NOTA 1255 maart 1981

Instituut voor Cultuurtechniek en Waterhuishouding
Wageningen

ASPECTEN van INFORMATIEVERWERKING

25

Nieuwe invoermogelijkheden voor een programma voor het berekenen van de kansverdeling van een som van twee stochastische variabelen

K. Oostindie

Nota's van het Instituut zijn in principe interne communicatiemiddelen, dus geen officiële publikaties.
Hun inhoud varieert sterk en kan zowel betrekking hebben op een eenvoudige weergave van cijferreeksen, als op een concluderende discussie van onderzoeksresultaten. In de meeste gevallen zullen de conclusies echter van voorlopige aard zijn omdat het onderzoek nog niet is afgesloten.
Bepaalde nota's komen niet voor verspreiding buiten het Instituut in aanmerking

A S P E C T E N v a n I N F O R M A T I E V E R W E R K I N G

25

De nota's handelende over Aspecten van Informatieverwerking bevatten inlichtingen over de ontwikkeling van de informatieverwerking binnen het Instituut. Naast meer concluderende en toelichtende beschouwingen zal aandacht worden besteed aan het gebruik van programma's en programmapakketten en zullen zakelijke inlichtingen over praktijkervaring met en toepassing van de informatieverwerking worden gegeven

INHOUD

1. INLEIDING

Tot de ICW-programma bibliotheek behoort het programma PROBSUM voor het bepalen van een kansverdeling van de som van twee stochastische variabelen, wanneer van elk de kansverdeling gegeven is. Dit programma is ontwikkeld door VAN GILS (1973) en de theorie hierover is beschreven door STOL (1973).

Het programma PROBSUM berekent de kansverdeling van:

$$\underline{z} = \underline{x} + \underline{y} \qquad (1.1)$$

waarin $\underline{z}$, $\underline{x}$ en $\underline{y}$ stochastische variabelen voorstellen.

De kansverdeling van $\underline{z}$ wordt berekend wanneer gegeven is:

- de (empirische) kansverdeling van $\underline{x}$
- de (empirische) kansverdeling van $\underline{y}$

De invoer van gegevens voor het programma is met nieuwe mogelijkheden uitgebreid. In deze nota zullen alle vormen van input, die thans mogelijk zijn, worden toegelicht.

2. TYPEN VAN KANSVERDELINGEN

Onderscheid dient te worden gemaakt tussen onderschrijding- en overschrijdingskansen. Er geldt:

cumulatieve overschrijdingskans = 1 - cumulatieve (onderschrijdings) kans

of

$$P(\underline{x} > x) = 1 - P(\underline{x} \leq x) \qquad (2.1)$$

Hierin is x de zogenaamde kritieke waarde en P een algemeen symbool voor kans.

2.1. D i s c r e t e   v e r d e l i n g e n

Discrete verdelingen hebben betrekking op integer variabelen. Deze nemen alleen gehele waarden aan bijvoorbeeld $k = 0,1,2,\ldots$ Het type van de bijbehorende kansverdeling is in dit geval:

$$P(\underline{k} = k) = p(k), \tag{2.1.1}$$

waarbij geldt, dat: $0 \leq p(k) \leq 1$. Tevens is het mogelijk de kansverdeling over alle waarden k_i $(i=1, \ldots, n)$ die $\underline{k}$ kan aannemen cumulatief te geven. De vorm van de kansverdeling is dan:

$$P(\underline{k} \leq k_j) = \sum_{i=1}^{j} p(k_i),\ j = 1, \ldots, n \tag{2.1.2}$$

Ook is het mogelijk de cumulatieve overschrijdingskans te geven. De vorm van de kansverdeling is dan:

$$P(\underline{k} > k_j) = 1-P(\underline{k} < k_j) = 1 - \sum_{i=1}^{j} p(k_i),\ j = 1,\ldots,n \tag{2.1.3}$$

of, wanneer men de ondergrens ook meeneemt

$$P(\underline{k} \geq k_j) = 1-P(\underline{k} \leq k_{j-1}) = 1 - \sum_{i=1}^{j-1} p(k_i), j = 2,\ldots,n \tag{2.1.4}$$

2.2. C o n t i n u e   v e r d e l i n g e n

Continue verdelingen hebben betrekking op continue variabelen. Deze kunnen zowel gehele-, als ook niet-gehele waarden aannemen. Het type van de bijbehorende kansverdeling is dan:

$$P(\underline{x} \leq x) = \int_{-\infty}^{x} f(t)dt \tag{2.2.1}$$

waarin $f(t)$ de kansdichtheid is. Wordt de kansverdeling op intervallen gedefinieerd, dan kan geschreven worden:

$$P(a \leq \underline{x} \leq b) = \int_{a}^{b} f(t)dt \tag{2.2.2}$$

waarin $\underline{x}$ ligt in het interval $[a,b]$. Ook bij continue variabelen is het mogelijk de cumulatieve onverschrijdingskans te geven en wel in de vorm van:

$$P(\underline{x} > x) = 1 - \int_{-\infty}^{x} f(t)dt \qquad (2.2.3)$$

3. OVERGANG VAN CONTINUE NAAR DISCRETE VERDELINGEN

Continue verdelingen kan men in een vorm brengen analoog aan discrete verdelingen. Wanneer van een variabele de verdeling niet bekend is en men alleen waarnemingsuitkomsten ter beschikking heeft zal men van deze voorstellingswijze gebruik moeten maken. De samenhang luidt dan:

$$P(a_i < x_{ci} \leq b_i) = p(x_{ci}) \qquad (3.1)$$

waarin

a_i = ondergrens van klasse (interval) i

b_i = bovengrens van klasse (interval) i

$x_{ci} = \frac{b_i - a_i}{2}$ (klasse midden van de i de klasse of centrale waarde)

en waarin

$$p(x_{ci}) = \int_{a_i}^{b_i} f(t)dt \qquad (3.2)$$

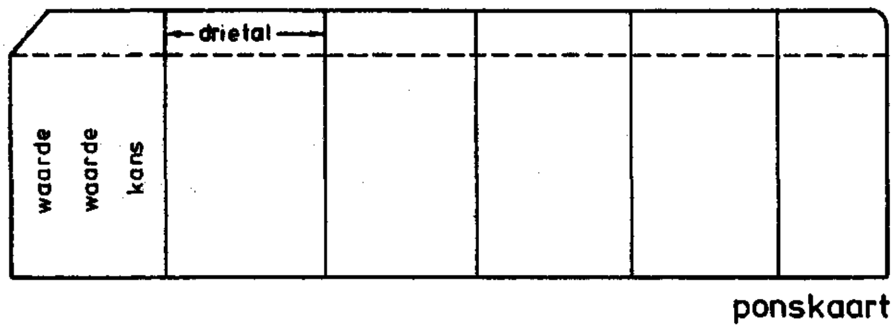
In dit geval wordt (3.1) de empirische verdeling genoemd wanneer waarden van p uit een steekproef worden gevonden. In dat geval wordt het rechterlid van (3.2) met p geschat.

4. ALGEMENE WERKWIJZE

De algemene werkwijze van het programma PROBSUM kan als volgt schematisch worden weergegeven.

- Een (empirische) kansverdeling wordt in principe gegeven door een reeks van k drietallen
- Onder 'drietal' zullen we verstaan: drie getallen die betrekking hebben op een klasse (discreet) of interval (continu) van een kansverdeling.
- Een drietal is in het algemeen opgebouwd uit 2 waarden van de gegeven kansvariabele en een getal voor de kans
- Het programma PROBSUM behandelt de invoergegevens per drietal
- Maximaal mogen er 6 drietallen op een invoerkaart voorkomen en 120 drietallen per kansverdeling
- Per berekening moeten 2 kansverdelingen gegeven zijn, welke worden gescheiden door het invoeren van spaties.

Schematische voorstelling van een ponskaart met 6 drietallen.



Opmerking: Er is verschil tussen nul en blank, zodat nullen in getallen moeten worden geponst

5. SCHEMATISCHE VOORSTELLING

Ter verduidelijking van de vorm waarin de kansverdelingen in het programma PROBSUM gegeven kunnen worden, volgen op de volgende pagina's de schematische voorstellingen, voor elk geval apart, van de inputkansverdeling.

	variabele	: discreet
Geval 1:	kansverdeling:	intervallen
	vorm	: kans per interval

De drietallen bestaan uit achtereenvolgens

1. ondergrens
2. bovengrens
3. kans

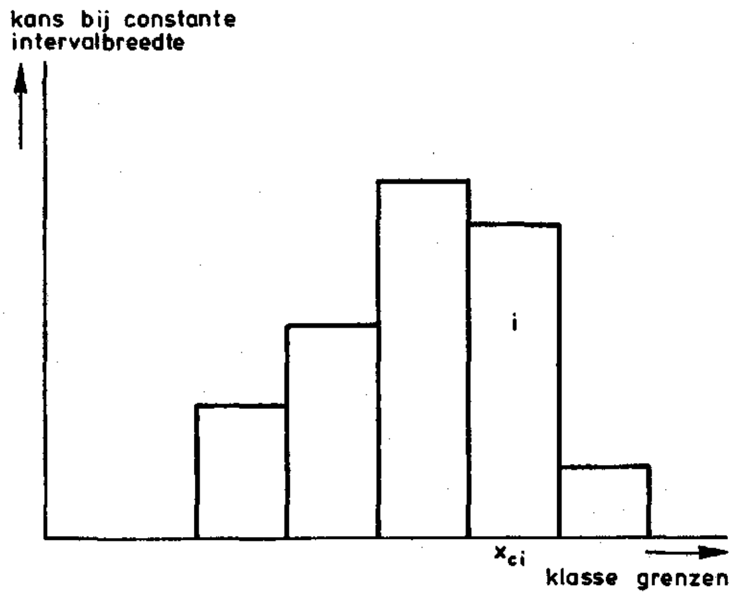


Fig. 1. Schematische voorstelling van een discrete kansverdeling ingedeeld in intervallen (histogram)

Per interval wordt een centrale waarde (x_{ci}) berekend. Deze centrale waarde komt overeen met het klasse-midden. Deze vorm van kansverdeling komt overeen met input mogelijkheid nr. 1 (blz.10).

	variabele	: discreet
Geval 2:	kansverdeling:	klassen
	vorm	: kans per klasse

De drietallen bestaan uit achtereenvolgens

1. klassemidden
2. kans per klasse
3. blank

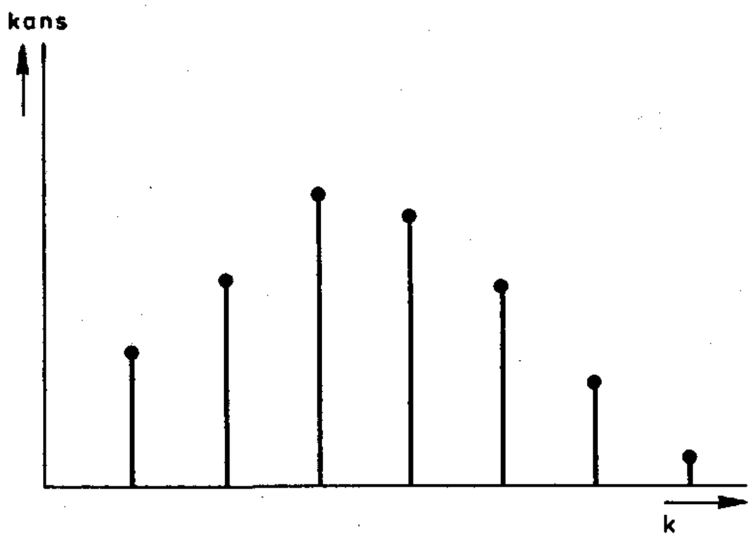


Fig. 2. Schematische voorstelling van een discrete kansverdeling ingedeeld in klassen (staafdiagram)

Deze vorm van kansverdeling komt overeen met inputmogelijkheid nr. 2 (blz. 11).

	variabele	: discreet
Geval 3:	kansverdeling:	intervallen
	vorm	: cum kans per interval

De drietallen bestaan uit achtereenvolgens

1. klasse - bovengrens
2. cumulatieve kans
3. blank

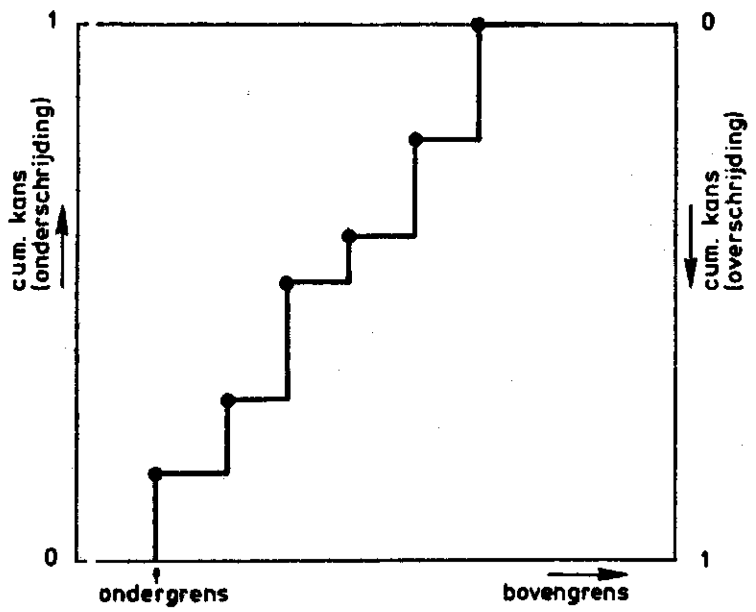


Fig. 3. Schematische voorstelling van een discrete cumulatieve kansverdeling ingedeeld in intervallen (trap-functie)

Deze vorm van kansverdeling komt overeen met inputmogelijkheid nr. 3 (blz.11).

variabele : continu

Geval 4: kansverdeling: intervallen

vorm : kans per interval

De drietallen bestaan uit achtereenvolgens

1. klasse ondergrens (x_1)
2. kans op waarde in interval
3. AAAA in eerste drietal
   blank in volgende drietallen

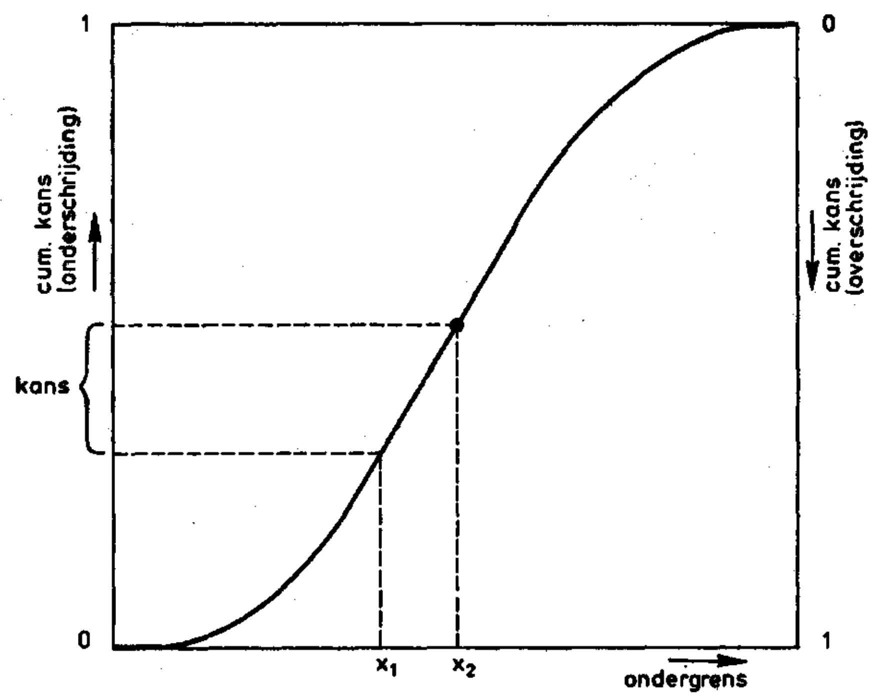


Fig. 4. Schematische voorstelling van een continue cumulatieve kansverdeling waarmee de ondergrens per klasse is gegeven

Deze vorm van kansverdeling komt overeen met inputmogelijkheid nr. 4 (blz. 12).

Geval 5: variabele : continu
kansverdeling: intervallen
vorm : kans per interval

De drietallen bestaan uit achtereenvolgens

1. klasse bovengrens (x_2)
2. kans op waarde in interval
3. BBBB in eerste drietal
   blank in volgende drietallen

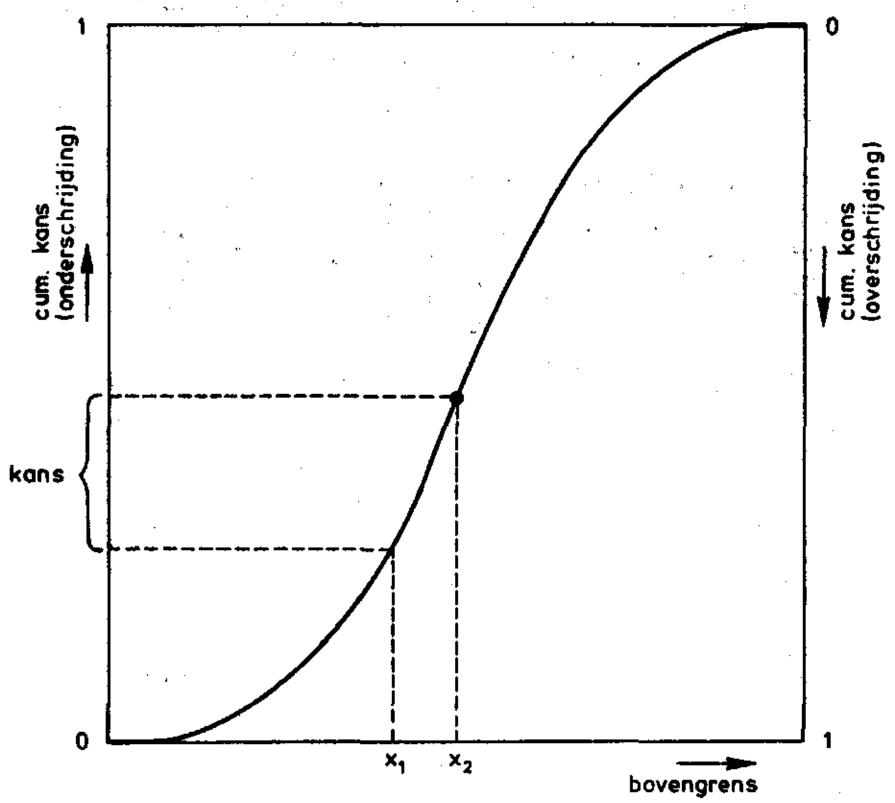


Fig. 5. Schematische voorstelling van een continue cumulatieve kansverdeling, waarvan de bovengrens per klasse is gegeven

Deze vorm van kansverdeling komt overeen met inputmogelijkheid nr. 5 (blz. 13).

6. VOORBEELDEN VAN INPUTMOGELIJKHEDEN

De input voor het programma PROBSUM bestaat uit twee (empirische) kansverdelingen. Deze behoeven niet in dezelfde vorm gegeven te zijn, elk mag dus in een eigen vorm opgegeven worden. Zoals in hoofdstuk 5 reeds werd behandeld zijn er 5 verschillende vormen van inputmogelijkheden. In elk voorbeeld worden twee gelijke verdelingen als invoer gegeven. In het algemeen is dit dus niet vereist evenmin als gelijkheid van vorm.

Geval 1 $P(x_1 < \underline{x} \leq x_2)$

Onder- en bovengrens met de kans (fig.1). De onder- en bovengrens moeten in niet dalende volgorde opgegeven worden.

Het format is: 6(F3.0 , F3.0 , 3X , F4.4)

```
        VOORBEELD 1 VAN DE INPUT VOOR PROGRAMMA PROBSUM.                    GEVAL 1
  1  1   1667  2  2   1666  3  3   1667  4  4   1667  5  5   1666  6  6   1667
  1  1   1667  2  2   1666  3  3   1667  4  4   1667  5  5   1666  6  6   1667
F3.0 F3.0 4x F4.4
<-- drietal 1 -->                  <-- drietal 4 -->
```

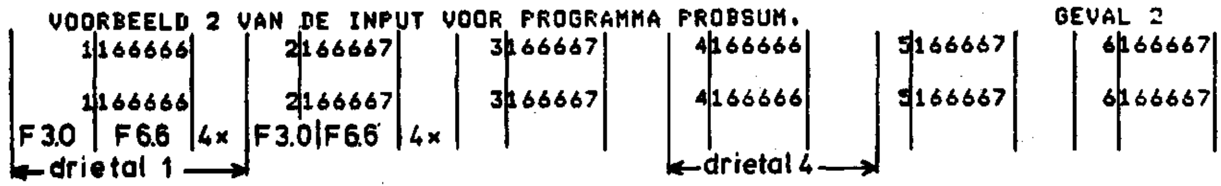
Ondergrens	Bovengrens	Kans
1	1	1/6
2	2	1/6
3	3	1/6
4	4	1/6
5	5	1/6
6	6	1/6

Zoals uit dit voorbeeld blijkt mogen onder- en bovengrens gelijk genomen worden, wat een bijzonder geval is.

Geval 2 $P(\underline{x} = x)$

Centrale waarde met de kans (fig. 2). De centrale waarde moet in stijgende volgorde opgegeven worden.

Het format is: 6(F3.0, F6.6, 4X)



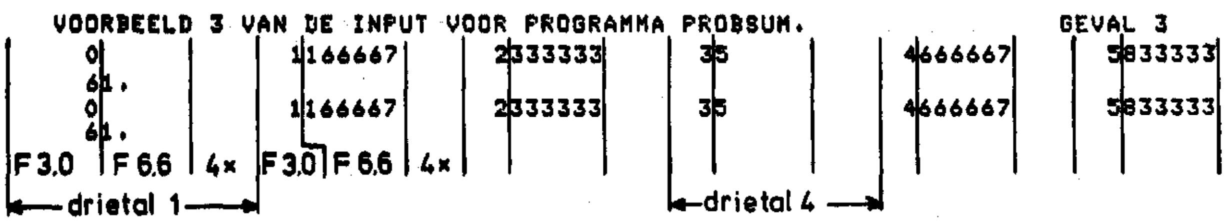
Centrale waarde	Kans
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Geval 3 $P(\underline{x} \leq x)$

Bovengrens met de cumulatieve kans (fig. 3).

De bovengrens en de cumulatieve kans dienen in stijgende volgorde opgegeven te worden, tevens moet het eerste drietal bestaan uit de laagste grenswaarde met de cumulatieve kans blank.

Het format is: 6(F3.0, F6.6, 4X)



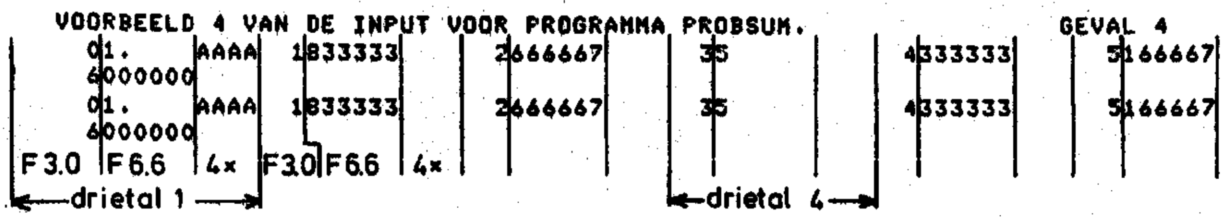
Bovengrens	Cumulatieve kans
0	blank
1	1/6
2	2/6
3	3/6
4	4/6
5	5/6
6	6/6

Geval 4 P($\underline{x}$ >x)

Ondergrens met de cumulatieve kans (fig. 4)

De ondergrens moet in stijgende volgorde opgegeven worden en de cumulatieve kans in dalende volgorde. Op het eerste drietal van een verdeling komt in kaartkolom 10 t/m 13: AAAA. Het laatste drietal van een verdeling bestaat uit de hoogste grenswaarde met een cumulatieve kans die gelijk is aan: ØØØØØØ.

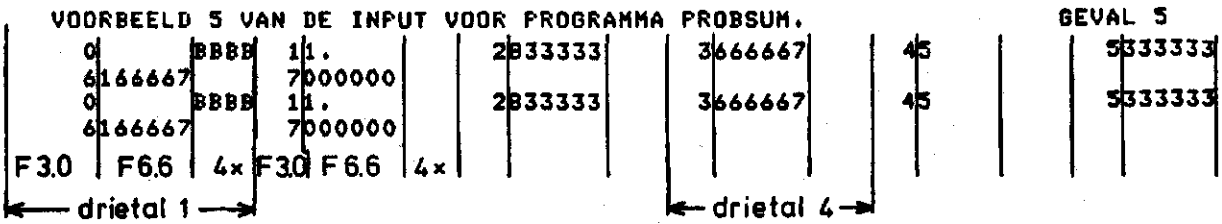
Het format is: 6(F3.0, F6.6, 4X)



Ondergrens	Cumulatieve kans
0	6/6
1	5/6
2	4/6
3	3/6
4	2/6
5	1/6
6	0

Geval 5 $P(\underline{x} \geq x)$

Ondergrens met de cumulatieve kans (fig. 5)
De ondergrens moet in stijgende volgorde opgegeven worden, de cumulatieve kans dalend. Op het eerste drietal van een verdeling komt in kaartkolom 10 t/m 13: BBBB. Het laatste drietal van een verdeling bestaat uit de hoogste grenswaarde met een cumulatieve kans die gelijk is aan: ØØØØØØ.
Het format is: 6(F3.0, F6.6, 4X)



Ondergrens	Cumulatieve kans
0	
1	6/6
2	5/6
3	4/6
4	3/6
5	2/6
6	1/6
7	0

7. AFSLUITEN VAN DE INPUT

Elke kansverdeling moet worden afgesloten met een blank drietal, eventueel op een nieuwe volgkaart.

De executie van het programma wordt gestopt, door na de tweede verdeling een geheel blanke kaart te laten lezen.

8. HET COMPLETE INPUT DEK

Een compleet input-dek wordt als volgt opgebouwd:

a. titelkaart bestaande uit maximaal 80 tekens tekst.

b. een frequentieverdeling volgens één van de vijf mogelijke vormen van input. De verdeling moet afgesloten worden met een blank drietal. Indien de laatste kaart met drietallen van de verdeling al helemaal gevuld is, moet een geheel blanke kaart volgen

c. een tweede frequentieverdeling volgens één van de vijf mogelijke vormen van input. Ook deze verdeling moet worden afgesloten met een blank drietal respectievelijk blanke kaart

d. herhaal de punten b en c, of voeg een blanke kaart aan de input toe. Deze blanke kaart zorgt ervoor, dat de executie van het programma gestopt wordt.

Een compleet input-dek zou er als volgt uit kunnen zien: de kansverdelingen zijn gegeven volgens voorbeeld 5.

Voorbeeld 6. Een compleet input-dek

NEERSLAG IN TIENDE MM. VOOR EEN TIJDVAKLENGTE VAN 1 DAG IN JANUARI

0		BBBB	1	1.	2	61573	3	56876	4	53027	5	50707
6	48556		7	46236	8	45048	9	44142	10	42954	15	41935
20	36276		25	31522	30	26881	35	23146	40	20147	45	18166
50	15789		55	13978	60	12280	65	11148	70	9620	75	8545
80	7187		85	6451	90	5602	95	5150	100	4244	110	3961
120	2943		130	2320	140	1980	150	1528	160	1358	170	1075
180	962		190	849	200	679	210	509	240	339	260	226
270	113		280	56	300	000000						
0		BBBB	1	1.	2	61573	3	56876	4	53027	5	50707
6	48556		7	46236	8	45048	9	44142	10	42954	15	41935
20	36276		25	31522	30	26881	35	23146	40	20147	45	18166
50	15789		55	13978	60	12280	65	11148	70	9620	75	8545
80	7187		85	6451	90	5602	95	5150	100	4244	110	3961
120	2943		130	2320	140	1980	150	1528	160	1358	170	1075
180	962		190	849	200	679	210	509	240	339	260	226
270	113		280	56	300	000000						

(blanke kaart)

Deze input bestaat uit twee keer een frequentieverdeling van neerslaggegevens in de maand januari, waarbij de tijdvaklengte bestaat uit één dag.

9. ENKELE OPMERKINGEN OVER DE OUTPUT

Een gedeelte van de output, van het in voorbeeld 6 genoemde inputdek, wordt hieronder afgedrukt. De output bestaat uit:

a. een tabel van de gegeven frequentieverdelingen van $\underline{x}$ en $\underline{y}$

b. een tabel van de frequentieverdeling van $\underline{x} + \underline{y}$.

Onder de theoretische veronderstelling waaronder de verdeling van $\underline{x} + \underline{y}$ is berekend wordt nu een kansverdeling verkregen van neerslaggegevens in de maand januari, maar waarbij de tijdvaklengte nu bestaat uit twee dagen. Aan de hand van de frequentieverdeling van gemeten neerslaggegevens in de maand januari over twee dagen kan de verkregen outputtabel vergeleken worden met de empirische verdeling. Hierbij komen we echter op het terrein van de specifieke toepassingen van het programma in onderzoekssituaties. Dit valt buiten het bestek van deze nota.

Voorbeeld 7. Een gedeelte van de output

X	P	I	Y	Q
0.	0.000000	1	0.	0.000000
1.	.384270	2	1.	.384270
2.	.046970	3	2.	.046970
3.	.038490	4	3.	.038490
4.	.023200	5	4.	.023200
5.	.021510	6	5.	.021510
6.	.023200	7	6.	.023200
7.	.011880	8	7.	.011880
8.	.009060	9	8.	.009060
9.	.011880	10	9.	.011880
10.	.010190	11	10.	.010190
15.	.056590	12	[illegible]	.056590
20.	.047540			[illegible]47540
25.	.046410			[illegible]0
		[illegible]	140.	[illegible]
		34	150.	.001[illegible]
	[illegible]30	35	160.	.002830
170.	.001130	36	170.	.001130
180.	.001130	37	180.	.001130
190.	.001700	38	190.	.001700
200.	.001700	39	200.	.001700
210.	.001700	40	210.	.001700
240.	.001130	41	240.	.001130
260.	.001130	42	260.	.001130
270.	.000570	43	270.	.000570
280.	.000560	44	280.	.000560

voorbeeld 7 (vervolg). Een gedeelte van de output

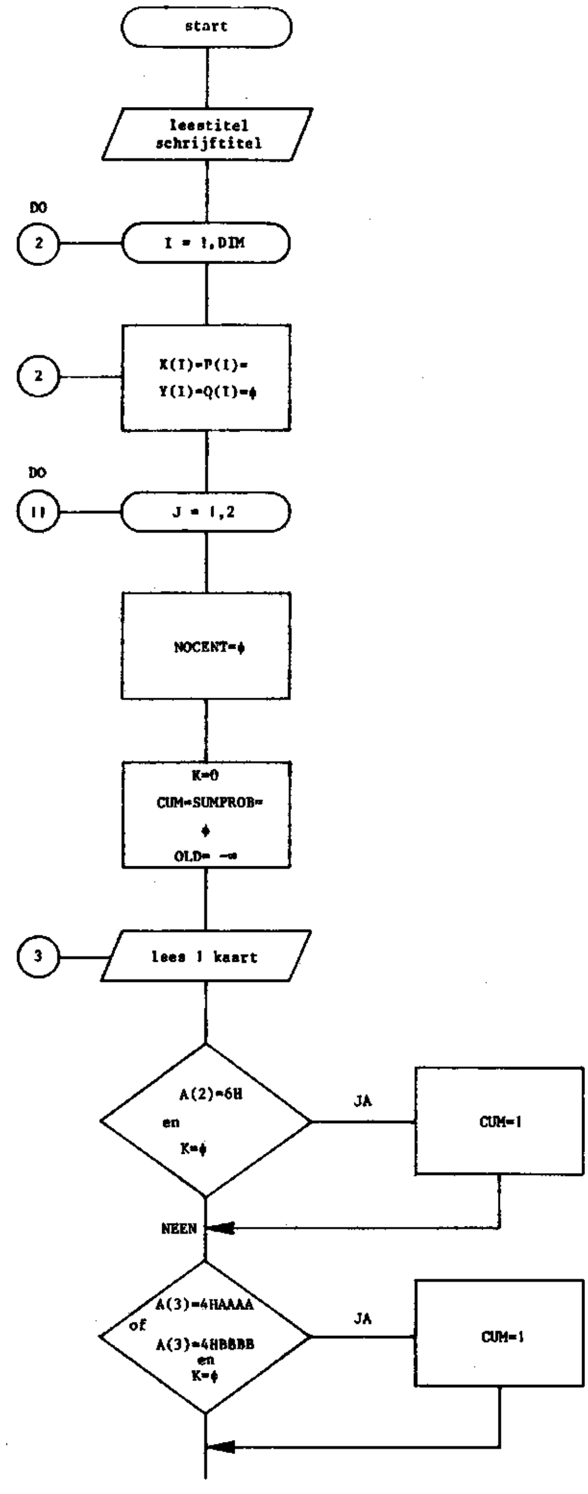
K	CRITICAL VALUE	(INTERVAL	)	* CUMULATIVE * PROBABILITY * DESCENDING *	* CUMULATIVE * PROBABILITY * ASCENDING *
1	1.0	1.0 ,	2.0	100.0000	0.0000
2	2.0	2.0 ,	3.0	85.2337	14.7663
3	3.0	3.0 ,	4.0	81.6238	18.3762
4	4.0	4.0 ,	5.0	78.4451	21.5549
5	5.0	5.0 ,	6.0	76.3005	23.6995
6	6.0	6.0 ,	7.0	74.2813	25.7187
7	7.0	7.0 ,	8.0	72.1176	27.8824
8	8.0	8.0 ,	9.0	70.7672	29.2328
9	9.0	9.0 ,	10.0	69.6809	30.3191
10	10.0	10.0 ,	11.0	68.4374	31.5626
11	15.0	15.0 ,	16.0	66.4812	33.5188
12	20.0	20.0 ,	21.0	60.4621	39.5379
13	25.0	25.0 ,	26.0	54.8215	45.1785
14	30.0	30.0 ,	31.0	49.0967	50.9033
15	35.0	35.0 ,	36.0	44.1012	55.8988
16	40.0	[illegible] ,	41.0	39.7697	60.2303
17	45.0		46.0	36.4704	63.5296
18	5[illegible]	[illegible]	[illegible]1.0	32.8525	67.1475
19		450.0 ,	[illegible]0	29.8184	70.1816
	[illegible].0	460.0 ,	4[illegible]	26.9563	73.0437
	470.0	470.0 ,	480.	24.7278	75.2722
66	480.0	480.0 ,	490.0	[illegible]2.1829	77.817[illegible]
67	490.0	490.0 ,	500.0	[illegible]	[illegible]
68	500.0	500.0 ,	510.0		[illegible]
69	510.0	510.0 ,	520.0	.0006	99.9994
70	520.0	520.0 ,	530.0	.0004	99.9996
71	530.0	530.0 ,	540.0	.0003	99.9997
72	540.0	540.0 ,	550.0	.0001	99.9999
73	550.0	550.0 ,	560.0	.0000	100.0000
74	560.0	560.0 ,	560.0	.0000	100.0000

REFERENTIES

GILS, J.B.H.M. VAN, 1973. Programma PROBSUM; computer programma voor het bepalen van de kansverdeling van de som van twee toevalsvariabelen wanneer de verdeling van elk gegeven is

STOL, PH.TH., 1973. Over de numerieke berekening van de kansverdeling van eenvoudige functies van twee kansvariabelen

Bijlage 1


start
leestitel
schrijftitel
DO
2
I = 1,DIM
2
X(I)=P(I)=
Y(I)=Q(I)=ϕ
DO
11
J = 1,2
NOCENT=ϕ
K=0
CUM=SUMPROB=
ϕ
OLD= -∞
3
lees 1 kaart
A(2)=6H
en
K=ϕ
JA
CUM=1
NEEN
A(3)=4HAAAA
of
A(3)=4HBBBB
en
K=ϕ
JA
CUM=1

DO-loop voor initiele waarden.

Alle getallen uit de verdelingen krijgen de uitgangswaarde.

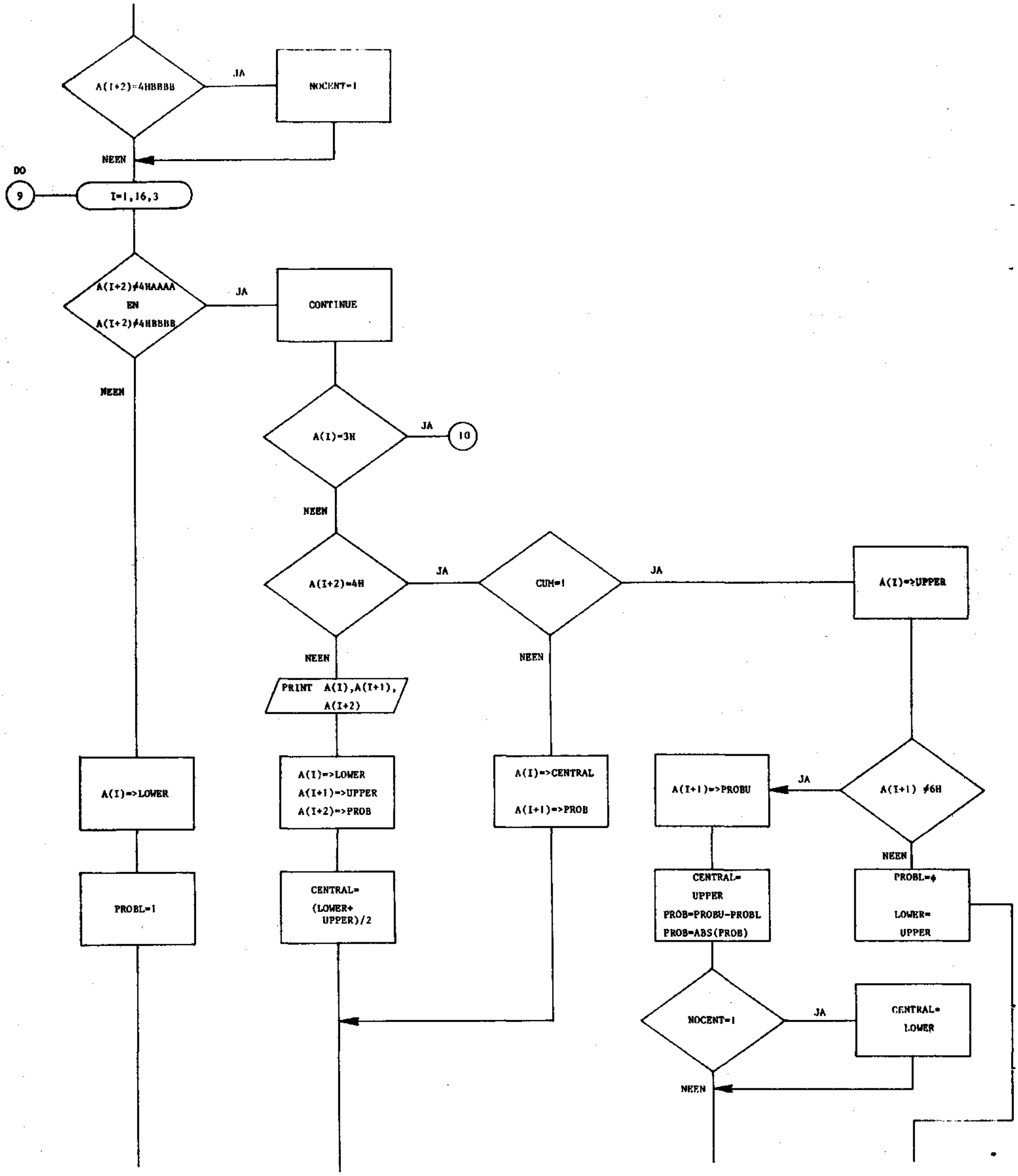
DO-loop voor verdelingen.

Besturingsteken wordt geactiveerd wanneer de gegeven waarde als ondergrens gaat fungeren.

K = een teller voor het aantal gegevens per verdeling. cum = besturingsteken. Wanneer de cum kans gegeven is wordt cum geactiveerd. SUMPROB = cum som van de kansen. OLD = variabele om te testen of de centrale waarden in niet-dalende volgorde staan.
Zie voorbeelden van Input.

Opzetten van besturingsteken cum.

Bijlage 1 vervolg



Test of het 3^e getal van het eerste drie-tal gelijk is aan BBBB, zo ja dan wordt het besturingsteken nocent opgezet.

DO-loop voor het decoderen van de gegevens per kaart en per drietal.

Test of het derde getal van een drietal gedecodeerd moet worden. Zo ja, dan is de verdeling onderschrijdend. Zo nee, dan is de verdeling overschrijdend.

Test of het eerste getal van een drietal blank is. Zo ja, dan worden er geen getallen meer, van die reeks, gelezen.

Test of het laatste getal van een drietal blank is — zo ja, — Test of het besturingsteken cum geaktiveerd is — zo ja, — eerste getal van een drietal wordt gedecodeerd als Upper

zo nee,

schrijf de drietallen

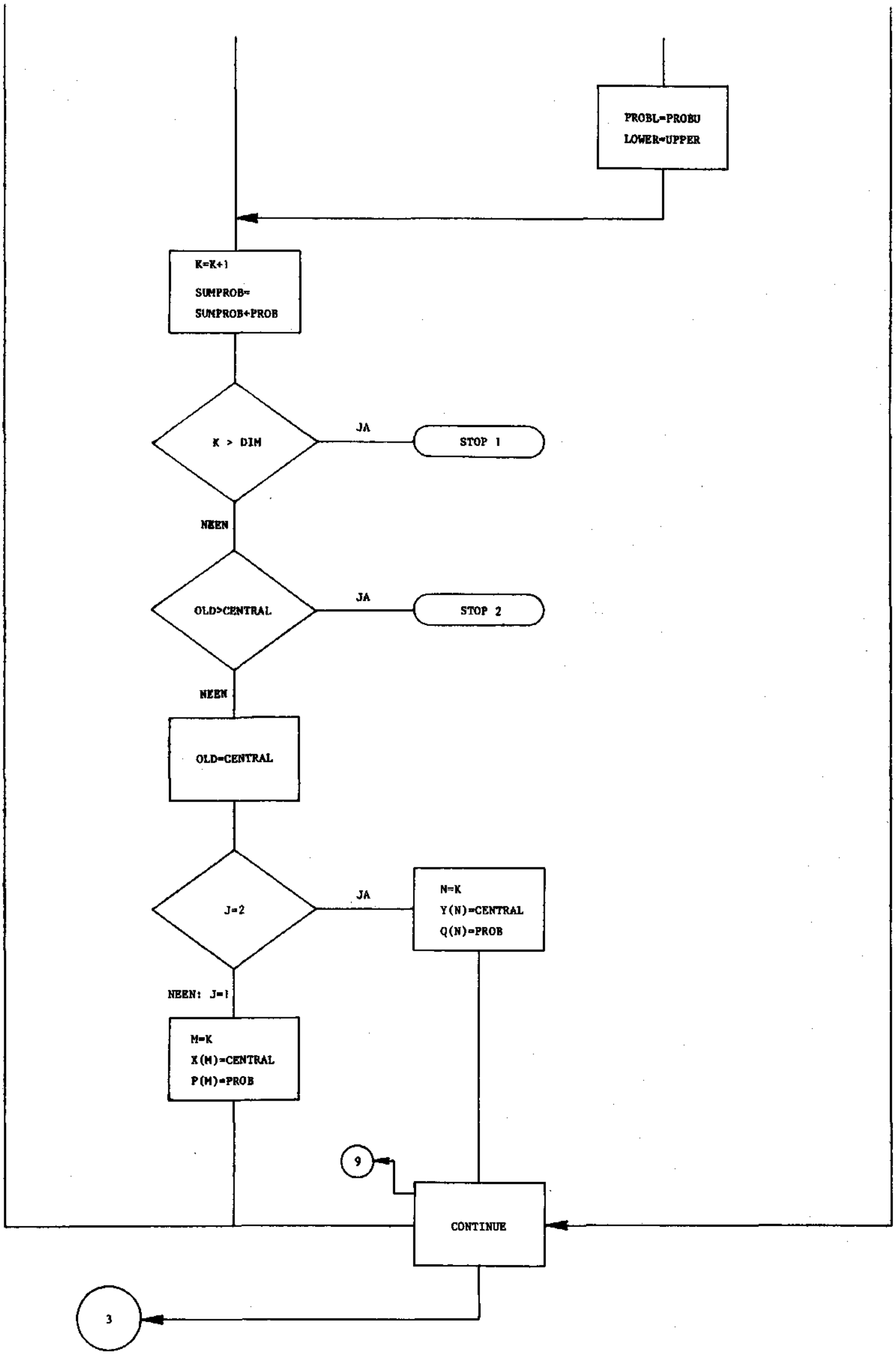
zo nee,

Decoderen van de drietallen naar de diverse variabelen ← zo ja, ← Test of het 2^e getal van een drietal blank is

Definiëren van variabelen en berekenen van centrale waarde

Test of het besturingsteken nocent opgezet is — zo ja, — dan is lower de centrale waarde

Bijlage 1 vervolg



Herdefiniëren van de variabelen
Probl en Lower

Teller wordt per drietal
met één verhoogd kans
wordt gesommeerd

Test of K groter is dan maximaal aantal zo ja, stop drietallen (nl. 120)

Test of de X respectievelijk Y waarden niet in stijgende volgorde staan
zo ja, stop

De nieuwe X respectievelijk Y waarde wordt de oude X respectievelijk Y
waarde

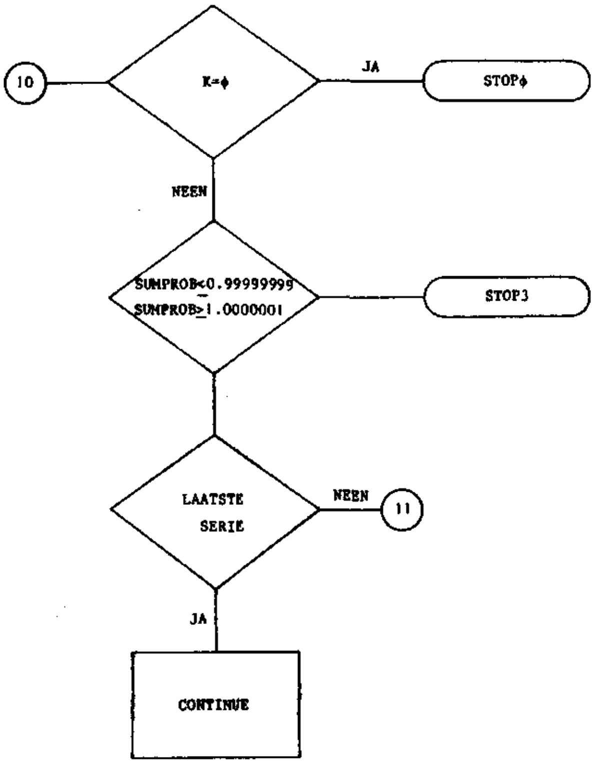
Test of dit de eerste of de tweede reeks is

Schrijf daarna voor de eerste (tweede) reeks het nummer van het drietal
weg op M(N), de waarde van de variabele op X(M), (Y(N)), en de kans op
P(M), (Q(N)).

Programma springt terug naar het begin van DO-loop met statement nr. 9

Ga naar statement nr. 3

Bijlage 1 vervolg



Komt terug via statementnummer 10.

Test of K gelijk is aan nul zo ja, stop: dan is er van de eerste of van de tweede reeks geen enkel drietal gelezen.

Test of de som van de kansen kleiner of groter is dan één, zo ja, dan is de verdeling niet compleet of bevat teveel gegevens.

Test of de tweede reeks gelezen is zo nee, programma springt terug naar DO-loop met statement nr. 11 om de tweede reeks te gaan bewerken.

Zo ja, dan wordt de kansverdeling van de som berekend.