STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING
LEIDING: PROF. DR D. VAN DANTZIG
ADVISEUR VOOR STATISTISCHE CONSULTATIE: PROF. DR J. HEMELRIJK

Rapport S 255 (M 81)

Het berekenen van kansen bij twee- en drie-dimensionale normale verdelingen met tabellen van D.B. OWEN en G.P. STECK.

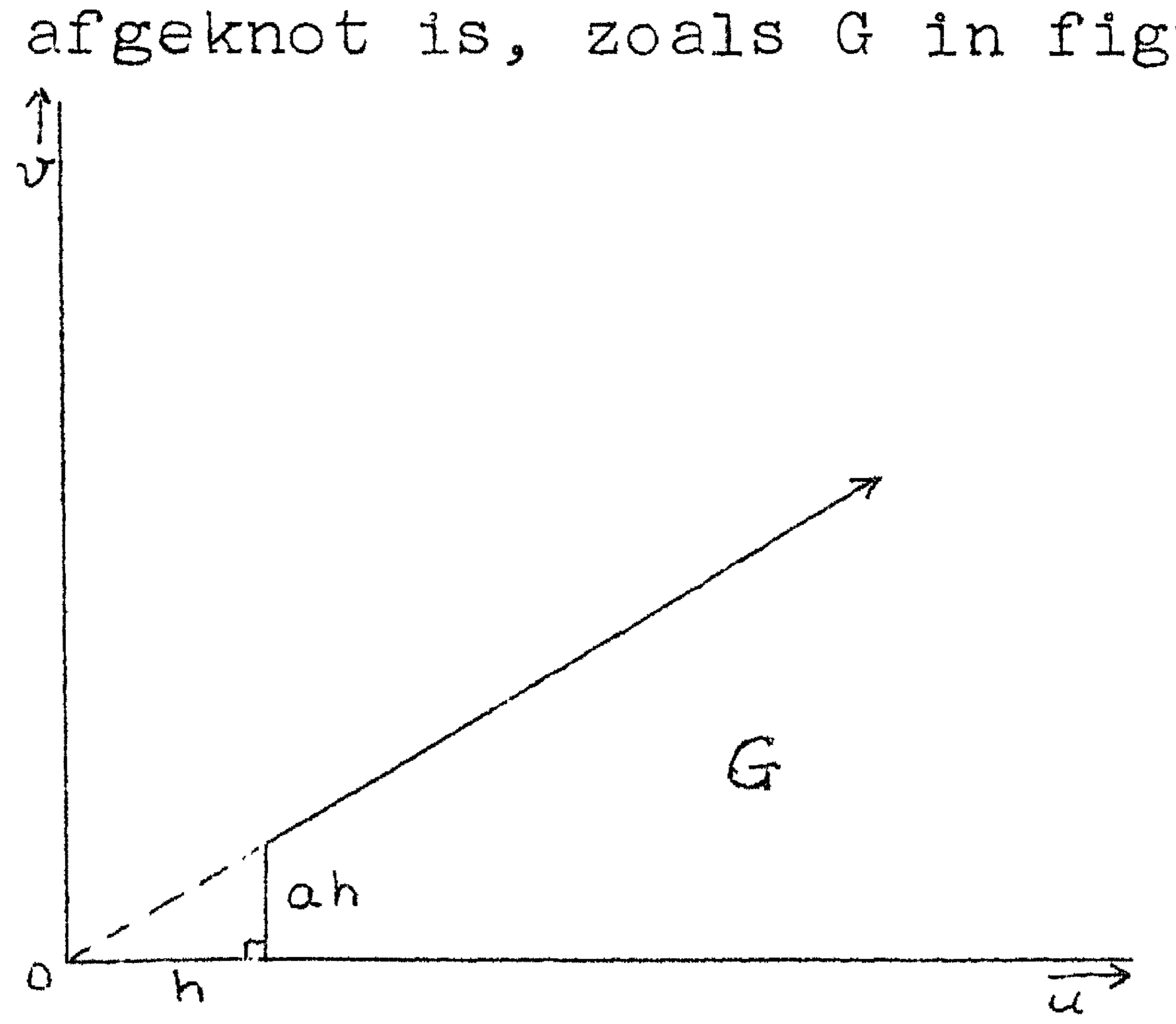
1959

Het berekenen van kansen bij twee- en drie-dimensionale normale verdelingen met tabellen van D.B. OWEN en G.P. STECK.

1. Inleiding. Bezitten $\underline{x}$ en $\underline{y}$ een twee-dimensionale normale verdeling met bekende parameters: $N(\mu_x; \mu_y; \sigma_x; \sigma_y; \rho)$ dan kan men met de tabellen van OWEN (1956) voor iedere veelhoek in het (x,y)-vlak de kans bepalen dat een waarnemingspaar $(\underline{x},\underline{y})$ in deze veelhoek zal liggen. In het algemene geval moet hiervoor eerst een transformatie naar onafhankelijke N(0;1) verdeelde grootheden worden uitgevoerd waarvoor in paragraaf 6 de formules te vinden zijn. In de paragrafen 2, 3 en 4 wordt de getabelleerde functie T(h,a) en de interpolatie in de tabellen besproken. In paragraaf 5 komt een speciaal geval nl. een rechthoek in het (x,y)-vlak ter sprake. Paragraaf 7 geeft een voorbeeld.

   Voor drie-dimensionale verdelingen geeft STECK (1958) formules en een tabel ter berekening van de kans $P\left[\underline{x} \leq h;\ \underline{y} \leq k;\ \underline{z} \leq m\right]$ (zie paragraaf 8).

2. De functie T(h,a). Zijn $\underline{u}$ en $\underline{v}$ stochastisch onafhankelijk en N(0;1) verdeeld dan geeft de getabelleerde functie T(h,a) de kans op een waarneming voor $(\underline{u},\underline{v})$ in een afgeknotte sector van het (u;v)-vlak, waarbij loodrecht op één van de zijden afgeknot is, zoals G in figuur 1.



$\underline{u}$ en $\underline{v}$ ond. onafh. N(0;1)

$P\left[(\underline{u},\underline{v}) \in G\right] = T(h,a)$

fig. 1

Daar de verdeling van $\underline{u}$ en $\underline{v}$ rotatie symmetrisch om O is behoort bij een gebied G', dat door draaiing om O uit G ontstaat, dezelfde kans T(h,a). De waarden van h en a zijn hier steeds positief te nemen. Enkele bijzondere waarden van $T(h,a)$ zijn:

$$(2.1)\qquad \begin{aligned} T(h,o) &= 0, \\ T(o,a) &= \frac{1}{2\pi} \text{ arctg } a \text{ (ook in de tabel op te zoeken)}, \\ T(h,1) &= \tfrac{1}{2}\, G(h) \left[1 - G(h)\right], \\ T(h,\infty) &= \tfrac{1}{2} \left[1 - G(h)\right], \quad h > o \end{aligned}$$

waarin $G(h)$ de verdelingsfunctie van de $N(0;1)$ verdeling voorstelt:

$$G(h) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{h} e^{-\frac{1}{2}t^2}\, dt.$$

Tenslotte geldt voor $a > 1$ de reductieformule:

$$(2.2)\qquad T(h,a) = \tfrac{1}{2}\left[G(h) + G(a\,h)\right] - G(h)\,G(a\,h) - T(a\,h, \tfrac{1}{a}),$$

zodat het voldoende was $T(h,a)$ voor $o < a \leq 1$ te tabelleren (OWEN, 1956).

3. Een willekeurige veelhoek in het (u,v)-vlak. Stel $\underline{u}$ en $\underline{v}$ zijn weer onderling onafhankelijk en $N(0;1)$ verdeeld. De kans op een waarnemingspaar $(\underline{u},\underline{v})$ in een willekeurig afgeknotte sector van het (u,v)-vlak is nu de som of het verschil van twee kansen $T(h,a)$; zie de figuren 2a en 2b.

$\underline{u}$ en $\underline{v}$ ond. onafh. $N(0;1)$

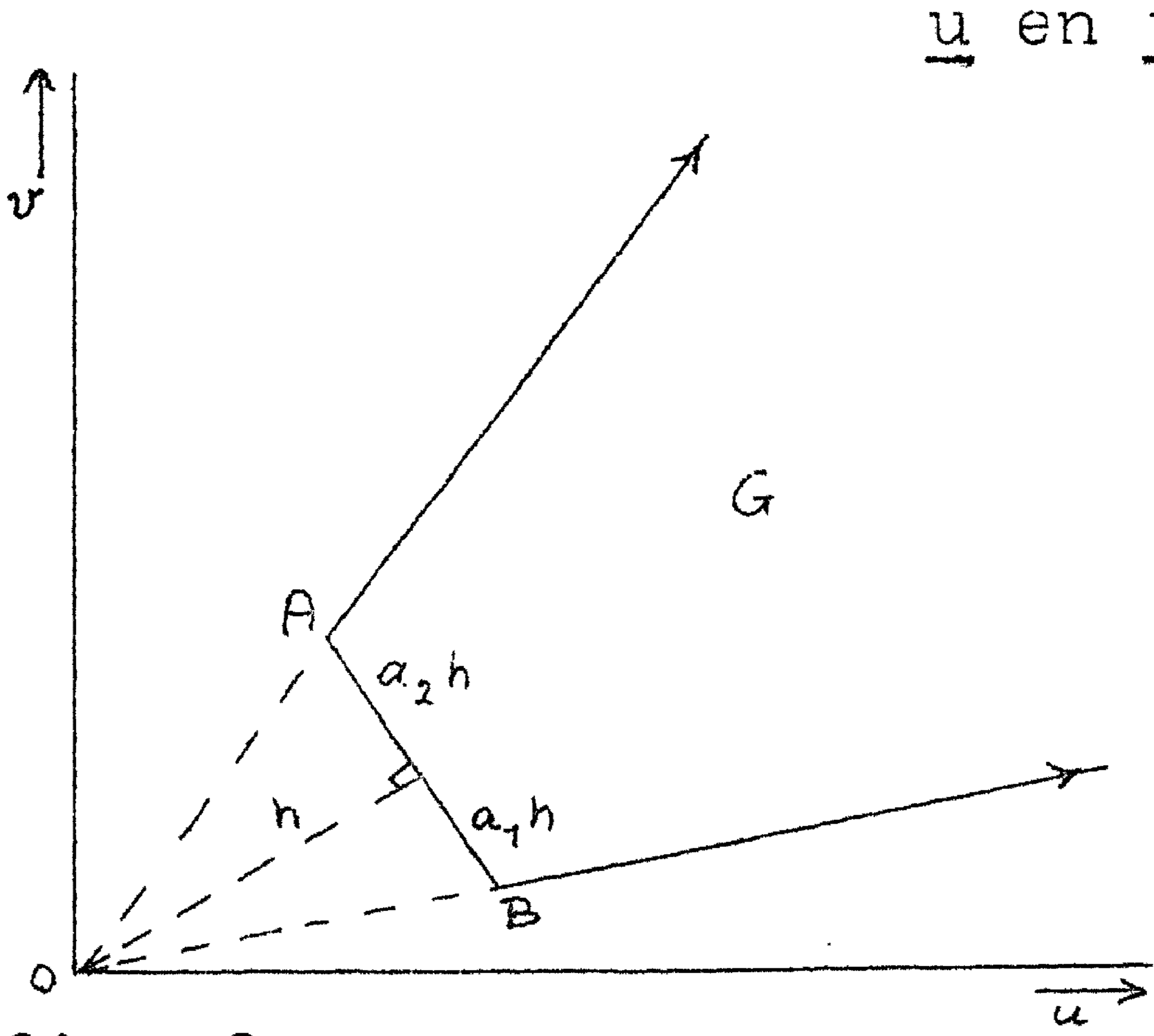


fig. 2a.

$P\left[(\underline{u},\underline{v}) \in G\right] = T(h,a_1) + T(h,a_2).$

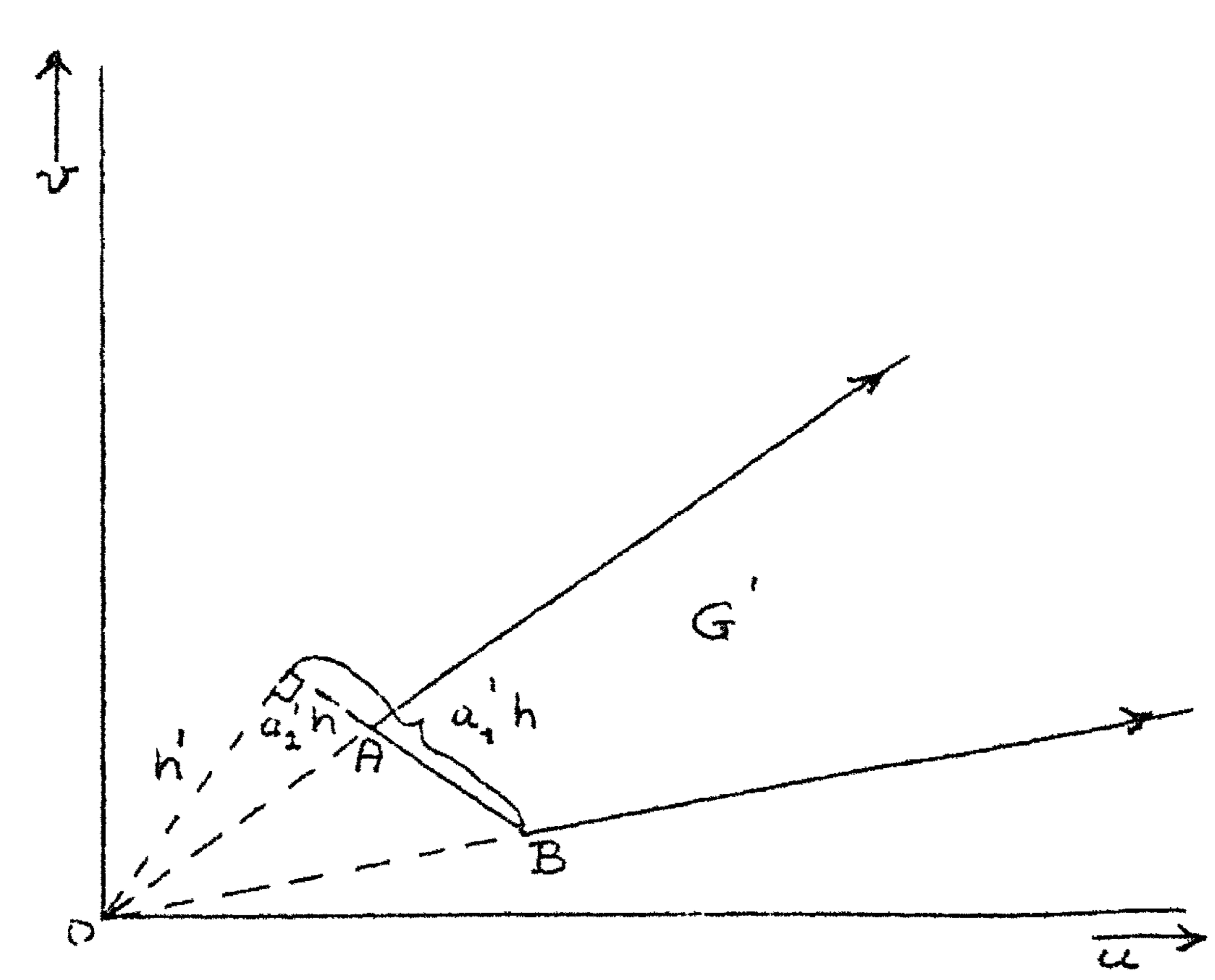


fig. 2b.

$P\left[(\underline{u},\underline{v}) \in G'\right] = T(h;a_1') - T(h;a_2')$

Zijn de hoekpunten A en B gegeven door de coördinaten (A_u,A_v) resp. (B_u,B_v) dan geldt voor h, a_1 en a_2:

$$(3.1)\quad h = \frac{|A_uB_v - A_vB_u|}{\sqrt{(A_u-B_u)^2 + (A_v-B_v)^2}}\ ; \qquad a_1 = \frac{|A_u(B_u-A_u) + A_v(B_v-A_v)|}{|A_uB_v - A_vB_u|}\ ;$$

$$a_2 = \frac{|B_u(B_u-A_u) + B_v(B_v-A_v)|}{|A_uB_v - A_vB_u|}\ .$$

De kans op een waarnemingspaar ($\underline{u},\underline{v}$) in een willekeurige veelhoek kan bepaald worden door het bovenstaande op elke zijde toe te passen en deze resultaten te combineren. De juiste combinatie kan bepaald worden aan de hand van een tekening van de veelhoek in het (u,v)-vlak.

4. Interpolatie in de tabellen voor T(h,a) van OWEN (1956).

De berekende waarden van T(h,a) zijn in 3 tabellen ondergebracht. In tabel A is een grove indeling voor a en een fijne indeling voor h gebruikt; in tabel B een fijne indeling voor a en een grove indeling voor h; in tabel C tenslotte is $h \geq 3$. In tabel C kan een gewone lineaire interpolatie gebruikt worden. Is $h < 3$ dan moet zowel tabel A als tabel B bij de interpolatie gebruikt worden. Deze verloopt dan als volgt: Stel dat $T(h_2,a_2)$ berekend moet worden. Kies dan a_1 en a_3 met $a_1 \leq a_2 < a_3$ uit de grove indeling voor a van tabel A en h_1 en h_3 met $h_1 \leq h_2 < h_3$ uit de grove indeling voor h van tabel B. Uit tabel A en B zijn dan de bijbehorende functiewaarden af te lezen (eventueel met toepassing van een lineaire interpolatie naar h in tabel A of naar a in tabel B):

h \ a	a_1	a_2	a_3
h_1	$T(h_1,a_1)$	$T(h_1,a_2)$	$T(h_1,a_3)$
h_2	$T(h_2,a_1)$	-	$T(h_2,a_3)$
h_3	$T(h_3,a_1)$	$T(h_3,a_2)$	$T(h_3,a_3)$

Bepaal verder $b = \dfrac{a_2-a_1}{a_3-a_1}$ en $c = \dfrac{h_2-h_1}{h_3-h_1}$ en hiermee de coëfficiënten matrix

$$\{w_{ij}\} = \begin{pmatrix} -(1-b)(1-c) & 1-c & -b(1-c) \\ 1-b & - & b \\ -(1-b)c & c & -bc \end{pmatrix}.$$

$T(h_2,a_2)$ kan dan bepaald worden met de formule:

$$T(h_2,a_2) = \sum_{i=1}^{3} \sum_{j=1}^{3} w_{ij}\ T(h_i,a_j).$$

Komt h_2 (of a_2) voor in de grove indeling voor h van tabel B (resp. voor a van tabel A) dan reduceert de beschreven interpolatiemethode tot een normale lineaire interpolatie in deze tabel B (of A).

5. <u>Rechthoeken bij een willekeurige twee-dimensionale normale verdeling</u>. De kans dat een waarnemingspaar (x;y') uit een willekeurige gegeven twee-dimensionale normale verdeling in een gegeven rechthoek met zijden evenwijdig aan de assen zal liggen kan bepaald worden als combinatie van kansen van de vorm $P\left[\underline{x}' \leq h';\ \underline{y}' \leq k'\right]$. Voor een dergelijke kans geeft OWEN een uitdrukking in de functies G en T. Stel $\underline{x}'$ en $\underline{y}'$ bezitten een $N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ verdeling. Bij x' en y' worden dan eerst de gestandaardiseerde grootheden x en y bepaald, dus:

$$x = \frac{x' - \mu_x}{\sigma_x} \quad \text{en} \quad y = \frac{y' - \mu_y}{\sigma_y}$$

dan is

$$P\left[\underline{x}' \leq h';\ \underline{y}' \leq k'\right] = P\left[\underline{x} \leq h;\ \underline{y} \leq k\right]$$

met

$$h = \frac{h' - \mu_x}{\sigma_x} \quad \text{en} \quad k = \frac{k' - \mu_y}{\sigma_y}.$$

Hiervoor geldt nu:

(5.1) $P\left[\underline{x} \leq h;\ \underline{y} \leq k\right] = B(h,k,\rho) =$

$$= \tfrac{1}{2}\left[G(h) + G(k)\right] - T(h,a_h) - T(k,a_k) - \begin{cases} 0, & \begin{cases} \text{als } h.k > 0 \\ \text{of } h.k = 0 \text{ èn } h+k \geq 0. \end{cases} \\ \tfrac{1}{2}, & \text{in andere gevallen} \end{cases}$$

waarbij

$$a_h = \frac{k}{h\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} \quad \text{en} \quad a_k = \frac{h}{k\sqrt{1-\rho^2}} - \frac{\rho}{\sqrt{1-\rho^2}} .$$

Indien hierbij negatieve waarden voor h, k, a_h of a_k optreden dan kan men gebruik maken van

$$\begin{aligned} T(h, -a) &= -T(h,a) \\ T(-h, a) &= T(h,a). \end{aligned} \tag{5.2}$$

De reductieformule (2.2) voor $a > 1$ mag alleen bij positieve a gebruikt worden, bij negatieve a moet dus eerst (5.2) daarna (2.2) toegepast worden.

6. Een willekeurige veelhoek bij een willekeurige normale verdeling. Dit geval wordt door transformatie teruggebracht tot dat van paragraaf 3. Hebben $\underline{x}$ en $\underline{y}$ een $N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ verdeling, dan zijn de transformatieformules:

$$u = \frac{1}{\sqrt{2+2\rho}} \left[\frac{x-\mu_x}{\sigma_x} + \frac{y-\mu_y}{\sigma_y} \right]$$

$$v = \frac{1}{\sqrt{2-2\rho}} \left[- \frac{x-\mu_x}{\sigma_x} + \frac{y-\mu_y}{\sigma_y} \right] .$$

Toegepast op de stochastische grootheden $\underline{x}$ en $\underline{y}$ geeft deze transformatie stochastische grootheden $\underline{u}$ en $\underline{v}$ die onderling onafhankelijk N(0;1) verdeeld zijn; terwijl de veelhoek in het (x,y)-vlak overgaat in een veelhoek in het (u,v)-vlak (tekening!). Na de transformatie kan dus de methode uit paragraaf 3 en 2 toegepast worden.

7. Voorbeelden.

a) Stel dat gegeven is dat $\underline{u}$ en $\underline{v}$ onderling onafhankelijk N(0,1) verdeeld zijn en dat gevraagd is de kans te berekenen op een waarnemingspaar binnen een hoek zoals in figuur 3 is aangegeven.

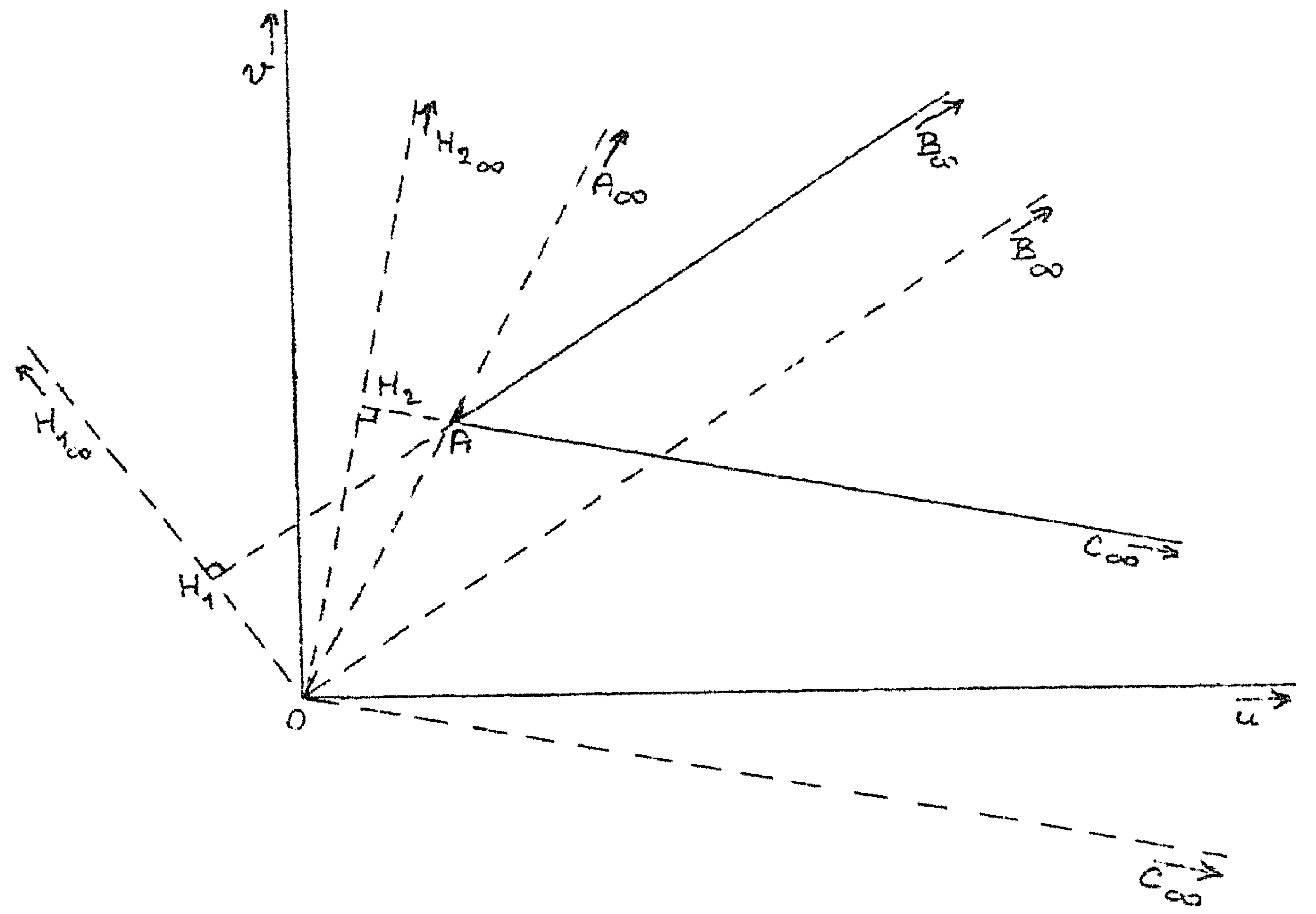


fig. 3
$\underline{u}$ en $\underline{v}$ ond.onafh.
N(0,1)

Hiervoor geldt dan:

$$P\left[(\underline{u},\underline{v})\in(B_\infty AC_\infty)\right] = P\left[(\underline{u},\underline{v})\in(C_\infty AA_\infty)\right] - P\left[(\underline{u},\underline{v})\in(B_\infty AA_\infty)\right] =$$

$$= P\left[(\underline{u},\underline{v})\in(C_\infty H_2H_{2\infty})\right] - P\left[(\underline{u},\underline{v})\in(A_\infty AH_2H_{2\infty})\right] +$$

$$- \left(P\left[(\underline{u},\underline{v})\in(B_\infty H_1H_{1\infty})\right] - P\left[(\underline{u},\underline{v})\in(A_\infty AH_1H_{1\infty})\right]\right) =$$

$$= \tfrac{1}{2}\left[1-G(OH_2)\right] - T\left(OH_2;\ \frac{H_2A}{OH_2}\right) - \left\{\tfrac{1}{2}\left[1-G(OH_1)\right] - T\left(OH_1;\ \frac{H_1A}{OH_1}\right)\right\} .$$

b) Stel $\underline{x}$ en $\underline{y}$ bezitten een twee dimensionale normale verdeling met $\mu_x = 2$; $\mu_y = 1$; $\sigma_x = 10$; $\sigma_y = 2$ en $\rho = 0{,}28$ en laat $A = (4;\ 3{,}8)$; $B = (13;\ 2)$ en $C = (9;\ 6)$ de hoekpunten zijn van een driehoek in het (x,y)-vlak (fig. 4a en 4b). Gevraagd wordt de waarschijnlijkheid dat een waarnemingspaar $(\underline{x},\underline{y})$ binnen deze driehoek zal liggen.

Om dit te berekenen moet volgens paragraaf 6 eerst de transformatie

$$\begin{cases} u = \dfrac{1}{\sqrt{2+2\rho}}\left(\dfrac{x-\mu_x}{\sigma_x} + \dfrac{y-\mu_y}{\sigma_y}\right) = \dfrac{1}{1{,}6}\left(\dfrac{x-2}{10} + \dfrac{y-1}{2}\right) \\[2ex] v = \dfrac{1}{\sqrt{2-2\rho}}\left(-\dfrac{x-\mu_x}{\sigma_x} + \dfrac{y-\mu_y}{\sigma_y}\right) = \dfrac{1}{1{,}2}\left(-\dfrac{x-2}{10} + \dfrac{y-1}{2}\right) \end{cases}$$

uitgevoerd worden, waardoor driehoek ABC overgaat in driehoek A'B'C' in het vlak van twee onderling onafhankelijke N(0;1) verdeelde grootheden. De nieuwe punten zijn:

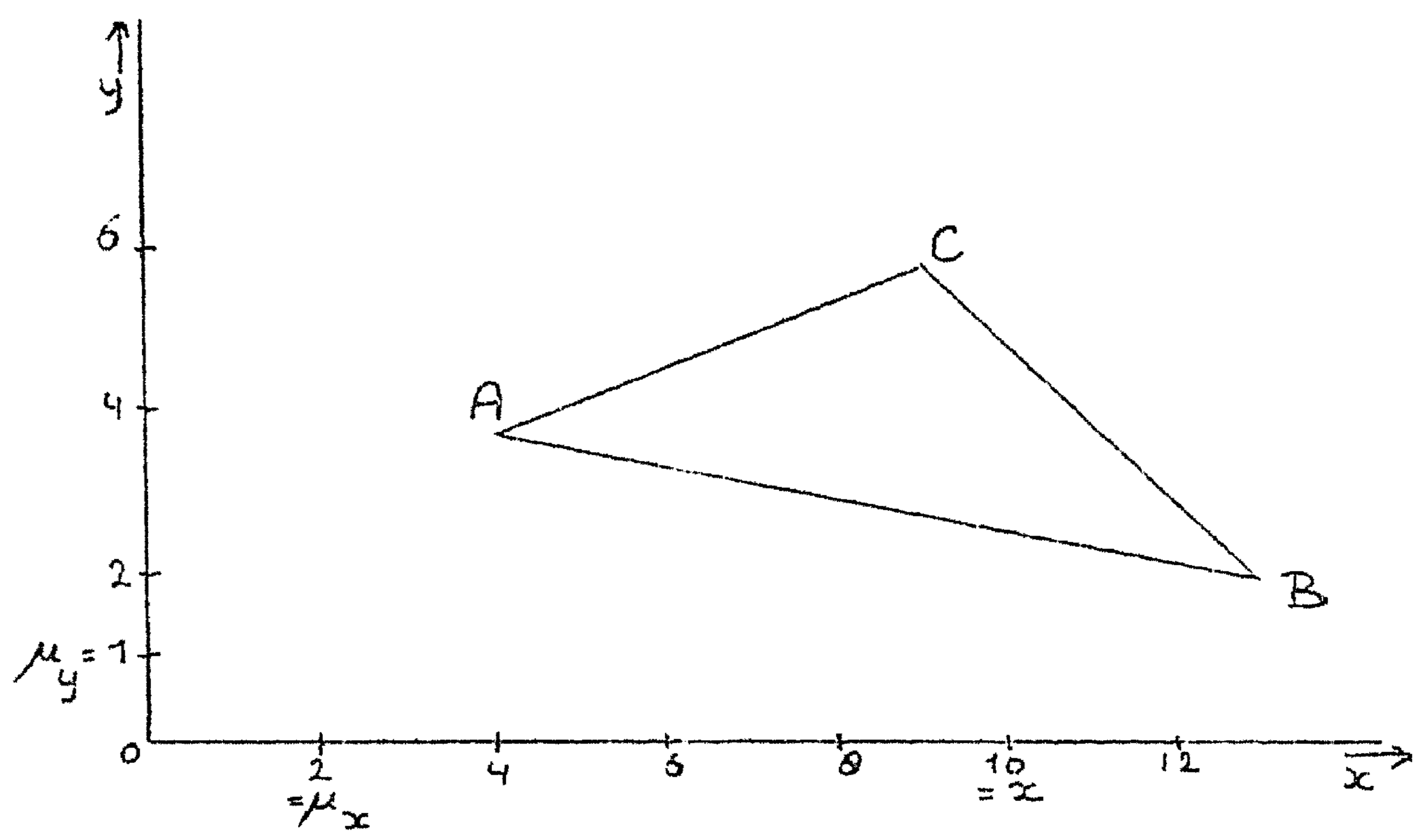


fig. 4a

$\underline{x}$ en $\underline{y}$ N (2; 1; 10; 2; 0,28)

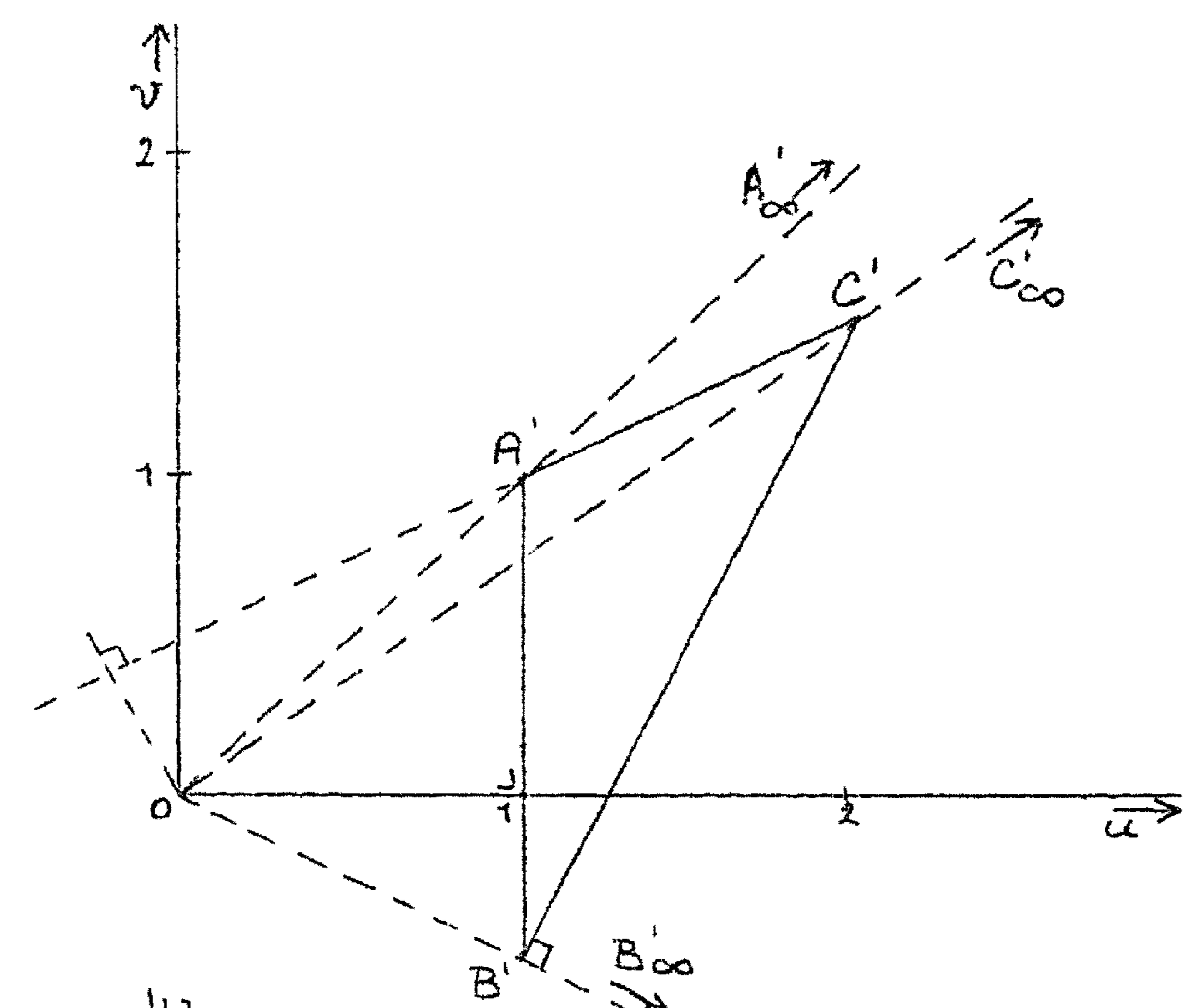


fig. 4b

$\underline{u}$ en $\underline{v}$ ond. onafh. N (0; 1)

A' = (1; 1); B' = (1; -0,5) en C' = (2; 1,5) (fig. 4b)

Uit figuur 4b kan afgelezen worden dat

$$(7.1) \quad P\left[(\underline{u},\underline{v}) \in \Delta(A'B'C')\right] =$$

$$= P\left[(\underline{u},\underline{v})\in(A'_\infty\ A'B'B'_\infty)\right] - P\left[(\underline{u},\underline{v})\in(A'_\infty\ A'C'C'_\infty)\right] +$$

$$- P\left[(\underline{u},\underline{v})\in(C'_\infty\ C'B'B'_\infty)\right] .$$

De drie kansen in het rechterlid kunnen volgens de methode uit paragraaf 3 bepaald worden. Voor de eerste kans volgt dan direct uit de figuur: $h = 1$; $a_1 = 1$ en $a_2 = 0,5$, dus

$$P\left[(\underline{u},\underline{v})\in(A'_\infty\ A'B'B'_\infty)\right] = T(1;\ 1) + T(1;\ 0,5) =$$

$$= 0,066742 + 0,043065 = 0,109807$$

is direct in de tabellen van OWEN op te zoeken.

Voor de tweede kans geldt:

$$h = \frac{|A'_u C'_v - C'_u A'_v|}{\sqrt{(A'_u - C'_u)^2 + (A'_v - C'_v)^2}} = \frac{0,5}{\sqrt{1,25}} = 0,2\sqrt{5}$$

$$a_1 = \frac{|A'_u(C'_u - A'_u) + A'_v(C'_v - A'_v)|}{|A'_u C'_v - C'_u A'_v|} = \frac{1,5}{0,5} = 3$$

$$a_2 = \frac{|C'_u(C'_u - A'_u) + C'_v(C'_v - A'_v)|}{|A'_u C'_v - C'_u A'_v|} = \frac{2,75}{0,5} = 5,5.$$

Dus:

(7.2) $P\left[(\underline{u},\underline{v}) \in (A'_\infty\ A'C'C'_\infty)\right] = T(0,2\sqrt{5};\ 5,5) - T(0,2\sqrt{5};3).$

Volgens de reductieformule (2,2) is nu:

$T(0,2\sqrt{5};\ 5,5) =$

$= \frac{1}{2}\left[G(0,2\sqrt{5}) + G(1,1\sqrt{5})\right] - G(0,2\sqrt{5}).G(1,1\sqrt{5}) - T(1,1\sqrt{5},\ \frac{1}{5,5}) =$

$= \frac{1}{2}\left[G(0,4472136) + G(2,4596748)\right] +$
$- G(0,4472136)\ .\ G(2,4596748) - T(2,4596748;\ 0,1818182) =$

$= \frac{1}{2}\left[0,672640 + 0,993047\right] - 0,667963 - T(2,4596748;\ 0,1818182) =$

$= 0,164881 - T(2,4596748;\ 0,1818182).$

Om de T te bepalen moet de interpolatiemethode van paragraaf 4 gebruikt worden. Uit de tabellen van OWEN worden daarbij de volgende waarden voor T opgezocht:

h \ a	0	0,1818182	0,25
2,25	0	0,002216	0,002948
2,4596748	0	-	0,001782
2,50	0	0,001216	0,001609

waarbij voor de middelste waarden van de eerste en derde rij in tabel B geïnterpoleerd is en voor de middelste waarde van de derde kolom in tabel A. Verder is

$b = \frac{0,1818182}{0,25} = 0,7272728$ en $C = \frac{2,4596748 - 2,25}{2,50 - 2,25} =$

0,8386992 zodat de coëfficiëntenmatrix, waarvan we de eerste kolom niet nodig hebben, gelijk is aan:

$$\left\{ w_{ij} \right\} = \begin{pmatrix} - & 0,1613008 & -0,1173097 \\ - & 0 & 0,7272728 \\ - & 0,8386992 & -0,6099631 \end{pmatrix}.$$

De gezochte T wordt nu

$$T(2,4596748;\ 0,1818182) = \sum_{i,j} w_{ij}\ T(h_i, a_j) = 0,001346$$

zodat tenslotte

$$T(0,2\sqrt{5};\ 55) = 0,164881 - 0,001346 = 0,163535.$$

Voor de tweede functie T van (7.2) moet ook de reductie-formule (2.2) gebruikt worden:

$$\begin{aligned} T\ (0,2\sqrt{5};3) &= \tfrac{1}{2}\left[G(0,2\sqrt{5})+G(0,6\sqrt{5})\right] - G(0,2\sqrt{5}),G(0,6\sqrt{5}) - T(0,6\sqrt{5};\tfrac{1}{3}) = \\ &= \tfrac{1}{2}\left[0,672640 + 0,910144\right] - 0,612199 - T(1,3416408; \\ &\qquad\qquad 0,333333) = \\ &= 0,179193 - T(1,3416408;\ 0,333333). \end{aligned}$$

T(1,3416408; 0,333333) moet weer door interpolatie bepaald worden met de T waarden:

h \ a	0,25	0,3333333	0,50
1,25	0,017569	0,022801	0,031828
1,3416408	0,015564	-	0,028017
1,50	0,012372	0,015975	0,022006

terwijl $b = \dfrac{0,3333333 - 0,25}{0,25} = 0,3333333$ en

$c = \dfrac{1,3416408 - 1,25}{0,25} = \dfrac{0,0916408}{0,25} = 0,3665632$ dus

$$\{w_{ij}\} = \begin{pmatrix} -\ 0,4222912 & 0,6334368 & -0,2111456 \\ 0,6666667 & 0 & 0,3333333 \\ -\ 0,2443755 & 0,3665632 & -0,1221877 \end{pmatrix}$$

dus

$$T(1,3416408;\ 0,3333333) = \sum_{i,j} w_{ij}\ T(h_i, a_j) = 0,020162.$$

Dit geeft

$$T(0,2\sqrt{5};\ 3) = 0,179193 - 0,020162 = 0,159031$$

en in (7,2) gesubstitueerd:

$$P\left[(\underline{u},\underline{v})\in(A'_{\infty}\ A'C'C'_{\infty})\right] = 0,163535 - 0,159031 = 0,004504.$$

Voor de laatste kans uit het rechterlid van (7.1) geldt tenslotte:

$$h = \frac{|B'_uC'_v - C'_uB'_v|}{\sqrt{(B'_u-C'_u)^2 + (B'_v-C'_v)^2}} = \frac{2,5}{\sqrt{5}} = 0,5\sqrt{5}$$

$$a_1 = \frac{|B'_u(C'_u-B'_u) + B'_v(C'_v-B'_v)|}{|B'_uC'_v - C'_uB'_v|} = \frac{0}{2,5} = 0$$

$$a_2 \quad \frac{|C'_u(C'_u-B'_u) + C'_v(C'_v-B'_v)|}{|B'_uC'_v - C'_uB'_v|} = \frac{5}{2,5} = 2.$$

Dus:

$$P\left[(\underline{u},\underline{v})\in(B'_{\infty}\ B'C'C'_{\infty})\right] = T(0,5\sqrt{5};\ 2) =$$

$$= \tfrac{1}{2}\left[G(0,5\sqrt{5}) + G(\sqrt{5})\right] - G(0,5\sqrt{5}).G(\sqrt{5}) - T(\sqrt{5};\ 0,5) =$$

$$= \tfrac{1}{2}\left[0,868224 + 0,987326\right] - 0,857220 - T(2,236068;\ 0,5) =$$

$$= 0,070555 - T(2,2360680;\ 0,5) =$$

$$= 0,070555 - 0,005059 = 0,065496.$$

Voor de gevraagde kans volgt hieruit, zie (7.1):

$$P\left[(\underline{x},\underline{y})\in(ABC)\right] = P\left[(\underline{u},\underline{v})\in(A'B'C')\right] = 0,109807-0,004504-0,065496 = 0,039807.$$

De waarden van de normale verdelingsfunctie $G(x)$ zijn bepaald met: Tables of probability functions, vol. I, Federal Works Agency, Works Projects Administration for the City of New York.

Generalisatie van paragraaf 5 voor drie-dimensionale verdelingen.
Bezitten $\underline{x}$, $\underline{y}$ en $\underline{z}$ een drie-dimensionale verdeling met verwachtingen 0, varianties 1 en correlatiecoëfficiënten ρ_{xy}, ρ_{xz}, ρ_{yz} dan kunnen kansen van de vorm:

$$P\left[\underline{x} \leq h;\ \underline{y} \leq k;\ \underline{z} \leq m\right]$$

met formules van G.P. STECK (1958) uitgedrukt worden in functies G, T en S. De laatste is een functie van drie variabelen, welke door STECK getabelleerd is. In deze tabel voor S kan lineair geïnterpoleerd worden; voor de functie S gelden overigens soortgelijke eigenschappen als voor T. Daar de formules tamelijk uitgebreid zijn, wordt volstaan met een verwijzing naar paragraaf 2 (pp. 781-782-783) van de beschrijving van G.P. STECK.

Litteratuur:

H.A. Bender (1955) Bivariate distribution. Bulletin Amer. Math. Soc., 61, pp. 561-562.

J.H. Cadwell (1951) The bivariate normal integral. Biometrika, 38, pp. 475-481.

C. Nicholson (1943) The probability integral for two variables. Biometrika, 33, pp. 59-72.

D.B. Owen (1956) Tables for computing bivariate normal probabilities. A.M.S., 27, pp. 1075-1090.

G.P. Steck (1958) A table for computing trivariate normal probabilities. A.M.S., 29, pp. 780-800.

D. Teichroew (1955) Numerical Analysis Research unpublished statistical tables. J.A.S.A., 50, pp. 550-556.