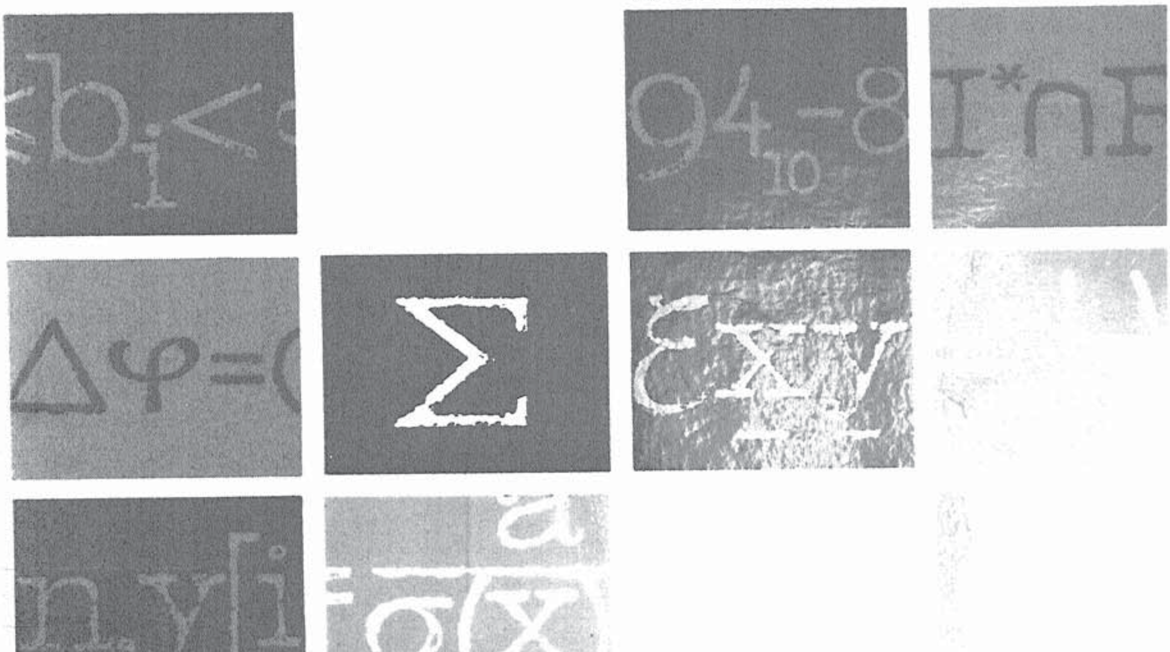
GRONDBEGRIPPEN VAN DE WAARSCHIJNLIJKHEIDSREKENING

J. FABIUS
W. R. VAN ZWET



M.C. SYLLABUS 10

M.C. SYLLABUS

10

M.C. SYLLABUS

10

NDBEGRIPPEN VAN DE WAARSCHIJNLIJKHEIDSREKENING

DOOR

J. FABIUS

W. R. VAN ZWET

SA

MATHEMATISCH CENTRUM AMSTERDAM

1970

VOORWOORD

Deze syllabus bevat de stof die wij in Leiden sinds 1965 als inleiding tot de colleges in de waarschijnlijkheidsrekening en de mathematische statistiek voor candidaten in de wiskunde doceren. Naar onze mening is dit de rijstebrijberg waar men zich doorheen moet eten om deze beide vakken te kunnen bestuderen op het niveau dat in de wiskunde aan onze universiteiten gebruikelijk is. Het behandelen van deze stof vergt drie uren per week gedurende één semester.

De voor de hand liggende voorbereiding op het bestuderen van deze syllabus is het afleggen van een candidaatsexamen in de wiskunde en het volgen van een college in de maat- en integratietheorie. Dit laatste is gewenst maar niet strikt noodzakelijk. In hoofdstuk 1 wordt een korte samenvatting gegeven van begrippen en stellingen uit de maat- en integratietheorie die in de waarschijnlijkheidsrekening een belangrijke rol spelen. Bewijzen worden hierbij vaak achterwege gelaten behalve waar het om stellingen gaat die niet in ieder college over maat- en integratietheorie worden behandeld.

Deze syllabus maakt geen enkele aanspraak op originaliteit. De titel is gestolen van Kolmogorov en de aanpak is grotendeels ontleend aan Loève. Alleen bij de behandeling van zwakke convergentie in § 2.9 hebben wij de voorkeur gegeven aan de fraaiere opzet à la Billingsley.

Gezien het doel van deze syllabus was de keuze van te bespreken onderwerpen voor ons geen probleem. Alleen die zaken worden behandeld die zowel in de waarschijnlijkheidsrekening als in de mathematische statistiek een belangrijke rol spelen.

Onze dank gaat uit naar de Raad van Beheer van het Mathematisch Centrum te Amsterdam die bereid was dit collegedictaat als MC syllabus uit te geven. Het manuscript werd getypt door mevrouw S.M.T. Hillebrand en mejuffrouw O.P. de Jong, de correctie werd uitgevoerd door de heren E.J. Sedoc en K.M. van Hee en de reproductie werd verzorgd door de heren D. Zwarst en J. Suiker. Wij zijn hun zeer erkentelijk voor hun nauwgezette en vlotte werk.

Leiden, november 1970

J. Fabius
W.R. van Zwet

INHOUD

1. MAAT- EN INTEGRATIETHEORIE

In dit hoofdstuk geven wij een samenvatting van de maat- en integratietheorie, voor zover nodig voor een goed begrip van de waarschijnlijkheidsrekening. Een uitgebreider behandeling is o.m. te vinden in de volgende boeken:

[1] Halmos, P.R., Measure Theory, D. Van Nostrand Company. Inc.
[2] Kingman, J.F.C., and Taylor, S.J., Introduction to measure and probability, Cambridge University Press.
[3] Loève, M., Probability Theory, D. Van Nostrand Company. Inc.
[4] Zaanen, A.C., Integration, North-Holland Publishing Company.

1.1. VERZAMELINGEN

Zij gegeven een niet-lege verzameling Ω, onze *ruimte*, die bestaat uit *elementen* of *punten* ω. Wij beschouwen in het volgende *deelverzamelingen* A, B, C, ... van Ω; hieronder valt ook de *lege verzameling* $\emptyset$ en de verzameling Ω zelf. Zoals gebruikelijk schrijven we $\omega \in A$ resp. $\omega \notin A$ als ω al dan niet een element van A is, en $A \subset B$ of $B \supset A$ als A een deelverzameling van B is, dwz. als ieder element van A ook in B ligt. We omschrijven deze situatie ook wel door te zeggen: A *is bevat in* B, of: B *bevat* of *omvat* A. Twee verzamelingen zijn *gelijk* als zij uit dezelfde elementen bestaan: A = B dan en dan alleen als $A \subset B$ en $B \subset A$. Een verzameling heet eindig indien hij uit een eindig aantal elementen bestaat.

Als $P(\omega)$ voor ieder punt $\omega \in \Omega$ een of andere bewering is, dan schrijven we $\{\omega: P(\omega)\}$ voor de verzameling van alle $\omega \in \Omega$, waarvoor $P(\omega)$ waar is. Als bv. $\Omega = R^1$, dan geldt $\{\omega: a < \omega \leq b\} = (a,b]$ voor willekeurige reële $a < b$. De verzameling die uit de punten $\omega_1, \omega_2, \ldots, \omega_n$ bestaat geven we aan met $\{\omega_1, \omega_2, \ldots, \omega_n\}$; de verzameling die uit één enkel punt ω bestaat met $\{\omega\}$.

De gebruikelijke verzamelingstheoretische operaties kunnen wij als volgt definiëren:

Vereniging:

$A \cup B$ = de verzameling van alle $\omega \in \Omega$, die tot A of B of beide behoren.

$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$ = de verzameling van alle $\omega \in \Omega$, die tot minstens één der verzamelingen $A_1, A_2, \ldots, A_n$ behoren.

$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \ldots$ = de verzameling van alle $\omega \in \Omega$, die tot minstens één der verzamelingen $A_1, A_2, \ldots$ behoren.

$\bigcup_{t \in T} A_t$ = de verzameling van alle $\omega \in \Omega$, met de eigenschap dat $\omega \in A_t$ voor minstens één $t \in T$. Hierbij is de *indexverzameling* T een willekeurige verzameling, niet noodzakelijk bevat in Ω, al dan niet eindig, al dan niet aftelbaar.

Doorsnede:

$A \cap B = AB$ = de verzameling van alle $\omega \in \Omega$, die tot A en B behoren.

$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n = A_1 A_2 \ldots A_n$ = de verzameling van alle $\omega \in \Omega$ die tot elk van de verzamelingen $A_1, A_2, \ldots, A_n$ behoren.

$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \ldots = A_1 A_2 \ldots$ = de verzameling van alle $\omega \in \Omega$, die tot elk van de verzamelingen $A_1, A_2, \ldots$ behoren.

$\bigcap_{t \in T} A_t$ = de verzameling van alle $\omega \in \Omega$ met de eigenschap dat $\omega \in A_t$ voor alle $t \in T$. Ook hier is T een willekeurige indexverzameling.

Twee verzamelingen heten *disjunct* als zij geen elementen gemeen hebben: A en B zijn disjunct dan en dan alleen als $AB = \emptyset$.

Complement:

$A^c = \{\omega : \omega \notin A\}$ = de verzameling van alle $\omega \in \Omega$ die niet tot A behoren.

Verschil:

$A - B = AB^c$ = de verzameling van alle $\omega \in \Omega$ die tot A maar niet tot B behoren. In het bijzonder geldt dus $A^c = \Omega - A$.

Symmetrisch Verschil:

$A \Delta B = (A-B) \cup (B-A) = (A \cup B) - AB$ = de verzameling van alle $\omega \in \Omega$ die tot precies één van de twee verzamelingen A en B behoren.

Uit deze definities kan men een aantal rekenregels afleiden. Zo blijken de operaties $\cup$ en $\cap$ commutatief, associatief en distributief ten opzichte van elkaar te zijn:

$A \cup B = B \cup A$; $\qquad A \cap B = B \cap A$;

$A \cup (B \cup C) = (A \cup B) \cup C$; $\qquad A \cap (B \cap C) = (A \cap B) \cap C$;

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $\qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Bovendien geldt

$$A \cup \emptyset = A;\ A \cup \Omega = \Omega;\ A \cap \emptyset = \emptyset;\ A \cap \Omega = A;$$

$$A \subset B \Longleftrightarrow A \cap B = A \Longleftrightarrow A \cup B = B.$$

Voor complementen geldt

$$(A^c)^c = A;\ \emptyset^c = \Omega;\ \Omega^c = \emptyset;$$

$$A \subset B \Longleftrightarrow B^c \subset A^c,$$

en voor het complement van een vereniging of doorsnede vindt men

$$(A \cup B)^c = A^c \cap B^c;\ (A \cap B)^c = A^c \cup B^c;$$

of, algemeen voor een willekeurige indexverzameling T,

$$(\bigcup_{t \in T} A_t)^c = \bigcap_{t \in T} A_t^c;\ (\bigcap_{t \in T} A_t)^c = \bigcup_{t \in T} A_t^c.$$

Een eindige of aftelbare vereniging kan men op de volgende manier altijd schrijven als een *disjuncte vereniging*, dwz. als een vereniging van paarsgewijs disjuncte verzamelingen:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup (A_1^c A_2) \cup (A_1^c A_2^c A_3) \cup \ldots \cup (A_1^c A_2^c \ldots A_{n-1}^c A_n),$$

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_1^c A_2) \cup (A_1^c A_2^c A_3) \cup \ldots .$$

Voor een oneindige rij verzamelingen $A_1, A_2, \ldots$ definiëert men

$$\lim\sup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m;\quad \lim\inf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

lim sup A_n bestaat derhalve uit alle $\omega \in \Omega$, die tot oneindig veel der verzamelingen A_n behoren, en lim inf A_n bestaat uit alle $\omega \in \Omega$ die tot bijna alle verzamelingen A_n, dwz. alle verzamelingen A_n op hoogstens een eindig getal na, behoren. In het algemeen zal dus lim inf $A_n \subset$ lim sup A_n. Als deze twee verzamelingen gelijk zijn noemt men de rij verzamelingen $A_1, A_2, \ldots$ *convergent* met

$$\lim A_n = \lim\sup A_n = \lim\inf A_n.$$

Een rij verzamelingen $A_1, A_2, \ldots$ heet (monotoon) *stijgend* als $A_1 \subset A_2 \subset A_3 \subset \ldots$ en (monotoon) *dalend* als $A_1 \supset A_2 \supset A_3 \supset \ldots$. Iedere monotone rij verzamelingen is convergent met $\lim A_n = \bigcup_{n=1}^{\infty} A_n$ als de rij stijgend is en $\lim A_n = \bigcap_{n=1}^{\infty} A_n$ als de rij daalt.

Een nuttig hulpmiddel bij het werken met verzamelingen is het begrip *indicatorfunctie* (ook: *karakteristieke functie*). De indicatorfunctie I_A van een verzameling $A \subset \Omega$ is een reële functie op Ω, gedefinieerd door

$$I_A(\omega) = \begin{cases} 1 & \text{als } \omega \in A, \\ 0 & \text{als } \omega \notin A. \end{cases}$$

Met behulp van zulke indicatorfuncties kan men alle verzamelingstheoretische relaties en operaties herleiden tot arithmetische

$$A \subset B \Longleftrightarrow I_A \leqq I_B; \qquad A = B \Longleftrightarrow I_A = I_B;$$

$$I_{AB} = \min(I_A, I_B) = I_A . I_B;$$

$$I_{A\cup B} = \max(I_A, I_B) = 1 - (1 - I_A)(1 - I_B) = I_A + I_B - I_{AB};$$

$$I_{\liminf A_n} = \liminf I_{A_n}; \; I_{\limsup A_n} = \limsup I_{A_n}.$$

Opgaven:

1. Bewijs voor willekeurige indexverzamelingen S en T:

$$A \cap \left(\bigcup_{t\in T} B_t \right) = \bigcup_{t\in T} (AB_t); \quad A \cup \left(\bigcap_{t\in T} B_t \right) = \bigcap_{t\in T} (A\cup B_t);$$

$$\left(\bigcup_{s\in S} A_s \right) \cap \left(\bigcup_{t\in T} B_t \right) = \bigcup_{s\in S} \bigcup_{t\in T} A_s B_t = \bigcup_{t\in T} \bigcup_{s\in S} A_s B_t;$$

$$\left(\bigcap_{s\in S} A_s \right) \cup \left(\bigcap_{t\in T} B_t \right) = \bigcap_{s\in S} \bigcap_{t\in T} (A_s \cup B_t) = \bigcap_{t\in T} \bigcap_{s\in S} (A_s \cup B_t).$$

2. Bewijs:

$$A - B = A - AB;$$

$$A \,\Delta\, B = B \,\Delta\, A;$$

$$(A \,\Delta\, B) \,\Delta\, C = A \,\Delta\, (B \,\Delta\, C);$$

$$A \,\Delta\, A = \emptyset; \; A \,\Delta\, A^c = \Omega; \; A \,\Delta\, \emptyset = A; \; A \,\Delta\, \Omega = A^c;$$

$$I_{A\Delta B} = \left| I_A - I_B \right| = I_A + I_B \pmod 2.$$

3. Bewijs:

$$\liminf A_n^c = (\limsup A_n)^c;$$

$$\limsup A_n^c = (\liminf A_n)^c.$$

4. Zij $\Omega = R^1$, $A_n = (a_n, b_n]$, $n = 1,2,\dots$. Bepaal lim inf A_n en lim sup A_n voor ieder van de volgende gevallen:

$a_n = n$, $b_n = n + 1$;

$a_n = -n$, $b_n = +n$;

$a_n = \frac{(-1)^n}{n}$, $b_n = 1 + \frac{1}{n}$;

$a_n = \frac{1}{n}$, $b_n = 1 - \frac{1}{n}$.

5. Zij $B_1 = A_1$ en $B_{n+1} = B_n \Delta A_{n+1}$ voor $n = 1,2,\dots$. Bewijs dat de rij B_1, $B_2,\dots$ dan en dan alleen convergeert als $\lim A_n = \emptyset$.

1.2. ALGEBRA'S EN σ-ALGEBRA'S

Een verzameling, waarvan de elementen zelf verzamelingen zijn, noemen wij een *klasse*. De klasse $M(A)$ van alle deelverzamelingen van een gegeven verzameling A heet de *machtsverzameling* van A. Iedere klasse van deelverzamelingen van Ω is dus bevat in $M(\Omega)$ en iedere deelklasse van $M(\Omega)$ is een klasse, die uit deelverzamelingen van Ω bestaat.

Een niet-lege klasse $F \subset M(\Omega)$ heet *algebra* (van Boole) als

(1.2.1) $\quad A \in F \Longrightarrow A^c \in F,$

(1.2.2) $\quad A,\ B \in F \Longrightarrow A \cup B \in F.$

Hieruit volgt dat iedere algebra de verzamelingen Ω en $\emptyset$ als elementen bevat. Daar $AB = (A^c \cup B^c)^c$ volgt ook

(1.2.3) $\quad A,\ B \in F \Longrightarrow AB \in F,$

en herhaalde toepassing van (1.2.2) en (1.2.3) geeft

(1.2.4) $\quad A_1, A_2, \dots, A_n \in F \Longrightarrow \bigcup_{i=1}^{n} A_i \in F,\ \bigcap_{i=1}^{n} A_i \in F.$

Een algebra is dus een niet-lege klasse die afgesloten is onder het vormen van complementen en eindige verenigingen en doorsneden.

Een niet-lege klasse $A \subset M(\Omega)$ heet een *σ-algebra* (van Boole) als

(1.2.5) $\quad A \in A \Longrightarrow A^c \in A,$

(1.2.6) $\quad A_1, A_2, \ldots \in \mathcal{A} \Longrightarrow \bigcup_{n=1}^{\infty} A_i \in \mathcal{A}.$

Evenals boven volgt dat iedere σ-algebra Ω en ∅ als elementen bevat en, daar $\bigcap A_i = (\bigcup A_i^c)^c$,

(1.2.7) $\quad A_1, A_2, \ldots \in \mathcal{A} \Longrightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$

Nemen wij $A_i = A_n$ voor $i \geq n$, dan gaan (1.2.6) en (1.2.7) over in (1.2.4) met $\mathcal{F}$ vervangen door $\mathcal{A}$. Een σ-algebra is dus een niet-lege klasse, die afgesloten is onder het vormen van complementen en eindige of aftelbare verenigingen en doorsneden.

Uiteraard is $\mathcal{M}(\Omega)$ een σ-algebra en dus ook een algebra. Verder is iedere doorsnede van (σ-)algebra's weer een (σ-)algebra. Hieruit volgt dat er bij iedere klasse $\mathcal{C} \subset \mathcal{M}(\Omega)$ een unieke minimale algebra $\mathcal{F}$ en een unieke minimale σ-algebra $\mathcal{A}$ bestaan, die $\mathcal{C}$ omvatten. We noemen $\mathcal{F}(\mathcal{A})$ de (σ-)algebra *voortgebracht* door $\mathcal{C}$. $\mathcal{F}(\mathcal{A})$ is de doorsnede van alle (σ-)algebra's die $\mathcal{C}$ omvatten.

Als in een ruimte Ω een σ-algebra $\mathcal{A}$ is gegeven, dan noemt men het paar $(\Omega, \mathcal{A})$ een *meetbare ruimte*, en de elementen van $\mathcal{A}$ *meetbare verzamelingen*.

Voorbeelden

1.2.1. Zij $\mathcal{C} \in \mathcal{M}(\Omega)$ de klasse van alle éénpuntsverzamelingen $\{\omega\}$. De door $\mathcal{C}$ voortgebrachte algebra $\mathcal{F}$ bestaat dan uit alle deelverzamelingen van Ω, die òf zelf eindig zijn, òf een eindig complement hebben. De door $\mathcal{C}$ voortgebrachte σ-algebra $\mathcal{A}$ bestaat uit alle deelverzamelingen van Ω die òf zelf eindig of aftelbaar zijn, òf een eindig of aftelbaar complement hebben. Steeds zal $\mathcal{C} \subset \mathcal{F} \subset \mathcal{A} \subset \mathcal{M}(\Omega)$. Is Ω aftelbaar, dan geldt $\mathcal{A} = \mathcal{M}(\Omega)$, en is Ω eindig, dan geldt $\mathcal{F} = \mathcal{A} = \mathcal{M}(\Omega)$.

1.2.2. Zij $\Omega = R^1$, en zijn $\mathcal{C}^1$ de klasse van alle *cellen*, dwz. intervallen van de vorm $(a,b]$ met $-\infty \leq a \leq b \leq \infty$. Hierbij stellen we $(a,\infty] = (a,\infty)$ en $(a,a] = \emptyset$. De doorsnede van twee cellen is weer een cel, maar $\mathcal{C}^1$ is niet gesloten onder het vormen van complementen en verenigingen. De door $\mathcal{C}^1$ voortgebrachte algebra $\mathcal{F}^1$ bestaat uit alle verzamelingen die als vereniging van eindig veel cellen geschreven kunnen worden. Een dergelijke verzameling kan altijd als vereniging van eindig veel disjuncte cellen geschreven worden. De door $\mathcal{C}^1$ voortgebrachte σ-algebra $\mathcal{B}^1$ heet de *σ-algebra van Borel in* R^1, zijn elementen de *Borel-verzamelingen* in R^1. $\mathcal{B}^1$ bevat onder meer: alle intervallen, alle open verzamelingen, alle eindige of aftelbare verzamelingen. Zie bv. [1]

voor een voorbeeld van een deelverzameling van R^1 die niet tot $\mathcal{B}^1$ behoort.

Als $\mathcal{D}^1$ de klasse van alle intervallen van de vorm $(-\infty,a]$ met $-\infty \leq a \leq \infty$ is, dan is $\mathcal{D}^1$ bevat in $\mathcal{C}^1$. De door $\mathcal{D}^1$ voortgebrachte (σ-)algebra is dus ook bevat in de door $\mathcal{C}^1$ voortgebrachte (σ-)algebra. Daar echter elke cel $(a,b] = (-\infty,b] \cap (-\infty,a]^c$, moet de door $\mathcal{D}^1$ voortgebrachte (σ-)algebra zeker $\mathcal{C}^1$ omvatten. Derhalve brengen $\mathcal{D}^1$ en $\mathcal{C}^1$ dezelfde (σ-)algebra voort.

1.2.3. Zij $\Omega = R^k$. We schrijven de elementen van Ω als vectoren: $\omega = (\omega_1,\ldots,\omega_k)$, $a = (a_1,\ldots,a_k)$ etc. Verder definiëren we $\infty = (\infty,\ldots,\infty)$, $-a = (-a_1,\ldots,-a_k)$ en betekent $a \leq b$ dat $a_i \leq b_i$ voor $i = 1,\ldots,k$. Als nu voor $-\infty \leq a \leq b \leq \infty$ de *cel* $(a,b]$ gedefiniëerd is door

$$(a,b] = \{\omega\colon\ \omega_i \in (a_i,b_i],\ i = 1,\ldots,k\},$$

en $\mathcal{C}^k$ de klasse van alle cellen is, dan is de door $\mathcal{C}^k$ voortgebrachte algebra $\mathcal{F}^k$ net als boven de klasse van alle deelverzamelingen van Ω die als eindige vereniging, of - wat op hetzelfde neerkomt - als eindige disjuncte vereniging van cellen geschreven kunnen worden. De door $\mathcal{C}^k$ voortgebrachte σ-algebra $\mathcal{B}^k$ heet de *σ-algebra van Borel in* R^k, zijn elementen de k-*dimensionale Borelverzamelingen*.

Definiëren we verder $\mathcal{D}^k$ als de klasse van alle cellen van de vorm $(-\infty,a]$ met $-\infty \leq a \leq \infty$, dan kunnen we een willekeurige cel $(a,b]$ weer uitdrukken in elementen van $\mathcal{D}^k$. Daartoe voeren we punten $c_j = (c_{j_1},\ldots,c_{j_k})$ in, gedefiniëerd door

$$c_{j_i} = \begin{cases} b_i & \text{als} \quad i \neq j, \\ a_i & \text{als} \quad i = j, \end{cases} \qquad i,j = 1,2,\ldots,k.$$

Nu is gemakkelijk in te zien dat

$$(a,b] = (-\infty,b] \cap \Big(\bigcup_{j=1}^{k} (-\infty,c_j]\Big)^c.$$

De door $\mathcal{D}^k$ voortgebrachte (σ-)algebra moet $\mathcal{C}^k$ dus omvatten. Omdat echter $\mathcal{D}^k \subset \mathcal{C}^k$, volgt ook nu dat $\mathcal{F}^k$ en $\mathcal{B}^k$ zowel door $\mathcal{D}^k$ als door $\mathcal{C}^k$ worden voortgebracht.

Opgaven

1. Bepaal voor ieder van de volgende definities van $\mathcal{C} \subset \mathcal{M}(\Omega)$ de door $\mathcal{C}$ voortgebrachte algebra en σ-algebra:
   a) $\mathcal{C}$ bestaat uit een enkele niet-lege verzameling $A \subset \Omega$
   b) $\mathcal{C}$ bestaat uit alle verzamelingen die een gegeven verzameling A omvatten;
   c) $\mathcal{C}$ bestaat uit alle verzamelingen die in een gegeven verzameling A bevat zijn;
   d) $\mathcal{C}$ bestaat uit alle verzamelingen die tenminste één punt gemeen hebben met een gegeven verzameling A.

2. Bewijs dat iedere eindige algebra een σ-algebra is.

3. Bepaal de door $\mathcal{C}$ voortgebrachte algebra en σ-algebra als $\Omega = R^2$ en $\mathcal{C}$ bestaat uit:
   a) Alle verzamelingen die bevat zijn in een verticale lijn, d.w.z. $A \in \mathcal{C}$ dan en slechts dan als er een $a \in R^1$ is, zodanig, dat $x = a$ als $(x,y) \in A$;
   b) Alle verzamelingen van de vorm $\{(x,y) : x \in (a,b]\}$ met $-\infty \leqq a \leqq b \leqq \infty$.

1.3. PRODUCTRUIMTEN

Als $A_1, A_2, \ldots, A_k$ willekeurige, niet noodzakelijk tot één en dezelfde ruimte behorende verzamelingen zijn, dan is hun *productverzameling* de verzameling van alle geordende k-tallen $(\omega_1, \omega_2, \ldots, \omega_k)$ met $\omega_1 \in A_1$, $\omega_2 \in A_2, \ldots,$ $\omega_k \in A_k$:

$$\prod_{i=1}^{k} A_i = A_1 \times A_2 \times \ldots \times A_k = \{(\omega_1, \omega_2, \ldots, \omega_k) : \omega_i \in A_i,\ i = 1,2,\ldots,k\}.$$

Als $\Omega_1, \Omega_2, \ldots, \Omega_k$ gegeven ruimten zijn, dan noemen wij $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_k$ hun *productruimte*. Nemen wij $A_i, B_i \subset \Omega_i$ voor $i = 1,2,\ldots,k$, dán zijn $A_1 \times A_2 \times \ldots \times A_k$ en $B_1 \times B_2 \times \ldots \times B_k$ deelverzamelingen van deze productruimte en

$$(1.3.1) \qquad (A_1 \times A_2 \times \ldots \times A_k) \cap (B_1 \times B_2 \times \ldots \times B_k) = (A_1 \cap B_1) \times (A_2 \cap B_2) \times \ldots \times (A_k \cap B_k),$$

(1.3.2.) $(A_1\times A_2\times\ldots\times A_k)^c = (A_1^c\times\Omega_2\times\ldots\times\Omega_k)\cup(A_1\times A_2^c\times\Omega_3\times\ldots\times\Omega_k)\cup\ldots$

$$\ldots\cup(A_1\times\ldots\times A_{k-2}\times A_{k-1}^c\times\Omega_k)\cup(A_1\times\ldots\times A_{k-1}\times A_k^c).$$

Uiteraard is niet iedere deelverzameling van Ω een productverzameling.

Als nu $F_1,F_2,\ldots,F_k$ algebra's van deelverzamelingen van resp. $\Omega_1,\Omega_2,\ldots,\Omega_k$ zijn, dan volgt uit het bovenstaande dat de klasse van alle eindige (disjuncte) verenigingen van verzamelingen van de vorm $A_1\times A_2\times\ldots\times A_k$ met $A_i\in F_i$, $i = 1,2,\ldots,k$, een algebra van deelverzamelingen van de productruimte $\Omega_1\times\Omega_2\times\ldots\times\Omega_k$ is. We noemen deze algebra de *productalgebra* van $F_1,F_2,\ldots,F_k$. De productalgebra van k σ-algebra's $A_1,A_2,\ldots,A_k$ in resp. $\Omega_1,\Omega_2,\ldots,\Omega_k$ is niet noodzakelijk een σ-algebra. De σ-algebra die door deze productalgebra wordt voortgebracht heet de *product-σ-algebra* $A_1\times A_2\times\ldots\times A_k$. De meetbare ruimte $(\Omega_1\times\ldots\times\Omega_k,\ A_1\times\ldots\times A_k)$ heet de *meetbare productruimte* van de meetbare ruimten (Ω_1,A_1) $(\Omega_2,A_2),\ldots,(\Omega_k,A_k)$.

Het voorgaande kan gemakkelijk worden gegeneraliseerd tot producten met oneindig veel "factoren". Wij beperken ons daarbij tot aftelbare producten: De *productverzameling* $A_1\times A_2\times\ldots$ is per definitie de verzameling van alle oneindige rijen $\omega = (\omega_1,\omega_2,\ldots)$ met $\omega_i\in A_i$ voor $i = 1,2,\ldots$.
Het product $\Omega_1\times\Omega_2\times\ldots$ van een rij gegeven ruimten $\Omega_1,\Omega_2,\ldots$ noemen wij weer hun *productruimte*. De verzamelingen van de vorm $A\times\Omega_{k+1}\times\ldots$ met $A\subset\Omega_1\times\Omega_2\times\ldots\times\Omega_k$ en $k < \infty$ noemen wij de *cylinderverzamelingen* in $\Omega_1\times\Omega_2\times\ldots$. Als hierbij de verzameling A, de *basis* van de cylinderverzameling, zelf een productverzameling $A_1\times A_2\times\ldots\times A_k$ met $A_1\in\Omega_1$, $A_2\in\Omega_2,\ldots,$ $A_k\in\Omega_k$ is, dan spreken wij van een *productcylinderverzameling* met *zijden* $A_1,A_2,\ldots,A_k$. Laten nu $F_1,F_2,\ldots$ algebra's in resp. $\Omega_1,\Omega_2,\ldots$ zijn. Dan is de klasse van alle eindige (disjuncte) verenigingen van productcylinders $A_1\times A_2\times\ldots\times A_k\times\Omega_{k+1}\times\Omega_{k+2}\times\ldots$ met zijden $A_i\in F_i$, $i = 1,2,\ldots,k$, $k < \infty$, een algebra in de productruimte $\Omega_1\times\Omega_2\times\ldots$: de *productalgebra* van $F_1,F_2,\ldots$. Deze productalgebra kan ook beschreven worden als de klasse van alle cylinderverzamelingen $A\times\Omega_{k+1}\times\Omega_{k+2}\times\ldots$ waarvan de basis tot de productalgebra van $F_1,F_2,\ldots,F_k$ behoort. De productalgebra van een rij σ-algebra's A_i in Ω_i, $i = 1,2,\ldots,$ is in het algemeen geen σ-algebra. De σ-algebra die door deze product algebra wordt voortgebracht is de *product-σ-algebra* $A_1\times A_2\times\ldots$, en $(\Omega_1\times\Omega_2\times\ldots,A_1\times A_2\times\ldots)$ is de *meetbare productruimte* van (Ω_1,A_1), $(\Omega_2,A_2),\ldots$.

Voorbeelden

1.3.1. Zij $\Omega_i = \mathrm{R}^1$, $D_i = D^1$, $C_i = C^1$, $F_i = F^1$, $A_i = B^1$ voor $i = 1,2,\ldots,k$ (Zie voorbeeld 1.2.2. en 1.2.3. voor de notatie).

Dan volgt:

$\Omega_1 \times \Omega_2 \times \ldots \times \Omega_k = \mathrm{R}^k$;

$D^k = \{A_1 \times A_2 \times \ldots \times A_k \colon A_i \in D^1,\ i = 1,\ldots,k\}$;

$C^k = \{A_1 \times A_2 \times \ldots \times A_k \colon A_i \in C^1,\ i = 1,\ldots,k\}$;

F^k is de productalgebra van $F_1, F_2, \ldots, F_k$ en wordt voortgebracht zowel door D^k als door C^k, en ook door de klasse $\{A_1 \times A_2 \times \ldots \times A_k \colon A_i \in F^1,\ i = 1,\ldots,k\}$.

$B^k = A_1 \times A_2 \times \ldots \times A_k$ en wordt voortgebracht door ieder van de hierboven genoemde klassen en ook door de klassen $\{A_1 \times A_2 \times \ldots \times A_k \colon A_i \in B^1,\ i = 1,\ldots,k\}$.

1.3.2. Zij $\Omega_i = \mathrm{R}^1$, $D_i = D^1$, $C_i = C^1$, $F_i = F^1$, $A_i = B^1$ voor $i = 1,2,\ldots$

De productruimte, die bestaat uit alle oneindige rijen van reële getallen noemen wij R^∞. Voorts definiëren we:

D^∞: de klasse van alle productcylinders in R^∞ met zijden in D^1.

C^∞: de klassen van alle productcylinders in R^∞ met zijden in C^1.

F^*: de klasse van alle productcylinderverzamelingen in R^∞ met zijden in F^1.

F^∞: de productalgebra van $F_1, F_2, \ldots$. F^∞ is de klasse van alle cylinderverzamelingen in R^∞ met basis in F^k, $k < \infty$ en al deze cylinderverzamelingen kunnen geschreven worden als eindige (disjuncte) verenigingen van verzamelingen uit F^*.

B^*: de klasse van alle productcylinderverzamelingen in R^∞ met zijden in B^1.

B^+: de productalgebra van $A_1, A_2, \ldots$. B^+ is geen σ-algebra en bestaat uit alle cylinderverzamelingen in R^∞ met basis in B^k, $k < \infty$.

B^∞: de *product-σ-algebra* van $A_1, A_2, \ldots$; We noemen B^∞ de *σ-algebra van Borel in* R^∞ en zijn elementen de *Borel-verzamelingen in* R^∞.

De productalgebra F^∞ wordt voortgebracht door ieder van de klassen D^∞, C^∞, F^*. De product-σ-algebra B^∞ wordt voortgebracht door ieder van de klassen D^∞, C^∞, F^*, F^∞, B^*, B^+.

1.3.3. Zij $\Omega_i = \{0,1\}$, $A_i = M(\Omega_i)$, $i = 1,2,\ldots$. De productruimte $\{0,1\}^k = \Omega_1\times\Omega_2\times\ldots\times\Omega_k$ bestaat uit alle geordende k-tallen $(\omega_1,\ldots,\omega_k)$ met $\omega_i = 0$ of 1, $i = 1,\ldots,k$, en $A_1\times A_2\times\ldots\times A_k = M(\Omega_1\times\Omega_2\times\ldots\times\Omega_k)$. De productruimte $\{0,1\}^\infty = \Omega_1\times\Omega_2\times\ldots$ bestaat uit alle oneindige rijen $\omega = (\omega_1,\omega_2,\ldots)$ met $\omega_i = 0$ of 1, $i = 1,2,\ldots$, en de productalgebra van $A_1,A_2,\ldots$ bestaat uit alle cylinderverzamelingen. Deze productalgebra is geen σ-algebra.

Opgaven

1. Stel dat $E = A\times B$, $E_1 = A_1\times B_1$ en $E_2 = A_2\times B_2$ niet-lege deelverzamelingen van $\Omega_1\times\Omega_2$ zijn. Bewijs dat $E = E_1\cup E_2$ en $E_1E_2 = \emptyset$ dan en slechts dan als òf $A = A_1\cup A_2$, $A_1A_2 = \emptyset$ en $B = B_1 = B_2$ òf $A = A_1 = A_2$, $B = B_1\cup B_2$ en $B_1B_2 = \emptyset$.
2. Zij A_i een (σ-)algebra van deelverzamelingen van Ω_i, die wordt voortgebracht door een klasse C_i, $i = 1,2,\ldots k$. Bewijs dat de product-(σ-)algebra wordt voortgebracht door de klasse $\{A_1\times A_2\times\ldots\times A_k\colon A_i\in C_i,\ i = 1,\ldots,k\}$.
3. Bewijs dat de σ-algebra van Borel B^k in R^k wordt voortgebracht door de klasse van de open verzamelingen.
4. Bewijs dat in de meetbare productruimte ($\Omega = \{0,1\}^\infty$, $A = A_1\times A_2\times\ldots$) van voorbeeld 1.3.3. de volgende verzamelingen meetbaar zijn:

$$\{\omega\colon \omega_i = 0 \text{ voor } i \geqq n\},$$

$$\{\omega\colon \sum_1^\infty \omega_i < \infty\},$$

$$\{\omega\colon \sum_1^\infty \omega_i\, 2^{-i} < \tfrac{1}{3}\},$$

$$\{\omega\colon \lim_{n\to\infty} \frac{1}{n} \sum_1^n \omega_i = \tfrac{1}{2}\}.$$

1.4. MEETBARE FUNCTIES

Zij f een functie op een ruimte Ω_1 met waarden in een tweede ruimte Ω_2. Het *inverse beeld* $f^{-1}(A)$ van een verzameling $A\subset\Omega_2$ is per definitie

$\{\omega_1 : f(\omega_1) \in A\}$. Gemakkelijk is in te zien dat f^{-1} alle verzamelingstheoretische operaties bewaart:

$$f^{-1}(A^c) = (f^{-1}(A))^c \qquad , \quad A \subset \Omega_2;$$

$$f^{-1}(\bigcup_{t\in T} A_t) = \bigcup_{t\in T} f^{-1}(A_t) \qquad , \quad A_t \in \Omega_2,\ t \in T;$$

$$f^{-1}(\bigcap_{t\in T} A_t) = \bigcap_{t\in T} f^{-1}(A_t) \qquad , \quad A_t \in \Omega_2,\ t \in T;$$

voor een willekeurige, eindige of oneindige indexverzameling T. Wij beschouwen functies op meetbare ruimte (Ω_1, A_1) met waarden in een tweede meetbare ruimte (Ω_2, A_2). Een dergelijke functie f heet *A_1-A_2-meetbaar*, of kortweg *meetbaar*, als $f^{-1}(A) \in A_1$ voor alle $A \in A_2$. Definiëren wij de door f in Ω_1 *geïnduceerde σ-algebra* $f^{-1}(A_2) = \{f^{-1}(A) : A \in A_2\}$ (dit is een σ-algebra!), dan kunnen wij ook zeggen dat f dan en slechts dan meetbaar is als $f^{-1}(A_2) \subset A_1$. Wanneer aan deze eis voldaan is voor een functie f die slechts op een (meetbare) deelverzameling Ω_1' van Ω_1, gedefinieerd is, dan zeggen we dat f meetbaar is op Ω_1'. Als de σ-algebra A_2 wordt voortgebracht door een klasse C_2, dan is de eis dat $f^{-1}(C_2) = \{f^{-1}(B) : B \in C_2\} \subset A_1$ nodig en voldoende voor de meetbaarheid van f: Nodig omdat $f^{-1}(C_2) \subset f^{-1}(A_2)$ en voldoende omdat de klasse $\{A \subset \Omega_2 : f^{-1}(A) \in A_1\}$ een σ-algebra is en C_2, en derhalve ook A_2, omvat. Als (Ω_i, A_i), $i = 1,2,3$, meetbare ruimten zijn en $f : \Omega_1 \to \Omega_2$ en $g : \Omega_2 \to \Omega_3$ resp. $A_1 - A_2$ - en $A_2 - A_3$ - meetbare functies zijn, dan is ook de *samengestelde functie* $g(f)$, gedefinieerd door $g(f)(\omega_1) = g(f(\omega_1))$ voor $\omega_1 \in \Omega_1$, meetbaar, en wel $A_1 - A_3$ - meetbaar.

Indien Ω_1 of Ω_2 de ruimte $R^k (1 \leq k \leq \infty)$ is, dan zal meetbaarheid in het hierna volgende zonder nadere aanduiding steeds betrekking hebben op de σ-algebra van Borel B^k. Daar B^k wordt voortgebracht door D^k volgt uit het voorafgaande:

Een reële functie f op een meetbare ruimte (Ω, A) is dan en slechts dan meetbaar als $\{\omega : f(\omega) \leq a\} \in A$ voor alle $a \in R^1$; daar steeds $\emptyset \in A$ en $\Omega \in A$, is de functie $f \equiv a$, $a \in R^1$, meetbaar ongeacht de keuze van A. Een functie f op een meetbare ruimte (Ω, A) met waarden $f(\omega) = (f_1(\omega), \ldots, f_k(\omega)) \in R^k$ $(k < \infty)$ of met waarden $f(\omega) = (f_1(\omega), f_2(\omega), \ldots) \in R^\infty$ is dan en slechts dan meetbaar als $\{\omega : f_i(\omega) \leq a\} \in A$ voor alle $a \in R^1$ in alle i, d.w.z. als ieder der reële functies f_i meetbaar is; de functie $f \equiv a$, $a \in R^k$ ($1 \leq k \leq \infty$) is steeds meetbaar. Een meetbare functie op R^m met waarden in R^k noemen wij een *Borelfunctie*. Voor een continue functie f op R^m met waarden in R^k is de verzameling $\{\omega\colon f_i(\omega) \leq a\}$ voor alle $a \in R^1$ en alle i gesloten; daar B^1 de gesloten verzamelingen in R^1 bevat, zijn alle continue functies Borelfuncties.

Wil men toelaten dat reële functies ook de waarden ∞ en $-\infty$ kunnen aannemen, dan dient men in het voorgaande R^1 overal te vervangen door zijn afsluiting $\bar{R}^1 = R^1 \cup \{\infty\} \cup \{-\infty\}$. In plaats van B^1 beschouwt men dan $\bar{B}^1$, de *σ-algebra van Borel in* $\bar{R}^1$, gedefiniëerd als de minimale σ-algebra die zowel B^1 als de beide verzamelingen $\{\infty\}$ en $\{-\infty\}$ bevat. $\bar{B}^1$ wordt dus voortgebracht door de klasse $\bar{D}^1$ van alle intervallen van de vorm $[-\infty,a]$ met $a \in R^1$. Geheel analoog kan men ook $\bar{R}^k$ en $\bar{B}^k$ definiëren. Wij merken nog op dat in $\bar{R}^1$ voor de elementen ∞ en $-\infty$ de volgende rekenregels die met limietovergangen corresponderen, gelden:

$$a \pm \infty = a \pm \lim_{x\to\infty} x = \pm\infty \qquad \text{voor } -\infty < a < \infty;$$

$$\infty + \infty = \infty \quad , \qquad -\infty - \infty = -\infty;$$

$$a.\infty = -a.(-\infty) = a \lim_{x\to\infty} x = \begin{cases} \infty \text{ voor } 0 < a < \infty \\ 0 \text{ voor } a = 0 \\ -\infty \text{ voor } -\infty < a < 0; \end{cases}$$

$$\infty.\infty = (-\infty).(-\infty) = \infty, \qquad -\infty \,.\, \infty = -\infty \;;$$

$$\frac{a}{\pm\infty} = \pm\, a \lim_{x\to\infty} \frac{1}{x} = 0 \qquad \text{voor } -\infty < a < \infty.$$

Met name geldt dus $0.\infty = 0$, doch $\infty-\infty$ en $\frac{\infty}{\infty}$ zijn niet gedefiniëerd.

Beschouw twee eindige reële meetbare functies f en g op een meetbare ruimte (Ω,A). De functies $h_1(x,y) = x+y$ en $h_2(x,y) = xy$ van de reële variabelen x en y zijn continu en dus Borelfuncties. Derhalve zijn de samengestelde functies $h_1(f,g) = f+g$ en $h_2(f,g) = f.g$ meetbaar. Evenzo is ook f/g meetbaar op de verzameling $\{\omega\colon g(\omega) \neq 0\}$ waar deze functie is gedefiniëerd. Indien wij voor f, g, f+g, f.g en f/g de waarden ∞ en $-\infty$ toelaten, dan zijn f+g, f.g en f/g weliswaar niet noodzakelijk overal (c.q. op $\{\omega\colon g(\omega) \neq 0\}$) gedefiniëerd, doch onze conclusies omtrent meetbaarheid blijven gehandhaafd.

Voor een rij reële meetbare functies f_1, f_2, ... op een meetbare ruimte $(\Omega,\mathcal{A})$ geldt

$$\{\omega: \sup_n f_n(\omega) \leq a\} = \bigcap_{n=1}^{\infty} \{\omega: f_n(\omega) \leq a\} \in \mathcal{A}.$$

zodat $\sup f_n$, en dus ook $\inf f_n = - \sup - f_n$, meetbaar zijn. Hetzelfde geldt voor $\lim_n \sup f_n$ en $\lim_n \inf f_n$, daar

$$\lim_n \sup f_n = \inf_n \sup_{m \geq n} f_m.$$

De verzameling C van alle $\omega \in \Omega$ waarvoor $\lim_{n\to\infty} f_n(\omega)$ bestaat is meetbaar, daar $\overline{f} = \lim \sup f_n$ en $\underline{f} = \lim \inf f_n$ meetbaar zijn en dus

$$C = \{\omega: \overline{f}(\omega) = \underline{f}(\omega) = \infty\} \cup \{\omega: \overline{f}(\omega) = \underline{f}(\omega) = -\infty\} \cup \{\omega: \overline{f}(\omega) - \underline{f}(\omega) = 0\} \in \mathcal{A}.$$

De functie $f = \lim_{n\to\infty} f_n$, gedefinieerd op C, is daar gelijk aan $\overline{f}$ en dus meetbaar op C. Ook $\sum f_n$ is dus meetbaar op de verzameling waar deze functie gedefinieerd is.

Iedere meetbare reële functie f op een meetbare ruimte $(\Omega,\mathcal{A})$ die slechts eindig veel waarden aanneemt, is van de vorm

$f = \sum_{i=1}^{n} a_i I_{A_i}$ met $A_1, \ldots, A_n$ disjunct, $A_i \in \mathcal{A}$ voor $i = 1, 2, \ldots, n$ en $n < \infty$. Wij noemen zulke functies *elementaire functies*. Iedere meetbare niet-negatieve functie f is de puntsgewijze limiet voor een niet-dalende rij niet-negatieve elementaire functies f_n. Men kan bijvoorbeeld

$$f_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{\{\omega: \frac{k-1}{2^n} < f(\omega) \leq \frac{k}{2^n}\}}$$

kiezen. Tenslotte is iedere meetbare reële functie f te schrijven als het verschil $f^+ - f^-$ van twee niet-negatieve meetbare functies f^+ en f^-, waarbij

$$f^+(\omega) = \max(f(\omega),0), \; f^-(\omega) = - \min(f(\omega),0).$$

We noemen f^+ en f^- het *positieve* resp. *negatieve deel* van f.

Stelling 1.4.1

Zij (Ω_1,A_1) een meetbare ruimte met de eigenschap dat A_1 de σ-algebra is, die in Ω_1 wordt geïnduceerd door een functie g op Ω_1 met waarden in een meetbare ruimte (Ω_2,A_2): $A_1 = g^{-1}(A_2)$. Dan is een reële functie f op (Ω_1,A_1) dan en slechts dan meetbaar als er een meetbare reële functie h op (Ω_2,A_2) is, zodanig dat $f = h(g)$.

Bewijs:

Als er een dergelijke functie h is, dan geldt voor iedere Borelverzameling $B \in B^1$:

$$f^{-1}(B) = g^{-1}(h^{-1}(B)) \in A_1,$$

wegens de meetbaarheid van g en h, en dus is f meetbaar. Zij nu gegeven dat f meetbaar is. We dienen dan de existentie van een functie h met de bovengenoemde eigenschappen aan te tonen. We beschouwen daartoe de volgende gevallen:

a) $f = I_A$, een indicatorfunctie, noodzakelijk met $A \in A_1$. Daar $A_1 = g^{-1}(A_2)$, is er een $B \in A_2$, zodanig dat $A = g^{-1}(B)$. Maar dit betekent dat $f(\omega_1) = I_B(g(\omega_1))$ zodat I_B een functie met de gezochte eigenschappen is.

b) $f = \sum_1^n a_i I_{A_i}$, een elementaire functie. Op grond van a) weten wij dat er meetbare functies h_i op Ω_2 zijn, zodanig dat $I_{A_i} = h_i(g)$, $i = 1,2,\dots,n$. Maar dan is ook de functie $h = \sum_1^n a_i h_i$ meetbaar op Ω_2 en $f = h(g)$.

c) f is niet-negatief. Er zijn dan elementaire functies f_n zodanig dat $f(\omega_1) = \lim_{n\to\infty} f_n(\omega_1)$, $\omega_1 \in \Omega_1$. Voorts is $f_n = h_n(g)$ met h_n meetbaar op Ω_2. Derhalve bestaat $\lim_{n\to\infty} h_n(\omega_2)$ voor alle $\omega_2 \in \{f(\omega_1): \omega_1 \in \Omega_1\}$. Definiëren wij nu

$$h(\omega_2) = \begin{cases} \lim_{n\to\infty} h_n(\omega_2) & \text{als deze limiet bestaat,} \\ 0 & \text{anders,} \end{cases}$$

dan is h meetbaar op de hele ruimte Ω_2, en $f = h(g)$.

d) Door c) tenslotte toe te passen op f^+ en f^- volgt de bewering voor een willekeurige meetbare f.

Opgaven

1. Bewijs dat een functie $f = (f_1, f_2, \ldots)$ op een meetbare ruimte $(\Omega, \mathcal{A})$ met waarden in een meetbare productruimte $(\Omega_1 \times \Omega_2 \times \ldots, \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots)$ dan en slechts dan meetbaar is als $f_i : (\Omega, \mathcal{A}) \to (\Omega_i, \mathcal{A}_i)$ meetbaar is voor $i = 1, 2, \ldots$.
2. Als f een continue reële functie op R^1 is met $\lim_{x \to \infty} f(x) = \lim_{x \to -\infty} f(x) = 0$, dan is er voor iedere $\varepsilon > 0$ een elementaire functie $g = \sum_1^n a_i I_{A_i}$ met de eigenschap dat de A_i eindige disjuncte intervallen zijn en dat $|f(x) - g(x)| < \varepsilon$ voor alle $x \in R^1$.

1.5. MATEN

Een *maat* μ is een reële functie op een klasse $\mathcal{C}$ van deelverzamelingen van een ruimte Ω met de eigenschappen

(i) $\mu(A)$ is eindig voor tenminste één $A \in \mathcal{C}$;

(ii) $\mu(A) \geq 0$ voor alle $A \in \mathcal{C}$;

(iii) μ is *σ-additief*, dwz. $\mu(\bigcup_1^\infty A_n) = \sum_1^\infty \mu(A_n)$ voor iedere disjuncte rij $A_1, A_2, \ldots \in \mathcal{C}$ waarvoor ook $\bigcup_1^\infty A_n \in \mathcal{C}$.

De maat μ heet *eindig* als $\mu(A)$ eindig is voor alle $A \in \mathcal{C}$, en *σ-finiet* als er verzamelingen $A_1, A_2, \ldots \in \mathcal{C}$ zijn, zodanig, dat $\Omega = \bigcup A_n$ en $\mu(A_n) < \infty$ voor alle n.

In veel gevallen blijkt het mogelijk een gegeven maat op een klasse $\mathcal{C}$ voort te zetten tot een maat op de door $\mathcal{C}$ voortgebrachte σ-algebra.

Stelling 1.5.1. (Carathéodory)

Als m een maat is op een algebra $\mathcal{F} \subset \mathcal{M}(\Omega)$, dan kan m worden uitgebreid tot een maat μ op de door $\mathcal{F}$ voortgebrachte σ-algebra $\mathcal{A}$ (d.w.z. er bestaat een maat μ op $\mathcal{A}$ met $\mu(A) = m(A)$ voor alle $A \in \mathcal{F}$). Als m σ-finiet is, dan is deze voortzetting uniek, en is er voor iedere $\varepsilon > 0$ en iedere $A \in \mathcal{A}$ met $\mu(A) < \infty$ een $B \in \mathcal{F}$ zodanig, dat $\mu(A \Delta B) < \varepsilon$, hetgeen inhoudt dat $|\mu(A) - \mu(B)| < \varepsilon$.

In het volgende beschouwen wij uitsluitend maten op σ-algebra's. Als

(Ω,A) een meetbare ruimte en μ een (eindige, σ-finiete) maat op A is, dan heet (Ω,A,μ) een (eindige, σ-finiete) *maatruimte*. Een verzameling $A \in A$ met $\mu(A) = 0$ heet een *μ-nulverzameling*, en een bewering $P(\omega)$ die juist is voor alle $\omega \in A^c$, waarbij A een μ-nulverzameling is, heet *μ-bijna overal* juist. Voor verzamelingen $A, B, A_1, A_2, \ldots \in A$ geldt:

(1.5.1) $\mu(\emptyset) = 0$;

(1.5.2) $\mu(A) + \mu(A^c) = \mu(\Omega)$;

(1.5.3) $\mu(A) \leq \mu(B)$ als $A \subset B$,
$\mu(B-A) = \mu(B) - \mu(A)$ als bovendien $\mu(A) < \infty$;

(1.5.4) $\mu(\bigcup_1^n A_m) \leq \sum_1^n \mu(A_m)$, $\mu(\bigcup_1^\infty A_n) \leq \sum_1^\infty \mu(A_n)$;

(1.5.5) $\mu(\lim A_n) = \lim \mu(A_n)$ als $A_1, A_2, \ldots$ een stijgende rij is, en ook als $A_1, A_2, \ldots$ en dalende rij is met $\mu(A_n) < \infty$ voor voldoende grote n;

(1.5.6) $\mu(\liminf A_n) \leq \liminf \mu(A_n)$;

(1.5.7) $\mu(\limsup A_n) \geq \limsup \mu(A_n)$ als $\mu(\bigcup_n^\infty A_m) < \infty$ voor voldoend grote n;

(1.5.8) $\mu(\lim A_n) = \lim \mu(A_n)$ als $\lim A_n$ bestaat en $\mu(\bigcup_n^\infty A_m) < \infty$ voor voldoend grote n.

Voorbeelden

1.5.1. Zij N een eindige of aftelbare deelverzameling van een ruimte Ω, zij $A \subset M(\Omega)$ een σ-algebra en zij $\mu(A)$ voor een willekeurige verzameling $A \in A$ gelijk aan het aantal elementen van $A \cap N$. μ is dan een maat op A. Men noemt μ een *telmaat*. In feite wordt μ geheel vastgelegd door de eis dat $\mu(N^c) = 0$ en $\mu\{\omega\} = 1$ voor iedere $\omega \in N$.

1.5.2. *Lebesgue maat op $(\mathbb{R}^1, B^1)$*

Voor iedere disjuncte vereniging $A = \bigcup_1^n (a_i, b_i]$ van cellen definiëren wij $l(A) = \sum_1^n (b_i - a_i) \leq \infty$. Hiermee is l als functie op F^1 ondubbelzinnig gedefinieerd. Bovendien blijkt l een σ-finiete maat op F^1 te zijn. Er is dus wegens stelling 1.5.1 een unieke voortzetting λ^1 van l, die een maat op de Borel verzamelingen is. λ^1 heet de Lebesgue maat op B^1 en is op grond van zijn constructie de enige maat op B^1 die aan ieder interval zijn lengte als maat toekent.

In feite kan l tot een maat op een nog grotere σ-algebra L, de σ-algebra van de *Lebesgue-meetbare verzamelingen*, worden uitgebreid. Men kan aantonen dat er bij iedere verzameling $L \in L$ een Borel verzameling B is zodanig, dat $L \Delta B$ bevat is in een λ^1-nulverzameling in B^1.

1.5.3. *Lebesgue-Stieltjes maat op* (R^1, B^1)

Zij F een reële, niet-dalende, rechtscontinue functie op R^1, en laten $F(\infty)$ en $F(-\infty)$ gedefinieerd zijn als de limieten van $F(x)$ voor $x \to \infty$ resp. $x \to -\infty$. Door nu $m(A) = \sum_1^n (F(b_i) - F(a_i))$ te stellen voor iedere disjuncte vereniging $A = \bigcup_1^n (a_i, b_i]$ van cellen, krijgen wij evenals boven een ondubbelzinnig gedefinieerde σ-finiete maat m op F^1. De unieke voortzetting μ op B^1 van m heet de Lebesgue-Stieltjes maat behorende bij de gegeven functie F. μ is de enige maat op B^1, die aan iedere cel $(a,b]$ het verschil $F(b) - F(a)$ als maat toekent. Neemt men $F(x) \equiv x$, dan wordt μ de Lebesguemaat op B^1.Ook hier geldt dat men m kan voortzetten tot een grotere σ-algebra B_μ, de σ-algebra van de *μ-meetbare verzamelingen*, die gekarakteriseerd wordt door de eigenschap dat er bij iedere $A \in B_\mu$ een $B \in B^1$ is, zodanig dat $A \Delta B$ in een μ-nulverzameling in B^1 bevat is.

Omgekeerd kan men bij iedere maat μ op B^1, die aan elk eindig interval eindige maat toekent, een niet-dalende, rechtscontinue functie F vinden zodanig dat $\mu(a,b] = F(b) - F(a)$ voor iedere cel $(a,b]$. Men neme bijv.

$$F(x) = \begin{cases} \mu(x_0, x] & \text{als } x \geqq x_0, \\ -\mu(x, x_0] & \text{als } x \leqq x_0, \end{cases}$$

waarbij x_0 een willekeurig doch vast gekozen reëel getal is. F wordt op een additieve constante na uniek bepaald door μ: Als F en G beide voldoen, dan is er een constante c zodanig, dat $F(x) \equiv G(x) + c$.

Iedere maat op B^1 die aan iedere eindige cel eindige maat toekent, is dus een Lebesgue-Stieltjes maat.

1.5.4. *Lebesgue-Stieltjes maat op* (R^k, B^k)

Analoog aan het voorgaande noemt men een maat μ op B^k, met de eigenschap dat $\mu(a,b]$ eindig is voor iedere eindige cel, d.w.z. voor iedere cel $(a,b]$ met $a = (a_1,\ldots,a_k)$ en $b = (b_1,\ldots,b_k)$ eindig, een Lebesgue-Stieltjes maat. Voor de constructie van een dergelijke maat gaan we uit van een reële functie F op R^k, die niet-dalend en rechtscontinu is in ieder van zijn argumenten. Dit is echter nog niet genoeg. Voor een eindige cel $(a,b]$ definiëren wij differentie-operatoren Δ_i door $\Delta_i\, F(x_1,\ldots,x_k) = F(x_1,\ldots,x_{i-1},b_i,x_{i+1},\ldots,x_k) - F(x_1,\ldots,x_{i-1},a_i,x_{i+1},\ldots,x_k)$ te stellen $(i=1,2,\ldots,k)$. We eisen nu dat $\Delta_1\, \Delta_2\, \ldots\, \Delta_k\, F(a_1,\ldots,a_k) \geq 0$ voor iedere eindige cel $(a,b]$. We definiëren vervolgens $m(a,b] = \Delta_1\, \Delta_2\, \ldots\, \Delta_k\, F(a_1,\ldots,a_k)$ voor alle cellen $(a,b]$, zo nodig via een limietovergang als één of meer van de coördinaten van a of b oneindig zijn. Gebruik makend van het feit dat iedere verzameling in $\mathcal{F}^k$ te schrijven is als een eindige disjuncte vereniging van cellen, kunnen we m via additiviteit ondubbelzinnig definiëren op $\mathcal{F}^k$. De aldus verkregen functie m blijkt een σ-finiete maat op $\mathcal{F}^k$ te zijn. Zijn unieke uitbreiding μ op $\mathcal{B}^k$ is de enige maat op $\mathcal{B}^k$ met $\mu(a,b] = \Delta_1\, \Delta_2\, \ldots\, \Delta_k\, F(a_1,\ldots,a_k)$ voor alle cellen $(a,b]$. Omgekeerd is er bij iedere Lebesgue-Stieltjes maat μ op $\mathcal{B}^k$ een functie F op R^k, die aan alle hierboven gestelde eisen voldoet, en die via de geschetste constructie weer tot μ voert.

Nemen wij $F(x_1,\ldots,x_k) = x_1\, x_2\, \ldots\, x_k$, dan is aan alle eisen voldaan. Voor eindige cellen $(a,b]$ blijkt
$\Delta_1\, \Delta_2\, \ldots\, \Delta_k\, F(a_1,\ldots,a_k) = \prod_1^k (b_i-a_i)$. De corresponderende maat λ^k heet de *Lebesgue-maat op* $\mathcal{B}^k$.

Heeft men een aantal σ-finiete maatruimten $(\Omega_i,\mathcal{A}_i,\mu_i)$, dan definieert men een *productmaat* $\mu = \mu_1 \times \mu_2 \times \ldots \times \mu_k$ op de productmeetbare ruimte $(\Omega,\mathcal{A}) = (\Omega_1\times\ldots\times\Omega_k,\mathcal{A}_1\times\ldots\times\mathcal{A}_k)$. Men gaat daarbij als volgt te werk. Eerst definieert men $m(A_1\times\ldots\times A_k) = \mu_1(A_1)\,\mu_2(A_2)\,\ldots\,\mu_k(A_k)$ voor alle product-verzamelingen $A_1 \times \ldots \times A_k$ met $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, k$. Vervolgens definieert men $m(A)$ voor A in de productalgebra $\mathcal{F}$ van $\mathcal{A}_1, \ldots, \mathcal{A}_k$ via additiviteit door A als eindige disjuncte vereniging van zulke productverzamelingen te schrijven. Deze definitie van m is ondubbelzinnig en maakt m tot een σ-finiete maat op $\mathcal{F}$. De productmaat μ is nu per definitie de unieke voortzetting van m op $\mathcal{A}$. μ is dus de unieke maat op $\mathcal{A}$ met de eigenschap dat

$\mu(A_1 \times \ldots \times A_k) = \mu_1(A_1)\ \mu_2(A_2)\ \ldots\ \mu_k(A_k)$ als $A_i \in A_i$, $i = 1, 2, \ldots, k$. Men noemt (Ω, A, μ) de *product maatruimte* van de (Ω_i, A_i, μ_i).

De product maatruimte (Ω, A, μ) van een oneindige rij maatruimten (Ω_i, A_i, μ_i), $i = 1, 2, \ldots$, kan men slechts op zinvolle wijze definiëren als $\mu_i(\Omega_i) = 1$ voor alle i. Men neemt dan $\Omega = \Omega_1 \times \Omega_2 \times \ldots$ en $A = A_1 \times A_2 \times \ldots$. Voorts definieert men $m(A) = \mu_1(A_1)\ \mu_2(A_2)\ \ldots\ \mu_k(A_k)$ voor iedere product cylinder $A = A_1 \times A_2 \times \ldots \times A_k \times \Omega_{k+1} \times \ldots$, met zijden $A_i \in A_i$, $i = 1, 2, \ldots, k < \infty$, waarna men deze definitie via additiviteit uitbreidt tot willekeurige verzamelingen in de product algebra F van $A_1, A_2, \ldots$. De zo verkregen functie m blijkt weer een ondubbelzinnig gedefinieerde maat op F te zijn met $m(\Omega) = 1$, en de productmaat μ is zijn unieke voortzetting op A.

Opgaven

1. Zij F een niet-dalende rechtscontinue functie op R^1 en zij μ de bijbehorende Lebesgue-Stieltjes maat op B^1.
   Bewijs voor willekeurige $a < b$:

$$\begin{aligned}
\mu(a,b) &= F(b-0) - F(a),\\
\mu[a,b) &= F(b-0) - F(a-0),\\
\mu[a,b] &= F(b) - F(a-0),\\
\mu\{a\} &= F(a) - F(a-0).
\end{aligned}$$

2. Laat zien dat (R^k, B^k, λ^k) de product-maatruimte van k exemplaren van (R^1, B^1, λ^1) is.
3. Zij F_i een niet-dalende rechtscontinue functie op R en zij μ_i de corresponderende Lebesgue-Stieltjes maat op B^1, $i = 1, 2, \ldots, k$. Bewijs dat de productmaat $\mu_1 \times \mu_2 \times \ldots \times \mu_k$ op B^k via de in voorbeeld 1.5.4 beschreven constructie uit de functie $F(x_1,\ldots,x_k) = F_1(x_1)\ F_2(x_2)\ldots F_k(x_k)$ verkregen kan worden.
4. Zij $F(x_1,\ldots,x_k) = \min(x_1,\ldots,x_k)$. Bewijs dat F aan alle in voorbeeld 1.5.4 gestelde eisen voldoet en dat de door F bepaalde Lebesgue-Stieltjes maat μ op B^k de eigenschap heeft dat $\mu(A) = 0$ voor iedere $A \in B^k$ die geen enkel punt gemeen heeft met de hoofddiagonaal $\Delta = \{(x_1,\ldots,x_k) : x_1 = x_2 = \ldots = x_k\}$ in R^k.

5. Als $(\Omega,\mathcal{A},\mu)$ de productmaatruimte van de maatruimten $(\Omega_i,\mathcal{A}_i,\mu_i)$ met $\mu_i(\Omega_i) = 1$, $i = 1, 2, \ldots$, dan geldt $A \in \mathcal{A}$ en $\mu(A) = \prod_1^\infty \mu_i(A_i)$ voor iedere productverzameling $A = A_1 \times A_2 \times \ldots$ met $A_i \in \mathcal{A}_i$ voor alle i. Bewijs dit.
6. Zij $(\Omega,\mathcal{A},\mu)$ de productruimte van de maatruimten $(\Omega_i,\mathcal{A}_i,\mu_i)$ met $\Omega_i = \{0,1\}$, $\mathcal{A}_i = M(\Omega_i)$ en waar μ_i gegeven wordt door $\mu_i\{0\} = \mu_i\{1\} = \frac{1}{2}$ $(i=1,2,\ldots)$.
   Bewijs:
   a) $\{\omega\} \in \mathcal{A}$ en $\mu\{\omega\} = 0$ voor iedere $\omega \in \Omega$.
   b) $E = \{\omega : \sum \omega_i < \infty\} \in \mathcal{A}$ en $\mu(E) = 0$.
   c) De functie $f(\omega) = \sum_1^\infty \omega_i\, 2^{-i}$ op Ω is meetbaar.
   d) Zij $\hat{\Omega} = E^c$, $\hat{\mathcal{A}} = \{A : A \in \mathcal{A},\ A \subset E^c\}$, $\hat{\mu}(A) = \mu(A)$ voor alle $A \in \hat{\mathcal{A}}$, en zij $\hat{f}$ de restrictie van f op $\hat{\Omega}$. $\hat{f}$ beeldt $\hat{\Omega}$ eenduidig af op $(0,1]$. $\hat{f}$ en zijn inverse functie zijn beide meetbaar.
   e) $\lambda^1(a,b] = \hat{\mu}(\hat{f}^{-1}(a,b]) = \mu(f^{-1}(a,b])$ voor iedere cel $(a,b] \subset (0,1]$.
   f) $\lambda^1(B) = \mu(f^{-1}(B))$ voor iedere Borel verzameling $B \subset [0,1]$.

1.6. INTEGRATIE

We beschouwen in het volgende al dan niet eindige, reële meetbare functies op een gegeven maatruimte $(\Omega,\mathcal{A},\mu)$. Voor de *integraal* van een dergelijke functie f over Ω schrijven wij

$$\int_\Omega f(\omega)\, d\mu(\omega), \text{ of kortweg } \int f\, d\mu.$$

Zulke integralen worden als volgt gedefinieerd:

Als f een *niet-negatieve elementaire functie* is, d.w.z. als $f = \sum_1^n a_i I_{A_i}$ met $n < \infty$, $0 \leq a_i \leq \infty$, $A_i \in \mathcal{A}$ $(i=1,2,\ldots,n)$ en $A_1, \ldots, A_n$ disjunct, dan

$$\int f\, d\mu = \sum_1^n a_i\, \mu(A_i),$$

met de conventie dat $0.\infty = \infty.0 = 0$ (zie §1.4).

Als f een *niet-negatieve meetbare functie* is, dan is er een rij niet-negatieve elementaire functies f_n, zodanig dat $0 \leq f_n(\omega) \uparrow f(\omega)$ voor $n \to \infty$, $\omega \in \Omega$ en

$$\int f\, d\mu = \lim_{n\to\infty} \int f_n\, d\mu .$$

Deze definitie is ondubbelzinnig.

Als f een willekeurige *meetbare functie* is en tenminste één van de integralen $\int f^+\, d\mu$ en $\int f^-\, d\mu$ eindig is, dan zeggen wij dat f *integreerbaar* is, of ook wel dat de *integraal* van f *bestaat*, en

$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu .$$

Is de aldus gedefinieerde integraal eindig, dan noemen we f *sommeerbaar*. De integraal van een meetbare functie f over een verzameling $A \in \mathcal{A}$ wordt gegeven door

$$\int_A f\, d\mu = \int_A f(\omega)\, d\mu(\omega) = \int I_A(\omega)\, f(\omega)\, d\mu(\omega)$$

mits het rechterlid gedefinieerd is. We zeggen dan dat f *integreerbaar op* $\mathcal{A}$ is, of, als de integraal eindig is, *sommeerbaar op* $\mathcal{A}$ is.

Uit deze definitie van integraal kan men de volgende stellingen afleiden.

Stelling 1.6.1.

a) Voor iedere constante $a \in \overline{R}^1$ en $A \in \mathcal{A}$ geldt

$$\int_A a\, d\mu = a\, \mu(A).$$

b) Als a, $b \in R^1$, $A \in \mathcal{A}$, en f en g integreerbaar op A zijn en af + bg op A gedefinieerd is, dan is af + bg ook integreerbaar op A met

$$\int_A (af + bg)\, d\mu = a \int_A f\, d\mu + b \int_A g\, d\mu,$$

mits het rechterlid van deze gelijkheid gedefinieerd is.

c) Als f en g integreerbaar op $A \in \mathcal{A}$ zijn en $f \leq g$ μ-bijna overal op A, dan geldt

$$\int_A f\, d\mu \leq \int_A g\, d\mu .$$

d) Als f integreerbaar op $A \in \mathcal{A}$ is, dan geldt

$$\left|\int_A f\, d\mu\right| \leq \int_A |f|\, d\mu = \int_A f^+\, d\mu + \int_A f^-\, d\mu .$$

Stelling 1.6.2. (monotone convergentie stelling)

Als $0 \leqq f_n(\omega) \uparrow f(\omega)$ voor $n \to \infty$ μ-bijna overal, dan geldt ook

$$\int f_n \, d\mu \uparrow \int f \, d\mu \text{ voor } n \to \infty.$$

Stelling 1.6.3. (lemma van Fatou)

Als $f_n \geqq 0$ μ-bijna overal voor $n = 1, 2, \ldots$, dan geldt

$$\int \liminf_n f_n \, d\mu \leqq \liminf_n \int f_n \, d\mu;$$

Als $f_n \leqq 0$ μ-bijna overal voor $n = 1, 2, \ldots$, dan geldt

$$\int \limsup_n f_n \, d\mu \geqq \limsup_n \int f_n \, d\mu.$$

Stelling 1.6.4. (gedomineerde convergentiestelling)

Als $f_n(\omega) \to f(\omega)$ voor $n \to \infty$ μ-bijna overal en er een sommeerbare functie g is zodanig dat $|f_n| \leqq g$ μ-bijna overal voor alle n, dan is f sommeerbaar en

$$\int |f - f_n| \, d\mu \to 0 \text{ voor } n \to \infty, \text{ zodat}$$

$$\int_A f_n \, d\mu \to \int_A f \, d\mu \quad \text{voor } n \to \infty, \text{ uniform in } A \in \mathcal{A}.$$

Voorbeelden

1.6.1. *Lebesgue integralen*

Als $(\Omega, \mathcal{A}, \mu) = (R^1, \mathcal{B}^1, \lambda^1)$, dan noemen wij $\int f \, d\mu$ de *Lebesgue integraal* van f. Voor een elementaire functie $f = \sum_1^n y_i \, I_{A_i}$ met de bijzondere eigenschap dat de verzamelingen A_i disjuncte eindige intervallen zijn volgt uit de definities dat de Lebesgue integraal en de Riemann integraal gelijk zijn. Daar de benaderingsprocedure in de definitie van beide typen integralen voor een continue functie op een eindig interval met elementaire functies van dit speciale type kan worden uitgevoerd, impliceert dit dat de Lebesgue integraal van een continue functie over een eindig interval en de overeenkomstige Riemann integraal gelijk zijn.

Nemen wij $(\Omega, \mathcal{A}, \mu) = (R^k, \mathcal{B}^k, \lambda^k)$, dan krijgen wij te maken met Lebesgue integralen over R^k. Deze vertonen een soortgelijke overeenkomst met Riemann integralen over R^k.

1.6.2. *Lebesgue-Stieltjes integralen*

Zij $(\Omega,\mathcal{A}) = (R^1,\mathcal{B}^1)$ en zij μ de Lebesgue-Stieltjes maat die correspondeert met een niet-dalende rechtscontinue functie F op R^1. We noemen $\int f\, d\mu$ dan een *Lebesgue-Stieltjes integraal*. Voor een elementaire functie $f = \sum_1^n y_i I_{A_i}$ met $A_i = (a_i,b_i]$ eindig en disjunct $(i=1,2,\ldots,n)$ is de Lebesgue-Stieltjes integraal gelijk aan de Riemann-Stieltjes integraal:

$$\int f\, d\mu = \sum_1^n y_i\ \mu(A_i) = \sum_1^n y_i\ (F(b_i) - F(a_i)) = \int f\, dF.$$

Evenals boven volgt hieruit dat de beide typen integraal identiek zijn voor continue functies over eindige intervallen.

Stelling 1.6.5. (ongelijkheid van Jensen)

Als g een reële meetbare en convexe functie op R^1 is en f een sommeerbare functie op een maatruimte $(\Omega,\mathcal{A},\mu)$ met $\mu(\Omega) = 1$ is, dan is de samengestelde functie $g\cdot f$ integreerbaar en

$$\int g(f)\, d\mu \geq g(\int f\, d\mu).$$

Bewijs:

Een reële functie g heet convex op een al dan niet eindig open interval I als

$$g(\tfrac{x+y}{2}) \leq \frac{g(x) + g(y)}{2}$$

voor alle $x, y \in I$. Dit is bv. het geval als g op I een niet-dalende afgeleide heeft. Men kan bewijzen dat iedere meetbare convexe functie ook continu is, en dat een continue functie g dan en dan alleen convex op I is als er bij iedere $a \in I$ een getal $m(a)$ is, zodanig dat $g(x) \geq g(a) + m(a)\ (x-a)$ voor alle $x \in I$. Uit het gegeven volgt dus voor willekeurige $a \in R^1$:

$$g(f(\omega)) \geq g(a) + m(a)\ (f(\omega) - a),\ \omega \in \Omega.$$

Daar het rechterlid een sommeerbare functie van ω is, is $g\cdot f$ integreerbaar met

$$\int g(f)\, d\mu \geq g(a) + m(a)\, \{\int f\, d\mu - a\},$$

waaruit de stelling volgt als men $a = \int f\, d\mu$ invult.

Stelling 1.6.6. (overplantingsstelling)

Zij $(\Omega,\mathcal{A},\mu)$ een maatruimte en zij $(\Omega',\mathcal{A}')$ een meetbare ruimte. Als nu f een $\mathcal{A} - \mathcal{A}'$-meetbare functie op Ω is met waarden in Ω', dan is de functie

$$\mu'(A') = \mu(f^{-1}(A')),\ A' \in \mathcal{A}'$$

een maat op $\mathcal{A}'$. Als verder g een meetbare reële functie op $(\Omega',\mathcal{A}')$ is, dan geldt

$$\int_\Omega g(f)\, d\mu = \int_{\Omega'} g\, d\mu'$$

in de zin dat de beide integralen bestaan en gelijk zijn zodra één van hen bestaat.

Bewijs:

De bewering dat μ' een maat is, is een direct gevolg van het feit dat f^{-1} alle verzamelingstheoretische operaties bewaart. De tweede bewering is in het speciale geval dat $g = I_{A'}$ een indicator functie is niets anders dan de definitie van μ'. Wegens de lineariteit van integralen volgt deze bewering dus voor elementaire functies g; wegens de monotone convergentiestelling voor niet-negatieve meetbare functies g, en tenslotte, via de splitsing van g in zijn positieve en negatieve delen, voor willekeurige meetbare g.

Stelling 1.6.7. (Fubini)

Laten $(\Omega_1,\mathcal{A}_1,\mu_1)$ en $(\Omega_2,\mathcal{A}_2,\mu_2)$ twee σ-finiete maatruimten zijn, en zij $(\Omega,\mathcal{A},\mu)$ de product-maatruimte. Als een reële functie f op $(\Omega,\mathcal{A},\mu)$ meetbaar en hetzij niet-negatief hetzij sommeerbaar is, dan zijn de integralen

$$\int_{\Omega_2} f(\omega_1,\omega_2)\, d\mu_2(\omega_2) \text{ en } \int_{\Omega_1} f(\omega_1,\omega_2)\, d\mu_1(\omega_1)$$

μ_1-bijna overal resp. μ_2-bijna overal gedefinieerd en $\mathcal{A}_1$ resp. $\mathcal{A}_2$ meetbaar, en

$$\int f\, d\mu = \int_{\Omega_1} \{\int_{\Omega_2} f(\omega_1,\omega_2)\, d\mu_2(\omega_2)\}\, d\mu_1(\omega_1) = \int_{\Omega_2} \{\int_{\Omega_1} f(\omega_1,\omega_2)\, d\mu_1(\omega_1)\}\, d\mu(\omega_2).$$

Zij (Ω,A) een meetbare ruimte en laten μ en ν twee maten op A zijn. Men noemt ν μ-*absoluut continu* als iedere μ-nulverzameling tevens een ν-nulverzameling is.

Stelling 1.6.8. (Radon-Nikodym)

Als μ σ-finiet en ν μ-absoluut continu is, dan is er een niet-negatieve μ-integreerbare functie g, zodanig dat

$$\nu(A) = \int_A g \, d\mu \text{ voor alle } A \in A.$$

Als g_1 en g_2 twee zulke functies zijn, dan is $g_1 = g_2$ μ-bijna overal. Als ook ν σ-finiet is, dan kan men g eindig kiezen. Men noemt g de *dichtheid* van ν ten opzichte van μ, of ook wel de *Radon-Nikodym afgeleide* van ν naar μ. Men schrijft ook wel $g = \frac{d\nu}{d\mu}$.

Zij (Ω,A) een meetbare ruimte en zij μ een maat en N een collectie maten op (Ω,A). Men zegt dat μ de collectie N *domineert* indien iedere $\nu \in N$ μ-absoluut continu is.

Stelling 1.6.9. (domineringsstelling)

Indien N wordt gedomineerd door een σ-finiete maat μ, dan bestaat er een rij maten $\nu_i \in N$ en een rij reële getallen $c_i \geq 0$ met $\sum_{i=1}^{\infty} c_i = 1$ zodanig dat de maat $\sum_{i=1}^{\infty} c_i \nu_i$ eveneens N domineert.

Bewijs:

Daar μ σ-finiet is bestaat er een rij disjuncte verzamelingen $A_n \in A$ met $\cup A_n = \Omega$ en $0 < \mu(A_n) < \infty$ voor alle n. De maat $\tilde{\mu}$ gedefinieerd door $\tilde{\mu}(A) = \sum_n \frac{\mu(A \cap A_n)}{\mu(A_n)\, 2^n}$ is eindig en domineert eveneens N. Zonder bezwaar kan μ dus eindig in plaats van σ-finiet worden verondersteld.

Zij R de klasse van alle maten $\rho = \sum_i c_i \nu_i$ met $\nu_i \in N$ en $c_i \geq 0$ voor alle i en $\sum_i c_i = 1$. R wordt gedomineerd door μ: zij r de dichtheid van $\rho \in R$ ten opzichte van μ. De stelling is bewezen indien wij een $\rho_0 \in R$ construeren die R (en dus N) domineert.

Zij A_0 de klasse van alle verzamelingen $C \in A$ waarvoor $\mu(C) > 0$ en waarvoor er een $\rho \in R$ bestaat met dichtheid $r(\omega) > 0$ voor μ-bijna alle $\omega \in C$. Indien $C_1, C_2, \ldots \in A_0$, en $\rho_1, \rho_2, \ldots \in R$ zo zijn gekozen dat $r_i(\omega) > 0$ voor μ-bijna alle $\omega \in C_i$, dan geldt voor $c_i > 0$ met $\sum_i c_i = 1$ dat

$\rho = \sum_i c_i \rho_i \in R$ dichtheid $r = \sum_i c_i r_i$ ten opzichte van μ bezit met $r(\omega) > 0$ voor μ-bijna alle $\omega \in \cup C_i$. Daar $\mu(C_i) > 0$ voor alle i, geldt $\mu(\cup C_i) > 0$, zodat $\bigcup_i C_i \in \mathcal{A}_0$. $\mathcal{A}_0$ is dus afgesloten onder aftelbare vereniging.

Kies nu een rij $C_i \in \mathcal{A}_0$ met de eigenschap dat $\lim \mu(C_i) = \sup_{C \in \mathcal{A}_0} \mu(C)$ (eindig!). Volgens het bovenstaande geldt $C_0 = \bigcup_i C_i \in \mathcal{A}_0$ en dus $\mu(C_0) = \sup_{C \in \mathcal{A}_0} \mu(C)$. Zij $\rho_0 \in R$ zo gekozen dat $r_0(\omega) > 0$ voor μ-bijna alle $\omega \in C_0$. Wij zullen aantonen dat deze ρ_0 R domineert.

Hiertoe beschouwen wij een willekeurige verzameling $A \in \mathcal{A}$ met $\rho_0(A) = 0$ en een willekeurige maat $\rho \in R$ met dichtheid r ten opzichte van μ en bewijzen dat $\rho(A) = 0$. Zij $C = \{\omega : r(\omega) > 0\}$ dan geldt $\rho(C^c) = 0$. Voorts is $\rho_0(A \cap C_0) = 0$ daar immers $\rho_0(A) = 0$ en aangezien $r_0(\omega) > 0$ voor μ-bijna alle $\omega \in C_0$, geldt $\mu(A \cap C_0) = 0$ en dus $\rho(A \cap C_0) = 0$ daar μ R domineert. Tenslotte: als $\mu(A \cap C_0^c \cap C) > 0$ zou zijn, dan zou $A \cap C_0^c \cap C \in \mathcal{A}_0$ daar immers $r > 0$ op C. Omdat $C_0 \in \mathcal{A}_0$ zou dan ook $C_0 \cup (A \cap C_0^c \cap C) \in \mathcal{A}_0$. Anderzijds volgt uit $\mu(A \cap C_0^c \cap C) > 0$ dat $\mu(C_0 \cup (A \cap C_0^c \cap C)) > \mu(C_0) = \sup_{C \in \mathcal{A}_0} \mu(C)$, hetgeen inhoudt dat $C_0 \cup (A \cap C_0^c \cap C) \notin \mathcal{A}_0$. Uit deze tegenspraak volgt dat $\mu(A \cap C_0^c \cap C) = 0$ dus $\rho(A \cap C_0^c \cap C) = 0$. Daar ook $\rho(C^c) = \rho(A \cap C_0) = 0$ geldt $\rho(A) = 0$ hetgeen te bewijzen was.

Opgaven

1. Bewijs: Als $f_1, f_2, \ldots$ reële sommeerbare functies op de maatruimte $(\Omega, \mathcal{A}, \mu)$ zijn en $\sum_1^\infty \int |f_n| \, d\mu < \infty$, dan is de reeks $\sum_1^\infty f_n(\omega)$ μ-bijna overal absoluut convergent met sommeerbare som en $\int (\sum f_n) \, d\mu = \sum \int f_n \, d\mu$.
2. Zij f een meetbare reële functie op $(\Omega, \mathcal{A}, \mu)$. Bewijs dat $f = 0$ μ-bijna overal als $\int |f| \, d\mu = 0$, en ook als $\int_A f \, d\mu = 0$ voor alle $A \in \mathcal{A}$.
3. Als f een integreerbare functie op $(\Omega, \mathcal{A}, \mu)$ is en $A_1, A_2, \ldots \in \mathcal{A}$ disjunct zijn, dan geldt:
$$\int_{\cup A_n} f \, d\mu = \sum_n \int_{A_n} f \, d\mu.$$
Bewijs dit.
4. Als ν een μ-absoluut continue maat is en μ σ-finiet is, dan geldt
$$\int f \, d\nu = \int f \frac{d\nu}{d\mu} \, d\mu$$

voor iedere reële meetbare functie f in de zin dat beide integralen bestaan en gelijk zijn zodra er één bestaat. Bewijs dit.

5. Laten μ_1 en μ_2 twee maten zijn op een meetbare ruimte $(\Omega,\mathcal{A})$, en zij $\nu(A) = \mu_1(A) + \mu_2(A)$ voor $A \in \mathcal{A}$. Bewijs dat ν een maat is en dat $\int f\,d\nu = \int f\,d\mu_1 + \int f\,d\mu_2$ voor alle functies f die zowel μ_1- als μ_2-sommeerbaar zijn.

2. WAARSCHIJNLIJKHEIDSREKENING

2.1. KANSRUIMTEN

Bij ieder experiment bestaat een *uitslagenruimte* Ω: de verzameling van alle mogelijke uitkomsten van het experiment. Een *eventualiteit* A behorende bij een experiment is een potentiële gebeurtenis, die bij de uitvoering van het experiment, afhankelijk van de uitslag ervan, al dan niet optreedt, met dien verstande dat het al dan niet optreden van A volledig bepaald wordt door de uitslag van het experiment. Dit betekent dat er een 1-1 correspondentie is tussen eventualiteiten die bij een gegeven experiment behoren enerzijds en deelverzamelingen van de uitslagenruimte Ω van dat experiment anderzijds. De met een eventualiteit A corresponderende verzameling bestaat uit juist die uitslagen ω, die het optreden van A tot gevolg hebben. Op grond van deze 1-1 correspondentie zullen wij in het volgende eventualiteiten en de daarmee corresponderende verzamelingen vereenzelvigen. De verzamelingstheoretische relaties en operaties worden zo relaties en operaties voor eventualiteiten:

Ω is de *zekere eventualiteit*, die bij iedere uitslag van het experiment optreedt;

$\emptyset$ is de *onmogelijke eventualiteit*, die bij geen enkele uitslag van het experiment optreedt;

$\cup A_n$ is de eventualiteit die dan en slechts dan optreedt als tenminste één van de A_n optreedt;

$\cap A_n$ is de eventualiteit die dan en slechts dan optreedt als alle eventualiteiten A_n optreden;

$A \Delta B$ is de eventualiteit die dan en slechts dan optreedt als precies één van de eventualiteiten A en B optreedt;

$\limsup A_n$ is de eventualiteit die dan en slechts dan optreedt als oneindig veel van de eventualiteiten A_n optreden;

$\liminf A_n$ is de eventualiteit die dan en slechts dan optreedt als al de eventualiteiten A_n op eindig veel na optreden;

$A \subset B$ betekent dat B optreedt als A optreedt, m.a.w. dat het optreden van A dat van B *impliceert*;

$AB = \emptyset$ betekent dat A en B niet beide kunnen optreden, m.a.w. dat A en B *elkaar uitsluiten*.

Bij een gegeven experiment wensen wij vaak niet de gehele klasse $M(\Omega)$ van alle deelverzamelingen van de uitslagenruimte Ω te beschouwen, doch slechts een deelklasse A hiervan. Wel zullen wij steeds veronderstellen dat A afgesloten is onder alle eindige en aftelbare verzamelingstheoretische operaties, d.w.z. dat A een σ-algebra is. Wij krijgen dus bij ieder experiment te maken met een *meetbare ruimte* (Ω,A). Wij spreken hierbij af dat slechts de meetbare verzamelingen eventualiteiten genoemd worden.

Stel dat wij een gegeven experiment E, steeds onder dezelfde initiële omstandigheden, kunnen herhalen. Zij n(A) het aantal malen dat een bij E behorende eventualiteit A optreedt indien het experiment n maal wordt uitgevoerd; n(A) wordt de *frequentie* van A genoemd. Het *frequentiequotiënt* of de *relatieve frequentie* van A in deze reeks van n realiseringen van het experiment is dan per definitie

$$\mathrm{fq}(A) = \frac{n(A)}{n}.$$

In het algemeen vertoont fq(A) bij aangroeiende n allerlei grillige fluctuaties. Voert men het experiment nogmaals n keer uit, dan zal men bovendien bij deze tweede reeks van n realiseringen veelal andere waarden voor fq(A) vinden dan in de eerste reeks. Nu doet zich echter in de praktijk bij vele experimenten de omstandigheid voor dat, naarmate n groter wordt, deze fluctuaties van fq(A) binnen één reeks realiseringen van een gegeven experiment, en ook de verschillen van fq(A) tussen verschillende reeksen realiseringen van hetzelfde experiment, steeds geringer worden. Men noemt dit verschijnsel de *empirische wet van de grote aantallen*. Het lijkt dus alsof fq(A) bij een steeds langer wordende reeks herhalingen van ons experiment convergeert naar een vaste limiet P(A). Dit getal P(A) nu zouden wij de kans van de eventualiteit A willen noemen. Dit is echter geen houdbare definitie van het begrip kans. Er is hier immers geen sprake van een limiet in de gebruikelijke zin van het woord omdat men een experiment nu eenmaal niet oneindig vaak kan herhalen. Men is er daarom toe overgegaan de waarschijnlijkheidsrekening op axiomatische grondslag op te bouwen door eenvoudig aan iedere eventualiteit A een getal P(A) toe te kennen en dit per definitie de *kans* of *waarschijnlijkheid* van A te noemen. Willen wij echter het gedrag van frequentiequotiënten in lange reeksen herhalingen van een gegeven experiment als achtergrond voor het begrip kans handhaven, dan zullen wij moeten eisen dat de aldus ingevoerde waarschijnlijkheid de volgende eigenschappen van frequentiequotiënten ook heeft:

$$fq(\emptyset) = 0,\ fq(\Omega) = 1$$

$$0 \leq fq(A) \leq 1 \text{ voor alle } A \in \mathcal{A}$$

$$fq(A \cup B) = fq(A) + fq(B) \text{ als } A,\ B \in \mathcal{A} \text{ disjunct zijn.}$$

Daar wij bovendien steeds werken met σ-algebra's van eventualiteiten ligt het voor de hand niet slechts additiviteit maar zelfs σ-additiviteit te eisen. Deze eisen komen er op neer dat de waarschijnlijkheid P een *genormeerde eindige maat* op $\mathcal{A}$ moet zijn, d.w.z. een maat op $\mathcal{A}$ met de eigenschap dat $P(\Omega) = 1$.

Op grond van deze overwegingen komt men tot de volgende definitie van wat men een *mathematisch model* voor een experiment zou kunnen noemen.

Definitie 2.1.1.

Een *kansruimte* of *waarschijnlijkheidsruimte* is een genormeerde maatruimte $(\Omega,\mathcal{A},P)$. De verzamelingen $A \in \mathcal{A}$ noemen wij *eventualiteiten*, de genormeerde eindige maat P noemen wij *waarschijnlijkheid* of *kans*.

De vraag hoe men Ω, $\mathcal{A}$ en P moet kiezen opdat de kansruimte $(\Omega,\mathcal{A},P)$ een redelijk bruikbaar model voor een gegeven experiment is, blijft bij deze axiomatische opzet van de waarschijnlijkheidsrekening dus geheel buiten beschouwing.

Voorbeelden

2.1.1. *Alternatief*

Een experiment dat slechts twee mogelijke uitkomsten heeft noemt men een *alternatief*. Veelal zullen wij de twee mogelijke uitkomsten "succes" en "mislukking" noemen. Ook duiden wij ze vaak aan met de cijfers 1 en 0. Een kansruimte die als model voor een alternatief kan dienen krijgen wij door $\Omega = \{0,1\}$ en $\mathcal{A} = \mathcal{M}(\Omega)$ te nemen. De kansmaat P op $\mathcal{A}$ wordt dan geheel vastgelegd door $P(1)$, de kans op een succes: Kiezen wij $P(1) = p$ met $0 \leq p \leq 1$, dan volgt $P(0) = 1 - p$.

Een voorbeeld van een alternatief is het kruis of munt gooien met een geldstuk, waarbij bv. de uitslag "kruis" een succes genoemd wordt. Als men de beide mogelijke uitkomsten even waarschijnlijk acht, zodat men $p = \frac{1}{2}$ stelt, dan spreekt men van een *zuivere munt*.

2.1.2. *Aselecte trekking*

Beschouwen wij nu een experiment dat een eindig aantal (r mogelijke) uitkomsten heeft. Een kansruimte voor een dergelijk experiment kan men construeren door $\Omega = \{1,2,\ldots,r\}$ en $A = M(\Omega)$ te kiezen. Een kansmaat P op A wordt dan geheel vastgelegd door de keuze van niet-negatieve getallen $P(1)$, $P(2)$, ..., $P(r)$, zodanig dat hun som 1 is. Bij veel experimenten van dit type, bv. het gooien met een dobbelsteen, het trekken van een kaart uit een goed geschud pak speelkaarten, het blindelings trekken van een loterijbriefje of een knikker uit een vaas met r goed doorelkaar geschudde genummerde briefjes of knikkers, zal men op grond van symmetrie alle mogelijke uitkomsten even waarschijnlijk achten, zodat men $P(i) = \frac{1}{r}$, $i = 1, 2, \ldots, r$ zal kiezen. Men spreekt dan van een worp met een *zuivere dobbelsteen* of van een *aselecte trekking* van 1 object uit een collectie van r objecten.

2.1.3. *Aselecte trekking van een getal tussen 0 en 1*

In voorbeeld 2.1.2 hebben wij het woord "aselect" gebruikt om aan te geven dat het experiment een zekere symmetrie vertoonde, met als gevolg dat alle mogelijke uitkomsten even waarschijnlijk geacht werden. Een dergelijke symmetrie kan echter ook aanwezig zijn bij experimenten met oneindig veel mogelijke uitkomsten. Wij kunnen hierbij denken aan een continu analogon van een roulette: een ronddraaiende wijzer, die op zeker moment tot rust komt, waarna men de stand van de wijzer afleest op een lineaire schaalverdeling van 0 tot 1 op de cirkel waarlangs de punt van de wijzer kan bewegen. Bij dit experiment hebben wij $\Omega = (0,1]$, en het ligt voor de hand voor A de σ-algebra van alle Borelverzamelingen in $(0,1]$ te kiezen. De veronderstelling dat alle mogelijke uitkomsten even waarschijnlijk zijn is nu echter niet voldoende om een kansmaat P op A vast te leggen. Uit deze veronderstelling volgt alleen dat iedere éénpuntsverzameling in $(0,1]$ de maat 0 moet hebben. De afwezigheid van enige voorkeur bij dit experiment kunnen wij echter ook omschrijven door te stellen dat de kans, dat de uitslag van het experiment in een gegeven interval $(a,b] \subset (0,1]$ valt, niet van de ligging, maar alleen van de lengte van dit interval afhangt. Maar dan moet deze kans evenredig zijn met de lengte van het beschouwde interval, en, daar Ω zelf lengte 1 heeft, zelfs gelijk zijn aan deze lengte. Hiermee is P geheel vastgelegd (zie voorbeeld 1.5.2): P is de Lebegue maat op A. Dit voorbeeld verklaart waarom voor de σ-algebra van eventualiteiten A niet steeds $M(\Omega)$ wordt gekozen. Weliswaar zou men in dit geval verder kunnen gaan dan de Borelverzamelingen en voor A de Lebesgue-meetbare verzamelingen kunnen kiezen, doch de Lebesgue maat is niet tot $M(\Omega)$ uit te breiden.

2.1.4. *Aselecte steekproef*

Wij beschouwen nu het experiment dat bestaat uit het nemen van een steekproef van n objecten uit een gegeven collectie van N objecten. Om de gedachten te bepalen denken wij hierbij aan een vaas met N genummerde knikkers, waaruit men een steekproef van n knikkers neemt. Men kan hierbij op verschillende manieren te werk gaan:

a) Na de vaas met inhoud goed geschud te hebben neemt men één voor één n knikkers uit de vaas. Het resultaat noemt men een geordende steekproef zonder teruglegging. De mogelijke uitkomsten van dit experiment kan men omschrijven als alle geordende n-tallen van de vorm $(i_1,i_2,\dots,i_n)$ waarin $i_1, i_2, \dots, i_n$ verschillend zijn en $1 \leq i_j \leq N$ voor $j = 1, 2, \dots, n$. Er zijn $N!/(N-n)!$ zulke geordende n-tallen, en het lijkt onder de gegeven omstandigheden redelijk alle mogelijke uitkomsten even waarschijnlijk te achten. Men komt zo tot een kansruimte (Ω,A,P), waarbij Ω uit $N!/(N-n)!$ punten bestaat, $A = M(\Omega)$ en waarbij de kansmaat P aan iedere éénpuntsverzameling de kans $(N-n)!/N!$ toekent; men spreekt dan van een *geordende aselecte steekproef zonder teruglegging*, of wel van n *aselecte trekkingen zonder teruglegging*.

b) Na goed schudden neemt men in één keer n knikkers uit de vaas. Men heeft dan een ongeordende steekproef zonder teruglegging. Nu kan men de mogelijke uitkomsten identificeren met alle verzamelingen van de vorm $\{i_1, i_2, \ldots, i_n\}$ met $i_1, \ldots, i_n$ verschillend en $1 \leq i_j \leq N$ voor $j = 1, 2, \ldots, n$. Daar er $\binom{N}{n}$ zulke verzamelingen zijn, en het wederom redelijk lijkt alle mogelijke uitkomsten even waarschijnlijk te achten komt men nu tot een kansruimte die uit $\binom{N}{n}$ punten bestaat, die ieder een kans $1/\binom{N}{n}$ hebben; men noemt dit een *aselecte steekproef zonder teruglegging*.

De onder a) en b) genoemde kansruimten zijn consistent in de volgende zin. De eventualiteit in model a) dat de steekproef uit de knikkers $i_1, i_2, \ldots, i_n$ bestaat ongeacht hun volgorde, bestaat uit n! punten, te weten de n! geordende n-tallen die men krijgt als men alle permutaties van $i_1, i_2, \ldots, i_n$ opschrijft. De kans van deze eventualiteit in model a) is dus $n! \frac{(N-n)!}{N!} = 1/\binom{N}{n}$, hetgeen gelijk is aan de kans die dezelfde eventualiteit in model b) heeft. Is men uitsluitend geïnteresseerd in eventualiteiten waarbij de volgorde van de knikkers in de steekproef geen rol speelt, dan kan men dus naar verkiezing met model a) of b) werken.

c) Men trekt de knikkers één voor één, maar steeds legt men de laatst getrokken knikker , na het nummer te hebben genoteerd, terug in de vaas, voordat men deze goed schudt en de volgende knikker neemt. Men krijgt zo een geordende steekproef met teruglegging. Waar in de gevallen a) en b) noodzakelijk is dat $n \leq N$, mag hier $n > N$ zijn. De mogelijke uitkomsten van het experiment kunnen we identificeren met de geordende n-tallen $(i_1, i_2, \ldots, i_n)$ met $1 \leq i_j \leq N$ voor $j = 1, 2, \ldots, n$, waarbij de i_j niet noodzakelijk verschillend hoeven te zijn. Het aantal mogelijke uitkomsten is derhalve N^n, en ook hier is het redelijk aan alle mogelijke uitkomsten dezelfde kans N^{-n} toe te kennen; men spreekt van een *geordende aselecte steekproef met teruglegging*, of wel van n *aselecte trekkingen met teruglegging*.

We veronderstellen nu dat r van de N knikkers in de vaas, bv. de knikkers met de nummers $1, 2, \ldots, r$, rood zijn en dat de overige $w = N - r$ knikkers wit zijn, en we vragen naar de kans p_k dat een aselecte steekproef van n knikkers precies k rode knikkers zal bevatten. Het antwoord op deze vraag hangt er uiteraard van af of de steekproef met of zonder teruglegging wordt genomen. In het laatste geval maakt het echter geen verschil of we met model a) of b) werken, omdat de volgorde van de knikkers in de steekproef hier geen rol speelt. (Dit zou wel het geval zijn als wij bv. vroegen naar de kans dat de eerste k knikkers in de steekproef rood zijn en de volgende n - k knikkers zwart.)

In het geval dat geen teruglegging plaats vindt zijn er onder de $\binom{N}{n}$ mogelijke ongeordende steekproeven precies $\binom{r}{k}\binom{w}{n-k}$ die uit k rode en n - k witte knikkers bestaan, zodat

$$p_k = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}.$$

Hierbij zij opgemerkt dat het aantal rode knikkers in de steekproef uiteraard niet groter dan min (n,r) kan zijn en dat het aantal witte knikkers in de steekproef niet groter dan min (n,w) kan zijn. Bij substitutie van een waarde voor k, zodanig dat $k > \min(n,r)$ of $n - k > \min(n,w)$, in onze uitdrukking voor p_k, vinden wij dan ook $p_k = 0$, dankzij de gebruikelijke conventies voor binomiaalcoëfficiënten.

Neemt men de steekproef met teruglegging, dan zijn er onder de N^n mogelijke resultaten $r^k\, w^{n-k}$ waarbij de steekproef k rode knikkers op k voorgeschreven plaatsen bevat en verder uit witte knikkers bestaat. Daar men k plaatsen in de steekproef op $\binom{n}{k}$ manieren kan voorschrijven, bestaat de eventualiteit dat de steekproef precies k rode knikkers bevat dus uit $\binom{n}{k} r^k\, w^{n-k}$ punten, zodat

$$p_k = \binom{n}{k} \left(\frac{r}{N}\right)^k \left(1 - \frac{r}{N}\right)^{n-k}.$$

Het aantal rode knikkers in de steekproef moet uiteraard tussen 0 en n liggen. Voor $k < 0$ en $k > n$ geeft onze formule dan ook $p_k = 0$.

Opgave

Geef een volledige opsomming van de in voorbeeld 2.1.4 a), b) en c) genoemde uitslagenruimte Ω voor $N = 4$, $n = 3$. Als de knikkers 1 en 2 rood, en de knikkers 3 en 4 wit zijn, geef dan in elk van de modellen a), b) en c) aan uit welke punten van Ω de volgende eventualiteiten bestaan:

a) De steekproef bevat alleen rode knikkers

b) De steekproef bevat 1 rode en 1 witte knikker

c) De steekproef bevat minstens 1 witte knikker.

Beschrijf evenzo voor model c) de eventualiteit dat de steekproef uit twee verschillende knikkers bestaat. Bereken in ieder van deze gevallen de kans van de beschouwde eventualiteit.

2.2. VOORWAARDELIJKE WAARSCHIJNLIJKHEID

Zij gegeven een kansruimte $(\Omega,\mathcal{A},P)$. Voor A, B $\in \mathcal{A}$ met $P(A) \neq 0$ definieert men de *voorwaardelijke waarschijnlijkheid van* B *gegeven* A:

$$P(B|A) = \frac{P(AB)}{P(A)}. \tag{2.2.1}$$

Gemakkelijk is na te gaan dat $P(B|A)$ bij vaste A als functie van B een kansmaat op $\mathcal{A}$ is en dat

$$P(B|A) = \begin{cases} 0 & \text{als } \; B \subset A^c, \\ \frac{P(B)}{P(A)} & \text{als } \; B \subset A. \end{cases}$$

De kansmaat $P(\cdot|A)$ wordt dus uit de oorspronkelijke maat P verkregen door de maat die P buiten A legt te verwijderen en de resterende maat op de in A bevatte deelverzamelingen opnieuw te normeren.

We kunnen deze definitie als volgt motiveren. Stel dat $(\Omega,\mathcal{A},P)$ een model is voor een experiment E. We beschouwen dan naast E een nieuw experiment E_A dat in feite gelijk is aan E, maar waarbij als extra voorwaarde wordt gesteld dat A moet optreden. Om E_A uit te voeren dient men dus allereerst E uit te voeren en te kijken of A daarbij optreedt. Is dit het geval, dan heeft men daarbij E_A uitgevoerd, en anders beschouwt men het experiment als mislukt en doet men het over. Willen we nu voor E_A een kansruimte $(\Omega_A,\mathcal{A}_A,P_A)$ construeren, dan mogen we ongestraft $\Omega_A = \Omega$ en $\mathcal{A}_A = \mathcal{A}$ nemen, als we tenminste $P_A(A^c) = 0$ stellen. Ter bepaling van $P_A(B)$ voor een willekeurige eventualiteit B $\in \mathcal{A}$ beschouwen wij een reeks van n realiseringen van het oorspronkelijke experiment E. Als A hierbij n(A) keer optreedt, dan bevat onze reeks van n realiseringen van E een reeks van n(A) realiseringen van E_A, en het aantal malen dat B optreedt bij deze n(A) realiseringen van E_A is gelijk aan n(AB), het aantal malen dat AB optreedt bij de n realiseringen van E.Het frequentiequotiënt $fq_A(B)$ van B in de reeks van n(A) realiseringen van E_A wordt dus gegeven door

$$fq_A(B) = \frac{n(AB)}{n(A)} = \frac{fq(AB)}{fq(A)}.$$

Als $(\Omega,\mathcal{A},P)$ een goed model voor E is, dan zullen fq(AB) en fq(A) bij grote n in de buurt van resp. P(AB) en P(A) liggen, zodat $fq_A(B)$ in de buurt van $P(B|A)$ komt. Derhalve dienen wij $P_A = P(\cdot|A)$ te kiezen.

Een andere schrijfwijze voor de definitie (2.2.1) is

$$P(AB) = P(A)\, P(B|A), \tag{2.2.2}$$

een gelijkheid die ook zinvol blijft als $P(A) = 0$, hoewel $P(B|A)$ in dat geval ongedefinieerd blijft. Door inductie volgt uit (2.2.2) de zogenaamde *productregel:*

$$P(\bigcap_1^n A_i) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1A_2)\, \dots\, P(A_n|\bigcap_1^{n-1} A_i). \tag{2.2.3}$$

Als een eindige of aftelbare collectie verzamelingen $A_1, A_2, \dots \in \mathcal{A}$ een *partitie* van Ω vormt (d.w.z. $\cup A_i = \Omega$ en $A_i\, A_j = \emptyset$ als $i \neq j$), dan volgt voor willekeurige $B \in \mathcal{A}$:

$$P(B) = \sum_i P(A_iB) = \sum_i P(A_i)\, P(B|A_i). \tag{2.2.4}$$

Als bovendien $P(B) \neq 0$, dan volgt hieruit

$$P(A_j|B) = \frac{P(A_j)\, P(B|A_j)}{\sum_i P(A_i)\, P(B|A_i)} \tag{2.2.5}$$

welke formule bekend staat als de *regel van Bayes*.

Zoals uit de volgende voorbeelden zal blijken zijn voorwaardelijke waarschijnlijkheden vaak zeer nuttig bij de constructie van kansruimten bij gegeven experimenten, daar in veel gevallen door de omschrijving van het experiment bepaalde voorwaardelijke kansen worden vastgelegd.

Voorbeelden

2.2.1. *Geordende aselecte steekproef*

De in voorbeeld 2.1.4 a) omschreven handelwijze laat zich ook als volgt omschrijven: Gegeven is een vaas met N knikkers, genummerd van 1 tot en met N. Hieruit neemt men aselect 1 knikker en vervolgens kiest men telkens aselect 1 knikker uit de nog in de vaas aanwezige knikkers, totdat men in totaal n knikkers uit de vaas genomen heeft. Zij nu $(i_1, i_2, \ldots, i_n)$ een geordend n-tal met $i_1, \ldots, i_n$ verschillend en $1 \leq i_j \leq N$ voor $j = 1, 2, \ldots, n$. Daar de eerste knikker aselect gekozen wordt uit alle N knikkers moeten we de kans dat de eerste knikker het nummer i_1 heeft gelijk aan $^1/N$ stellen. Als reeds bekend is dat de eerste k knikkers de nummers $i_1, \ldots, i_k$ hebben, dan wordt de volgende knikker aselect gekozen uit de nog resterende N - k knikkers, waaronder knikker i_{k+1}. De voorwaardelijke kans dat de (k+1)-ste knikker het nummer i_{k+1} heeft, gegeven dat de eerste k knikkers de nummers $i_1, i_2, \ldots, i_k$ hebben, moeten wij dus gelijk aan $^1/N-k$ stellen $(k=1,2,\ldots,n-1)$. Toepassing van de productregel (2.2.3) geeft nu

$$P(i_1,\ldots,i_n) = \frac{1}{N} \cdot \frac{1}{N-1} \cdots \frac{1}{N-n+1} = \frac{(N-n)!}{N!} ,$$

hetgeen in overeenstemming is met de eerder gemaakte keuze voor deze kans.

Neemt men de steekproef met teruglegging, dan is iedere trekking opnieuw een aselecte trekking van 1 knikker uit onze vaas met N knikkers. Met behulp van de productregel volgt hieruit dat we voor ieder geordend n-tal $(i_1,\ldots,i_n)$ met $1 \leq i_j \leq N$ voor $j = 1, 2, \ldots, n$,

$$P(i_1,\ldots,i_n) = \frac{1}{N} \cdot \frac{1}{N} \cdots \frac{1}{N} = \frac{1}{N^n} ,$$

moeten kiezen, hetgeen weer in overeenstemming is met de in voorbeeld 2.1.4 c) gedane keuze.

2.2.2. Gegeven zijn twee vazen met knikkers: Vaas I bevat 2 witte en 8 rode knikkers, vaas II bevat 4 witte en 1 rode knikker. Men kiest aselect, bv.

door met een zuivere munt te gooien, één van deze vazen en trekt daaruit aselect 1 knikker. Zij W de eventualiteit dat deze knikker wit is. We kunnen P(W) dan berekenen met behulp van (2.2.4). Immers, als A_1 resp. A_2 de eventualiteit is dat vaas I resp. II gekozen wordt, dan is $P(A_1) = P(A_2) = \frac{1}{2}$ wegens de aselecte keuze van de vaas. Omdat de knikker aselect getrokken wordt uit de gekozen vaas geldt

$$P(W|A_1) = \frac{1}{5}\,, P(W|A_2) = \frac{4}{5}\,.$$

Hieruit volgt

$$P(W) = \frac{1}{2} \times \frac{1}{5} + \frac{1}{2} \times \frac{4}{5} = \frac{1}{2}\,.$$

Door toepassing van de regel van Bayes vinden wij verder

$$P(A_1|W) = \frac{\frac{1}{2} \times \frac{1}{5}}{\frac{1}{2}} = \frac{1}{5}\,, P(A_2|W) = \frac{4}{5}\,.$$

Op grond van het voorgaande kunnen wij dit als volgt interpreteren: Voert men het gehele experiment, inclusief de keuze van de vaas, een groot aantal malen uit, dan mag men verwachten, dat in ongeveer de helft van de gevallen de gekozen knikker wit zal zijn, in ongeveer $\frac{1}{5}$ van de gevallen waarin de gekozen knikker wit is, zal deze knikker uit vaas I afkomstig zijn.

Zie voor meer voorbeelden en opgaven bv.[1] en [2].

2.3. ONAFHANKELIJKHEID

Zij gegeven een kansruimte (Ω, A, P).

Definitie 2.3.1.

Twee eventualiteiten A, B heten *(onderling) onafhankelijk* (afkorting: o.o.) als $P(AB) = P(A)\,P(B)$.

Twee klassen eventualiteiten $C_1, C_2 \subset A$ heten (onderling) onafhankelijk als ieder paar eventualiteiten $A_1 \in C_1$, $A_2 \in C_2$ o.o. is.

Een eventualiteit A waarvoor $P(A) = 0$ of 1 is dus o.o. met betrekking tot ieder eventualiteit B. Daar $P(AB) = P(A)\,P(B)$ impliceert dat

$P(A^cB) = P(A^c)\ P(B)$, $P(AB^c) = P(A)\ P(B^c)$ en $P(A^cB^c) = P(A^c)\ P(B^c)$ (zie opgave 2), houdt onafhankelijkheid van A en B in dat de klassen $\{\emptyset, A, A^c, \Omega\}$ en $\{\emptyset, B, B^c, \Omega\}$ o.o. zijn. Indien $0 < P(A) < 1$, dan is onafhankelijkheid van A en B dus equivalent met $P(B|A) = P(B|A^c) = P(B)$. Hieraan ontleent het begrip onafhankelijkheid zijn intuitieve interpretatie: informatie over het al dan niet optreden van A brengt geen verandering in de kans die wij aan B toekennen en leert ons dus niets over het optreden van B. Indien $P(A) = 0$ of 1, dan geeft het al dan niet optreden van A geen enkele informatie. Daar de definitie van onafhankelijkheid symmetrisch in A en B is mogen wij in deze interpretatie de rollen van A en B verwisselen.

Een eerste uitbreiding van deze definitie tot meer dan twee (klassen van) eventualiteiten is de volgende.

Definitie 2.3.2.

Eventualiteiten A_t, $t \in T$ heten *paarsgewijs onafhankelijk* als ieder paar eventualiteiten A_s, A_t met $s, t \in T$, $s \neq t$, o.o. is.

Klassen van eventualiteiten $\mathcal{C}_t \subset \mathcal{A}$, $t \in T$ heten paarsgewijs onafhankelijk als elk tweetal klassen $\mathcal{C}_s$, $\mathcal{C}_t$ met $s, t \in T$, $s \neq t$, o.o. is.

Als A, B en C drie paarsgewijs onafhankelijke eventualiteiten zijn, dan kan er toch nog een zekere mate van afhankelijkheid tussen A, B en C bestaan, zoals blijkt uit het volgende voorbeeld.

Voorbeeld 2.3.1.

Beschouwen wij het experiment dat bestaat uit het doen van twee worpen met een zuivere dobbelsteen. Zij Ω de bijbehorende uitslagenruimte met $\mathcal{A} = \mathcal{M}(\Omega)$ en zij, voor $i = 1, 2$, $\mathcal{C}_i$ de klasse van alle eventualiteiten waarvan het al of niet optreden geheel bepaald wordt door het resultaat van de i-de worp. Het ligt dan voor de hand P zo te kiezen, dat $\mathcal{C}_1$ en $\mathcal{C}_2$ o.o. zijn. In combinatie met de veronderstelling dat de dobbelsteen zuiver is, leidt dit tot het toekennen van gelijke kansen aan de 36 punten in Ω.

Zij nu A de eventualiteit dat het resultaat van de eerste worp even is, B de eventualiteit dat het resultaat van de tweede worp even is, en C de eventualiteit dat het totale aantal ogen even is. Een eenvoudige berekening leert dan dat A, B en C paarsgewijs onafhankelijk zijn met $P(A) = P(B) = P(C) = \frac{1}{2}$. Uit de definitie van A, B en C volgt verder

$$C = AB \cup A^cB^c,$$

zodat

$$AC = BC = AB = ABC.$$

Derhalve hebben wij

$$P(C|AB) = P(B|AC) = P(A|BC) = 1.$$

Hoewel informatie over het al of niet optreden van één van de eventualiteiten A, B en C ons niets verteld over het al of niet optreden van de andere twee, impliceert het optreden van twee van deze eventualiteiten, dat ook de derde optreedt.

Gezien het voorgaande zullen wij, als wij elke afhankelijkheid tussen een aantal eventualiteiten A_t, $t \in T$ willen uitsluiten, moeten eisen dat iedere voorwaardelijke kans van de vorm $P(A_{t_1}|A_{t_2} \dots A_{t_k})$, voor zover gedefinieerd, gelijk is aan de corresponderende onvoorwaardelijke waarschijnlijkheid $P(A_{t_1})$.

Definitie 2.3.3.

Eventualiteiten A_t, $t \in T$ heten *onderling onafhankelijk* als

$$P(\bigcap_{i=1}^{k} A_{t_i}) = \prod_{i=1}^{k} P(A_{t_i})$$

voor iedere eindige deelverzameling $\{t_1, t_2, \dots, t_k\}$ van T.

Klassen van eventualiteiten $\mathcal{C}_t \subset \mathcal{A}$, $t \in T$ heten onderling onafhankelijk als voor iedere keuze van $A_t \in \mathcal{C}_t$, $t \in T$ de eventualiteiten A_t, $t \in T$ o.o. zijn.

Stelling 2.3.1.

Als algebra's $\mathcal{F}_t \subset \mathcal{A}$, $t \in T$ onderling onafhankelijk zijn, dan zijn de daardoor voortgebrachte σ-algebra's dat ook.

Bewijs:

Laat $A_i \in \mathcal{A}_{t_i}$, $t_i \in T$, $i = 1, 2, \dots, k < \infty$, en zij $\varepsilon > 0$. Voor iedere $\eta > 0$ en $i = 1, 2, \dots, k$ is er volgens stelling 1.5.1 een verzameling $B_i \in \mathcal{F}_{t_i}$ zodanig, dat $P(A_i \Delta B_i) < \eta$.

Daar

$$(\bigcap_{i=1}^{k} A_i) \,\Delta\, (\bigcap_{i=1}^{k} B_i) \subset \bigcup_{i=1}^{k} (A_i \Delta B_i),$$

volgt dan

$$|P(\bigcap_1^k A_i) - P(\bigcap_1^k B_i)| \leq \sum_1^k P(A_i \Delta B_i) < k.\eta.$$

Kiezen wij $\eta > 0$ zo klein dat $k.\eta < {}^{\varepsilon}/2$ en bovendien

$$|\prod_1^k P(A_i) - \prod_1^k P(B_i)| < {}^{\varepsilon}/2,$$

dan volgt

$$|P(\bigcap_1^k A_i) - \prod_1^k P(A_i)| < \varepsilon,$$

en dus, daar $\varepsilon > 0$ willekeurig klein genomen kan worden,

$$P(\bigcap_1^k A_i) = \prod_1^k P(A_i).$$

Voorbeeld 2.3.2. *Onafhankelijke experimenten*

Zij E een samengesteld experiment dat uit n deelexperimenten E_i met kansruimten $(\Omega_i, \mathcal{A}_i, P_i)$ bestaat $(i=1,2,\ldots,n)$. Willen wij bij E een kansruimte $(\Omega, \mathcal{A}, P)$ construeren, dan ligt het voor de hand $(\Omega, \mathcal{A})$ gelijk te stellen aan de product meetbare ruimte van de $(\Omega_i, \mathcal{A}_i)$. Voor $i = 1, 2, \ldots, n$ zij $\mathcal{A}_i^*$ de klasse van alle eventualiteiten waarvan het al of niet optreden geheel bepaald wordt door de uitslag van E_i. Gemakkelijk is in te zien dat $\mathcal{A}_i^*$ een σ-algebra is en uit alle verzamelingen van de vorm $A_i^* = A_1 \times A_2 \times \ldots \times A_n$, met $A_j = \Omega_j$ voor $j \neq i$ en $A_i \in \mathcal{A}_i$, bestaat. Uiteraard moeten wij P zo kiezen dat $P(A_i^*) = P_i(A_i)$ voor $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, n$, wil ons model $(\Omega, \mathcal{A}, P)$ in overeenstemming zijn met de gegeven modellen $(\Omega_i, \mathcal{A}_i, P_i)$. We veronderstellen nu dat de deelexperimenten E_i op een zodanige wijze worden uitgevoerd, dat zij elkaar niet kunnen beïnvloeden, met als consequentie dat in ons model $(\Omega, \mathcal{A}, P)$ de σ-algebra's $\mathcal{A}_i^*$ o.o. moeten zijn. Daar $A_1 \times A_2 \times \ldots \times A_n = A_1^* \cap A_2^* \cap \ldots \cap A_n^*$ voor $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, n$, volgt hieruit

$$P(A_1 \times A_2 \times \ldots \times A_n) = P(\bigcap_{i=1}^{n} A_i^*) = \prod_{i=1}^{n} P(A_i^*) = \prod_{i=1}^{n} P_i(A_i)$$

voor iedere productverzameling $A_1 \times A_2 \times \ldots \times A_n$ met $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, n$. Maar hiermee is P geheel vastgelegd: P is de productmaat $P_1 \times P_2 \times \ldots \times P_n$ en $(\Omega, \mathcal{A}, P)$ is dus de productmaatruimte van de kansruimten $(\Omega_i, \mathcal{A}_i, P_i)$. Wij noemen in dit geval de deelexperimenten $E_1, \ldots, E_n$ onderling onafhankelijk.

De zojuist gegeven redenering blijft geldig als wij niet van een eindige maar van een oneindige rij experimenten uitgaan. We kunnen $\mathcal{A}_i^*$ dan beschrijven als de klasse van alle eventualiteiten $A_i^* = A_1 \times A_2 \times \ldots$ met $A_j = \Omega_j$ voor $j \neq i$ en $A_i \in \mathcal{A}_i$. De veronderstelling dat de σ-algebra's $\mathcal{A}_i^*$, $i = 1, 2, \ldots$ o.o. zijn houdt dan in dat voor een willekeurige product cylinderverzameling A met zijden $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, k < \infty$

$$P(A) = P(\bigcap_{i=1}^{k} A_i^*) = \prod_{i=1}^{k} P(A_i^*) = \prod_{i=1}^{k} P_i(A_i),$$

zodat $P = P_1 \times P_2 \times \ldots$.

Opgaven

1. Gegeven zijn twee vazen met knikkers. De ene vaas bevat 1 witte en 4 rode knikkers, de andere 4 witte en 1 rode. Men kiest aselect een van de twee vazen en neemt daaruit met teruglegging een aselecte steekproef van n knikkers. Zij A_i de eventualiteit dat de i-de knikker in de steekproef wit is $(i=1,2,\ldots,n)$. Ga na of $A_1, A_2, \ldots, A_n$ onderling (paarsgewijs) onafhankelijk zijn.
2. Bewijs: Als in een kansruimte $(\Omega, \mathcal{A}, P)$ een collectie o.o. klassen $\mathcal{C}_t \subset \mathcal{A}$, $t \in T$ gegeven is, dan zijn ook de klassen $\mathcal{C}_t' = \{A: A \in \mathcal{C}_t$ of $A^c \in \mathcal{C}_t\}$ o.o.
3. Bewijs: n eventualiteiten $A_1, A_2, \ldots, A_n$ zijn dan en dan alleen o.o. als

$$P(\bigcap_{i=1}^{n} \tilde{A}_i) = \prod_{i=1}^{n} P(\tilde{A}_i)$$

voor ieder van de 2^n manieren waarop men hierin $\widetilde{A}_i = A_i$ of $\widetilde{A}_i = A_1^c$ kan substitueren.

4. Bewijs: n aselecte trekkingen met teruglegging vormen o.o. experimenten; voor aselecte trekkingen zonder teruglegging is dit niet het geval.
5. Construeer een kansruimte voor een oneindige rij o.o. alternatieven met succeskans p.

2.4. KANSVERDELINGEN

Een kansmaat P op$(R^1,\mathcal{B}^1)$ wordt een *(kans)verdeling* op R^1 of ook wel een één-dimensionale (kans)verdeling genoemd.

Definitie 2.4.1.

Onder een *verdelingsfunctie* op R^1 (één-dimensionale verdelingsfunctie) verstaan wij een reële niet-dalende, rechtscontinue functie F op R^1 met $\lim_{x\to\infty} F(x) = 1$ en $\lim_{x\to-\infty} F(x) = 0$.

Voor iedere kansverdeling P op R^1 is de functie F, gegeven door

$$F(x) = P((-\infty,x]), \; -\infty < x < \infty, \tag{2.4.1}$$

een verdelingsfunctie. Omgekeerd is er bij iedere verdelingsfunctie F op R^1 een unieke kansverdeling P op R^1 die voldoet aan (2.4.1) (zie voorbeeld 1.5.3). Er is dus een ééneenduidig verband tussen kansverdelingen en verdelingsfuncties op R^1. De in (2.4.1) gedefinieerde F wordt de verdelingsfunctie van P genoemd.

Als een kansverdeling P op R^1 absoluut continu is ten opzichte van een σ-finitie maat μ op $\mathcal{B}^1$, dan heeft P volgens de stelling van Radon-Nikodym (stelling 1.6.8) een dichtheid f ten opzichte van μ: er is een eindige niet-negatieve meetbare functie f op R^1 zodanig dat

$$P(B) = \int_B f(x)\, d\mu(x), \; B \in \mathcal{B}^1, \tag{2.4.2}$$

zodat in het bijzonder

$$(2.4.3) \qquad \int_{R^1} f(x)\, d\mu(x) = 1;$$

twee versies van f zijn μ-bijna overal gelijk. Omgekeerd is iedere eindige niet-negatieve meetbare functie f op R^1 die aan (2.4.3) voldoet, een dichtheid ten opzichte van μ van een kansverdeling P op R^1 gegeven door (2.4.2). Men noemt f een *(kans)dichtheid* van P ten opzichte van μ. Uit (2.4.1) en (2.4.2) volgt de relatie tussen verdelingsfunctie en kansdichtheid van P:

$$(2.4.4) \qquad F(x) = \int_{(-\infty,x]} f(y)\, d\mu(y), \; -\infty < x < \infty.$$

Een kansmaat P op $(R^k,\mathcal{B}^k)$, $1 \leq k < \infty$, wordt een (kans)verdeling op R^k of ook wel een k-dimensionale (kans)verdeling genoemd. Een verdelingsfunctie op R^k wordt, analoog aan definitie 2.4.1, als volgt gedefinieerd (zie voorbeeld 1.5.4 voor de notatie).

Definitie 2.4.2.

Een reële functie F op R^k heet een verdelingsfunctie op R^k (k-dimensionale verdelingsfunctie) als $F(x_1,\dots,x_k)$ een niet-dalende rechtscontinue functie van ieder van zijn argumenten is met

$$\lim_{x_i\to-\infty} F(x_1,\dots,x_k) = 0 \quad \text{voor } i = 1, 2, \dots, k,$$

$$\lim_{\substack{x_i\to\infty \\ i=1,2,\dots,k}} F(x_1,\dots,x_k) = 1,$$

terwijl bovendien $\Delta_1 \Delta_2 \dots \Delta_k F(a_1,\dots,a_k) \geq 0$ voor alle $a = (a_1,\dots,a_k)$ en $b = (b_1,\dots,b_k)$ in R^k met $a_i < b_i$ voor $i = 1, 2, \dots, k$.

De verdelingsfunctie F van een kansverdeling P op R^k wordt weer door (2.4.1) gedefinieerd, nu met $x \in R^k$ zodat $(-\infty,x] = \{y: y_i \leq x_i,\ i = 1, 2, \dots, k\}$ een cel in R^k voorstelt. Er is een ééneénduidig verband tussen kansverdelingen en verdelingsfuncties op R^k (zie voorbeeld 1.5.4).

Als een kansverdeling P op R^k absoluut continu is ten opzichte van een σ-finiete maat μ op $\mathcal{B}^k$, dan is er een eindige niet-negatieve meetbare functie f op R^k zodanig dat (2.4.2) geldt voor alle $B \in \mathcal{B}^k$ en dus in het bijzonder

$$(2.4.5) \qquad \int_{R^k} f(x)\, d\mu(x) = 1.$$

Men noemt f een kansdichtheid van P ten opzichte van μ. Twee versies van f zijn μ-bijna overal gelijk en iedere eindige niet-negatieve meetbare functie f waarvoor (2.4.5) geldt is een kansdichtheid van een kansverdeling P op R^k. Ook nu geldt (2.4.4) voor $x \in R^k$.

De kansverdelingen waar men zich in de waarschijnlijkheidsrekening en de mathematische statistiek mee bezig houdt behoren merendeels tot één van twee typen, die discrete respectievelijk continue verdelingen worden genoemd.

Definitie 2.4.3.

Een kansverdeling P op R^k, $k \geq 1$, heet *discreet* als er een eindige of aftelbare verzameling $D \subset R^k$ is met $P(D) = 1$.

Voor een discrete kansverdeling P op R^k geldt

$$P(B) = \sum_{x \in B \cap D} P(\{x\}), \qquad B \in \mathcal{B}^k,$$

waarbij de som in het rechterlid steeds hoogstens aftelbaar veel termen bevat. Dit betekent dat P absoluut continu is ten opzichte van de telmaat μ, gegeven door

$$\mu(B) = \text{aantal elementen van } B \cap D, \quad B \in \mathcal{B}^k;$$

voor de kansdichtheid van P ten opzichte van μ kan worden gekozen

$$f(x) = P(\{x\}), \qquad x \in R^k.$$

Omgekeerd is ook iedere kansverdeling die absoluut continu is ten opzichte van een telmaat op een eindige of aftelbare verzameling $D \subset R^k$, discreet. De verdelingsfunctie F van een discrete kansverdeling P kan men beschrijven als een sprongfunctie van $k = 1$ en een "terrasfunctie" voor $k > 1$:

$$F(x) = \sum_{\substack{y \leq x \\ y \in D}} P(\{x\}), \quad x \in R^k,$$

en iedere kansverdeling met een dergelijke verdelingsfunctie is discreet.

Definitie 2.4.4.

Een kansverdeling P op R^k, $k \geq 1$, heet *continu* als hij absoluut continu is ten opzichte van de Lebesgue maat λ^k op $\mathcal{B}^k$ en bovendien de dichtheid f van P ten opzichte van λ^k op een open verzameling $C \subset R^k$ met $\lambda^k(C^c) = 0$ continu kan worden gekozen.

Uit de λ^k-absolute continuïteit van een continue verdeling P op R^k volgt dat de bijbehorende verdelingsfunctie F continu is. Voor iedere $x \in R^k$ geldt immers

$$0 \leq F(x) - F(x-0) \leq P(\bigcup_{i=1}^{k} \{y : y_i = x_i\})$$

en

$$\lambda^k(\bigcup_{i=1}^{k} \{y : y_i = x_i\}) = 0.$$

Wij mogen dit echter niet omkeren: Er zijn kansverdelingen met continue verdelingsfuncties die niet λ^k-absoluut continu zijn, laat staan continu.

Stelling 2.4.1.

Een kansverdeling P op R^k is dan en dan alleen continu als de bijbehorende verdelingsfunctie F op een open verzameling $C \subset R^k$ met $\lambda^k(C^c) = 0$ een continue k-voudige partiële afgeleide $\frac{\partial^k F(x)}{\partial x_1 \, \partial x_2 \ldots \partial x_k}$ heeft met de eigenschap dat de (eventueel oneigenlijke) Riemann integraal

$$\int_C \frac{\partial^k F(x)}{\partial x_1 \, \partial x_2 \ldots \partial x_k} \, dx = 1.$$

De dichtheid f van P ten opzichte van λ^k kunnen wij dan op C gelijk stellen aan deze afgeleide.

Bewijs:

Zij P een continue kansverdeling op R^k met verdelingsfunctie F en kansdichtheid f ten opzichte van λ^k en laat $x \in C$, de in definitie 2.4.4

genoemde open verzameling. Het punt x heeft dan een omgeving $O_x \subset C$ waarop wij f continu mogen veronderstellen. Binnen deze omgeving kiezen wij een punt $a < x$. Als nu $y \geq a$, $y \in O_x$, dan zijn de Lebesgue integraal en de Riemann integraal van f over $(a,y]$ gelijk:

$$F(y) - F(a) = \int_{(a,y]} f(t)\, d\lambda^k(t) = \int_{(a,y]} f(t)\, dt,$$

zodat

$$\frac{\partial^k F(x)}{\partial x_1\, \partial x_2 \ldots \partial x_k} = f(x), \quad x \in C,$$

wegens de hoofdstelling van de integraalrekening. Schrijven wij de open verzameling C als de vereniging van hoogstens aftelbaar veel disjuncte begrensde cellen $C_i = (a^{(i)}, b^{(i)}]$ met $[a^{(i)}, b^{(i)}] \subset C$, dan volgt uit het voorgaande

$$\int_C f(x)dx = \sum_i \int_{C_i} f(x)dx = \sum_i \int_{C_i} f(x)d\lambda^k(x) = \sum_i P(C_i) = P(C) = 1.$$

Zij nu omgekeerd P een kansverdeling die aan de gestelde eisen voldoet en zij

$$f(x) = \begin{cases} \dfrac{\partial^k F(x)}{\partial x_1\, \partial x_2 \ldots \partial x_k} & \text{als } x \in C \\[2ex] 0 & \text{als } x \in C^c. \end{cases}$$

Schrijven wij C weer als de vereniging van aftelbaar veel disjuncte begrensde cellen $C_i = (a^{(i)}, b^{(i)}]$ met $[a^{(i)}, b^{(i)}] \subset C$, dan volgt

$$P(C) = \sum_i P(C_i) = \sum_i \{F(b^{(i)}) - F(a^{(i)})\} = \sum_i \int_{C_i} f(x)dx = \int_C f(x)dx = 1.$$

Voor een willekeurige cel $(a,b]$ geldt dus $P((a,b]) = P((a,b] \cap C)$ en splitsing van $(a,b] \cap C$ in aftelbaar veel disjuncte begrensde cellen $\tilde{C}_i = (c^{(i)}, d^{(i)}]$ met $[c^{(i)}, d^{(i)}] \subset C$ geeft

$$P((a,b]) = \sum_i \{F(d^{(i)}) - F(c^{(i)})\} = \sum_i \int_{\tilde{C}_i} f(x)dx =$$

$$= \sum_i \int_{\tilde{C}_i} f(x)d\lambda^k(x) = \int_{(a,b]} f(x)d\lambda^k(x).$$

Hiermee is bewezen dat de kansmaat P en de kansmaat P^* gegeven door

$$P^*(B) = \int_B f(x)d\lambda^k(x), \quad B \in \mathcal{B}^k,$$

op de klasse van alle cellen in R^k overeenstemmen, zodat zij ook op $\mathcal{B}^k$ overeenstemmen. P is dus λ^k-absoluut continu en heeft een continue kansdichtheid f op C.

Voor een continue kansverdeling P op R^1 kunnen wij dus als kansdichtheid ten opzichte van de Lebesgue maat λ kiezen:

$$f(x) = \begin{cases} F'(x) & \text{voor } x \in C. \\ 0 & \text{voor } x \in C^c. \end{cases}$$

Uit het bewijs van stelling 2.4.1 blijkt dan dat

$$F(x) = \int_{-\infty}^{x} f(y)dy, \qquad -\infty < x < \infty.$$

Voor een continue kansverdeling P op R^k is

$$f(x) = \begin{cases} \dfrac{\partial^k F(x)}{\partial x_1\, \partial x_2 \cdots \partial x_k} & \text{voor } x \in C \\ 0 & \text{voor } x \in C^c \end{cases}$$

een kansdichtheid van P ten opzichte van λ^k en

$$F(x) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f(y_1,\dots,y_k)\, dy_1 \dots dy_k, \quad x \in R^k.$$

De verdelingsfunctie van een continue verdeling op R^k, $k \geq 1$, kan dus door

Riemann integratie uit de kansdichtheid worden verkregen.

Opgaven

1. Bewijs dat een verdelingsfunctie op R^1 hoogstens aftelbaar veel discontinuïteiten heeft.
2. Bewijs: Iedere verdelingsfunctie F op R^1 kan geschreven worden als een convexe combinatie $F(x) = \rho\, F_c(x) + (1-\rho)\, F_d(x)$, $x \in R^1$, met $\rho \in [0,1]$, waarin F_d een discrete en F_c een continue verdelingsfunctie is.

2.5. STOCHASTISCHE GROOTHEDEN EN VECTOREN

Zijn wij bij een experiment geïnteresseerd in een bepaald quantificeerbaar, d.w.z. in een getal uit te drukken aspect van de uitslag van het experiment, dan ligt het voor de hand dit aspect in het model $(\Omega,\mathcal{A},P)$ voor het experiment te representeren door een functie X, die aan iedere $\omega \in \Omega$ een getal $X(\omega)$ toevoegt. Daarbij zullen wij steeds eisen dat voor ieder interval $(a,b]$ de verzameling $\{\omega\colon X(\omega) \in (a,b]\}$ een eventualiteit is, zodat X een meetbare functie op Ω is.

Definitie 2.5.1.

Een reële, eindige, $\mathcal{A}$-$\mathcal{B}^1$-meetbare functie op een kansruimte $(\Omega,\mathcal{A},P)$ heet een *stochastische grootheid.*

Als regel zullen wij stochastische grootheden aanduiden met hoofdletter X, Y, etc. Dit om duidelijk onderscheid te maken tussen een stochastische grootheid, d.w.z. een functie X en een mogelijke waarde $x = X(\omega)$ ervan. (Een andere, speciaal in ons land in zwang zijnde conventie is om stochastische grootheden met onderstreepte symbolen als $\underline{x}$, $\underline{y}$, etc. aan te geven.) Wij zullen steeds gebruik maken van de volgende verkorte notatie:

$$\{X \in B\} = \{\omega\colon X(\omega) \in B\}$$

$$P(X \in B) = P(\{X \in B\}) = P(\{\omega\colon X(\omega) \in B\}), \quad B \in \mathcal{B}^1,$$

$$\{X \le x\} = \{\omega\colon X(\omega) \le x\}$$

$$P(X \le x) = P(\{X \le x\}) = P(\{\omega\colon X(\omega) \le x\}), \quad x \in R^1, \text{ etc.}$$

Uit de overplantingsstelling (stelling 1.6.6) volgt dat voor iedere stochastische grootheid X de functie P_X gegeven door

$$(2.5.1) \qquad P_X(B) = P(X\in B), \quad B \in \mathcal{B}^1,$$

een kansmaat op $\mathcal{B}^1$, d.w.z. een kansverdeling op R^1 is. Wij noemen P_X de *(kans)verdeling van de stochastische grootheid* X. De verdelingsfunctie F_X van P_X wordt gedefinieerd door (2.4.1):

$$(2.5.2) \qquad F_X(x) = P_X((-\infty,x]) = P(X\leq x), \quad -\infty < x < \infty;$$

F_X wordt ook de *verdelingsfunctie van de stochastische grootheid* X genoemd. Als P_X absoluut continu is ten opzichte van een σ-finiete maat μ op $\mathcal{B}^1$ met kansdichtheid f_X ten opzichte van μ, dan geldt

$$(2.5.3) \qquad P_X(B) = P(X\in B) = \int_B f_X(x)\, d\mu(x), \qquad B \in \mathcal{B}^1,$$

$$(2.5.4) \qquad F_X(x) = P(X\leq x) = \int_{(-\infty,x]} f_X(y)\, d\mu(y), \qquad x \in R^1,$$

$$(2.5.5) \qquad \int_{R^1} f_X(x)\, d\mu(x) = 1.$$

Men noemt f_X een *(kans)dichtheid van de stochastische grootheid* X ten opzichte van μ.

Als $X_1, X_2, \ldots, X_k$ stochastische grootheden op een gemeenschappelijke kansruimte $(\Omega,\mathcal{A},P)$ zijn, dan ligt het voor de hand $X = (X_1,X_2,\ldots,X_k)$ een k-dimensionale stochastische vector te noemen. Dit is dan een functie op $(\Omega,\mathcal{A},P)$ met waarden $X(\omega) = (X_1(\omega),X_2(\omega),\ldots,X_k(\omega)) \in R^k$, met de eigenschap dat de reële functies $X_1, X_2, \ldots, X_k$ meetbaar zijn. De volgende formulering is hiermee equivalent (zie §1.4).

Definitie 2.5.2.

Een $\mathcal{A}$-$\mathcal{B}^k$-meetbare functie op een kansruimte $(\Omega,\mathcal{A},P)$ met waarden in R^k heet een *k-dimensionale stochastische vector*.

Evenals een stochastische grootheid bepaalt ook een stochastische vector $X = (X_1,X_2,\ldots,X_k)$ een kansverdeling P_X gegeven door (2.5.1), doch nu met

$B \in \mathcal{B}^k$. Deze kansverdeling P_X op R^k wordt de *(kans)verdeling van de stochastische vector* X of ook wel de *simultane (kans)verdeling van de stochastische grootheden* $X_1, X_2, \ldots, X_k$ genoemd. De verdelingsfunctie F_X van P_X wordt gedefinieerd door

$$(2.5.6) \qquad F_X(x) = P_X((-\infty,x]) = P(X \leqq x) = P(\bigcap_{i=1}^{k} \{X_i \leq x_i\}),$$
$$x = (x_1,\ldots,x_k) \in R^k;$$

F_X wordt ook de *verdelingsfunctie van de stochastische vector* X of de *simultane verdelingsfunctie van de stochastische grootheden* $X_1, X_2, \ldots, X_k$ genoemd. Als P_X absoluut continu is ten opzichte van een σ-finiete maat μ op $\mathcal{B}^k$ met kansdichtheid f_X ten opzichte van μ, dan geldt (2.5.3) voor $B \in \mathcal{B}^k$, (2.5.4) voor $x \in R^k$ en

$$(2.5.7) \qquad \int_{R^k} f_X(x)\, d\mu(x) = 1.$$

Men noemt f_X een *(kans)dichtheid van de stochastische vector* X of een *simultane (kans)dichtheid van de stochastische grootheden* $X_1, X_2, \ldots, X_k$ ten opzichte van μ.

Wanneer een stochastische vector $X = (X_1,X_2,\ldots,X_k)$, $k \geq 1$, een discrete kansverdeling P_X op een eindige of aftelbare verzameling $D \subset R^k$ bezit (zie definitie 2.4.3) dan is

$$f_X(x) = P_X(\{x\}) = P(X=x), \qquad x \in R^k,$$

een kansdichtheid van X ten opzichte van de telmaat op D, en

$$P_X(B) = P(X \in B) = \sum_{x \in B \cap D} P(X=x), \qquad B \in \mathcal{B}^k,$$

$$F_X(x) = P(X \leqq x) = \sum_{\substack{y \leqq x \\ y \in D}} P(X=y), \qquad x \in R^k,$$

$$\sum_{x \in D} P(X=x) = 1.$$

De kansen $P(X=x)$ voor $x \in D$ leggen de verdeling van een discreet verdeelde stochastische vector X dus geheel vast.

Wanneer een stochastische vector $X = (X_1,X_2,\ldots,X_k)$, $k \geq 1$, een continue kansverdeling P_X op een open verzameling $C \subset R^k$ bezit (zie definitie 2.4.4) dan is

$$(2.5.8)\qquad f_X(x) = \begin{cases} \dfrac{\partial^k F_X(x)}{\partial x_1\,\partial x_2 \ldots \partial x_k} & \text{voor } x \in C \\ 0 & \text{voor } x \in C^c \end{cases}$$

een kansdichtheid van X ten opzichte van de Lebesgue maat λ^k op $\mathcal{B}^k$ en

$$F_X(x) = P(\bigcap_{i=1}^{k} \{X_i \leq x_i\}) = \int_{-\infty}^{x_k} \ldots \int_{-\infty}^{x_1} f_X(y_1,\ldots,y_k)\, dy_1 \ldots dy_k, \quad x \in R^k.$$

Merk op dat voor een continu verdeelde stochastische vector X steeds $P(X=x) = 0$ voor alle $x \in R^k$; deze kansen verschaffen in dit geval dus geen enkele informatie over de verdeling van X.

Uit de kansverdeling van $X = (X_1,\ldots,X_k)$ kan de verdeling van $\tilde{X} = (X_1,\ldots,X_m)$, $m < k$, eenvoudig worden afgeleid. Daar voor $B \in \mathcal{B}^m$, $\{\tilde{X} \in B\} = \{X \in B \times R^{k-m}\}$, geldt

$$P_{\tilde{X}}(B) = P_X(B \times R^{k-m}), \qquad B \in \mathcal{B}^m,$$

en hieruit volgt voor $x = (x_1,\ldots,x_m) \in R^m$

$$F_{\tilde{X}}(x) = \lim_{\substack{x_i \to \infty \\ i > m}} F_X(x_1,\ldots,x_k).$$

In dit verband noemt men de verdeling $P_{\tilde{X}}$ de *marginale (kans)verdeling* van $\tilde{X}$. Als de verdeling van X discreet is, dan geldt voor $x = (x_1,\ldots,x_m) \in R^m$

$$P(\tilde{X}=x) = \sum_{x_{m+1}} \ldots \sum_{x_k} P(X=(x_1,\ldots,x_k))$$

waarbij de sommatie zich uitstrekt over alle $(x_1,\ldots,x_k)$ waarvoor

$(x_1,\ldots,x_m) = x$. Als de verdeling van X continu is, dan geldt voor $x = (x_1,\ldots,x_m)$ volgens het bovenstaande

$$f_{\tilde{X}}(x) = \frac{\partial^m}{\partial x_1 \ldots \partial x_m} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{x_m} \cdots \int_{-\infty}^{x_1} f_X(y_1,\ldots,y_k)\, dy_1 \ldots dy_k =$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1,\ldots,x_k)\, dx_{m+1} \ldots dx_k.$$

Zoals wij een k-tal stochastische grootheden op één kansruimte als een k-dimensionale stochastische vector kunnen beschouwen, zo kunnen wij ook een oneindige rij $X = (X_1,X_2,\ldots)$ van stochastische grootheden op een kansruimte $(\Omega,\mathcal{A},P)$ opvatten als een *∞-dimensionale stochastische vector*, d.w.z. als een $\mathcal{A}$-$\mathcal{B}^\infty$-meetbare functie op $(\Omega,\mathcal{A},P)$ met waarden $X(\omega) = (X_1(\omega),X_2(\omega),\ldots)$ in $(R^\infty,\mathcal{B}^\infty)$. De $\mathcal{A}$-$\mathcal{B}^\infty$-meetbaarheid van X is immers equivalent met de $\mathcal{A}$-$\mathcal{B}'$-meetbaarheid van de componenten X_i, $i = 1, 2,\ldots$. Onder de verdeling van zulk een ∞-dimensionale stochastische vector, of de simultane verdeling van de stochastische grootheden X_i, $i = 1, 2, \ldots$ verstaan wij de kansmaat P_X op $\mathcal{B}^\infty$, gegeven door

$$P_X(B) = P(X\in B), \quad B \in \mathcal{B}^\infty.$$

Nu wordt de σ-algebra $\mathcal{B}^\infty$ voortgebracht door de algebra $\mathcal{B}^+$ die bestaat uit alle cylinderverzamelingen van de vorm

$$B' = \{(x_1,x_2,\ldots) : (x_1,x_2,\ldots,x_k) \in B\},\ B \in \mathcal{B}^k,\ k = 1, 2, \ldots .$$

Daar voor een dergelijke verzameling

$$P_X(B') = P(x_1,\ldots,x_k) \in B),$$

betekent dit, dat P_X bepaald wordt door de verdelingen P_k van de stochastische vectoren $(X_1,\ldots,X_k)$, $k = 1, 2, \ldots$, en dus ook door hun verdelingsfuncties F_k. Uiteraard zijn deze verdelingen *consistent* in de zin dat

$$P_k(B) = P_{k+1}(B\times R^1),\quad B \in \mathcal{B}^k,\ k = 1,\ 2,\ \dots,$$

of, in termen van de verdelingsfuncties,

$$(2.5.9)\qquad F_k(x_1,\dots,x_k) = \lim_{y\to\infty} F_{k+1}(x_1,\dots,x_k,y),\ x = (x_1,\dots,x_k) \in R^k$$
$$k = 1,\ 2,\ \dots .$$

Omgekeerd zegt de zogenaamde *consistentiestelling van Kolmogorov* dat er bij iedere consistente rij verdelingsfuncties F_k op R^k, $k = 1, 2, \dots$, een kansmaat P op $\mathcal{B}^\infty$ bestaat, zodanig dat

$$F_k(x) = P\{(y_1,y_2,\dots):(y_1,\dots,y_k)\leqq(x_1,\dots,x_k)\},$$
$$x = (x_1,\dots,x_k) \in R^k$$
$$k = 1,\ 2,\ \dots .$$

Het bewijs van deze stelling is o.m. te vinden in [4] en [5].

In 1.4 wordt gesproken over de σ-algebra, die door een meetbare functie in zijn domein wordt geïnduceerd. Zo induceert een k-dimensionale ($1\leqq k\leqq\infty$) stochastische vector X op een kansruimte $(\Omega,\mathcal{A},P)$ een σ-algebra $\mathcal{A}_X \subset \mathcal{A}$ in Ω, die bestaat uit alle verzamelingen van de vorm $\{X \in B\}$ met $B \in \mathcal{B}^k$. Stellen wij ons $(\Omega,\mathcal{A},P)$ voor als een model voor een experiment E, dan kunnen wij $\mathcal{A}_X$ omschrijven als de collectie van alle eventualiteiten met de eigenschap dat het al of niet optreden ervan alleen van de uitslag ω van het experiment afhangt via de daarbij behorende waarde $X(\omega)$ van X. Deze interpretatie van de door een stochastische vector in Ω geïnduceerde σ-algebra ligt ten grondslag aan het begrip onafhankelijkheid van stochastische vectoren.

Definitie 2.5.3.

Zij $(\Omega,\mathcal{A},P)$ een kansruimte, T een willekeurige indexverzameling, en zij X_t voor iedere $t \in T$ een stochastische vector op $(\Omega,\mathcal{A},P)$. De stochastische vectoren X_t, $t \in T$ heten dan *onderling (paarsgewijs) onafhankelijk* als de door hen in Ω geïnduceerde σ-algebra's $\mathcal{A}_{X_t}$, $t \in T$ dit zijn.

Stelling 2.5.1.

Voor een eindige of oneindige rij stochastische grootheden $X_1, X_2, \ldots$ op een kansruimte $(\Omega,\mathcal{A},P)$ zijn de volgende drie beweringen equivalent:

(i) $X_1, X_2, \ldots$ zijn o.o.;

(ii) De simultane verdeling P_X van $X_1, X_2, \ldots$ is de product kansmaat van hun marginale verdelingen;

(iii) Voor iedere $k = 1, 2, \ldots$ is de simultane verdelingsfunctie F_k van $X_1, \ldots, X_k$ het product van hun marginale verdelingsfuncties:

$$F_k(x_1,\ldots,x_k) = \prod_{i=1}^{k} F_{X_i}(x_i),\ (x_1,\ldots,x_k) \in R^k,\ k = 1, 2, \ldots .$$

Bewijs:

Zij P_k de simultane verdeling van $X_1, \ldots, X_k$ $(k=1,2,\ldots)$. Uit de definities 2.3.3 en 2.5.3 volgt dan dat onderlinge onafhankelijkheid van $X_1, X_2, \ldots$ equivalent is met de eigenschap dat

$$P_k(B_1\times B_2\times\ldots\times B_k) = P\Big(\bigcap_{i=1}^{k}\{X_i \in B_i\}\Big) = \prod_{i=1}^{k} P_{X_i}(B_i),$$

$$\text{als } B_i \in \mathcal{B}^1,\ i = 1, 2, \ldots, k,\ k = 1, 2, \ldots$$

De equivalentie van (i) en (ii) volgt dus rechtstreeks uit de definitie van productmaat. De equivalentie van (ii) en (iii) is een gevolg van de eeneenduidigheid van het verband tussen verdelingen en verdelingsfuncties (zie 1.5, opgave 3).

Stelling 2.5.2.

Zij $X = (X_1,\ldots,X_k)$ een k-dimensionale stochastische vector met de eigenschap dat de marginale verdelingen P_{X_i} dichtheden f_{X_i} ten opzichte van σ-finiete maten μ_i hebben $(i=1,2,\ldots,k)$. Nodig en voldoende voor onderlinge onafhankelijkheid van $X_1, X_2, \ldots, X_k$ is dan dat hun simultane verdeling P_X een dichtheid heeft ten opzichte van de productmaat $\mu = \mu_1 \times \mu_2 \times \ldots \times \mu_k$, gegeven door

$$f_X(x_1,\ldots,x_k) = \prod_{i=1}^{k} f_{X_i}(x_i),\quad (x_1,\ldots,x_k) \in R^k.$$

Bewijs:

De gestelde voorwaarde is, wegens de stelling van Fubini, equivalent met

$$P_X(B_1\times B_2\times\ldots\times B_k) = \int_{B_1\times\ldots\times B_k} f_X(x)d\mu(x) = \prod_{i=1}^{k}\int_{B_i} f_{X_i}(x_i)d\mu_i(x_i) =$$

$$= \prod_{i=1}^{k} P_{X_i}(B_i) \text{ voor } B_i \in \mathcal{B}^1,\ i = 1, 2, \ldots, k,$$

m.a.w. met de eis dat $P_X = P_{X_1} \times P_{X_2} \times \ldots \times P_{X_k}$.

Stelling 2.5.3.

Zij X_t, $t \in T$ een collectie o.o. stochastische vectoren op een kansruimte $(\Omega,\mathcal{A},P)$. Zij k_t de dimensie van X_t en zij f_t een Borelfunctie op R^{k_t} met waarden in R^{m_t} $(t\in T)$. Dan zijn ook de stochastische vectoren $Y_t = f_t(X_t)$, $t \in T$ onderling onafhankelijk.

Bewijs:

Voor $t \in T$ en $B \in \mathcal{B}^{m_t}$ geldt $f_t^{-1}(B) \in \mathcal{B}^{k_t}$, en dus

$$\{Y_t \in B\} = \{X_t \in f_t^{-1}(B)\} \in \mathcal{A}_{X_t},$$

zodat $\mathcal{A}_{Y_t} \subset \mathcal{A}_{X_t}$, $t \in T$.

Wij besluiten deze paragraaf met enige opmerkingen over de kansverdeling van een functie van een stochastische vector. Zij $X = (X_1,\ldots,X_k)$ een stochastische vector met kansverdeling P_X en zij ϕ een meetbare functie op R^k met waarden in R^m. Dan is $Z = \phi(X)$ een m-dimensionale stochastische vector waarvan de kansverdeling als volgt uit die van X kan worden gevonden:

$$P_Z(B) = P(\phi(X)\in B) = P_X(\{x:\phi(x)\in B\}),\quad B \in \mathcal{B}^m. \tag{2.5.10}$$

Indien X een kansdichtheid f_X ten opzichte van een σ-finiete maat μ op $\mathcal{B}^k$ bezit dan geldt dus

$$(2.5.11) \qquad P_Z(B) = \int_{\{x:\phi(x)\in B\}} f_X(x)d\mu(x), \qquad B \in \mathcal{B}^m.$$

Als X discreet verdeeld is, dan is de verdeling van Z eveneens discreet en wordt bepaald door

$$(2.5.12) \qquad P(Z=z) = \sum_{\{x:\phi(x)=z\}} P(X=x), \qquad z \in R^m;$$

als X continu verdeeld is met kansdichtheid f_X gegeven door (2.5.8) en bovendien ϕ continu is, dan geldt

$$(2.5.13) \qquad F_Z(z) = \int_{\{x:\phi(x)\leq z\}} f_X(x)dx, \qquad z \in R^m,$$

waarbij het rechterlid een (oneigenlijke) Riemann integraal voorstelt.

Het eenvoudigste voorbeeld van het bovenstaande is het volgende. Indien X een stochastische grootheid met verdelingsfunctie F_X voorstelt, dan wordt de verdelingsfunctie van $Y = a + bX$, $b > 0$, gegeven door

$$F_Y(x) = P(a+bX\leq x) = F_X(\tfrac{x-a}{b}), \qquad -\infty < x < \infty.$$

De grafiek van F_Y ontstaat dus uit die van F_X door verschuiving naar rechts (c.q. links) over een afstand a (c.q - a) en vermenigvuldiging van de schaal langs de horizontale as met een factor b^{-1}(d.w.z. bij behoud van dezelfde schaal "uitrekken" van de grafiek in horizontale richting met een factor b). De klasse van alle kansverdelingen van de vorm $F_X(b^{-1}(x-a))$, $-\infty < a < \infty$, $b > 0$, noemt men de door F_X voortgebrachte *familie van kansverdelingen met plaatsparameter a en schaalparameter b*. Evenzo noemt men de klassen $\{F_X(x-a)\colon -\infty < a < \infty\}$ resp. $\{F_X(b^{-1}x)\colon b > 0\}$ de door F_X voortgebrachte familie van kansverdelingen met plaatsparameter a resp. schaalparameter b. Bij de in de waarschijnlijkheidsrekening meest voorkomende families heeft men één bepaalde representant vastgelegd ten opzichte waarvan de plaats- en/ of schaalparameter worden gedefinieerd. Deze representant wordt de *standaard-verdeling* van de familie genoemd. Indien P_X continu is met kansdichtheid $f_X = F'(x)$ dan heeft de verdeling van $Y = a + bX$, $b > 0$, als kansdichtheid

$$f_Y(x) = F_Y'(x) = \frac{1}{b} f_X(\frac{x-a}{b}).$$

De grafiek van f_Y wordt uit die van f_X verkregen door verschuiven over een afstand a, "uitrekken" in horizontale richting met een factor b en delen van de functiewaarden door b. De klassen van kansdichtheden $\{b^{-1} f_X(b^{-1}(x-a)): -\infty < a < \infty, b > 0\}$, $\{f_X(x-a): -\infty < a < \infty\}$ en $\{b^{-1} f_X(b^{-1}x): b > 0\}$ worden de door f_X voortgebrachte families en kansdichtheden met plaatsparameter a en/of schaalparameter b genoemd.

Een tweede probleem van dit type is het bepalen van de kansverdeling van $Z = X + Y$ waarbij X en Y o.o. stochastische grootheden zijn. Wanneer $P_{X,Y} = P_X \times P_Y$ de simultane kansverdeling van X en Y voorstelt, dan volgt uit (2.5.10) en de stelling van Fubini (stelling 1.6.7)

$$\begin{aligned} F_Z(z) &= \int_{R^2} I_{\{(x,y):x+y\leq z\}}(x,y)\, dP_{X,Y}(x,y) = \\ &= \int_{R^1} dP_X(x) \int_{R^1} I_{\{(x,y):y\leq z-x\}}(x,y)\, dP_Y(y) = \\ &= \int_{R^1} F_Y(z-x)\, dP_X(x), \qquad z \in R^1. \end{aligned} \tag{2.5.14}$$

Indien X en Y beide discreet verdeeld zijn, dan wordt de verdeling van Z bepaald door

$$P(Z=z) = \sum_x P(Y=z-x)\, P(X=x), \qquad z \in R^1. \tag{2.5.15}$$

Als X continu verdeeld is met $f_X = F_X'$, dan geldt

$$F_Z(z) = \int_{-\infty}^{\infty} F_Y(z-x)\, f_X(x)dx, \qquad z \in R^1. \tag{2.5.16}$$

Wanneer ook Y continu verdeeld is met $f_Y = F_Y'$, dan iz Z eveneens continu verdeeld. Uit (2.5.16) en de stelling van Fubini (stelling 1.6.7) volgt immers

$$F_Z(z) = \int_{-\infty}^{\infty} f_X(x)dx \int_{-\infty}^{z} f_Y(y-x)dy = \int_{-\infty}^{z} dy \int_{-\infty}^{\infty} f_Y(y-x)\, f_X(x)dx,$$

zodat Z continu is met kansdichtheid

$$(2.5.17) \qquad f_Z(z) = F_Z'(z) = \int_{-\infty}^{\infty} f_Y(z-x)\, f_X(x)dx, \qquad z \in R^1.$$

De kansverdeling P_Z (respectievelijk de verdelingsfunctie F_Z of de kansdichtheid f_Z) wordt de *convolutie* van de kansverdelingen P_X en P_Y (resp. van de verdelingsfuncties F_X en F_Y of de kansdichtheden f_X en f_Y) genoemd.

Uit (2.5.10) blijkt dat de kansverdeling P_Y van een functie $Y = \phi(X)$ van een stochastische vector X slechts afhangt van de kansverdeling P_X van X; de kansruimte $(\Omega,\mathcal{A},P)$ waarop X gedefinieerd is en de functie X zelf doen hierbij niet ter zake. Om de verdeling P_Y te bepalen kunnen wij dus naar believen een kansruimte $(\tilde\Omega,\tilde{\mathcal{A}},\tilde P)$ construeren, hierop een stochastische vector $\tilde X$ definiëren met de vereiste verdeling $\tilde P_{\tilde X} = P_X$ en hieruit de verdeling $\tilde P_{\tilde Y} = P_Y$ van $\tilde Y = \phi(\tilde X)$ berekenen. Daar een geschikte keuze van $(\tilde\Omega,\tilde{\mathcal{A}},\tilde P)$ en $\tilde X$ de berekeningen vaak aanzienlijk vereenvoudigt, wordt van dit *invariantie principe* veelvuldig gebruik gemaakt (zie §2.6).

Opgaven

1. Als X een stochastische grootheid is met een continue verdelingsfunctie F, dan is $U = F(X)$ een stochastische grootheid met $P(U \in B) = \lambda(B)$ voor iedere Borelverzameling $B \subset [0,1]$. Bewijs dit.
2. Zij F een verdelingsfunctie op R^1 en zij U een stochastische grootheid met $P(U \leq u) = u$ voor $0 \leq u \leq 1$. Als verder $F^{-1}(y) = \inf\{x: F(x) \geq y\}$ voor $0 < y < 1$, dan is $X = F^{-1}(U)$ een stochastische grootheid met F als verdelingsfunctie. Bewijs dit.
3. Bewijs dat een stochastische vector $X = (X_1,X_2,\ldots,X_k)$ dan en dan alleen een discrete verdeling heeft als de marginale verdelingen van $X_1, X_2, \ldots, X_k$ discreet zijn.
4. Zij $X = (X_1,X_2,\ldots,X_k)$ een stochastische vector met een dichtheid f_X ten opzichte van λ^k, en zij $Y = (X_1,X_2,\ldots,X_m)$ met $m < k$. Dan heeft Y een λ^m-absoluut continue verdeling met dichtheid

$$f_Y(x_1,\ldots,x_m) = \int_B f_X(u)\, d\lambda^k(u), \qquad (x_1,\ldots,x_m) \in R^m,$$

waarin

$$B = \{u: u = (u_1,\ldots,u_k),\ u_i = x_i \text{ voor } i = 1, 2, \ldots, m\}.$$

Bewijs dit.

5. Als $X_1, X_2, \ldots$ o.o. stochastische grootheden zijn met $P(X_n=0) = P(X_n=1) = \frac{1}{2}$ voor alle n, dan is ook

$$Y = 2 \sum_{n=1}^{\infty} \frac{X_n}{3^n}$$

een welgedefinieerde stochastische grootheid.

Bewijs:

a) $P(0 \leq Y \leq 1) = 1$;

b) $P(X_1=0) = P(0 \leq Y \leq \frac{1}{3})$, $P(X_1=1) = P(\frac{2}{3} \leq Y \leq 1)$;

c) Als $k \geq 1$ en $x_n = 0$ of 1 voor $n = 1, 2, \ldots, k$, dan geldt

$$P(2 \sum_1^k \frac{x_n}{3^n} \leq Y \leq 2 \sum_1^k \frac{x_n}{3^n} + \frac{1}{3^k}) = P(\bigcap_{n=1}^k \{X_n = x_n\}) = 2^{-k};$$

d) $P(\frac{i}{3^k} + \frac{1}{3^{k+1}} < Y < \frac{i}{3^k} + \frac{2}{3^{k+1}}) = 0$ voor $i = 0, 1, \ldots, 3^k - 1$, $k = 0, 1, 2, \ldots$;

e) Er is een gesloten verzameling $C \subset R^1$ met $\lambda(C) = 0$ en $P(Y \in C) = 1$;

f) De verdelingsfunctie van Y is continu en is op C^c differentieerbaar met afgeleide 0.

6. Als de stochastische grootheden $X_1, X_2, \ldots, X_k$ o.o. zijn en alle dezelfde verdelingsfunctie F bezitten, dan bezit $Y = \min(X_1,\ldots,X_k)$ de verdelingsfunctie $1 - (1-F)^k$ en $Z = \max(X_1,\ldots,X_k)$ de verdelingsfunctie F^k. Bewijs dit.

2.6. VOORBEELDEN VAN KANSVERDELINGEN

DIRECTE VERDELINGEN

2.6.1. *Gedegenereerde verdeling*

Een stochastische grootheid X bezit een gedegenereerde verdeling indien voor zekere $x_0 \in R^1$ geldt $P(X=x_0) = 1$.

2.6.2. *Alternatieve verdeling*

Een stochastische grootheid X bezit een alternatieve verdeling indien voor zekere $x_0, x_1 \in R^1$ en $0 < p < 1$ geldt $P(X=x_0) = p$. $P(X=x_1) = 1 - p$.

2.6.3. *Binomiale verdeling*

Beschouw een samengesteld experiment (Ω, A, P) bestaande uit n o.o. (deel) experimenten (Ω_i, A_i, P_i), $i = 1, 2, \ldots, n$. Zij S_i een eventualiteit waarvan het al dan niet optreden door de uitslag van het i^e experiment wordt bepaald (d.w.z. $S_i \in A_i$) en waarvoor $P_i(S_i) = p$, $i = 1, 2, \ldots, n$. Indien S_i optreedt zeggen wij dat het i^e experiment een succes oplevert. Zij X het aantal der eventualiteiten $S_1, S_2, \ldots, S_n$ dat bij uitvoering van het samengestelde experiment optreedt. Wegens de onafhankelijkheid van de (deel) experimenten is de kans op het optreden van precies x voorgeschreven eventualiteiten $S_{i_1}, S_{i_2}, \ldots, S_{i_x}$ gelijk aan $p^x(1-p)^{n-x}$. Daar er $\binom{n}{x}$ verschillende deelverzamelingen $\{i_1,i_2,\ldots,i_x\} \subset \{1,2,\ldots,n\}$ zijn, komen wij tot het volgende resultaat: Het *aantal successen* X *bij* n *o.o. experimenten met kans* p *op succes* bezit de kansverdeling

$$P(X=x) = \binom{n}{x} p^x(1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Deze verdeling wordt de binomiale verdeling met parameters n en p genoemd. Het hier beschreven model heeft in de praktijk vrijwel steeds betrekking op n o.o. uitvoeringen van hetzelfde experiment waarbij $S_1, S_2, \ldots, S_n$ steeds dezelfde gebeurtenis bij de verschillende uitvoeringen van het experiment aangeven. In voorbeeld 2.1.4 c) vonden wij reeds dat het aantal rode knikkers bij n aselecte trekkingen met teruglegging uit een vaas met r rode en (N-r) witte knikkers een binomiale verdeling met parameters n en $\frac{r}{N}$ bezit. Daar aselecte trekkingen met teruglegging o.o. experimenten zijn (opgave 4 in §2.3) en bij iedere trekking de kans op een rode knikker (succes) gelijk is aan $\frac{r}{N}$ is dit een speciaal geval van het hier behandelde algemene model.

Zij X_i het "aantal successen" bij het i^e experiment, d.w.z.

$$X_i = \begin{cases} 1 & \text{als } S_i \text{ optreedt} \\ 0 & \text{als } S_i^c \text{ optreedt,} \end{cases}$$

dan geldt $P(X_i=1) = p$, $P(X_i=0) = 1 - p$, $i = 1, 2, \ldots, n$, terwijl bovendien $X_1, X_2, \ldots, X_n$ o.o. zijn. Daar $X = \sum X_i$, is hiermee aangetoond dat de in ons model geconstrueerde binomiaal verdeelde stochastische grootheid X kan worden geschreven als de som van n o.o. stochastische grootheden die alle dezelfde alternatieve verdeling bezitten. Toepassing van het in § 2.5 genoemde invariantie principe levert: de som van n o.o. stochastische grootheden $X_1, X_2, \ldots, X_n$ met dezelfde alternatieve verdeling $P(X_i=1) = p$,

$P(X_i=0) = 1 - p$, $i = 1, 2, \ldots, n$, bezit een binomiale verdeling met parameters n en p. Indien (m+n) o.o. experimenten met kans p op succes worden uitgevoerd en X (resp. Y) het aantal successen bij de eerste , (resp. de laatste n) experimenten voorstelt, dan zijn X en Y o.o. en binomiaal verdeeld met parameters m en p resp. n en p, terwijl $Z = X + Y$ een binomiale verdeling met parameters (m+n) en p bezit. Volgens het invariantie principe geldt dus: de som van twee o.o. stochastische grootheden met binomiale verdelingen met parameters m en p resp. n en p bezit een binomiale verdeling met parameters (m+n) en p.

2.6.4. *Hypergeometrische verdeling*

Uit een verzameling van N elementen, *populatie* genaamd, waarvan r elementen een kenmerk R bezitten wordt zonder teruglegging een aselecte steekproef van n elementen getrokken. Indien X het aantal elementen in de steekproef voorstelt die het kenmerk R bezitten, dan geldt (zie voorbeeld 2.1.4 c))

$$P(X=x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},$$

voor $\max(0,n-N+r) \leq x \leq \min(r,n)$, x geheel. Deze verdeling wordt de hypergeometrische verdeling genoemd; N, r en n zijn de parameters van deze verdeling. Voor vaste n en x geldt

$$\lim_{\substack{r\to\infty \\ N-r\to\infty}} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \left[\binom{n}{x}\left(\frac{r}{N}\right)^x\left(1-\frac{r}{N}\right)^{n-x}\right]^{-1} =$$

$$= \lim_{\substack{r\to\infty \\ N-r\to\infty}} \frac{r!}{(r-x)!\,r^x} \cdot \frac{(N-r)!}{(N-r-n+x)!\,(N-r)^{n-x}} \cdot \frac{(N-n)!\,N^n}{N!} = 1;$$

indien zowel r als (N-r) veel groter zijn dan n, kan de binomiale verdeling met parameters n en $\frac{r}{N}$ dus als benadering dienen voor de hypergeometrische verdeling met parameters N, r en n. Daar de binomiale verdeling veelal eenvoudiger te berekenen is dan de hypergeometrische verdeling, wordt deze benadering veel gebruikt. Wanneer wij bedenken dat X bij trekkingen met teruglegging een binomiale verdeling met parameters n en $\frac{r}{N}$ bezit (zie voorbeeld 2.6.3), ligt deze benadering ook wel voor de hand: als r en (N-r) veel groter zijn dan n, dan is het intuïtief wel duidelijk dat al dan niet terugleggen weinig verschil maakt.

2.6.5. *Poisson verdeling*

Een stochastische grootheid X bezit een Poisson verdeling met parameter μ (>0) indien

$$P(X=x) = e^{-\mu}\frac{\mu^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

Deze verdeling kan worden opgevat als de limiet van een rij binomiale verdelingen met parameters n en p_n voor $n \to \infty$ en $np_n \to \mu$ (dus $p_n \to 0$). Immers, voor vaste x,

$$\lim_{\substack{n\to\infty \\ np_n\to\mu}} \binom{n}{x} p_n^x (1-p_n)^{n-x} = \frac{1}{x!}\lim \frac{n!\, p_n^x}{(n-x)!}(1-p_n)^{n-x} =$$

$$= \frac{1}{x!}\lim (np_n)^x (1-p_n)^n = \frac{1}{x!}\mu^x e^{-\mu}.$$

Voor grote n en kleine p kan de Poisson verdeling met $\mu = np$ dus als benadering dienen voor de binomiale verdeling met parameters n en p. Deze benadering wordt veel gebruikt.

Er zijn echter ook situaties waarin deze limietovergang reeds op natuurlijke wijze in het model plaatsvindt en waarbij dan in dit model een stochastische grootheid optreedt die niet bij benadering, doch exact een Poisson verdeling bezit. Beschouw hiertoe een proces dat zich in de tijd afspeelt en waarbij op ieder tijdstip een gebeurtenis S kan optreden (bijv. gespreksaanvragen bij een telefooncentrale). Zo'n proces wordt een *Poisson proces* met parameter λ (>0) genoemd indien aan de volgende voorwaarden is voldaan (de notatie $f(x) = o(g(x))$ voor $x \to 0$ betekent dat $\lim_{x\to 0} f(x)/g(x) = 0$):

a) In een tijdsinterval ter lengte Δt is de kans dat S éénmaal optreedt gelijk aan $\lambda\Delta t + o(\Delta t)$ voor $\Delta t \to 0$; de kans dat S in dit tijdsinterval meer dan éénmaal optreedt is $o(\Delta t)$ voor $\Delta t \to 0$.

b) Het optreden van S in disjuncte tijdsintervallen is onafhankelijk.

Zij X het aantal malen dat S optreedt in een tijdsinterval ter lengte T. Om de verdeling van X te bepalen verdelen wij het tijdsinterval in n gelijke delen ter lengte $\frac{T}{n}$ en definiëren $X_{i,n}$ als het aantal malen dat S optreedt in het i^e van deze deelintervallen. Uit a) en b) volgt dat voor $n \to \infty$ $P(X_{i,n}=1) = \frac{\lambda T}{n} + o(n^{-1})$, $P(X_{i,n}>1) = o(n^{-1})$ en dus $P(X_{i,n}=0) = 1 - \frac{\lambda T}{n} + o(n^{-1})$ voor $i = 1, 2, \ldots, n$, en voorts dat $X_{1,n}, \ldots, X_{n,n}$ o.o. zijn voor iedere n. Dus geldt voor gehele $x \geqq 0$

$$P(X=x) \geqq P(\{X=x\} \cap \{X_{i,n} \leqq 1 \text{ voor } i=1,2,\ldots,n\}) =$$

$$= P(\{x \text{ der } X_{i,n} \text{ zijn 1 en } (n-x) \text{ der } X_{i,n} \text{ zijn } 0\}) =$$

$$= \binom{n}{x}\left(\frac{\lambda T}{n} + o(n^{-1})\right)^x \left(1 - \frac{\lambda T}{n} + o(n^{-1})\right)^{n-x} \quad \text{voor } n \to \infty.$$

Nu geldt voor $n \to \infty$ en vaste x

$$\binom{n}{x} = \frac{n^x}{x!}(1+o(1)); \quad \left(\frac{\lambda T}{n} + o(n^{-1})\right)^x = \left(\frac{\lambda T}{n}\right)^x (1+o(1))$$

$$\left(1 - \frac{\lambda T}{n} + o(n^{-1})\right)^{n-x} = \left(1 - \frac{\lambda T}{n}\right)^n (1+o(1)) = e^{-\lambda T}(1+o(1)),$$

zodat

$$P(X=x) \geqq e^{-\lambda T}\frac{(\lambda T)^x}{x!}(1+o(1)) \to e^{-\lambda T}\frac{(\lambda T)^x}{x!}$$

voor $n \to \infty$. Daar echter $\sum_{x=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^x}{x!} = 1$, geldt

$$P(X=x) = e^{-\lambda T}\frac{(\lambda T)^x}{x!}, \qquad x = 0, 1, \ldots;$$

X bezit dus een Poisson verdeling met parameter $\mu = \lambda T$.

Zij X (resp. Y) het aantal malen dat S optreedt in het tijdsinterval $[0,\mu)$ (resp. $[\mu,\mu+\nu)$) bij een Poisson proces met parameter $\lambda = 1$. Dan zijn X en Y o.o. (veronderstelling b) en Poisson verdeeld met parameter μ resp. ν, terwijl $Z = X + Y$ een Poisson verdeling met parameter $(\mu+\nu)$ bezit. Toepassing van het invariantie principe levert: de som van twee o.o. stochastische grootheden met Poisson verdelingen met parameter μ resp. ν bezit een Poisson verdeling met parameter $(\mu+\nu)$.

2.6.6. *Negatief binomiale verdeling*

Beschouw een rij o.o. experimenten met kans op $p > 0$ op succes en zij X het nummer van het experiment waarbij het k^e succes optreedt, d.w.z. het aantal experimenten dat moet worden uitgevoerd om k successen te verkrijgen. Daar $\{X = x\} = \{(k-1)$ successen bij de eerste $(x-1)$ experimenten$\} \cap \{$succes bij het x^e experiment$\}$ en de eventualiteiten in het rechterlid kans $\binom{x-1}{k-1} p^{k-1} (1-p)^{x-k}$ resp. p bezitten en bovendien o.o. zijn, geldt

$$P(X=x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \qquad x = k, k+1, \ldots . \tag{2.6.1}$$

Deze verdeling wordt de negatief binomiale verdeling met parameters k en p genoemd. Deze naam wordt in de literatuur echter ook wel gebruikt voor de verdeling van $\tilde{X} = X - k$, het aantal mislukkingen dat aan het k^e succes voorafgaat,

$$\begin{aligned} P(\tilde{X}=x) &= \binom{x+k-1}{k-1} p^k (1-p)^x = \\ &= (-1)^x \binom{-k}{x} p^k (1-p)^x, \qquad x = 0, 1, \ldots . \end{aligned} \tag{2.6.2}$$

en is in feite aan deze laatste schrijfwijze ontleend. Voor $k = 1$ vindt men

$$P(X=x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots, \tag{2.6.3}$$

$$P(\tilde{X}=x) = p(1-p)^x, \qquad x = 0, 1, \ldots; \tag{2.6.4}$$

beide verdelingen worden in de literatuur aangeduid met de naam *geometrische verdeling* of ook wel *Pascal verdeling* met parameter p. Om misverstanden te vermijden zullen wij de naam negatief binomiale verdeling (c.q. geometrische of Pascal verdeling) uitsluitend gebruiken om de verdeling van X, zoals gegeven in (2.6.1) (c.q. (2.6.3)) aan te duiden.

Bij een rij o.o. experimenten met kans p op succes definiëren wij nu

X_1, $X_1 + X_2$, ..., $X_1 + X_2 + \ldots + X_k$ als de nummers van de experimenten waarbij het 1^e, 2^e, ..., k^e succes optreedt. X_1 is dus het aantal experimenten dat moet worden uitgevoerd om het eerste succes te verkrijgen, terwijl voor $i = 2, 3, \ldots, k$, X_i het aantal experimenten voorstelt dat na het verkrijgen van het $(i-1)^e$ succes nog moet worden uitgevoerd om het i^e succes te verkrijgen. De simultane verdeling van $X = (X_1,\ldots,X_k)$ wordt gevonden door op te merken dat voor iedere vector $x = (x_1,\ldots,x_k)$ van positieve gehele getallen de eventualiteit $\{X = x\}$ dan en alleen dan optreedt als rijen van (x_1-1), (x_2-1), ...,(x_k-1) mislukkingen steeds worden gevolgd door één succes. Dus geldt

$$P(X=x) = p^k\,(1-p)^{\sum x_i - k}, \qquad x_i \geqq 1,\ i = 1, 2, \ldots, k,$$

waaruit volgt dat X_1, X_2, ..., X_k o.o. zijn en geometrische verdelingen met parameter p bezitten. Daar $\sum X_i$ een negatief binomiale verdeling met parameters k en p bezit, volgt uit het invariantie principe dat de som van k o.o. stochastische grootheden met Pascal verdelingen met parameter p een negatief binomiale verdeling met parameters k en p bezit. Evenzo geldt dat de som van twee o.o. stochastische grootheden met negatief binomiale verdeling met parameters k en p resp. m en p, een negatief binomiale verdeling met parameters (k+m) en p bezit.

2.6.7. *Multinomiale verdeling*

Juist zoals in voorbeeld 2.6.3 gaan wij uit van een samengestelde experiment (Ω,A,P) bestaande uit n o.o. (deel) experimenten (Ω_i,A_i,P_i), $i = 1, 2, \ldots, n$. In plaats van twee eventualiteiten S_i, $S_i^c \in A_i$ beschouwen wij nu, voor iedere $i = 1, 2, \ldots, n$, k paarsgewijs disjuncte eventualiteiten $S_{i,1}$, $S_{i,2}$, ..., $S_{i,k}$ met $\bigcup_j S_{i,j} = \Omega_i$ en $P_i(S_{i,j}) = p_j$, $\sum p_j = 1$, voor $i = 1, 2, \ldots, n$. Zij X_j het aantal der eventualiteiten $S_{1,j}$, $S_{2,j}$, ..., $S_{n,j}$ dat bij uitvoering van het samengestelde experiment optreedt en zij $X = (X_1,X_2,\ldots,X_k)$. Indien $x = (x_1,x_2,\ldots,x_k)$ een vector van van niet-negatieve gehele getallen met $\sum x_j = n$ voorstelt, dan is de kans dat voor iedere $j = 1, 2, \ldots, k$ precies x_j voorgeschreven eventualiteit uit $\{S_{1,j}, S_{2,j}, \ldots, S_{n,j}\}$ optreden gelijk aan $\prod_j p_j^{x_j}$. Daar de verzameling

$\{1, 2, \ldots, n\}$ op $n!\ (\prod_j x_j!)^{-1}$ verschillende manieren te splitsen is in k disjuncte deelverzamelingen waarvan de j^e x_j elementen bevat, komen wij tot het volgende resultaat:

$$P(X=x) = \frac{n!}{x_1!\ x_2!\ \ldots\ x_k!}\ p_1^{x_1}\ p_2^{x_2} \ldots p_k^{x_k}$$

voor iedere vector $x = (x_1,x_2,\ldots,x_k)$ van niet-negatieve gehele getallen met $\sum x_j = n$. Deze verdeling wordt de multinomiale verdeling met parameters n en $p = (p_1,p_2,\ldots,p_k)$ genoemd. Uit het model volgt rechtstreeks dat de marginale verdeling van X_j een binomiale verdeling met parameters n en p_j is. In de praktijk heeft het hier beschreven model vrijwel steeds betrekking op n o.o. uitvoeringen van hetzelfde experiment waarbij voor iedere j $S_{1,j}, S_{2,j}, \ldots, S_{n,j}$ steeds dezelfde gebeurtenis bij de verschillende uitvoeringen van het experiment aangeven; X_j is dan het aantal malen dat deze gebeurtenis optreedt.

Continue verdelingen

Voor continue verdelingen van stochastische grootheden (c.q. vectoren) X zal de kansdichtheid f_X steeds worden gedefinieerd door (2.5.8) voor $k = 1$ (c.q. voor $k > 1$); f_X is dus steeds een kansdichtheid van X ten opzichte van de Lebesgue maat op $\mathcal{B}^1$ (c.q. op $\mathcal{B}^k$).

2.6.8. *Homogene verdeling*

Een stochastische grootheid X bezit een homogene verdeling (ook wel *uniforme verdeling* genaamd) op (a,b) indien voor $a < b \in R^1$

$$f_X(x) = \begin{cases} (b-a)^{-1} & \text{voor } a < x < b \\ 0 & \text{anders.} \end{cases}$$

Dan geldt

$$F_X(x) = \begin{cases} 0 & \text{voor } x \leq a \\ (x-a)\ (b-a)^{-1} & \text{voor } a < x < b \\ 1 & \text{voor } x \geq b. \end{cases}$$

2.6.9. *Gamma verdeling*

Een Poisson proces met parameter λ (zie voorbeeld 2.6.5) wordt vanaf het tijdstip 0 waargenomen. Zij X het tijdstip waarop voor de k^e maal de gebeurtenis S optreedt, d.w.z. de *wachttijd* tot de k^e gebeurtenis. Voor $x > 0$ treedt de eventualiteit $\{X \leq x\}$ dan en slechts dan op indien de gebeurtenis S in het tijdsinterval $(0,x]$ ten minste k maal optreedt. Indien Y het aantal malen voorstelt dat S optreedt in het tijdsinterval $(0,x]$, dan bezit Y een Poisson verdeling met parameter $\mu = \lambda x$ (zie voorbeeld 2.6.5). Dus geldt

$$(2.6.5) \qquad \begin{aligned} F_X(x) &= P(\underline{X} \leq x) = P(\underline{Y} \geq k) = 1 - P(Y<k) = \\ &= 1 - \sum_{j=0}^{k-1} e^{-\lambda x} \frac{(\lambda x)^j}{j!}, \qquad x > 0. \end{aligned}$$

Daar F_X continu differentieerbaar is op $(0,\infty)$ is X continu verdeeld met kansdichtheid

$$(2.6.6) \qquad \begin{aligned} f_X(x) &= F_X'(x) = \lambda \sum_{j=0}^{k-1} e^{-\lambda x} \frac{(\lambda x)^j}{j!} - \lambda \sum_{j=1}^{k-1} e^{-\lambda x} \frac{(\lambda x)^{j-1}}{(j-1)!} = \\ &= \frac{\lambda^k}{(k-1)!} e^{-\lambda x} x^{k-1}, \qquad x > 0. \end{aligned}$$

De verdelingsfunctie F_X is dus ook te schrijven als

$$(2.6.7) \qquad F_X(x) = \frac{\lambda^k}{(k-1)!} \int_0^x e^{-\lambda y} y^{k-1}\, dy = \frac{1}{(k-1)!} \int_0^{\lambda x} e^{-y} y^{k-1}\, dy, \; x > 0,$$

waaruit (2.6.5) door herhaalde partiële integratie kan worden teruggevonden. Daar de integraal in het rechterlid van (2.6.7) bekend staat als een onvolledige gamma functie, wordt deze kansverdeling met parameters k en λ genoemd. Voor $k = 1$ ontstaat de zogenaamde *exponentiële verdeling* met parameter λ waarvan de kansdichtheid en de verdelingsfunctie worden gegeven door

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x > 0,$$

$$F_X(x) = 1 - e^{-\lambda x}, \qquad x > 0.$$

Laten nu $Z_1 = X_1$, $Z_2 = X_1 + X_2$, ..., $Z_k = X_1 + X_2 + \dots + X_k$ de tijdstippen voorstellen waarop de gebeurtenis S voor de 1^e, 2^e, ..., k^e maal optreedt. X_1 is dus de wachttijd tot de eerste gebeurtenis, terwijl voor $i = 2, 3, \dots, k$, X_i de wachttijd van de $(i-1)^e$ tot de i^e gebeurtenis voorstelt. Om de simultane kansdichtheid f_Z van $Z_1, Z_2, \dots, Z_k$ te bepalen merken wij op dat de eventualiteit $\bigcap_i \{z_i < Z_i \leq z_i + h_i\}$, $0 < z_1 < z_2 < \dots < z_k$, $0 < h_i \leq z_{i+1} - z_i$, $i = 1, 2, \dots, k$, dan en slechts dan optreedt als S niet optreedt in de tijdsintervallen $(0,z_1]$, $(z_1+h_1, z_2]$, ..., $(z_{k-1}+h_{k-1}, z_k]$ en precies éénmaal optreedt in ieder van de intervallen $(z_i, z_i+h_i]$, $i = 1, 2, \dots, k$. Uit de definitie van het Poisson proces en zijn relatie tot de Poisson verdeling (zie voorbeeld 2.6.5) volgt dat voor $h_i \to 0$ voor $i = 1, 2, \dots, k$

$$\int_{z_k}^{z_k+h_k} \dots \int_{z_1}^{z_1+h_1} f_Z(y_1,\dots,y_k)\, dy_1 \dots dy_k = P(\bigcap_i \{z_i < Z_i \leq z_i + h_i\}) =$$

$$= e^{-\lambda z_1}\, e^{-\lambda(z_2-z_1-h_1)} \dots e^{-\lambda(z_k-z_{k-1}-h_{k-1})} \prod_{i=1}^{k} (\lambda h_i) + o(\prod_{i=1}^{k} h_i).$$

Deling door Πh_i gevolgd door de limietovergang $h_i \to 0$, $i = 1, 2, \dots, k$, levert

$$f_Z(z_1,\dots,z_k) = \lambda^k e^{-\lambda z_k}, \qquad 0 < z_1 < z_2 < \dots < z_k.$$

De simultane verdelingsfunctie van $X_1, \dots, X_k$ in het punt $(x_1,\dots,x_k)$ wordt gevonden door f_Z te integreren over de verzameling $\{(z_1,\dots,z_k)\colon z_1 \leq x_1,$ $z_i - z_{i-1} \leq x_i$, $i = 2, 3, \dots, k\}$. Daar de determinant van Jacobi van de transformatie $z_1 = x_1$, $z_2 = x_1 + x_2$, ..., $z_k = x_1 + x_2 + \dots + x_k$ gelijk is aan 1, levert overgang op de nieuwe variabelen $x_1, x_2, \dots, x_k$ in de integraal

$$P(\bigcap_{i=1}^{k} \{X_i \leq x_i\}) = \lambda^k \int_0^{x_k} \dots \int_0^{x_1} e^{-\lambda \sum_{i=1}^{k} x_i}\, dx_1 \dots dx_k =$$

$$= \prod_{i=1}^{k} (1-e^{-\lambda x_i}), \quad x_i > 0,\ i = 1, 2, \dots, k.$$

$X_1, X_2, \ldots, X_k$ zijn dus o.o. en identiek verdeeld met exponentiële verdelingen met parameter λ. Daar Z_k een gamma verdeling met parameters k en λ bezit, volgt uit het invariantie principe dat de som van k o.o. stochastische grootheden met exponentiële verdelingen met parameter λ een gamma verdeling met parameters k en λ bezit. Evenzo geldt dat de som van twee o.o. stochastische grootheden met gamma verdelingen met parameters k en λ resp. m en λ, een gamma verdeling met parameters (k+m) en λ bezit. Dat de relatie tussen de gamma verdeling en de Poisson verdeling een sterke analogie vertoont met die tussen de negatief binomiale en de binomiale verdeling, zal na het voorafgaande wel duidelijk zijn.

Uit (2.6.5), (2.6.6) of (2.6.7) volgt dat voor vaste k de gamma verdelingen een familie van kansverdelingen met schaalparameter λ^{-1} vormen (zie §2.5); als standaard verdeling wordt de verdeling met $\lambda = 1$ gekozen met als kansdichtheid

$$f(x) = \frac{1}{(k-1)!}\, e^{-x}\, x^{k-1}, \qquad x > 0.$$

Indien $\tilde{X}$ een standaard gamma verdeling met parameter k bezit, dan heeft $X = \lambda^{-1}\tilde{X}$ dus een gamma verdeling met parameters k en λ. Dit is ook zonder meer wel duidelijk uit de definitie van het Poisson proces. Een Poisson proces met parameter 1 gaat over in een Poisson proces met parameter λ wanneer de eenheid van tijd met een factor λ wordt vermenigvuldigd.

Bij de definitie van de gamma verdeling hebben wij ons tot nu toe beperkt tot positieve gehele waarden van de parameter k. Naar analogie van (2.6.6) noemen wij nu voor willekeurige reële $k > 0$ en $\lambda > 0$ de kansverdeling met dichtheid

$$f_X(x) = \frac{\lambda^k}{\Gamma(k)}\, e^{-\lambda x}\, x^{k-1}, \qquad x > 0, \tag{2.6.8}$$

de gamma verdeling met parameters k en λ. Hierin is

$$\Gamma(k) = \int_0^\infty e^{-x}\, x^{k-1}\, dx, \qquad k > 0,$$

de (volledige) gamma functie. Met behulp van partiële integratie vindt men $\Gamma(k+1) = k\,\Gamma(k)$ voor $k > 0$; daar $\Gamma(1) = 1$ geldt $\Gamma(k) = (k-1)!$ voor positieve gehele k, zodat (2.6.8) in dit geval overgaat in (2.6.6) en deze beide definities dus consistent zijn. Hoewel voor niet gehele k de relatie met het Poisson proces is verbroken, kan men rechtstreeks met behulp van (2.5.17) aantonen dat de som van twee o.o. stochastische grootheden met gamma verdelingen met parameters k en λ resp. m en λ ook voor niet gehele k en m een gamma verdeling met parameters (k+m) en λ bezit. Ook de conclusie dat λ^{-1} een schaalparameter van de gamma verdeling is blijft voor niet gehele k geldig; als standaard gamma verdeling met parameter k wordt ook nu het geval $\lambda = 1$ gekozen.

2.6.10. *Dubbel-exponentiële verdeling*

Een stochastische grootheid X bezit een dubbel-exponentiële verdeling met parameter $\lambda > 0$ (ook wel *Laplace verdeling* genaamd) indien

$$f_X(x) = \frac{1}{2}\lambda\, e^{-\lambda|x|}, \qquad -\infty < x < \infty.$$

Daar de integraal van f_X over R^1 gelijk is aan 1, is f_X inderdaad een kansdichtheid. Juist zoals bij de exponentiële verdeling is λ^{-1} een schaalparameter; als standaard verdeling wordt het geval $\lambda = 1$ gekozen. De naam Laplace verdeling kan tot verwarring aanleiding geven aangezien in het Frans de normale verdeling (zie voorbeeld 2.6.12) naar Laplace wordt genoemd.

2.6.11. *Cauchy verdeling*

Een stochastische grootheid X bezit een standaard Cauchy verdeling als

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty,$$

zodat

$$F_X(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{dy}{1+y^2} = \frac{1}{2} + \frac{1}{\pi}\,\mathrm{arctn}\, x, \qquad -\infty < x < \infty.$$

Daar $F_X(\infty) = 1$ is dit inderdaad een kansverdeling. De door deze verdeling voortgebrachte familie met plaatsparameter a en schaalparameter $b > 0$ wordt de familie van Cauchy verdelingen genoemd.

2.6.12. *Normale verdeling*

Een stochastische grootheid X bezit een normale verdeling met parameters $-\infty < \mu < \infty$ en $\sigma^2 > 0$ indien

$$(2.6.9) \qquad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty,$$

waarbij σ de positieve wortel uit σ^2 voorstelt. Deze verdeling wordt symbolisch als $N(\mu,\sigma^2)$-verdeling aangeduid. De dichtheid (2.6.9) is symmetrisch om het punt $x = \mu$, neemt zijn maximum aan in $x = \mu$ en bezit twee buigpunten in $x = \mu \pm \sigma$. Door overgang op poolcoördinaten blijkt

$$\left(\int_{-\infty}^{\infty} f_X(x)dx\right)^2 = \frac{1}{2\pi}\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx\right)^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx\, dy =$$

$$= \frac{1}{2\pi}\int_0^{2\pi} d\phi \int_0^{\infty} r\, e^{-\frac{1}{2}r^2} dr = 1$$

en daar $f_X \geqq 0$, is f_X dus inderdaad een kansdichtheid voor iedere $-\infty < \mu < \infty$ en iedere $\sigma^2 > 0$.

De normale verdelingen vormen een familie van kansverdelingen met plaatsparameter μ en schaalparameter σ, waarbij als *standaard normale verdeling* de $N(0,1)$-verdeling wordt gekozen. De standaard normale dichtheid en de standaard normale verdelingsfunctie worden met ϕ resp. Φ aangegeven:

$$(2.6.10) \qquad \phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}, \qquad -\infty < x < \infty,$$

$$(2.6.11) \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy, \qquad -\infty < x < \infty.$$

De integraal (2.6.11) kan niet in elementaire functies worden uitgedrukt. Er bestaan echter uitgebreide tabellen van de functie Φ. Wegens de symmetrie van $\phi(x)$ om het punt $x = 0$ geldt voor alle x

$$(2.6.12) \qquad \Phi(-x) = 1 - \Phi(x),$$

waardoor met een tabel van Φ voor positieve waarden van het argument kan worden volstaan.

Indien X een $N(\mu,\sigma^2)$-verdeling bezit, dan wordt de verdelingsfunctie van X gegeven door

$$(2.6.13)\qquad F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}y^2} dy = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

zodat de waarden van F_X voor iedere normale verdeling in een tabel van Φ kunnen worden gevonden. Voor $b > 0$ wordt de verdelingsfunctie van $Y = a + bX$ gegeven door

$$F_Y(y) = P(a+bX\leq y) = F_X\left(\frac{y-a}{b}\right) = \Phi\left(\frac{y-a-b\mu}{b\sigma}\right),$$

zodat $Y = a + bX$ voor $b > 0$ en $N(a+b\mu, b^2\sigma^2)$-verdeling bezit. Merk op dat dit laatste resultaat zowel als (2.6.13) niet anders dan een herhaling is van de reeds uit (2.6.9) getrokken conclusie dat de normale verdelingen een familie kansverdelingen vormen met plaatsparameter μ en schaalparameter σ. Uit (2.6.12) volgt nu nog dat de verdelingsfunctie van $Y = a + bX$ voor $b < 0$ wordt gegeven door

$$F_Y(y) = P(a+bX\leq y) = 1 - \Phi\left(\frac{y-a-b\mu}{b\sigma}\right) = \Phi\left(\frac{y-a-b\mu}{|b|\sigma}\right),$$

zodat $Y = a + bX$ voor alle $b \neq 0$ en $N(a+b\mu, b^2\sigma^2)$-verdeling bezit. Voor $a = -\mu/\sigma$, $b = 1/\sigma$ volgt dat $(X-\mu)/\sigma$, de *gestandaardiseerde van* X, een standaard normale verdeling bezit; deze uitspraak is equivalent aan (2.6.13).

In voorbeeld 2.6.13 zal worden aangetoond dat de som van twee o.o. stochastische grootheden met resp. een $N(\mu_1,\sigma_1^2)$- en een $N(\mu_2,\sigma_2^2)$-verdeling, een $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$-verdeling bezit.

2.6.13. *Meerdimensionale normale verdeling*

Zij $A = (a_{ij})$ een symmetrische reële $(k\times k)$-matrix. A heet positief definiet als voor iedere reële kolomvector $c = (c_1,\ldots,c_k)'$ de kwadratische vorm $c'Ac = \sum_i \sum_j a_{ij} c_i c_j > 0$ is, tenzij $c = 0$. Daar A symmetrisch is, bestaat er een orthogonale matrix P zodanig dat $P'AP = (\lambda_i \delta_{ij})$, een diagonaal-matrix met elementen $\lambda_1, \lambda_2, \ldots, \lambda_k$; voor iedere orthogonale P met deze eigenschap vinden wij dezelfde waarden voor de λ_i hoewel mogelijk in een andere volgorde. $\lambda_1, \lambda_2, \ldots, \lambda_k$ heten de eigenwaarden van A. Als A positief definiet is, dan is ook $P'AP$ positief definiet want voor $c \neq 0$ is $d = Pc \neq 0$ zodat $c'P'APc = d'Ad > 0$. Echter $c'P'APc = \sum \lambda_i c_i^2$ en uit $\sum \lambda_i c_i^2 > 0$ voor alle $c > 0$ volgt $\lambda_i > 0$ voor $i = 1, 2, \ldots, k$. Omgekeerd: Als $\lambda_i > 0$ voor $i = 1, \ldots, k$, dan geldt voor iedere $c \neq 0$ dat $b = P'c \neq 0$ en dus dat $c'Ac = c'P(P'AP)\,P'c = \sum \lambda_i b_i^2 > 0$. Een symmetrische matrix A is dus dan en slechts dan positief definiet als zijn eigenwaarden λ_i alle positief zijn. Daar $|A| = |P'AP| = \prod_i \lambda_i$ heeft een positief definiete matrix A een positieve determinant $\prod_i \lambda_i$ en is dus niet-singulier. Daar $P'A^{-1}P$ de inverse is van $P'AP = (\lambda_i \delta_{ij})$, geldt $P'A^{-1}P = (\lambda_i^{-1}\delta_{ij})$. De inverse van een positief definiete matrix A is dus positief definiet en zijn eigenwaarden zijn de inversen van de eigenwaarden van A.

Zij $\Sigma = (\sigma_{ij})$ een reële symmetrische positief definiete $(k\times k)$-matrix met inverse $\Sigma^{-1} = (\sigma^{ij})$ en zij $\mu = (\mu_1, \mu_2, \ldots, \mu_k)'$ een reële kolomvector. De stochastische vector $X = (X_1, X_2, \ldots, X_k)$ bezit een k-dimensionale normale verdeling als

$$f_X(x_1,\ldots,x_k) = \frac{1}{(2\pi)^{k/2}(|\Sigma|)^{1/2}}\, e^{-\frac{1}{2}Q(x_1,\ldots,x_k)} \tag{2.6.14}$$

voor $-\infty < x_i < \infty$, $i = 1, 2, \ldots, k$, waarbij

$$Q(x_1,\ldots,x_k) = (x-\mu)'\,\Sigma^{-1}\,(x-\mu) = \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma^{ij}\,(x_i-\mu_i)(x_j-\mu_j), \tag{2.6.15}$$

en $|\Sigma|^{\frac{1}{2}}$ de positieve wortel uit de determinant van Σ voorstelt. Deze verdeling wordt symbolisch als $N(\mu, \Sigma)$-verdeling aangeduid. Voor $k = 1$ is (2.6.14) de dichtheid van de (één-dimensionale) $N(\mu_1, \sigma^{11})$-verdeling.

Laten $\lambda_1, \lambda_2, \ldots, \lambda_k$ de eigenwaarden van Σ voorstellen en zij P de orthogonale matrix waarvoor geldt $P'\,\Sigma^{-1}\,P = (\lambda_i^{-1}\delta_{ij})$. Beschouw de transformatie $z = P'(x-\mu)$, $x - \mu = Pz$. Dan geldt

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1,\ldots,x_k)\, dx_1 \ldots dx_k =$$

$$= \frac{1}{(2\pi)^{k/2}(\prod_i \lambda_i)^{1/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}z'P'\Sigma^{-1}Pz}\; ||P||\, dz_1 \ldots dz_k =$$

$$= \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\lambda_i}} \int_{-\infty}^{\infty} e^{-z_i^2/2\lambda_i}\, dz_i = 1,$$

daar de i^e factor in dit product de integraal van de dichtheid van de $N(0,\lambda_i)$-verdeling voorstelt. Daar $f_X > 0$ is f_X dus inderdaad een kansdichtheid.

Indien $X = (X_1,\dots,X_k)$ een $N(\mu,\Sigma)$-verdeling bezit dan wordt de kansdichtheid van $\tilde{X} = (X_1,\dots,X_{k-1})$ gegeven door (zie §2.5):

$$f_{\tilde{X}}(x_1,\dots,x_{k-1}) = \int_{-\infty}^{\infty} f_X(x_1,\dots,x_k)\, dx_k =$$

$$= C \exp\{-\frac{1}{2} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} \sigma^{ij} (x_i-\mu_i)(x_j-\mu_j)\}.$$

$$\cdot \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} \sigma^{kk}(x_k-\mu_k)^2 - \sum_{j=1}^{k-1} \sigma^{kj}(x_j-\mu_j)(x_k-\mu_k)\}\, dx_k =$$

$$= C \exp\{-\frac{1}{2} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} (\sigma^{ij} - \frac{\sigma^{ki}\sigma^{kj}}{\sigma^{kk}}) (x_i-\mu_i)(x_j-\mu_j)\}.$$

$$\cdot \int_{-\infty}^{\infty} \exp\{-\frac{1}{2} \sigma^{kk}[x_k-\mu_k + \sum_{j=1}^{k-1} \frac{\sigma^{kj}}{\sigma^{kk}} (x_j-\mu_j)]^2\}\, dx_k =$$

$$= \tilde{C} \exp\{-\frac{1}{2} \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} (\sigma^{ij} - \frac{\sigma^{ki}\sigma^{kj}}{\sigma^{kk}}) (x_i-\mu_i)(x_j-\mu_j)\},$$

waarin C en $\tilde{C}$ constanten voorstellen. Zij $\tilde{\Sigma}$ de $(k-1) \times (k-1)$-matrix die uit Σ ontstaat door weglaten van de k^e kolom en de k^e rij. Dan is $\tilde{\Sigma}^{-1}$ de matrix met elementen $\sigma^{ij} - \sigma^{ki}\sigma^{kj}(\sigma^{kk})^{-1}$. Immers, voor $i, j = 1, 2, \dots, k-1$,

$$\sum_{s=1}^{k-1} (\sigma^{is} - \frac{\sigma^{ki}\sigma^{ks}}{\sigma^{kk}})\, \sigma_{sj} = \delta_{ij} - \sigma^{ik}\sigma_{kj} - \frac{\sigma^{ki}}{\sigma^{kk}} (-\sigma^{kk}\sigma_{kj}) = \delta_{ij}.$$

Dus geldt

$$f_{\tilde{X}}(x_1,\dots,x_{k-1}) = \tilde{C} \exp\{-\frac{1}{2} (\tilde{x}-\tilde{\mu})' \tilde{\Sigma}^{-1} (\tilde{x}-\tilde{\mu})\}$$

waarbij $\tilde{x} = (x_1,\ldots,x_{k-1})'$ en $\tilde{\mu} = (\mu_1,\ldots,\mu_{k-1})'$. Daar $f_{\tilde{X}}$ een kansdichtheid is en $\tilde{\Sigma}$ positief definiet is, leidt het vergelijken van deze uitdrukking voor $f_{\tilde{X}}$ met (2.6.14) tot de conclusie dat

$$\tilde{C} = \{(2\pi)^{\frac{k-1}{2}} (|\tilde{\Sigma}|)^{\frac{1}{2}}\}^{-1}.$$

$\tilde{X}$ is dus $N(\tilde{\mu},\tilde{\Sigma})$-verdeeld. Door permutatie van $X_1, \ldots, X_k$ en herhaling van het bovenstaande procedé vinden wij: Indien $X = (X_1,\ldots,X_k)$ een $N(\mu,\Sigma)$-verdeling bezit, dan bezit $\tilde{X} = (X_{\nu_1},X_{\nu_2},\ldots,X_{\nu_m})$, $1 \leqq \nu_1 < \nu_2 < \ldots < \nu_m \leqq k$ een $\tilde{N}(\mu,\tilde{\Sigma})$-verdeling, waarbij $\tilde{\mu} = (\mu_{\nu_1},\mu_{\nu_2},\ldots,\mu_{\nu_m})'$ en $\tilde{\Sigma} = (\sigma_{\nu_i,\nu_j})$, $i, j = 1, 2, \ldots, m$. Door in dit resultaat $m = 1$ te kiezen vinden wij dat de stochastische grootheid X_i een (ééndimensionale) $N(\mu_i,\sigma_{ii})$-verdeling bezit, $i = 1, 2, \ldots, k$. De vector μ en de diagonaal-elementen σ_{ii} van Σ zijn dus juist de parameters van de marginale normale verdelingen van $X_1, X_2, \ldots, X_k$.

De overige parameters σ_{ij}, $i \neq j$, van de $N(\mu,\Sigma)$-verdeling bepalen, zoals wij later zullen zien, de mate van afhankelijkheid van $X_1, X_2, \ldots, X_k$. In dit stadium constateren wij slechts dat $X_1, X_2, \ldots, X_k$ dan en slechts dan o.o. zijn indien $\sigma_{ij} = 0$ voor ieder paar indices $i \neq j$. Nodig en voldoende voor onafhankelijkheid is immers dat

$$\prod_{i=1}^{k} f_{X_i}(x_i) = \frac{1}{(2\pi)^{k/2}(\prod_i \sigma_{ii})^{1/2}} \exp\{-\frac{1}{2}\sum_{i=1}^{k}\sigma_{ii}^{-1}(x_i-\mu_i)^2\}$$

een versie van f_X is en hieraan is dan en slechts dan voldaan als $\sigma_{ij} = 0$ voor alle $i \neq j$.

Tenslotte tonen wij aan dat meerdimensionale normaliteit door een lineare transformatie van de stochastische vector niet verloren gaat. Zij $X = (X_1,\ldots,X_k)'$ een stochastische vector met een $N(\mu,\Sigma)$-verdeling, $B = (b_{ij})$ een niet-singuliere $(k\times k)$-matrix en $a = (a_1,\ldots,a_k)'$. Beschouw de stochastische vector $Z = BX + a$, d.w.z. $Z = (Z_1,\ldots,Z_k)'$ en $Z_i = \sum_j b_{ij} X_j + a_i$, $i = 1, 2, \ldots, k$. De verdelingsfunctie van Z in het punt $z = (z_1,\ldots,z_k)'$ wordt gegeven door

$$F_Z(z) = \int \cdots \int_{Bx + a \leq z} f_X(x_1, \ldots, x_k)\, dx_1 \ldots dx_k .$$

In deze integraal gaan wij over op nieuwe variabelen $y = Bx + a$, d.w.z. $x = B^{-1}(y-a)$. De kwadratische vorm $(x-\mu)' \Sigma^{-1} (x-\mu)$ in de exponent in (2.6.14) gaat over in

$$(B^{-1}(y-a)-\mu)' \Sigma^{-1} (B^{-1}(y-a)-\mu) = (y-B\mu-a)' (B^{-1})' \Sigma^{-1} B^{-1}(y-B\mu-a) =$$

$$= (y-\mu^*)' \Sigma^{*^{-1}} (y-\mu^*),$$

waarbij

(2.6.16) $\qquad \mu^* = B\mu + a, \ \Sigma^* = B \Sigma B'.$

Dus geldt

$$F_Z(z) = \int_{-\infty}^{z_k} \cdots \int_{-\infty}^{z_1} \frac{1}{(2\pi)^{k/2} (|\Sigma|)^{1/2}} \exp\{-\tfrac{1}{2}(y-\mu^*)' \Sigma^{*^{-1}} (y-\mu^*)\}$$

$$. \ ||B^{-1}||\ dy_1 \ldots dy_k =$$

$$= \int_{-\infty}^{z_k} \cdots \int_{-\infty}^{z_1} \frac{1}{(2\pi)^{k/2} (|\Sigma^*|)^{1/2}} \exp\{-\tfrac{1}{2}(y-\mu^*)' \Sigma^{*^{-1}} (y-\mu^*)\}$$

$$. \ dy_1 \ldots dy_k ,$$

aangezien $(|\Sigma^*|)^{\frac{1}{2}} = (|\Sigma|)^{\frac{1}{2}}\, ||B|| = (|\Sigma|)^{\frac{1}{2}}\, ||B^{-1}||^{-1}$. Daar Σ^* positief definiet is, hebben wij dus het volgende resultaat verkregen: Indien $X = (X_1, \ldots, X_k)'$ een $N(\mu, \Sigma)$-verdeling bezit en B een niet-singuliere $(k \times k)$-matrix voorstelt, dan bezit $Z = BX + a$ een $N(\mu^*, \Sigma^*)$-verdeling met μ^* en Σ^* gegeven door (2.6.16).

Een tweetal rechtstreekse gevolgen van dit resultaat verdienen bijzondere aandacht. Daar B zodanig kan worden gekozen dat $\Sigma^* = B \Sigma B'$ een diagonaal-matrix is (d.w.z. $\sigma^*_{ij} = 0$ voor alle $i \neq j$), is het steeds mogelijk om uit een $N(\mu, \Sigma)$-verdeelde stochastische vector $X = (X_1, \ldots, X_k)'$ door een lineaire transformatie een stochastische vector $Z = BX + a$ te verkrijgen waarvan de componenten $Z_1, \ldots, Z_k$ o.o. zijn en (ééndimensionale) normale verdelingen bezitten. Merk op dat dit verschijnsel al impliciet werd uitgebuit in het bewijs dat f_X een kansdichtheid is. Daar Σ positief definiet is kan men B zelfs zo kiezen dat $\Sigma^* = B \Sigma B' = I$ (de eenheidsmatrix); kiest men dan tevens $a = -B\mu$, dan volgt dat $Z = B(X-\mu)$ een $N(0,I)$-verdeling bezit, d.w.z. dat $Z_1, \ldots, Z_k$ o.o. zijn en standaard normale verdelingen bezitten.

Een tweede gevolg van ons resultaat is dat $Z_1 = \sum_j b_{1j} X_j + a_1$ een (ééndimensionale) $N(\mu_1^*, \sigma_{11}^*)$-verdeling bezit. Daar iedere $b = (b_1, \ldots, b_k) \neq 0$ als eerste rij van een niet-singuliere matrix B kan fungeren, volgt hieruit dat, indien $X = (X_1, \ldots, X_k)$ een $N(\mu, \Sigma)$-verdeling bezit, de stochastsiche grootheid $Y = \sum_j b_j X_j + a$ een $N(\sum_j b_j \mu_j + a, \sum_i \sum_j \sigma_{ij} b_i b_j)$-verdeling bezit. Beperken wij ons tot het geval dat Σ een diagonaal-matrix is, dan kunnen wij dit ook als volgt formuleren: Indien de stochastische grootheden $X_1, \ldots, X_k$ o.o. zijn en X_j een $N(\mu_j, \sigma_j^2)$-verdeling bezit, $j = 1, 2, \ldots, k$, dan bezit $Y = \sum_j b_j X_j + a$ een $N(\sum_j b_j \mu_j + a, \sum_j b_j^2 \sigma_j^2)$-verdeling. Dit resultaat werd reeds in voorbeeld 2.6.12 aangekondigd.

Opgaven

1. De som van twee o.o. stochastische grootheden met binomiale verdelingen met parameters m en p resp. n en p bezit een binomiale verdeling met parameters (m+n) en p (zie voorbeeld 2.6.3). Bewijs dit, zonder gebruik te maken van het invariantie principe, rechtstreeks met behulp van (2.5.15).
2. Uit een populatie van N elementen waarvan r een kenmerk R bezitten wordt zonder teruglegging een aselecte steekproef van n elementen getrokken. Zij, voor $i = 1, 2, \ldots, n$,

$$X_i = \begin{cases} 1 & \text{als } i^e \text{ element in steekproef kenmerk R bezit} \\ 0 & \text{anders.} \end{cases}$$

   Geef de kansverdeling van X_i en de simultane kansverdeling van X_i en X_j, $i \neq j$. Zijn X_i en X_j o.o.?
3. Voor de hypergeometrische verdeling uit voorbeeld 2.6.4 geldt

$$P(X=x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{n}{x}\binom{N-n}{r-x}}{\binom{N}{r}} .$$

Geef een kanstheoretische interpretatie van deze laatste uitdrukking en leid een benadering voor $P(X=x)$ af voor het geval zowel n als $(N-n)$ veel groter zijn dan r.

4. De som van twee o.o. stochastische grootheden met Poisson verdelingen met parameter μ resp. ν bezit een Poisson verdeling met parameters $(\mu+\nu)$ (zie voorbeeld 2.6.5). Bewijs dit rechtstreeks met behulp van (2.5.15).
5. Geef een Poisson benadering van de binomiale verdeling voor het geval n groot is en p dichtbij 1 ligt.
6. Bewijs dat de som van de kansen in de negatief binomiale verdeling (2.6.1) gelijk is aan 1. Wat betekent dit resultaat?
7. Bepaal de kansverdeling van de som van twee o.o. stochastische grootheden die beide een homogene verdeling op $(0,1)$ bezitten.
8. Indien X een binomiale verdeling met parameters n en p bezit, dan geldt

$$P(X \leqq k) = \frac{n!}{k!(n-k-1)!} \int_p^1 y^k(1-y)^{n-k-1}\, dy, \quad k = 0, 1, \ldots, n.$$

Indien X een Poisson verdeling met parameter μ bezit dan geldt

$$P(X \leqq k) = \frac{1}{k!} \int_\mu^\infty e^{-y}\, y^k\, dy, \qquad k = 0, 1, \ldots .$$

Bewijs deze resultaten door herhaalde partiële integratie. Het tweede resultaat volgt echter ook rechtstreeks uit de relatie tussen de Poisson- en de gamma verdeling (zie voorbeeld 2.6.9).
9. De som van twee o.o. stochastische grootheden met gamma verdelingen met parameters k en λ resp. m en λ bezit ook voor niet gehele positieve k en m een gamma verdeling met parameters $(k+m)$ en λ. Bewijs dit met behulp van (2.5.17).
10. Indien X een dubbel-exponentiële verdeling met parameter λ bezit, dan heeft $|X|$ een exponentiële verdeling met parameter λ. Bewijs dit.
11. Indien X en Y o.o. zijn en beide een standaard Cauchy verdeling bezit-

ten, dan heeft hun gemiddelde $Z = \frac{1}{2}(X+Y)$ eveneens een standaard Cauchy verdeling. Bewijs dit door de dichtheid van $X + Y$ met behulp van contour integratie te berekenen.

12. Indien X en Y o.o. zijn en beide een standaard normale verdeling bezitten, dan bezit hun gemiddelde $Z = \frac{1}{2}(X+Y)$ een $N(0,\frac{1}{2})$-verdeling. Bewijs dit rechtstreeks met behulp van (2.5.17).
13. Indien X en Y o.o. zijn en beide een standaard normale verdeling bezitten, dan heeft $Z = {}^X/Y$ een standaard Cauchy verdeling. Bewijs dit.

2.7. VERWACHTING EN MOMENTEN

Zij X een stochastische grootheid gedefiniëerd op een kansruimte $(\Omega,\mathcal{A},P)$.

Definitie 2.7.1.
Indien X P-integreerbaar is noemt men

$$(2.7.1) \qquad EX = \int_\Omega X(\omega)\, dP$$

de *verwachting* van X. Indien X P-sommeerbaar is spreekt men van een stochastische grootheid X met eindige verwachting.

Daar $P_X(B) = P(X^{-1}(B))$ voor $B \in \mathcal{B}^1$, volgt uit de overplantingsstelling (stelling 1.6.6)

$$(2.7.2) \qquad EX = \int_{R^1} x\, dP_X.$$

EX hangt dus alleen van de kansverdeling P_X van X af, d.w.z. stochastische grootheden met dezelfde kansverdeling hebben ook dezelfde verwachting. Men spreekt dan ook wel over de verwachting van een kansverdeling in plaats van de verwachting van een stochastische grootheid die deze kansverdeling bezit. Soms zullen wij EX door één symbool willen aangeven. Hiervoor zal dan in het algemeen de letter μ worden gebruikt. Daarbij zal uiteraard steeds worden aangegeven op welke stochastische grootheid of op welke kansverdeling deze verwachting betrekking heeft.

Zij $Y = (Y_1,\ldots,Y_k)$ een stochastische vector met kansverdeling P_Y en beschouw de stochastische grootheid $X = \phi(Y) = \phi(Y_1,\dots,Y_k)$, ϕ eindig en meetbaar. Dan geldt, wederom volgens de overplantingsstelling

$$(2.7.3) \qquad EX = E\phi(Y) = \int_\Omega \phi(Y(\omega))dP = \int_{R^k} \phi(y)\, dP_Y.$$

Wij kunnen de verwachting van $X = \phi(Y)$ dus op drie verschillende wijzen bepalen: als integraal over Ω m.b.t. P, als integraal over R^1 m.b.t. zijn kansverdeling P_X en als integraal over R^k m.b.t. de kansverdeling P_Y van Y. In de vaak voorkomende situatie waarin P_Y gegeven is, is (2.7.3) meestal de aangewezen methode; het voordeel boven (2.7.2) is dat wij ons de moeite van het bepalen van P_X besparen.

Indien X een kansdichtheid f_X ten opzichte van een σ-finiete maat μ op $(R^1,\mathcal{B}^1)$ bezit, kan (2.7.2) volgens de stelling van Radon-Nikodym ook geschreven worden als

$$(2.7.4) \qquad EX = \int_{R^1} x\, f_X(x)\, d\mu(x).$$

Als X een discrete verdeling op een eindige of aftelbare verzameling $D \subset R^1$ bezit, kan voor μ de telmaat op D worden gekozen waardoor (2.7.4) overgaat in

$$(2.7.5) \qquad EX = \sum_{x\in D} x\, P(X=x).$$

Voor continu verdeelde X met kansdichtheid $f_X = F_X'$ kan voor μ de Lebesgue maat worden gekozen waardoor (2.7.4) overgaat in

$$(2.7.6) \qquad EX = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

Op analoge wijze kan (2.7.3) worden geschreven als

$$(2.7.7) \qquad EX = E\,\phi(Y) = \int_{R^k} \phi(y)\, f_Y(y)\, d\nu(y)$$

wanneer Y een dichtheid f_Y ten opzichte van een σ-finiete maat ν op $(R^k,\mathcal{B}^k)$ bezit. Als Y discreet is op $D \subset R^k$ vinden wij

$$(2.7.8) \qquad EX = E\,\phi(Y) = \sum_{y \in D} \phi(y)\,P(Y=y),$$

terwijl voor continu verdeelde Y met $f_Y(y) = \dfrac{\partial^k F_Y(y)}{\partial y_1 \ldots \partial y_k}$ en continue ϕ geldt

$$(2.7.9) \qquad EX = E\,\phi(Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)\,f_Y(y)\,dy_1 \ldots dy_k.$$

De volgende eigenschappen van de verwachting volgen rechtstreeks uit die van de integraal (zie stelling 1.6.1).

Stelling 2.7.1.

a. Als $X(\omega) = I_A(\omega)$. $A \in A$, dan geldt $EX = P(A)$.

b. Als $P(X=c) = 1$, $c \in R^1$, dan geldt $EX = c$.

c. Voor stochastische grootheden X_1 en X_2 en reële a, b en c geldt $E(aX_1+bX_2+c) = a\,EX_1 + b\,EX_2 + c$ mits het rechterlid van deze gelijkheid gedefiniëerd is.

d. Als $P(X_1 \leq X_2) = 1$ dan geldt $EX_1 \leq EX_2$ mits beide verwachtingen gedefiniëerd zijn; met name volgt uit $P(X \leq c) = 1$, $c \in R^1$, dat $EX \leq c$.

e. Indien EX gedefiniëerd is, geldt $|EX| \leq E|X| = EX^+ + EX^-$.

Natuurlijk bezitten ook de andere stellingen over integralen een analogon voor verwachtingen. Wij noemen hier slechts de ongelijkheid van Jensen en een speciaal geval van de stelling van Fubini.

Stelling 2.7.2. (Jensen)

Zij X een stochastische grootheid met eindige verwachting en zij ϕ een reële meetbare convexe functie op R^1. Dan geldt $E\,\phi(X) \geq \phi(EX)$.

Stelling 2.7.3. (Fubini)

Als X_1 en X_2 o.o. stochastische grootheden zijn waarvoor hetzij $P(X_1X_2 \geq 0) = 1$ hetzij $E|X_1X_2| < \infty$, dan geldt $EX_1X_2 = EX_1EX_2$.

De verwachting van X kan worden opgevat als een "gemiddelde waarde" van de stochastische grootheid. Evenzo kunnen wij "gemiddelden" van zekere functies van X beschouwen.

Definitie 2.7.2.

Indien X^k voor zekere gehele $k > 0$ P-integreerbaar is, noemt men

$$EX^k = \int X^k(\omega)\ dP$$

het k^e *moment* van X; als bovendien $\mu = EX$ gedefinieerd en eindig is, noemt men

$$E(X-\mu)^k = \int (X(\omega)-\mu)^k\ dP$$

het k^e *centrale moment* van X. Voor iedere reële $r > 0$ heet

$$E|X|^r = \int |X(\omega)|^r\ dP$$

het *absolute moment* van de orde r van X.

In tegenstelling tot absolute momenten worden momenten en centrale momenten slechts voor gehele waarden van k gedefiniëerd. Deze beperking komt voort uit het feit dat wij slechts verwachtingen van reële functies wensen te beschouwen. Merk op dat voor iedere stochastische grootheid X alle absolute momenten en alle even momenten gedefiniëerd zijn; als X een eindige verwachting bezit, zijn ook alle even centrale momenten gedefinieerd. Dit houdt natuurlijk geenszins in dat deze momenten ook eindig zouden zijn.

Stelling 2.7.4.

Indien voor zekere reële $r_0 > 0$ het absolute moment van de orde r_0 van een stochastische grootheid X eindig is, dan is ook het absolute moment van de orde r van X eindig voor alle $0 < r < r_0$. Tevens zijn dan het k^e moment en het k^e centrale moment van X gedefinieerd en eindig voor alle positieve gehele $k \leq r_0$.

Bewijs:

Voor $0 < r < r_0$ geldt

$$0 \leq \int |X(\omega)|^r\ dP \leq \int \{|X(\omega)|^{r_0} + 1\}\ dP < \infty.$$

Voor gehele positieve $k \leq r_0$ volgt uit de hierboven aangetoonde P-sommeerbaarheid van $|X|^k$ en de meetbaarheid van X^k tevens de P-sommeerbaarheid van het positieve en van het negatieve deel van X^k en dus ook die van X^k

zelf. Het k^e moment is dus gedefiniëerd en eindig. Het bestaan van een positieve gehele $k \leqq r_0$ impliceert dat $r_0 \geqq 1$, zodat volgens het voorafgaande $\mu = EX$ gedefiniëerd en eindig is. Dus is ook

$$(X-\mu)^k = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mu^{k-j} X^j$$

P-sommeerbaar, zodat het k^e centrale moment gedefiniëerd en eindig is.

Behalve het eerste moment (de verwachting) waarvan in stellingen 2.7.1 - 2.7.3 reeds eigenschappen werden genoemd, zullen wij in deze paragraaf slechts het tweede centrale moment nader beschouwen. Dit moment wordt ook de variantie genoemd en aangegeven met het symbool σ^2; deze notatie als kwadraat is geoorloofd daar het tweede centrale moment, mits gedefinieerd, steeds niet negatief is.

Definitie 2.7.3.
De *variantie* van een stochastische grootheid X met eindige verwachting $\mu = EX$ wordt gegeven door

$$(2.7.10) \qquad \sigma^2(X) = E(X-\mu)^2 = \int (X(\omega)-\mu)^2 \, dP;$$

de *standaardafwijking* van X wordt gegeven door de niet negatieve wortel uit de variantie van X:

$$(2.7.11) \qquad \sigma(X) = [E(X-\mu)^2]^{\frac{1}{2}} \geqq 0.$$

Volgens stelling 2.7.4 is de variantie van X gedefinieerd en eindig indien $EX^2 < \infty$. Als X discreet verdeeld is op $D \subset R^1$ dan geldt volgens (2.7.8)

$$(2.7.12) \qquad \sigma^2(X) = \sum_{x \in D} (x-\mu)^2 \, P(X=x);$$

als X continu verdeeld is met $f_X = F'(x)$ dan geldt volgens (2.7.9)

$$(2.7.13) \qquad \sigma^2(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \, f_X(x) \, dx.$$

De standaardafwijking van X, als zijnde de wortel uit de verwachte kwadratische afwijking van X van zijn verwachting, heeft dezelfde "fysische dimensie" als de stochastische grootheid X zelf en wordt algemeen gehanteerd als een maat voor de "spreiding" van de kansverdeling van X.

De variantie bezit de volgende elementaire eigenschappen.

Stelling 2.7.5.

a. $\sigma^2(X) = 0$ dan en slechts dan indien $P(X=c) = 1$ voor zekere reële c.

b. $\sigma^2(X) = EX^2 - (EX)^2$ indien het linkerlid gedefiniëerd is.

c. Indien $\sigma^2(X)$ gedefiniëerd is geldt voor alle reële a en b

$$\sigma^2(a+bX) = b^2\,\sigma^2(X) \text{ en dus } \sigma(a+bX) = |b|\,\sigma(X).$$

Bewijs:

a. $\sigma^2(X) = \int (X(\omega)-\mu)^2 = 0$, $\mu = EX$, impliceert dat $(X(\omega)-\mu)^2 = 0$ P-bijna overal, d.w.z. $P(X=\mu) = 1$. Omgekeerd: als $P(X=c) = 1$ dan geldt $\mu = EX = c$ zodat $(X(\omega)-\mu)^2 = 0$ P-bijna overal en dus $\sigma^2(X) = 0$.

b. $\sigma^2(X) = E(X-\mu)^2 = E(X^2-2\mu X+\mu^2) = EX^2 - 2\mu\, EX + \mu^2 = EX^2 - (EX)^2$ waarin $\mu = EX$.

c. Voor $\mu = EX$ geldt $E(a+bX) = a + b\mu$ zodat $\sigma^2(a+bX) = E(a+bX-a-b\mu)^2 =$ $= Eb^2(X-\mu)^2 = b^2\,\sigma^2(X)$.

Uit c. en stelling 2.7.1 volgt dat indien X eindige verwachting $\mu = EX$ en eindige varaintie $\sigma^2 = \sigma^2(X)$ bezit, de stochastische grootheid

$$X^* = \frac{X-\mu}{\sigma}$$

verwachting 0 en variantie 1 bezit. Men noemt X^* de *gestandaardiseerde* van X. Uit c. blijkt tevens dat de standaardafwijking de eigenschappen bezit die wij van een maat voor "spreiding" mogen eisen: hij is invariant onder verschuiving van de kansverdeling van X en reageert lineair op schaalveranderingen.

Definitie 2.7.4.

Indien de stochastische grootheden X_1 en X_2 eindige verwachtingen μ_1 resp. μ_2 bezitten en $X_1\, X_2$ P-integreerbaar is, wordt de *covariantie* van X_1 en X_2

gegeven door

$$\mathrm{cov}(X_1,X_2) = E(X_1-\mu_1)(X_2-\mu_2).$$

De stochastische grootheden X_1 en X_2 heten *ongecorreleerd* indien $\mathrm{cov}(X_1,X_2) = 0$.

Merk op dat $\mathrm{cov}(X,X) = \sigma^2(X)$

Stelling 2.7.6.

a. $\mathrm{cov}(X_1,X_2) = EX_1X_2 - EX_1EX_2$ indien het linkerlid gedefiniëerd is.

b. Indien X_1 en X_2 eindige varianties bezitten is $\mathrm{cov}(X_1,X_2)$ gedefiniëerd en eindig.

c. Indien X_1 en X_2 o.o. zijn en een eindige covariantie bezitten dan zijn X_1 en X_2 ongecorreleerd.

d. Voor reële $a_1, a_2, \ldots, a_n$ en stochastische grootheden $X_1, X_2, \ldots, X_n$ met eindige varianties geldt

$$\sigma^2\left(\sum_{i=1}^n a_iX_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i\, a_j\, \mathrm{cov}(X_i,X_j) =$$

$$= \sum_{i=1}^n a_i^2\, \sigma^2(X_i) + 2 \sum_{i} \sum_{< j} \mathrm{cov}(X_i,X_j).$$

Indien $X_1, X_2, \ldots, X_n$ bovendien paarsgewijs ongecorreleerd zijn geldt

$$\sigma^2\left(\sum_{i=1}^n a_iX_i\right) = \sum_{i=1}^n a_i^2\, \sigma^2(X_i).$$

N.B. Met nadruk zij er op gewezen dat deel c van deze stelling niet omkeerbaar is: ongecorreleerde stochastische grootheden X_1 en X_2 zijn niet noodzakelijk o.o. (zie opgave 1).

Bewijs:

a. Daar $\mathrm{cov}(X_1,X_2)$ gedefiniëerd is zijn $\mu_1 = EX_1$ en $\mu_2 = EX_2$ beide eindig zodat

$$\mathrm{cov}(X_1,X_2) = E(X_1X_2-\mu_1X_2-\mu_2X_1+\mu_1\mu_2) =$$

$$= EX_1X_2 - \mu_1EX_2 - \mu_2EX_1 + \mu_1\mu_2 = EX_1X_2 - EX_1EX_2$$

aangezien het rechterlid gedefiniëerd is (stelling 2.7.1). Merk op dat hieruit volgt dat eindigheid van μ_1 en μ_2 en P-integreerbaarheid van $X_1\ X_2$ ook P-integreerbaarheid van $(X_1-\mu_1)\ (X_2-\mu_2)$ inhoudt, waarmee definitie 2.7.4 gerechtvaardigd is.

b. Daar $\sigma^2(X_1) < \infty$ en $\sigma^2(X_2) < \infty$, zijn EX_1, EX_1^2, EX_2 en EX_2^2 alle eindig. Aangezien $|2\ X_1\ X_2| \leqq X_1^2 + X_2^2$ is dus ook $X_1\ X_2$ P-sommerbaar zodat het gestelde uit a. volgt.

c. Uit stelling 2.7.3 volgt dat in dit geval $EX_1 \cdot X_2 = EX_1 \cdot EX_2$ zodat volgens a. $\mathrm{cov}(X_1,X_2) = 0$.

d. Uit de eindigheid der varianties volgt de eindigheid der verwachtingen en covarianties zodat $E \sum a_i\ X_i = \sum a_i\ \mu_i$ waarbij $\mu_i = EX_i$, en

$$\sigma^2(\sum a_i X_i) = E\{\sum a_i(X_i-\mu_i)\}^2 = E \sum_i \sum_j a_i\ a_j\ (X_i-\mu_i)\ (X_j-\mu_j) =$$

$$= \sum_i \sum_j a_i\ a_j\ \mathrm{cov}(X_i,X_j).$$

Hieruit volgen de overige beweringen onder d.

Uit stelling 2.7.1 c. en stelling 2.7.6 c. en d. volgt onmiddellijk

Stelling 2.7.7.

Indien $X_1, X_2, \ldots, X_n$ stochastische grootheden zijn met dezelfde eindige variantie $\sigma^2 = \sigma^2(X_i)$ en verwachting $\mu = EX_i$, dan geldt voor hun som $S = \sum_{i=1}^{n} X_i$ en hun gemiddelde $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$ES = n\mu, \quad E\overline{X} = \mu.$$

Als $X_1, X_2, \ldots, X_n$ bovendien paarsgewijs ongecorreleerd zijn (dus a fortiori wanneer zij paarsgewijs of onderling onafhankelijk zijn) dan geldt

$$\sigma^2(S) = n\sigma^2, \quad \sigma^2(\overline{X}) = \frac{\sigma^2}{n}.$$

Definitie 2.7.5.
Zij $X = (X_1, X_2, \dots, X_n)$ een stochastische vector waarvan iedere component eindige variantie bezit. De *covariantie-matrix* Σ_X van deze stochastische vector is de $(n \times n)$-matrix met elementen $\sigma_{i,j} = \operatorname{cov}(X_i, X_j)$, $i, j = 1, 2, \dots, n$.

N.B. Het diagonaal element $\sigma_{i,i}$ stelt de variantie van X_i voor en niet de standaardafwijking van X_i zoals de notatie zou doen vermoeden.

Stelling 2.7.8.
Zij $X = (X_1, X_2, \dots, X_n)$ een stochastische vector waarvan iedere component eindige variantie bezit. Dan is de covariantie-matrix Σ_X symmetrisch en positief semi-definiet. Σ_X is dan en slechts dan singulier indien er een niet triviale lineaire combinatie van $X_1, X_2, \dots, X_n$ bestaat die een gedegenereerde verdeling bezit, d.w.z. indien voor zekere reële $a_1, a_2, \dots, a_n$ en c waarvoor $\sum a_i^2 \neq 0$, geldt

$$P\left(\sum_{i=1}^{n} a_i X_i = c\right) = 1.$$

Bewijs:
Σ_X is symmetrisch daar $\operatorname{cov}(X_i, X_j) = \operatorname{cov}(X_j, X_i)$. Aangezien een variantie steeds niet-negatief is volgt uit stelling 2.7.6. d

$$0 \leq \sigma^2\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{i,j}\, a_i\, a_j, \quad \sigma_{i,j} = \operatorname{cov}(X_i, X_j),$$

voor alle reële $a_1, a_2, \dots, a_n$, zodat Σ_X positief semi-definiet is. Σ_X is dus dan en slechts dan singulier, indien er reële getallen $a_1, a_2, \dots, a_n$ met $\sum a_i^2 \neq 0$ bestaan, waarvoor de bovenstaande kwadratische vorm gelijk is aan nul, d.w.z. waarvoor $\sigma^2(\sum a_i X_i) = 0$. Toepassing van stelling 2.7.5. a voltooit het bewijs.

Past men het bovenstaande toe voor $n = 2$, dan vindt men dat voor twee stochastische grootheden X_1 en X_2 met eindige varianties de matrix

$$(2.7.14) \qquad \begin{pmatrix} \sigma^2(X_1) & \operatorname{cov}(X_1, X_2) \\ \operatorname{cov}(X_1, X_2) & \sigma^2(X_2) \end{pmatrix}$$

positief semi-definiet is. De determinant van deze matrix is dus niet negatief, d.w.z.

(2.7.15) $\qquad |\mathrm{cov}(X_1,X_2)| \leqq \sigma(X_1)\ \sigma(X_2);$

gelijkheid geldt hierbij dan en slechts dan indien de matrix (2.7.14) singulier is, d.w.z. indien voor zekere a_1, a_2 en c met $a_1^2 + a_2^2 \neq 0$ geldt

(2.7.16) $\qquad P(a_1X_1+a_2X_2=c) = 1.$

De ongelijkheid (2.7.15) - die in feite niet anders is dan de ongelijkheid van Schwartz - is aanleiding tot de volgende definitie.

Definitie 2.7.6.
De *correlatie-coefficiënt* van twee stochastische grootheden X_1 en X_2 met eindige positieve varianties wordt gegeven door

$$\rho(X_1,X_2) = \frac{\mathrm{cov}(X_1,X_2)}{\sigma(X_1)\sigma(X_2)}$$

Stelling 2.7.9.
Voor twee stochastische grootheden X_1 en X_2 met eindige positieve varianties geldt steeds

$$-1 \leqq \rho(X_1,X_2) \leqq 1.$$

$\rho(X_1,X_2) = 0$ dan en slechts dan indien X_1 en X_2 ongecorreleeerd zijn;
$\rho(X_1,X_2) = 1$ (resp. -1) dan en slechts dan als (2.7.16) geldt voor zekere a_1, a_2 en c waarvoor $a_1.a_2 < 0$ (resp. > 0).

Bewijs:
De beide eerste beweringen volgen uit (2.7.15), het feit dat beide varianties positief en eindig zijn en uit de definitie van het begrip ongecorreleerd. Daar de varianties positief zijn, zijn de verdelingen van X_1 en X_2 beide niet gedegenereerd. Dus geldt $|\rho(X_1,X_2)| = 1$ dan en slechts dan indien (2.7.16) vervuld is voor zekere a_1 en a_2 waarvoor $a_1.a_2 \neq 0$. In dit

geval geldt

$$a_1 \cdot a_2 \; cov(X_1,X_2) = cov(a_1X_1,a_2X_2) = cov(a_1X_1,c-a_1X_1) = - a_1^2 \; \sigma^2(X_1) < 0,$$

zodat $\rho(X_1,X_2) = 1$ (resp. $- 1$) en dus $cov(X_1,X_2) > 0$ (resp. < 0) impliceert dat $a_1 \cdot a_2 < 0$ (resp. > 0).

Een correlatiecoëfficiënt 1 of − 1 geeft dus aan dat er tussen beide stochastische grootheden met kans 1 en stijgend respectievelijk dalend lineair verband bestaat. In het algemeen wordt $|\rho(X_1,X_2)|$ opgevat als een aanduiding van de mate van lineaire afhankelijkheid tussen X_1 en X_2.

Voorbeelden

Deze paragraaf wordt besloten met het berekenen van verwachtingen, varianties en covarianties van de in §2.6 behandelde kansverdelingen.

2.7.1. *Gedegenereerde verdeling*

Indien $P(X=x_0) = 1$ dan geldt $EX = x_0$ en $\sigma^2(X) = 0$ volgens stellingen 2.7.1. en 2.7.5.

2.7.2. *Alternatieve verdeling in 0 en 1*

Als $P(X=1) = p$ en $P(X=0) = 1 - p$ dan geldt $EX = p$, $EX^2 = p$ en dus $\sigma^2(X) = EX^2 - (EX)^2 = p - p^2 = p(1-p)$.

2.7.3. *Binomiale verdeling*

Zij X binomiaal verdeeld met parameters n en p, d.w.z.

$$P(X=x) = \binom{n}{x} p^x \, (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Zoals in §2.6 werd opgemerkt is de verdeling van X dezelfde als die van

$$\sum_{i=1}^{n} X_i$$

waarbij $X_1, X_2, \ldots, X_n$ o.o. zijn en alle de in het vorige voorbeeld besproken alternatieve verdeling bezitten. Dus geldt volgens stelling 2.7.7

(2.7.17) $\qquad EX = np, \quad \sigma^2(X) = np(1-p).$

2.7.4. *Hypergeometrische verdeling*

Veronderstel dat X de hypergeometrische verdeling

$$P(X=x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

bezit. Een voorbeeld van een dergelijke stochastische grootheid X is het aantal rode knikkers in een aselecte steekproef zonder teruglegging van n knikkers uit een vaas waarin zich r rode en (N-r) witte knikkers bevinden. Definiëren wij

$$X_i = \begin{cases} 1 & \text{als bij de } i^e \text{ trekking een rode knikker wordt getrokken} \\ 0 & \text{anders,} \end{cases}$$

dan geldt $X = \sum_{i=1}^{n} X_i$ voor deze X. Bij het beantwoorden van opgave 2 van §2.6 wordt gevonden (ga dit na)

$$P(X_i=1) = \frac{r}{N},\quad P(X_i=0) = 1 - \frac{r}{N},\qquad i = 1, 2, \ldots, n$$

$$P(X_iX_j=1) = \frac{r(r-1)}{N(N-1)},\quad P(X_iX_j=0) = 1 - \frac{r(r-1)}{N(N-1)},\quad i \neq j,$$

zodat $X_1, X_2, \ldots, X_n$ afhankelijk zijn doch alle dezelfde alternatieve verdeling bezitten. Uit het bovenstaande volgt

$$EX_i = \frac{r}{N},\quad \sigma^2(X_i) = \frac{r(N-r)}{N^2},$$

$$\mathrm{cov}(X_i,X_j) = EX_iX_j - EX_iEX_j = \frac{r(r-1)}{N(N-1)} - \left(\frac{r}{N}\right)^2 = -\frac{r(N-r)}{N^2(N-1)}.$$

Uit het eerste deel van stelling 2.7.7 en uit stelling 2.7.6 d volgt nu

$$(2.7.18)\qquad EX = \frac{nr}{N},\quad \sigma^2(X) = \frac{n(N-n)\, r(N-r)}{N^2(N-1)},$$

aangezien

$$\sigma^2(X) = \sigma^2\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \sigma^2(X_i) + 2\sum_{i}\sum_{<\, j} \mathrm{cov}(X_i,X_j) =$$

$$= n\,\frac{r(N-r)}{N^2} - n(n-1)\,\frac{r(N-r)}{N^2(N-1)} = \frac{n(N-n)\;r(N-r)}{N^2(N-1)}\;.$$

2.7.5. *Poisson verdeling*

Indien X een Poisson verdeling met parameter μ bezit, d.w.z.

$$P(X=x) = e^{-\mu}\,\frac{\mu^x}{x!}, \qquad\qquad x = 0,\ 1,\ \ldots,$$

dan geldt

$$(2.7.19) \qquad EX = \mu, \quad \sigma^2(X) = \mu.$$

Immers

$$EX = e^{-\mu} \sum_{x=0}^{\infty} x\,\frac{\mu^x}{x!} = e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} = e^{-\mu}\,\mu\,e^{\mu} = \mu,$$

$$EX(X-1) = e^{-\mu} \sum_{x=0}^{\infty} x(x-1)\,\frac{\mu^x}{x!} = e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^x}{(x-2)!} = e^{-\mu}\,\mu^2\,e^{\mu} = \mu^2,$$

$$\sigma^2(X) = EX^2 - (EX)^2 = EX(X-1) + EX - (EX)^2 = \mu^2 + \mu - \mu^2 = \mu.$$

2.7.6. *Negatief binomiale verdeling*

Als X een negatief binomiale verdeling met parameters k en p bezit, d.w.z.

$$P(X=x) = \binom{x-1}{k-1}\,p^k\,(1-p)^{x-k}, \qquad\qquad x = k,\ k+1,\ \ldots,$$

dan geldt

$$(2.7.20) \qquad EX = \frac{k}{p}\,, \quad \sigma^2(X) = \frac{k(1-p)}{p^2}\,.$$

Immers, voor k = 1 geldt

$$EX = \sum_{x=1}^{\infty} x\,p(1-p)^{x-1} = -\,p\,\frac{d}{dp} \sum_{x=0}^{\infty} (1-p)^x = -\,p\,\frac{d}{dp}\,(p^{-1}) = \frac{1}{p},$$

$$EX(X+1) = \sum_{x=1}^{\infty} x(x+1)\,p(1-p)^{x-1} = p\,\frac{d^2}{dp^2} \sum_{x=-1}^{\infty} (1-p)^{x+1} = \frac{2}{p^2},$$

$$\sigma^2(X) = EX^2 - (EX)^2 = EX(X+1) - EX - (EX)^2 = \frac{1-p}{p^2}.$$

Aangezien de som van k o.o. stochastische grootheden met negatief binomiale verdelingen met parameters 1 en p een negatief binomiale verdeling met parameters k en p bezit, volgt (2.7.20) uit stelling 2.7.7.

2.7.7. *Multinomiale verdeling*

Zij $X = (X_1, X_2, \dots, X_k)$ een stochastische vector met een multinomiale verdeling met parameters n en $p = (p_1, p_2, \dots, p_k)$, d.w.z.

$$P(X=x) = \frac{n!}{x_1!\, x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

voor iedere vector $x = (x_1, x_2, \dots, x_k)$ van niet-negatieve gehele getallen met $\sum x_j = n$. Daar de marginale verdeling van X_j een binomiale verdeling met parameters n en p_j is, geldt

$$(2.7.21) \qquad EX_j = np_j,\ \sigma^2(X_j) = np_j(1-p_j), \qquad j = 1, 2, \dots, k.$$

Voorts is voor $j \neq m$

$$EX_jX_m = \sum x_j x_m \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} =$$

$$= n(n-1)\, p_j\, p_m \sum \frac{(n-2)!}{y_1! \dots y_k!} p_1^{y_1} \dots p_k^{y_k} = n(n-1)\, p_j\, p_m,$$

waarbij de eerste sommatie loopt over alle $x = (x_1, \dots, x_k)$ met niet-negatieve gehele componenten met $\sum x_i = n$ en de tweede sommatie over alle $y = (y_1, \dots, y_k)$ met niet-negatieve gehele componenten met $\sum y_i = n - 2$. De laatste som is derhalve gelijk aan 1. Uit het bovenstaande volgt

$$(2.7.22) \qquad \operatorname{cov}(X_j, X_m) = -n\, p_j\, p_m, \qquad j \neq m.$$

2.7.8. *Homogene verdeling op (0,1)*

De kansdichtheid van een stochastische grootheid X die homogeen verdeeld is op (0,1) is

$$f_X(x) = \begin{cases} 1 & \text{voor } 0 < x < 1 \\ 0 & \text{anders.} \end{cases}$$

Daar $\int x\ f_X(x)\ dx = \frac{1}{2}$ en $\int x^2\ f_X(x)\ dx = \frac{1}{3}$, geldt

(2.7.23) $\quad EX = \frac{1}{2}, \quad \sigma^2(X) = \frac{1}{12}.$

2.7.9. *Gamma verdeling*

De kansdichtheid van een stochastische grootheid X die een gamma verdeling met parameters k en λ bezit wordt gegeven door

$$f_X(x) = \frac{\lambda^k}{\Gamma(k)}\ e^{-\lambda x}\ x^{k-1}, \qquad x > 0.$$

Daar

$$\int_{-\infty}^{\infty} x\ f_X(x)\ dx = \frac{\lambda^k}{\Gamma(k)} \int_0^{\infty} e^{-\lambda x}\ x^k\ dx = \frac{1}{\lambda\Gamma(k)} \int_0^{\infty} e^{-x}\ x^k\ dx$$

$$= \frac{\Gamma(k+1)}{\lambda\Gamma(k)} = \frac{k}{\lambda},$$

$$\int_{-\infty}^{\infty} x^2\ f_X(x)\ dx = \frac{\lambda^k}{\Gamma(k)} \int_0^{\infty} e^{-\lambda x}\ x^{k+1}\ dx = \frac{\Gamma(k+2)}{\lambda^2\Gamma(k)} = \frac{k(k+1)}{\lambda^2},$$

geldt

(2.7.24) $\quad EX = \frac{k}{\lambda}, \quad \sigma^2(X) = \frac{k}{\lambda^2}.$

2.7.10. *Dubbel-exponentiële verdeling*

X bezit een dubbel-exponentiële verdeling met parameter λ als

$$f_X(x) = \frac{1}{2}\ \lambda\ e^{-\lambda|x|}.$$

Dan geldt

(2.7.25) $\quad EX = 0, \quad \sigma^2(X) = \frac{2}{\lambda^2}.$

Het eerste volgt uit de integreerbaarheid van x $f_X(x)$ en de symmetrie van f_X om nul, terwijl

$$\sigma^2(X) = EX^2 = \frac{1}{2}\lambda \int_{-\infty}^{\infty} x^2 e^{-\lambda|x|} dx = \lambda \int_{0}^{\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

2.7.11. *Cauchy verdeling*

X bezit een standaard Cauchy verdeling als

$$f_X(x) = \frac{1}{\pi(1+x^2)} .$$

Voor reële $r \geq 1$ is de functie $|x|^r (1+x^2)^{-1}$ niet sommeerbaar terwijl voor oneven $k \geq 1$ de functie $x^k (1+x^2)^{-1}$ zelfs niet integreerbaar is. De Cauchy verdeling bezit dan ook geen enkel eindig moment van een orde ≥ 1. De oneven momenten en alle centrale momenten zijn niet gedefiniëerd; de even momenten en alle andere absolute momenten van een orde $r \geq 1$ zijn alle gelijk aan $+\infty$.

2.7.12. *Normale verdeling*

X bezit een normale verdeling met parameters μ en σ^2 als

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

Zoals de notatie reeds doet vermoeden zijn verwachting en variantie van X juist gelijk aan de parameters μ en σ^2 van de verdeling:

$$(2.7.26) \qquad EX = \mu, \quad \sigma^2(X) = \sigma^2.$$

Het eerste volgt uit de integreerbaarheid van x $f_X(x)$ en de symmetrie van f_X om μ, terwijl partiële integratie levert

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

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = \sigma^2.$$

2.7.13. *Meerdimensionale normale verdeling*

De stochastische vector $X = (X_1, X_2, \ldots, X_k)$ bezit een k-dimensionale $N(\mu, \Sigma)$-verdeling als

$$f_X(x) = \frac{1}{(2\pi)^{k/2}\,(|\Sigma|)^{1/2}} \exp((x-\mu)'\,\Sigma^{-1}(x-\mu)).$$

In §2.6 werd aangetoond dat X_i in dit geval een ééndimensionale $N(\mu_i, \sigma_{ii})$-verdeling bezit, zodat

$$(2.7.27) \qquad EX_i = \mu_i, \quad \sigma^2(X_i) = \sigma_{ii}, \qquad i = 1, 2, \ldots, k.$$

Voorts werd bewezen dat $\sum b_j X_j$ een ééndimensionale $N(\sum b_j\mu_j, \sum_i \sum_j \sigma_{ij} b_i b_j)$-verdeling bezit. Hieruit volgt dat $\sigma^2(X_i+X_j) = \sigma_{ii} + \sigma_{jj} + 2\sigma_{ij} = \sigma^2(X_i) + \sigma^2(X_j) + 2\sigma_{ij}$ voor $i \neq j$, aangezien de matrix Σ symmetrisch is. Vergelijking met stelling 2.7.6 d leert nu dat

$$(2.7.28) \qquad \operatorname{cov}(X_i, X_j) = \sigma_{ij} (= \sigma_{ji}), \qquad i \neq j.$$

De parameters μ en Σ van de meerdimensionale normale verdeling zijn dus juist de vector van verwachting en de covariantie-matrix van de stochastische vector X. In dit licht bezien betekent de eis dat Σ symmetrisch en positief-definiet is (zie §2.6) slechts uitsluiting van het geval waarin een lineaire combinatie van de componenten van X een gedegenereerde verdeling bezit.

Tenslotte werd in §2.6 aangetoond dat $X_1, X_2, \ldots, X_k$ dan en slechts dan o.o. zijn indien Σ een diagonaal-matrix is. Voor stochastische grootheden die een simultane normale verdeling bezitten zijn de begrippen "(paarsgewijs) ongecorreleerd" een "onderling onafhankelijk" dus equivalent.

Opgaven

1. De stochastische grootheden X en Y bezitten de volgende simultane verdeling

$$P(X=0,Y=1) = P(X=0,Y=-1) = P(X=1,Y=0) = P(X=-1,Y=0) = \tfrac{1}{4}.$$

Toon aan dat X en Y ongecorreleerd doch niet o.o. zijn.

2. Indien $X_1, X_2, \ldots, X_n$ eindige varianties bezitten dan geldt voor reële $a_1, \ldots, a_n, b_1, \ldots, b_n$

$$\operatorname{cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i\, a_j \operatorname{cov}(X_i, X_j).$$

Bewijs dit.

3. Indien de stochastische grootheden X en Y dezelfde eindige variantie bezitten, dan zijn de stochastische grootheden X + Y en X - Y ongecorreleerd. Toon dit aan.

4. Bepaal verwachting en variantie van de binomiale en van de hypergeometrische verdeling rechtstreeks uit de betreffende kansverdelingen zonder gebruik te maken van splitsing in alternatief verdeelde stochastische grootheden.

5. In een flatgebouw met (k+1) woonlagen, van beneden af genummerd 0, 1, ..., k, vertrekt een lift met n personen van woonlaag 0 naar boven. Aangenomen wordt dat op de woonlagen 1, 2, ..., k geen personen wensen in te stappen, zodat alleen wordt gestopt om de n personen te laten uitstappen. Voorts wordt verondersteld dat deze n personen het nummer van de woonlaag waar zij uitstappen aselect en onafhankelijk van elkaar uit de getallen 1, 2, ..., k kiezen. Bereken verwachting en variantie van het aantal malen dat de lift stopt zonder de kansverdeling van dit aantal malen te bepalen.

6. X bezit een standaard normale verdeling. Bewijs dat voor positieve gehele k geldt

$$EX^k = \begin{cases} 0 & \text{als k oneven is,} \\ \dfrac{k!}{2^{k/2}.(k/2)!} & \text{als k even is.} \end{cases}$$

Leid hieruit een uitdrukking voor $E(Y-\mu)^k$ af, wanneer Y een $N(\mu,\sigma^2)$-verdeling bezit.

2.8. DE ZWAKKE WET VAN DE GROTE AANTALLEN

Stelling 2.8.1.

Zij Y een stochastische grootheid waarvoor $P(Y \geq 0) = 1$ en zij ϕ een niet-negatieve en niet-dalende functie gedefiniëerd op $[0,\infty)$. Dan geldt voor iedere $c > 0$ waarvoor $\phi(c) > 0$,

$$(2.8.1) \qquad P(Y \geq c) \leq \frac{E\phi(Y)}{\phi(c)}.$$

Bewijs:

$$E\phi(Y) = \int \phi(y)\, dP_Y = \int_{[0,c)} \phi(y)\, dP_Y + \int_{[c,\infty)} \phi(y)\, dP_Y \geq$$

$$\geq \phi(c) \int_{[c,\infty)} dP_Y = \phi(c)\, P(Y \geq c),$$

waaruit de stelling volgt daar $\phi(c) > 0$.

Ongelijkheden van het type (2.8.1) worden *ongelijkheden van Chebyshev* genoemd. Zij zijn alleen dan niet triviaal indien $E\phi(Y) < \phi(c)$; zeker dient dus $E\phi(Y)$ eindig te worden verondersteld. Meestal hanteert men hierbij functies ϕ waarvoor $\lim_{y\to\infty} \phi(y) = \infty$. Men gebruikt dan de ongelijkheid om te concluderen dat voor alle niet-negatieve stochastische grootheden Y waarvoor $E\phi(Y)$ een gegeven eindige waarde bezit, de kans $P(Y \geq c)$ uniform naar boven begrensd is door een functie van c die naar 0 nadert voor $c \to \infty$. Voor praktische toepassing zijn ongelijkheden van Chebyshev in het algemeen niet voldoende scherp; als bewijstechniek zijn zij echter van grote waarde. De volgende speciale gevallen zijn hierbij van bijzonder belang.

Door voor een willekeurige stochastische grootheid X, $Y = |X|$ en $\phi(y) = y^r$, $r > 0$, te kiezen ontstaat de zogenaamde *ongelijkheid van Markov:* Voor iedere stochastische grootheid X waarvan het absolute moment van de orde r (> 0) eindig is, geldt voor iedere $c > 0$

$$(2.8.2) \qquad P(|X| \geq c) \leq \frac{E|X|^r}{c^r}.$$

Dit resultaat wordt vooral voor $r = 1$ en $r = 2$ veel gebruikt.

Door voor een stochastische grootheid X met eindige verwachting μ en variantie σ^2, $Y = |X - \mu|$ en $\phi(y) = y^2$ te kiezen ontstaat de *ongelijkheid van Bienaymé - Chebyshev:*
Voor iedere stochastische grootheid X waarvoor $\mu = EX$ en $\sigma^2 = \sigma^2(X)$ eindig zijn, geldt voor iedere $c > 0$

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2} \; ; \tag{2.8.3}$$

als $\sigma^2 > 0$ kunnen wij hiervoor ook schrijven

$$P\left(\frac{|X-\mu|}{\sigma} \geq c\right) \leq \frac{1}{c^2} \; : \tag{2.8.4}$$

de kans dat de gestandaardiseerde van een stochastische grootheid in absolute waarde $\geq c$ is, is nooit groter dan c^{-2}.

Definitie 2.8.1.
Een rij stochastische grootheden $X_1, X_2, \ldots$ *convergeert in waarschijnlijkheid* naar een reëel getal a (notatie: $X_n \overset{P}{\to} a$ voor $n \to \infty$) indien voor iedere $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n-a| > \varepsilon) = 0.$$

Een rij stochastische grootheden $X_1, X_2, \ldots$ convergeert in waarschijnlijkheid naar een stochastische grootheid X (notatie: $X_n \overset{P}{\to} X$ voor $n \to \infty$) indien $X_n - X \overset{P}{\to} 0$ voor $n \to \infty$, d.w.z. indien voor iedere $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n-X| > \varepsilon) = 0.$$

Merk op dat in het laatste geval de stochastische grootheden $X, X_1, X_2, \ldots$ op dezelfde kansruimte gedefiniëerd dienen te zijn; voor convergentie in waarschijnlijkheid naar een reëel getal behoeft de rij $X_1, X_2, \ldots$ niet op dezelfde kansruimte gedefiniëerd te zijn: het betreft in dit geval een eigenschap van de rij kansverdelingen $P_{X_1}, P_{X_2}, \ldots$.

Stelling 2.8.2.

Indien X_1, X_2, ... o.o. stochastische grootheden zijn die alle dezelfde verwachting μ en dezelfde eindige variantie σ^2 bezitten, dan convergeert de rij der gemiddelden

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad n = 1, 2, \ldots,$$

voor $n \to \infty$ in waarschijnlijkheid naar μ.

Bewijs:

Daar σ^2 gedefiniëerd is, is μ eindig en volgens stelling 2.7.7 geldt

$$E\overline{X}_n = \mu, \quad \sigma^2(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Toepassing van de ongelijkheid van Bienaymé-Chebyshev op $\overline{X}_n$ levert voor iedere $\varepsilon > 0$

$$P(|\overline{X}_n - \mu| > \varepsilon) \leqq \frac{\sigma^2}{\varepsilon^2 n},$$

zodat voor iedere $\varepsilon > 0$

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

Deze stelling staat bekend als de *zwakke wet van de grote aantallen*. De stelling is overigens een zwakke versie van deze zwakke wet aangezien de eis van gelijke eindige varianties onnodig restrictief is (zie opgave 1). Indien men veronderstelt dat X_1, X_2, ... o.o. en identiek verdeeld zijn met eindige verwachting $\mu = EX_i$, behoeft zelfs geen enkele andere voorwaarde te worden opgelegd om de zwakke wet te bewijzen. Een interessant speciaal geval van stelling 2.8.2 is het volgende.

Stelling 2.8.3. (Bernoulli)

Beschouw een oneindige rij van o.o. uitvoeringen van een experiment waarbij bij ieder (deel) experiment de kans op het optreden van een eventualiteit S gelijk is aan p. Zij $X^{(n)}$ het aantal malen dat S optreedt bij de eerste n experimenten. Dan convergeert de rij der frequentie-quotiënten van S,

$$\frac{X^{(n)}}{n}, \qquad n = 1, 2, \ldots,$$

voor $n \to \infty$ in waarschijnlijkheid naar p.

Bewijs:
Zij

$$X_i = \begin{cases} 1 & \text{als S bij het } i^e \text{ experiment optreedt} \\ 0 & \text{anders,} \end{cases}$$

dan zijn $X_1, X_2, \ldots$ o.o. en identiek verdeeld met $EX_i = p$ en $\sigma^2(X_i) = p(1-p) < \infty$. Daar

$$\frac{X^{(n)}}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

volgt de stelling rechtstreeks uit stelling 2.8.2.

Stelling 2.8.3 vertoont een duidelijke analogie met de empirische wet van de grote aantallen die het intuïtieve uitgangspunt bij de opbouw van het kansbegrip vormt. De empirische wet houdt in dat in lange reeksen herhalingen van een experiment - steeds zorgvuldig onder dezelfde omstandigheden uitgevoerd - het frequentie-quotiënt van een gebeurtenis S zich vaak stabiliseert in de nabijheid van een getal p dat wij intuïtief als de kans op de gebeurtenis S opvatten. Daar een op deze intuïtie gebaseerde definitie van het begrip kans als limiet van het frequentie-quotiënt op ernstige moeilijkheden betreffende het bestaan van deze limiet stuit, hebben wij gekozen voor een axiomatische opbouw van het kansbegrip waarbij de empirische wet slechts als intuïtieve motivering van de keuze der axioma's een rol speelt. Nu blijkt dat uit de ogenschijnlijk zwakke axioma's (kans is een genormeerde maat) een stelling volgt die redelijk goed overeenkomt met de empirische wet, mag men hopen dat het door de axioma's gedefiniëerde wiskundige model van het kansbegrip tot praktisch bruikbare resultaten zal leiden. Men hoede zich echter voor de vergissing te menen dat de zwakke wet een bewijs van de empirische wet is. Empirisch gevonden resultaten zijn langs wiskundige weg niet te bewijzen.

Opgave

1. Laten $X_1, X_2, \ldots$ o.o. stochastische grootheden met eindige varianties voorstellen die alle dezelfde verwachting μ bezitten. Indien

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2(X_i) = 0,$$

dan convergeert de rij der gemiddelden

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \qquad n = 1, 2, \ldots,$$

in waarschijnlijkheid naar μ.

2.9. ZWAKKE CONVERGENTIE VAN KANSVERDELINGEN

Wij beschouwen in deze paragraaf kansverdelingen op $(R^k,\mathcal{B}^k)$ met $1 \leqq k < \infty$. Wij gebruiken daarbij de volgende notatie:
Voor willekeurige $x, y \in R^k$ en $A \subset R^k$ is

$$d_k(x,y) = \{\sum_{i=1}^{k} (x_i-y_i)^2\}^{\frac{1}{2}}$$

de afstand van x tot y, is

$$d_k(x,A) = \inf_{y\in A} d_k(x,y)$$

de afstand van x tot A, en stelt

$$\partial A = \{x\colon d_k(x,A) = d_k(x,A^c) = 0\}$$

de rand van A voor. Merk op dat ∂A gesloten is, zodat $\partial A \in \mathcal{B}^k$ voor alle $A \subset R^k$. Verder schrijven wij $C(R^k)$ voor de verzameling van alle reële functies op R^k, die begrensd en continu zijn.

Definitie 2.9.1

Als P een kansverdeling op $(R^k,\mathcal{B}^k)$ is en $A \in \mathcal{B}^k$ de eigenschap heeft dat $P(\partial A) = 0$, dan noemen wij A een *P-continuïteitsverzameling*.

Definitie 2.9.2

Een rij kansverdelingen $P_1, P_2, \ldots$ op $(R^k,\mathcal{B}^k)$ *convergeert zwak* naar een kansverdeling P op $(R^k,\mathcal{B}^k)$ (notatie: $P_n \xrightarrow{zw} P$ voor $n \to \infty$) als

$$\lim_{n\to\infty} \int_{R^k} f(x)\, dP_n(x) = \int_{R^k} f(x)\, dP(x)$$

voor iedere functie $f \in C(R^k)$. Een rij verdelingsfuncties $F_1, F_2, \ldots$ op R^k convergeert zwak naar een verdelingsfunctie F op R^k (notatie: $F_n \xrightarrow{zw} F$ voor $n \to \infty$) als de met de F_n corresponderende rij kansverdelingen zwak convergeert naar de kansverdeling, die bij F behoort.

Stelling 2.9.1

De volgende vijf beweringen betreffende kansverdelingen $P_1, P_2, \ldots, P$ op $(R^k, \mathcal{B}^k)$ met verdelingsfunties $F_1, F_2, \ldots, F$ zijn equivalent:

(i) $P_n \xrightarrow{zw} P$ voor $n \to \infty$;

(ii) $\limsup_{n\to\infty} P_n(A) \leq P(A)$ voor iedere gesloten verzameling $A \subset R^k$;

(iii) $\liminf_{n\to\infty} P_n(A) \geq P(A)$ voor iedere open verzameling $A \subset R^k$;

(iv) $\lim_{n\to\infty} P_n(A) = P(A)$ voor iedere P-continuïteitsverzameling $A \subset R^k$;

(v) $\lim_{n\to\infty} F_n(x) = F(x)$ voor ieder punt $x \in R^k$ waar de functie F continu is.

Bewijs:

Indien wij bij een gesloten verzameling $A \subset R^k$ functies $f_1, f_2, \ldots \in C(R^k)$ definiëren door

$$f_m(x) = \begin{cases} 1 - md_k(x,A) & \text{als } d_k(x,A) \leq \frac{1}{m}, \\ 0 & \text{als } d_k(x,A) \geq \frac{1}{m}, \end{cases}$$

dan geldt

$$0 \leq I_A(x) \leq f_m(x) \leq 1,\ x \in R^k,\ m = 1, 2, \ldots,$$

en

$$\lim_{m\to\infty} f_m(x) = I_A(x),\ x \in R^k.$$

Hieruit volgt

$$P_n(A) \leq \int_{R^k} f_m(x)\, dP_n(x),\ m, n = 1, 2, \ldots,$$

en dus impliceert (i) op grond van de definitie van zwakke convergentie

$$\limsup_{n\to\infty} P_n(A) \leq \int_{R^k} f_m(x)\, dP(x),\ m = 1, 2, \ldots.$$

Laten wij nu $m \to \infty$, dan gaat het rechter lid van deze ongelijkheid wegens de gedomineerde convergentie van f_m naar I_A over in $P(A)$. Daar wij voor A iedere gesloten deelverzameling van R^k kunnen nemen, is hiermee aangetoond dat (ii) uit (i) volgt. De equivalentie van (ii) en (iii) is gemakkelijk te bewijzen door over te gaan op complementen. Schrijven wij A° voor het inwendige, en $\overline{A}$ voor de afsluiting van een verzameling $A \in \mathcal{B}^k$, dan volgt uit (ii) en (iii)

$$P(A^\circ) \leq \liminf P_n(A^\circ) \leq \liminf P_n(A) \leq$$

$$\leq \limsup P_n(A) \leq \limsup P_n(\overline{A}) \leq P(\overline{A}).$$

Daar voor iedere P-continuïteitsverzameling A per definitie $P(A^\circ) = P(\overline{A})$ volgt (iv) dus uit (ii) en (iii). Verder volgt (v) uit (iv), omdat continuïteit van de verdelingsfunctie F in een punt x betekent, dat de cel $(-\infty,x]$ een P-continuïteitsverzameling is (ga dit na!). Tenslotte moeten wij nog bewijzen dat (i) uit (v) volgt. Gemakshalve geven wij het bewijs alleen voor het ééndimensionale geval (k=1). Laat $f \in C(R^1)$, zodat $M = \sup_x |f(x)| < \infty$, en zij $\varepsilon > 0$. Wij kiezen dan een interval $(a,b]$ zo groot dat $P((a,b]) \geq 1 - \varepsilon$ en bovendien zo, dat de verdelingsfunctie F continu is in ieder van de punten a en b. Dit laatste is mogelijk omdat een verdelingsfunctie op R^1 hoogstens aftelbaar veel discontinuïteiten heeft (zie 2.4, opgave 1). Nu geldt

$$(2.9.1) \qquad \left| \int_{(a,b]^c} f(x)\, dP(x) \right| \leq M.P((a,b]^c) \leq M.\varepsilon,$$

en

$$(2.9.2) \qquad \left| \int_{(a,b]^c} f(x)\, dP_n(x) \right| \leq M.P_n((a,b]^c),\ n = 1, 2, \ldots .$$

Verder volgt uit de continuïteit van f, dat f uniform continu is op het gesloten interval $[a,b]$. Er is dus een getal $\delta > 0$ met de eigenschap, $|f(x) - f(y)| \leq \varepsilon$ voor alle $x, y \in [a,b]$ met $|x - y| \leq \delta$. We splitsen het interval $(a,b]$ nu in een aantal deelintervallen $I_j = (x_{j-1},x_j]$, $j = 1, 2, \ldots, m$ met $a = x_0 < x_1 < \ldots < x_m = b$, waarbij we er voor zorgen dat $x_j - x_{j-1} \leq \delta$ voor alle j en dat F continu is in ieder van de punten x_j. Nu volgt

$$(2.9.3) \qquad |\int_{(a,b]} f(x)\, dP(x) - \sum_{j=1}^{m} f(x_j)\, P(I_j)| \leq \varepsilon$$

en evenzo,

$$(2.9.4) \qquad |\int_{(a,b]} f(x)\, dP_n(x) - \sum_{j=1}^{m} f(x_j)\, P_n(I_j)| \leq \varepsilon,\ n = 1, 2, \ldots,$$

daar

$$|f(x) - f(x_j)| \leq \varepsilon,\ x \in I_j,\ j = 1, 2, \ldots, m.$$

Uit (2.9.1), (2.9.2), (2.9.3) en (2.9.4) concluderen wij

$$(2.9.5) \qquad |\int_{R^1} f(x)\, dP_n(x) - \int_{R^1} f(x)\, dP(x)| \leq$$

$$\leq 2\varepsilon + M \sum_{j=1}^{m} |P_n(I_j) - P(I_j)| + MP_n((a,b]^c) + M\varepsilon,$$

$$n = 1, 2, \ldots .$$

Daar F continu is in de punten x_j, volgt uit (v)

$$\lim_{n\to\infty} P_n(I_j) = \lim_{n\to\infty} (F_n(x_j)-F_n(x_{j-1})) = P(I_j),\ j = 1, 2, \ldots, m,$$

en ook

$$\lim_{n\to\infty} P_n((a,b]^c) = P((a,b]^c) \leq \varepsilon.$$

Derhalve impliceert (2.9.5)

$$\limsup_{n\to\infty} |\int_{R^1} f(x)\, dP_n(x) - \int_{R^1} f(x)\, dP(x)| \leq 2(M+1)\ \varepsilon,$$

en dus volgt (i) daar $\varepsilon > 0$ willekeurig klein gekozen kan worden. Het bewijs voor het meerdimensionale geval $(k>1)$ verloopt in principe net zo. Men kiest eerst een cel $(a,b]$ die zo groot is, dat $P((a,b]) \geqq 1 - \varepsilon$, en vervolgens splitst men deze cel met behulp van hypervlakken evenwijdig aan de zijvlakken van $(a,b]$ in een aantal deelcellen, die zo klein zijn, dat f op iedere deelcel minder dan ε variëert. Men dient echter $(a,b]$ en de splitsing daarvan in deelcellen zo te kiezen, dat F continu is in alle 2^k hoekpunten van $(a,b]$ en in alle 2^k hoekpunten van iedere deelcel. Alleen dan kan men immers uit (v) concluderen dat $P_n((a,b]) \to P((a,b])$ voor $n \to \infty$ en evenzo voor de deelcellen. Men kan aan deze eis voldoen door bij de constructie uitsluitend hypervlakken te gebruiken die P-nulverzamelingen zijn (zie opgave 1).

Uit stelling 2.9.1 kunnen wij concluderen dat de limiet van een zwak convergente rij kansverdelingen uniek is, daar volgens (v) zijn verdelingsfunctie vast ligt (zie opgave 1). Hieruit volgt de volgende stelling.

Stelling 2.9.2
Twee kansverdelingen P en Q op $(R^k, \mathcal{B}^k)$ met de eigenschap dat

$$\int_{R^k} f(x)\, dP(x) = \int_{R^k} f(x)\, dQ(x),\ f \in C(R^k),$$

zijn identiek.

Bewijs:
Stellen wij $P_n = P$, $n = 1, 2, \ldots$, dan volgt $P_n \xrightarrow{zw} P$ en $P_n \xrightarrow{zw} Q$ voor $n \to \infty$, en dus $P = Q$.

Op grond van stelling 2.9.1 beschikken wij nu over vijf karakteriseringen van het begrip zwakke convergentie. De karakterisering waarop wij onze definitie gebaseerd hebben is uit theoretisch oogpunt wellicht de belangrijkste. Voor de practijk zijn (iv) en (v) echter van groot belang, daar zij betekenen dat men bij zwakke convergentie van P_n naar P, voor grote waarden van n en een ruime klasse verzamelingen, $P(A)$ als benadering voor $P_n(A)$ kan beschouwen. Vooral in de statistiek maakt men hier vaak gebruik van. Men moet echter wel bedenken dat zwakke convergentie van P_n naar P

niet noodzakelijk voor alle $A \in \mathcal{B}^k$ impliceert dat $P_n(A) \to P(A)$ als $n \to \infty$. Als bijv. P_n voor $n = 1, 2, \ldots$ de discrete homogene verdeling op de verzameling $\{\frac{i}{n} : i = 0, 1, 2, \ldots, n\}$ is, zodat de verdelingsfunctie van P_n gegeven wordt door

$$F_n(x) = \begin{cases} 0 & \text{als } x < 0, \\ \frac{i}{n+1} & \text{als } \frac{i-1}{n} \leq x < \frac{i}{n}, \ i = 1, 2, \ldots, n, \\ 1 & \text{als } 1 \leq x, \end{cases}$$

dan convergeert P_n voor $n \to \infty$ zwak naar de homogene verdeling op $[0,1]$. Voor de verzameling A van alle irrationale getallen geldt echter $P_n(A) = 0$, $n = 1, 2, \ldots$, terwijl de limietverdeling aan deze verzameling de maat 1 toekent. Verder mag men niet uit het oog verliezen dat in stelling 2.9.1 verondersteld wordt dat F een verdelingsfunctie is. Zonder deze veronderstelling mag men uit (v) niet concluderen dat de bij de F_n behorende rij kansverdelingen zwak convergent is. Zo geldt (v) bijvoorbeeld voor iedere rij functies $F_1, F_2, \ldots$ en iedere functie F die nergens continu is. Een minder triviaal voorbeeld van een situatie, waarin (v) geldt zonder dat er sprake is van zwakke convergentie, wordt gegeven door

$$F_n(x) = \frac{1}{3}\,\Phi(x-n) + \frac{1}{3}\,\Phi(x) + \frac{1}{3}\,\Phi(x+n), \ x \in R^1, \ n = 1, 2, \ldots,$$

waarin Φ de verdelingsfunctie van de $N(0,1)$-verdeling is. Hier convergeert $F_n(x)$ voor $n \to \infty$ naar

$$F(x) = \frac{1}{3} + \frac{1}{3}\,\Phi(x), \ x \in R^1. \tag{2.9.6}$$

Deze functie F is geen verdelingsfunctie, hoewel hij wel een Lebesgue-Stieltjes maat op R^1 bepaalt. Deze maat kent aan de hele ruimte R^1 slechts de maat $\frac{1}{3}$ toe. Men zegt in dergelijke situaties wel dat er voor $n \to \infty$ massa naar het oneindige verdwijnt.

Definitie 2.9.3
Wij noemen een collectie kansverdelingen Π op $(R^k, \mathcal{B}^k)$ *relatief compact* als iedere oneindige rij kansverdelingen $P_n \in \Pi$, $n = 1, 2, \ldots$ een zwak convergente deelrij bevat.

Stelling 2.9.3
Een rij kansverdelingen $P_1, P_2, \ldots$ op $(R^k, \mathcal{B}^k)$ is dan en slechts dan zwak convergent als de collectie $\Pi = \{P_n : n = 1, 2, \ldots\}$ relatief compact is en alle zwak convergente deelrijen van de rij $P_1, P_2, \ldots$ dezelfde limiet hebben.

Bewijs:
Zij Π relatief compact, zodat de rij $P_1, P_2, \ldots$ tenminste één zwak convergente deelrij heeft, en stel dat alle zwak convergente deelrijen van de rij $P_1, P_2, \ldots$ een gemeenschappelijke limiet P hebben. Voor een willekeurige functie $f \in C(R^k)$ vormen de getallen

$$a_n = \int_{R^k} f(x)\, dP_n(x),\ n = 1, 2, \ldots$$

dan een begrensde getallenrij. Zij $a_{n_1}, a_{n_2}, \ldots$ een willekeurige deelrij van deze getallenrij. Uit de relatieve compactheid van Π volgt, dat de rij $P_{n_1}, P_{n_2}, \ldots$ een zwak convergente deelrij $P_{m_1}, P_{m_2}, \ldots$ bevat. Daar deze deelrij tevens een deelrij van de rij $P_1, P_2, \ldots$ is, is zijn limiet volgens onze veronderstelling P. Op grond van de definitie van zwakke convergentie geldt dus

$$\lim_{j\to\infty} a_{m_j} = a = \int_{R^k} f(x)\, dP(x).$$

Hiermee is aangetoond, dat de getallenrij $a_1, a_2, \ldots$ de eigenschap heeft, dat iedere deelrij een verdere deelrij bevat, die naar a convergeert. Maar dit betekent dat $a_n \to a$ voor $n \to \infty$ en dus volgt dat $P_n \xrightarrow{zw} P$ voor $n \to \infty$.
Als omgekeerd gegeven is dat de rij $P_1, P_2, \ldots$ zwak convergeert naar een limiet P, dan zijn uiteraard alle deelrijen van deze rij zwak convergent

met dezelfde limiet P. Als verder $P_{n_1}, P_{n_2}, \ldots$ een oneindige rij in Π is, dan zijn er twee mogelijkheden: óf er is een kansverdeling P_n die oneindig vaak voorkomt in de rij $P_{n_1}, P_{n_2}, \ldots$, zodat deze rij de zwak convergente deelrij $P_n, P_n, \ldots$ bevat, óf iedere P_n komt slechts eindig vaak voor in de rij $P_{n_1}, P_{n_2}, \ldots$. In het laatste geval bevat de rij $P_{n_1}, P_{n_2}, \ldots$ echter een deelrij die tevens een deelrij is van de rij $P_1, P_2, \ldots$ en als zodanig zwak convergeert. Iedere oneindige rij in Π bevat dus een zwak convergente deelrij, m.a.w. Π is relatief compact.

Definitie 2.9.4

Een reële functie F op R^k, die niet een verdelingsfunctie is, heet een *defectieve verdelingsfunctie* als $F(x)$ een niet-dalende rechtscontinue functie van ieder van de coördinaten x_i, $i = 1, 2, \ldots, k$ van $x \in R^k$ is met $0 \leq F(x) \leq 1$ voor alle $x \in R^k$, terwijl bovendien $\Delta_1 \Delta_2 \ldots \Delta_k F(a) \geqq 0$ voor iedere $a, b \in R^k$ met $a \leqq b$ (zie voorbeeld 1.5.4 voor de notatie).

Een defectieve verdelingsfunctie op R^k bepaalt een Lebesgue-Stieltjes maat P op $(R^k, \mathcal{B}^k)$ met $P(R^k) < 1$. Men noemt een maat met deze eigenschap wel een defectieve kansverdeling. De in (2.9.6) gegeven functie F is een voorbeeld van een defectieve verdelingsfunctie op R^1.

Stelling 2.9.4 *(Selectiestelling van Helly)*

Bij iedere rij verdelingsfuncties $F_1, F_2, \ldots$ op R^k is er een deelrij $F_{n_1}, F_{n_2}, \ldots$ en een, eventueel defectieve, verdelingsfunctie F op R^k, zodanig dat $F_{n_j}(x) \to F(x)$ voor $j \to \infty$ en alle $x \in R^k$ waar F continu is.

Bewijs:

Wij geven het bewijs hier alleen voor het ééndimensionale geval (k=1). Het bewijs voor het meerdimensionale geval (k>1) is analoog, zij het iets ingewikkelder. Zij $R_0^1 = \{r_1, r_2, \ldots\}$ een aftelbare overal dicht liggende verzameling in R^1. We bewijzen nu eerst, dat er voor iedere $i = 1, 2, \ldots$ een stijgende rij natuurlijke getallen n_{ij}, $j = 1, 2, \ldots$, is, zodanig dat $F_{n_{ij}}(r_i)$ voor $j \to \infty$ convergeert, terwijl bovendien voor iedere $i = 1, 2, \ldots$ de rij $n_{i+1,j}$, $j = 1, 2, \ldots$, een deelrij van de rij

n_{ij}, $j = 1, 2, \ldots$, is. We bewijzen dit door inductie naar i. De existentie van een stijgende rij natuurlijke getallen n_{1j}, $j = 1, 2, \ldots$, zodanig dat $F_{n_{1j}}(r_1)$ voor $j \to \infty$ convergeert, is duidelijk, daar $0 \leqq F_n(r_1) \leqq 1$ voor alle n. Evenzo kunnen wij bij een gegeven stijgende rij natuurlijke getallen n_{ij}, $j = 1, 2, \ldots$, een deelrij $n_{i+1,j}$, $j = 1, 2, \ldots$, vinden met de eigenschap dat $F_{n_{i+1,j}}(r_{i+1})$ voor $j \to \infty$ convergeert. Nu wij ons overtuigd hebben van de existentie van getallen n_{ij}, $i, j = 1, 2, \ldots$, met de boven opgesomde eigenschappen, beschouwen wij de diagonaalrij $n_j = n_{jj}$, $j = 1, 2, \ldots$. Deze rij is voor iedere $i = 1, 2, \ldots$, afgezien van zijn eerste $(i-1)$ elementen, een deelrij van de rij n_{ij}, $j = 1, 2, \ldots$, en dus convergeert $F_{n_j}(r_i)$ voor $j \to \infty$ en iedere $i = 1, 2, \ldots$ naar een limiet, die wij $L(r_i)$ noemen. Laat nu

$$F(x) = \inf_{\substack{r>x \\ r \in R_0^1}} L(r), \quad x \in R^1.$$

Men gaat gemakkelijk na dat de zo gedefinieerde functie F op R^1 niet-dalend en rechtscontinu is met $0 \leqq F(x) \leqq 1$ voor alle $x \in R^1$, zodat F een, eventueel defectieve, verdelingsfunctie is. Als F continu is in een punt x en $\varepsilon > 0$, dan zijn er getallen $r, s \in R_0^1$, zodanig dat $r < x < s$ en

$$F(x) - \varepsilon < L(r) \leqq L(s) < F(x) + \varepsilon.$$

Daar verder de beide uiterste leden van de ongelijkheid

$$F_{n_j}(r) \leqq F_{n_j}(x) \leqq F_{n_j}(s)$$

voor $j \to \infty$ naar $L(r)$ resp. $L(s)$ convergeren, volgt hieruit, dat $F_{n_j}(x)$ voor alle voldoend grote j minder dan ε van $F(x)$ afwijkt, zodat $F_{n_j}(x) \to F(x)$ voor $j \to \infty$.

Definitie 2.9.5

Een collectie kansverdelingen Π op $(R^k, \mathcal{B}^k)$ heet *uniform beperkt* (Engels: tight) als er voor iedere $\varepsilon > 0$ een compacte verzameling $K \subset R^k$ is met de eigenschap dat $P(K) \geqq 1 - \varepsilon$ voor alle $P \in \Pi$.

Stelling 2.9.5 *(Prohorov)*
Een collectie kansverdelingen Π op $(R^k, \mathcal{B}^k)$ is dan en slechts dan relatief compact als hij uniform beperkt is.

Bewijs:
Zij Π uniform beperkt en zij $P_1, P_2, \ldots$ een rij kansverdelingen in Π met verdelingsfuncties $F_1, F_2, \ldots$. Te bewijzen is nu dat deze rij een zwak convergente deelrij bevat. Toepassing van stelling 2.9.4 geeft ons een deelrij P_{n_j}, $j = 1, 2, \ldots$, en een eventueel defectieve verdelingsfunctie F op R^k, zodanig dat $F_{n_j}(x) \to F(x)$ voor $j \to \infty$ in alle continuïteitspunten x van F. Neem nu $\varepsilon > 0$ en een compacte $K \subset R^k$ zo, dat $P(K) \geqq 1 - \varepsilon$ voor alle $P \in \Pi$. Kies $a, b \in R^k$ met $a \leqq b$ zo, dat $K \subset (a,b]$ en zo, dat F continu is in a en b. Dan geldt

$$F(a) = \lim_{j\to\infty} F_{n_j}(a) \leqq \sup_{P\in\Pi} P((a,b]^c) \leqq \varepsilon,$$

$$F(b) = \lim_{j\to\infty} F_{n_j}(b) \geqq \inf_{P\in\Pi} P((a,b]) \geqq 1 - \varepsilon.$$

Daar $\varepsilon > 0$ willekeurig klein gekozen kan worden, volgt hieruit dat F een verdelingsfunctie is, zodat $F_{n_j} \xrightarrow{zw} F$ voor $j \to \infty$. Het omgekeerde bewijzen wij uit het ongerijmde. Stel dat Π relatief compact maar niet uniform beperkt is. Er is dan een $\varepsilon > 0$ zo, dat er bij iedere compacte $K \subset R^k$ een $P \in \Pi$ is met $P(K) < 1 - \varepsilon$. Geven wij voor $r > 0$ de k-dimensionale gesloten bol om de oorsprong met straal r aan met S_r, dan is er dus een rij $P_1, P_2, \ldots$ in Π met $P_n(S_n) < 1 - \varepsilon$ voor $n = 1, 2, \ldots$. Wegens de relatieve compactheid van Π bevat deze rij een deelrij P_{n_j}, $j = 1, 2, \ldots$, die zwak convergeert naar een limiet P. Kies nu $r > 0$ zo, dat S_r een P-continuïteitsverzameling is. Dan geldt

$$P(S_r) = \lim_{j\to\infty} P_{n_j}(S_r) \leqq \limsup_{n\to\infty} P_n(S_n) \leqq 1 - \varepsilon,$$

daar $S_r \subset S_n$ voor alle $n \geqq r$. Maar ook kunnen wij r zo groot kiezen, dat

$P(S_r) > 1 - \varepsilon$, waarmee wij een tegenspraak gekregen hebben.

In de waarschijnlijkheidsrekening en de mathematische statistiek past men de theorie over zwakke convergentie van verdelingen vaak toe op verdelingen van stochastische vectoren, waarbij het deze stochastische vectoren zijn, waarin men primair geïnteresseerd is. Men heeft daarom de volgende terminologie ingevoerd.

Definitie 2.9.6

Een rij k-dimensionale stochastische vectoren $X_1, X_2, \ldots$ *convergeert in verdeling* naar een k-dimensionale stochastische vector X (notatie: $X_n \xrightarrow{D} X$ voor $n \to \infty$) als de bijbehorende verdelingen P_{X_n} voor $n \to \infty$ zwak convergeren naar de verdeling P_X van X. Als daarbij de limietverdeling P_X gedegenereerd is in een punt $a \in R^k$, d.w.z. als $P(X=a) = 1$ voor een zekere $a \in R^k$, dan zeggen wij ook wel dat de rij $X_1, X_2, \ldots$ in verdeling naar a convergeert (notatie: $X_n \xrightarrow{D} a$ voor $n \to \infty$).

Uit de definities 2.9.2 en 2.9.6 volgt, dat convergentie in verdeling van X_n naar X equivalent is met convergentie van de verwachtingen $Ef(X_n)$ naar $Ef(X)$ voor alle $f \in C(R^k)$. Ook de andere, in stelling 2.9.1 opgesomde karakteriseringen van zwakke convergentie van verdelingen kan men met behulp van definitie 2.9.6 formuleren in termen van convergentie in verdeling van stochastische vectoren.
Naast convergentie in verdeling kennen wij ook convergentie in waarschijnlijkheid. Dit begrip hebben wij weliswaar in 2.8 alleen voor het ééndimensionale geval gedefinieerd, maar de definitie laat zich gemakkelijk generaliseren.

Definitie 2.9.7

Een rij k-dimensionale stochastische vectoren $X_1, X_2, \ldots$ *convergeert in waarschijnlijkheid* naar een punt $a \in R^k$ (notatie: $X_n \xrightarrow{P} a$ voor $n \to \infty$) als voor iedere $\varepsilon > 0$

$$\lim_{n\to\infty} P(d_k(X_n,a) > \varepsilon) = 0.$$

De rij X_1, X_2, ... convergeert in waarschijnlijkheid naar een k-dimensionale stochastische vector X als $d_k(X_n,X) \overset{P}{\to} 0$ voor $n \to \infty$, d.w.z. als voor iedere $\varepsilon > 0$

$$\lim_{n\to\infty} P(d_k(X_n,X) > \varepsilon) = 0.$$

Merk op dat een rij stochastische vectoren X_1, X_2, ... slechts dan in waarschijnlijkheid naar een stochastische vector X kan convergeren, als X_1, X_2, ..., X alle op één kansruimte gedefinieerd zijn, terwijl dit laatste niet nodig is voor convergentie in verdeling. Hieruit volgt al dat convergentie in verdeling in het algemeen niet samen hoeft te gaan met convergentie in waarschijnlijkheid. Zelfs als X_1, X_2, ..., X wel alle op één kansruimte zijn gedefinieerd, dan nog impliceert convergentie in verdeling van X_n naar X niet dat ook $X_n \overset{P}{\to} X$ voor $n \to \infty$. Als bijvoorbeeld $P(X=1) = P(X=-1) = \frac{1}{2}$ en $P(X_n=(-1)^n X) = 1$ voor $n = 1, 2, \ldots$, dan convergeert X_n voor $n \to \infty$ in verdeling naar X, maar $P(|X_n-X|>\varepsilon)$ heeft geen limiet voor $n \to \infty$ als $0 < \varepsilon < 2$, zodat X_n niet in waarschijnlijkheid naar X convergeert.

Stelling 2.9.6
Uit convergentie in waarschijnlijkheid van een rij stochastische vectoren volgt convergentie in verdeling naar dezelfde limiet. Omgekeerd impliceert convergentie in verdeling naar een constante vector a, dat de rij ook in waarschijnlijkheid naar a convergeert.

Bewijs:
Als $X_n \overset{P}{\to} X$ voor $n \to \infty$ en $A \in \mathcal{B}^k$, dan geldt voor iedere $\varepsilon > 0$ en $n = 1, 2, \ldots$,

$$P(X_n\in A) = P(X_n\in A,\ d_k(X_n,X)>\varepsilon) + P(X_n\in A,\ d_k(X_n,X)\leqq\varepsilon) \leqq$$

$$\leqq P(d_k(X_n,X)>\varepsilon) + P(d_k(X,A)\leqq\varepsilon),$$

zodat

$$\limsup_{n\to\infty} P(X_n\in A) \leqq P(d_k(X,A)\leqq\varepsilon).$$

Daar deze ongelijkheid voor iedere $\varepsilon > 0$ geldt, volgt voor elke gesloten verzameling $A \subset R^k$,

$$\limsup_{n\to\infty} P(X_n \in A) \leqq P(X \in A),$$

en dus (stelling 2.9.1) geldt $X_n \overset{\mathcal{D}}{\to} X$ voor $n \to \infty$.
Het voorgaande geldt in het bijzonder als de verdeling van X de in een punt $a \in R^k$ gedegenereerde verdeling P_a is, d.w.z. als gegeven is dat $X_n \overset{P}{\to} a$ voor $n \to \infty$. De conclusie is dan dat $X_n \overset{\mathcal{D}}{\to} a$ voor $n \to \infty$. Als omgekeerd dit laatste gegeven is, dan volgt voor iedere $\varepsilon > 0$,

$$\limsup_{n\to\infty} P(d_k(X_n,a)>\varepsilon) \leqq \limsup_{n\to\infty} P(d_k(X_n,a)\geqq\varepsilon) =$$

$$= \limsup_{n\to\infty} P_{X_n}(\{x\colon d_k(x,a)\geqq\varepsilon\}) \leqq P_a(\{x\colon d_k(x,a)\geqq\varepsilon\}) = 0,$$

daar de verzamelingen $\{x\colon d_k(x,a) \geqq \varepsilon\}$ gesloten zijn. Dus geldt $X_n \overset{P}{\to} a$ voor $n \to \infty$.

Stelling 2.9.7
Als twee rijen k-dimensionale stochastische vectoren $X_1, X_2, \ldots$ en $Y_1, Y_2, \ldots$ de eigenschap hebben dat $X_n \overset{\mathcal{D}}{\to} X$ voor $n \to \infty$, terwijl $d_k(X_n,Y_n) \overset{P}{\to} 0$ voor $n \to \infty$, dan geldt ook $Y_n \overset{\mathcal{D}}{\to} X$ voor $n \to \infty$.

Merk op, dat het gegeven impliceert dat voor iedere $n = 1, 2, \ldots$ de twee stochastische vectoren X_n en Y_n op dezelfde kansruimte gedefinieerd zijn.

Bewijs:
Voor iedere $A \in \mathcal{B}^k$, $\varepsilon > 0$, $n = 1, 2, \ldots$, geldt

$$P(Y_n\in A) = P(Y_n\in A,\ d_k(X_n,Y_n)>\varepsilon) + P(Y_n\in A,\ d_k(X_n,Y_n)\leqq\varepsilon) \leqq$$

$$\leqq P(d_k(X_n,Y_n)>\varepsilon) + P(d_k(X_n,A)\leqq\varepsilon),$$

zodat

$$\limsup_{n\to\infty} P(Y_n\in A) \leqq \limsup_{n\to\infty} P(d_k(X_n,A)\leqq\varepsilon) \leqq P(d_k(X,A)\leqq\varepsilon),$$

daar de verzameling $\{x: d_k(x,A) \leq \varepsilon\}$ gesloten is. Omdat dit voor iedere $\varepsilon > 0$ geldt, volgt hieruit voor alle gesloten verzamelingen $A \subset R^k$,

$$\limsup_{n\to\infty} P(Y_n \in A) \leq P(X \in A),$$

en dus $Y_n \xrightarrow{D} X$ voor $n \to \infty$.

Uit de definities valt gemakkelijk af te leiden, dat convergentie in verdeling van een rij k-dimensionale stochastische vectoren X_1, X_2, ... naar een stochastische vector X impliceert, dat voor iedere continue functie g op R^k met waarden in R^m, de rij $g(X_1)$, $g(X_2)$, ... in verdeling convergeert naar $g(X)$. De volgende stelling zegt dat dit ook nog waar is voor Borel functies g op R^k met waarden in R^m die niet overal continu zijn, mits de verzameling D_g van alle punten $x \in R^k$, waar g niet continu is, een P_X-nulverzameling is. Wil deze voorwaarde zinvol zijn, dan moet uiteraard $D_g \in \mathcal{B}^k$. Dit is echter altijd het geval, zelfs als de functie g geen Borel functie is. Dit is als volgt in te zien. Zij $A_{\varepsilon,\delta}$ voor een gegeven functie g en voor $\varepsilon > 0$, $\delta > 0$ de verzameling van alle punten $x \in R^k$ met de eigenschap dat er punten $y, z \in R^k$ zijn waarvoor $d_k(x,y) < \delta$, $d_k(x,z) < \delta$ en $d_m(g(y),g(z)) \geq \varepsilon$. Deze verzameling $A_{\varepsilon,\delta}$ is dan open, zodat $A_{\varepsilon,\delta} \in \mathcal{B}^k$, en tevens geldt

$$D_g = \bigcup_{\varepsilon} \bigcap_{\delta} A_{\varepsilon,\delta},$$

waar de vereniging resp. doorsnede genomen wordt over alle rationale $\varepsilon > 0$ resp. $\delta > 0$. Derhalve is D_g een Borel verzameling in R^k.

Stelling 2.9.8

Als een rij k-dimensionale stochastische vectoren X_1, X_2, ... in verdeling convergeert naar een stochastische vector X en g een Borel functie op R^k is met waarden in R^m, waarvoor $P(X \in D_g) = 0$, dan convergeert de rij $g(X_1)$, $g(X_2)$, ... in verdeling naar $g(X)$.

Bewijs:

Als A een gesloten verzameling in R^m is, dan geldt voor de afsluiting van $g^{-1}(A)$,

$$\overline{g^{-1}(A)} \subset g^{-1}(A) \cup D_g,$$

en dus

$$P(X \in \overline{g^{-1}(A)}) = P(g(X) \in A).$$

Derhalve volgt uit het gegeven

$$\limsup_{n\to\infty} P(g(X_n) \in A) = \limsup_{n\to\infty} P(X_n \in g^{-1}(A)) \leqq$$

$$\leqq \limsup_{n\to\infty} P(X_n \in \overline{g^{-1}(A)}) \leqq P(X \in \overline{g^{-1}(A)}) = P(g(X) \in A),$$

en dus $g(X_n) \overset{\mathcal{D}}{\to} g(X)$ voor $n \to \infty$.

Stelling 2.9.9

Als een rij k-dimensionale stochastische vectoren $X_1, X_2, \ldots$ in verdeling convergeert naar een stochastische vector X, en een rij m-dimensionale stochastische vectoren $Y_1, Y_2, \ldots$ in waarschijnlijkheid naar een constante $a \in R^m$ convergeert, terwijl voor iedere $n = 1, 2, \ldots$, de twee stochastische vectoren X_n en Y_n op één kansruimte zijn gedefinieerd, dan convergeert de rij (k+m)-dimensionale stochastische vectoren $(X_1,Y_1), (X_2,Y_2), \ldots$ in verdeling naar (X,a).

Bewijs:

Uit het gegeven volgt met behulp van stelling 2.9.8 dat $(X_n,a) \overset{\mathcal{D}}{\to} (X,a)$ voor $n \to \infty$. Verder geldt

$$d_{k+m}((X_n,Y_n), (X_n,a)) = d_m(Y_n,a) \overset{P}{\to} 0, \ n \to \infty,$$

en dus volgt het gestelde uit stelling 2.9.7.

Opgaven

1. Zij P een kansverdeling op $(R^k, \mathcal{B}^k)$ met verdelingsfunctie F. Bewijs:
   a) Onder alle hypervlakken H in R^k, die loodrecht staan op een gegeven richting, zijn er hoogstens aftelbaar veel met $P(H) > 0$;
   b) Als $\mathcal{S} \subset \mathcal{B}^k$ een collectie verzamelingen is met de eigenschap dat $\partial A \cap \partial B = \emptyset$ voor alle $A, B \in \mathcal{S}$ met $A \neq B$, dan bevat $\mathcal{S}$ hoogstens aftelbaar veel verzamelingen die niet P-continuïteitsverzamelingen zijn;
   c) De verzameling van alle punten waar F continu is ligt overal dicht in R^k;
   d) Als F continu is in alle hoekpunten van een cel $(a,b]$, dan is deze cel een P-continuïteitsverzameling. Laat door middel van een tegenvoorbeeld zien dat het omgekeerde niet algemeen geldt.
2. Bewijs stelling 2.9.2 rechtstreeks, zonder stelling 2.9.1 te gebruiken.
3. Bewijs dat zwakke convergentie van een rij k-dimensionale verdelingsfuncties $F_1, F_2, \ldots$ naar een continue verdelingsfunctie F impliceert dat $F_n(x) \to F(x)$ voor $n \to \infty$ uniform in x.
4. Bewijs stelling 2.9.4 voor $k > 1$.
5. Zij P_x voor $x \in R^k$ de gedegenereerde kansverdeling op $(R^k, \mathcal{B}^k)$ gegeven door $P_x(A) = I_A(x)$, $A \in \mathcal{B}^k$. Bewijs dat $P_{x_n} \xrightarrow{zw} P$ voor $n \to \infty$ dan en dan alleen als de rij $x_1, x_2, \ldots$ in R^k convergeert naar een limiet x en $P = P_x$. Onderzoek de betekenis van relatieve compactheid en uniforme beperktheid voor gedegenereerde verdelingen en formuleer de stellingen 2.9.3, 2.9.4 en 2.9.5 voor zulke verdelingen als stellingen van rijen en verzamelingen in R^k.
6. De Lévy afstand $\lambda(F,G)$ tussen twee verdelingsfuncties F en G op R^1 wordt gedefinieerd als het infimum van alle reële h met de eigenschap dat voor alle $x \in R^1$

$$F(x-h) - h \leq G(x) \leq F(x+h) + h.$$

   Bewijs:
   a) $\lambda(F,G)$ is op een factor $\sqrt{2}$ na de grootste afstand tussen de grafieken van F en G, gemeten langs lijnen met richtingscoëfficiënt -1;

b) λ is een metriek;

c) De verzameling van alle verdelingsfuncties op R^1 met de metriek λ is een volledige metrische ruimte;

d) $\lambda(F_n,F) \to 0$ voor $n \to \infty$ dan en dan alleen als $F_n \xrightarrow{zw} F$ voor $n \to \infty$.

7. Als voor iedere $n = 1, 2, \ldots$ twee o.o. stochastische vectoren X_n en Y_n gegeven zijn, zodanig dat $X_n \overset{\mathcal{D}}{\to} X$ en $Y_n \overset{\mathcal{D}}{\to} Y$ voor $n \to \infty$, dan zijn er twee o.o. stochastische vectoren X' en Y' zo, dat $(X_n,Y_n) \overset{\mathcal{D}}{\to} (X',Y')$ voor $n \to \infty$. Bewijs dit.

2.10 KARAKTERISTIEKE FUNCTIES

In het voorgaande (stelling 2.9.2) hebben wij gezien , dat men een kansverdeling P op $(R^k,\mathcal{B}^k)$ geheel kan beschrijven door middel van de integralen van alle begrensde, reële, continue functies op R^k. In deze paragraaf zal blijken dat P in feite al wordt vastgelegd door alle integralen van de vorm

$$\int_{R^k} \cos(\sum_1^k t_j x_j)\, dP(x), \quad \int_{R^k} \sin(\sum_1^k t_j x_j)\, dP(x), \quad t_1, \ldots, t_k \in R^1.$$

Ter vereenvoudiging van de notatie gebruiken wij complexe integralen, d.w.z. integralen, waarbij de integrand een functie met complexe waarden is. Dit gebruik is echter niet meer dan een afkorting: Als $f = u + iv$, waarin u en v reële meetbare functies zijn, dan schrijven wij

$$\int f dP = \int u dP + i \int v dP.$$

Bovendien zullen wij voortdurend vectornotatie gebruiken. Daarbij beschouwen wij elementen van R^k, en ook stochastische vectoren, als kolomvectoren. Wij zullen in de paragraaf derhalve integralen van de vorm

$$\int_{R^k} e^{it'x}\, dP(x) = \int_{R^k} \cos(\sum_1^k t_j x_j)\, dP(x) + i \int_{R^k} \sin(\sum_1^k t_j x_j)\, dP(x)$$

met $t = (t_1,\ldots,t_k)' \in R^k$, of, in termen van een stochastische vector $X = (X_1,\ldots,X_k)'$,

$$Ee^{it'X} = E\cos(\sum_1^k t_jX_j) + i\,E\sin(\sum_1^k t_jX_j)$$

beschouwen.

Definitie 2.10.1

De *karakteristieke functie* van een kansverdeling P op $(R^k,\mathcal{B}^k)$ is een functie ϕ op R^k, gegeven door

$$\phi(t) = \int_{R^k} e^{it'x}\, dP(x),\ t \in R^k.$$

De karakteristieke functie ϕ_X van een k-dimensionale stochastische vector X is de karakteristieke functie van de verdeling P_X van X, zodat

$$\phi_X(t) = \int_{R^k} e^{it'x}\, dP_X(x) = Ee^{it'X},\ t \in R^k.$$

Uit de definitie volgt onmiddellijk, dat iedere karakteristieke functie ϕ op R^k continu is met

$$\phi(0) = 1,$$

$$|\phi(t)| \leq 1, \quad t \in R^k,$$

$$\phi(-t) = \overline{\phi(t)},\ t \in R^k,$$

Als $X = (X_1,\ldots,X_k)'$ een k-dimensionale stochastische vector is, a een m-dimensionale vector is en B een $(m\times k)$-matrix is, dan is a + BX een m-dimensionale stochastische vector met karakteristieke functie

$$\phi_{a+BX}(t) = Ee^{it'a+it'BX} = e^{it'a}\,\phi_X(B't),\ t \in R^m. \tag{2.10.1}$$

Uit (2.10.1) met m = 1, a = 0 volgt in het bijzonder

$$\phi_{t'X}(u) = \phi_X(tu),\ t \in R^k,\ u \in R^1. \tag{2.10.2}$$

De karakteristieke functie ϕ_X van de stochastische vector X bepaalt dus de karakteristieke functie $\phi_{t'X}$ van iedere lineaire combinatie t'X van de componenten van X, en omgekeerd wordt ϕ_X vastgelegd door de karakteristieke functies $\phi_{t'X}$ van alle lineaire combinaties $t'X$, $t \in R^k$.

Stelling 2.10.1

Als $X = (X_1,\ldots,X_k)'$ een k-dimensionale stochastische vector is met o.o. componenten $X_1, \ldots, X_k$, dan geldt

$$(2.10.3) \qquad \phi_X(t) = \prod_{j=1}^{k} \phi_{X_j}(t_j), \qquad t = (t_1,\ldots,t_k)' \in R^k;$$

$$(2.10.4) \qquad \phi_{t'X}(u) = \prod_{j=1}^{k} \phi_{X_j}(t_j u), \quad t = (t_1,\ldots,t_k)' \in R^k,\ u \in R^1;$$

en dus in het bijzonder

$$(2.10.5) \qquad \phi_{\sum_1^k X_j}(u) = \prod_{j=1}^{k} \phi_{X_j}(u), \quad u \in R^1.$$

Bewijs:

(2.10.3) volgt uit stelling 2.5.3 en stelling 2.7.3:

$$\phi_X(t) = E \prod_{j=1}^{k} e^{it_j X_j} = \prod_{j=1}^{k} Ee^{it_j X_j} = \prod_{j=1}^{k} \phi_{X_j}(t_j),$$

$$t = (t_1,\ldots,t_k)' \in R^k,$$

en de beide andere beweringen volgen hieruit door substitutie in (2.10.2).

Ter illustratie van het voorgaande berekenen wij de karakteristieke functies van een aantal normale verdelingen. Voor de karakteristieke functie ϕ van de $N(0,1)$-verdeling geldt per definitie

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx-\frac{1}{2}x^2} dx = \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} dx, \ t \in R^1.$$

Nu kan men met behulp van contour integratie gemakkelijk inzien dat

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} dx = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx, \ t \in R^1,$$

en dus wordt de karakteristieke functie van de $N(0,1)$-verdeling gegeven door

$$(2.10.6) \qquad \phi(t) = e^{-\frac{1}{2}t^2}, \quad t \in R^1.$$

Als een stochastische grootheid X een standaard normale verdeling heeft, dan heeft $\mu + \sigma X$ met μ en σ reëel een $N(\mu,\sigma^2)$-verdeling. Op grond van (2.10.1) en (2.10.6) kunnen wij dus besluiten dat de karakteristieke functie van de $N(\mu,\sigma^2)$-verdeling gegeven wordt door

$$(2.10.7) \qquad \phi(t) = e^{it\mu-\frac{1}{2}\sigma^2t^2}, \quad t \in R^1.$$

Als tenslotte een stochastische vector $X = (X_1,\ldots,X_k)'$ een $N(\mu,\Sigma)$-verdeling heeft, waarbij nu $\mu \in R^k$ en waarbij Σ een symmetrische positief definiete $(k\times k)$-matrix is, dan heeft iedere lineaire combinatie $t'X$ van de componenten van X een $N(t'\mu,\ t'\Sigma t)$-verdeling (zie 2.6.13). Uit (2.10.2) volgt derhalve dat de karakteristieke functie van de $N(\mu,\Sigma)$-verdeling gegeven wordt door

$$(2.10.8) \qquad \phi(t) = e^{it'\mu-\frac{1}{2}t'\Sigma t}, \quad t \in R^k.$$

In de aanhef van deze paragraaf zinspeelden wij reeds op het feit dat een kansverdeling door zijn karakteristieke functie geheel vastgelegd wordt. Om dit te kunnen bewijzen hebben wij de volgende gelijkheid nodig, die men wel de *gelijkheid van Parseval* voor karakteristieke functies noemt.

Stelling 2.10.2
Als P_1 en P_2 twee kansverdelingen op $(R^k,\mathcal{B}^k)$ zijn met karakteristieke functies ϕ_1 en ϕ_2, dan geldt

$$(2.10.8) \qquad \int_{R^k} \phi_1(y)\ dP_2(y) = \int_{R^k} \phi_2(x)\ dP_1(x).$$

Bewijs:
Bij substitutie van de definitie voor ϕ_1 en ϕ_2 blijkt (2.10.8) equivalent te zijn met de gelijkheid

$$\int_{R^k} \{ \int_{R^k} e^{iy'x}\, dP_1(x)\}\, dP_2(y) = \int_{R^k} \{ \int_{R^k} e^{ix'y}\, dP_2(y)\}\, dP_1(x),$$

en deze volgt uit de stelling van Fubini.

Als X en Y twee k-dimensionale stochastische vectoren zijn met verdelingen $P_X = P_1$ en $P_Y = P_2$, dan kunnen wij (2.10.8) schrijven als

$$(2.10.9) \qquad E\phi_X(Y) = E\phi_Y(X).$$

Vervangen wij hierin X door X-t met $t \in R^k$, dan vinden wij met behulp van (2.10.1)

$$Ee^{-it'Y}\, \phi_X(Y) = Ee^{-iY't}\, \phi_X(Y) = E\phi_{X-t}(Y) = E\phi_Y(X-t),\ t \in R^k.$$

Uit stelling 2.10.2 volgt derhalve de ogenschijnlijk sterkere bewering dat

$$(2.10.10) \qquad \int_{R^k} e^{-it'y}\, \phi_1(y)\, dP_2(y) = \int_{R^k} \phi_2(x-t)\, dP_1(x),\ t \in R^k.$$

Stelling 2.10.3 *(Inversiestelling)*
Als ϕ de karakteristieke functie is van een kansverdeling P op $(R^k, \mathcal{B}^k)$, dan geldt

$$(2.10.11) \qquad P(A) = \lim_{\sigma\to\infty} \frac{1}{(2\pi)^k} \int_A \{\int_{R^k} e^{-it'y}\, \phi(y)\, e^{-\frac{1}{2\sigma^2} y'y}\, dy\}\, d\lambda^k(t)$$

voor iedere P-continuïteitsverzameling A. Als bovendien

$$\int_{R^k} |\phi(t)| dt < \infty,$$

dan is P een continue verdeling met dichtheid

$$(2.10.12) \qquad f(x) = \frac{1}{(2\pi)^k} \int_{R^k} e^{-ix'y}\, \phi(y)\, dy,\ x \in R^k.$$

Merk op dat deze stelling impliceert dat een kansverdeling P op $(R^k, \mathcal{B}^k)$ door zijn karakteristieke functie geheel vastgelegd wordt.

Bewijs:

Wij bewijzen 2.10.11 hier alleen voor het ééndimensionale geval (k=1). Het bewijs voor het meerdimensionale geval is analoog. Substitutie in (2.10.10) van P voor P_1 en van de $N(0,\sigma^2)$-verdeling voor P_2 geeft, na vermenigvuldigen met $\sigma/\sqrt{2\pi}$,

$$(2.10.13) \qquad \frac{1}{2\pi}\int_{R^1} e^{-ity}\,\phi(y)\,e^{-\frac{y^2}{2\sigma^2}}\,dy = \frac{\sigma}{\sqrt{2\pi}}\int_{R^1} e^{-\frac{1}{2}\sigma^2(x-t)^2}\,dP(x),\ t\in R^1.$$

Nu blijkt het rechterlid van deze gelijkheid bij nadere inspectie juist gelijk te zijn aan de dichtheid van de convolutie van P met een $N(0,1/\sigma^2)$-verdeling. Anders gezegd, als X en Y twee o.o. stochastische grootheden zijn, X met verdeling P en Y met een $N(0,1/\sigma^2)$-verdeling, dan heeft $T = X + Y$ een continue verdeling met dichtheid $f_T(t)$ gegeven door het rechterlid van (2.10.13) (Ga dit na!). Derhalve geldt voor iedere $A \in \mathcal{B}^1$

$$P(T\in A) = \int_A f_T(t)\,d\lambda^1(t) = \frac{1}{2\pi}\int_A \{\int_{R^1} e^{-ity}\,\phi(y)\,e^{-\frac{y^2}{2\sigma^2}}\,dy\}\,d\lambda^1(t).$$

Verder is gemakkelijk in te zien dat $Y \xrightarrow{P} 0$ voor $\sigma\to\infty$, zodat (stelling 2.9.7) $T \xrightarrow{\mathcal{D}} X$ voor $\sigma\to\infty$. Hiermee is (2.10.11) bewezen.

Als ϕ een λ^k-sommeerbare functie is (gezien de continuïteit van ϕ komt de gestelde voorwaarde daarop neer), dan is f door (2.10.12) welgedefinieerd en continu op R^k. Verder volgt dan uit (2.10.11) en de gedomineerde convergentiestelling

$$P(A) = \frac{1}{(2\pi)^k}\int_A \lim_{\sigma\to\infty}\{\int_{R^k} e^{-it'y}\,\phi(y)\,e^{-\frac{1}{2\sigma^2}y'y}\,dy\}\,d\lambda^k(t) = \int_A f(t)d\lambda^k(t),$$

voor iedere P-continuïteitsverzameling A. Voor de verdelingsfunctie F van P geldt dus

$$(2.10.14) \qquad F(x) = \int_{(-\infty,x]} f(t)\,d\lambda^k(t) = \int_{(-\infty,x]} f(t)\,dt,$$

voor alle $x \in R^k$ waar F continu is. Maar dit houdt in dat F overal continu is, zodat (2.10.14) voor alle $x \in R^k$ geldt.

De inversiestelling en (2.10.2) impliceren samen, dat de verdeling van een stochastische vector X wordt vastgelegd door de verdelingen van alle lineaire combinaties $t'X$ van de componenten van X. Wij kunnen k-dimensionale verdelingen dus volledig beschrijven met behulp van één-dimensionale verdelingen. Zo kunnen wij de meerdimendionale normale verdeling karakteriseren door te stellen dat een k-dimensionale stochastische vector X dan en slechts dan $N(\mu,\Sigma)$-verdeeld is, als voor iedere $t \in R^k$ de stochastische grootheid $t'X$ een één-dimensionale $N(t'\mu,t'\Sigma t)$-verdeling heeft.
Tot nu toe hebben wij - sprekend over meerdimensionale normale verdelingen - altijd verondersteld, dat de covariantie-matrix Σ positief definiet is. De zojuist gegeven karakterisering van de $N(\mu,\Sigma)$-verdeling definieert echter ook een verdeling op $(R^k,\mathcal{B}^k)$ als wij slechts veronderstellen dat Σ een symmetrische reële positief semi-definiete $(k\times k)$-matrix is, mits wij per conventie de ééndimensionale $N(\mu,0)$-verdeling gelijk stellen aan de in het punt μ gedegenereerde verdeling. We kunnen dit als volgt inzien. Als Σ een reële symmetrische positief semidefiniete $(k\times k)$-matrix is, dan is er een orthogonale matrix P, zodanig dat $P'\Sigma P = \Lambda$ een diagonaal-matrix is met diagonaal-elementen $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$. Laten $Y_1, Y_2, \ldots, Y_k$ nu k o.o. stochastische grootheden zijn, waarbij voor iedere $j = 1, 2, \ldots, k$ de stochastische grootheid Y_j een $N(0,\lambda_j)$-verdeling heeft, met dien verstande dat $P(Y_j=0) = 1$ als $\lambda_j = 0$. De karakteristieke functies van $Y_1, Y_2, \ldots, Y_k$ worden dan gegeven door

$$\phi_{Y_j}(u) = e^{-\frac{1}{2}\lambda_j u^2}, \quad u \in R^1, \; j = 1, 2, \ldots, k$$

en dus (stelling 2.10.1) vinden wij voor de karakteristieke functie van de stochastische vector $Y = (Y_1,Y_2,\ldots,Y_k)'$,

$$\phi_Y(t) = e^{-\frac{1}{2}\sum_{j=1}^{k}\lambda_j t_j^2} = e^{-\frac{1}{2}t'\Lambda t}, \quad t = (t_1,t_2,\ldots,t_k)' \in R^k.$$

Als nu $\mu \in R^k$ en $X = \mu + PY$, dan voldoet de verdeling van X aan de gestelde eisen. Uit (2.10.1) volgt immers

$$\phi_X(t) = e^{it'\mu}\,\phi_Y(P't) = e^{it'\mu-\frac{1}{2}t'P\Lambda P't} = e^{it'\mu-\frac{1}{2}t'\Sigma t} ,$$

en dus ook

$$\phi_{t'X}(u) = e^{it'\mu u-\frac{1}{2}t'\Sigma t u^2}, \; u \in R^1, \; t \in R^k,$$

zodat $t'X$ voor iedere $t \in R^k$ de voorgeschreven ééndimensionale verdeling heeft. Wij noemen de verdeling van X de $N(\mu,\Sigma)$-verdeling, ook als Σ niet positief definiet maar alleen positief semidefiniet is. In het laatste geval spreken wij van een *singuliere normale verdeling*. Steeds is de karakteristieke functie van de vorm (2.10.8). Uit de beschreven constructie volgt, dat in het singuliere geval één of meer componenten van Y met kans 1 gelijk aan 0 zijn. Dit betekent dat de singuliere $N(\mu,\Sigma)$-verdeling kans 1 toekent aan een verzameling, die door de transformatie $x = \mu + Py$, m.a.w. door een translatie en een rotatie, ontstaat uit een lineaire deelruimte van R^k met dimensie gelijk aan de rang van Σ en dus kleiner dan k. Daar dergelijke verzamelingen λ^k-nulverzamelingen zijn, zijn singuliere normale verdelingen niet absoluut continu. Een singuliere normale verdeling heeft dus geen dichtheid!

Stelling 2.10.4
Een rij kansverdelingen $P_1, P_2, \ldots$ op $(R^k,\mathcal{B}^k)$ met karakteristieke functies $\phi_1, \phi_2, \ldots$ convergeert dan en dan alleen zwak naar een verdeling P met karakteristieke functie ϕ, als $\phi_n(t) \to \phi(t)$ voor $n \to \infty$ en alle $t \in R^k$.

Bewijs:
Daar $e^{it'x}$ voor iedere vaste $t \in R^k$ een begrensde continue functie van $x \in R^k$ is, impliceert zwakke convergentie van P_n naar P per definitie dat $\phi_n(t) \to \phi(t)$ voor $n \to \infty$. Het omgekeerde volgt uit stelling 2.10.5.

Stelling 2.10.5 *(Continuïteitsstelling van Lévy)*
Als de karakteristieke functies $\phi_1, \phi_2, \ldots$ van een rij kansverdelingen $P_1, P_2, \ldots$ op $(R^k,\mathcal{B}^k)$ de eigenschap hebben dat $\phi_n(t)$ voor iedere $t \in R^k$ en $n \to \infty$ naar een limiet $\phi(t)$ convergeert, die als functie van t continu is in de oorsprong, dan is ϕ de karakteristie functie van een kansverdeling P, en $P_n \xrightarrow{zw} P$ voor $n \to \infty$.

Bewijs:

We bewijzen eerst dat de collectie verdelingen $\Pi = \{P_1, P_2, \ldots\}$ uniform beperkt is. Als dat niet het geval was, en als S_r voor $r > 0$ de gesloten k-dimensionale bol met straal r om de oorsprong voorstelt, dan zou er een $\varepsilon > 0$ zijn zodanig dat er bij iedere $r > 0$ een natuurlijk getal m is met $P_m(S_r) < 1 - \varepsilon$. Dit betekent, dat er een stijgende rij $n_1, n_2, \ldots$ zou zijn, zodanig dat

$$(2.10.15) \qquad P_{n_j}(S_j) < 1 - \varepsilon, \; j = 1, 2, \ldots .$$

Geven wij, voor $\sigma > 0$, de k-dimensionale $N(0,\sigma^2 I)$-verdeling aan met Q_σ, dan geldt volgens stelling 2.10.2

$$(2.10.16) \qquad \int_{R^k} \phi_n(t)\, dQ_\sigma(t) = \int_{R^k} e^{-\frac{1}{2}\sigma^2 x'x}\, dP_n(x) =$$

$$= \int_{S_r} e^{-\frac{1}{2}\sigma^2 x'x}\, dP_n(x) + \int_{S_r^c} e^{-\frac{1}{2}\sigma^2 x'x}\, dP_n(x) \leqq$$

$$\leqq P_n(S_r) + e^{-\frac{1}{2}\sigma^2 r^2}, \quad n = 1,2,\ldots, \; \sigma>0, \; r>0.$$

Merk op, dat het eerste lid van (2.10.16) dus reëel is. Uit (2.10.15) en (2.10.16) zou dan volgen dat

$$\int_{R^k} \phi_{n_j}(t)\, dQ_\sigma(t) \leqq 1 - \varepsilon + e^{-\frac{1}{2}\sigma^2 j^2}, \; j = 1, 2, \ldots, \; \sigma > 0,$$

en dus, wegens de gedomineerde convergentie van ϕ_{n_j} naar ϕ,

$$(2.10.17) \qquad \int_{R^k} \phi(t)\, dQ_\sigma(t) \leqq 1 - \varepsilon.$$

Nu is ϕ echter continu in de oorsprong met $\phi(0) = \lim \phi_n(0) = 1$. Er is dus een $\delta > 0$ zo, dat $\mathrm{Re}\, \phi(t) > 1 - \varepsilon/2$ voor alle $t \in S_\delta$, en dus geldt

$$\int_{R^k} \phi(t)\, dQ_\sigma(t) = \int_{R^k} \mathrm{Re}\, \phi(t)\, dQ_\sigma(t) \geqq (1 - \tfrac{\varepsilon}{2})\, Q_\sigma(S_\delta) + Q_\sigma(S_\delta^c).$$

Daar Q_σ voor $\sigma \to 0$ zwak convergeert naar de verdeling die gedegenereerd is in de oorsprong, volgt hieruit

$$\liminf_{\sigma\to 0} \int_{R^k} \phi(t)\, dQ_\sigma(t) \geqq (1 - \frac{\varepsilon}{2}),$$

in strijd met (2.10.17). Π is dus uniform beperkt en derhalve (stelling 2.9.5) relatief compact. De rij $P_1, P_2, \ldots$ heeft dus minstens één zwak convergente deelrij. Uit het reeds bewezen deel van de vorige stelling volgt, dat de bij deze zwak convergente deelrij behorende karakteristieke functies naar de karakteristieke functie van de limietverdeling convergeren. Daar zij ook naar ϕ convergeren, kunnen wij hieruit concluderen dat ϕ de karakteristieke functie van een kansverdeling P is. Maar nu volgt met behulp van de vorige stelling en de inversiestelling, dat iedere zwak convergente deelrij van de rij $P_1, P_2, \ldots$ dezelfde limiet P heeft, en dus (stelling 2.9.3) dat $P_n \xrightarrow{zw} P$ voor $n \to \infty$.

Stelling 2.10.6

Als de eerste n momenten van een kansverdeling P op $(R^1, \mathcal{B}^1)$ bestaan en eindig zijn, dan heeft de karakteristieke functie ϕ van die verdeling n continue afgeleiden, gegeven door

$$(2.10.18) \qquad \phi^{(k)}(t) = i^k \int_{R^1} x^k e^{itx}\, dP(x), \; t \in R^1, \; k = 1, 2, \ldots, n.$$

Bewijs:

Stellen wij $\phi^{(0)} = \phi$, dan geldt (2.10.18) voor $k = 0$. Verder volgt uit (2.10.18) voor een $k < n$ dat (2.10.18) ook geldt voor $k + 1$. Het gegeven betekent immers, dat

$$\int_{R^1} |x|^k\, dP(x) < \infty, \; k = 1, 2, \ldots, n.$$

en dus impliceren (2.10.18) en de gedomineerde convergentiestelling

$$\phi^{(k+1)}(t) = \lim_{h\to 0} i^k \int_{R^1} x^k\, e^{itx}\, \frac{e^{ihx}-1}{h}\, dP(x) = i^{k+1} \int_{R^1} x^{k+1} e^{itx}\, dP(x),$$

daar

$$\left| \frac{e^{ihx}-1}{h} \right| \leqq |x|,\ x \in R^1,\ h \in R^1,\ h \neq 0.$$

De continuïteit van $\phi^{(k)}$ voor $k = 1, 2, \ldots, n-1$ volgt uiteraard uit de differentieerbaarheid. De continuïteit van $\phi^{(n)}$ kan men rechtstreeks met behulp van de gedomineerde convergentiestelling bewijzen.

Karakteristieke functies worden veel gebruikt als hulpmiddel bij het bestuderen van verdelingen van sommen van o.o. stochastische grootheden. Ter illustratie geven wij hier de volgende stellingen. De eerste is een al eerder genoemde versie van de zwakke wet van de grote aantallen, die aanzienlijk sterker is dan stelling 2.8.2, zij het dan dat hier de stochastische grootheden in kwestie niet alleen o.o. maar ook identiek verdeeld worden verondersteld.

Stelling 2.10.7

Als $X_1, X_2, \ldots$ o.o. en identiek verdeelde stochastische grootheden zijn met eindige verwachting μ, dan convergeren de gemiddelden

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j,\ n = 1, 2, \ldots,$$

voor $n \to \infty$ in waarschijnlijkheid naar μ.

Bewijs:

Als ϕ de karakteristieke functie van de X_j, $j = 1, 2, \ldots$ is, dan wordt de karakteristieke functie van $\overline{X}_n$ volgens stelling 2.10.1 gegeven door

$$\phi_{\overline{X}_n}(t) = (\phi(\tfrac{t}{n}))^n,\ t \in R^1,\ n = 1, 2, \ldots .$$

Daar ϕ volgens stelling 2.10.6 differentieerbaar is met $\phi'(0) = i\mu$, impliceert dit voor vaste doch willekeurige t en $n \to \infty$,

$$\phi_{\overline{X}_n}(t) = (1 + \frac{i\mu t}{n} + o(\tfrac{1}{n}))^n \to e^{i\mu t}.$$

Nu is $e^{i\mu t}$ de karakteristieke functie van de verdeling die gedegenereerd is in het punt μ, en dus (stelling 2.10.4) geldt $\overline{X}_n \overset{D}{\to} \mu$, zodat ook $\overline{X}_n \overset{P}{\to} \mu$ voor $n \to \infty$.

De volgende stelling is een eenvoudige versie van de zogenaamde *centrale limietstelling*. Deze zegt dat onder zekere voorwaarden sommen van grote aantallen o.o. stochastische grootheden bij benadering normaal verdeeld zijn. Als gevolg hiervan speelt de normale verdeling een belangrijke rol in vele situaties, waar men met sommen of gemiddelden van vele waarnemingen te doen heeft.

Stelling 2.10.8 *(Lindeberg-Lévy)*

Als X_1, X_2, ... o.o. en identiek verdeelde stochastische grootheden zijn met eindige verwachting μ en variantie σ^2, dan convergeert de verdeling van de gestandaardiseerde som

$$Z_n = \frac{\sum_{j=1}^{n} X_j - n\mu}{\sigma\sqrt{n}}, \quad n = 1, 2, \ldots$$

voor $n \to \infty$ zwak naar de $N(0,1)$-verdeling.

Bewijs:

Als ϕ de karakteristieke functie van $X_j - \mu$, $j = 1, 2, \ldots$ is, dan wordt de karakteristieke functie van Z_n gegeven door

$$\phi_{Z_n}(t) = \left(\phi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n, \quad t \in R^1, \; n = 1, 2, \ldots .$$

Uit het gegeven volgt verder, wegens stelling 2.10.6 dat ϕ twee keer differentieerbaar is met $\phi'(0) = iE(X_j-\mu) = 0$ en $\phi''(0) = -E(X_j-\mu)^2 = -\sigma^2$. Voor vaste doch willekeurige $t \in R^1$ en $n \to \infty$ geldt dus

$$\phi_{Z_n}(t) = \left(1 - \tfrac{1}{2}\frac{t^2}{n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-\frac{1}{2}t^2},$$

en zo volgt het gestelde uit stelling 2.10.4 en (2.10.6).

Een interessant en historisch belangrijk speciaal geval van de laatste stelling is het volgende (zie ook stelling 2.8.3).

Stelling 2.10.9 *(de Moivre-Laplace)*
Als $X^{(n)}$, $n = 1, 2, \ldots$, het aantal successen in n o.o. alternatieven is, waarbij voor ieder alternatief de kans op een succes gelijk is aan $p = 1-q$ met $pq > 0$, dan convergeren de gestandaardiseerde grootheden

$$Z_n = \frac{X^{(n)}-np}{\sqrt{npq}}, \quad n = 1, 2, \ldots,$$

voor $n \to \infty$ naar een $N(0,1)$-verdeelde stochastische grootheid, zodat

$$P(np+a\sqrt{npq}<X^{(n)}\leq np+b\sqrt{npq}) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}t^2}\,dt$$

voor $n \to \infty$ en alle $a, b \in R^1$ met $a \leq b$.

Op deze laatste stelling berust de zogenaamde *normale benadering* voor binomiale verdelingen. Als een stochastische grootheid X een binomiale verdeling heeft met parameters n en $p = 1-q$ met $pq > 0$, dan geldt voor $k < l$ en grote waarden van n

(2.10.19) $$P(k<X\leq l) \approx \Phi\left(\frac{l-np}{\sqrt{npq}}\right) - \Phi\left(\frac{k-np}{\sqrt{npq}}\right),$$

waarin Φ de verdelingsfunctie van de $N(0,1)$-verdeling is. Voor gehele waarden van k en l vervangt men (2.10.19) vaak door

(2.10.20) $$P(k<X\leq l) \approx \Phi\left(\frac{l+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{npq}}\right).$$

In de meeste gevallen geeft dit een iets betere benadering. In het bijzonder krijgt men dan voor gehele k,

$$P(X=k) \approx \Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{npq}}\right) - \Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{npq}}\right).$$

Men zegt dat (2.10.20) uit (2.10.19) onstaat door het aanbrengen van een *continuïteitscorrectie*.

Opgaven

1. Ga na dat de karakteristieke functie van

a) de binomiale verdeling met parameters n en p door

$$\phi(t) = (pe^{it}+q)^n, \quad -\infty < t < \infty;$$

b) de Poisson verdeling met parameter λ door

$$\phi(t) = e^{-\lambda(1-e^{it})}, \quad -\infty < t < \infty;$$

c) de negatief binomiale verdeling met parameters k en p door

$$\phi(t) = \left(\frac{pe^{it}}{1-qe^{it}}\right)^k, \quad -\infty < t < \infty;$$

d) de multinomiale verdeling met parameters n en $p_1, p_2, \ldots, p_k$ door

$$\phi(t_1,\ldots,t_k) = \left(\sum_{j=1}^{k} p_j e^{it_j}\right)^n, \quad -\infty < t_j < \infty, \; j = 1, 2, \ldots, k;$$

e) de homogene verdeling op $(-1,1)$ door

$$\phi(t) = \frac{\sin t}{t}, \quad -\infty < t < \infty;$$

f) de standaard gamma verdeling met parameter k door

$$\phi(t) = \frac{1}{(1-it)^k}, \quad -\infty < t < \infty;$$

g) de standaard dubbel-exponentiële verdeling door

$$\phi(t) = \frac{1}{1+t^2}, \quad -\infty < t < \infty;$$

h) de standaard Cauchy verdeling door

$$\phi(t) = e^{-|t|}, \quad -\infty < t < \infty,$$

gegeven wordt.

2. Bewijs:

   a) Als X_λ een Poisson verdeling met parameter λ heeft, dan convergeert $(X_\lambda-\lambda)/\surd\lambda$ voor $\lambda \to \infty$ in verdeling naar een $N(0,1)$-verdeelde stochastische grootheid.

   b) Als X_k een standaard gamma verdeling met parameter k heeft, dan convergeert $(X_k-k)/\surd k$ voor $k \to \infty$ in verdeling naar een $N(0,1)$-verdeelde grootheid.

3. Laat zien dat het gemiddelde van n o.o. standaard Cauchy verdeelde stochastische grootheden zelf ook een standaard Cauchy verdeling heeft.

4. Bewijs dat een rij k-dimensionale stochastische vectoren $X_1, X_2, \ldots$ dan en dan alleen in verdeling naar een stochastische vector X convergeert als $t'X_n \xrightarrow{\mathcal{D}} t'X$ voor $n \to \infty$ en alle $t \in R^k$.

5. Geef generalisaties van de stellingen 2.10.7, 2.10.8 en 2.10.9 voor o.o. en identiek verdeelde k-dimensionale stochastische vectoren $X_1, X_2, \ldots$.

2.11. VOORWAARDELIJKE VERWACHTING

Zij gegeven een kansruimte $(\Omega,\mathcal{A},P)$. Voor $A, C \in \mathcal{A}$ met $P(A) \neq 0$ wordt de voorwaardelijke waarschijnlijkheid van C gegeven A gedefinieerd door (zie §2.2)

$$(2.11.1) \qquad P(C|A) = \frac{P(AC)}{P(A)} .$$

Indien $V = (V_1,\ldots,V_m)$ een stochastische vector op $(\Omega,\mathcal{A},P)$ voorstelt, kunnen wij de voorwaardelijke kansverdeling van V gegeven A definiëren door

$$(2.11.2) \qquad P_{V|A}(B) = P(V\in B|A), \ B \in \mathcal{B}^m.$$

Aangezien $P(C|A)$ voor vaste A als functie van C een kans op $(\Omega,\mathcal{A})$ is, is $P_{V|A}$ een kansverdeling op R^m. Als $Y = (Y_1,\ldots,Y_k)$ een discreet verdeelde stochastische vector op $(\Omega,\mathcal{A},P)$ voorstelt, dan wordt voor iedere $y \in R^k$ waarvoor $P(Y=y) \neq 0$ de voorwaardelijke kansverdeling van V gegeven $Y = y$ gedefinieerd door

(2.11.3) $\qquad P_{V|Y=y}(B) = P(V\in B|Y=y),\ B \in \mathcal{B}^m.$

Wij interpreteren $P_{V|Y=y}$ als de kansverdeling van V als bekend is dat Y bij het experiment de waarde y heeft aangenomen. Als ϕ een reële Borelfunctie op R^m is, dan wordt de verwachting van $\phi(V)$ met betrekking tot de verdeling (2.11.3) de voorwaardelijke verwachting van $\phi(V)$ gegeven $Y = y$ genoemd:

(2.11.4) $$E(\phi(V)|Y=y) = \int_{R^m} \phi(v)\ dP_{V|Y=y}(v).$$

Definitie (2.11.3) wordt mogelijk gemaakt doordat $P(Y=y) \neq 0$. Indien echter Y een continue verdeling bezit en dus $P(Y=y) = 0$ voor alle $y \in R^k$, moeten wij op een andere wijze te werk gaan. Als het paar (V,Y) bijvoorbeeld continu verdeeld is met dichtheid $f_{V,Y}$, dan is ook Y continu verdeeld met dichtheid f_Y. Men zou dan kunnen overwegen om, onder het opleggen van voldoende regulariteitsvoorwaarden om de onderstaande limietovergangen te rechtvaardigen, (2.11.3) te vervangen door

$$P_{V|Y=y}(B) = \lim_{h\to 0} P(V\in B|Y\in[y,y+h]) =$$
$$= \lim_{h\to 0} \frac{\int_{[y,y+h]}\int_B f_{V,Y}(v,u)dvdu}{\int_{[y,y+h]} f_Y(u)du} = \frac{\int_B f_{V,Y}(v,y)dv}{f_Y(y)}$$

voor alle y waarvoor $f_Y(y) \neq 0$. Dit komt neer op een definitie waarbij de voorwaardelijke verdeling $P_{V|Y=y}$ wordt gegeven door zijn dichtheid

(2.11.5) $$f_{V|Y=y}(v) = \frac{f_{V,Y}(v,y)}{f_Y(y)},\ f_Y(y) \neq 0,$$

hetgeen een duidelijke analogie met (2.11.3) vertoont. Voor een reële Borelfunctie ϕ op R^m zou men nu voorts kunnen definiëren

(2.11.6) $$E(\phi(V)|Y=y) = \int_{R^m} \phi(v)\ f_{V|Y=y}(v)dv.$$

Een dergelijke aanpak is echter weinig algemeen. Men geeft er daarom de voorkeur aan de begrippen voorwaardelijke verdeling en verwachting op de stelling van Radon-Nikodym te baseren. Op grond van de overweging dat een kans als een verwachting kan worden opgevat - namelijk als verwachting van een indicator functie - stelt men hierbij de definitie van voorwaardelijke verwachting centraal.

Zij X een stochastische grootheid gedefinieerd op een kansruimte (Ω,A,P) en veronderstel dat $E|X| < \infty$. Zij voorts A_0 een deel-σ-algebra van A.

Stelling 2.11.1
Er bestaat een A_0-meetbare stochastische grootheid X' zodanig dat voor iedere $A \in A_0$

$$(2.11.5) \qquad \int_A XdP = \int_A X'dP.$$

Ieder tweetal A_0-meetbare stochastische grootheden met de eigenschap (2.11.5) is P-bijna overal gelijk.

Bewijs:
Veronderstel eerst dat $X \geq 0$ op Ω en definieer voor alle $A \in A_0$

$$\nu(A) = \int_A XdP.$$

Men verifieert gemakkelijk dat ν een eindige maat op (Ω,A_0) is die P-absoluut continu is. Uit de stelling van Radon-Nikodym volgt nu het bestaan van een niet-negatieve, eindige en A_0-meetbare functie X' waarvoor

$$\nu(A) = \int_A X'dP, \quad A \in A_0;$$

twee functies met deze eigenschappen zijn P-b.o. gelijk.

Indien X niet niet-negatief is op Ω schrijven wij $X = X^+ - X^-$, waarbij $X^+ = \max(X,0)$ en $X^- = -\min(X,0)$ beide niet-negatief en P-sommeerbaar zijn. Toepassing van het voorafgaande op X^+ en X^- levert het bestaan van niet-negatieve, eindige en A_0-meetbare functies $X^{+'}$ en $X^{-'}$ zodanig dat voor iedere $A \in A_0$

$$\int_A X^+ dP = \int_A X^{+'} dP, \quad \int_A X^- dP = \int_A X^{-'} dP$$

en hieruit volgt dat de A_0-meetbare functie $X' = X^{+'} - X^{-'}$ voldoet aan (2.11.5).

Als X' en X" beiden A_0-meetbaar zijn en aan (2.11.5) voldoen, dan geldt voor iedere $A \in A_0$

$$\int_A X' dP = \int_A X'' dP.$$

Door hierin voor A de verzamelingen $\{\omega: X'(\omega) > X''(\omega)\}$ respectievelijk $\{\omega: X'(\omega) < X''(\omega)\}$ te kiezen ziet men dat deze beide verzamelingen P-maat nul bezitten, zodat $P(X'=X'') = 1$.

Definitie 2.11.1
Een A_0-meetbare stochastische grootheid X' zodanig dat voor alle $A \in A_0$ aan (2.11.5) is voldaan wordt een *voorwaardelijke verwachting van* X *gegeven* A_0 genoemd en aangeduid door de notatie $X' = E(X|A_0)$. Indien wij de afhankelijkheid van $\omega \in \Omega$ in de notatie tot uitdrukking wensen te brengen schrijven wij $X'(\omega) = E(X|A_0)(\omega)$.

In plaats van over een voorwaardelijke verwachting spreekt men ook wel over een *versie* van de voorwaardelijke verwachting. Volgens stelling 2.11.1 zijn twee versies P-bijna overal gelijk.

De volgende stelling geeft een tweede karakterisering van $E(X|A_0)$ die echter in wezen nauwelijks van de eerste verschilt.

Stelling 2.11.2
$X' = E(X|A_0)$ dan en slechts dan indien X' A_0-meetbaar is en

$$\int XZ dP = \int X'Z dP \tag{2.11.6}$$

(d.w.z. $EXZ = EX'Z$) voor iedere A_0-meetbare stochastische grootheid Z waarvoor XZ P-sommeerbaar is (d.w.z. $E|XZ| < \infty$).

Bewijs:

Voor $A \in \mathcal{A}_0$ is de stochastische grootheid $Z = I_A$ $\mathcal{A}_0$-meetbaar en daar $E|X| < \infty$ is XZ P-sommeerbaar. Uit (2.11.6) volgt dus (2.11.5). Het bewijs dat ook omgekeerd (2.11.6) uit (2.11.5) volgt verloopt in vier stappen.

a) Als Z een elementaire $\mathcal{A}_0$-meetbare functie is, is het te bewijzene triviaal.

b) Als $X \geqq 0$ en $Z \geqq 0$ op Ω, dan kiezen wij een rij niet-negatieve elementaire $\mathcal{A}_0$-meetbare functies $Z_n \uparrow Z$. Volgens a) geldt $\int XZ_n dP = \int X'Z_n dP$ voor alle n. Uit het bewijs van stelling 2.11.1 volgt dat $X \geqq 0$ op Ω impliceert dat $X' \geqq 0$ P-bijna overal. Toepassing van de monotone convergentie stelling levert $\int XZ_n dP \to \int XZdP$ en $\int X'Z_n dP \to \int X'ZdP$, waaruit (2.11.6) volgt.

c) Als $X \geqq 0$ op Ω passen wij b) toe op Z^+ en Z^- en vinden $\int XZ^+ dP = \int X'Z^+ dP$ en $\int XZ^- dP = \int X'Z^- dP$. Daar XZ P-sommeerbaar is zijn deze integralen eindig zodat (2.11.6) geldt.

d) Voor willekeurige X en Z kan c) worden toegepast op X^+ en X^-, daar de P-sommeerbaarheid van XZ die van X^+Z en X^-Z impliceert. Dit geeft $\int X^+ZdP = \int (X^+)'ZdP$ en $\int X^-ZdP = \int (X^-)'ZdP$. Daar deze integralen eindig zijn en uit het bewijs van stelling 2.11.1 volgt dat $X' = (X^+)'-(X^-)'$ P-bijna overal, geldt (2.11.6) ook nu.

In het voorafgaande kwamen herhaaldelijk beweringen voor die P-bijna overal juist zijn. In een meer op de waarschijnlijkheidsrekening afgestemde terminologie zegt men dat een dergelijke bewering *met kans 1* of *bijna zeker* (afgekort: b.z.) juist is (in het Engels: *almost surely* of a.s.). In de volgende stelling worden een aantal eenvoudige eigenschappen van voorwaardelijke verwachtingen opgesomd. Daar voorwaardelijke verwachtingen door ons slechts zijn gedefinieerd voor stochastische grootheden met eindige verwachtingen, wordt hierbij stilzwijgend aangenomen dat X, X_1 en X_2 inderdaad eindige verwachtingen bezitten.

Stelling 2.11.3

a. $E(E(X|A_0)) = EX$.

b. $E(X|A_0) = X$ b.z. als X A_0-meetbaar is.

c. $E(XZ|A_0) = ZE(X|A_0)$ b.z. als Z A_0-meetbaar is met $E|XZ| < \infty$.

d. $E(X|A_0) = EX$ b.z. als A_0 en de door X geïnduceerde σ-algebra A_X o.o. zijn.

e. $E(E(X|A_0)|A_1) = E(X|A_0) = E(E(X|A_1)|A_0)$ b.z. als $A_0 \subset A_1 \subset A$.

f. Als $X = c$ b.z. dan geldt $E(X|A_0) = c$ b.z..

g. $E(aX_1+bX_2|A_0) = aE(X_1|A_0) + bE(X_2|A_0)$ b.z..

h. Als $X \geqq 0$ b.z. dan geldt $E(X|A_0) \geqq 0$ b.z..

i. $|E(X|A_0)| \leqq E(|X||A_0)$ b.z..

Bewijs:

a, b, f, g en de eerste gelijkheid in e volgen rechtstreeks uit stelling 2.11.1 en de daarop aansluitende definitie 2.11.1. h volgt uit het bewijs van stelling 2.11.1. Uit g en h volgt i daar $E(X|A_0) = E(X^+|A_0) - E(X^-|A_0)$ b.z. en $E(|X||A_0) = E(X^+|A_0) + E(X^-|A_0)$ b.z.. Door in stelling 2.11.2 Z te vervangen door ZI_A, $A \in A_0$, verkrijgt men c. Als A_0 en A_X o.o. zijn en Z A_0-meetbaar is, zijn X en Z o.o. zodat volgens stelling 2.7.3 $EXZ = EXEZ$ indien $E|XZ| < \infty$. Toepassing van stelling 2.11.2 levert d. Daar

$$\int_A XdP = \int_A E(X|A_1)dP,\ A \in A_1,$$

$$\int_A XdP = \int_A E(X|A_0)dP,\ A \in A_0,$$

zijn de beide rechterleden gelijk voor iedere $A \in A_0$ als $A_0 \subset A_1 \subset A$. Daar $E(X|A_0)$ A_0-meetbaar is en $E(E(X|A_0)) = EX$ eindig is levert toepassing van stelling 2.11.1 en definitie 2.11.1 de tweede gelijkheid in e.

Men kan $E(X|A_0)$ blijkens zijn definitie opvatten als een vergroving van X tot een A_0-meetbare stochastische grootheid waarbij de gemiddelde waarde over A_0-meetbare verzamelingen met betrekking tot P behouden blijft. Als $A_0 = A$ treedt geen vergroving op daar volgens stelling 2.11.3b $E(X|A) = X$ b.z.. Naarmate de σ-algebra A_0 echter kleiner wordt treedt volgens stelling 2.11.3e stapgewijs een steeds verder gaande vergroving op.

Dit proces eindigt als men voor A_0 de kleinste σ-algebra in A kiest, namelijk $A_0 = \{\phi,\Omega\}$. Men gaat met behulp van definitie 2.11.1 of stelling 2.11.3d eenvoudig na dat in dit geval $E(X|A_0) = EX$ b.z. zodat X nu tot een P-bijna overal constante stochastische grootheid is teruggebracht.

Hoewel $E(X|A_0)$ niet als een integraal is gedefinieerd corresponderen de eigenschappen f-i in stelling 2.11.3 met eigenschappen van EX (zie stelling 2.7.1 b-e). Ook de hierna volgende stellingen wijzen erop dat een voorwaardelijke verwachting zich, grof gesproken, bijna zeker als een verwachting gedraagt (verg. stellingen 1.6.2-1.6.5).

Stelling 2.11.4 *(monotone convergentie stelling)*
Als $0 \leq X_n \uparrow X$ b.z. en $EX < \infty$, dan geldt ook $E(X_n|A_0) \uparrow E(X|A_0)$ b.z..

Bewijs:

Daar X_n b.z. een niet-dalende rij niet-negatieve stochastische grootheden is, geldt hetzelfde voor de rij $X_n' = E(X_n|A_0)$ volgens stelling 2.11.3g en h. De rij X_n' convergeert dus b.z. naar een A_0-meetbare functie Z. Toepassing van stelling 1.6.2 op $X_n I_A$ en XI_A respectievelijk $X_n' I_A$ en ZI_A levert voor iedere $A \in A_0$

$$\int_A X_n dP \to \int_A X dP, \quad \int_A X_n' dP \to \int_A Z dP.$$

Daar de beide linkerleden voor iedere n aan elkaar gelijk zijn, zijn ook de rechterleden gelijk. Aangezien Z A_0-meetbaar is, volgt hieruit dat $Z = E(X|A_0)$ b.z..

Stelling 2.11.5 *(lemma van Fatou)*
Als $X_n \geq 0$ b.z. en $EX_n < \infty$ voor $n = 1, 2, \ldots$ en als voorts $E(\liminf_n X_n) < \infty$, dan geldt

$$E(\liminf_n X_n|A_0) \leq \liminf_n E(X_n|A_0) \text{ b.z..}$$

Als $X_n \leq 0$ b.z. en $EX_n > -\infty$ voor $n = 1, 2, \ldots$ en als voorts $E(\limsup X_n) > -\infty$, dan geldt

$$E(\limsup_n X_n|A_0) \geq \limsup_n E(X_n|A_0) \text{ b.z..}$$

Bewijs:

Het tweede deel van de stelling volgt uit het eerste door X_n door $-X_n$ te vervangen.Om het eerste deel te bewijzen merken wij op dat

$$\liminf_n X_n = \lim_{n\to\infty} Y_n,\ Y_n = \inf_{m\geq n} X_m.$$

Daar Y_n een niet-dalende rij b.z. niet-negatieve stochastische grootheden is volgt uit stelling 2.11.4

$$E(\liminf_n X_n | A_0) = \lim_{n\to\infty} E(Y_n | A_0).$$

Voor iedere $m \geq n$ geldt $Y_n \leq X_m$. Volgens stelling 2.11.3g en h geldt dus $E(Y_n|A_0) \leq E(X_m|A_0)$ b.z. voor iedere $m \geq n$, zodat

$$E(Y_n|A_0) \leq \inf_{m\geq n} E(X_m|A_0) \text{ b.z..}$$

Hieruit volgt dat met kans 1

$$E(\liminf_n X_n|A_0) \leq \lim_{n\to\infty} \inf_{m\geq n} E(X_m|A_0) = \liminf_n E(X_n|A_0).$$

In de formuleringen van stellingen 2.11.4 en 2.11.5 wordt de eindigheid van een aantal verwachtingen geëist terwijl overeenkomstige voorwaarden in stellingen 1.6.2 en 1.6.3 ontbreken. De reden hiervan is gelegen in het feit dat wij voorwaardelijke verwachtingen slechts gedefinieerd hebben voor stochastische grootheden die eindige verwachtingen bezitten: de bedoelde voorwaarden in stellingen 2.11.4 en 2.11.5 dienen uitsluitend om het bestaan van de in deze stellingen genoemde voorwaardelijke verwachtingen te garanderen. Men kan echter de definitie van voorwaardelijke verwachting gemakkelijk uitbreiden tot stochastische grootheden die P-integreerbaar zijn, d.w.z. waarvoor de verwachting gedefinieerd doch niet noodzakelijk eindig is. Doet men dit, dan blijven stellingen 2.11.4 en 2.11.5 ook zonder de bedoelde voorwaarden gelden. Stelling 2.11.4 is dan ook niet meer een speciaal geval van de hierna volgende stelling 2.11.6, hetgeen thans wel het geval is.

Stelling 2.11.6 *(gedomineerde convergentie stelling)*
Als $X_n \to X$ b.z. en er een stochastische grootheid Y met $E|Y| < \infty$ bestaat zodanig dat $|X_n| \leqq Y$ b.z. voor alle n, dan geldt

$$E(|X_n - X|\,|A_0) \to 0 \text{ b.z., zodat}$$

$$E(X_n|A_0) \to E(X|A_0) \text{ b.z..}$$

Bewijs:
Uit het gegeven volgt dat de in de stelling en in het onderstaande bewijs voorkomende stochastische grootheden eindige verwachtingen bezitten zodat hun voorwaardelijke verwachtingen gedefinieerd zijn. Zij $Y_n = |X_n - X|$. Dan is $2Y - Y_n \geqq 0$ b.z. zodat volgens stelling 2.11.3g en stelling 2.11.5

$$E(2Y|A_0) - E(\limsup_n Y_n|A_0) = E(\liminf_n (2Y-Y_n)|A_0) \leqq$$

$$\leqq \liminf_n E(2Y-Y_n|A_0) = E(2Y|A_0) - \limsup_n E(Y_n|A_0) \text{ b.z.,}$$

zodat volgens stelling 2.11.3h en f

$$0 \leqq \limsup_n E(Y_n|A_0) \leqq E(\limsup_n Y_n|A_0) = 0 \text{ b.z..}$$

De tweede uitspraak van de stelling volgt door toepassing van stelling 2.11.3 i en g.

In het bovenstaande werd stelling 2.11.5 uit stelling 2.11.4 en stelling 2.11.6 uit stelling 2.11.5 afgeleid. Merk op dat men op precies dezelfde manier stelling 1.6.3 uit stelling 1.6.2 en stelling 1.6.4 uit stelling 1.6.3. kan afleiden.

Stelling 2.11.7 *(ongelijkheid van Jensen)*
Zij g een reële meetbare en convexe functie op R^1. Indien $E|X| < \infty$ en $E|g(X)| < \infty$ dan geldt

$$E(g(X)|A_0) \geqq g(E(X|A_0)) \text{ b.z..}$$

Bewijs:

Daar g meetbaar en convex is, is g ook continu op R^1 en bestaat er voor iedere $a \in R^1$ een reëel getal $m(a)$ zodanig dat $g(x) \geqq g(a) + m(a)(x-a)$ voor alle $x \in R^1$ (zie pag. 24). Aangezien g begrensd is op begrensde intervallen geldt hetzelfde voor de functie m; men verifieert zelfs eenvoudig dat m monotoon niet-dalend is op R^1. Uit $g(X) \geqq g(a) + m(a)(X-a)$ volgt volgens stelling 2.11.3 dat iedere $a \in R^1$

$$E(g(X)|A_0) \geqq g(a) + m(a)(E(X|A_0)-a) \text{ b.z..}$$

Zij D een aftelbare verzameling die dicht ligt in R^1. Daar een aftelbare vereniging van P-nulverzamelingen zelf ook een P-nulverzameling is, geldt

$$(2.11.7) \qquad E(g(X)|A_0) \geqq \sup_{d \in D} \{g(d)+m(d)(E(X|A_0)-d)\} \text{ b.z..}$$

Beschouw de verzameling $A \in A_0$ van die $\omega \in \Omega$ waarvoor $E(X|A_0)(\omega)$ eindig is en waarvoor

$$E(g(X)|A_0)(\omega) \geqq \sup_{d \in D} \{g(d)+m(d)(E(X|A_0)(\omega)-d)\}.$$

Daar $E(X|A_0)$ b.z. eindig is volgt uit (2.11.7) dat $P(A) = 1$. Voor een willekurige vaste $\omega \in A$ kiezen wij een rij $d_n \in D$ die voor $n \to \infty$ naar $E(X|A_0)(\omega)$ convergeert. Daar g continu is op R^1 en m begrensd is op begrensde intervallen vinden wij voor iedere $\omega \in A$

$$E(g(X)|A_0)(\omega) \geqq \lim_{n\to\infty} \{g(d_n)+m(d_n)(E(X|A_0)(\omega)-d_n)\} =$$

$$= g(E(X|A_0)(\omega)).$$

Beschouw een stochastische vector $Y = (Y_1,\dots,Y_k)$, $1 \leqq k \leqq \infty$, gedefinieerd op dezelfde kansruimte (Ω,A,P) waarop ook de stochastische grootheid X is gedefinieerd. Zij A_Y de door Y in Ω geïnduceerde σ-algebra. Indien $E|X| < \infty$, definiëren wij de *voorwaardelijke verwachting van* X *gegeven* Y als de voorwaardelijke verwachting van X gegeven A_Y:

(2.11.8) $$E(X|Y) = E(X|A_Y).$$

Dit is een A_Y-meetbare stochastische grootheid, hetgeen volgens stelling 1.4.1 equivalent is met het bestaan van een reële Borelfunctie g op R^k zodanig dat $E(X|Y) = g(Y)$. Uit definitie 2.11.1 volgt nu dat (2.11.8) equivalent is met:

Definitie 2.11.2

Een stochastische grootheid X' wordt een voorwaardelijke verwachting van X gegeven Y genoemd (notatie: $X' = E(X|Y)$) indien er een reële Borelfunctie g op R^k bestaat zodanig dat $X' = g(Y)$ en indien voorts

(2.11.9) $$\int_A X dP = \int_A X' dP$$

voor iedere $A \in A_Y$, d.w.z. voor iedere $A = Y^{-1}(B)$, $B \in B^k$.

Merk op dat volgens de overplantingsstelling

$$\int_{Y^{-1}(B)} g(Y) dP = \int_B g(y) dP_Y(y),$$

zodat (2.11.9) in definitie 2.11.2 mag worden vervangen door

(2.11.10) $$\int_{Y^{-1}(B)} X dP = \int_B g dP_Y \qquad \text{voor alle } B \in B^k.$$

Wegens (2.11.8) zijn twee versies $g_1(Y)$ en $g_2(Y)$ van $E(X|Y)$ als functies op Ω P-bijna overal gelijk. Dit is equivalent met de bewering dat de bij deze versies behorende functies g_1 en g_2 op R^k P_Y-bijna overal gelijk zijn.

Indien twee stochastische vectoren Y en $\tilde{Y}$ dezelfde σ-algebra $A_Y = A_{\tilde{Y}}$ in Ω induceren, dan geldt volgens (2.11.8) $E(X|Y) = E(X|\tilde{Y})$ b.z.. Dit betekent dat versies van deze voorwaardelijke verwachtingen $g(Y(\omega)) = E(X|Y)(\omega)$ en $\tilde{g}(\tilde{Y}(\omega)) = E(X|\tilde{Y})(\omega)$ als functie van ω P-bijna overal gelijk zijn. De functies g en $\tilde{g}$ zijn in het algemeen natuurlijk niet gelijk; zij kunnen zelfs op ruimten met van elkaar verschillende dimensies gedefinieerd zijn.

Uit (2.11.8) volgt ook dat $E(X|Y)$ alle in stellingen 2.11.2-2.11.7 opgesomde eigenschappen van $E(X|A_0)$ bezit. In de twee volgende stellingen noemen wij hiervan slechts de met stellingen 2.11.2 en 2.11.3c-e corresponderende eigenschappen, daar deze met behulp van stelling 1.4.1 formeel een ander aanzien krijgen (eigenschap b in stelling 2.11.3 is een speciaal geval van c).

Stelling 2.11.8

$X' = E(X|Y)$ dan en slechts dan indien er een reële Borelfunctie g op R^k bestaat zodanig dat $X' = g(Y)$ en indien voorts $EXh(Y) = EX'h(Y)$ voor iedere reële Borelfunctie h op R^k waarvoor $E|Xh(Y)| < \infty$.

Stelling 2.11.9

Op een kansruimte (Ω,A,P) zijn gedefinieerd een stochastische grootheid X met $E|X| < \infty$ en stochastische vectoren $Y = (Y_1,\ldots,Y_k)$ en $\tilde{Y} = (\tilde{Y}_1,\ldots,\tilde{Y}_m)$, $1 \leqq k,\ m \leqq \infty$, zodanig dat $Y = \phi(\tilde{Y})$ waarbij ϕ een Borelfunctie op R^m met waarden in R^k voorstelt. Zij h een reële Borelfunctie op R^k. Dan geldt:

1. $E(Xh(Y)|Y) = h(Y)E(X|Y)$ b.z. als $E|Xh(Y)| < \infty$.
2. $E(X|Y) = EX$ b.z. als X en Y o.o. zijn.
3. $E(E(X|Y)|\tilde{Y}) = E(X|Y) = E(E(X|\tilde{Y})|Y)$ b.z..

Wij wensen nu het begrip *voorwaardelijke verwachting van* X *gegeven* $Y = y$ (notatie: $E(X|Y=y)$) voor $y \in R^k$ in te voeren. Hiertoe kiezen wij een versie $g(Y)$ van $E(X|Y)$ en definiëren $E(X|Y=y) = g(y)$ voor alle $y \in R^k$. Keuze van een andere versie van $E(X|Y)$ leidt voor P_Y-bijna alle $y \in R^k$ tot dezelfde definitie van $E(X|Y=y)$. Volgens definitie 2.11.2 en het daarna volgende is iedere reële Borelfunctie g op R^k die aan (2.11.10) voldoet een versie van de functie $E(X|Y=\cdot)$ op R^k. De eigenschappen van $g(y) = E(X|Y=y)$ volgen rechtstreeks uit die van $g(Y) = E(X|Y)$, waarbij beweringen die voor $E(X|Y)$ bijna zeker gelden overgaan in beweringen over $E(X|Y=y)$ voor P_Y-bijna alle $y \in R^k$.

Naar analogie van de relatie $P(A) = EI_A$ kunnen wij ook voorwaardelijke waarschijnlijkheden als voorwaardelijke verwachtingen van de corresponderende indicator functies definiëren. Als boven beschouwen wij een kansruimte (Ω,A,P), een deel-σ-algebra A_0 van A en een stochastische vector

$Y = (Y_1,\ldots,Y_k)$, $1 \leqq k \leqq \infty$, op (Ω,A,P) en definiëren de volgende begrippen. Voor $A \in A$ is

$$P(A|A_0) = E(I_A|A_0) \tag{2.11.11}$$

een *voorwaardelijke waarschijnlijkheid van* A *gegeven* A_0. Volgens definitie 2.11.1 is $P(A|A_0)$ voor iedere vaste $A \in A$ gedefinieerd als een A_0-meetbare stochastische grootheid waarvoor geldt

$$P(A \cap A_0) = \int_{A_0} P(A|A_0)dP \quad \text{voor iedere } A_0 \in A_0. \tag{2.11.12}$$

Twee versies van $P(A|A_0)$ zijn P-bijna overal gelijk. Voor $A \in A$ is

$$P(A|Y) = E(I_A|Y) = E(I_A|A_Y) \tag{2.11.13}$$

een *voorwaardelijke waarschijnlijkheid van* A *gegeven* Y. Volgens definitie 2.11.2 en (2.11.10) is $P(A|Y)$ voor iedere vaste $A \in A$ gedefinieerd als een reële Borelfunctie g van Y waarvoor geldt

$$P(A \cap Y^{-1}(B)) = \int_{Y^{-1}(B)} P(A|Y)dP = \int_B g(y)dP_Y(y) \tag{2.11.14}$$

voor alle $B \in B^k$. Indien g(Y) een versie van $P(A|Y)$ is, definiëren wij voor alle $y \in R^k$ de *voorwaardelijke waarschijnlijkheid van* A *gegeven* $Y = y$ als $P(A|Y=y) = g(y)$; $P(A|Y=\cdot)$ is dus een reële Borelfunctie op R^k waarvoor geldt

$$P(A \cap Y^{-1}(B)) = \int_B P(A|Y=y)dP_Y(y) \quad \text{voor iedere } B \in B^k. \tag{2.11.15}$$

Zij $V = (V_1,\ldots,V_m)$, $1 \leqq m \leqq \infty$, een tweede stochastische vector op (Ω,A,P). Geheel analoog aan de notatie $P_V(B)$ of $P(V \in B)$ voor de (onvoorwaardelijke) waarschijnlijkheid $P(\{\omega : V(\omega) \in B\}) = P(V^{-1}(B))$ voor $B \in B^m$ (zie §2.5) voeren wij nu overeenkomstige notatie in voor de voorwaardelijke waarschijnlijkheid van de eventualiteit $\{V \in B\}$. Voor iedere $B \in B^m$ is

(2.11.16) $\qquad P_{V|A_0}(B) = P(V\in B|A_0) = P(V^{-1}(B)|A_0)$,

(2.11.17) $\qquad P_{V|Y}(B) = P(V\in B|Y) = P(V^{-1}(B)|Y)$,

(2.11.18) $\qquad P_{V|Y=y}(B) = P(V\in B|Y=y) = P(V^{-1}(B)|Y=y)$.

De relaties die deze drie begrippen definiëren volgen rechtstreeks uit (2.11.12), (2.11.14) en (2.11.15) door hierin A te vervangen door $V^{-1}(B)$, $B \in B^m$. Zo is bijvoorbeeld $P_{V|Y=.}(B)$ voor iedere vaste $B \in B^m$ gedefinieerd als een reële Borelfunctie op R^k waarvoor geldt

(2.11.19) $$P(\{V\in B\}\cap\{Y\in\tilde{B}\}) = \int_{\tilde{B}} P_{V|Y=y}(B)dP_Y(y)$$

voor iedere $\tilde{B} \in B^k$. De functie $P_{V|A_0}$ (respectievelijk $P_{V|Y}$ en $P_{V|Y=y}$) op B^m wordt een *voorwaardelijke kansverdeling van* V *gegeven* A_0 (respectievelijk Y en Y = y) genoemd.

De hierboven ingevoerde terminologie suggereert dat er voor iedere $A \in A$ een zodanige versie van $P(A|A_0)$ zou kunnen worden gekozen dat $P(A|A_0)(\omega)$ voor iedere (of P-bijna alle) vaste $\omega \in \Omega$ als functie van A een waarschijnlijkheid op (Ω,A) zou zijn. Indien dit inderdaad mogelijk is noemt men een dergelijke keuze een *reguliere versie* van de voorwaardelijke waarschijnlijkheid $P(.|A_0)$. Over het al dan niet bestaan van een reguliere versie merken wij het volgende op. Daar $0 \leqq I_A \leqq 1$ is $P(A|A_0) = E(I_A|A_0)$ voor iedere $A \in A$ gedefinieerd en b.z. niet-negatief volgens stelling 2.11.3h. Er is dus voor iedere $A \in A$ een versie waarvoor $P(A|A_0)(\omega) \geqq 0$ voor alle $\omega \in \Omega$. Uit (2.11.12) volgt dat er een versie van $P(\Omega|A_0)$ is waarvoor $P(\Omega|A_0)(\omega) = 1$ voor alle $\omega \in \Omega$. Voorts geldt volgens stelling 2.11.4 voor een disjuncte rij $A_1, A_2, \ldots \in A$

(2.11.20) $$P(\bigcup_{n=1}^{\infty} A_n|A_0) = \sum_{n=1}^{\infty} P(A_n|A_0) \text{ b.z..}$$

De P-nulverzameling van die $\omega \in \Omega$ waarvoor de bovenstaande gelijkheid niet geldt zal echter in het algemeen van de beschouwde rij $A_1, A_2, \ldots$ afhangen. De verzameling van die $\omega \in \Omega$ waarvoor $P(.|A_0)(\omega)$ niet σ-additief is op A, is dus de vereniging van een - mogelijk overaftelbaar - aantal P-nulverzamelingen en dus niet noodzakelijk een P-nulverzameling. Men kan aan de hand van een voorbeeld aantonen dat deze complicatie tot gevolg kan hebben dat er geen reguliere versie van $P(.|A_0)$ bestaat. Tevens kan men echter bewijzen dat er voor een stochastische vector $V = (V_1,\ldots,V_m)$, $1 \leqq m \leqq \infty$, wel steeds een reguliere versie van zijn voorwaardelijke kansverdeling $P_{V|A_0}$ op $(R^m,\mathcal{B}^m)$ bestaat. Wij spreken af dat wij voor een voorwaardelijke kansverdeling van een stochastische vector steeds een reguliere versie zullen kiezen. Voor P-bijna alle ω, c.q. P_Y-bijna alle y, zijn $P_{V|A_0}(B)(\omega)$, $P_{V|Y}(B)(\omega)$ en $P_{V|Y=y}(B)$ als functie van B dus voortaan waarschijnlijkheden op $(R^m,\mathcal{B}^m)$.

In het voorafgaande hebben wij voorwaardelijke verwachtingen en voorwaardelijke waarschijnlijkheden impliciet, d.w.z. als oplossing van een vergelijking gedefinieerd. De twee volgende stellingen geven een methode om in de meest voorkomende gevallen deze functies expliciet te bepalen.

Stelling 2.11.10

Laten $V = (V_1,\ldots,V_m)$ en $Y = (Y_1,\ldots,Y_k)$, $1 \leqq m$, $k < \infty$, twee stochastische vectoren op een kansruimte $(\Omega,\mathcal{A},P)$ voorstellen, μ_1 en μ_2 twee σ-finiete maten op $(R^m,\mathcal{B}^m)$ respectievelijk $(R^k,\mathcal{B}^k)$ en $\mu = \mu_1 \times \mu_2$ hun productmaat op $(R^{m+k},\mathcal{B}^{m+k})$. Veronderstel dat de simultane kansverdeling $P_{V,Y}$ van het paar (V,Y) een kansdichtheid $f_{V,Y}$ ten opzichte van μ bezit. Dan is

$$f_Y(y) = \int_{R^m} f_{V,Y}(v,y)\, d\mu_1(v),\ y \in R^k, \tag{2.11.21}$$

een kansdichtheid van Y met betrekking tot μ_2. Voor alle $v \in R^m$ en $y \in R^k$ definiëren wij

$$f_{V|Y=y}(v) = \begin{cases} \dfrac{f_{V,Y}(v,y)}{f_Y(y)} & \text{als } 0 < f_Y(y) < \infty \\ 0 & \text{anders.} \end{cases} \tag{2.11.22}$$

Dan is

$$P_{V|Y=y}(B) = \int_B f_{V|Y=y}(v)\, d\mu_1(v),\ y \in R^k,\ B \in \mathcal{B}^m, \tag{2.11.23}$$

een reguliere versie van de voorwaardelijke kansverdeling van V gegeven $Y = y$. $f_{V|Y=y}$ wordt een voorwaardelijke dichtheid van V gegeven $Y = y$ ten opzichte van μ_1 genoemd.

Bewijs:
Voor $\tilde{B} \in \mathcal{B}^k$ geldt volgens de stelling van Fubini (stelling 1.6.7)

$$P(Y\in\tilde{B}) = \int_{R^m\times\tilde{B}} f_{V,Y}(v,y)\, d\mu(v,y) =$$

$$= \int_{\tilde{B}} \{\int_{R^m} f_{V,Y}(v,y)\, d\mu_1(v)\}\, d\mu_2(y),$$

zodat (2.11.21) inderdaad een dichtheid van Y ten opzichte van μ_2 is. Zij $C = \{y|y \in R^k, 0 < f_Y(y) < \infty\}$ zodat $C \in \mathcal{B}^k$ en $P_Y(C) = 1$. Door wederom de stelling van Fubini toe te passen vinden wij dat het rechterlid van (2.11.23) voor iedere $B \in \mathcal{B}^m$ een Borelfunctie van y is en dat voor iedere $B \in \mathcal{B}^m$ en $\tilde{B} \in \mathcal{B}^k$

$$\int_{\tilde{B}} \{\int_B f_{V|Y=y}(v)\, d\mu_1(v)\}\, dP_Y(y) =$$

$$= \int_{\tilde{B}C} \{\int_B \frac{f_{V,Y}(v,y)}{f_Y(y)}\, d\mu_1(v)\}\, f_Y(y)\, d\mu_2(y) =$$

$$= \int_{B\times\tilde{B}C} f_{V,Y}(v,y)\, d\mu(v,y) = P(\{V\in B\}\cap\{Y\in\tilde{B}C\}) =$$

$$= P(\{V\in B\}\cap\{Y\in\tilde{B}\}).$$

Volgens (2.11.19) is (2.11.23) dus inderdaad een versie van $P_{V|Y=y}$. Voorts is $f_{V|Y=y}$ niet-negatief terwijl voor iedere $y \in C$

$$\int_{R^m} f_{V|Y=y}(v)\, d\mu_1(v) = \frac{f_Y(y)}{f_Y(y)} = 1,$$

zodat $f_{V|Y=y}$ voor P_Y-bijna alle $y \in R^k$ een kansdichtheid is. (2.11.23) is dus een reguliere versie.

Stelling 2.11.11

Laat $V = (V_1,\dots,V_m)$, $1 \leqq m \leqq \infty$, een stochastische vector op een kansruimte $(\Omega,\mathcal{A},P)$ voorstellen, ϕ een reële Borelfunctie op R^m waarvoor $E|\phi(V)| < \infty$ en $\mathcal{A}_0$ een deel-σ-algebra van $\mathcal{A}$. Dan geldt voor een reguliere versie van $P_{V|\mathcal{A}_0}$

$$(2.11.24)\qquad E(\phi(V)|\mathcal{A}_0) = \int_{R^m} \phi(v)\ dP_{V|\mathcal{A}_0}(v) \text{ b.z..}$$

Merk op dat het rechterlid slechts gedefinieerd is als $P_{V|\mathcal{A}_0}$ een maat is; voor een reguliere versie is dit b.z. het geval.

Bewijs:

Als $\phi = I_B$, $B \in \mathcal{B}^m$, dan is $\phi(V) = I_{V^{-1}(B)}$ zodat volgens (2.11.11) en (2.11.16)

$$E(\phi(V)|\mathcal{A}_0) = E(I_{V^{-1}(B)}|\mathcal{A}_0) = P(V^{-1}(B)|\mathcal{A}_0) = P_{V|\mathcal{A}_0}(B) \text{ b.z.}$$

terwijl

$$\int_{R^m} \phi(v)\ dP_{V|\mathcal{A}_0}(v) = \int_B dP_{V|\mathcal{A}_0} = P_{V|\mathcal{A}_0}(B) \text{ b.z..}$$

Uit stelling 2.1.3g volgt nu dat (2.11.24) juist is indien ϕ een elementaire Borelfunctie is. Voor een willekeurige niet-negatieve Borelfunctie ϕ bewijst men (2.11.24) door ϕ te benaderen door een niet-dalende rij niet-negatieve elementaire Borelfuncties en de monotone convergentie stellingen 1.6.2 en 2.11.4 toe te passen. Voor een willekeurige Borelfunctie ϕ volgt (2.11.24) het bovenstaande toe te passen op ϕ^+ en ϕ^-.

Als $Y = (Y_1,\dots,Y_k)$, $1 \leqq k \leqq \infty$, een tweede stochastische vector op $(\Omega,\mathcal{A},P)$ voorstelt dan volgt uit (2.11.24) dat

$$(2.11.25)\qquad E(\phi(V)|Y=y) = \int_{R^m} \phi(v)\ dP_{V|Y=y}(v)$$

voor P_Y-bijna alle $y \in R^k$. Indien de voorwaarden in stelling 2.11.10 vervuld zijn, kunnen wij dus expliciet een versie van $E(\phi(V)|Y=y)$ aangeven, namelijk

$$(2.11.26) \qquad E(\phi(V)|Y{=}y) = \int_{R^m} \phi(v)\, f_{V|Y=y}(v)\, d\mu_1(v),$$

waarin $f_{V|Y=y}$ wordt gegeven door (2.11.22).

In het begin van deze paragraaf werden in (2.11.3)-(2.11.6) voor een tweetal eenvoudige gevallen rechtstreeks definities van voorwaardelijke kansverdeling en voorwaardelijke verwachting gegeven en de vraag rijst of deze wel in overeenstemming zijn met de daarna gegeven algemene definities. Het antwoord op deze vraag is ja. Uit (2.11.19) volgt voor discreet verdeelde Y dat $P_{V|Y=y}(B)$ voor alle y met $P(Y{=}y) \neq 0$ ondubbelzinnig gedefinieerd is en door (2.11.3) wordt bepaald. Indien de simultane verdeling van (V,Y) continu is, zijn de voorwaarden in stelling 2.10.10 vervuld (μ_1, μ_2 en μ zijn nu Lebesgue maten) zodat de juistheid van (2.11.5) uit deze stelling volgt. Tenslotte worden de voorwaardelijke verwachtingen in (2.11.4) en (2.11.6) gedefinieerd als verwachtingen met betrekking tot voorwaardelijke kansverdelingen hetgeen volgens (2.11.25) en (2.11.26) juist is.

Tot besluit willen wij nog de aandacht vestigen op enige consequenties van (2.11.19) en stelling 2.11.10. Kiezen wij $\tilde{B} = R^k$, dan gaat (2.11.19) over in

$$(2.11.27) \qquad P_V(B) = \int_{R^k} P_{V|Y=y}(B)\, dP_Y(y),$$

een formule die grote overeenkomst vertoont met, en voor discrete Y zelfs een speciaal geval is van (2.2.4). Uit stelling 2.11.10 volgt onder de gestelde voorwaarden dat een dichtheid van P_V met betrekking tot de maat μ_1 gegeven wordt door

$$(2.11.28) \qquad f_V(v) = \int_{R^k} f_{V|Y=y}(v)\, f_Y(y)\, d\mu_2(y).$$

Als V en Y beide discreet zijn is ook (2.11.28) een speciaal geval van (2.2.4). Verder volgt uit stelling 2.11.10 een analogon van (2.2.5), de regel van Bayes:

$$(2.11.29) \qquad f_{Y|V=v}(y) = \frac{f_{V|Y=y}(v)\ f_Y(y)}{\int_{R^k} f_{V|Y=u}(v)\ f_Y(u)\ d\mu_2(u)}$$

voor P_V-bijna alle v en P_Y-bijna alle y. Als V en Y discreet zijn is dit zelfs een speciaal geval van (2.2.5).

In het voorgaande hebben wij (2.11.19) en stelling 2.11.10 gebruikt om de voorwaardelijke verdeling van V gegeven Y = y te definiëren in termen van de simultane verdeling van V en Y. In de practijk gebruikt men (2.11.19), stelling 2.11.10 en de daaruit afgeleide formules (2.11.27)-(2.11.29) vaak in omgekeerde richting om de simultane verdeling verdeling van V en Y, de marginale verdeling van V, of de voorwaardelijke verdeling van Y gegeven V te bepalen als de marginale verdeling van Y en de voorwaardelijke verdeling van V gegeven Y bekend zijn. Deze situatie doet zich met name voor bij de constructie van kansruimten bij experimenten die uit twee of meer afhankelijke deelexperimenten bestaan. Bij wijze van voorbeeld beschouwen wij het volgende experiment: Men kiest aselect een getal y uit het eenheidsinterval en vervolgens voert men n o.o. alternatieven uit, alle met het eerder gekozen getal y als succeskans. Het hierbij behaalde aantal successen noemt men v. Wij interpreteren deze omschrijving in een mathematisch model door te spreken over twee stochastische grootheden, V en Y, zodanig dat Y de homogene verdeling op (0,1) bezit, terwijl de voorwaardelijke verdeling van V gegeven Y = y voor $y \in (0,1)$ de binomiale verdeling met parameters n en y is. Nemen wij voor μ_1 de telmaat op de gehele getallen en voor μ_2 de Lebesgue maat op R^1, dan hebben wij dus in de notatie van stelling 2.11.10

$$f_Y(y) = \begin{cases} 1 & \text{als } y \in (0,1), \\ 0 & \text{anders;} \end{cases}$$

en

$$f_{V|Y=y}(v) = \binom{n}{v} y^v (1-y)^v \text{ voor } v \in \{0,1,\dots,n\},\ y \in (0,1).$$

Met behulp van (2.11.22), (2.11.28) en (2.11.29) vinden wij derhalve

$$f_{V,Y}(v,y) = \binom{n}{v} y^v(1-y)^{n-v} \text{ voor } v \in \{0,1,\ldots,n\},\ y \in (0,1);$$

$$f_V(v) = \binom{n}{v} \int_0^1 y^v(1-y)^{n-v}\, dy = \frac{1}{n+1} \text{ voor } v \in \{0,1,\ldots,n\}$$

$$f_{Y|V=v}(y) = (n+1)\binom{n}{v} y^v(1-y)^{n-v} \text{ voor } v \in \{0,1,\ldots,n\},\ y \in (0,1).$$

De verwachting van V kunnen wij op twee manieren berekenen:

$$E(V) = \int_{R^1} vf_V(v)\, d\mu_1(v) = \sum_{v=0}^{n} \frac{v}{n+1} = \frac{n}{2},$$

en

$$E(V) = E(E(V|Y)) = E(nY) = \frac{n}{2}.$$

Opgaven

1. Bepaal de voorwaardelijke verdeling van X gegeven Y als X en Y twee stochastische grootheden zijn wier simultane verdeling een twee-dimensionale $N(\mu,\Sigma)$-verdeling is met $|\Sigma| > 0$ dan wel $|\Sigma| = 0$.
2. $X_1, X_2, \ldots, X_{n+1}$ zijn o.o. identiek verdeelde stochastische grootheden en $S_k = \sum_{i=1}^{k} X_i$, $k = 1, 2, \ldots, n+1$. Bepaal de voorwaardelijke verdeling van de stochastische vector $(S_1,\ldots,S_n)$ gegeven S_{n+1} als X_i
   a) exponentiëel
   b) geometrisch
   verdeeld is.
3. Bepaal de voorwaardelijke verdeling van X gegeven Z = X+Y als X en Y o.o. stochastische grootheden zijn die
   a) Poisson verdelingen met parameters μ resp. ν
   b) binomiale verdelingen met parameters m en p resp. n en p
   bezitten.
4. Bepaal de voorwaardelijke verdeling van X_1 gegeven $M = \max(X_1,\ldots,X_n)$ als $X_1, \ldots, X_n$ o.o. stochastische grootheden zijn die homogeen verdeeld zijn op (0,1).

5. Voor een stochastische grootheid X waarvoor $EX^2 < \infty$ en een stochastische vector Y geldt

$$\sigma^2(X) = E\sigma^2(X|Y) + \sigma^2(E(X|Y)),$$

waarin $\sigma^2(X|Y)$ de variantie van de voorwaardelijke verdeling van X gegeven Y voorstelt en $\sigma^2(E(X|Y))$ de variantie van $E(X|Y)$. Bewijs dit.

6. Bij een samengesteld experiment neemt men eerst de waarde van een stochastische grootheid Y waar, waarbij Y een gamma verdeling met parameters k en λ bezit (k geheel). Indien Y de waarde y aanneemt verricht men vervolgens een waarneming van een stochastische grootheid met een Poisson verdeling met parameter y. Bepaal de (onvoorwaardelijke) kansverdeling van het resultaat van deze tweede waarneming.

7. Beantwoord opgave 6 als Y $N(\mu,\sigma^2)$ verdeeld is en in tweede instantie een $N(y,\tau^2)$ verdeelde stochastische grootheid wordt waargenomen.

Literatuur

[1] Feller, W., An introduction to probability theory and its applications, Vol. I, Wiley.

[2] Gangolli, R.A., and Ylvisaker, D., Discrete probability, Harcourt, Brace and World, Inc.

[3] Hennequin, P.L., et Tortrat, A., Théorie des probabilités et quelques applications, Masson.

[4] Kingman, J.F.C., and Taylor, S.J., Introduction to measure and probability, Cambridge University Press.

[5] Loève, M., Probability Theory, Van Nostrand.