

MATHEMATISCH CENTRUM

Afdeling Toegepaste Wiskunde

R A P P O R T T. W. no.7

Auteur: J.H. Kemperman

Titel: De verdelingsfunctie van het
aantal inversies in de test van
Mann en Whitney.

Datum: Maart 1950.

- - -

Corrigenda bij rapport T.W. no. 7.

blz. 1	regel 18	Staat: $\frac{1}{\sqrt{2}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.	Lees: $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.
blz. 3	regel 14	Vedder	Verder
	formule (14)	" $(1-t)^2$	" $(1-t)$
blz. 5	formule (17)	" $\frac{\alpha_{n,j}(\varepsilon)}{(1-\varepsilon t)^{j+1}}$	" $\frac{\alpha_{n,j}(\varepsilon)}{(1-\varepsilon^{-1}t)^{j+1}}$
	formule (18)	" ε^h	" ε^{-h}
blz. 9	regel 16	" $F_n(v)$	" $F_n(v)$
	regel 5 v.o.	" $1 \leq h < k \leq n$	" $\sum_{1 \leq h < k \leq n}$
blz. 10	laatste kolom tabel	+15	-15
blz. 13	regel 14 v.o.	" 5. Appendix	" Appendix
blz. 14	regel 16 v.o.	" variabelen	" variabelen η_j
	formule (68)	" $=mn \{ \dots$	" $\sigma^2 = mn \{ \dots$
blz. 15	regel 19	" $d\xi d\eta dx$	" $dx d\xi d\eta$
	formule (67)	" $(x + \mu\sqrt{\frac{2}{3}})^2$	" $(x + \mu\sqrt{\frac{2}{3}})^2$
blz. 16	regel 11		" $\frac{1}{2} \phi(-\frac{1}{2}\lambda\sqrt{3})$
	formule (70)		" $\phi(-\frac{1}{2}\lambda\sqrt{3})$
blz. 17	regel 11 v.o.		" $(1 - \exp(-\frac{5}{8} \frac{\lambda^2}{\cot^2 \varphi})) d\varphi$

1. Inleiding.

Probleem ¹⁾. Laten $x_1, \dots, x_m; y_1, \dots, y_n$ onafhankelijke stochastische variabelen zijn, waarin m en n vaste natuurlijke getallen voorstellen. Hierin hebben $x_1, \dots, x_m$ resp. $y_1, \dots, y_n$ een zelfde continue waarschijnlijkheidsverdeling $F(x)$ resp. $G(y)$. Elk paar x_i, y_k met $y_k < x_i$ wordt als één inversie geteld. Zij $P(u, m, n)$ de kans op u inversies en

$$(1) \quad Q(u, m, n) = \sum_{v=0}^u P(v, m, n)$$

de kans op ten hoogste u inversies. Gevraagd wordt de grootheid $Q(u, m, n)$ te bepalen.

In dit rapport zullen we een oplossing geven van het gestelde probleem in de twee volgende bijzondere gevallen.

1. Het geval, dat $F(x) \equiv G(x)$, de z.g. gelijkverdeling. Onderzocht wordt hiervoor ook het gedrag van $Q(u, m, n)$ voor grote waarden van m resp. grote waarden van m en n beide.

2. Het geval dat $G(x) = F(x - \mu)$ waarin

$$\begin{aligned} F(x) &= 1 \quad \text{voor } 0 \leq x \leq 1 \\ &= 0 \quad \text{voor } x < 0 \text{ en voor } x > 1. \end{aligned}$$

In het geval, dat

$$G(x + \mu) = F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

worden in een appendix bruikbare formules gegeven voor de spreiding σ van de waarschijnlijkheidsverdeling van het aantal inversies, zowel voor grote als voor kleine waarden μ .

2. Het geval van gelijkverdeling.

Neem aan, dat $F(x) \equiv G(x)$. Zij $z_1 < z_2 < \dots < z_{n+m}$ de geordende vereniging van de elementen $x_1, \dots, x_m; y_1, \dots, y_n$. Elke manier om m elementen uit de rij $z_1, \dots, z_{n+m}$ aan te duiden met x_j en de overige met y_k levert een rangschikking met dezelfde waarschijnlijkheid. Het totaal aantal mogelijkheden is $\binom{m+n}{n}$. Op een factor $\frac{1}{\binom{m+n}{n}}$ na wordt dus de waarschijnlijkheid $P(u, m, n)$, dat u het aantal inversies is, gegeven door het aantal mogelijkheden om de n elementen y_k te verdelen in $m+1$ groepen van n_0 , resp. $n_1, \dots$, resp. n_m elementen, zodanig dat de rangschikking

(2) $\left[n_m \text{ elem. } y \right] < x_1 < \left[n_{m-1} \text{ el. } y \right] < x_2 < \dots < x_m < \left[n_0 \text{ el. } y \right]$
een situatie oplevert met u inversies. Als dus $A(u, m, n)$ voorstelt het aantal partities

$$(3) \quad n = n_0 + n_1 + \dots + n_m$$

van n met n_i geheel ≥ 0 voor $i = 0, 1, \dots, m$, en onder de bijvoorwaarde dat

1) vgl. H. B. Mann and D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18(1947), 50-60

$$(4) \quad u = n_1 + 2n_2 + \dots + mn_m,$$

dan volgt

$$(5) \quad P(u, m, n) = \frac{1}{\binom{m+n}{n}} A(u, m, n).$$

Nu geldt

$$(6) \quad \frac{1}{(1-\xi)(1-\xi t)\dots(1-\xi t^m)} = \sum_{n=0}^{\infty} \xi^n \sum_{u=0}^{\infty} A(u, m, n) t^u.$$

want ontwikkelt men het linkerlid van (6) volgens

$$\prod_{j=1}^m \sum_{j=0}^{\infty} \xi^j t^{1j}$$

dan blijkt hierin de coëfficiënt van $\xi^n t^u$ juist gelijk te zijn aan het aantal mogelijkheden om n zodanig te schrijven als een som (3), dat tegelijkertijd aan (4) voldaan is. Deze coëfficiënt is dus gelijk aan $A(u, m, n)$.

Stel nu

$$(7) \quad F(\xi) = \frac{1}{(1-\xi)(1-\xi t)\dots(1-\xi t^m)} = \sum_{n=0}^{\infty} c_n \xi^n$$

Wegens (6) is hierin de coëfficiënt c_n dus gelijk aan de ontwikkeling

$$(8) \quad \mathcal{O}_{m,n} = \sum_{u=0}^{\infty} A(u, m, n) t^u$$

Uit (7) volgt

$$F(\xi t) = \frac{1-\xi}{1-\xi t^{m+1}} F(\xi)$$

en dus

$$(1-\xi t^{m+1}) \sum_{n=0}^{\infty} c_n \xi^n t^n = (1-\xi) \sum_{n=0}^{\infty} c_n \xi^n.$$

Vergelijken we de coëfficiënt van ξ^n in beide leden, dan volgt

$$c_n t^n - c_{n-1} t^{m+n} = c_n - c_{n-1}$$

ofwel

$$\frac{c_n}{c_{n-1}} = \frac{1-t^{m+n}}{1-t^n} \quad (n=1, 2, 3, \dots)$$

en wegens $c_0 = 1$ volgt hieruit voor $\mathcal{O}_{m,n} = c_n$

$$(9) \quad \mathcal{O}_{m,n} = \sum_{u=0}^{\infty} A(u, m, n) t^u = \frac{(1-t^{m+1}) \dots (1-t^{m+n})}{(1-t)(1-t^2) \dots (1-t^n)}$$

Opm. Deze formule is ook als volgt te bewijzen. Door voor de rangschikking (2) de mogelijkheden, dat het eerste element een element x_j resp. een element y_k is, afzonderlijk te beschouwen ziet men in dat geldt voor $(m, n) \neq (0, 0)$

$$(10) \quad A(u, m, n) = A(u, m-1, n) + A(u-m, m, n-1)$$

mits we stellen

$$A(u, m, n) = 0$$

voor $u < 0$, voor $m < 0$ en voor $n < 0$. Hieruit volgt voor $(m, n) \neq (0, 0)$

$$\mathcal{O}_{m,n} = \sum_{u=0}^{\infty} A(u, m, n) t^u = \sum_{u=0}^{\infty} (A(u, m-1, n) + A(u-m, m, n-1)) t^u$$

$$= \mathcal{O}_{m-1,n} + \sum_{u=-m}^{\infty} A(u,m,n-1)t^{u+m}$$

ofwel voor $(m,n) \neq (0,0)$

$$(11) \quad \mathcal{O}_{m,n} = \mathcal{O}_{m-1,n} + t^m \mathcal{O}_{m,n-1}$$

Wegens $\mathcal{O}_{0,0} = 1$ en $\mathcal{O}_{m,n} = 0$ voor $m < 0$ of $n < 0$ is $\mathcal{O}_{m,n}$ door de re-
cursieformule (11) eenduidig bepaald. Het rechterlid van (9) voldoet
eveneens aan (11) en genoemde beginvoorwaarden, waarmee een tweede bewijs
van formule (9) is gevonden.

Uit (9) volgen vele eigenschappen van $A(u,m,n)$ (en dus wegens (5)
ook voor $P(u,m,n)$) welke overigens ook direct uit de definitie van
 $A(u,m,n)$ zijn af te leiden. Vooreerst volgt uit

$$\mathcal{O}_{m,n} = \frac{(1-t^{m+1}) \dots (1-t^{m+n})}{(1-t)(1-t^2) \dots (1-t^n)} = \frac{(1-t^{n+1}) \dots (1-t^{m+n})}{(1-t) \dots (1-t^m)} = \mathcal{O}_{n,m}$$

dat

$$(12) \quad A(u,m,n) = A(u,n,m).$$

Verder geldt wegens

$$t^{m+n} \mathcal{O}_{m,n} \left\{ \frac{1}{t} \right\} = \mathcal{O}_{m,n}(t)$$

de symmetrieeigenschap t.o.v. de waarde $u = \frac{mn}{2}$

$$(13) \quad A(mn-u,m,n) = A(u,m,n)$$

Daar $A(u,m,n) = 0$ voor $u < 0$, volgt dus ook dat $A(u,m,n) = 0$ voor $u > mn$.

Uit (9) blijkt, dat voor $u \leq m$ de grootheid $A(u,m,n)$ van m niet af-
hangt. Is bovendien $u \leq n$, dan is $A(u,m,n) = p(u)$ een alleen van u af-
hankelijke grootheid. En wel geldt

$$\sum_{u=0}^{\infty} p(u) t^u = \frac{1}{(1-t)(1-t^2)(1-t^3) \dots}$$

waarin $p(u)$ het aantal partities voorstelt van u in natuurlijke getallen.
Van deze functie $p(u)$ zijn vele eigenschappen bekend (vgl. Hardy and
Wright, An introduction to the theory of numbers, Chapt. XIX: Partitions).

Wegens (5) en (9) geldt

$$(14) \quad \sum_{u=0}^{\infty} P(u,m,n)t^u = \frac{1}{\binom{m+n}{n}} \cdot \frac{(1-t^{m+1}) \dots (1-t^{m+n})}{(1-t)^2(1-t^2) \dots (1-t^n)}$$

Verder hebben we wegens (1)

$$\begin{aligned} \sum_{u=0}^{\infty} Q(u,m,n)t^u &= \sum_{u=0}^{\infty} t^u \sum_{v=0}^u P(u,m,n) \\ &= \sum_{v=0}^{\infty} P(v,m,n)t^v \sum_{u=v}^{\infty} t^{u-v} \\ &= \frac{1}{1-t} \sum_{v=0}^{\infty} P(v,m,n) t^v \end{aligned}$$

en wegens (14) volgt dus

$$(15) \quad \sum_{u=0}^{\infty} Q(u,m,n)t^u = \frac{1}{\binom{m+n}{n}} \frac{(1-t^{m+1}) \dots (1-t^{m+n})}{(1-t)^2(1-t^2)(1-t^3) \dots (1-t^n)}$$

waarmee een methode gevonden is voor de berekening van $Q(u, m, n)$.

Voorbeeld. We berekenen de verdelingsfunctie $Q(u, m, n)$ voor het geval $m=5$ en $n=3$. Dan hebben we

$$\sum_{u=0}^{\infty} Q(u, 5, 3) t^u = \frac{1}{\binom{8}{5}} \frac{(1-t^6)(1-t^7)(1-t^8)}{(1-t)^2(1-t^2)(1-t^3)}.$$

Wegens $\frac{1}{(1-t)^2} = \sum_{k=0}^{\infty} (k+1)t^k$ verloopt de berekening als in het volgende schema. Hierin is

$$R_n(t) = \frac{1}{(1-t)^2(1-t^2)(1-t^3)\dots(1-t)^n}$$

Voor elke functie uit de eerste kolom staan in de overeenkomstige rij de coëfficiënten uit de machtreeksontwikkeling volgens opklimmende machten van t .

$\rightarrow t$	0	1	2	3	4	5	6	7	8
$R_1(t)$	1	2	3	4	5	6	7	8	9
$t^2 R_1(t)$			1	2	3	4	5	6	7
$t^4 R_1(t)$					1	2	3	4	5
$t^6 R_1(t)$							1	2	3
$t^8 R_1(t)$									1
$R_2(t)$	1	2	4	6	9	12	16	20	25
$t^3 R_2(t)$				1	2	4	6	9	12
$t^6 R_2(t)$							1	2	4
$R_3(t)$	1	2	4	7	11	16	23	31	41
$-t^6 R_3(t)$							-1	-2	-4
$-t^7 R_3(t)$								-1	-2
$-t^8 R_3(t)$									-1
$\binom{8}{5} \sum_{u=0}^{\infty} Q(u, 5, 3) t^u$	1	2	4	7	11	16	22	28	34
$Q(u, 5, 3) \rightarrow$	$\frac{1}{56}$	$\frac{2}{56}$	$\frac{4}{56}$	$\frac{7}{56}$	$\frac{11}{56}$	$\frac{16}{56}$	$\frac{22}{56}$	$\frac{28}{56}$	$\frac{34}{56}$

Wegens de symmetrieeigenschap (13) is de functie $Q(u, 5, 3)$ nu volledig bekend.

Uit dit voorbeeld blijkt wel, dat het van vitaal belang is de volgende ontwikkeling te bepalen.

$$(16) \frac{1}{(1-t)^2(1-t^2)\dots(1-t^n)} = \sum_{h=0}^{\infty} a_{n,h} t^h.$$

Dit kunnen we doen door het linkerlid in partiële breuken te splitsen. De polen van het linkerlid zijn de eenheidswortels.

$$\xi = \exp. (2\pi i \frac{k}{l}).$$

met k en l geheel onderling ondeelbaar en $0 \leq k < l \leq n$. Hierin heeft

$\xi = 1$ de multipliciteit $\mu(1) = n+1$.

De andere polen hebben een multipliciteit

$$\mu\left(e^{2\pi i \cdot \frac{k}{l}}\right) = \left[\frac{n}{l}\right] \leq \left[\frac{n}{2}\right].$$

We krijgen dus een splitsing in partiële breuken

$$(17) \quad \frac{1}{(1-t)^2(1-t^2)\dots(1-t^n)} = \sum_{\xi} \sum_{j=0}^{\mu(\xi)-1} \frac{\alpha_{n,j}(\xi)}{(1-\xi t)^{j+1}}$$

waaruit wegens (16) en

$$\frac{1}{(1-z)^{j+1}} = \sum_{h=0}^{\infty} \binom{h+j}{h} z^h$$

volgt dat

$$(18) \quad a_{n,h} = \sum_{\xi} \sum_{j=0}^{\mu(\xi)-1} \alpha_{n,j}(\xi) \binom{h+j}{h} \xi^h$$

Wegens $|\xi| = 1$, $\mu(1) = n+1$, $\mu(\xi) \leq \left[\frac{n}{2}\right]$ voor $\xi \neq 1$ en

$$\binom{h+j}{h} \sim \frac{(h)^j}{j!}$$

voor $h \rightarrow \infty$, volgt uit (18) de asymptotische formule

$$a_{n,h} = \sum_{j=0}^{\left[\frac{n}{2}\right]} \alpha_{n,j}(1) \cdot \binom{h+j}{h} + O(h^{\left[\frac{n}{2}\right]-1})$$

Er zijn dus constanten $\beta_{n,j}$ te vinden, zodanig dat voor $h \rightarrow \infty$ geldt

$$(19) \quad a_{n,h} = \sum_{j=0}^{\left[\frac{n}{2}\right]} \beta_{n,j} h^j + O(h^{\left[\frac{n}{2}\right]-1})$$

Wegens $\alpha_{n,n}(1) = \frac{1}{n!}$ en $\binom{h+n}{h} \sim \frac{h^n}{n!}$ geldt voor de hoofdterm

$$\beta_{n,n} \text{ in (19), dat } \beta_{n,n} = \left(\frac{1}{n!}\right)^2.$$

Voor kleine waarden van n kan men zonder veel moeite de splitsing (17) in partiële breuken uitvoeren en vindt men wegens (18) exacte formules voor $a_{n,h}$. Op deze wijze zijn de volgende formules afgeleid

$$(20) \quad a_{1,h} = h + 1$$

$$(21) \quad a_{2,h} = \left(\frac{h+2}{2}\right)^2 \quad \text{als } h \text{ even}$$

$$a_{2,h} = \left(\frac{h+1}{2}\right) \left(\frac{h+3}{2}\right) \quad \text{als } h \text{ oneven}$$

$$(22) \quad a_{3,h} = \frac{1}{72} \left[(h+1)(2h^2+19h+47) + \theta \right]$$

$$\text{met } \theta = \begin{matrix} 25 & 8 & 9 & 16 & 17 & 0 \end{matrix}$$

$$\text{voor } h \equiv \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 \end{matrix} \pmod{6}$$

Wegens $0 \leq \frac{\theta}{72} \leq \frac{25}{72}$ is $a_{3,h}$ het gelijk aan het grootste gehele getal

$\leq \frac{1}{72}(h+1)(2h^2+19h+47)$. Verder geldt de formule

$$(23) \quad a_{4,h} = \frac{1}{576} \left[h^4 + 22h^3 + 166h^2 + 504h + 518 + \Theta \right]$$

als h even en

$$(2) \quad a_{4,h} = \frac{1}{576} [h^4 + 22h^3 + 166h^2 + 486h + 420 + \theta] \quad \text{als } h \text{ oneven,}$$

$$\begin{array}{rcccl} \text{waarin } \theta & = & 58 & 57 & -78 & -15 & +58 & -7 & -14 & -15 & -6 & +57 & -14 & -79 \\ \text{voor } h & = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array}$$

(mod. 12)

Aan het eind van dit rapport is nog een tabel toegevoegd van $a_{n,h}$ voor $1 \leq n \leq 5$ en $0 \leq h \leq 40$ ($a_{5,h}$ berekent men uit $a_{4,h}$ met de recursieformule $a_{5,h} = a_{5,h-5} + a_{4,h}$). Men kan deze tabel gebruiken voor de berekening van $Q(u,m,n)$, mits geldt $0 \leq u \leq 40$ en $1 \leq n \leq 5$ (m is verder willekeurig).

Voorbeeld. Gevraagd wordt voor $m=26$ en $n=5$ de kans $Q(34,26,5)$ op hoogstens 34 inversies. Wegens (15) is deze kans gelijk aan de coëfficiënt van t^{34} in

$$\frac{1}{\binom{31}{5}} \frac{(1-t^{27})(1-t^{28})(1-t^{29})(1-t^{30})(1-t^{31})}{(1-t)^2(1-t^2)(1-t^3)(1-t^4)(1-t^5)}$$

Wegens (16) volgt dus

$$Q(34,26,5) = \frac{1}{\binom{31}{5}} [a_{5,34} - a_{5,7} - a_{5,6} - a_{5,5} - a_{5,4} - a_{5,3}]$$

en aan de hand van de tabel vinden we

$$Q(34,26,5) = \frac{8728}{\binom{31}{5}} = 0,051363.$$

3. Gelijkverdeling met grote waarden m of n .

Teneinde de verdelingsfunctie $Q(u,m,n)$ voor grote waarden van m of n te bestuderen, beschouwen we de karakteristieke functie

$$(24) \quad \varphi_{m,n}(y) = \sum_{u=0}^{\infty} P(u,m,n) e^{i(u - \frac{mn}{2})y}$$

Uit (14) volgt

$$\sum_{u=0}^{\infty} P(u,m,n) t^{u - \frac{mn}{2}} = \frac{1}{\binom{m+n}{n}} \frac{(t^{\frac{m+1}{2}} - t^{\frac{m-1}{2}}) \dots (t^{\frac{m+n}{2}} - t^{\frac{m+n-1}{2}})}{(t^{\frac{1}{2}} - t^{-\frac{1}{2}}) \dots (t^{\frac{n}{2}} - t^{-\frac{n}{2}})}$$

en stellen we hierin $t = e^{iy}$ dan vinden we

$$(25) \quad \varphi_{m,n}(y) = \frac{1}{\binom{m+n}{n}} \frac{\sin \frac{m+1}{2}y \sin \frac{m+2}{2}y \dots \sin \frac{m+n}{2}y}{\sin \frac{1}{2}y \sin y \dots \sin \frac{n}{2}y}$$

Als we $\varphi_{m,n}(y)$ ontwikkelen volgens

$$(26) \quad \varphi_{m,n}(y) = \sum_{r=0}^{\infty} \mu_r \frac{(iy)^r}{r!}$$

dan is wegens (24) μ_r gelijk aan

$$\sum_{u=0}^{\infty} P(u,m,n) (u - \frac{mn}{2})^r$$

dus juist het r -de moment van de verdelingsfunctie $Q(u,m,n)$ t.o.v. het gemiddelde $u = \frac{mn}{2}$.

Wegens (25) en (26) hebben we

$$(27) \quad \frac{1}{\binom{m+n}{n}} \frac{\sin \frac{m+1}{2} y \dots \sin \frac{m+n}{2} y}{\sin \frac{1}{2} y \dots \sin \frac{n}{2} y} = \sum_{r=0}^{\infty} \frac{\mu_r}{r!} (iy)^r$$

Nu is $\frac{\sin z}{z} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \dots$

en $\frac{z}{\sin z} = 1 + \frac{1}{6} z^2 + \frac{7}{360} z^4 + \dots$

en uit (27) volgt nu

$$(28) \quad \sum_{r=0}^{\infty} \frac{\mu_r}{r!} (iy)^r = \prod_{k=1}^n \left\{ 1 - \frac{(m+k)^2}{24} y^2 + \frac{(m+k)^4}{1920} y^4 - \dots \right\} \\ \cdot \prod_{k=1}^n \left\{ 1 + \frac{k^2}{24} y^2 + \frac{7k^4}{5760} y^4 + \dots \right\}$$

hieruit volgt onder meer $\mu_0 = 1$, $\mu_{2r+1} = 0$ ($r=0,1,\dots$) en

$$-\frac{1}{2} \mu_2 = -\sum_{k=1}^n \frac{(m+k)^2}{24} + \sum_{k=1}^n \frac{k^2}{24}$$

ofwel $\mu_2 = \frac{1}{12} \left\{ \sum_{k=1}^{m+n} k^2 - \sum_{k=1}^m k^2 - \sum_{k=1}^n k^2 \right\}$

en wegens $\sum_{k=1}^p k^2 = \frac{1}{6} p(p+1)(2p+1)$ vinden we

$$(29) \quad \mu_2 = \frac{1}{12} mn(m+n+1);$$

analoog zouden we uit (28) de hogere momenten kunnen bepalen. De stochastische variabele u heeft dus een gemiddelde $\frac{mn}{2}$ en een spreiding

$$(30) \quad \sigma = \frac{1}{2} m \sqrt{\frac{1}{3} n \left(1 + \frac{n+1}{m}\right)}$$

De grootheid

$$(31) \quad v = \frac{u - \frac{mn}{2}}{\sigma}$$

heeft een gemiddelde 0, een spreiding 1, en heeft als karakteristieke functie t.o.v. het gemiddelde

$$\frac{\sin \frac{m+1}{m} \frac{y}{\sqrt{\frac{1}{3} n \left(1 + \frac{n+1}{m}\right)}} \dots \sin \frac{m+n}{m} \frac{y}{\sqrt{\frac{1}{3} n \left(1 + \frac{n+1}{m}\right)}}}{\binom{m+n}{n} \sin \frac{y}{m \sqrt{\frac{1}{3} n \left(1 + \frac{n+1}{m}\right)}} \dots \sin \frac{ny}{m \sqrt{\frac{1}{3} n \left(1 + \frac{n+1}{m}\right)}}}$$

Voor vastgehouden n en voor $m \rightarrow \infty$ convergeert de teller naar $\left(\sin \frac{y}{\sqrt{\frac{1}{3} n}}\right)^n$ en de noemer naar $\left(\frac{y}{\sqrt{\frac{1}{3} n}}\right)^n$ wegens $\binom{m+n}{n} \sim \frac{m^n}{n!}$. Hieruit volgt

$$(32) \quad \lim_{m \rightarrow \infty} \varphi_{m,n} \left(\frac{y}{\sigma} \right) = \left(\frac{\sin \frac{y}{\sqrt{\frac{1}{3} n}}}{\frac{y}{\sqrt{\frac{1}{3} n}}} \right)^n = \psi_n(y).$$

Nu is voor $n \rightarrow \infty$ $\frac{\sin \frac{y}{\sqrt{\frac{1}{3} n}}}{\frac{y}{\sqrt{\frac{1}{3} n}}} = 1 - \frac{y^2}{2n} + O\left(\frac{y^4}{n^2}\right)$

en dus volgt

$$(33) \quad \lim_{n \rightarrow \infty} \psi_n(y) = \lim_{\substack{m \rightarrow \infty \\ n \rightarrow \infty}} \varphi_{m,n} \left(\frac{y}{\sigma} \right) = e^{-y^2/2}$$

Er geldt nu de volgende stelling (vgl. Cramer, Mathematical methods of statistics, blz.96).

Zij $G_k(v)$ met $k=1,2,3,\dots$ een oneindige rij verdelingsfuncties met $\varphi_k(y)$ ($k=1,2,3,\dots$) als de respectievelijke karakteristieke functies, zodanig dat aan de volgende voorwaarden is voldaan

1) Voor elke reële y bestaat de limiet

$$\lim_{k \rightarrow \infty} \varphi_k(y) = \varphi(y)$$

2) $\varphi(y)$ is continu voor $y=0$. Dan geldt voor elke waarde v , waarvoor $G(v)$ continu is,

$$\lim_{k \rightarrow \infty} G_k(v) = G(v),$$

waarin $G(v)$ een verdelingsfunctie is met karakteristieke functie $\varphi(y)$

Verder merken we op, dat

$$\left(\frac{\sin \frac{y}{\sqrt{4/3n}}}{\frac{y}{\sqrt{4/3n}}} \right)^n$$

de karakteristieke functie is bij een verdelingsfunctie $F_n(v)$, die kan worden opgevat als de superpositie van n homogene verdelingen, elk met een gemiddelde nul en breedte $2 \cdot \sqrt{\frac{3}{n}}$. Bovendien is $e^{-y^2/2}$ de karakteristieke functie bij de normale verdeling $F(v)$ met gemiddelde 0 en spreiding 1.

Gebruik makend van (32) en (33) volgt uit deze opmerkingen dat voor de verdelingsfunctie

$$F_{m,n}(v) = Q \left(\frac{mn}{2} + \sigma v, m, n \right)$$

van $v = \frac{u - \frac{mn}{2}}{\sigma}$ de volgende formules gelden ¹⁾.

$$(34) \quad \lim_{m \rightarrow \infty} Q \left(\frac{mn}{2} + \sigma v, m, n \right) = F_n(v)$$

en

$$(35) \quad \lim_{\substack{m \rightarrow \infty \\ n \rightarrow \infty}} Q \left(\frac{mn}{2} + \sigma v, m, n \right) = F(v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^v e^{-t^2/2} dt$$

De laatste betrekking drukt uit, dat de verdelingsfunctie $Q(u, m, n)$ asymptotisch normaal is (voor $m \rightarrow \infty$ en $n \rightarrow \infty$). Wanneer m en n groot zijn, en wel ongeveer van de zelfde orde, dan zal in goede benadering gelden

$$(36a) \quad Q(u, m, n) \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^v e^{-t^2/2} dt$$

1) De formules (34) en (35) volgen ook uit algemene stellingen van B.L. van der Waerden, toegepast op het bijzondere geval van gelijkverdeling vgl. Rapport T.W. nr.6 (stellingen 4 en 5).

waarin

$$(36 \text{ b}) \quad v = \frac{u - \frac{mn}{2} + \frac{1}{2}}{\sigma} = \frac{u - \frac{mn}{2} + \frac{1}{2}}{\sqrt{\frac{1}{12}mn(m+n+1)}}$$

(de term $+\frac{1}{2}$ is toegevoegd, omdat een normale verdeling, die de "trapvormige" verdelingsfunctie $Q(u, m, n)$ goed benadert, de "treden" ongeveer in het midden zal snijden, terwijl we juist een goede benadering willen hebben in de sprongpunten).

Geldt daarentegen $m \gg n$, dan moeten we verwachten, dat in goede benadering geldt $Q(u, m, n) \sim F_n(v)$, waarin v weer door (36 b) wordt gegeven.

Zij $H_n(w)$ de verdelingsfunctie van $w = w_1 + w_2 + \dots + w_n$, waarin $w_1, w_2, \dots, w_n$ onafhankelijk stochastische variabelen zijn met waarschijnlijkheidsdichtheid

$$f(w) = \begin{cases} 1 & \text{voor } 0 \leq w \leq 1 \\ 0 & \text{voor } w < 0 \text{ en } w > 1. \end{cases}$$

Dan geldt de formule (vgl. Cramer l.c. blz. 245)

$$(37) \quad H_n(w) = \frac{1}{n!} \sum_{0 \leq k \leq w} (-1)^k \binom{n}{k} (w-k)^n$$

Wanneer een stochastische variabele v de verdelingsfunctie $F_n(v)$ heeft, dan heeft de grootheid

$$(38) \quad w = \frac{v + n \sqrt{\frac{3}{n}}}{2 \sqrt{\frac{3}{n}}}$$

de verdelingsfunctie $H_n(w)$ (d.w.z. $F_n(v) = H_n(w)$), zoals direct volgt uit de definitie van $F_n(v)$.

Substitueren we (36 b) in (38) dan vinden we

$$(39) \quad w = \frac{n}{2} + \frac{u - \frac{nm}{2} + \frac{1}{2}}{\sqrt{m(m+n+1)}}$$

Voor $m \gg n$ zal (37) een goede benadering zijn van $Q(u, m, n)$, wanneer w door (39) wordt gegeven.

Opm. Rekening houdend met de definitie (16) van $a_{n,h}$ volgt uit (15) de exact geldende formule

$$Q(u, m, n) = \frac{1}{\binom{m+n}{n}} \left[a_{n,u} - \sum_{h=1}^n a_{n,u-m-h} + \sum_{1 \leq h < k \leq n} a_{n,u-2m-h-k} - \dots \right]$$

Wanneer $m \gg n$ dan vervangen we eens deze formule ruwweg door

$$(40) \quad Q(u, m, n) \sim \frac{1}{\binom{m+n}{n}} \left[a_{n,u} - n a_{n,u-m} + \binom{n}{2} a_{n,u-2m} - \dots \right]$$

Nu volgt uit (19) voor $h \rightarrow \infty$

$$a_{n,h} = \left(\frac{1}{n!} \right)^2 h^n + O(h^{n-1}) \quad \text{en wegens } \binom{m+n}{n} \sim \frac{m^n}{n!} \text{ volgt uit}$$

(40) ruw gesproken, dat

$$Q(u, m, n) \sim \frac{1}{n!} \sum_{0 \leq k < \frac{u}{m}} (-1)^k \binom{n}{k} \left(\frac{u}{m} - k \right)^n$$

Deze zelfde formule verkrijgt men uit (37), als men opmerkt dat voor $m \gg n$ uit (39) volgt $w \sim \frac{u}{m}$. De boven gehouden heuristische redenering versterkt dus de verwachting, dat de benadering $Q(u, m, n) \sim H_n(w)$, waarin w door (39) wordt gegeven, bruikbaar zal zijn.

In de volgende tabel is voor $m=10$ en $n=3$ de exacte waarde $Q(u, 10, 3)$ aangegeven evenals de benaderde waarden zoals die zouden volgen uit de formules (36) en (37). Het blijkt dat de benadering (37) beter voldoet.

$Q(u, 10, 3)$					
u	exact	met (36)	fout $\times 10^4$	met (37)	fout $\times 10^4$
0	0,0035	0,0071	+36	0,0034	-1
1	0,0070	0,0112	+42	0,0077	+7
2	0,0140	0,0173	+33	0,0145	+5
3	0,0245	0,0260	+15	0,0245	0
4	0,0385	0,0380	-5	0,0383	-2
5	0,0559	0,0541	-18	0,0565	+6
6	0,0804	0,0754	-50	0,0796	+8
7	0,1084	0,1024	-60	0,1083	-1
8	0,1433	0,1359	-75	0,1432	-1
9	0,1853	0,1762	-91	0,1849	-4
10	0,2343	0,2234	-109	0,2331	-12
11	0,2867	0,2771	-96	0,2868	+1
12	0,3462	0,3363	-99	0,3447	+15
13	0,4056	0,3999	-57	0,4056	0
14	0,4685	0,4663	-22	0,4683	-2
15	0,5315	0,5337	+22	0,5317	+2

Zij ε een positief getal ≤ 1 , en zij $u_\varepsilon(m, n)$ de gehele waarde u waarvan $Q(u, m, n)$ de waarde ε het meest nabij komt. Zij $w_\varepsilon(n)$ de waarde, waarvoor $H_n(w_\varepsilon) = \varepsilon$. Een goede benadering van u_ε is dan wegens (37) en (39)

$$(41) \quad u_\varepsilon(m, n) \sim \frac{nm-1}{2} - \sqrt{m(m+n+1)} \left(\frac{n}{2} - w_\varepsilon(n) \right)$$

Analoog als v_ε de waarde is, waarvoor

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v_\varepsilon} e^{-t^2/2} dt = \varepsilon$$

dan volgt uit (36) de benadering

$$(42) \quad u_\varepsilon(m, n) \sim \frac{nm-1}{2} + v_\varepsilon \sqrt{\frac{1}{12} mn(m+n+1)}$$

Voor de numerieke berekening is formule (42) het gemakkelijkst. Evenwel mag zij alleen gebruikt worden als m en n ongeveer van dezelfde grootteorde zijn. Voor $n=3$ en $m=40$ geeft formule (42) bijvoorbeeld

$$U \frac{1}{10} \sim 41,263$$

en men zou dus verwachten, dat $U \frac{1}{10} = 41$, terwijl wegens $Q(40) = 0,097413$ en $Q(41) = 0,10429$ in werkelijkheid geldt $U \frac{1}{10} = 40$. Formule (41) geeft in dit geval $U \frac{1}{10} \sim 40,3838$ en levert dus $U \frac{1}{10} = 40$ en duidt bovendien aan, dat

$$\frac{Q(41) - \varepsilon}{\varepsilon - Q(40)} \sim \frac{0,6162}{0,3838} = 1,61$$

(het linkerlid is in werkelijkheid 1,66). In alle door mij gecontroleerde getallen bleek formule (41) het juiste resultaat te geven.

4. Het geval van verschoven homogene verdelingen.

Onder $F(x, \mu)$ zullen we de volgende verdelingsfunctie verstaan

$$(43) \quad F(x, \mu) = \begin{cases} 0 & \text{als } x \leq \mu \\ x - \mu & \text{" } \mu \leq x \leq 1 + \mu \\ 1 & \text{" } 1 + \mu \leq x. \end{cases}$$

Laten $x_1, \dots, x_m$ stochastische variabelen zijn, elk met de waarschijnlijkheidsverdeling $F(x, 0)$, en analoog $y_1, \dots, y_n$ stochastische variabelen elk met een waarschijnlijkheidsverdeling $F(y, \mu)$. Voor een steekproef $x_1, \dots, x_m; y_1, \dots, y_n$ stellen we

$$\begin{aligned} \delta_{i,j} &= +1 & \text{en } \varepsilon_{i,j} &= +1 & \text{als } y_j < x_i \\ \delta_{i,j} &= 0 & \text{en } \varepsilon_{i,j} &= -1 & \text{als } y_j > x_i \end{aligned}$$

$$u = \sum_{i,j} \delta_{i,j}$$

$$(44) \quad v = \sum_{i,j} \varepsilon_{i,j}$$

Wegens $\varepsilon_{i,j} = 2 \delta_{i,j} - 1$ geldt $v = 2u - mn$. Wanneer we niet de waarschijnlijkheidsverdeling van u , doch die van v weergeven, dan worden de formules iets eleganter. Zij $p(v, m, n)$ resp. $q(v, m, n)$ de kans, dat het rechterlid van (44) de waarde v resp. hoogstens de waarde v aanneemt. Deze grootheden zijn natuurlijk ook van μ afhankelijk. Verder geldt

$$(45) \quad q(v, m, n) = \sum_{v'=-\infty}^v p(v', m, n)$$

In het geval, dat de elementen x_i en y_j gelijk verdeeld zijn geldt voor $v = 2u - mn$ dat $p(v, m, n) = P(u, m, n)$ en $q(v, m, n) = Q(u, m, n)$.

Uit (14) en (15) volgt dus in het geval van gelijkverdeling

$$(46) \quad \sum_{v=-\infty}^{\infty} p(v, m, n) z^v = \phi_{m, n}(z)$$

$$(47) \quad \sum_{v=-\infty}^{\infty} q(v, m, n) z^v = \frac{\phi_{m, n}(z)}{1-z}$$

waarin

$$(48) \quad \phi_{m, n}(z) = \frac{1}{\binom{m+n}{n}} \frac{(z^{m+1} - z^{-m-1}) \dots (z^{m+n} - z^{-m-n})}{(z - z^{-1}) \dots (z^m - z^{-m})}$$

Beschouwen we nu het geval dat de grootheden x_i resp. y_j stochastisch verdeeld zijn met verdelingsfunctie $F(x, 0)$ resp. $F(y, \mu)$. Als $\mu \geq +1$ (resp. $\mu \leq -1$) dan is steeds $v = -mn$ (resp. $v = +mn$) en valt er niets te onderzoeken. We veronderstellen dus, dat $|\mu| < 1$ en stellen dan

$$(49) \quad q = |\mu| \quad \text{en} \quad p = 1 - q$$

De waarschijnlijkheid, dat r elementen x_i en s elementen y_j (met $0 \leq r \leq m$ en $0 \leq s \leq n$) gelegen zijn in het interval $[0, p]$ wordt gegeven door

$$\binom{m}{r} p^r q^{m-r} \quad \binom{n}{s} p^s q^{n-s}.$$

In dit geval leveren de $(mn - rs)$ paren x_i, y_j waarvoor of x_i of y_j buiten het interval $[0, p]$ ligt een bijdrage tot $v = \sum_{i,j} \varepsilon_{i,j}$ die gelijk is aan $\gamma(rs - mn)$, waarin

$$(50) \quad \begin{aligned} \gamma &= +1 & \text{als } \mu \geq 0 \\ \gamma &= -1 & \text{" } \mu \leq 0 \end{aligned}$$

In het interval $[0, p]$ nemen zowel de stochastische variabelen x_i als y_j elke waarde met even grote waarschijnlijkheidsdichtheid aan. In de veronderstelling dat r elementen x_i en s elementen y_j in het interval $[0, p]$ liggen wordt de voorwaardelijke kans dat $\sum \varepsilon_{i,j}$ de waarde v aanneemt dus gegeven door de coëfficiënt van z^v in

$$\phi_{r,s}(z) \cdot z^{\gamma(rs - mn)}$$

waarin $\phi_{r,s}(z)$ door (48) wordt gegeven. We vinden nu

$$(51) \quad \sum_{v=-\infty}^{\infty} p(v, m, n) z^v = \sum_{r=0}^m \sum_{s=0}^n \binom{m}{r} \binom{n}{s} p^{r+s} q^{m+n-r-s} \phi_{r,s}(z) z^{\gamma(rs - mn)}$$

en analoog wegens (45)

$$(52) \quad \sum_{v=-\infty}^{\infty} q(v, m, n) z^v = \sum_{r=0}^m \sum_{s=0}^n \binom{m}{r} \binom{n}{s} p^{r+s} q^{m+n-r-s} \frac{\phi_{r,s}(z) z^{\gamma(rs - mn)}}{1-z}$$

Zij $\beta(v, m, n)$ de z.g. "powerfunction" gedefinieerd door

$$(53) \quad \beta(v, m, n) = \sum_{v'=-\infty}^v (p(v', m, n) + p(-v', m, n)).$$

Dit is dus de kans, dat $\sum \varepsilon_{i,j}$ een der waarden $\pm mn, \pm(mn-1), \dots, \pm v$

aanneemt (wegens $v=2u-mn$ worden overigens alleen waarden v aangenomen met $mn-v$ even).

Wegens $\phi(z) = \phi(z^{-1})$ geldt

$$\sum_{v=-\infty}^{\infty} p(-v, m, n) z^v = \sum_{r=0}^m \sum_{s=0}^n \binom{m}{r} \binom{n}{s} p^{r+s} q^{m+n-r-s} \phi_{r,s}(z) z^{-\delta(r,s)}$$

en wegens (50) hebben we nu zowel van $\mu \geq 0$ als $\mu \leq 0$

$$\sum_{v=-\infty}^{\infty} \{p(v, m, n) + p(-v, m, n)\} z^v = \sum_{r=0}^m \sum_{s=0}^n \binom{m}{r} \binom{n}{s} p^{r+s} q^{m+n-r-s} \frac{\phi_{r,s}(z)}{1-z} [z^{rs-mn} + z^{-rs+mn}]$$

en uit (53) volgt nu

$$(54) \quad \sum_{v=-\infty}^{\infty} \beta(v, m, n) z^v = \sum_{r=0}^m \sum_{s=0}^n \binom{m}{r} \binom{n}{s} p^{r+s} q^{m+n-r-s} \frac{\phi_{r,s}(z)}{1-z} [z^{rs-mn} + z^{-rs+mn}]$$

In verband met (49) is de "powerfunctie" $\beta(v, m, n)$ dus een even functie van μ , en wel een polynoom in μ^2 met een graad $\leq \left\lfloor \frac{m+n}{2} \right\rfloor$.

Uit (54) volgt bijvoorbeeld voor de coëfficiënt

$$g_v = \left[\frac{d^2}{d\mu^2} \beta(v, m, n) \right]_{\mu=0}$$

$$\text{dat } \sum_{v=-\infty}^{\infty} g_v z^v = 2 \binom{m+n}{n} \frac{\phi_{m,n}}{1-z} - (m+n-1) \left[\frac{\phi_{m-1,n}}{1-z} + n \frac{\phi_{m,n-1}}{1-z} \right] (z + z^{-1})$$

$$+ \binom{m}{2} \frac{\phi_{m-2,n}}{1-z} + mn \frac{\phi_{m-1,n-1}}{1-z} + \binom{n}{2} \frac{\phi_{m,n-2}}{1-z} (z^2 + z^{-2})$$

waaruit g_v zonder veel moeite is te berekenen.

5. Appendix.

We beschouwen het algemene probleem genoemd in de inleiding, waarin $F(x)$ en $G(y)$ voorlopig nog willekeurig. Voor het gemiddelde $\bar{u}$ en de spreiding σ van de waarschijnlijkheidsverdeling van het aantal inversies ¹⁾

$$u = \sum_{i,j} \delta_{i,j}$$

heeft Prof. B.L. van der Waerden de volgende formules gevonden (Rapport T.W. nr.6, blz.11)

$$(55) \quad \bar{u} = mn \alpha$$

$$(56) \quad \sigma^2 = \overline{(u-\bar{u})^2} = mn \{ \alpha \beta + (m-1)(A - \alpha^2) + (n-1)(B - \alpha^2) \}$$

waarin

$$(57) \quad \alpha = 1 - \beta = \int_{-\infty}^{\infty} \{G(x)\}^2 dF(x)$$

$$(58) \quad A = \int_{-\infty}^{\infty} \{1-F(x)\}^2 dG(x)$$

1) Voor de definitie van $\delta_{i,j}$ zie begin van laatste paragraaf.

$$(59) \quad B = \int_{-\infty}^{+\infty} \{G(x)\}^2 dF(x)$$

Daar de verdelingsfunctie van u voor grote waarden van m en n na-
genoeg normaal is (vgl. Rapport T.W. nr.6) zijn de grootheden $\bar{u}$
en G^2 (vooral voor grote waarden m en n) van bijzonder belang.

Lemma 1. Als de reële getallen λ en μ bestaan met

$$F(y) + F(\lambda - y) = 1$$

en

$$G(x) = F(x - \mu)$$

dan volgt $A = B$.

Bewijs. Stellen we $x = \lambda + \mu - \xi$ dan hebben we

$$\begin{aligned} A &= \int_{-\infty}^{+\infty} \{1 - F(x)\}^2 dG(x) = \int_{-\infty}^{+\infty} \{F(\lambda - x)\}^2 dF(x - \mu) \\ &= \int_{-\infty}^{+\infty} \{F(\xi - \mu)\}^2 dF(\lambda - \xi) = \int_{-\infty}^{+\infty} \{F(\xi - \mu)\}^2 dF(\xi) = B \end{aligned}$$

Lemma 2. Onder de voorwaarden van lemma 1 is G^2 een even functie van μ .

Bewijs. Stellen we

$$\xi = \lambda - x \quad \text{en} \quad \eta = \lambda - y$$

dan komt met een steekproef $x_1, \dots, x_m; y_1, \dots, y_n$ met u inversies een
steekproef $\xi_1, \dots, \xi_m; \eta_1, \dots, \eta_n$ cveeren met $u' = mn - u$ inversies

De respectievelijke verdelingsfuncties van u en u' hebben dus dezelfde
spreiding. Elk der variabelen ξ_i heeft een verdelingsfunctie

$$1 - F(\lambda - \xi) = F(\xi)$$

en elk der variabelen heeft een verdelingsfunctie

$$1 - F(\lambda - \eta - \mu) = F(\eta + \mu)$$

waarmee het gestelde bewezen is.

In de twee volgende gevallen, waarin we G^2 zullen berekenen is
inderdaad aan de voorwaarden van lemma 1 voldaan en geldt dus

$$(60) \quad G^2 = mn \left\{ \alpha\beta + (m+n-2)(A - \alpha^2) \right\}$$

Wegens lemma 2 mogen we ons bij de berekening beperken tot het geval
 $\mu \geq 0$.

I. Beschouw vooreerst het geval, dat

$$F(x) = F(x, 0) \quad \text{en} \quad G(x) = F(x, \mu)$$

waarin $F(x, \mu)$ volgens (43) is gedefinieerd.

Voor $0 \leq \mu \leq 1$ geldt nu

$$A = \int_{\mu}^1 (x - \mu)^2 dx = \frac{1}{3} (1 - \mu)^3$$

$$\text{en} \quad (61) \quad \alpha = \int_{\mu}^1 (x - \mu) dx = \frac{1}{2} (1 - \mu)^2$$

waaruit met (56) en lemma 2 volgt voor $|\mu| = q \leq 1$

$$(62) \quad G^2 = \frac{mn}{12} \left[6q^2(1-q)^2 + (m+n+1)(1-q)^3(1+3q) \right].$$

Uit (55) en (61) volgt voor $0 \leq \mu \leq 1$

$$\bar{u} = \frac{mn}{2} (1 - \mu)^2$$

en analoog vinden we voor $-1 \leq \mu \leq 0$

$$\bar{u} = \frac{mn}{2} (1 - \mu^2)$$

II. We beschouwen nu het geval van verschoven normale verdelingen, waarin

$$(63) \quad F(x) = \Phi(x) \quad \text{en} \quad G(x) = \Phi(x - \mu).$$

Hierin is

$$(64) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

dus een normale verdeling met gemiddelde 0 en spreiding 1. Uit (55) volgt

$$\alpha = \int_{-\infty}^{\infty} \Phi(x - \mu) d\Phi(x) = \left(\frac{1}{\sqrt{2\pi}} \right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{x-\mu} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) dx dy$$

ofwel

$$(65) \quad \alpha = 1 - \Phi\left(\frac{1}{2}\mu\sqrt{2}\right) = \Phi\left(-\frac{1}{2}\mu\sqrt{2}\right)$$

en wegens (55) is dus

$$(66) \quad \bar{u} = mn \Phi\left(-\frac{1}{2}\mu\sqrt{2}\right).$$

In verband met (60) kunnen we de berekening van de spreiding herleiden tot de berekening van

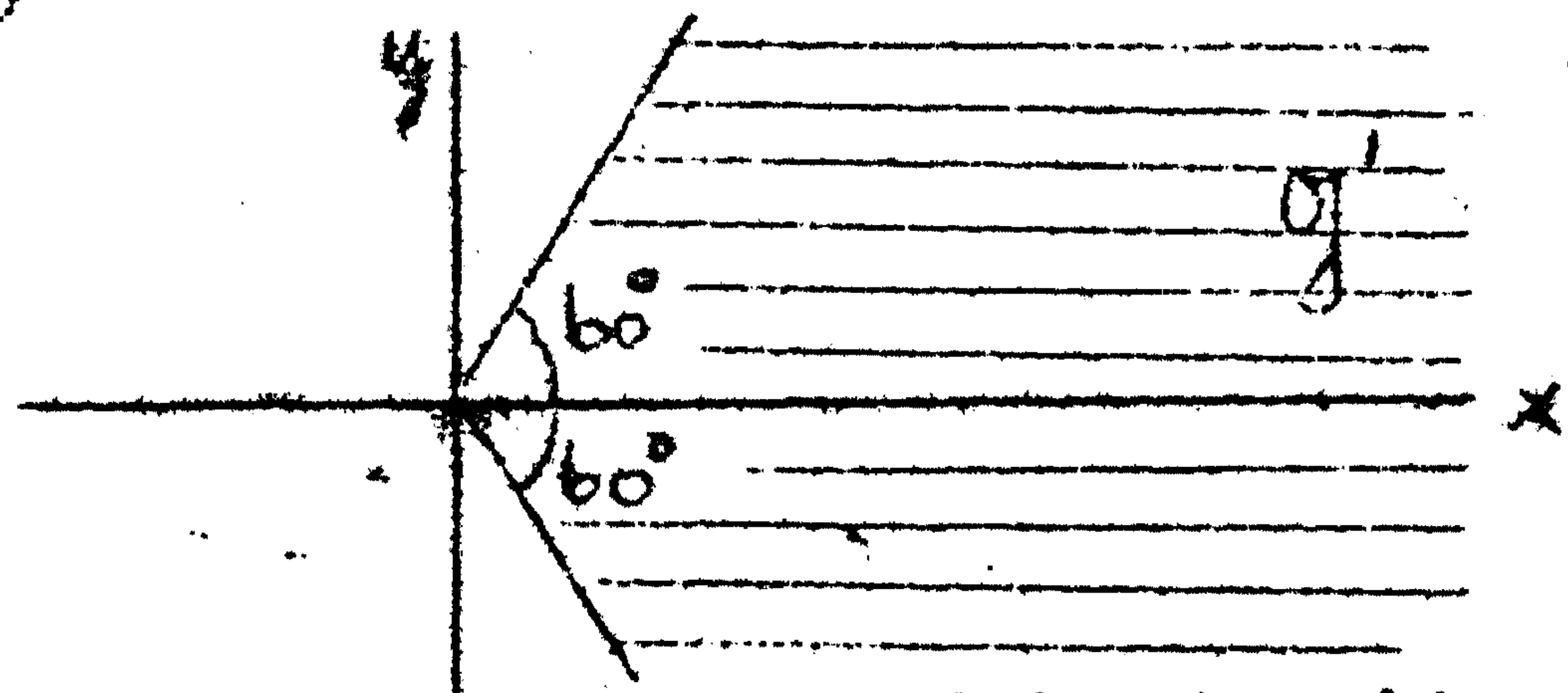
$$A = \int_{-\infty}^{\infty} \{1 - F(x)\}^2 dG(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} \right)^3 \exp\left(-\frac{1}{2}(\xi^2 + \eta^2 + (x - \mu)^2)\right) d\xi d\eta dx$$

Het integratiegebied is een tweevlakshoek ingesloten door de vlakken

$\xi = x$ en $\eta = x$, en wel is de standhoek gelijk aan 120° . De ribbe is de lijn $\xi = \eta = x$. Als we door een orthogonale transformatie deze lijn tot coördinaatsas maken, dan vinden we

$$(67) \quad A = \left(\frac{1}{\sqrt{2\pi}} \right)^3 \int_{-\infty}^{\infty} dz \int_{\mathcal{G}} dx dy \exp - \frac{1}{2} \left\{ (x - \mu\sqrt{\frac{2}{3}})^2 + y^2 + z^2 \right\}$$

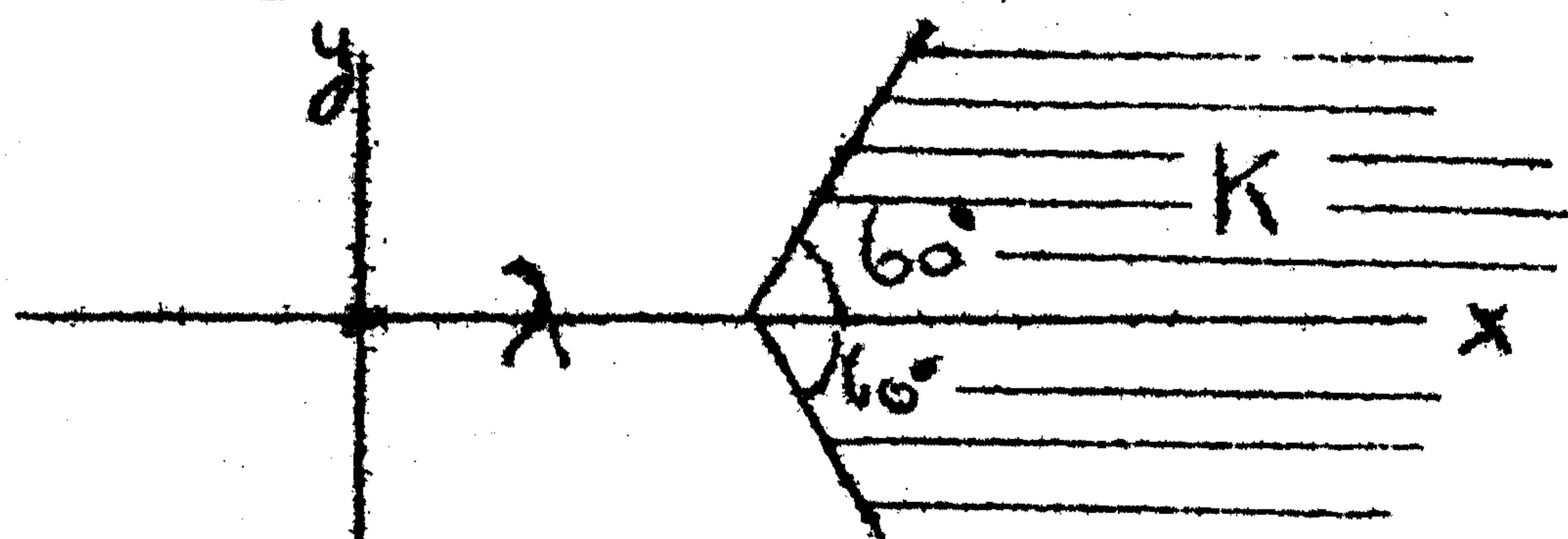
waarin we onder $\mathcal{G}$ het volgende gearceerd gebied in het x-y-vlak verstaan



De integratie naar z kunnen we direct uitvoeren. Stellen we voor $\lambda \geq 0$

$$(68) \quad C(\lambda) = \frac{1}{2\pi} \int_{\mathcal{K}} e^{-\frac{1}{2}(x^2 + y^2)} dx dy$$

waarin $\mathcal{K}$ het volgende gebied aangeeft



dan volgt uit (67) met de substitutie $x' = x + \mu\sqrt{\frac{2}{3}}$ dat voor $\mu \geq 0$

$$(69) \quad A = C\left(\mu\sqrt{\frac{2}{3}}\right)$$

" I " " DGH " " 90°

" I₂ " " FOH " " 30°

" Δ de driehoek EOG.

Dan geldt

$$K_1 = I_1 - I_2 + \underline{\Delta}$$

Nu is

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dx dy = \frac{1}{2} \Phi\left(-\frac{1}{2} \lambda \sqrt{3}\right)$$

$$\frac{1}{2\pi} \int_{I_2} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dx dy = \frac{1}{12}$$

waaruit volgt

mit folgt

$$(70) \quad c(\lambda) = 2 \cdot \frac{1}{2\pi} \int_{K_1} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dx dy$$

$$= \Phi\left(-\frac{1}{2}\lambda\sqrt{3}\right) - \frac{1}{6} + \frac{1}{\pi} \int_{K_1} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dx dy.$$

Hierin is

$$J = \frac{1}{\pi} \int_{\Delta} \exp\left(-\frac{1}{2}(x^2+y^2)\right) dx dy = \frac{1}{\pi} \int_0^{13} \int_0^{\frac{2\lambda}{\cos\varphi}} \exp\left(-\frac{1}{2}r^2\right) r dr d\varphi$$

$$= \frac{1}{\pi} \int_0^{13} \left(1 - \exp\left(-\frac{1}{2} \frac{\lambda^2}{\cos^2\varphi}\right)\right) d\varphi$$

waaruit volgt voor willekeurige waarden λ

$$J = \frac{1}{\pi} \int_0^{\pi/3} \left[\frac{3}{8} \frac{\lambda^2}{\cos^2 \varphi} - \frac{1}{2!} \left(\frac{3}{8} \right)^2 \frac{\lambda^4}{\cos^4 \varphi} + \dots + \frac{(-1)^{n-1}}{n!} \left(\frac{3}{8} \right)^n \frac{\lambda^{2n}}{\cos^{2n} \varphi} + \dots \right] d\varphi$$

of wel

$$(71) \quad J = \frac{1}{n} \sum_{n=1}^{\infty} (-1)^{n-1} \left(\frac{3}{8}\right)^n \frac{I_n}{n!} \lambda^{2n}$$

met

et

(72)
$$I_n = \int_0^{\pi/3} \frac{d\varphi}{\cos^{2n} \varphi}.$$

Wegens

$$\frac{d}{d\psi} \left[\frac{\sin \psi}{\cos^{2n-1} \psi} \right] = \frac{1}{\cos^{2n-2} \psi} + (2n-1) \frac{\sin^2 \psi}{\cos^{2n} \psi} - (2n-2) \frac{1}{\cos^{2n-2} \psi}$$

geldt de recursieformule

$$x^{2n-1} \sqrt{x} = (2n-1) I_n - (2n-2) I_{n-1} .$$

Daar $I_1 = \sqrt{3}$ volgt nu met inductie

$$(73) \quad I_n = \frac{\sqrt{3}}{2n-1} \left[\lambda^{2n-2} + \sum_{k=1}^{n-1} \frac{(2n-2)(2n-4)\dots(2n-2k)}{(2n-3)(2n-5)\dots(2n-2k-1)} \lambda^{2n-2k-2} \right]$$

voor $n = 1, 2, 3, \dots$

Breekt men de reeks (71) na N termen af, dan volgt voor de bovengrens van de gemaakte fout

$$\Delta_N = \left| J - \sum_{n=1}^{\infty} (-1)^{n-1} \frac{I_n}{n!} \left(\frac{3}{8} \lambda^2 \right)^n \right| = \frac{1}{\pi} \int_0^{\pi/6} \left(\frac{3}{8} \frac{\lambda^2}{\cos^2 \varphi} \right)^{N+1} e^{-\frac{3}{8} \theta \frac{\lambda^2}{\cos^2 \varphi}} d\varphi$$

waarin $0 \leq \theta(\varphi) \leq 1$. Dus is

$$(74) \quad \Delta_N < \frac{1}{\pi} \frac{\left(\frac{3}{8} \lambda^2 \right)^{N+1}}{(N+1)!} I_{N+1}$$

en de gemaakte fout is dus kleiner dan de eerst verwaarloosde term. Uit (73) vinden we als ruwe bovengrens

$$I_N \leq \sqrt{3} \sum_{k=0}^{n-1} \lambda^{2n-2k-1} \leq \frac{1}{3} \sqrt{3} 4^n$$

en wegens (74) is voor $\lambda \leq \sqrt{2/3}$ de reeks (71) even goed bruikbaar voor de berekening van J , als de reeks $\sum_{n=0}^{\infty} \frac{1}{n!}$ geschikt is voor de berekening van e .

Uit (70) en (71) volgt nu voor $\lambda \geq 0$.

$$(75) \quad c(\lambda) = \Phi\left(-\frac{1}{2} \lambda \sqrt{3}\right) - \frac{1}{6} + \frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{I_n}{n!} \left(\frac{3}{8} \lambda^2 \right)^n$$

en wegens (69) vinden we voor $\mu \geq 0$

$$(76) \quad A = \Phi\left(-\frac{1}{2} \mu \sqrt{2}\right) - \frac{1}{6} + \frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{I_n}{n!} \left(\frac{1}{4} \mu^2 \right)^n$$

Wegens (60) en (65) hebben we nu

$$(77) \quad G^2 = mn \left[\alpha(1-\alpha) + (m+n-2) \left\{ \alpha(1-\alpha) - \frac{1}{6} + \frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{I_n}{n!} \left(\frac{1}{4} \mu^2 \right)^n \right\} \right]$$

waarin $\alpha = \Phi\left(-\frac{1}{2} \mu \sqrt{2}\right)$ en waarin I_n door (73) wordt gegeven. De afleiding van (77) geldt alleen voor $\mu \geq 0$.

In verband met lemma 2 en

$$\Phi(x) + \Phi(-x) = 1$$

zijn beide leden van (77) even functies van μ en geldt (77) dus voor willekeurige waarden μ . Breken we de oneindige som in het rechterlid van (77) af na de N^{de} term dan maken we een fout die kleiner is dan de eerst verwaarloosde term.

Voor $\mu = 0$ is $\alpha = \frac{1}{2}$ en gaat (77) over in $G^2 = mn \left[\frac{1}{4} + (m+n-2) \left(\frac{1}{4} - \frac{1}{6} \right) \right] = \frac{mn(m+n+1)}{12}$ dus juist de formule (29) (aldaar is $\mu_2 = G^2$).

$C(\lambda)$ voor grote waarden λ

In de door (68) gedefinieerde integraal $C(\lambda)$ transformeren we de integratievariabelen op poolcoördinaten

$$x = r \cos \varphi$$

$$y = r \sin \varphi$$

De grens van het integratiegebied wordt dan gegeven door

$$r(\varphi) = \frac{\frac{1}{2} \lambda \sqrt{3}}{\sin(\pi/3 - \varphi)}$$

en we vinden nu

$$\begin{aligned} C(\lambda) &= \frac{1}{2\pi} \int_K e^{-\frac{1}{2}x^2} dx dy = \frac{1}{\pi} \int_0^{\pi/3} \int_0^{r(\varphi)} e^{-\frac{1}{2}r^2} r d\varphi dr \\ &= \frac{1}{\pi} \int_0^{\pi/3} \exp\left(-\frac{3}{8} \frac{\lambda^2}{\sin^2(\pi/3 - \varphi)}\right) d\varphi = \frac{1}{\pi} \int_0^{\pi/3} \exp\left(-\frac{3}{8} \frac{\lambda^2}{\sin^2 \psi}\right) d\psi \end{aligned}$$

waarbij we de substitutie $\psi = \frac{\pi}{3} - \varphi$ hebben toegepast. Stellen we nu

$$t = \frac{1}{\sin^2 \psi} - \frac{1}{\sin^2 \frac{\pi}{3}} = \frac{1}{\sin^2 \psi} - \frac{4}{3}$$

dan volgt

$$\begin{aligned} (78) \quad C(\lambda) &= \frac{1}{\pi} e^{-\frac{1}{2}\lambda^2} \int_{-\infty}^0 \exp\left(-\frac{3}{8}\lambda^2 t\right) d\left(\text{bg} \sin \sqrt{\frac{1}{t+\frac{4}{3}}}\right) \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}\lambda^2} \int_0^{\infty} \exp\left(-\frac{3}{8}\lambda^2 t\right) \frac{dt}{\left(t+\frac{4}{3}\right) \sqrt{t+\frac{4}{3}}} \end{aligned}$$

Nu geldt voor $t \geq 0$

$$(79) \quad \frac{1}{\left(t+\frac{4}{3}\right) \sqrt{t+\frac{4}{3}}} = a_0 + a_1 t + \dots + a_{n-1} t^{n-1} + B_n t^n$$

(met $a_0 = 3/4 \sqrt{3}$, $a_1 = -\frac{27}{16} \sqrt{3}$ enz.)

waarin voor zekere ξ met $0 \leq \xi \leq t$ geldt

$$B_n = \frac{1}{n!} \frac{d^n}{d\xi^n} \frac{1}{\left(\xi+\frac{4}{3}\right) \sqrt{\xi+\frac{4}{3}}} = \frac{1}{2\pi i} \int_C \frac{1}{\left(t+\frac{4}{3}\right) \sqrt{t+\frac{4}{3}}} \frac{dt}{(t-\xi)^{n+1}}$$

waarin de integratieweg C een in positieve zin omlopen kleine cirkel om het punt $t = \xi$ voorstelt (gelegen in het complexe t -vlak). Voor $n \geq 0$ mag men deze cirkel deformeren in de (van beneden naar boven doorlopen) lijn

$$t = -\frac{1}{3} + i y \quad ; \quad -\infty < y < +\infty$$

en voor $\xi \geq 0$ vinden we dan de bovengrens

$$|B_n| \leq \frac{1}{2\pi} \int_{-\frac{1}{3}-i\infty}^{-\frac{1}{3}+i\infty} \frac{1}{\left|t+\frac{4}{3}\right| \left|t+\frac{1}{3}\right|^{\frac{1}{2}}} \frac{d|t|}{|t|^{n+1}}$$

Hierin is voor $t = -\frac{1}{3} + iy$

$$|t + \frac{4}{3}| = \sqrt{y^2 + 1} \quad \text{en} \quad |t| \geq \frac{1}{3}$$

en dus volgt

$$(80) \quad |B_n| \leq \frac{3^{n+1}}{\pi} \int_0^\infty \frac{dy}{\sqrt{y(y^2+1)}}$$

Substitutie van (79) in (78) levert nu voor $\lambda \geq 0$

$$C(\lambda) = \frac{1}{2\pi i} e^{-\frac{1}{2}\lambda^2} \left[\int_0^\infty \exp(-\frac{3}{8}\lambda^2 t) \sum_{k=0}^{n-1} a_k t^k dt + R_n \right]$$

ofwel

$$(81) \quad C(\lambda) = \frac{1}{2\pi i} e^{-\frac{1}{2}\lambda^2} \left[\sum_{k=0}^{n-1} \frac{k! a_k}{(\frac{3}{8}\lambda^2)^{k+1}} + R_n \right]$$

waarin de coëfficiënten a_k uit (79) volgen en waarin

$$R_n = \int_0^\infty \exp(-\frac{3}{8}\lambda^2 t) B_n(t) t^n dt$$

Uit (80) volgt nu dat

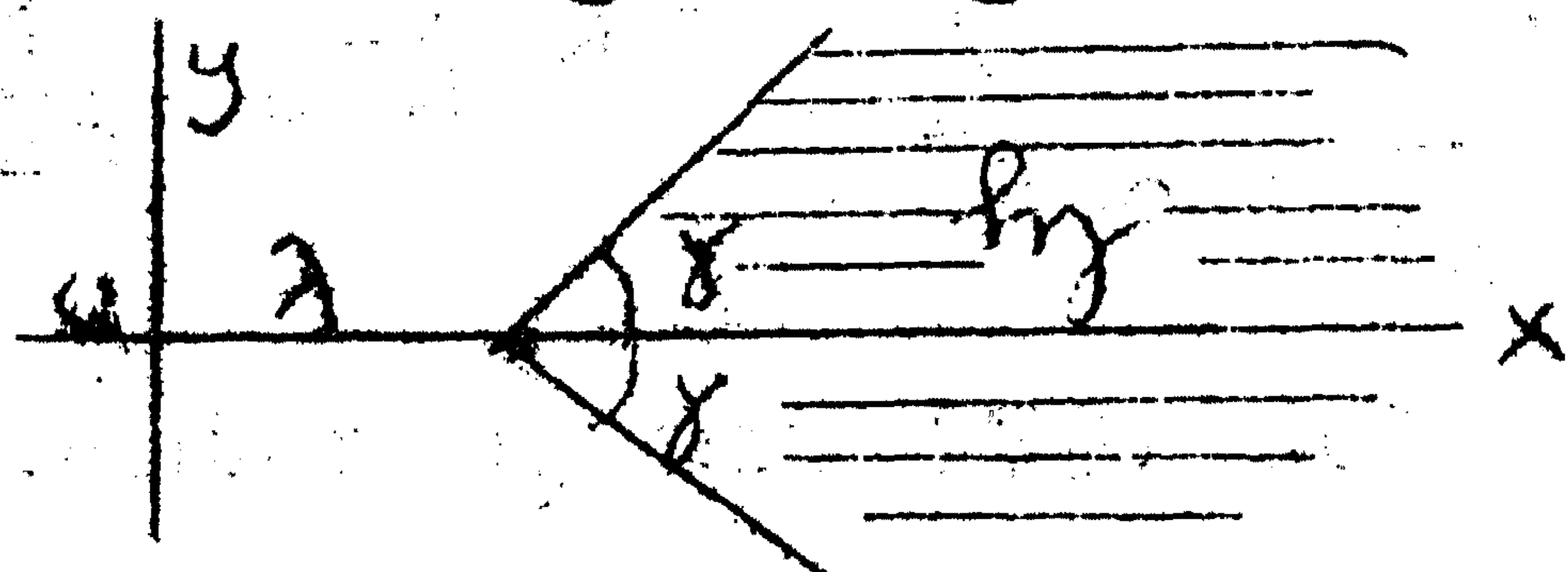
$$(82) \quad |R_n| \leq C 3^{n+1} \frac{n!}{(\frac{3}{8}\lambda^2)^{n+1}}$$

waarin

$$(83) \quad C = \frac{1}{\pi} \int_0^\infty \frac{dy}{\sqrt{y(y^2+1)}}$$

Wegens (82) levert dus de ontwikkeling (81) een asymptotische ontwikkeling van $C(\lambda)$ voor grote waarden van λ . Daar de oneindig voortgezette reeks (79) een convergentiestraal $\frac{1}{3}$ heeft is wegens (82) de orde van de fout (voor $n \rightarrow \infty$) ongeveer de orde van de eerst verwaarloosde term.

Opm. Beschouw het volgende gebied in het x-y-vlak



waarbij $0 < \gamma < \frac{\pi}{2}$ en $\lambda \geq 0$. Vorm nu

$$(84) \quad D(\lambda, \gamma) = \frac{1}{2\pi} \int_{\Omega} \exp(-\frac{1}{2}x^2 - \frac{1}{2}y^2) dx dy$$

Dus is $C(\lambda) = D(\lambda, \frac{\pi}{3})$. Analooq met (81) bewijst men de asymptotische ontwikkeling voor grote λ

$$(85) \quad D(\lambda, \gamma) \sim \frac{1}{2} e^{-\frac{1}{2}\lambda^2} \sum_{k=0}^{\infty} \frac{k! b_k}{(\frac{1}{2}\lambda^2 \sin^2 \gamma)^{k+1}}$$

waarin de coëfficiënten b_k volgen uit de ontwikkeling

$$(86) \quad \frac{1}{(t + \frac{1}{\sin^2 \gamma}) \sqrt{t + \cot^2 \gamma}} = \sum_{k=0}^{\infty} b_k t^k$$

Met (69) en (81) vinden we voor $\mu \geq 0$

$$A = \frac{1}{2\pi} e^{-\frac{i}{3}\mu^2} \left[\sum_{k=0}^{n-1} \frac{k! a_k}{(\frac{1}{4}\mu^2)^{k+1}} + S_n \right]$$

met

$$(87) \quad |S_n| \leq C 3^{n+1} \frac{n!}{(\frac{1}{4}\mu^2)^{n+1}}$$

en uit (60) volgt nu de asymptotische ontwikkeling

$$(88) \quad Q^2 = mn \left[\alpha(1-\alpha) + (m+n-2) \left\{ -\alpha^2 + \frac{1}{2\pi} e^{-\frac{i}{3}\mu^2} \left(\sum_{k=0}^{n-1} \frac{k! a_k}{(\frac{1}{4}\mu^2)^{k+1}} + S_n \right) \right\} \right]$$

waarin $\alpha = \Phi(-\frac{1}{2}|\mu|\sqrt{2})$ (analoog met (77) blijkt (88) voor willekeurige waarden μ te gelden).

Desgewenst kan men ook α en α^2 asymptotisch ontwikkelen; voor α^2 gebruikt men dan formule (85) met $\gamma = \frac{\pi}{4}$.

Tabel voor de coëfficiënt $a_{n,h}$ in $\frac{1}{(1-t)^2(1-t^2)\dots(1-t^n)} = \sum_{h=0}^{\infty} a_{n,h} t^h$

voor $1 \leq n \leq 5$ en $0 \leq h \leq 40$.

$\rightarrow n$						$\rightarrow n$					