

Arbeidsnotater

S T A T I S T I S K S E N T R A L B Y R Å

Dronningensgt. 16, Oslo-Dep., Oslo 1. Tlf. 41 38 20

IO 76/35

7. desember 1976

TO STANDARDISERTE MASKINPROGRAMMER.

TREKKING AV UTVALG

GENERERING AV TILFELDIGE TALL

av

Kjetil Sørli

INNHold

	Side
1. Innledning	1
2. Et maskinprogram for trekking av utvalg	1
2.1. Generelt	1
2.2. Trekkemetoden	1
2.3. Stratifisering	2
2.4. Input - output	3
2.5. Parameterkort	4
2.6. Bruk av egen subrutine for tildeling av stratumnumre	5
2.7. Kontroll av trekkingen	6
2.8. Et bevis for trekkemetodens gyldighet	6
3. En maskinell rutine for generering av tilfeldige tall	8
3.1. Generelt	8
3.2. Teknisk bruk ved programmering	8
3.3. Eksempel på generering av tilfeldige seks-sifrede tall ved hjelp av rutinen	8
3.4. Tallgeneratoren. Metode og egenskaper	9
3.5. Tabell med tilfeldige tall	12
Referenser	14
Vedlegg: Nødvendige styrekort for trekkeprogrammet til bruk på H 6060	15

Ikke for offentliggjøring. Dette notat er et arbeidsdokument og kan siteres eller refereres bare etter spesiell tillatelse i hvert enkelt tilfelle. Synspunkter og konklusjoner kan ikke uten videre tas som uttrykk for Statistisk Sentralbyrås oppfatning.

1. Innledning

Denne publikasjon inneholder beskrivelse av to maskinprogrammer. I kapittel 2 beskrives et generelt program for trekking av utvalg, og kapittel 3 beskriver en maskinell rutine for generering av tilfeldige tall. Rutinens navn er TFLDG. Eldre beskrivelser [3] av de to maskinprogrammene vil bli erstattet av denne.

Det er naturlig her å ta med en kommentar til et generelt problem som man stadig møter, når man vil utarbeide program- eller systembeskrivelser. Saken er at i slike beskrivelser må man henvende seg til to typer lesere samtidig. Vi har på den ene side de som er mest interessert i sak og metode og på den annen de som interesserer seg mest for den praktiske bruk. Da det er en tendens i dag til at gruppen av de som må interessere seg for begge aspekter samtidig blir større, finner jeg det naturlig å samle all dokumentasjon under ett. Det problemet som da oppstår er å finne et balansepunkt i terminologi og form som vil gjøre stoffet tilgjengelig for flest mulig lesere. Her har jeg forsøkt å oppnå det ved i så stor grad som mulig å rendyrke hvert enkelt avsnitt for de forskjellige leserformål. Jeg har også forsøkt å gjøre dette klart ved valg av navn på de forskjellige avsnittene.

Ved praktisk bruk er det en fordel at beskrivelsen er mest mulig kortfattet og oversiktlig. Dette har jeg generelt forsøkt å ta hensyn til og spesielt i de avsnitt hvor kravet til dette er størst, avsnitt 2.4 og 2.5. Forøvrig har jeg forsøkt å lage beskrivelsene mest mulig uavhengige av maskintype. Det er bare avsnitt 2.4 og vedlegget jeg har vært nødt til å tilpasse for spesiell bruk i Statistisk Sentralbyrå.

2. Maskinprogram for trekking av utvalg

2.1. Generelt

Programmet trekker et spesifisert antall elementer (m) tilfeldig ut av et kjent totaltall (n), hvor $0 \leq m \leq n$. Dette kan gjøres for flere strata (grupper av elementer) samtidig, og slik at alle mulige utvalg har den samme sannsynlighet for å bli trukket ut. Det forutsettes at hvert element er representert ved en record (enhet) på datafilen.

For hvert stratum spesifiseres m og n på parameterkort, dette er nødvendig bare for strata hvor $m > 0$. Hvis totalt antall records for et stratum på filen avviker fra det som spesifiseres på parameterkort, går trekningen galt. Det kan kontrolleres automatisk om dette skjer.

2.2. Trekkemetoden

Vi ser nå på ettstratum hvor elementene ligger i en vilkårlig rekkefølge. Hvert element behandles nå ett og ett om gangen i den rekkefølgen de ligger, etter følgende rutine:

- det simuleres et tilfeldig rektangulær fordelt tall x i intervallet $[0,1]$

- la element nr. $k+1$ være med i utvalget hvis $x \leq \frac{m-j}{n-k}$

hvor j er antall elementer som er blitt trukket ut blant de k foregående elementene ($k = 0,1,2,\dots,n-1$).

Dette betyr at hvis det er blitt trukket j blant de k første elementene, blir det neste elementet trukket ut med en betinget sannsynlighet nettopp lik $m-j/n-k$.

Gangen i trekkerutinen er illustrert i nedenstående tabell. Vi antar at vi har $m = 5$ og $n = 10$ og altså at ett og ett element behandles om gangen.

element nr.	k	j	$\frac{m-j}{n-k}$	x	er $x \leq \frac{m-j}{n-k}$
1	0	0	0.500	0.036	JA
2	1	1	0.444	0.711	NEI
3	2	1	0.500	0.875	NEI
4	3	1	0.571	0.179	JA
5	4	2	0.500	0.341	JA
6	5	3	0.400	0.950	NEI
7	6	3	0.500	0.763	NEI
8	7	3	0.667	0.855	NEI
9	8	3	1.000	0.007	JA
10	9	4	1.000	0.845	JA

Element nr. 1, 4, 5, 9 og 10 ble altså trukket.

Denne trekkeметoden er ikke ny, den er bl.a. beskrevet av Cassel [1]. Hos Sverdrup [2] antydes det hvordan man kan vise at hvert element har den samme ubetingete sannsynlighet m/n for å bli trukket ut. Beviset er her tatt med i avsnitt 2.8.

2.3. Stratifisering

Det er altså mulig å trekke simultant for flere strata samtidig. Det er ikke nødvendig at elementene innen hvert stratum ligger samlet på datafilen. Hvert stratum representeres med et stratumnummer som brukeren selv identifiserer strataene med. Hvis massen det skal trekkes fra utgjør akkurat ett stratum, trenger ikke brukeren spesifisere nummeret. Ingen stratumnummer må overstige 3000, og dette er også maksimalgrensen for antall strata.

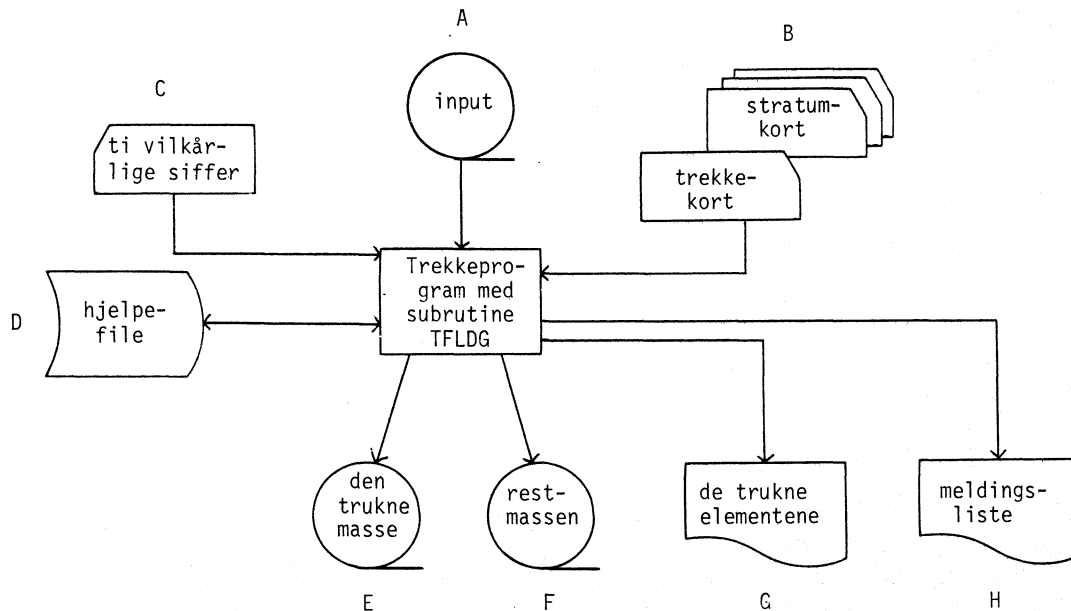
For hvert stratum må brukeren spesifisere et stratumkort, som inneholder stratumnummer, totalt antall elementer (n) og antall elementer som skal trekkes (m). Redigeringen av disse kortene er beskrevet i avsnitt 2.5. Hvis det ligger records på datafilen som ikke tilhører noen av de strata det skal trekkes for, er det ikke nødvendig med stratumkort for disse. Stratumnummer må likevel alle records ha, selv om de altså tilhører et stratum det ikke skal trekkes for. Overstiger stratumnummeret 3000 gis feilmelding uten at det får konsekvenser for trekkingen i andre strata.

Hver record, element, identifiseres altså med stratumnummeret i programmet. Dette kan gjøres på én av to måter. Enten ligger stratumnummeret i recorden direkte, da angir brukeren nummerets startposisjon og lengde på et parameterkort (trekkekortet). Den andre muligheten foreligger hvis stratumindelingen er slik at det ikke finnes noe naturlig entydig nummer i recorden, som identifiserer stratumet. Derimot kan det f.eks. hende at to eller flere kjennemerker tilsammen gir et identifikasjonsbegrep, eller at stratumindelingen ligger i en fordeling av antall records (f.eks. 100 records i hvert stratum). I slike tilfelle er det muligheter for å lage seg en egen subrutine (STRNUM) som tildeler recordene et lovlig entydig stratumnummer avhengig av stratifiseringsbetingelsene. Selv om man vanligvis er i den første situasjonen med stratumnummer direkte i recorden, kan det være greitt å ha muligheten for å foreta en enkel omkodning uten å måtte kjøre gjennom filen en gang ekstra.

I avsnitt 2.6 beskrives hvordan subrutine STRNUM lages og hvordan den fungerer. Det gis to eksempler, ett i COBOL og ett i FORTRAN.

Hvis brukeren ikke kjenner antall elementer, records, i hvert stratum på forhånd, må han kanskje kjøre gjennom filen for å telle dem opp. Hvis det er mange strata, kan det da lønne seg å lage stratumkortene (og også trekkekortet) samtidig ved å skrive dem ut på en midlertidig file.

2.4. Input - output



Figuren viser input og output til trekkeprogrammet.

A. Datafile, organisert med én record pr. trekkeelement

Filekode I1

Recordlengde max 1908, System Standard Format

Ingen spesiell sortering

B. Ett trekkekort + stratumkortene

Filekode Ix

Detaljert beskrivelse i avsnitt 2.5

C. Parameterkort til rutine som genererer tilfeldige tall

Filekode O1 (null - én)

Redigering:

Kol. 1 - 10 ti vilkårlige siffer

" 11 - 80 blanke

D. Hjelpfile som allokeres bare hvis det ønskes en kontroll på om trekkingen er gått riktig

Filekode W1

Se ellers beskrivelse av trekkekortet, avsnitt 2.5

E. Outputfile med de trukne records

Filekode U1

Filen allokeres etter behov, se beskrivelse av trekkekortet, avsnitt 2.5

Følgende midlertidige ALTER legges ved for å få ønsket recordlengde:

A ALTER 29,29 02 FILLER PIC X (·), ellers fås recordlengde 1908

F. Outputfile med de records som ikke blir trukket

Filekode U2

Filen allokeres etter behov, se beskrivelse av trekkekortet, avsnitt 2.5

Følgende midlertidige ALTER legges ved for å få ønsket recordlengde:

A ALTER 32,32 02 FILLER PIC X (·), ellers fås recordlengde 1908

G. Printfile med de trukne records og/eller disse's recordnumre fra inputfilen, nummerert fra 1 og oppover

Filekode P1

Redigering:

Pos. 1- 6 recordnummeret fra inputfilen

" 7 blank

" 8 [x]

" 9-132 pos. 1-124 fra inputrecorden

Filen allokeres etter behov med eller uten selve recorden i pos. 9-132, se ellers beskrivelse av trekkekortet, avsnitt 2.5

H. Meldinger

Filekode Px

Innhold:

- antall records på inputfilen
- antall records som trekkes totalt
- antall records på restfilen, hvis filen allokeres
- melding fra kontrollen om trekkingen har gått riktig (hvis kontrollen ønskes)

2.5. Parameterkort

Parameterkortene består av ett trekkekort og like mange stratumkort som antall strata det skal trekkes fra. Trekkekortet må alltid ligge først. Innhold og redigering er (se figuren i avsnitt 2.4):

- a) Kol. 1: [1] hvis outputfile E ønskes
[] ellers
- b) Kol. 2: [1] hvis outputfile G ønskes med pos. 1-124 fra inputrecorden på lista
[2] hvis outputfile G ønskes uten pos. 1-124 fra inputrecorden på lista (bare recordnummer)
[] ingen outputliste
- c) Kol. 3 er en kode for stratuminnndeling: [1] hele inputfilen utgjør ett stratum
[2] stratumnummer ligger i recorden (max. 4 siffer, ≤ 3000)
[3] stratumnummer tildeles av subrutine STRNUM, som brukeren lager selv
- d) Hvis kol. 3 = [1] eller [3], settes kol. 4-7 blanke. Hvis kol. 3 = [2] opplyses i kol. 4-7:
Kol. 4-6: startposisjon til stratumnummeret på inputfilen
" 7: antall siffer i stratumnummeret (max. 4)
- e) Kol. 8: [K] hvis det ønskes kontroll om trekkingen er gått riktig
[] ellers

- f) Kol. 9: [R] hvis outputfile F ønskes
[] ellers

Kol. 10-80 settes blanke

Etter trekkekortet følger så stratumkortene, ett for hvert stratum:

- Kol. 1- 4 Stratumnummer (ikke nødvendig å fylle ut hvis hele massen utgjør ett stratum)
" 5-10 antall records i dette stratum (n)
" 11-16 antall records som skal trekkes i dette stratum (m)
" 17-80 blanke

Det er viktig at det er overensstemmelse mellom stratumnumrene i parameterkortene og dem som står på filen (eller som genereres i STRNUM).

2.6. Bruk av egen subrutine for tildeling av stratumnumre

Hvis stratumnummeret ikke naturlig kan velges eller er et maximum 4-sifret tall på inputfilen, kan man skrive en egen subrutine - STRNUM - som programmet kaller opp for hver innlest record. Kol. 3 i trekkekortet må da = [3]. Rutinens oppgave er å tildele recorden et stratumnummer som er avhengig av visse opplysninger på inputfilen.

Best kan vel dette illustreres ved et par eksempler. Anta at vi har en personfile med kode for kjønn i kol. 61, kode for aldersgruppe i kol. 82 og kode for fylke i kol. 8-9. Gyldige koder er for kjønn 1, 2 - for aldersgrupper 1-8 og for fylker 01-20. Filens recordlengde er 222. Anta at personer i samme fylke med samme kjønn og i samme aldersgruppe utgjør et stratum. Da får vi $20 \times 2 \times 8 = 320$ strata. En entydig stratumnummerering har vi da ved å la kodene for kjønn, aldersgruppe og fylke i denne rekkefølge utgjøre vårt stratumnummer. Disse numrene blir på 4 siffer og alle er mindre enn 3 000 (det største er 2 820).

Under følger subrutine STRNUM, som tildeler records stratumnumre etter ovenstående oppskrift, skrevet i COBOL.

```

0001 IDENTIFICATION DIVISION.
0002 PROGRAM-ID. STRATA.
0003 ENVIRONMENT DIVISION.
0004 CONFIGURATION SECTION.
0005 SOURCE-COMPUTER. 6000 WITH EIS.
0006 OBJECT-COMPUTER. 6000 WITH EIS.
0007 DATA DIVISION.
0008 WORKING-STORAGE SECTION.
0009 77 J PIC 9(6) COMP-1.
0010 01 REK.
0011     02 FILL PIC X(7).
0012     02 FYLKE PIC XX.
0013     02 FILL PIC X(51).
0014     02 KJØNN PIC X.
0015     02 FILL PIC X(20).
0016     02 ALDERSGR PIC X.
0017     02 FILL PIC X(140).
0018 01 J-ALFANUM.
0019     02 KJ PIC X.
0020     02 ALD PIC X.
0021     02 FYL PIC XX.
0022 01 J-RE REDEFINES J-ALFANUM PIC 9(4).
```

```

0023  PROCEDURE DIVISION.
0024      ENTER LINKAGE.
0025      ENTRY STRNUM USING REK GIVING J.
0026      ENTER COBOL.
0027  TILDEL-NR.
0028      MOVE KJONN TO KJ.
0029      MOVE ALDERSGR TO ALD.
0030      MOVE FYLKE TO FYL.
0031      MOVE J-RE TO J.
0032  UT.
0033      EXIT STRNUM.
0034      END OF PROGRAM.

```

Subrutinens parametre er altså inputrecorden REK og stratumnummeret J. I COBOL må J alltid defineres COMP-1.

Hvis man f.eks. ønsker å sette 100 på-hverandre følgende records i hvert stratum, f.eks. for å sikre en trend i trekkeettheten av inputmassen, kan følgende FORTRAN-rutinge brukes:

```

SUBROUTINE STRNUM (REK,J)
  IF (K-9) 1,2,1
1  J = 1
  I = 0
  K = 9
2  I = I + 1
  IF (I-100) 4,4,3
3  I = 1
  J = J + 1
4  RETURN
  END

```

Her nummereres stratumene, hver bestående av 100 records, fortløpende fra 1 og oppover.

2.7. Kontroll av trekkingen

Ved å punche [K] i kol. 8 i trekkekortet blir det kontrollert at antall records trukket og antall records totalt stemmer med det som er angitt i stratumkortene. Det meldes fra om alt er i orden eller hvor eventuelle uoverensstemmelser ligger. Uoverensstemmelser skyldes enten at antall records i stratum på filen ikke er som angitt, eller at stratumnumrene selv er feilpunchet i stratumkortene.

2.8. Bevis for trekke metodens gyldighet

Med et induksjonsbevis skal vi nå vise at hvert element har ubetinget sannsynligheten $\frac{m}{n}$ for å bli trukket ut, når element nr. $k+1$ blir trukket med betinget sannsynlighet $\frac{m-j}{n-k}$, $k = 0, 1, \dots, n-1$.

Først vil vi for illustrasjonens skyld vise at dette er riktig for de 3 første elementene. For første element er det greitt, ubetinget sannsynlighet er lik den betingete.

For de to neste har vi følgende:

$$\begin{aligned}
& P \{2. \text{ element skal trekkes} \} = \\
& P \{1. \text{ element ble trukket \& 2. element blir trukket} \} \\
& + P \{1. \text{ element ikke ble trukket \& 2. element blir trukket} \} \\
& = P \{1. \text{ element ble trukket} \} \cdot P \{2. \text{ element trekkes} \mid 1. \text{ ble trukket} \} \\
& + P \{1. \text{ element ikke ble trukket} \} \cdot P \{2. \text{ element trekkes} \mid 1. \text{ ikke ble trukket} \} \\
& = \frac{m}{n} \left(\frac{m-1}{n-1} \right) + \left(1 - \frac{m}{n} \right) \frac{m}{n-1} = \frac{m}{n} \\
& P \{3. \text{ element skal trekkes} \} = \\
& P \{1. \text{ ble trukket \& 2. ble trukket \& 3. blir trukket} \} \\
& + P \{1. \text{ ble trukket \& 2. ikke ble trukket \& 3. blir trukket} \} \\
& + P \{1. \text{ ikke ble trukket \& 2. ble trukket \& 3. blir trukket} \} \\
& + P \{1. \text{ ikke ble trukket \& 2. ikke ble trukket \& 3. blir trukket} \} \\
& = \frac{m}{n} \left(\frac{m-1}{n-1} \right) \left(\frac{m-2}{n-2} \right) + \frac{m}{n} \left(1 - \frac{m-1}{n-1} \right) \left(\frac{m-1}{n-2} \right) \\
& + \left(1 - \frac{m}{n} \right) \left(\frac{m}{n-1} \right) \left(\frac{m-1}{n-2} \right) + \left(1 - \frac{m}{n} \right) \left(1 - \frac{m}{n-1} \right) \frac{m}{n-2} = \frac{m}{n}
\end{aligned}$$

Vi skal nå anta at vi vet at det blant de k første elementene hele veien er blitt trukket med ubetinget sannsynlighet $\frac{m}{n}$. Vi skal da vise at også element nr. $k+1$ vil ha ubetinget sannsynlighet $\frac{m}{n}$ for å bli trukket ut.

$$P \{ \text{element nr. } k+1 \text{ skal bli trukket} \} =$$

$$P \{ U [j \text{ elementer ble trukket blant de første } k \text{ \& element nr. } k+1 \text{ trekkes}] \} =$$

$$j \in [0, \min(m, k)]$$

$$(1) \quad \sum_{j=0}^{\min(m, k)} P \{ j \text{ elementer ble trukket blant de første } k \} \cdot P \{ \text{element nr. } k+1 \text{ trekkes} \mid j \text{ elementer ble trukket blant de første } k \text{ elementene} \}$$

At j elementer trekkes blant k med sannsynlighet $\frac{m}{n}$ betyr direkte at j er hypergeometrisk fordelt med parametre k, m, n :

$$P \{ j \text{ elementer er trukket blant } k \} = \frac{\binom{k}{j} \binom{n-k}{m-j}}{\binom{n}{m}}$$

(1) kan da skrives

$$(2) \quad \sum_{j=0}^k \frac{\binom{k}{j} \binom{n-k}{m-j}}{\binom{n}{m}} \cdot \frac{m-j}{n-k}$$

(hvis $k > m$ gir verdier av $j > m$ bidrag 0 til summen)

(2) gir nå videre

$$\sum_{j=0}^k \frac{\binom{k}{j} \frac{(n-k)!}{(m-j)! (n-k-m+j)!}}{\frac{n!}{m! (n-m)!}} \cdot \frac{m-j}{n-k}$$

$$= \sum_{j=0}^k \binom{k}{j} \frac{(n-k-1)!}{(m-j-1)! (n-k-m+j)!} \cdot \frac{m-j}{m! (n-m)!}$$

$$= \frac{m}{n} \sum_{j=0}^k \binom{k}{j} \frac{\binom{n-k-1}{m-j-1}}{\binom{n-1}{m-1}!} \frac{(m-1)! (n-m)!}{(m-1)! (n-m)!}$$

$$= \frac{m}{n} \sum_{j=0}^k \binom{k}{j} \frac{\binom{n-k-1}{m-j-1}}{\binom{n-1}{m-1}} = \frac{m}{n}$$

idet uttrykket under summetegnet er den hypergeometriske sannsynlighetsfordelingen med parametre $k, m-1, n-1$.

Vi vet at antagelsen vår gjelder for $k=3$. Hermed er bevist at den også gjelder for $k=4$. Men da gjelder den også for $k=5, 6, \dots, n-1$.

3. En maskinell rutine for generering av tilfeldige tall

3.1. Generelt

Dette kapitlet gir ingen nødvendig tilleggsinformasjon om trekkeprogrammet, som er beskrevet i kapittel 2. Her beskrives subrutine TFLDG helt generelt, men som et eksempel er det vist hvordan rutinen er brukt i trekkeprogrammet (i avsnitt 3.3).

Ved hvert kall av rutinen blir det hentet fram et tilfeldig helt tall blant tallene 1, 2, 3, ..., 1 000. Hvis man ønsker en serie tilfeldige tall, kalles rutinen opp tilsvarende mange ganger. En får da en serie tall som er stokastisk uavhengige og multinomisk fordelte, alle med sannsynlighet 1:1 000. En beskrivelse av tallgeneratoren samt en beskrivelse av hvordan uavhengighet og fordeling påvises finnes i avsnitt 3.4.

3.2. Teknisk bruk ved programmering

Subrutinen er meget enkel å bruke. I COBOL kalles den opp ved

```
CALL TFLDG USING K, M.
```

I FORTRAN er kallet av den

```
CALL TFLDG (K, M)
```

Både K og M må være binært representert, dvs. definert COMP-1 i COBOL. I FORTRAN skjer dette automatisk.

- K er en "bryter" som gis verdi 0 i hovedprogrammet før første kall av rutinen. Deretter brukes ikke K noe mer i hovedprogrammet.
- M er det tilfeldige tallet som blir hentet frem ($M = 1, 2, 3, \dots, 1\ 000$)
- Rutinen krever ett parameterkort som den leser første gang den kalles opp (når bryteren K er lik 0). Dette er tilknyttet filekode 01 (null - én) og skal inneholde ti vilkårlige sifre i kol. 1-10.

3.3. Eksempel på generering av tilfeldige seks-sifrede tall ved hjelp av rutinen

Hvis en ønsker å hente frem tilfeldig et av tallene 0, 1, 2, 3, ..., 999 999, kan dette gjøres ved to kall av subrutinen på følgende måte:

- i) kall rutinen, plasser tallet i en variabel N
- ii) kall rutinen på nytt, plasser tallet i en annen variabel M
- iii) beregn et tall L:

$$L = 1\ 000 (N - 1) + M - 1.$$

Når M og N har verdier blant 1, 2, 3, ..., 1 000 og fordi M og N er stokastisk uavhengige, vil L ha verdi blant 0, 1, 2, ..., 999 999 hvor alle tallene har sannsynlighet 1 : 1 000 000 til å opptre. Denne teknikken blir brukt i trekkeprogrammet. Ved sammenligning av et tilfeldig tall x ($x = 0, 0.001, 0.002, \dots, 0.999$) hvor x blir generert av TFLDG og en brøk b med som oftest flere enn 3 desimaler, kommer en gjennomsnittlig én gang av tusen i en situasjon hvor

$$x < b < x + 0.001.$$

Da genereres det tre nye desimaler bak de tre foregående og ny sammenligning av brøken b og det seks-sifrete, tilfeldige tallet blir foretatt.

På den måten oppnås en nøyaktighet på seks desimaler uten at det som regel er nødvendig å generere mer enn tre. En slik teknikk gir merkbar tidsbesparelse og bør brukes i lignende situasjoner.

3.4. Tallgeneratoren. Metode og egenskaper

Metoden er utviklet under utarbeidelsen av trekkeprogrammet og er ikke beskrevet noe annet sted. Den baserer seg på summariske roteringer av 4 tallserier som alle inneholder et primtall antall elementer (23, 37, 97, 991). Dette innebærer at det må genereres $23 \times 37 \times 97 \times 991 = 81\,804\,077$ tall før samme tallserie kan dukke opp på nytt. Sannsynligheten er da 1 : 1 000 for at den andre tallserien ikke skal være skalaforskøvet i forhold til den første. Gangen i, og sammenhengen mellom roteringene fremgår av figuren på neste side.

Fotskrift 0 står for forrige verdi, fotskrift 1 for neste verdi av S, K, I, L og N under roteringene. M gir etter hvert kall det tilfeldige tallet.

Parameterkortet med de ti vilkårlige siffer gir som figuren viser en del initialverdier. Ved bruk av samme parameterkortet flere ganger fås samme tallrekken ut. Ved å bruke sifre fra dato og klokkeslett i kall-øyeblikket er det selvfølgelig mulig å unngå dette. Men det er ikke noe poeng, snarere kan det være en fordel å ha muligheten til å finne tilbake til den samme tallrekken igjen.

Stokastisk uavhengighet mellom tallene i en forsøksrekke oppnås som en konsekvens av fordelingen. Når en realisasjon av et kall, L, har sannsynlighet 1 : 1 000 for å opptre, lar vi L representere differensen mellom to påhverandre følgende tall. Når vi antar at M_0 er forrige realisasjon, får vi da det neste lik

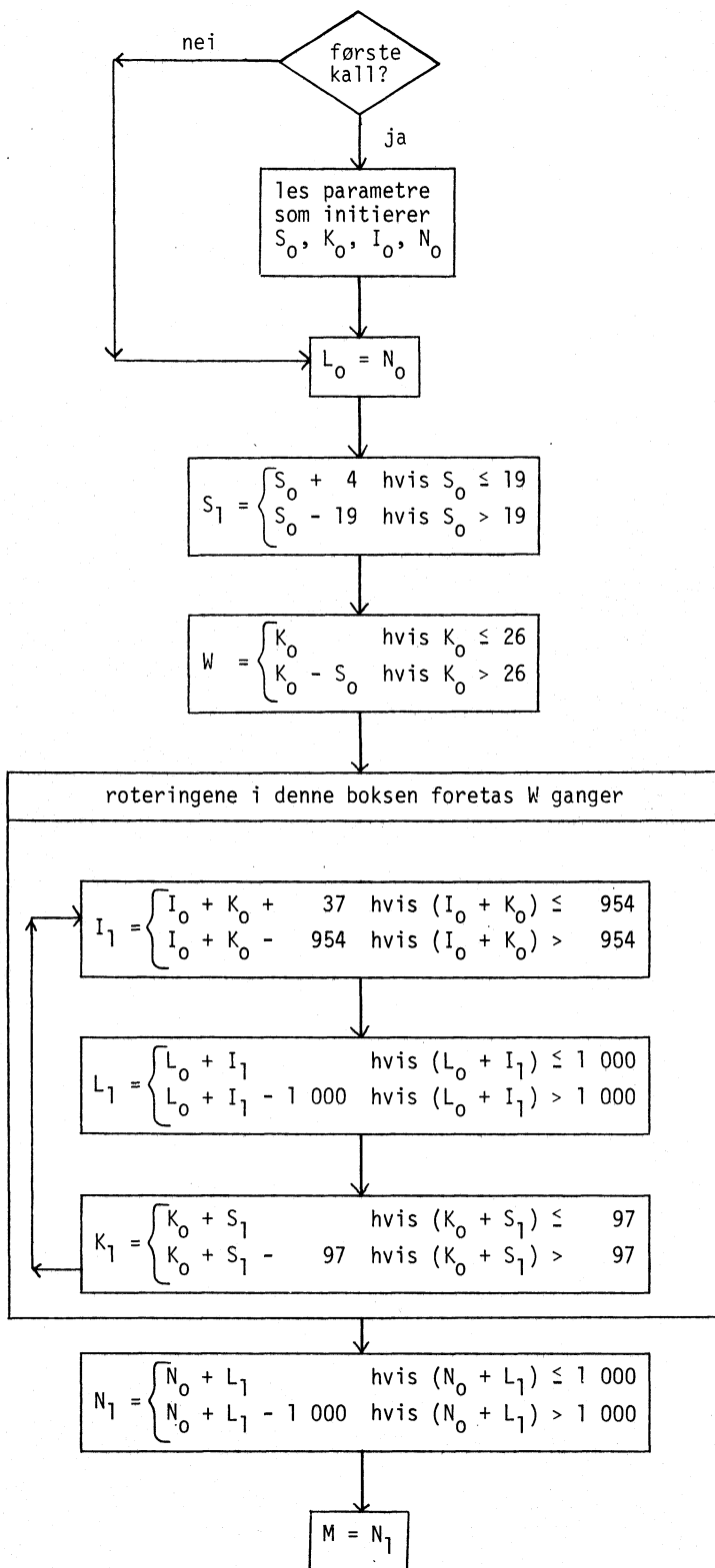
$$M_1 = \begin{cases} M_0 + L & \text{hvis } M_0 + L \leq 1\,000 \\ M_0 + L - 1\,000 & \text{hvis } M_0 + L > 1\,000 \end{cases}$$

Når L er et tilfeldig tall, følger da automatisk at M_0 og M_1 vil bli stokastisk uavhengige.

Da gjenstår det bare å vise at tallene virkelig er tilfeldig multinomisk fordelte. På grunn av systematikken kan dette bare oppnås tilnærmet. For å teste tilnærmelsen er det tatt utgangspunkt i hyppighetsfordelingen for tilfeldige multinomiske tall. I forsøksrekker på hhv. 1 000, 2 000, 3 000 osv. til og med 10 000 kall er det talt opp antall forekomster av hver av de 1 000 verdiene. Disse er så testet mot Poissonfordelingene med parameter lik hhv. 1, 2, 3, ..., 10. De er identiske med hyppighetsfordelingene man får, når forsøksrekkene virkelig fremkommer etter vilkårlig trekking med tilbakelegging. De tallmessige resultatene av disse testene tas ikke med her, jeg vil bare nevne at resultatene var forbausende gode.

Testene ble utført på både de tallene som blir brukt som differenser, altså L-ene i eksemplet over, og de tilfeldige tallene, M-ene.

Det er verdt å merke seg at selv om tallene innen en forsøksrekke er uavhengige, kan det være systematisk avhengighet mellom forskjellige forsøksrekker. Hvis et siffer i parameterkortet endres, blir det endret en startverdi. De ti sifrene gir tilsammen fire startverdier: utgangsposisjon, utgangshastighet og -aksellerasjon, og en utgangsverdi for antall rotasjoner. Hvis bare utgangsposisjonen endres, oppnås bare en skalaforskyving av L-ene.





3.5. TABELL MED TILFELDIGE TALL

37	443	52	526	919	869	291	438	495	653	489	949	755	426	75
712	765	440	778	935	98	461	869	626	215	34	298	902	579	646
876	683	575	214	1000	995	446	559	215	347	910	386	160	954	472
180	943	35	337	980	524	384	806	202	672	1000	547	871	229	144
342	795	621	963	65	332	688	806	247	40	11	555	450	743	601
951	859	44	261	832	983	982	954	399	87	615	145	633	118	239
764	279	263	678	774	81	479	827	526	763	419	825	325	250	52
856	96	706	233	58	56	459	383	284	658	138	3	136	547	390
8	863	423	661	695	967	587	644	520	870	803	999	798	522	833
846	664	466	3	573	215	401	741	322	888	482	911	291	716	36
288	580	457	30	497	646	776	72	890	471	317	2	908	829	908
500	305	73	790	440	277	792	820	715	963	742	562	309	370	906
336	837	314	540	815	377	221	895	799	224	280	618	381	555	928
668	659	468	810	400	903	651	88	877	528	203	825	642	96	984
398	916	44	904	83	386	665	780	722	314	179	665	657	467	150
760	742	691	320	930	581	720	119	452	266	649	796	692	91	988
850	558	888	226	767	406	793	438	668	962	579	898	552	905	908
170	666	616	710	103	114	310	791	346	893	974	720	443	680	902
802	856	363	202	198	771	317	750	160	597	140	651	471	714	376
442	671	129	807	111	259	133	601	941	215	901	690	548	683	103
270	62	894	14	897	175	735	675	387	522	367	811	698	376	131
642	311	720	606	209	297	51	562	221	944	306	722	637	550	676
584	454	993	308	893	51	321	629	406	336	916	496	936	151	346
837	699	10	862	64	510	123	758	250	361	245	682	377	517	544
569	615	26	161	76	462	596	985	281	102	904	733	848	992	847
507	47	363	641	984	565	798	262	28	257	121	632	30	482	114
836	251	313	789	600	357	939	294	299	13	567	501	978	281	130
471	245	69	578	921	650	509	633	523	175	657	590	117	223	569
987	54	117	374	255	226	276	292	793	448	378	736	112	540	533
12	749	260	81	512	260	665	164	16	336	27	478	777	230	111
260	806	212	890	702	297	524	664	925	497	196	391	511	10	644
607	14	770	469	834	635	618	908	400	209	603	617	612	919	853
663	306	862	954	363	384	649	161	793	164	725	858	770	601	511
173	866	856	634	394	505	123	89	715	725	410	548	568	450	949
910	628	776	605	918	321	989	696	551	363	630	252	290	314	839
86	865	1	75	378	158	941	168	322	201	957	691	950	95	496
938	136	64	152	431	63	784	969	408	324	734	863	473	774	790
858	888	938	828	408	611	538	244	119	418	83	927	211	525	360
939	915	475	870	820	303	371	654	966	800	969	171	977	99	110
547	208	832	322	249	90	376	413	294	723	545	169	973	98	44
428	411	975	269	54	175	698	22	494	22	599	588	788	917	351
112	722	217	791	458	871	795	443	737	760	988	89	496	673	125
471	609	331	130	102	835	76	689	326	813	502	173	300	936	95
899	77	769	172	15	789	997	838	859	958	832	682	800	584	450
575	856	561	366	966	661	37	264	161	527	476	736	746	564	16

TALLENE ER GENERERT MED TFLDG.

TALLSERIEN FREMKOMMER KOLONNEVIS VED Å SETTE NULLER (ELLER BLANKE) I KOL. 1-10 I PARAMETERKORTET.

642	783	989	591	381	155	270	773	892	429	952	90	929	552	597
41	381	875	389	694	234	927	840	352	510	850	379	187	190	70
122	329	779	70	265	388	471	439	782	983	548	26	968	411	381
904	233	935	374	226	323	895	188	695	828	463	105	508	820	495
403	606	989	935	891	764	980	344	402	132	572	990	283	525	383
102	202	598	551	770	172	464	521	805	978	983	410	741	513	982
975	674	581	166	774	142	777	941	104	963	382	4	687	578	650
457	835	388	112	65	71	162	945	502	90	768	126	921	145	247
472	119	24	424	323	130	646	308	281	878	370	271	644	200	267
21	68	421	839	820	848	760	482	790	220	842	28	340	913	798
309	246	944	966	383	862	494	912	8	478	353	562	809	893	802
627	692	31	315	996	192	128	193	417	937	545	16	501	567	686
46	557	204	761	647	556	449	482	616	719	615	681	99	412	823
612	599	754	606	422	39	3	909	293	624	546	440	657	135	328
674	113	695	722	870	239	205	847	274	283	673	676	927	539	163
735	980	700	333	405	410	595	165	501	6	382	328	130	244	405
571	552	476	380	401	658	46	548	763	898	938	666	631	226	529
858	591	740	158	911	600	103	894	741	829	713	231	237	218	240
147	673	353	327	533	713	192	963	231	861	731	223	125	376	650
568	140	990	922	273	459	283	605	792	413	689	23	944	146	214
493	238	146	903	218	339	903	522	974	329	975	297	737	910	565
563	866	961	317	279	763	432	694	646	750	608	115	758	950	681
944	952	327	996	423	153	353	219	264	602	1	61	979	199	764
105	108	526	421	910	534	108	313	573	489	508	600	953	803	815
872	638	642	384	246	834	789	812	407	228	623	101	86	765	977
56	748	929	762	704	91	37	578	784	121	701	113	468	412	277
12	343	140	356	824	55	539	317	867	5	480	895	169	524	138
321	66	680	804	685	921	585	123	597	663	498	281	203	594	371
925	815	202	561	263	84	367	405	698	891	12	875	274	267	610
970	878	482	408	41	871	190	411	225	283	291	111	645	779	504
918	628	177	298	830	801	893	492	446	65	89	182	23	439	666
530	158	969	783	546	830	371	307	758	530	338	895	699	965	6
649	282	773	933	262	950	450	752	649	131	480	593	962	285	453
27	186	793	625	634	62	576	683	998	196	236	376	370	140	963
145	43	232	769	969	959	270	446	619	645	216	242	325	396	520
933	350	77	163	738	552	109	966	722	366	229	38	345	772	826
79	36	573	890	116	167	785	253	492	620	14	417	54	453	36
260	680	184	363	711	7	914	6	48	280	311	870	993	80	186
514	775	389	475	169	399	525	628	511	376	645	409	502	598	347
831	740	515	357	771	727	377	977	919	109	831	815	77	156	694
261	669	434	764	362	598	847	232	538	387	961	896	107	780	801
130	721	434	137	96	411	522	721	296	489	363	15	241	487	483
376	422	35	244	15	507	645	972	4	122	465	743	638	744	867
621	662	413	448	413	42	288	958	249	820	989	258	261	568	256
136	226	337	584	552	880	800	227	898	574	727	508	309	402	534

Referenser

- [1] Per Gunnar Cassel: Urvalsdraging på datamaskin med sekventielle metoder. Statistisk Tidsskrift 1967. s. 486-489
- [2] Erling Sverdrup: Lov og tilfeldighet, bind I, s. 50-51
- [3] Kjetil Sørli: "Trekkeprogram P C265039" og "TFLDG, subrutine for generering av tilfeldige tall", to notater i Statistisk Sentralbyrå. A-270-0, 13.11.74

Nødvendige styrekort for trekkeprogrammet til bruk på HB 6000

```

A IDENT      - - - - -
A USERID    - - - - -
A OPTION     FORTRAN
A COBOL      NDECK
A PRMFL      S*, R/W, S, SSB/SOURSE/S5000/SC265039
A UPDATE
A INCODE     IBMEL
A ALTER      29,29      }
                        } gjelder file U1
  02 FILLER PIC X(·)
A ALTER      32,32      }
                        } gjelder file U2
  02 FILLER PIC X(·)
A COBOL      NDECK      }
                        } hvis subrutine STRNUM brukes
  kortdekk til STRNUM
A LIBRARY    A4
A EXECUTE
A PRMFL      A4, R, R, SSB/SUBLIB
A LIMITS     , 27K
A TAPE       I1, ....
(*) A TAPE    U1, ....
(*) A TAPE    U2, ....
(*) A SYSOUT  P1, ....
(*) A FILE     W1, ....

  1 trekkekort
  stratumkortene
A DATA      01 (null - én)
  vilkårlige sifre i kol. 1-10
A ENDJOB

```

(*) : hvis filen skal allokeres (valgfri file).

