NOTA 759 september 1973

Instituut voor Cultuurtechniek en Waterhuishouding
Wageningen

OVER DE NUMERIEKE BEREKENING VAN DE KANSVERDELING VAN EENVOUDIGE FUNCTIES VAN TWEE KANSVARIABELEN

ir. Ph.Th. Stol

I N H O U D

1. INLEIDING

In de cultuurtechniek doet zich het probleem voor van de som van stochastische variabelen de kansverdeling te bepalen. Beschouw bijvoorbeeld de neerslag als kansvariabele dan behoort hierbij een kansverdeling. Vervolgens kan van de verdamping eveneens de kansverdeling in beschouwing worden genomen. Nu is de volgende stap de kansverdeling van neerslag min verdamping te bepalen.

Een ander voorbeeld is het optreden van peilen in open water gecombineerd met een oploop door windinvloed. Beide grootheden, namelijk peilen en oploop, zijn stochastische variabelen die samen het uiteindelijke waterpeil bepalen. Ook nu kan gevraagd worden naar de kansverdeling van het som-effect wanneer de kansverdeling van de afzonderlijke termen bekend is.

Veelal zullen de elementaire kansverdelingen slechts empirisch bekend zijn en niet door middel van een functie zijn gegeven. In dat geval moet een methode voor numerieke integratie worden toegepast.

In deze nota zal het gestelde probleem in het kort worden toegelicht. Een algemene behandeling maakt het mogelijk ook uitdrukkingen voor de kansverdeling van andere eenvoudige functies op te stellen.

2. FORMULERING VAN HET PROBLEEM

Stel dat twee stochastische variabelen worden beschouwd bijvoorbeeld $\underline{x}$ en $\underline{y}$ en dat de kansverdelingen achtereenvolgens zijn

$$P(\underline{x} \leq x) = \int_{-\infty}^{x} f(x)\, dx = F(x) \qquad (2.1)$$

$$P(\underline{y} \leq y) = \int_{-\infty}^{y} g(y)\, dy = G(y) \qquad (2.2)$$

Bijvoorbeeld
voor de neerslag $\underline{N}$ van 10-daagse sommen zou kunnen gelden

$$P(\underline{N} \leq 5 \text{ mm}) = p_n \,\%$$

en voor de verdamping $\underline{E}$ van 10-daagse sommen

$$P(\underline{E} \leq 1 \text{ mm}) = p_e \,\%$$

Hoewel het voorgaande uitgelegd kan worden als gebeurtenissen met een herhalingsperiode van gemiddeld 1 x per 10 nieuwe gebeurtenissen, betekent dit niet dat $P(N - E \leq 4 \text{ mm})$ eveneens een herhalingsperiode van gemiddeld 1 x per 10 nieuwe gebeurtenissen heeft. Er zijn namelijk, behalve de hier genoemde, meer combinaties van N en E mogelijk die tot uitkomst een waarde $\leq$ 4 mm hebben.

Het hangt dus af van de kans waarmede $\underline{N}$ en $\underline{E}$ in deze combinaties voorkomen wat de uiteindelijke kans $P(N - E \leq 4 \text{ mm})$ voor waarde zal aannemen.

De algemene probleemstelling is dus, gegeven de uitdrukkingen (2.1) en (2.2) de kansverdeling te bepalen voor $\underline{x} + \underline{y}$.

3. DE KANSVERDELING VAN DE SOM $\underline{z}$ VAN TWEE STOCHASTISCHE VARIABELEN

De gebruikelijke afleiding voor $P(\underline{x} + \underline{y})$ voor continue kansvariabelen verloopt als volgt. Stel

$$z = x + y$$

zodat $$y = z - x$$

zie fig. 1.

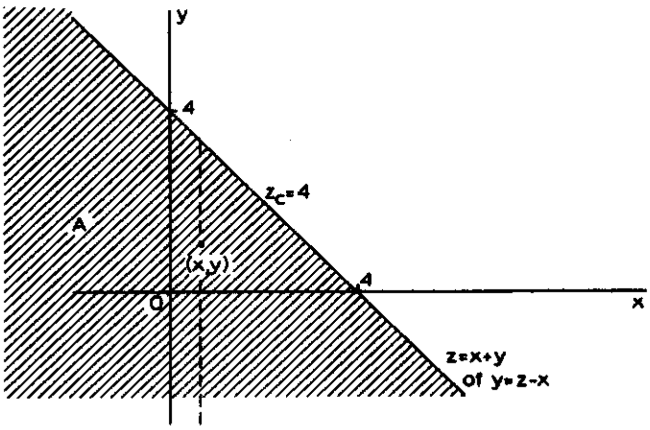


Fig. 1. Voorstelling van de gebeurtenis G_3 dat $(\underline{x}, \underline{y}) \in A$ welke identiek is met de elementaire gebeurtenissen

$$G_1 \stackrel{d}{=} \{ -\infty < \underline{x} < +\infty \} \text{ èn } G_2 \stackrel{d}{=} \{ \underline{y} \leq (4 - x) \}$$

De volgende gedefinieerde gebeurtenissen zijn nu equivalent

$$\left.\begin{array}{l} \underline{G}_1 \stackrel{d}{=} \{ -\infty < \underline{x} < +\infty \} \\ \\ \text{èn} \\ \\ \underline{G}_2 \stackrel{d}{=} \{ \underline{y} \leq (z_c - x) \} \end{array}\right\} \text{ met } \underline{G}_3 \stackrel{d}{=} \{ (\underline{x}, \underline{y}) \in A \}$$

Nu nemen we aan dat $\underline{x}$ en $\underline{y}$ onafhankelijke kansvariabelen zijn zodat voor het optreden van de combinatie van gebeurtenissen de produktregel voor kansen van toepassing is, en dus

$$P(\underline{G}_3) = \iint_A f(x)\ g(y)\ dxdy \qquad (3.1)$$

wat dus tevens de kans is dat $\underline{z} < z_c$.

Integratie van (3.1) kan plaatsvinden door naar y te integreren over het traject $-\infty < y \leq z - x$ bij gefixeerde waarde voor x (zie fig. 1).

Deze integratie loopt dan over het totale waardebereik van x dus over $-\infty < x < +\infty$.

Er komt dan:

$$P(\underline{z} \leq z) = \int_{-\infty}^{+\infty} f(x) \int_{-\infty}^{z-x} g(y)\, dydx$$

of
$$P(\underline{z} \leq z) = \int_{-\infty}^{+\infty} f(x)\, G(z - x)\, dx \overset{d}{=} H(z) \qquad (3.2)$$

waarin $G(z - x) = \int_{-\infty}^{z-x} g(y)\, dy$, weer volgens definitie (2.2), de kansverdeling van $\underline{y}$ voorstelt.

Wordt echter eerst over x geïntegreerd en daarna over het gehele waardenbereik van y, dan wordt verkregen met

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

de uitkomst

$$P(\underline{z} \leq z) = \int_{-\infty}^{+\infty} g(y)\, F(z - y)\, dy \qquad (3.3)$$

welke gelijkwaardig is aan (3.2).

De berekening van de integraal (3.2) respectievelijk (3.3) is voor een aantal verdelingen wel mogelijk; het hangt hierbij sterk af van de gedaante van de functies in de integrand of de integratie slaagt. Is de integratie niet uitvoerbaar, dan moet numeriek worden geïntegreerd. Voor kansverdelingen gebaseerd op empirische frequentieverdelingen is een numerieke integratie, gebaseerd op (3.2) of (3.3) steeds de aangewezen weg om tot een oplossing te komen.

4. DE KANSDICHTHEID VAN $\underline{z}$

Met (3.2) en ook met (3.3) is de kansverdeling van de som van twee kansvariabelen bepaald. De integraties vinden echter plaats over respectievelijk x en y. Ten einde een uitdrukking te vinden waarbij integratie over z plaats vindt, de kansdichtheid dus van z zelf, moet een transformatie worden toegepast die als volgt verloopt.

We definiëren

$$P(\underline{z} \leq z) = \int_{-\infty}^{z} h(z)\, dz = H(z)$$

en gevraagd wordt uit (3.2) de functie h(z) te bepalen.

Uit (3.2) volgt nu, door over te gaan op

$$z = y + x \qquad \text{met} \qquad dz = dy$$

als x constant gehouden wordt en transformatie van de integratiegrenzen wordt toegepast door invullen van de grenzen voor y in z = y + x, dus

$$y\Big]_{-\infty}^{z-x} \rightarrow z\Big]_{-\infty}^{z}$$

dat

$$G(z - x) = \int_{-\infty}^{z-x} g(y)\, dy = \int_{-\infty}^{z} g(z - x)\, dz$$

Dit wordt ingevuld in (3.2) waarna het verwisselen van de integratievolgorde leidt tot

$$H(z) = P(\underline{z} \leq z) = \int_{-\infty}^{z} \int_{-\infty}^{+\infty} f(x)\, g(z - x)\, dx\, dz$$

waaruit weer volgt dat

$$h(z) = \int_{-\infty}^{+\infty} f(x)\ g(z - x)\ dx \tag{4.1}$$

wat de kansdichtheidsfunctie van $\underline{z}$ is.

De verwachtingswaarde van $\underline{z}$ volgt nu uit (4.1) door berekening van

$$E(\underline{z}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} zf(x)\ g(z - x)\ dxdz \stackrel{d}{=} \xi$$

en de variantie uit

$$E(\underline{z} - \xi)^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (z - \xi)^2\ f(x)\ g(z - x)\ dxdz$$

Op analoge wijze kan worden verkregen over y (vergelijk met (4.1))

$$h(z) = \int_{-\infty}^{+\infty} g(y)\ f(z - y)\ dy \tag{4.2}$$

als kansdichtheid met analoge formules voor de verwachtingswaarde en de variantie.

5. DE KANSVERDELING VAN $\underline{z}$ ALS $\underline{x}$ EN $\underline{y}$ BEIDE POSITIEVE VARIABELEN ZIJN

Voor het geval dat het waardenbereik van x en y beperkt is tot positieve waarden zoals veelal met cultuurtechnische grootheden het geval is, wordt een enigszins afwijkend eindresultaat verkregen zoals toegelicht wordt met fig. 2.

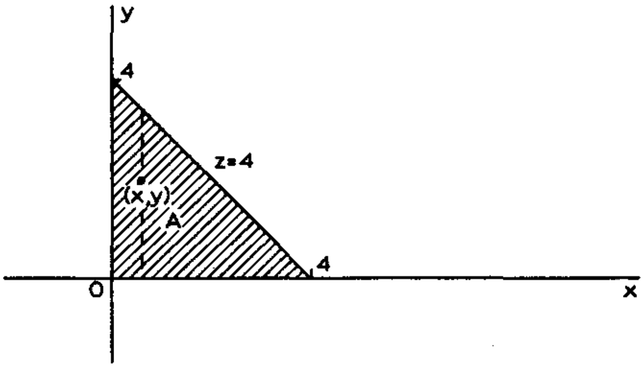


Fig. 2. Voorstelling van de gebeurtenis G_3 dat $(\underline{x}, \underline{y}) \in A$ welke identiek is met de elementaire gebeurtenissen

$\underline{G}_1 \overset{d}{=} \{0 \leq \underline{x} \leq 4\}$ en $\underline{G}_2 \overset{d}{=} \{0 \leq \underline{y} \leq 4\}$

Opgemerkt wordt dat nu de volgende gebeurtenissen equivalent zijn

$$G_1 \overset{d}{=} \{0 \leq \underline{x} \leq 4\}$$

en

$$G_2 \overset{d}{=} \{0 \leq \underline{y} \leq 4\} \qquad \underline{G}_3 \overset{d}{=} (\underline{x}, \underline{y}) \in A$$

Integratie kan nu plaatsvinden naar y over het interval $0 \leq y \leq z - x$. Deze integratie loopt dan vervolgens weer over het totale waardebereik van x in dit geval gelijk aan $0 \leq x \leq z$ (zie fig. 2). Hogere waarden van x kunnen niet voorkomen aangezien $x + y$ beperkt moet blijven tot een waarde z en y positief is.

Hieruit volgt nu dat

$$H(z) = P(\underline{z} \leq z | 0 \leq z) = \int_0^z f(x)\ G(z - x)\ dx \qquad (5.1)$$

en

$$h(z) = \int_0^z f(x)\ g(z - x)\ dx \qquad (5.2)$$

naar anologie van (3.2) en (4.1).

6. TOEPASSING OP EMPIRISCHE VERDELINGEN

Zijn van de kansverdeling van $\underline{x}$ en $\underline{y}$ niet de formules bekend, dan kan tot numerieke integratie worden overgegaan waarbij continue verdelingen omgezet worden in discrete verdelingen. Deze procedure zal toegelicht worden met een tweetal figuren.

In fig. 3 staan twee denkbeeldige kansvariabelen $\underline{x}$ en $\underline{y}$ tegen elkaar uitgezet. Langs de x-as staat het histogram voor x getekend, langs de y-as dat voor y. Het histogram geeft weer dat bijvoorbeeld waarden die in de tweede klasse van de $\underline{x}$-variabele vallen met een kans van 8 % zullen voorkomen.

kans in %	y								
		8.00	24.00	36.00	16.00	8.00		8.00	100.00 %
6	0	48	144	216	96	48	0	48	6.00
15	0	120	360	540	240	120	0	120	15.00
6	0	48	144	216	96	48	0	48	6.00
10	0	80	240	360	160	80	0	80	10.00
20	0	160	480	720	320	160	0	160	20.00
25	0	200	600	900	400	200	0	200	25.00
12	0	96	288	432	192	96	0	96	12.00
6	0	48	144	216	96	48	0	48	6.00
	0	0	0	0	0	0	0	0	
		8	24	36	16	8		8	100.00 %

kans in % op voorkomen in aangegeven klasse

Fig. 3. Kansverdelingen van $\underline{x}$ en $\underline{y}$ gegeven als histogrammen.
Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld, dan kan de produktregel worden toegepast voor het gecombineerd optreden van klassen. Produkten van kansen (x 100) staan in de marges van de tabel vermeld.
De marginale verdelingen sommeren weer tot 100 %.

Waarden die in de tweede klasse van de $\underline{y}$-variabele vallen komen met een kans van 6 % voor.

Nu nemen we aan dat $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld zijn. Dit houdt in dat de gebeurtenis: 'een combinatie ($\underline{x}$, $\underline{y}$) valt in de tweede klasse van x ên in de tweede klasse van y' zal voorkomen met het produkt van de kansen van de klassen afzonderlijk dus gelijk is aan 0.08 x 0.06 = 0.48 %.

In het x, y-vlak is nu dus een 2-dimensionale kansverdeling gedefinieerd door de produkten van kansen.

De marginale verdelingen zijn weer gelijk aan de oorspronkelijke histogrammen zoals algemeen blijkt uit de volgende beschouwing.

Stel $\underline{x}$ is verdeeld in de klassen $F_1, \dots, F_n$ en $\underline{y}$ in de klassen $G_1, \dots, G_m$. We definiëren de kansen als

$$P(\underline{x} \in F_i) = p_i \ , \ i = 1, \dots, n \tag{6.1}$$

$$P(\underline{y} \in G_i) = q_i \ , \ i = 1, \dots, m \tag{6.2}$$

We vonden

$$P\{(\underline{x}, \underline{y}) \in (F_i, G_j)\} = p_i q_j \ , \ \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, m \end{matrix}$$

De marginale verdeling van bijvoorbeeld $\underline{x}$ wordt gevonden door over alle klassen van $\underline{y}$ te sommeren dus

$$P(\underline{x} \in F_i) = \sum_{j=1}^{m} p_i q_j = p_i \sum_{j=1}^{m} q_j = p_i$$

aangezien de som van q_j over alle y-klassen gelijk is aan de eenheid. De uitkomst komt overeen met (6.1) waarmee dus de kansverdeling van $\underline{x}$ is terugverkregen.

De volgende stap is nu x en y in relatie met elkaar te brengen door bijvoorbeeld te beschouwen

$$\underline{z} = \underline{x} + \underline{y} \tag{6.3}$$

Er wordt nu een opteltabel verkregen met behulp van de klassemiddens zoals weergegeven in fig. 4. Bij elk veld in de tabel van fig. 4 is volgens fig. 3 een kans gedefinieerd.

kans	waarde	y								
6	25	26	27	28	29	30	35	40	45	
15	20	21	22	23	24	25	30	35	40	
6	15	16	17	18	19	20	25	30	35	
10	10	11	12	13	14	15	20	25	30	
20	5	6	7	8	9	10	15	20	25	
25	4	5	6	7	8	9	14	19	24	
12	3	4	5	6	7	8	13	18	23	
6	2	3	4	5	6	7	12	17	22	
0	1	2	3	4	5	6	11	16	21	x
		1	2	3	4	5	10	15	20	waarde
		0	8	24	36	16	8	0	8	kans

Fig. 4. Voorbeeld van het gebruik van klassemiddens voor het bepalen van kritieke waarden van $\underline{z} = \underline{x} + \underline{y}$. Ingetekend zijn de grenzen van de kritieke waarden voor respectievelijk $\underline{z} \leq 5$ en $\underline{z} \leq 20$ (volgetrokken). De streeplijn geeft de begrenzing van het gebied $\underline{x} \leq 5 \wedge \underline{y} \leq 5$

De kansverdeling van $\underline{z}$ kan dus worden verkregen door sommatie van kansen uit fig. 3 volgens kritieke waarden voor $\underline{z}$.

Uit fig. 4 volgt dan bijvoorbeeld in verband met fig. 3

$$P(\underline{x} \leq 5) = 84\ \% \qquad \text{(fig. 3)}$$

$$P(\underline{y} \leq 5) = 63\ \% \qquad \text{(fig. 3)}$$

$$P(\underline{x} + \underline{y} \leq 5) = 2.88\ \% \qquad \text{(fig. 4)}$$

maar $$P(\underline{x} \leq 5 \text{ én } \underline{y} \leq 5) = 52.92\ \% \qquad \text{(fig. 3 en 4)}$$

namelijk $$0.84 \times 0.63 = 52.92\ \%$$

weergegeven in fig. 4 met een streeplijn.

Een ander voorbeeld is nog

$$P(\underline{x} + \underline{y} \leq 20) = 72.20\ \% \qquad \text{(fig. 3 en 4)}$$

Dat hiermede een cumulatieve kansverdeling is verkregen volgt uit

$$\sum_{i=1}^{n} \sum_{j=1}^{m} p_i q_j = \sum_{i=1}^{n} p_i \sum_{j=1}^{m} q_j = \sum_{i=1}^{n} p_i = 1$$

7. ANDERE FUNCTIES VAN VARIABELEN

Op de variabelen x en y kunnen ook andere bewerkingen worden toegepast. Een voorbeeld wordt gegeven in fig. 5. waarbij een delingstabel is gegeven dus z = y/x.

kans	waarde	y								
6	25	25	12.5	8.3	6.2	5	2.5	1.7	1.2	
15	20	20	10	6.7	5	4	2.0	1.3	1.0	
6	15	15	7.5	5	3.8	3	1.5	1.0	.8	
10	10	10	5	3.3	2.5	2	1.0	.7	.5	
20	5	5	2.5	1.7	1.2	1	.5	.3	.3	
25	4	4	2	1.3	1	.8	.4	.3	.2	
12	3	3	1.5	1	.8	.6	.3	.2	.2	
6	2	2	1	.7	.5	.4	.2	.1	.1	
0	1	1	.5	.3	.2	.2	.1	.1	.1	x
		1	2	3	4	5	10	15	20	waarde
		0	8	24	36	16	8	0	8	kans

Fig. 5. Voorbeeld van het gebruik van klassemiddens voor het bepalen van kritieke waarden van $\underline{z} = \dfrac{\underline{y}}{\underline{x}}$

Ingetekend zijn de grenzen van de kritieke waarden voor respectievelijk

$\underline{z} \leq 5$ en $\underline{z} \leq 20$

Op deze tabel kan weer hetzelfde kansveld van fig. 3 worden toegepast. Nu echter moet het sommeren van kansen over andere grenzen plaatsvinden zodat een andere kansverdeling wordt verkregen.

Nagegaan kan bijvoorbeeld worden dat

$$P(\frac{\underline{y}}{\underline{x}} \leq 5) = 90.64\ \%$$

$$P(\frac{\underline{y}}{\underline{x}} \leq 20) = 100\ \%$$

Het is van belang op te merken dat dus het k a n s v e l d bepaald wordt door

- de kansverdeling van $\underline{x}$
- de kansverdeling van $\underline{y}$
- de (aangenomen) onafhankelijkheid tussen $\underline{x}$ en $\underline{y}$

De k a n s v e r d e l i n g van $\underline{z} = \phi(\underline{x}, \underline{y})$ wordt bepaald door

- de vorm van de functie ϕ
- het waardenbereik van x en y

8. ALGEMENE AFLEIDING VAN DE KANSVERDELING VAN $\phi(\underline{x}, \underline{y})$

Hieronder wordt puntsgewijs de algemene afleiding voor de kansverdeling van $\underline{z}$ gegeven waarbij gedacht wordt aan de volgende functies

ϕ_1: $z = x + y$
ϕ_2: $z = x - y$
ϕ_3: $z = xy$
ϕ_4: $z = x/y$

Verwezen wordt naar KENDALL and STUART (1958 pp. 263, 265), HUTTINGTON (1939) en STAM (1964, pp. 251 e.v.) die ook gevallen behandelt waarin een van beide variabelen niet continu is.

1. De kansverdelingen van $\underline{x}$ en $\underline{y}$ worden per definitie gegeven door

$$P(\underline{x}) = \int_{x_o}^{x_b} f(x)\, dx = 1 \tag{8.1}$$

$$P(\underline{y}) = \int_{y_o}^{y_b} g(y)\, dy = 1 \tag{8.2}$$

waarin integratie plaats vindt over het gehele waardenbereik van x namelijk van $x_{ondergrens}$ tot $x_{bovengrens}$ (x_o, x_b) en voor y analoog.

2. Aangenomen wordt dat er een simultane verdeling w bestaat zodat

$$P(\underline{x}, \underline{y}) = \int_{x_o}^{x_b} \int_{y_o}^{y_b} w(x, y)\, dydx = 1 \tag{8.3}$$

3. Onder de veronderstelling dat $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld zijn volgt uit (8.3)

$$P(\underline{x}, \underline{y}) = \int_{x_o}^{x_b} f(x)\, dx \int_{y_o}^{y_b} g(y)\, dy = 1$$

of

$$P(\underline{x}, \underline{y}) = \int_{x_o}^{x_b} \int_{y_o}^{y_b} f(x)\, g(y)\, dydx = 1 \tag{8.4}$$

Aangenomen zal steeds worden dat de integratievolgorde mag worden verwisseld. Verder zal in het volgende eerste integratie plaatsvinden over x bij constante y, daarna over y.

De kansverdeling (8.4) geeft aanleiding tot de volgende opmerkingen.

3a De kansverdeling van x is de marginale verdeling van $P(\underline{x}, \underline{y})$ indien y alle waarden aanneemt tussen y_o en y_b dus, voor de integrand van x,

$$\int_{y_o}^{y_b} f(x)\, g(y)\, dy = f(x) \int_{y_o}^{y_b} g(y)\, dy = f(x)$$

aangezien $\int_{y_o}^{y_b} g(y)\, dy = 1$ volgens (8.2).

3b Evenzo kan de marginale verdeling van y gevonden worden door alleen over x te integreren.

3c Het probleem de verdeling van $\underline{z} = \phi(\underline{x}, \underline{y})$ te vinden kan opgelost worden door over de relevante combinatie van integratiegrenzen te integreren over een deelgebied met bovengrens $\phi(\underline{x}, \underline{y})$

Opgemerkt wordt dat in deze opvatting 3a het voorbeeld geeft van de functie $\underline{z} = \phi(\underline{x}, \underline{y}) = \underline{x}$, voor alle y.

4. Gevraagd wordt de kansverdeling van $\underline{z}$ te bepalen als $\underline{z} = \phi(\underline{x}, \underline{y})$. De volgende transformaties worden ingevoerd:

$$\begin{array}{lcl} z = \phi(x, y) & & x = \Psi(z, y) \\ & \text{met inverse functies} & \\ y = y & & y = y \end{array}$$

De Jacobiaan van deze transformatie luidt

$$J = \begin{vmatrix} \dfrac{\partial \Psi}{\partial z} & \dfrac{\partial \Psi}{\partial y} \\ & \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial y} \end{vmatrix} = \frac{\partial \Psi}{\partial z} \text{ of ook } \frac{\partial x}{\partial z}$$

waarin $\partial x/\partial z$ een functie is van z en y.

Een oppervlakte elementje dydx gaat dus over in $\frac{\partial x}{\partial z}$ dydz.

5. Wordt de relatie ϕ ingevoerd in (8.4) dan wordt verkregen de integraal over het gevraagde deelgebied, namelijk

$$P(\underline{x},\ \underline{y} \mid \phi(\underline{x},\ \underline{y}) < z) = \int_{y_o}^{y_b} \int_{x_o}^{\Psi(z,y)} f(x)\ g(y)\ dxdy \qquad (8.5)$$

wat gelijk is aan

$$P(\underline{z} \leq z \mid z \geq z_o\}$$

Zie hiervoor fig. 6. De voorwaarde dat $z \geq z_{ondergrens}$ zal in het volgende als vanzelfsprekend worden aangenomen.
Met (8.5) is dus gevonden dat

$$P(\underline{z} \leq z) = \int_{y_o}^{y_b} F\{\Psi(z,\ y)\}\ g(y)\ dy \qquad (8.6)$$

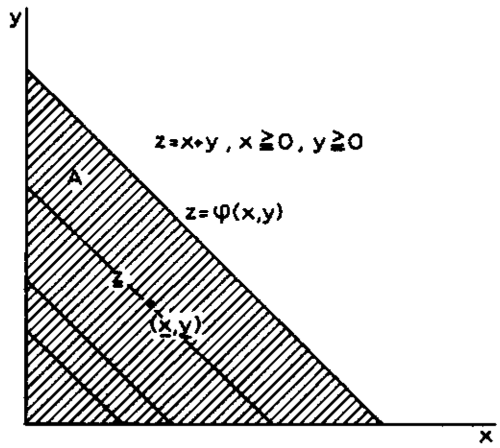


Fig. 6. Voorstelling van de identieke gebeurtenissen

G_1: $(\underline{x},\ \underline{y}) \in A$

G_2: $\underline{z} < z$

6. De uitkomst (8.6) kan herleid worden tot een integratie over z (fig. 6) waarvoor het noodzakelijk is de transformatie uit 4 toe te passen op (8.5).

Er volgt dan, overgaand van (y, x) op(y, z)

$$\int_{y_o}^{y_b} \int_{x_o}^{\Psi(z,y)} f(x)\, g(y)\, dxdy = \int_{y_o}^{y_b} \int_{z_o}^{z} f\{\Psi(z,\, y)\}\, g(y)\, \frac{\partial x}{\partial z}\, dzdy$$

verwisseling van integratie volgorde geeft dan

$$H(z) = P(\underline{z} \leq z) = \int_{z_o}^{z} \int_{y_o}^{y_b} f\{\Psi(z,\, y)\}\, g(y)\, \frac{\partial x}{\partial z}\, dydz \qquad (8.7)$$

Opgemerkt wordt dat de integratie (8.7) slechts kan worden uitgevoerd als $J \neq 0$. Hiermede moet met het vaststellen van de integratiegrenzen rekening worden gehouden. Deze moeten zo gekozen worden dat in het open interval $(z_o,\, z)$ geldt $J > 0$ (KENDALL and STUART).

De kansdichtheid van z is gelijk aan de integrand van dz en luidt dus

$$h(z) = \int_{y_o}^{y_b} f\{\Psi(z,\, y)\}\, g(y)\, \frac{\partial x}{\partial z}\, dy \qquad (8.8)$$

9. HET GEVAL DAT $\partial x/\partial z = \Psi'(y)$

Is $\frac{\partial x}{\partial z}$ alléén een functie van y, wat veelal het geval is, dan kan de algemene vorm nog een stap verder worden uitgewerkt door toepassing van de definitie

$$F(x) = \int_{x_o}^{x} f(x)\, dx \qquad (9.1)$$

en wel als volgt

Stel $\frac{\partial x}{\partial z} = \Psi'(y)$ zodat (8.8) luidt

$$h(z) = \int_{y_o}^{y_b} f\{\Psi(z, y)\}\ g(y)\ \Psi'(y)\ dy \qquad (9.2)$$

De kansverdeling van $\underline{z}$ is dus

$$H(z) = \int_{z_o}^{z} \int_{y_o}^{y_b} f\{\Psi(z, y)\}\ g(y)\ \Psi'(y)\ dydz$$

Verwisseling van integratie volgorde geeft

$$H(z) = \int_{y_o}^{y_b} \int_{z_o}^{z} f\{\Psi(z, y)\}\ dz\ \ g(y)\ \Psi'(y)\ dy$$

wat in verband met (9.1) wordt

$$H(z) = \int_{y_o}^{y_b} F\{\Psi(z, y)\}\ g(y)\ \Psi'(y)\ dy \qquad (9.3)$$

welke functie dan een bijzonder geval is van (8.7).

10. SAMENVATTING RESULTATEN

De algemene formules (8.7) en (8.8) respectievelijk (9.2) en (9.3) kunnen toegepast worden op de functies genoemd in de inleiding van par. 8. Van belang is steeds de grenzen waarover geïntegreerd moet worden zorgvuldig na te gaan.

10.1. De som-functie

$$z = x + y \quad \text{zodat} \quad \Psi(z, y) = z - y \quad \text{en} \quad \Psi'(y) = 1$$

Indien $(-\infty < x < +\infty)$ en $(-\infty < y < +\infty)$
dan wordt geïntegreerd tussen respectievelijk

$$(-\infty < x < z - y) \text{ en } (-\infty < y < +\infty)$$

zodat

$$h(z) = \int_{-\infty}^{+\infty} f(z - y)\, g(y)\, dy$$

en

$$H(z) = \int_{-\infty}^{+\infty} F(z - y)\, g(y)\, dy$$

Indien $(0 \leq x < +\infty)$ en $(0 \leq y < +\infty)$ dan wordt geïntegreerd tussen respectievelijk

$$(0 \leq x < z - y) \text{ en } (0 \leq y \leq z)$$

zodat

$$h(z) = \int_{0}^{z} f(z - y)\, g(y)\, dy$$

en

$$h(z) = \int_{0}^{z} F(z - y)\, g(y)\, dy$$

10.2. De verschil-functie

$$z = x - y \quad \text{zodat} \quad \Psi(z, y) = z - y \quad \text{en } \Psi'(y) = 1$$

Indien $(-\infty < x < +\infty)$ en $(-\infty < y < +\infty)$
dan wordt geïntegreerd tussen

$$(-\infty < x < z + y) \text{ en } (-\infty < y < +\infty)$$

zodat

$$h(z) = \int_{-\infty}^{+\infty} f(z + y)\, g(y)\, dy$$

en

$$H(z) = \int_{-\infty}^{+\infty} F(z + y)\, g(y)\, dy$$

10.3. De produkt-functie

$$z = xy \quad \text{zodat } \Psi(z, y) = \frac{z}{y} \quad \text{en} \quad \Psi'(y) = \frac{1}{y}$$

Indien $(0 \leq x < +\infty)$ en $(0 \leq y < +\infty)$
dan wordt geïntegreerd tussen

$$(0 \leq x < \frac{z}{y}) \text{ en } (0 \leq y < +\infty)$$

zodat

$$h(z) = \int_{0}^{\infty} f(\frac{z}{y})\, g(y)\, \frac{1}{y}\, dy$$

en

$$H(z) = \int_{0}^{\infty} F(\frac{z}{y})\, g(y)\, \frac{1}{y}\, dy$$

10.4. D e   q u o t i ë n t - f u n c t i e

$z = \frac{x}{y}$ zodat $\Psi(z, y) = zy$ en $\Psi'(y) = y$

Indien $(-\infty < x < +\infty)$ en $(0 \leq y < +\infty)$ dan wordt geïntegreerd tussen

$$(-\infty < x < zy) \quad \text{en} \quad (0 \leq y < +\infty)$$

zodat

$$h(z) = \int_0^\infty f(zy)\, g(y)\, ydy$$

en

$$H(z) = \int_0^\infty F(zy)\, g(y)\, ydy$$

Een grafische voorstelling van de integratiegebieden wordt gegeven in fig. 7, welke verder geen toelichting meer behoeft.

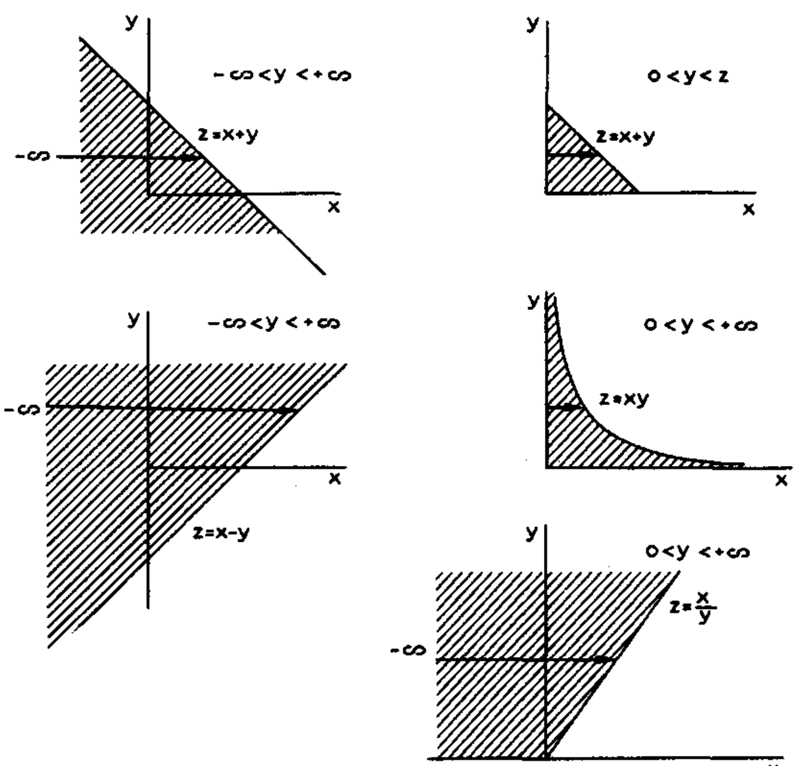


Fig. 7. Voorbeelden van gebieden waarover het kansveld voor $\underline{x}$ en $\underline{y}$ geïntegreerd moet worden voor het bepalen van de kansverdeling van $\underline{z} = \phi(\underline{x}, \underline{y})$. De pijl geeft het traject van eerste integratie aan bij constante waarde van y

11. DE HERHALINGSPERIODE VAN GEBEURTENISSEN $\underline{z} > z_c$

Het aantal waarnemingen T, zó groot dat gemiddeld in zo'n reeks de gebeurtenis $\{\underline{z} > z_c\}$ 1 x voorkomt, wordt gegeven door

$$T(z_c) = \frac{1}{1 - H(z_c)} \tag{11.1}$$

Hierin stelt z_c de kritieke waarde voor waarvan, volgens (9.3), de onderschrijdingskans is

$$H(z_c) = \int_{y_o}^{y_b} F\{\Psi(z_c, y)\}\, g(y)\, \Psi'(y)\, dy \tag{11.2}$$

Opgemerkt wordt nu dat ook als algemene formule geldt

$$T(x_c) = \frac{1}{1 - F(x_c)} \tag{11.3}$$

en

$$T(y_c) = \frac{1}{1 - G(y_c)} \tag{11.4}$$

met de definities analoog aan die voor z.

Wil men de herhalingsperiode $T(z_c)$ uit $T(x_c)$ en $T(y_c)$ berekenen dan dient eerst (11.2) opgelost te worden met de transformaties (11.3) en (11.4). Dit levert geen eenvoudig toepasbare uitdrukking op.

Slechts onder de volgende voorwaarden is een dergelijke oplossing eenvoudig mogelijk, namelijk als

- $F\{\Psi(z_c, y)\}$ geen functie is van y
- $\Psi'(y)$ constant

zodat $\int_{y_o}^{y_b} g(y)\, \Psi'(y)\, dy = 1$

Dit wordt bereikt met

$$\frac{\partial \Psi}{\partial z} = \Psi'(y) = C_1$$

en dus

$$\Psi(y) = C_1 z + C_2$$

zodat de transformatie luidt

$$z = \frac{x - C_2}{C_1}$$

waaruit voor (11.2) dan weer volgt

$$H(z_c) = F(C_1 z_c + C_2)$$

zodat rechtstreekse herleiding alleen mogelijk is onder lineaire transformatie van de stochastische variabele, waarvoor de oplossing triviaal is.

12. OPLOSSING DOOR NUMERIEKE INTEGRATIE

In een computerprogramma waarin de kansverdeling van $\underline{z} = \underline{x} + \underline{y}$ numeriek wordt bepaald wordt ervan uitgegaan dat de integratie onder- en bovengrenzen van x en y eindige waarden zijn. Dit betekent dat steeds het schema van fig. 2 van toepassing is; voor niet-continue variabelen - waartoe het probleem voor de numerieke integratie wordt getransformeerd - geeft dan fig. 4 het grondprincipe van de berekening.

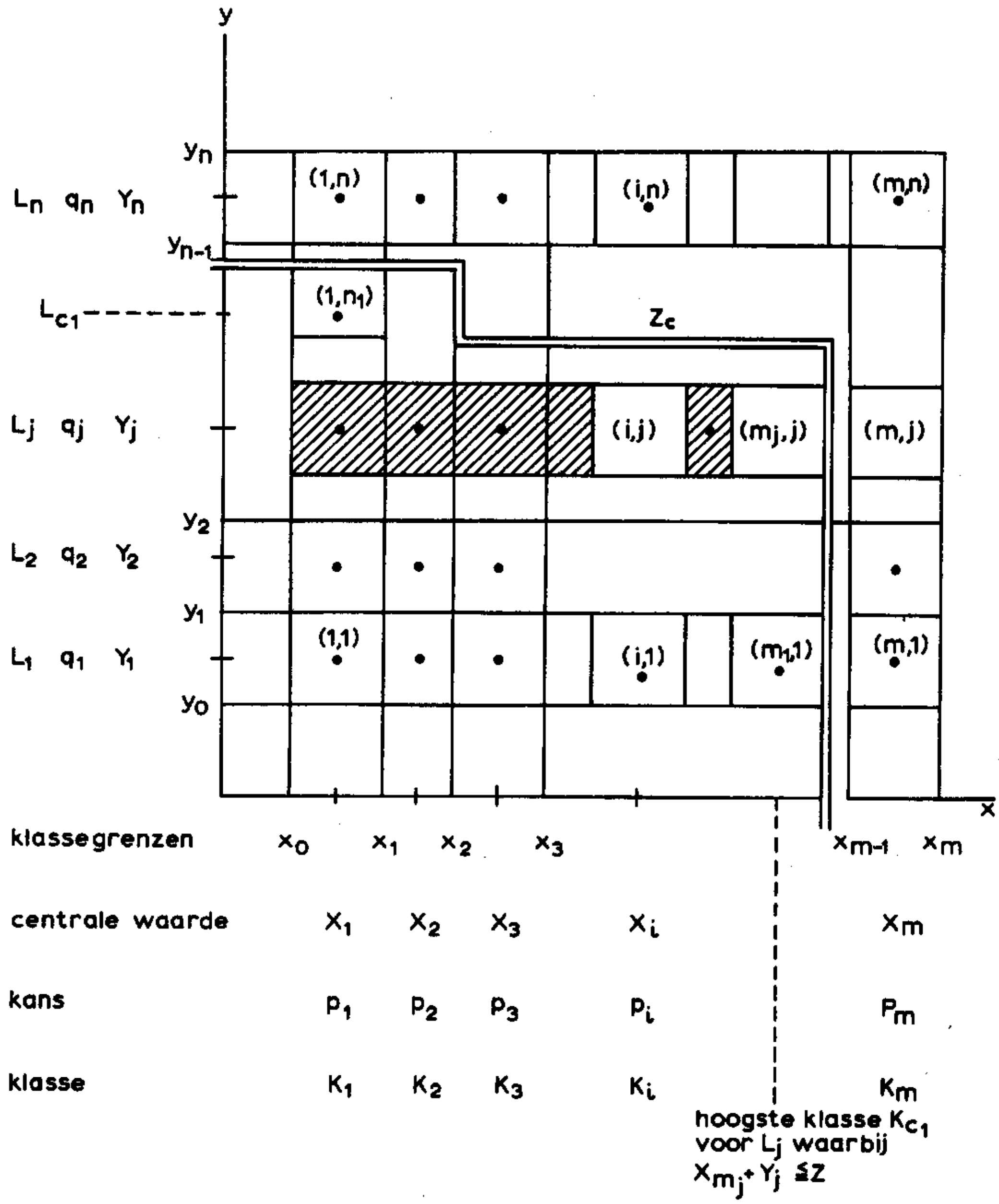
Als gegevens is vereist de empirische verdeling van $\underline{x}$ en $\underline{y}$ dus bekend verondersteld worden

$$P(x_{i-1} < \underline{x} \leq x_i) = p_i \ , \ i = 1, \ldots , m$$

en

$$P(y_{j-1} < \underline{x} \leq y_j) = q_j \ , \ j = 1, \ldots , n$$

Bij elke klasse worden nu centrale waarden gedefinieerd (fig. 8) volgens


y
x
y_n
L_n q_n Y_n
y_{n-1}
L_{c1}
L_j q_j Y_j
y_2
L_2 q_2 Y_2
y_1
L_1 q_1 Y_1
y_0
(1,n)
(i,n)
(m,n)
$(1,n_1)$
Z_c
(i,j)
(m_j,j)
(m,j)
(1,1)
(i,1)
$(m_1,1)$
(m,1)
klassegrenzen x_0 x_1 x_2 x_3 x_{m-1} x_m
centrale waarde X_1 X_2 X_3 X_i X_m
kans p_1 p_2 p_3 p_i p_m
klasse K_1 K_2 K_3 K_i K_m
hoogste klasse K_{c_1}
voor L_j waarbij
$X_{m_j} + Y_j \leq Z$

$$X_i = \frac{1}{2}(x_{i-1} + x_i)$$

$$Y_j = \frac{1}{2}(y_{j-1} + y_j)$$

en een waarde van x binnen een klasse i wordt gelijkgesteld aan de centrale waarde. In formule:

indien $\quad x_{i-1} < x \leq x_i \quad$ dan $\quad x = X_i$

verder wordt dan gezegd dat in dit geval x een element is van de klasse K_i zodat

$$x \in K_i$$

en voor y analoog.

Feitelijk wordt hiermede via de klasse-indeling, waarvoor geldt dat

$$P(x \in K_i) = p_i,$$

het histogram vervangen door de discrete verdeling

$$P(x = X_i) = p_i$$

De verdere bewerking bestaat nu uit de volgende stappen

1. De kritieke waarden die z kan aannemen moeten voldoen aan de volgende voorwaarde

$$(X_1 + Y_1) \leq z_c \leq (X_m + Y_n)$$

   waarop in een computerprogramma kan worden getoetst.
   Voor een bepaalde kritieke waarde z_c worden nu de verdere bewerkingen uitgevoerd:

2. Bepaald wordt wat de hoogste klasse L_{c_1} van y is waarvoor Y_j binnen z_c valt als $x = X_1$. Deze wordt als volgt gevonden

$$y_c = z_c - X_1$$

Bepaal de waarde n_1 waarvoor

$$y_{n_1 - 1} < y_c \leq y_{n_1}$$

zodat y_c dus een element is van klasse L_{n_1}.

Het kan echter voorkomen dat $z_c - X_1 > y_n$ in welk geval de hoogste klasse van y in de bewerking moet worden opgenomen. Samengevat: bepaal

$$y_c = z_c - X_1$$

indien $\quad y_n \leq y_c \quad$ dan $\quad L_{c_1} = L_n$

indien $\quad y_{n_1 - 1} < y_c \leq y_{n_1} \quad$ dan $\quad L_{c_1} = L_{n_1}$

zodat c_1 de kleinste waarde van de getallen n_1 en n aanneemt.

3. Voor een constante waarde van y wordt gesommeerd over alle relevante waarden van x. De constante waarden van y zijn volgens 2 de volgende, met Y_{c_1} de centrale waarde van de klasse L_{c_1}

$$Y_1, \ldots, Y_j, \ldots, Y_{c_1}$$

met kansen $\quad q_1, \ldots, q_j, \ldots, q_{c_1}$

We beschouwen klasse L_j en bepalen de waarden van x waarover de kansen gesommeerd moeten worden. Allereerst wordt de hoogste klasse K_{c_j} bepaald die binnen z_c ligt. Deze wordt als volgt gevonden bepaal

$$x_c = z_c - Y_j$$

indien $\quad x_m \leq x_c \quad$ dan $\quad K_{c_j} = K_m$

indien $\quad x_{m_j - 1} < x_c \leq x_{m_j} \quad$ dan $\quad K_{c_j} = K_{m_j}$

zodat nu c_j de kleinste is van de getallen m_j en m.

4. Het resultaat kan nu dus in de volgende formules worden samengevat:

$$P(\underline{x} + \underline{y} \leq z_c) = \sum_{j=1}^{c_1} \left\{ q_j \sum_{i=1}^{c_j} p_i \right\} \qquad (12.1)$$

Hierin is $\qquad c_1 = \text{Min}(n_1 \ , n)$

en $\qquad c_j = \text{Min}(m_j \ , m)$

Zie ook fig. 8.

5. Ter controle nemen we

$$z_c = X_m + Y_n$$

waarmede aan 1 voldaan is.

Voor 2 wordt verkregen

$$y_c = X_m + Y_n - X_1 > y_n$$

zodat $\qquad L_{c_1} = L_n \quad$ en dus $\quad z_c \in L_n$

Voor 3 wordt verkregen, voor de hoogste y-klasse

$$x_c = X_m + Y_n - Y_n = X_m$$

Aangezien

$$x_{m-1} < X_m \leq x_m$$

is $\qquad K_{m_n} = K_m \quad$ zodat ook $z_c \in K_m$

zodat de sommatie wordt

$$P(\underline{z} \leq z_c) = \sum_{j=1}^{n} \left\{ q_j \sum_{i=1}^{m} p_i \right\} = \sum_{j=1}^{n} q_j = 1$$

Aangezien namelijk in de hoogste y-klasse over alle x-klassen moet worden gesommeerd, geldt dit ook voor alle lagere y-klassen. Het totaal aan kansen sommeert, zoals vereist voor dit geval, tot de eenheid.

Tenslotte

$$z_c = X_1 + Y_1$$

zodat $\quad y_c = X_1 + Y_1 - X_1 = Y_1 \quad$ zodat $\quad L_{c_1} = L_1$

en dus $\quad z_c \in L_1$

en $\quad x_c = X_1 + Y_1 - Y_1 = X_1 \quad$ zodat $\quad K_{c_1} = K_1$

Eveneens geldt $z_c \in K_1$, waaruit volgt

$$P(\underline{z} \leq z_c) = \sum_{j=1}^{1} \left\{ q_j \sum_{i=1}^{1} p_i \right\} = p_1 q_1$$

13. DE CUMULATIEVE VERDELING VAN $\underline{z}$ UIT NUMERIEKE INTEGRATIE

De kansverdeling van $\underline{z}$ zoals deze uit numerieke integratie wordt verkregen is een stapfunctie. Door uit te gaan van discrete waarden van x en y zal z eveneens een discrete variabele zijn. Dit betekent dat de kans op een bepaalde waarde groter is dan 0. Het is nu interessant bij een gegeven kritieke waarde van z het interval te bepalen waarbinnen de kans gelijk blijft.

Stel dat de kritieke waarde gelijk is aan z_c. Dan volgt uit (12.1)

$$P(\underline{z} \leq z_c) = p_c$$

waarin p_c de met (12.1) berekende kans voorstelt.

Gevraagd wordt nu z_o en z_b zo te bepalen dat

$$P(\underline{z} \leq z_o) = P(\underline{z} < z_b) = p_c$$

of

$$P(\underline{z} \leq z_c \mid z_o \leq z_c < z_b) = p_c$$

Uit een beschouwing van fig. 8 volgt dat de oplossing wordt verkregen door voor elke laatste x-klasse de waarde x + y te noteren en tenslotte hieruit de hoogste te kiezen. Deze is dan de gevraagde waarde voor z_o.

Ook blijkt dat indien van elke op de hoogste volgende x-klasse de waarde x + y genoteerd wordt dat dan tenslotte de kleinste waarde de gevraagde z_b is.

De oplossing kan gegeven worden in de volgende algemene vorm voor de ondergrens

$$z_o = \underset{j=1, \ldots, c_1}{\mathrm{Max}} \{X_k + Y_j\}$$

waarin $c_1 = \mathrm{Min}(n_1 , n)$

en $k = \mathrm{Min}(m_j , m)$

en evenzo voor de bovengrens

$$z_b = \underset{j=1, \ldots, c_1}{\mathrm{Min}} \{X_k + Y_j\}$$

waarin $c_1 = \mathrm{Min}(n_1 , n)$

en $k = \mathrm{Min}(m_j + 1, m)$

De cumulatieve verdeling van $\underline{z}$ kan verkregen worden door te beginnen bij $z_c = X_1 + Y_1$ en hierbij volgens de bovenomschreven procedure z_b vast te stellen. Vervolgens kan als volgende kritieke waarde z_b gekozen worden enz. Om te vermijden dat hiermede een te groot aantal waarden van z moet worden doorgerekend kan eventueel het interval $(X_1 + Y_1 , X_m + Y_n)$ in bijvoorbeeld 10 delen worden opgedeeld. Bij elke zo verkregen waarde van z_c kunnen de bijbehorende z_o en z_b uitgerekend worden.

Hiermede is het dan mogelijk de waarden van z_c te vervangen door ronde getallen met dezelfde onderschrijdingskans.

14. HERLEIDING TOT ANDERE FUNCTIES

In de vorige paragrafen is ook weer het verschil tussen de berekening van het kansveld en het vaststellen van de integratiegrenzen duidelijk naar voren gekomen. Dit betekent dat een rekenprocedure voor een som-functie herleid kan worden tot een procedure voor een andere functie door de bepaling van de integratiegrenzen volgens 2e en 3e uit par. 12 aan de nieuwe functie aan te passen.

De onderdelen van fig. 7 kunnen hiertoe als handleiding dienst doen.

LITERATUUR

KENDALL, M.G. and A. STUART (1958). The advanced theory of statistics vol. 1. Distribution theory Griffin, London.

HUTTINGTON, E.V., 1939. Frequency distribution of product and quotient. The annuals of mathemathical statistics Volume X, pp. 195-198.

STAM, A.J., 1964. Inleiding tot de waarschijnlijkheidsrekening. De Technische Uitgeverij H. Stam, Haarlem.