.

Ausarbeitung zum Vortrag

Lineare Kongruenzgeneratoren

von Dirk Deters

vom 5. November 2002

Im Seminar “Perlen der Theoretischen Informatik”

bei Prof. Köbler und O. Beyersdorff

Inhaltsverzeichnis

1 Einführung

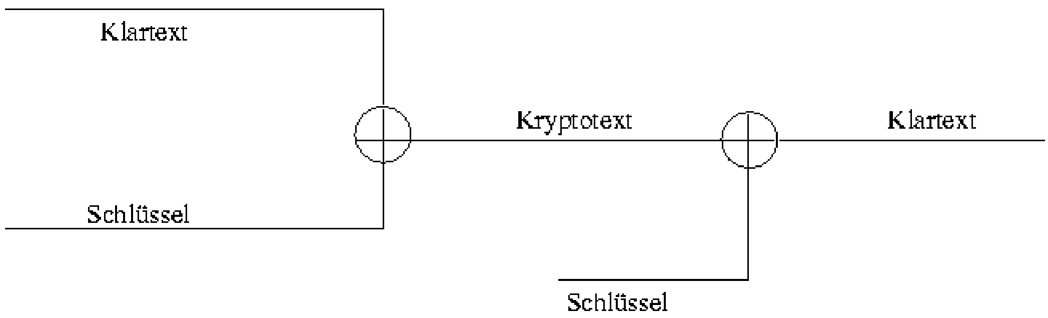
1.1 Anwendung von Zufallszahlen

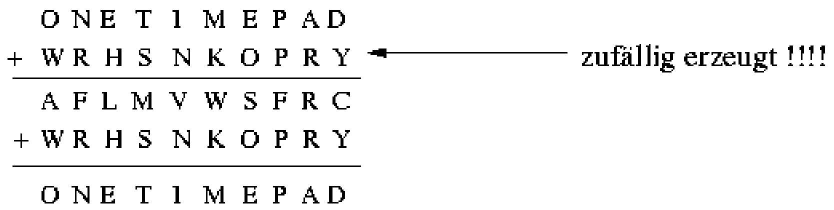
Es gibt eine Reihe von Feldern, auf denen Zufallszahlen Anwendung finden, zum Beispiel

- bei der Simulation von Vorgängen (Warteschlangenmodelle etc.),
- bei Stichproben, um ein typischen Verhalten einer Gesamtheit zu überprüfen,
- in der Numerik,
- bei der Programmierung (Zufallszahlen als Eingabedaten für Tests),
- in Randomisierten Algorithmen (Treffen von Entscheidungen)
- in der Kryptographie (One-Time Pad)
- oder beim Glücksspiel (Roulette, Würfeln, ...).

1.2 Der One-Time Pad als Anwendung von Zufallszahlen in der Kryptographie

"Das" sichere Verfahren zur Verschlüsselung von Daten ist der One-Time Pad. Hierbei wird zu jedem Klartextzeichen ein zufällig erzeugter Wert addiert. Der legale Empfänger addiert dann zu dem Empfangenen Kryptotext den Schlüsselstrom und erhält wiederum den Klartext. Wichtig ist dabei, dass der sogenannte Schlüsselstrom nur einmal verwendet werden darf, da sonst Regelmäßigkeiten auftauchen, die sich ein Kryptoanalytiker zu Nutze machen kann. Beachtet man dies, so ist dieses Verfahren absolut sicher. Das Problem hierbei ist nur, dass man für jedes Zeichen das man verschlüsseln möchte, ein zufällig erzeugtes Zeichen benötigt, das man über einen sicheren Kanal mit dem legalen Empfänger austauschen muss, was einen hohen Aufwand mit sich bringt.





1.3 Eigenschaften von Zufallszahlen

Zunächst kann man sich die Frage stellen, was eigentlich “zufällig” bedeutet. Ist zum Beispiel die Zahl 5 eine zufällige Zahl ? Man betrachtet eher eine Sequenz von zufälligen Zahlen aus einer bestimmten Menge mit einer vorgegebenen Verteilung. Oft, wie zum Beispiel beim Roulette, wünscht man sich eine Gleichverteilung, das heißt, dass jede Zahl aus der Menge mit der gleichen Wahrscheinlichkeit auftritt. Bei einer Sequenz von Dezimalzahlen hätte dann jede Zahl von 0 bis 9 die Wahrscheinlichkeit $\frac{1}{10}$. Es kommt natürlich in jeder dieser Sequenzen nur sehr selten jede Zahl gleich oft vor. Betrachten wir die Generierung einer Sequenz von einer Million Dezimalzahlen. Nehmen wir an, die ersten 999.999 Zahlen waren alle Nullen. Die Wahrscheinlichkeit, dass nun auch die letzte Zahl eine Null wird, ist wieder genau $\frac{1}{10}$. Es ist daher nicht einfach, zu entscheiden, ob eine gegebene Sequenz zufällig erzeugt wurde oder nicht. Man kann nur eine Wahrscheinlichkeit angeben, mit der man von einer Sequenz sagen kann, ob sie zufällig erzeugt wurde, oder nicht.

Früher wurden Zufallszahlen durch das Ziehen einer Karte, das Werfen einer Münze oder durch Würfeln erzeugt. Später gab es Tabellen von Zufallszahlen. So erschien 1927 eine solche mit 40.000 Zufallszahlen. 1939 konstruierte Kendall eine Maschine, die eine Tafel mit 100.000 Zufallszahlen erzeugte und 15 Jahre später gab es bereits eine, die 1 Million Zufallszahlen erzeugte. Seit der Einführung von Computern, sucht man aber nach effizienteren Wegen Zufallszahlen mit Computern zu generieren. Eine Möglichkeit sind Tabellen, in denen man Werte nachschlägt. Diese benötigen aber viel Speicherplatz und man erhält nur kurze Zahlenfolgen.

1.4 Pseudozufallsfolgen

Neumann hatte die Idee, die arithmetischen Operationen eines Computers auszunutzen, um Folgen von Zufallszahlen zu erzeugen. So entwickelte er die "Middle Square" Methode. Möchte man beispielsweise Zufallszahlen zwischen 100 und 999 generieren, so wählt man zunächst zufällig eine Zahl (Tabelle, Uhrzeit,...) zwischen 100 und 999. Sei diese in unserem Beispiel X_0=125. X_1 ensteht nun aus X_0 wie folgt: Man quadriert X_0 und erhält 15625; nun wählt man die drei mittleren Ziffern dieser Zahl und erhält X_1= 562. X_1^2 ist 315844 und daraus ergibt sich für X_2 der Wert 158 oder 584 (es muss vorher festgelegt werden, ob man die "linke" Mitte, oder die "rechte" Mitte wählt). Genau daran sieht man aber schon, dass bis auf die erste Zahl alle folgenden Zahlen deterministisch erzeugt werden. Ist X_0 festgelegt, so ist auch die gesamte Folge $\langle X_n \rangle$ eindeutig bestimmt. Solch eine Folge scheint also nur eine zufällig erzeugte Folge zu sein. Daher spricht man von Pseudozufallsfolgen.

Pseudozufallsfolgen sind Folgen von Zahlen, die zwar deterministisch bestimmt werden, jedoch wichtige Eigenschaften von wirklichen Zufallsfolgen aufweisen sollen. Man kann dann, je nach dem, ob die Pseudozufallsfolge einer zufällig erzeugten Folge ähnelt oder eher nicht, von guten und schlechten Pseudozufallsfolgen sprechen. Ein Zufallsgenerator ist ein deterministischer Algorithmus, der eine "kleine" Menge von wirklichen Zufallszahlen als Eingabe nimmt (diese werden "seed" genannt) und daraus eine "große" Menge von Pseudozufallszahlen generiert.

Die "Middle Square" Methode scheint nur oberflächlich "gut" zu sein. In Wahrheit hat sie jedoch mehrere Mängel. Taucht zum Beispiel eine Null in einem der X_i auf, so werden auch viele weitere X_r eine Null enthalten.

Zum Beispiel wenn X_0 609 ist :

$$X_0=609,\ X_0^2=370881$$
$$X_1=708,\ X_1^2=501264$$
$$X_2=012,\ X_2^2 = \text{...........}$$

Im Folgenden soll es nun darum gehen, Algorithmen vorzustellen, die "besser" sind, als die "Middle Square" Methode.

2 Lineare Kongruenzgeneratoren

Die Linearen Kongruenzgeneratoren wurden bereits 1949 entwickelt und stellen immer noch die bei weitem beliebteste Methode zur Generierung von Pseudozufallszahlen dar. Zur Generierung einer solchen Folge werden vier Zahlen gewählt. Von diesen Zahlen hängt ab, wie gut die Pseudozufallsfolge sein wird. Wie man diese Zahlen geschickt auswählt, soll dann später erläutert werden.

2.1 Das Prinzip der Linearen Kongruenzgeneratoren

Die folgenden Zahlen werden im gesamten Text mit den nun eingeführten Variablen gebraucht.

m Modul

a Faktor

c Inkrement

X_0 Startwert

Die Folge $\langle X_n \rangle$ von Pseudozufallszahlen ensteht nun induktiv wie folgt:

$$X_{n+1} = (aX_n + c) \bmod m, \text{ für alle } n \geq 0$$

Zur Veranschaulichung sollen nun ein paar Beispiele folgen.

1. Wir wählen m=10 und X_0=a=c=7. Daraus ensteht die Folge
   7, 6, 9, 0, 7, 6, 9, 0, ...

2. Wir wählen m=13 und X_0=a=c=3. Daraus entsteht die Folge
   3, 12, 0, 3, 12, 0, ...

3. Wir wählen m=7 und X_0=0, a=1, sowie c=1. Daraus entsteht
   0, 1, 2, 3, 4, 5, 6, 0, 1, ...

4. Wir wählen m=9, X_0=1, c=5 und a=7. Dann erhalten wir
   1, 3, 8, 7, 0, 5, 4, 6, 2, 1, 3, ...

Wie man sehen kann, entsteht bei jeder dieser Folgen eine Schleife/Periode. Hier haben wir Perioden der Länge vier, drei, sieben bzw. neun. Die Voraussetzung dafür, dass eine Folge $<X_n>$ gut ist, ist, dass sie eine lange Periode hat. In dieser Beziehung sind Beispiel 3) und 4) optimal. Die Periode kann aber nie länger sein, als m. Daher sollte m stets größer als die erwartete Anzahl benötigter Zufallszahlen sein. Aber wie wir an Beispiel 3) sehen können ist die Länge der Periode nicht das einzige Kriterium für eine "gute"

Folge. Die Periode ist hier zwar maximal (bzgl. m), aber die Folge erscheint trotzdem keineswegs zufällig.

Auch wird es darauf ankommen die vier Zahlen so zu wählen, dass nicht nur eine "gute" Folge entsteht, sondern auch so, dass möglichst viele Zufallszahlen möglichst schnell erzeugt werden können. Bei den meisten Anwendungen werden nämlich extrem viele Zufallszahlen benötigt. Daher ist auch die Zeit, die zu Ihrer Erzeugung benötigt wird, wichtig.

Nun wollen wir uns den vier Zahlen widmen, vor allem aber m,c und a und untersuchen, wie wir sie geschickt wählen, um unsere Ziele zu erreichen.

2.2 Die Periodenlänge

Wie schon erwähnt ist die maximale Periodenlänge m. Man sollte dies bei der Wahl von m von vornherein berücksichtigen. Eine lange Periode ist zwar für eine gute Zufallsfolge notwendig, aber nicht hinreichend. Jedoch sind die meisten Perioden kürzer als m. Wie man nun durch geschickte Wahl der anderen Parameter auf die maximale Periodenlänge kommt, wird in folgendem zentralen Satz gezeigt:

Der Lineare Kongruenzgenerator $\mathrm{X}_{i+1} = \mathrm{aX}_i + \mathrm{c} \bmod \mathrm{m}$ erzeugt genau dann alle Reste modulo m, hat also Periodenlänge m, wenn die folgenden Bedingungen erfüllt sind

1. ggT (c,m) = 1
2. b = a -1 ist ein Vielfaches von p, für alle Primfaktoren p von m
3. b ist ein Vielfaches von 4, wenn m ein Vielfaches von 4 ist.

Mit Hilfe dieses Theorems kann man nun sowohl einen Linearen Kongruenz-Generator konstruieren, der Periodenlänge m hat, als auch von einem gegebenen Generator sagen, ob er optimal in Bezug auf die Periodenlänge ist. Es soll noch einmal darauf hingewiesen werden, dass auch die Folge 0, 1, 2, ..., m-1, 0, 1,.... optimal in Bezug auf die Periodenlänge ist, jedoch keineswegs den von uns intuitiv geforderten Kriterien für eine Pseudozufallsfolge entspricht. Man beachte deshalb auch die in 2.3 angegebenen Spezialfälle. Noch ein Hinweis : die Periodenlänge hängt bei maximaler Periodenlänge m wie erwartet nicht vom Startwert X_0 ab.

Beweis des Satzes (in Auszügen)

Zunächst muss man das folgende Lemma zeigen:

Sei $m = p_1^{e_1} \cdot .... \cdot p_t^{e_t}$ die Primfaktorenzerlegung von m. Die Periodenlänge λ einer durch $(X_0,\ a,\ c,\ m)$ definierten linearen Kongruenzfolge ist das kgV der Periodenlängen der durch (X_0 mod $p_j^{e_j}$, a mod $p_j^{e_j}$, c mod $p_j^{e_j}$, $p_j^{e_j}$), $1 \leq j \leq t$, definierten linearen Kongruenzfolgen.

Dieses Lemma zeigt man durch Induktion über t, wobei man als Induktionsanfang für m=$m_1 m_2$ mit ggT(m_1, m_2) = 1 wählt. Nun muss man den Satz nur noch für alle m der Form m=p^e, p prim, zeigen. Die Periodenlänge λ ist genau dann gleich m, wenn jede mögliche Zahl $0 \leq x < m$ in der Periode auftaucht, da kein Wert in der Periode zweimal auftaucht. Man nimmt o.B.d.A. nun an, dass X_0=0 ist und erhält dann für X_n = $\left(\frac{a^n-1}{a-1}\right)$ c mod m. Wenn nun ggT(c,m) $\neq 1$ wäre, dann könnte X_n nie 1 werden. Daher braucht man Bedingung 1 aus dem Satz. Der kleinste Wert für n>0, für den $X_n = X_0 = 0$ ist muss nun m sein. Um dies zu zeigen wird die Bedingung 1 und die Gleichung für X_n verwendet, um abschließend noch folgendes Lemma zu zeigen:

Sei 1 < a < p^e, p prim. Sei λ die kleinste positive natürliche Zahl, für die $\frac{a^\lambda-1}{a-1} \equiv 0$ mod p^e, dann ist $\lambda = p^e$ genau dann wenn

$$\begin{cases} a \equiv 1 \ mod \ p, \ wenn \ p > 2 \\ a \equiv 1 \ mod \ 4, \ wenn \ p = 2. \end{cases}$$

Hier werden also die Bedingungen zwei und drei in den Beweis eingebracht. Dies ist also in groben Zügen der Beweis dieses zentralen Satzes.

Nun aber zu einigen Beispielen:

- Unser voriges Beispiel mit m=10, X_0 = a = c = 7 hatte eine Periodenlänge von 4≤m. Es gilt zwar Bedingung 1. (ggT(7,10)=1), aber a - 1 = 6 ist kein Vielfaches der Primfaktoren 2 und 5 von m. Daher ist Bedingung 2. verletzt und die Periode kann nicht die Länge 10 haben.

- a = 24298, c = 99991, m = 199017 = $3^7 \cdot 7 \cdot 13$. Da werder 3, 7 noch 13 Teiler von c sind ist die Bedingung 1. erfüllt. Da a-1 = 24297 = 8099·3, 24297 = 3471 ·7 und 24797 = 1869 · 13 ist auch Bedingung 2. erfüllt. Bedingung 3. gilt trivialer Weise. Wir erhalten also eine optimale Periodenlänge m. Dieser Zufallsgenerator ist der von der Firma Texas-Instruments im Taschenrechner TI-59 eingesetzte Zufallsgenerator.

2.3 Was bei der Wahl von a,c und m noch zu beachten ist

Es sollen nun noch (verkürzt) einige wichtige Regeln für die Wahl der Parameter (abgesehen von der Periodenlänge) angegeben werden, die in der Praxis wichtig sind. Der

Startwert ist problemlos frei wählbar, was auch wichtig ist, da dieser Wert ja wirklich zufällig erzeugt werden soll (Uhr).

2.3.1 Die Wahl von a

Die Wahl von a=0 ist wie schon erwähnt offensichtlich unsinnig, da hier die Folge X_0, b mod m, b mod m, ... entsteht. Aber auch für a=1 erhalten wir eine unbrauchbare Folge, nämlich X_0, X_0 + b mod m, X_0 + 2b mod b, X_0 + 3b mod m, ...

2.3.2 Die Wahl von c

Ist a=1 und c=0 so erhalten wir die konstante und daher schlechte Folge X_i=X_0. Allgemein sprechen wir bei c=0 von einem sogenannten multiplikativen Kongruenzgenerator, der nicht unbedingt unbrauchbar ist.

2.3.3 Die Wahl von m

Oft wird für m einfach die Wortlänge des zu benutzenden Computers gewählt, da somit die Operation mod m trivial und damit schnell ausfürbar ist, was für die Erzeugung von langen Zufallsfolgen wichtig ist. Man muss auf jeden Fall beachten, dass m größer ist als die Anzahl der gewünschten Zufallszahlen, da sonst eine Regelmäßigkeit auftreten wird (wegen der Periode). Dies wäre vor allem in der Kryptographie fatal (siehe One-Time Pad).

3 Statistische Tests

3.1 Der χ^2-Test

Lineare Kongruenzgeneratoren wurden theoretischen Betrachtungen unterzogen. Bei empirischen Untersuchungen ergab sich, dass diese Generatoren gute Pseudozufallsfolgen erzeugen. Daher ist es sinnvoll solche Generatoren zu verwenden und sie den vielen schwer bis gar nicht analysierbaren vorzuziehen. In vielen statistischen Tests wurde die Güte der LKG gezeigt.

Wir haben gesehen, dass man bei der Wahl der drei Parameter darauf achten muss, dass die maximale Periodenlänge m erreicht wird und dass man einige triviale ungünstige Fälle leicht ausschließen kann. Wir wollen aber auch zeigen, wie man von einem LKG sagen kann, ob er nun hinreichend zufällige Zahlenfolgen erzeugt. Wir haben einige Folgen gesehen, von denen wir intuitiv sagen würden, dass sie “zufällig” sind. Das reicht aber in den meisten Fällen nicht aus. Wir benötigen objektive Tests, die uns bestätigen (oder auch nicht),ob ein LKG wirklich unseren Erwartungen entspricht. Wir können nun zwar einige Tests ausführen, die vermuten lassen, dass eine gewisse Folge wohl zufällig ist. Ein weiterer Test könnte diese Vermutung jedoch ablehnen. Ganz sicher können wir also nie sein, jedoch gibt uns das Durchführen einiger Tests eine größere Sicherheit. Wir verwenden daher eine Teststatistik, um für eine Folge von Zahlen zu zeigen, ob diese eher zufällig ist, oder nicht.

Der χ^2-Test ist einer der bekanntesten statistischen Tests. An einem Beispiel soll kurz die Arbeitsweise dieses Tests erläutert werden. Wir betrachten das Werfen von zwei Würfeln und deren Augensumme s:

Wert von s	2	3	4	5	6	7	8	9	10	11	12
Wahrscheinlichkeit von s	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Wenn wir nun diese Würfel n mal werfen, erwarten wir, dass die Augensumme s ungefähr np_s mal auftaucht. Bei 120 Durchführungen erwarten wir den Wert 4 ca. 10 mal. Wir führen nun 144 mal unser Experiment durch und nennen Y_s die von uns beobachtete Anzahl und np_s die erwartete:

Wert von s	2	3	4	5	6	7	8	9	10	11	12
Beobachtete Anzahl Y_s	2	4	10	12	22	29	21	15	14	9	6
Erwartete Anzahl np_s	4	8	12	16	20	24	20	16	12	8	4

Die beobachtete Zahl entspricht, wie wir sehen können nur selten der erwarteten Anzahl. Alle 36^{144} mögliche Folgen sind gleich wahrscheinlich. Trotzdem würde man das vorherige Experiment wohl als zufällig ansehen. Angenommen wir hätten es aber nun mit manipulierten Würfeln zu tun. Können wir dies entscheiden ? Könnte unser Experiment nicht auch mit manipulierten Würfeln durchgeführt worden sein, die etwa die Augensumme 12 oder 7 begünstigen ? Mann kann diese Frage nicht eindeutig mit ja oder nein beantworten, aber wir können eine wahrscheinliche Antwort angeben und auch sagen wie wahrscheinlich diese ist. Wir addieren zu diesem Zweck die Quadrate der Differenzen zwischen den Y_i und den np_i (wir nehmen die Quadrate, um Abweichungen in beide Richtungen feststellen zu können). Wir erhalten dann

$$V = \sum_{i=1}^{12} (Y_i - np_i)^2.$$

Haben wir nun zwei “schlechte” Würfel, so wird der Wert V höher sein, also für zwei “gute”. Nun können wir uns fragen, wie wahrscheinlich es ist, dass V so hoch ist und die Würfel regulär sind. Wenn diese Wahrscheinlichkeit sehr klein wird, zum Beispiel 5%, dann wissen wir, dass sich in ungefähr fünf von 100 Fällen ein so großes V ergeben kann. Wir können nun den Test öfter wiederholen, um diese Wahrscheinlichkeit noch zu drücken.

Nun müssen wir noch berücksichtigen, dass die einzelnen $(\mathrm{Y}_i - \mathrm{np}_i)^2$ alle gleich gewichtet sind, obwohl $(Y_7 - np_7)^2$ wahrscheinlich deutlich höher ist, als $(Y_2 - np_2)^2$, da sieben als Augensumme wesentlich wahrscheinlicher ist als die zwei. Wir gewichten daher die einzelnen Summanden in unserer Formel wie folgt:

$$\sum_{i=2}^{12} \frac{(Y_i - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=2}^{12} \left(\frac{Y_s^2}{p_s} \right) - n$$

Dies ist nun die χ^2-Statistik für unsere Beobachtungen Y_i. Für unser Experiment ergibt sich nun für $V = \frac{(2-4)^2}{4} + \frac{(4-8)^2}{8} + ... + \frac{(9-8)^2}{8} + \frac{(6-4)^2}{4} = 7\frac{7}{48}$. Ist dieser Wert nun hoch oder niedrig bzw. gut oder schlecht ? Dazu benötigen wir eine Tabelle der χ^2-Verteilung, die hier auszugsweise angegeben wird.

	p = 1%	p = 5%	p = 25%	p = 50%	p = 75%	p = 95%	p = 99%
$\nu = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu = 2$	0.02010	0.1026	0.5753	1.386	2.773	5.991	9.210
ν =3	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu = 15$	5.229	7.261	11.04	14.34	18.25	25.00	30.58
$\nu = 20$	8.260	10.85	15.45	19.34	23.83	31.41	37.57
$\nu = 30$	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$\nu = 50$	29.71	34.76	42.94	49.33	56.33	67.50	76.15
$\nu = 16383$	15964.8	16086.4	16260.5	16382.3	16504.7	16681.9	16807.0

Was kann man nun aus dieser Tabelle erfahren ? Zunächst einmal haben wir k = 11 mögliche Ausgänge unseres Experiments. Daraus berechnen sich die Freiheitsgrade ν = k -1. Wir müssen also in der Spalte $\nu = 10$ nachsehen. In der Wert x der Spalte p sagt nun, dass der Wert V mit Wahrscheinlichkeit p kleiner oder gleich x ist. Unser Wert für V war 7.15. Mit einer Wahrscheinlichkeit von 1 - p = 1 - 0.25 = 0.75 ist der Wert bei $\nu = 10$ größer als 6.737. Unser wert liegt also ganz gut. Noch kurz zwei Beispiele zum selben Experiment (wieder 144 "Versuche"):

Wert von s	2	3	4	5	6	7	8	9	10	11	12
Y_{s_1}	4	10	10	13	20	18	18	11	13	14	13
Y_{s_1}	3	7	11	15	19	24	21	17	13	9	5
np_s	4	8	12	16	20	24	20	16	12	8	4

Im ersten Fall ergibt sich $V_1 = 29\frac{59}{120} \approx 29,492$ und im zweiten errechnet man $V_2 = 1\frac{17}{120} \approx 1.142$. Aus der Tabelle erfahren wir, dass nur in 1% der Fälle der Wert größer als 23.21 ist. Daher wissen wir, dass das Experiment 1 fast sicher mit gezinkten Würfeln ausgeführt wurde, da V_1 größer als 23.21 ist und dass das zweite Experiment fast sicher mit echten Würfeln ausgeführt wurde. Auch aus den Ausgängen des Experiments hätte man das vielleicht folgern können: man sieht, dass der erste Fall sehr stark von np_s abweicht, besonders bei der Augensumme 12, dass aber das zweite Experiment nur geringfügig von unseren erwarteten Werten abweicht.

Man nimmt an, dass man es mit gezinkten Würfeln zu tun hat, wenn der Wert p für unser V und das ν größer als 95% ist. (Genauer gesagt: wir lehnen dann die Hypothese, dass wir es mit echten Würfeln zu tun haben ab). Wir müssen die Würfel bzw. den LKG aber auch verwerfen, wenn wir einen Wert kleiner als 5% erhalten, weil dann die Verteilung zu regelmäßig ist.

Abschließend ist zu sagen, dass dieser Test relativ zuverlässig ist, aber dass darauf geachtet werden muss, dass die Anzahl der Versuche hoch ist (jedes np_s sollte mindestens fünf sein).

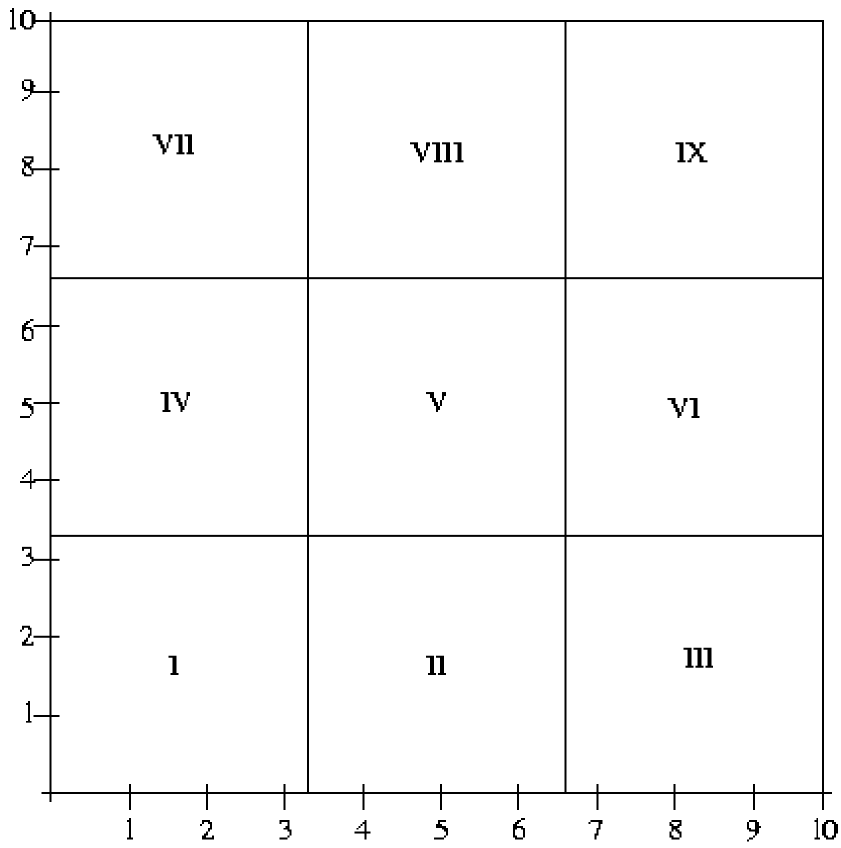
Wie kann man nun mit Hilfe dieses Tests von einem LKG sagen, ob er für uns "gute" Zufallsfolgen generiert ? Zunächst muss noch hinzugefügt werden, dass der χ^2 -Test nicht der einzige Test ist der durchgeführt werden muss, um die Güte eines LKG zu zeigen, es gibt noch Tests auf Unabhängigkeit etc. Ein zu testender LKG muss vielmehr *alle* Tests bestehen. Hier soll aber nur als Beispiel der χ^2-Test für LKG erläutert werden.

3.2 Höherdimensionale Zufallsfolgen

Auch eine Zufallsfolge mit maximaler Periodenlänge m ist überhaupt nicht zufällig. Wir erhalten jede Zahl zwischen 0 und m genau einmal, was sicher von uns nicht als zufällig angesehen werden kann. Deswegen sind in der Praxis höherdimensionale Zufallsfolgen anzuwenden, bei denen auch mal die gleiche Zahl zweimal hintereinander auftritt, oder wir aber eine Teilfolge der Form 1, 2, 3 erhalten.

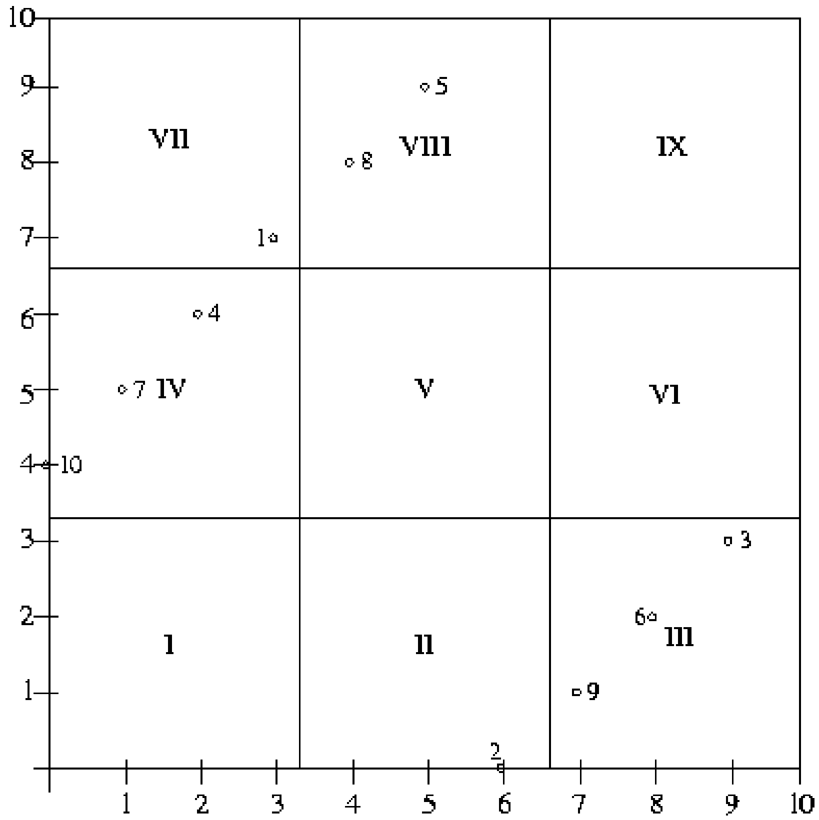
Hierzu erzeugt man mit einem LKG n Zufallsfolgen der Länge m, wobei der Startwert X_0 stets wirklich zufällig gewählt wird. Wir erhalten dann m n-Tupel. Wir teilen dann das Koordinatensystem in k Intervalle ein und tragen die einzelnen Tupel hier ein. Die erste nun hoffentlich wirklich zufällige Zahl erhalten wir wie folgt: das r-te Tupel, das in das i-te Intervall fällt ist unsere r-te Zufallszahl, die den Wert i erhält.

An einem Beispiel (im $\Re^2$) soll diese Art der Generierung nun durchgeführt werden und anschließend auf das Ergebnis der χ^2-Test angewandt werden. Wir wählen zunächst einen LKG der eine maximale Periode erzeugt. Sei a=11, c=3 und m=10. Die drei Bedingungen aus dem Satz sind erfüllt, da ggT(c,m)=1 und $a \equiv 1$ mod p für die zwei Primfaktoren 2 und 5 von 10 ist. Für die dritte Bedingung fehlt die Voraussetzung. Wir teilen nun das Koordinatensystem in neun gleich große Intervalle. Das erste hat die Eckpunkte (0,0), $\left(3\frac{1}{3},\ 0\right)$, $\left(0,\ 3\frac{1}{3}\right)$, $\left(3\frac{1}{3},\ 0\right)$, das zweite $\left(3\frac{1}{3},\ 0\right)$, $\left(6\frac{2}{3},\ 0\right)$, $\left(6\frac{2}{3},\ 3\frac{1}{3}\right)$ und $\left(3\frac{1}{3},\ 3\frac{1}{3}\right)$. Welcher Eckpunkt zu welchem Intervall gehört, ist hier egal, da wir nur ganzzahlige Koordinaten eintragen, diese aber nie zu zwei Quadraten gehören können. Hier die zugehörige Zeichnung:



Nun müssen wir noch zufällig zwei Startwerte wählen. Wir erhalten zum Beispiel für $X_{0_1} = 3$ und damit die Folge 3, 6, 9, 2, 5, 8, 1, 4, 7, 0 und für $X_{0_2} = 7$ die Folge 7, 0, 3, 6, 9, 2, 5, 8, 1, 4. Unsere daraus entstehenden Tupel lauten $\begin{pmatrix} 3 \\ 7 \end{pmatrix}$, $\begin{pmatrix} 6 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$,

$\begin{pmatrix} 2 \\ 6 \end{pmatrix}, \begin{pmatrix} 5 \\ 9 \end{pmatrix}, \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 \\ 8 \end{pmatrix}, \begin{pmatrix} 7 \\ 1 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 4 \end{pmatrix}$. Wir tragen diese nun in das Koordinatensystem ein und achten dabei auf die Reihenfolge.



Wir erhalten daher als Zufallsfolge 7, 2, 3, 4, 8, 3, 4, 8, 3, 4, was auf den ersten Blick gar nicht besonders zufällig aussieht, was auch an der sogenannten Gitterstruktur der LKG liegt. Wir können jetzt diese Folge mit dem χ^2-Test überprüfen. Müssen wir diese folge als nicht zufällig ablehnen, so ist vielleicht unsere Koordinateneinteilung schlecht, oder aber einer unserer Parameter.

Werte	I	II	III	IV	V	VI	VII	VIII	IX
Beobachtete Anzahl	0	1	3	3	0	0	1	2	0
Erwartete Anzahl np	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$	$\frac{10}{9}$

Jetzt rechnen wir das gewichtete V aus. Wir erhalten $V = \frac{\left(0-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(1-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(3-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(3-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(0-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(0-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(1-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(2-\frac{10}{9}\right)^2}{\frac{10}{9}} + \frac{\left(0-\frac{10}{9}\right)^2}{\frac{10}{9}} = 11.6$.

Unser ist$\nu = 9 - 1 = 8$. Hier nun der Tabelleneintrag für acht Freiheitsgrade:

	p = 1%	p = 5%	p = 25%	p = 50%	p = 75%	p = 95%	p = 99%
$\nu = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09

In immerhin 25% der Fälle ist der Wert größer als 10.22 und in 5% der Fälle größer als 15.51. Unser Wert 11.5 ist also gar nicht so schlecht. Wir können sagen, dass dieser Test unsere Folge nicht als nicht-zufällig ablehnen würde. Ob das für alle Folgen dieses Generators gilt, oder ob andere Tests wegen der auffälligen Regelmäßigkeit unserer Folge und damit vielleicht auch den gesamten LKG als "schlecht" ablehnt, steht auf einem anderen Blatt.

Diese gewisse Regelmäßigkeit in der Folge liegt an den großen Quadraten und an der kleinen Dimension. Man kann diese erheblich erhöhen, was aber unser Beispiel zu umfangreich gemacht hätte. Auch erzeugt der im Verhältnis zum Faktor a kleine Inkrement c zu einer solch offensichtlichen Regelmäßigkeit. Man muss also die Parameter a, c, m, Dimension und Quadratgröße verändern, um bessere Resultate zu erlangen.

4 Literatur

D.E. Knuth: The Art of Computer Programming, Band 2, Addison-Wesley, 1982

Lothar Afflerbach: Lineare Kongruenz-Generatoren, Dissertation, Darmstadt, 1983

Julius Duffner: Statistik mit SAS, 2. Auflage, Teubner, Stuttgart 2002

Mathar, Pfeifer: Stochastik für Informatiker, Teubner, Stuttgart 1990