een fq in een lange reeks waarnemingen gerechtvaardigd is, terwijl bovendien dit succes van de theorie het vertrouwen in verdere resultaten - d.w.z. het vertrouwen dat de toepassingen daarvan in de praktijk bruikbaar zullen blijken te zijn - vergroot. De op blz. 24 gemaakte opmerking, "dat in het kansmodel het fq als zodanig, met zijn statistische fluctuaties, op natuurlijke wijze weer ingevoerd kan worden", is hiermede waar gemaakt.

De praktische betekenis van stelling 12.2 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^{de} 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 (12.5) 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 fysische 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 is de variabiliteit van een gemiddelde van een aantal onafhankelijke waarnemingen aanzienlijk kleiner, dus de nauwkeurigheid aanzienlijk groter, dan van één enkele waarneming.

De grootheid μ, waarbij $\bar{x}_n$ in de buurt komt, hangt vaak mede van de 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 spreidingen 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 grotere 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 12.2 ook blijft gelden als de varianties ongelijk zijn, mits deze begrensd zijn. Is $\sigma^2(x_i) \leq \sigma_0^2$ 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 in deel 3 van deze syllabusserie te behandelen schattingstheorie.

13. De centrale limietstelling.

Wij hebben al gezien (stelling 11.4) dat een som van normaal verdeelde grootheden weer normaal verdeeld is. Met behulp van deze stelling kunnen allerlei vraagstukken op eenvoudige wijze opgelost worden. Voor grootheden met een andere verdeling dan de normale, gaat dit in het algemeen niet zo eenvoudig. Vaak kan men dan gebruik maken van de volgende uiterst belangrijke stelling, die bekend staat onder de naam centrale limietstelling en welke hier geformuleerd wordt voor onderling onafhankelijke stochastische grootheden.

Stelling 13.1

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

$$(13.1) \qquad \underline{x}(n) \overset{\text{def}}{=} \sum_{i=1}^{n} \underline{x}_i \; ; \quad \mu(n) \overset{\text{def}}{=} \sum_{i=1}^{n} \mu_i \; ; \quad \sigma^2(n) \overset{\text{def}}{=} \sum_{i=1}^{n} \sigma_i^2 \; .$$

De zg. LINDEBERG-voorwaarde, dat voor elke $\varepsilon > 0$ geldt dat de afgeknotte grootheden

$$(13.2) \qquad \underline{u}_i \overset{\text{def}}{=} \begin{cases} \underline{x}_i - \mu_i & \text{als } |x_i - \mu_i| \leq \varepsilon\sigma(n), \\ 0 & \text{anders} \end{cases}$$

voldoen aan

$$(13.3) \qquad \lim_{n\to\infty} \frac{1}{\sigma^2(n)} \sum_{i=1}^{n} \mathcal{E}\, \underline{u}_i^2 = 1 \; ,$$

is dan nodig en voldoende opdat de grootheid

$$(13.4) \qquad \underline{x}(n)^* \overset{\text{def}}{=} \frac{\underline{x}(n) - \mu(n)}{\sigma(n)}$$

voor $n \to \infty$ asymptotisch $N(0,1)$ verdeeld is; d.w.z. dat voor iedere a geldt

$$(13.5) \qquad \lim_{n\to\infty} P\left[\underline{x}(n)^* \leq a\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}u^2}\, du \; .$$

Opmerkingen

De grootheid $\underline{x}(n)^*$ is de gestandaardiseerde van $\underline{x}(n)$, want deze heeft $\mu(n)$ als verwachting en $\sigma(n)$ als spreiding en wij kunnen de inhoud van de stelling dus ook als volgt formuleren:
Onder bepaalde, vrij algemeen geldende voorwaarden bezit de gestandaardiseerde som van een rij stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots$ asymptotisch een N(0,1)-verdeling. Vaak is men minder exact en dan zegt men dat "$\underline{x}(n)$ asymptotisch normaal verdeeld" is, 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 $\underline{x}_1, \underline{x}_2, \ldots$ alle dezelfde verdeling bezitten, dan mag men de LINDEBERG-voorwaarde weglaten. De stelling krijgt dan de volgende vorm:

Stelling 13.2 (centrale limietstelling voor onderling onafhankelijke, identiek verdeelde stochastische grootheden)

Zijn $\underline{x}_1, \underline{x}_2, \ldots$ onderling onafhankelijk verdeeld volgens dezelfde verdeling met verwachting μ en spreiding σ, dan is

$$\text{(13.6)} \qquad \frac{\sum_{i=1}^{n} \underline{x}_i - n\mu}{\sigma\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} \underline{x}_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

voor $n \to \infty$ asymptotisch N(0,1) verdeeld.

Uit deze stelling volgt direct, dat,wanneer $\underline{x}_1, \underline{x}_2, \ldots$ een steekproef voorstellen uit een willekeurige verdeling met verwachting μ en spreiding σ, het gemiddelde

$$\text{(13.7)} \qquad \overline{\underline{x}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i$$

asymptotisch normaal verdeeld is voor $n \to \infty$ met verwachting μ en spreiding $\frac{\sigma}{\sqrt{n}}$.

In deze leergang houden wij ons slechts bezig met verdelingen waarvan alle momenten eindig zijn, zodat wij stelling 13.2 steeds zonder meer kunnen toepassen. Een speciaal geval van stelling 13.2 komt nu aan de orde.

Stelling 13.3 (stelling van De Moivre en Laplace)

Bezit $\underline{x}$ een binomiale verdeling (7.6) met parameters n en p, dan is $\underline{x}$ voor $n \to \infty$ asymptotisch normaal verdeeld. Dus voor iedere a geldt

$$\lim_{n\to\infty} P\left[\frac{\underline{x} - np}{\sqrt{npq}} \leq a\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}u^2}\, du \,. \tag{13.8}$$

Bewijs:

In voorbeeld 10.9 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^{de} 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 13.2, met $\mu = \mu(\underline{x}_i) = p$ en $\sigma^2 = \sigma^2(\underline{x}_i) = pq$.

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 de verwachting np en de spreiding $\sqrt{npq}$. De normale verdeling met dezelfde verwachting en spreiding wordt nu de aangepaste normale verdeling genoemd.

In figuur 13.1 zijn voor p = 0,1 en p = 0,36 en voor n = 4, 9, 25 en 100 de kansen van de binomiale verdeling geschetst, evenals de verdelingsdichtheid van de aangepaste normale verdeling. De x-schaal is daarbij steeds zo gekozen, dat de verdelingsdichtheid dezelfde vorm bezit (de spreidingen zijn dus in cm. gelijk gemaakt). Verder zijn, behalve voor n = 100, ook de verdelingsfuncties getekend.

Stelling 3.3 wordt geïllustreerd door het feit dat de benadering duidelijk beter wordt bij toenemende n. Tevens blijkt uit de figuren,

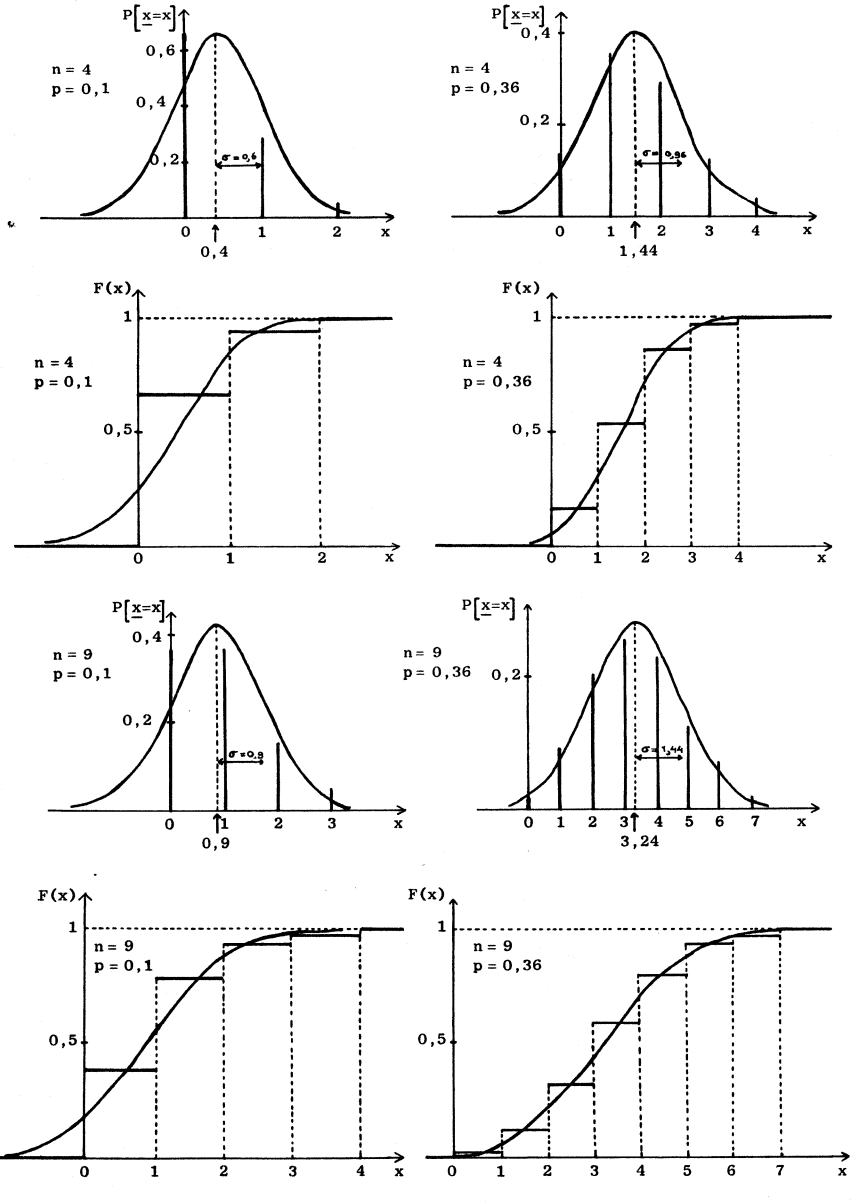


fig. 13.1

Convergentie van de binomiale verdeling naar de normale

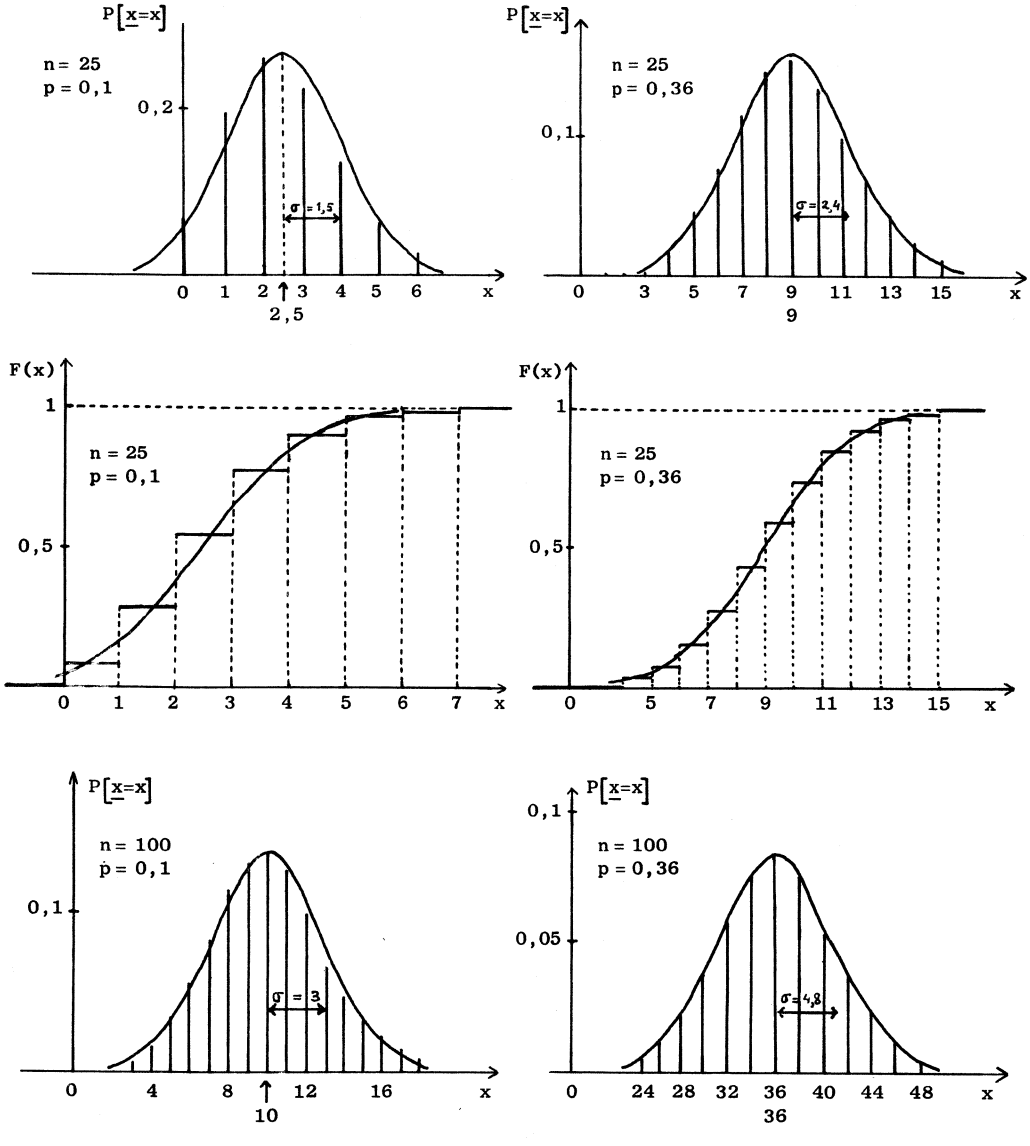


fig. 13.1 (vervolg)

Convergentie van de binomiale verdeling naar de normale

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 = 0,5, hetgeen men reeds kan vermoeden op grond van het feit dat de binomiale verdeling dan symmetrisch is, evenals de normale. Voor n = 100 is het aantal stappen van de cumulatieve binomiale verdeling zo groot, dat de figuur op deze schaal niet meer getekend kan worden. Overigens blijkt reeds uit de figuur van de cumulatieve verdelingen bij p = 0,36 en n = 25, dat de aanpassing daar heel behoorlijk is; voor n = 100 valt de trapfunctie in nog sterkere mate samen met de vloeiende lijn van de cumulatieve normale verdeling.

Voorbeeld 13.1

In paragraaf 2 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 0,5. Het belang van kansen van dit type, die overschrijdingskansen genoemd worden, zullen wij later leren kennen.

Oplossing

Het aantal jongens, $\underline{x}$, heeft een binomiale verdeling met $n = 1927$ en $p = 0,5$, dus met $\mathcal{E}\underline{x} = np = 963,5$ en $\sigma(\underline{x}) = \sqrt{npq} = \frac{1}{2}\sqrt{1927} = 21,9$.

De gevraagde kans is $P\left[\underline{x} \geq 974 \mid n = 1927,\ p = 0,5\right]$, waarin nu achter de streep geen voorwaarden, maar gegevens staan aangegeven. Wij berekenen deze kans nu - bij benadering - met behulp van (13.9) op de volgende wijze (zonder de gegevens over n en p telkens weer te vermelden):

$$P\left[\underline{x} \geq 974\right] = P\left[\frac{\underline{x} - np}{\sqrt{npq}} \geq \frac{974-np}{\sqrt{npq}}\right] \approx P\left[\underline{u} \geq \frac{974-963,5}{21,9}\right] = P\left[\underline{u} \geq 0,478\right],$$

waarin $\underline{u}$ een N(0,1) verdeelde grootheid voorstelt en $\approx$ betekent, dat wij met een benadering te maken hebben.

Volgens de tabel van de N(0,1)-verdeling is de kans in het laatste lid gelijk aan 0,316. Onze uitkomst is dus

$$P\left[\underline{x} \geq 974 \mid n = 1927,\ p = 0{,}5\right] \approx 0{,}316 \ .$$

Wij kunnen de precisie van dit antwoord nog wat opvoeren door de zg. continuïteitscorrectie van YATES toe te passen. Bij het benaderen van de discrete binomiale verdeling is het redelijk niet uit te gaan van $\underline{x} \geq 974$, maar van $\underline{x} \geq 973{,}5$; dit blijkt door omgekeerd de continu N(963,5;21,9) verdeelde grootheid af te ronden op het dichtstbijzijnde gehele getal. Men vindt dan

$$P\left[\underline{x} \geq 973{,}5\right] = P\left[u \geq \frac{973{,}5-963{,}5}{21{,}9}\right] = 0{,}324337 \ .$$

Het exacte antwoord, berekend door de X1 Electrologica computer, is 0,324342.

Men ziet dat de benadering voor n = 1927 en p = 0,5 vrijwel perfect is. Reeds vanaf n = 80 is de precisie voor de meeste doeleinden voldoende, tenzij de verdeling erg scheef is (p dicht bij 0 of 1). Dan kan men de hier niet bewezen stelling gebruiken, dat de verdeling van $\underline{x}$ voor kleine p en grote n benaderd mag worden door de Poisson-verdeling (7.9) met $\lambda = np$ als parameter.

Tabel A

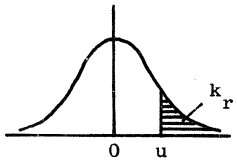
Overzicht van de kansen resp. verdelingsdichtheden, verwachtingen en varianties van enige belangrijke verdelingsfuncties

naam van de verdeling	kansen resp. verdelingsdichtheden	verwachting	variantie
alternatief	$P[\underline{x}=0] = p \; ; \; P[\underline{x}=1] = 1-p = q$	q	pq
binomiale	$P[\underline{x}=k] = \binom{n}{k} p^k q^{n-k} \quad (k=0,\ldots,n)$	np	npq
hypergeometrische	$P[\underline{x}=k] = \dfrac{\binom{M}{k}\binom{N}{n-k}}{\binom{M+N}{n}}$ $(\max(n-N,0) \leq k \leq \min(M,n))$	$\dfrac{Mn}{M+N}$	$\dfrac{nMN(M+N-n)}{(M+N)^2(M+N-1)}$
negatief binomiale	$P[\underline{x}=x] = \binom{x-1}{k-1} p^k q^{x-k} \quad (x=k,k+1,\ldots)$	$\dfrac{k}{p}$ *)	$\dfrac{kq}{p^2}$
Poisson	$P[\underline{x}=k] = \dfrac{\lambda^k e^{-\lambda}}{k!} \quad (k=0,1,\ldots)$	λ	λ
(discrete) homogene	$P[\underline{x}=k] = \dfrac{1}{n+1} \quad (k=0,\ldots,n)$	$\dfrac{n}{2}$	$\dfrac{1}{12}n(n+2)$
(continue) homogene	$\dfrac{1}{\theta_2-\theta_1}$ voor $\theta_1 \leq x \leq \theta_2$	$\dfrac{1}{2}(\theta_1+\theta_2)$	$\dfrac{1}{12}(\theta_2-\theta_1)^2$
normale	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$	μ	σ^2
exponentiële	$\lambda e^{-\lambda x} \quad (0 \leq x < \infty)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
χ^2	$\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (0 \leq x < \infty)$	n	$2n$
gamma	$\dfrac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \quad (0 \leq x < \infty)$	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$
bêta	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad (0 \leq x \leq 1)$	$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
logarithmisch normale	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \cdot \dfrac{1}{x} \quad (0 \leq x < \infty)$	$e^{\mu+\frac{1}{2}\sigma^2}$	$e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}$

*) Vergelijk de voetnoot op blz. 92.

Tabel B

N(0,1)-verdeling



Waarden van $k_r.10^4$ voor u = 0,00(0,01)3,49.

$$k_r = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{1}{2}v^2}\, dv \ .$$

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 dichtstbijgelegen oneven cijfer afgerond moet worden. Een 5 wordt naar het dichtstbijzijnde even cijfer afgerond.

Literatuur over statistiek

Elementaire inleidingen

J.L. HODGES Jr. and E.L. LEHMANN, Basic concepts of probability and statistics, Holden-Day, San Francisco, 1964 (een voortreffelijke inleiding).

W.A. WALLIS and H.V. ROBERTS, Statistics: a new approach, The Free Press, Glencoe, Illinois, 1956 (een boek met talrijke uitgewerkte voorbeelden en vraagstukken).

H. DE JONGE, Inleiding tot de medische statistiek, Nederlands Instituut voor Praeventieve Geneeskunde, Leiden, deel I (2e druk), 1963; deel II, 1960 (een goed "receptenboek" zonder veel theorie).

W.J. DIXON and F.J. MASSEY, An introduction to statistical analysis, McGraw-Hill, New York, 2e druk, 1957 (idem).

M.L. WIJVEKATE, Verklarende statistiek, Aula pocket nr. 39, Het Spectrum, Utrecht, 1961 en later (een populaire inleiding voor niet wiskundig geschoolden).

M.J. MORONEY, Facts from figures, Pelican book A 236, Harmondsworth, Middlesex, 3e druk, 1962 (idem).

Wiskundige leerboeken en handboeken

M.G. KENDALL and A. STUART, The advanced theory of statistics, Griffin, London, deel I, 1958; deel II, 1960; deel III, 1966 (een belangrijk handboek).

H. CRAMÉR, Mathematical methods of statistics, Princeton University Press, Princeton, 1946 (een duidelijk, maar enigszins verouderd leerboek).

S.S. WILKS, Mathematical statistics, Wiley, New York, 1962 (een modern handboek).

UITGAVEN IN DE SERIE MC SYLLABUS

Onderstaande uitgaven zijn verkrijgbaar bij het Mathematisch Centrum, 2e Boerhaavestraat 49 te Amsterdam-1005, tel. 020-947272.

MCS 1.1 F. GÖBEL & J. VAN DE LUNE, *Leergang Besliskunde, deel 1: Wiskundige basiskennis*, 1965. ISBN 90 6196 014 2.

MCS 1.2 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 2: Kansberekening*, 1965. ISBN 90 6196 015 0.

MCS 1.3 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 3: Statistiek*, 1966. ISBN 90 6196 016 9.

MCS 1.4 G. DE LEVE & W. MOLENAAR, *Leergang Besliskunde, deel 4: Markovketens, en wachttijden*, 1966. ISBN 90 6196 017 7.

MCS 1.5 J. KRIENS & G. DE LEVE, *Leergang Besliskunde, deel 5: Inleiding tot de mathematische besliskunde*, 1966. ISBN 90 6196 018 5.

MCS 1.6a B. DORHOUT & J. KRIENS, *Leergang Besliskunde, deel 6a: Wiskundige programmering 1*, 1968. ISBN 90 6196 032 0.

MCS 1.7a G. DE LEVE, *Leergang Besliskunde, deel 7a: Dynamische programmering 1*, 1968. ISBN 90 6196 033 9.

MCS 1.7b G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7b: Dynamische programmering 2*, 1970. ISBN 90 6196 055 X.

MCS 1.7c G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7c: Dynamische programmering 3*, 1971. ISBN 90 6196 066 5.

MCS 1.8 J. KRIENS, F. GÖBEL & W. MOLENAAR, *Leergang Besliskunde, deel 8: Minimaxmethode, netwerkplanning, simulatie*, 1968. ISBN 90 6196 034 7.

MCS 2.1 G.J.R. FÖRCH, P.J. VAN DER HOUWEN & R.P. VAN DE RIET, *Colloquium stabiliteit van differentieschema's, deel 1*, 1967. ISBN 90 6196 023 1.

MCS 2.2 L. DEKKER, T.J. DEKKER, P.J. VAN DER HOUWEN & M.N. SPIJKER, *Colloquium stabiliteit van differentieschema's, deel 2*, 1968. ISBN 90 6196 035 5.

MCS 3.1 H.A. LAUWERIER, *Randwaardeproblemen, deel 1*, 1967. ISBN 90 6196 024 X.

MCS 3.2 H.A. LAUWERIER, *Randwaardeproblemen, deel 2*, 1968. ISBN 90 6196 036 3.

MCS 3.3 H.A. LAUWERIER, *Randwaardeproblemen, deel 3*, 1968. ISBN 90 6196 043 6.

MCS 4 H.A. LAUWERIER, *Representaties van groepen*, 1968. ISBN 90 6196 037 1.

MCS 5 J.H. VAN LINT, J.J. SEIDEL & P.C. BAAYEN, *Colloquium discrete wiskunde*, 1968. ISBN 90 6196 044 4.

MCS 6 K.K. KOKSMA, *Cursus ALGOL 60*, 1969. ISBN 90 6196 045 2.

MCS 7.1 *Colloquium Moderne rekenmachines, deel 1*, 1969. ISBN 90 6196 046 0.

MCS 7.2 *Colloquium Moderne rekenmachines, deel 2*, 1969. ISBN 90 6196 047 9.

MCS 8 H. BAVINCK & J. GRASMAN, *Relaxatietrillingen*, 1969. ISBN 90 6196 056 8.

MCS 9.1 T.M.T. COOLEN, G.J.R. FÖRCH, E.M. DE JAGER & H.G.J. PIJLS, *Elliptische differentiaalvergelijkingen, deel 1*, 1970. ISBN 90 6196 048 7.

MCS 9.2 W.P. VAN DEN BRINK, T.M.T. COOLEN, B. DIJKHUIS, P.P.N. DE GROEN, P.J. VAN DER HOUWEN, E.M. DE JAGER, N.M. TEMME & R.J. DE VOGELAERE, *Colloquium Elliptische differentiaalvergelijkingen, deel 2*, 1970. ISBN 90 6196 049 5.

MCS 10 J. FABIUS & W.R. VAN ZWET, *Grondbegrippen van de waarschijnlijkheidsrekening*, 1970. ISBN 90 6196 057 6.

MCS 11 H. BART, M.A. KAASHOEK, H.G.J. PIJLS, W.J. DE SCHIPPER & J. DE VRIES, *Colloquium Halfalgebra's en positieve operatoren*, 1971. ISBN 90 6196 067 3.

MCS 12 T.J. DEKKER, *Numerieke algebra*, 1971. ISBN 90 6196 068 1.

MCS 13 F.E.J. KRUSEMAN ARETZ, *Programmeren voor rekenautomaten; De MC ALGOL 60 vertaler voor de EL X8*, 1971. ISBN 90 6196 069 x.

MCS 14 H. BAVINCK, W. GAUTSCHI & G.M. WILLEMS, *Colloquium Approximatietheorie*, 1971. ISBN 90 6196 070 3.

MCS 15.1 T.J. DEKKER, P.W. HEMKER & P.J. VAN DER HOUWEN, *Colloquium Stijve differentiaalvergelijkingen, deel 1*, 1972. ISBN 90 6196 078 9.

MCS 15.2 P.A. BEENTJES, K. DEKKER, H.C. HEMKER, S.P.N. VAN KAMPEN & G.M. WILLEMS, *Colloquium Stijve differentiaalvergelijkingen, deel 2*, 1973. ISBN 90 6196 079 7.

MCS 15.3 P.A. BEENTJES, K. DEKKER, P.W. HEMKER & M. VAN VELDHUIZEN, *Colloquium Stijve differentiaalvergelijkingen, deel 3*, 1975. ISBN 90 6196 118 1.

MCS 16.1 L. GEURTS, *Cursus Programmeren, deel 1: De elementen van het programmeren*, 1973. ISBN 90 6196 080 0.

MCS 16.2 L. GEURTS, *Cursus Programmeren, deel 2: De programmeertaal ALGOL 60*, 1973. ISBN 90 6196 087 8.

MCS 17.1 P.S. STOBBE, *Lineaire algebra, deel 1*, 1974. ISBN 90 6196 090 8.

MCS 17.2 P.S. STOBBE, *Lineaire algebra, deel 2*, 1974. ISBN 90 6196 091 6.

MCS 18 F. VAN DER BLIJ, H. FREUDENTHAL, J.J. DE IONGH, J.J. SEIDEL & A. VAN WIJNGAARDEN, *Een kwart eeuw wiskunde 1946-1971, Syllabus van de Vakantiecursus 1971*, 1974. ISBN 90 6196 092 4.

MCS 19 A. HORDIJK, R. POTHARST & J.TH. RUNNENBURG, *Optimaal stoppen van Markovketens*, 1974. ISBN 90 6196 093 2.

MCS 20 T.M.T. COOLEN, P.W. HEMKER, P.J. VAN DER HOUWEN & E. SLAGT, *ALGOL 60 procedures voor begin- en randwaardeproblemen*, 1976. ISBN 90 6196 094 0.

MCS 21 J.W. DE BAKKER (red.), *Colloquium Programmacorrectheid*, 1974. ISBN 90 6196 103 3.

* MCS 22 R. HELMERS, F.H. RUYMGAART, M.C.A. VAN ZUYLEN & J. OOSTERHOFF, *Asymptotische methoden in de statistiek*, 1976. ISBN 90 6196 104 1.

* MCS 23.1 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 1*, 1976. ISBN 90 6196 105 X.

* MCS 23.2 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 2*, 1976. ISBN 90 6196 115 7.

MCS 24.1 P.J. VAN DER HOUWEN, *Numerieke integratie van differentiaalvergelijkingen, deel 1: Eenstapsmethoden*, 1974. ISBN 90 6196 106 8.

MCS 25 *Colloquium Structuur van programmeertalen*, 1976. ISBN 90 6196 116 5.

MCS 26.1 N.M. TEMME (red.), *Nonlinear Analysis, volume 1*, 1976. ISBN 90 6196 117 3.

MCS 26.2 N.M. TEMME (red.), *Nonlinear Analysis, volume 2*, ISBN 90 6196 121 1.

MCS 27 M. BAKKER, P.W. HEMKER, P.J. VAN DER HOUWEN, S.J. POLAK & M. VAN VELDHUIZEN, *Colloquium Discreteringmethoden*, 1976. ISBN 90 6196 124 6.

De met een * gemerkte uitgaven moeten nog verschijnen.