Bernd Ulitzka:

(Pseudo-) Zufallszahlen

"Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin." JOHN VON NEUMANN 1951.

1. Einführung

Zufällig erzeugte Zahlen finden heute Verwendung in den verschiedensten Anwendungsbereichen:

a) Simulation:
In der Atomphysik versucht man durch Modellvorstellungen z.B. die unregelmäßigen Zusammenstöße von Atomen oder Molekülen in Gasen zu untersuchen.
In der Unternehmensforschung (operation research) geht es z.B. um die Belastung der Abfertigungsschalter eines Flughafens oder um das Einrichten einer grünen Welle in der Verkehrsplanung.

b) Numerische Mathematik:
Z.B. Berechnung von komplizierten Flächen- und Rauminhalten mit der sogenannten Monte-Carlo-Methode.

c) Informatik:
Zufallszahlen als Ausgangsdaten zum Testen von Programmen.

d) Entscheidungen fällen:
Z.B. Münzwurf bei der Platzverteilung einer Sportveranstaltung, Herausfinden von optimalen Strategien in der mathematischen Spieltheorie, sowie bei Qualitätskontrollen oder Zollabfertigungen.

e) Unterhaltung:
Glücksspielautomaten, Computerspiele, Lotto.

Was heißt überhaupt z u f ä l l i g ?
Wir wollen darunter verstehen, daß eine Folge von Zahlen auf gut Glück erzeugt wird. D.h. jede Zahl ist von den anderen unabhängig und besitzt eine bestimmte Wahrscheinlichkeit (chance), Werte aus einem vorgegebenen Bereich anzunehmen. Der Begriff der Wahrscheinlichkeit wird aber nur durch Eigenschaften erklärt (z.B. durch die Axiome von KOLMOGOROFF, vergl. Jahrgangsstufe 12) und überprüft. Konstruktiv kann er nicht gegeben werden.

Beispiel 1:
Jede der Zahlen 0,1,....,9 soll die gleiche Chance haben, in einem Experiment gezogen zu werden. Man spricht dann von Gleichverteilung. Führt man diesen Versuch 1000 000 - mal durch, so erwartet man von jeder dieser Zahlen, daß sie etwa 100 000 - mal auftritt. Die Möglichkeit aber, daß jede dieser Zahlen g e n a u 100 000 - mal auftritt, ist ä u ß e r s t g e r i n g . Führt man dieses Experiment (1000 000 Ziehungen) sehr oft durch und bildet jeweils die Mittelwerte der Häufigkeiten, so erhält man Werte, die nahe bei 100 000 liegen (siehe auch die Würfelsimulation unter 4.Anhang). Jede Zusammensetzung der 1000 000 Zahlen ist genauso wahrscheinlich, wie z.B. 999 999 Nullen hintereinander. Die Chance beim nächsten Ziehen wieder eine Null zu ziehen, ist immer noch $\frac{1}{10}$.

Als Ausblick in die Stochastik soll hier noch die oben angesprochene Wahrscheinlichkeit angegeben werden:

$$p = \frac{\frac{1000\,000!}{(100\,000!)^{10}}}{10^{1000\,000}} \approx 2{,}55 \cdot 10^{-26},$$

d.h. die Chance, jeweils genau 100 000 Zahlen zu bekommen, ist ungefähr $1 : 4 \cdot 10^{25}$. Diese Berechnung erfolgte mit der STIRLINGschen Formel (vergl.

[2] Seite 138).

2. Erzeugung von Zufallszahlen

1939 lassen KENDALL und BABINGTON-SMITH eine mechanische Maschine bauen, mit der sie 100 000 Zufallszahlen erzeugen, die sie dann veröffentlichten. 1955 veröffentlichte die RAND Corporation 1000 000 mit einer neuen mechanischen Maschine erzeugte Zufallszahlen. Weitere bekannte mechanische Vorrichtungen sind die Lottomaschinen(in England unter dem Namen ERNIE bekannt).
Parallel dazu entstanden elektronische Datenverarbeitungsanlagen (Computer), die die mechanischen Maschinen ersetzen sollten, andererseits aber auch Verwender von Zufallszahlen wurden (siehe 1.). Um Programme erfolgreich zu testen, war es erforderlich, öfters die völlig gleiche Zufallszahlenfolge zu erzeugen, was mit einer mechanischen Vorrichtung nicht möglich war.

1946 schlug JOHN VON NEUMANN die Quadrat-Mitten-Methode (QM) vor:

2.1 Definition: Ausgehend von der letzten erzeugten k-stelligen natürlichen Zahl x_n (n,k aus N, k gerade) bildet man ihr Quadrat, das höchstens 2k-stellig ist und nimmt dann diejenige Zahl als x_{n+1}, die nur durch die mittleren k Ziffern gebildet wird. Die Anordnung soll immer rechtsbündig sein.

Beispiel 2:
x_n = 5772156649 (k = 10); x_n^2 = 33317|7923805949|09201, also x_{n+1}=7923805949.

Bemerkung: 2.1 zeigt, daß x_{n+1} von x_n sicher nicht unabhängig ist, also wird dies keine Folge von zufälligen Zahlen. Da aber i.a. der Benutzer die Erzeugungsart seiner Zufallszahlen nicht kennt, scheinen sie ihm zufällig; daher spricht man von p s e u d o - Z u f a l l s z a h l e n .
Intuitiv wird man diese Methode als nicht gerade schlecht empfinden, da die Ausgangszahl ganz schön "durcheinandergewirbelt" wird. Trotzdem zeigt das nächste Beispiel, daß es sich bei 2.1 um eine schlechte Zufallszahlenquelle handelt:

Beispiel 3:
Untersuchung von 2-stelligen natürlichen Zahlen als Ausgangszahlen der QM-Methode: Nach wenigen Schritten erhält man eine Konstante:

36, 29, 84, 05, 02, 00, 00, ...
34, 15, 22, 48, 30, 90, 10, 10,....
42, 76, 77, 92, 46, 11, 12, 14, 19, 36, 29, 84, 05, 02, 00, 00, ...

2.2 Formaldarstellung der QM-Methode: Ausgangszahl sei die k-stellige natürliche Zahl x_n (k-stellig). Dann ist nach 2.1

$$x_{n+1} = [x_n^2 : 10^{k:2}] - [x_n^2 : 10^{1,5 \cdot k}] \cdot 10^k,$$

wobei [x] diejenige größte ganze Zahl ist, die kleiner oder gleich x ist (in Basic: INT(X)).

Beispiel 4:
k = 4, x_n = 4567 sei vorgegeben; dann ist x_n^2 = 20|8574|89 also x_{n+1}=8574; andererseits erhält man nach 2.2

$$x_{n+1} = [20857489:10^2] - [20857489:10^6] \cdot 10^4 =$$
$$= 208574 - 200\,000 = 8574.$$

Auf einen Beweis von 2.2 wird verzichtet. Siehe auch Programm MATHSEM3 in 4.

Bemerkung: Die Folge kann schnell periodisch werden oder degenerieren und nur noch die Zahl null erzeugen. METROPOLIS zeigte 1950, daß es bei der Benutzung von 20-stelligen Dualzahlen 13 verschiedene Perioden gibt, wobei die längste Periode aus 142 Zahlen besteht. Bei 38-stelligen Dualzahlen kann man aber ca 750 000 Zahlen erzeugen, bis Degeneration eintritt. Vergleiche 4.

Die nächste zur Diskussion gestellte Methode ist besser als die QM-Methode. Die mit ihr erzeugten Pseudo-Zufallszahlen erfüllen viele Zufallseigenschaften und Degeneration ist nicht möglich:

1949 stellte D.H.LEHMER seine lineare Kongruenzmethode vor, die heute noch sehr populär ist und z.B. in Mikrocomputern verwendet wird.

2.3 Definition: (Die Schreibweisen sind im gleichen Heft unter Smolka zu finden). Sei m aus $\mathbb{N}$ der Modul, a aus $\mathbb{N}$ ein Faktor mit $0 \leq a < m$ (manchmal auch $a > m$), c aus $\mathbb{N}$ der Zuwachs mit $0 \leq c < m$ und x_n die Zufallszahl mit $0 \leq x_n < m$, wobei n aus $\mathbb{N}_0$ ist; dann heißt die Folge

$$x_{n+1} \equiv (ax_n + c) \bmod m \qquad (1)$$

lineare Kongruenzfolge.

Durch $\frac{x_n}{m}$ erhält man die üblicherweise verwendeten Zufallszahlen aus dem Bereich [0;1].

Beispiel 5: $m = 10$; $x_0 = a = c = 7$:

$x_1 \equiv (7 \cdot 7 + 7) \bmod 10 \equiv 6$
$x_2 \equiv (7 \cdot 6 + 7) \bmod 10 \equiv 9$
$x_3 \equiv (7 \cdot 9 + 7) \bmod 10 \equiv 0$
$x_4 \equiv (7 \cdot 0 + 7) \bmod 10 \equiv 7 \equiv x_0$ usw.

Auch die lineare Kongruenzmethode (LK-Methode) führt zu einer Periode; deshalb fodert man m maximal, also möglichst groß, damit die Periode möglichst groß wird.

Spezialfälle:
a) $c = 0$:
Die Berechnung von x_{n+1} geht schneller, aber die Periode wird kürzer.
Bei LEHMER war $c = 0$ und $c \neq 0$ wurde von ihm nur empfohlen.
Untersuchungen mit $c \neq 0$ wurden 1958 von THOMSON und ROTENBERG veröffentlicht.

b) $a = 1$
2.4 Satz: $x_{n+1} \equiv (x_n + c) \bmod m \equiv (x_0 + n \cdot c) \bmod m \qquad (2)$
Hierzu zunächst

Beipsiel 6: $a = 1$, $x_0 = 7$, $c = 2$, $m = 10$

	Berechnung nach 2.3	Berechnung nach 2.4
x_1	$\equiv (7 + 2) \bmod 10 \equiv 9$	$(7 + 1 \cdot 2) \bmod 10 \equiv 9$
x_2	$\equiv (9 + 2) \bmod 10 \equiv 1$	$(7 + 2 \cdot 2) \bmod 10 \equiv 1$
x_3	$\equiv (1 + 2) \bmod 10 \equiv 3$	$(7 + 3 \cdot 2) \bmod 10 \equiv 3$
x_4	$\equiv (3 + 2) \bmod 10 \equiv 5$	$(7 + 4 \cdot 2) \bmod 10 \equiv 5$
x_5	$\equiv (5 + 2) \bmod 10 \equiv 7 \equiv x_0$	$(7 + 5 \cdot 2) \bmod 10 \equiv 7$

d.h. die Periodenlänge beträgt 5 (sie ist nicht maximal, weil $5 < 10$.
Versuchen Sie $a = 1$, $x_0 = c = 7$, $m = 10$!

Beweis zu 2.4 nach dem Index n:
a) (2) ist richtig für $n = 1$, weil $x_1 \equiv (x_0 + c) \bmod m \equiv (x_0 + 1 \cdot c) \bmod m$.

b) Annahme (2) sei richtig für n-1, d.h. es gelte
$x_n \equiv (x_{n-1} + c) \bmod m \equiv (x_0 + (n-1)c) \bmod m$, d.h. es gibt ein ganzzahliges k mit
$x_n - (x_0 + (n-1)c) = k \cdot m$.

c) Induktionsschluß:
$x_{n+1} :\equiv (x_n + c) \bmod m \equiv (x_o + (n-1)c + km + c) \bmod m \equiv (x_o + nc) \bmod m$
wegen b). Also gilt (2) für alle n aus !N.

Auswahl des Moduls m:
m sollte möglichst groß gewählt werden, weil die Periode höchstens m verschiedene Elemente haben kann. Da im Dualsystem gerechnet wird, wäre $m = 2^w$ mit einem natürlichen w interessant. Division mit m bedeutet dann nur eine Kommaverschiebung, d.h. bewirkt eine hohe Rechengeschwindigkeit. KNUTH zeigt in [1], daß bei dieser Division die rechte Hälfte der Ziffern von x_{n+1} viel weniger zufällig ist als die linke. Er schlägt daher vor:
$m = 2^w + 1$, z.B. $2^{16} + 1 = 65537$ für die Verwendung in Mikrocomputern. Diese Zahl ist prim, $2^{16} - 1 = 3 \cdot 5 \cdot 17 \cdot 257$ dagegen nicht.

Auswahl des Faktors a:
KNUTH beweist in [1], daß die lineare Kongruenzfolge dann und nur dann die Periodenlänge m hat, wenn
a) c teilerfremd mit a ist,
b) b = a - 1 ein Vielfaches von jedem in m enthaltenen Primfaktor ist.
c) Falls m ein Vielfaches von 4 ist, muß auch b = a-1 Vielfaches von 4 sein.

Zum Schluß noch eine wichtiges Mahnung von KNUTH: Zufallszahlen sollten nie mit einer Methode erzeugt sein, die "zufällig" ausgewählt wurde. Mathematische Theorie muß immer die Auswahl unterstützen. Unter 4. findet man drei Basic-Programme, die die QM- und LK-Methode mit dem IBM-PC eingebauten Zufallsgenerator vergleichen lassen. Simuliert wird ein Würfel und die absoluten und die relativen Häufigkeiten ausgedruckt.

3. Literatur

[1] Knuth, D.E.: The Art of Computer Programming, vol 2, Reading Massachusetts Addison-Wesley (1981).

[2] Brinstein-Semendjajew: Taschenbuch der Mathematik, Frankfurt/Main Verlag Harri Deutsch (1965).

4. Anhang

```
100 REM *** QUADRADATMITTENMETHODE ***
110 REM *** MATHSEM3 ***
120 CLS
130 PRINT "QUADRATMITTENMETHODE"
140 PRINT
150 INPUT "GEBEN SIE EINE 2-, 4-, 6-, 8- ODER 10-STELLIGE NATÜRLICHE ZAHL AN ";X
#
160 K=LEN(STR$(X#))-1 '*** BESTIMMUNG DER STELLENZAHL ***
170 PRINT
180 N=N+1
190 PRINT USING "####       ################";N,X#:Z#=X#
200 Y#=X#*X#            '*** QUADRIERUNG ***
210 X#=INT(Y#/10^(K/2))-INT(Y#/10^(1.5*K))*10^K '*** MITTENBILDUNG ***
220 IF X#<>Z# THEN 180 '*** ABBRUCH BEI DEGENERATION , PERIODENERKENNUNG FEHLT
230 END
```

```
RUN
QUADRATMITTENMETHODE

GEBEN SIE EINE 2-, 4-, 6-, 8- ODER 10-STELLIGE NATÜRLICHE ZAHL AN ? 36

   1                    36
   2                    29
   3                    84
   4                     5
   5                     2
   6                     0
Ok
```

```
RUN
QUADRATMITTENMETHODE

GEBEN SIE EINE 2-, 4-, 6-, 8- ODER 10-STELLIGE NATÜRLICHE ZAHL AN ? 1234

   1                    1234
   2                    5227
   3                    3215
   4                    3362
   5                    3030
   6                    1809
   7                    2724
   8                    4201
   9                    6484
  10                     422
  11                    1780
  12                    1684
  13                    8358
  14                    8561
  15                    2907
  16                    4506
  17                    3040
  18                    2416
  19                    8370
  20                     569
  21                    3237
  22                    4781
  23                    8579
  24                    5992
  25                    9040
  26                    7216
  27                     706
  28                    4984
  29                    8402
  30                    5936
  31                    2360
  32                    5696
  33                    4444
  34                    7491
  35                    1150
  36                    3225
  37                    4006
  38                     480
  39                    2304
  40                    3084
  41                    5110
  42                    1121
  43                    2566
  44                    5843
  45                    1406
  46                    9768
  47                    4138
  48                    1230
  49                    5129
  50                    3066
  51                    4003
  52                     240
  53                     576
  54                    3317
  55                      24
  56                       5
  57                       0
Ok
```

```
100 REM *** SIMULATION EINES WÜRFELS NACH ZWEI METHODEN ***
110 REM *** MATHSEM4 ***
120 REM *** QUADRADATMITTENMETHODE ***
130 CLS
140 PRINT "QUADRATMITTENMETHODE"
150 PRINT
160 INPUT "GEBEN SIE EINE 2-, 4- ODER 6-STELLIGE NATÜRLICHE ZAHL AN ";X
170 K=LEN(STR$(X))-1
180 W$="                    "
190 CLS
200 INPUT "ANZAHL DER WÜRFE ";E
210 LOCATE 1,29:PRINT "WURFNR.:"
220 PRINT:RANDOMIZE TIMER
230 PRINT "QM - METHODE  X =";X;TAB(41)"I B M"
240 PRINT "------------------------------------------------------------"
250 FOR I=1 TO E : LOCATE 1,40 : PRINT I
260    Y=X*X                                        '*** QM - METHODE ***
270    X=INT(Y/10^(K/2))-INT(Y/10^(1.5*K))*10^K   '*** MITTENBILDUNG ***
280    Z=INT(X/10^K*6+1)          ' *** 0 < Z < 7  ***
290    ON Z GOTO 300,310,320,330,340,350
300    N(1)=N(1)+1:GOTO 360
310    N(2)=N(2)+1:GOTO 360
320    N(3)=N(3)+1:GOTO 360
330    N(4)=N(4)+1:GOTO 360
340    N(5)=N(5)+1:GOTO 360
350    N(6)=N(6)+1
360    FOR J=1 TO 6            ' *** BILDSCHIRMDARSTELLUNG ***
370       W(J)=N(J)/I:LOCATE 4+2*J,2:PRINT J;": ";N(J),W$:LOCATE 4+2*J,15:PRINT
W(J)
380    NEXT J
390    Z=INT(RND(1)*6)+1       ' *** IBM ZUFALLSGENERATOR ***
400    ON Z GOTO 410,420,430,440,450,460
410    M(1)=M(1)+1:GOTO 470
420    M(2)=M(2)+1:GOTO 470
430    M(3)=M(3)+1:GOTO 470
440    M(4)=M(4)+1:GOTO 470
450    M(5)=M(5)+1:GOTO 470
460    M(6)=M(6)+1
470    FOR J=1 TO 6            ' *** BILDSCHIRMDARSTELLUNG ***
480       W(J)=M(J)/I:LOCATE 4+2*J,40:PRINT J;": ";M(J),W$:LOCATE 4+2*J,53:PRINT
 J)
490    NEXT J
500 NEXT I:LOCATE 19,1
510 END
```

```
ANZAHL DER WÜRFE ? 1000      WURFNR.:     1000

QM - METHODE  X = 123456                  I B M
------------------------------------------------------------

  1 :  186      .186                      1 :  175      .175

  2 :  170      .17                       2 :  161      .161

  3 :  144      .144                      3 :  156      .156

  4 :  183      .183                      4 :  161      .161

  5 :  169      .169                      5 :  167      .167

  6 :  148      .148                      6 :  180      .18
```

```
100 REM *** SIMULATION EINES WÜRFELS NACH ZWEI METHODEN ***
110 REM *** MATHSEM1 ***
120 W$="                    "
130 CLS
140 INPUT "ANZAHL DER WÜRFE ";E
150 A=11879 :C=36777!:M=65537!' *** PARAMETER DES ZUFALLSZAHLENGENERATORS ***
160 X=20000   '*** ANFANGSWERT ***
170 LOCATE 1,29:PRINT "WURFNR.:"
180 PRINT:RANDOMIZE TIMER
190 PRINT "KONGRUENZMETHODE                                  I B M"
200 PRINT "-------------------------------------------------------------"
210 FOR I=1 TO E : LOCATE 1,40 : PRINT I
220    H=A*X+C:X=H-INT(H/M)*M ' *** KONGRUENZMETHODE ***
230    Z=INT(X/M*6+1)          ' *** 0 < Z < 7  ***
240    ON Z GOTO 250,260,270,280,290,300
250    N(1)=N(1)+1:GOTO 310
260    N(2)=N(2)+1:GOTO 310
270    N(3)=N(3)+1:GOTO 310
280    N(4)=N(4)+1:GOTO 310
290    N(5)=N(5)+1:GOTO 310
300    N(6)=N(6)+1
310    FOR J=1 TO 6            ' *** BILDSCHIRMDARSTELLUNG ***
320       W(J)=N(J)/I:LOCATE 4+2*J,2:PRINT J;": ";N(J),W$:LOCATE 4+2*J,15:PRINT
W(J)
330    NEXT J
340    Z=INT(RND(1)*6)+1       ' *** IBM ZUFALLSGENERATOR ***
350    ON Z GOTO 360,370,380,390,400,410
360    M(1)=M(1)+1:GOTO 420
370    M(2)=M(2)+1:GOTO 420
380    M(3)=M(3)+1:GOTO 420
390    M(4)=M(4)+1:GOTO 420
400    M(5)=M(5)+1:GOTO 420
410    M(6)=M(6)+1
420    FOR J=1 TO 6            ' *** BILDSCHIRMDARSTELLUNG ***
430       W(J)=M(J)/I:LOCATE 4+2*J,40:PRINT J;": ";M(J),W$:LOCATE 4+2*J,53:PRINT
 W(J)
440    NEXT J
450 NEXT I:LOCATE 19,1
460 END
```

```
ANZAHL DER WÜRFE ? 1000      WURFNR.:     1000

KONGRUENZMETHODE                                  I B M
-------------------------------------------------------------

  1 :   188      .188                          1 :   174      .174

  2 :   163      .163                          2 :   183      .183

  3 :   155      .155                          3 :   149      .149

  4 :   170      .17                           4 :   174      .174

  5 :   179      .179                          5 :   171      .171

  6 :   145      .145                          6 :   149      .149
```