Zeitschrift für Anomalistik Band 3 (2003), S. 56-67

Kann ein Pseudozufallsgenerator eine akausale Korrelation vortäuschen?

VOLKER GUIARD

Zusammenfassung – Die in den Experimenten von Hagel und Tschapke (2002) gefundenen akausalen Korrelationen werden auf die Möglichkeit hin untersucht, ob es sich um durch Autokorrelationen in Pseudo-Zufallsgeneratoren bedingte Artefakte handeln könnte. Diese Vermutung konnte jedoch nicht bestätigt werden, so dass für die von Hagel und Tschapke dokumentierte Anomalie nach wie vor eine Erklärung aussteht.

Schlüsselbegriffe: Akausale Korrelation – Zufallszahlengeneratoren – Autokorrelation – Artefakte – Interaktions-Anomalie – Parapsychologie

Can a pseudo-random number generator create the illusion of an acausal correlation?

Abstract – The acausal correlations found in the experiments by Hagel and Tschapke (2002) are examined whether they can be explained as artefacts due to autocorrelations in pseudo-random number generators. However, this supposition was not confirmed, so the anomaly documented by Hagel and Tschapke is still waiting for an explanation.

Keywords: acausal correlation – random number generators – autocorrelation – artefacts – interaction anomaly – parapsychology

Problemstellung

Im Experiment von Hagel und Tschapke (2002) wurde durch einen Pseudozufallsgenerator über die Stellung einer Weiche in einer Modelleisenbahnanlage entschieden. In Abhängigkeit von der Weichenstellung wird eine Modelllokomotive über den Außenkreis oder über eine Abkürzung fahren. Im Außenkreis erleidet die Lokomotive eine starke Belastung durch eine zweimalige Stromumpolung, wodurch die Lok aus voller Fahrt zweimal die Fahrtrichtung wechselt. Während des weiteren Kreisdurchlaufs durchfährt die Lok einen Schienenkontakt, wodurch die Abfrage der nächsten Zufallszahl und damit die Entscheidung über die Weichenstellung ausgelöst wird. Obwohl der Pseudozufallsgenerator theoretisch beide Weichenstellungen mit der gleichen Wahrscheinlichkeit p=0,5 liefert, zeigte sich im Experiment, dass über eine Dauer von 100.000 Umläufen der belastende Außenkreis etwas seltener durchfahren wurde. Dieses führte zu der Interpretation, dass die Lokomotive selbst über den genauen Zeitpunkt der Schienenkontaktdurchfahrt „entscheidet“, und damit auch darüber, welche Zufallszahl abgerufen wird. Will man dieser Interpretation nicht voreilig zustimmen, so hat man nach anderen Ursachen für diesen Effekt zu suchen. Eine der Ursachen könnte darin bestehen, dass der Pseudozufallszahlengenerator deterministisch und nicht zufällig ist und somit unerwartete Abweichungen von der theoretisch unterstellten Zufallsverteilung aufweisen könnte. Diesem Verdacht wird im Folgenden nachgegangen.

Der Pseudozufallszahlengenerator

Hagel und Tschapke (2002) verwenden in ihrem Experiment einen Pseudozufallszahlengenerator der Programmiersprache MS-QBASIC45. Solch ein Generator hat eine Zykluslänge M, d.h. nach M Zahlen werden alle bisherigen Zahlen in gleicher Reihenfolge wiederholt. Siebenborn (2002) wies darauf hin, dass für den hier betrachteten Generator $M = 2^{24} = 16777216$ gilt. Dieses konnte vom Autor bestätigt werden. Weiterhin wurde ermittelt, dass es sich um einen gemischten Generator mit der Rekursionsformel

$$c_{i+1} = \mathrm{mod}(a \cdot c_i + b\,,M) \qquad (1)$$

handelt, wobei die „modulo“-Funktion $\mathrm{mod}(x, z)$ den Rest liefert, welcher sich ergibt, wenn man x ganzzahlig durch z dividiert. Für a und b ergaben sich die Werte

$$a = 16598013 \quad \text{bzw.} \quad b = 12820163 \qquad (2)$$

(siehe Anhang 1). Dieser Generator gehört zu einem Generatorentyp, der durch folgende Bedingungen charakterisiert ist:

$$\left.\begin{array}{c} M = 2^K, \quad K > 2 \\ \mathrm{mod}(a,4) = 1 \\ b \text{ ungerade} \end{array}\right\} \qquad (3)$$

Bei einem Generator dieses Typs durchlaufen die c_i-Werte in irregulärer Reihenfolge alle Werte von 0 bis $M-1$ (Zielinski 1972; Knuth 1997). Die eigentlichen „Zufallszahlen“ ergeben sich aus $z_i = c_i / M$ womit $0 \le z_i < 1$ folgt. Die c_i-Werte von 0 bis $M/2-1$ machen genau die Hälfte des Zyklusumfanges aus, die c_i-Werte von $M/2$ bis $M-1$ bilden die andere Hälfte. Definiert man die Variable B wie folgt:

$$B = -1 \text{, falls } z < 0{,}5 \qquad \text{bzw.} \qquad B = +1 \text{, falls } z \ge 0{,}5\ , \qquad (4)$$

dann wird man bei einem Zyklusdurchlauf beide B-Werte genau gleich oft erhalten bzw. die (Pseudo-)Wahrscheinlichkeit des Ereignisses „B=1“ ist genau p=0,05. Hagel und Tschapke (2002) definierten B gemäß

$$B = -1 \text{, falls } 0 < z < 0{,}5 \ (\Rightarrow \text{ Innenkreis})$$
$$B = 1 \text{ , falls } 0{,}5 < z < 1 \ (\Rightarrow \text{ Außenkreis}).$$

Hierbei bleibt offen, welche Entscheidung in den Fällen $z = 0$ oder $z = 0{,}5$ gefällt wurde. Dieses dürfte aber keinen wesentlichen Einfluss auf das Endergebnis gehabt haben.

Es gilt, sofern man B als zufällig akzeptiert, dass B den Erwartungswert Null und die Varianz 1 hat. Aus der Tatsache, dass während eines kompletten Zyklusdurchlaufes die Werte $B = 1$ und $B = -1$ genau gleich oft auftreten, zeigt sich die Besonderheit eines Pseudo-Zufallsgenerators. Denn üblicherweise würde man für die Summe von M B-Werten die Varianz $V\left(\sum_{i=1}^{M} B_i\right)=M$ erwarten, bei einem kompletten Zyklusdurchlauf folgt aber

$$V\left(\sum_{i=1}^{M} B_i\right)=V(0)=0\,. \tag{5}$$

Wodurch könnte eine akausale Korrelation verursacht werden ?

Wie im vorherigen Abschnitt gezeigt wurde, ist die Wahrscheinlichkeit der beiden Werte von B (die den Weichenstellungen entsprechen) genau gleich. Hierin kann also keine Ursache für die akausale Korrelation gesehen werden. In dem Experiment werden aber die Zufallszahlen mit einem gewissen zeitlichen Abstand nacheinander aufgerufen. In diesem Zeitintervall mögen L (=lag) weitere Zufallszahlen erzeugt worden sein, wovon die letzte wieder zur Weichenstellung verwendet wird. Es entsteht nun die Frage, ob ein Zufallswert B_i mit dem L-ten Wert B_{i+L} der darauf folgenden Werte korreliert ist. Die entsprechende Autokorrelation bezüglich des Abstandes L ergibt sich aus

$$R(L)=\operatorname{cov}(B_i,B_{i+L})/\sqrt{V(B_i)\cdot V(B_{i+L})}=\frac{1}{2}\sum_{i=1}^{M} B_i\cdot B_{i+L}$$

$$=\frac{1}{M}\left[S_{--}-S_{-+}-S_{+-}+S_{++}\right]. \tag{6}$$

Dabei ist S_{--} die Anzahl Wertepaare $(B_{i+L},B_i)=(-1,-1)$. Die weiteren S sind analog definiert. Diese Summen können in einer 2×2-Tafel dargestellt werden:

Tabelle 1: 2×2-Tafel der Häufigkeiten der Paare (B_{i+L},B_i)

		$B_i=-1$	+1	Summe
$B_{i+L}=$	-1	S_{--}	S_{-+}	$M/2$
	+1	S_{+-}	S_{++}	$M/2$
	Summe	$M/2$	$M/2$	M

Da $S_{--}+S_{-+}$ der Anzahl der Werte $B_i=-1$ entspricht, muss diese Summe, wie oben gezeigt, genau gleich $M/2$ sein. Analog erklären sich die weiteren Randsummen. Bezeichnen wir mit $\Delta=S_{++}-M/4$ die Abweichung der Summe S_{++} von $M/4$, dann muss, wegen der in Tabelle 1 angegebenen Randsummen,

$$S_{++}=S_{--}=\frac{M}{4}+\Delta \qquad \text{und} \qquad S_{-+}=S_{+1}=\frac{M}{4}-\Delta \tag{7}$$

gelten, womit wegen (6)

$$R(L) = \frac{4\Delta}{M} = \frac{4S_{++}}{M} - 1 \tag{8}$$

folgt. Die Wahrscheinlichkeit für "$B_{i+L} = 1$" unter der Bedingung, dass vorher $B_i = 1$ (bzw. $B_i = -1$) galt, ergibt sich nun aus

$$P_{+/+}(L) = \frac{S_{++}}{S_{-+} + S_{++}} = \frac{2S_{++}}{M} = 0,5 + \frac{2\Delta}{M} = 0,5[1 + R(L)] \tag{9}$$

bzw.

$$P_{+/-}(L) = \frac{2S_{-+}}{M} = 0,5 - \frac{2\Delta}{M} = 0,5[1 - R(L)]\,. \tag{10}$$

Analog werden die bedingten Wahrscheinlichkeiten $P_{-/+}(L) = 1 - P_{+/+}(L)$ und $P_{-/-}(L) = 1 - P_{+/-}(L)$ für "$B_{i+L} = -1$" unter der Bedingung $B_i = 1$ bzw. $B_i = -1$ definiert. Diese Wahrscheinlichkeiten sollten im Idealfall gleich 0,5 sein, was bei einem Pseudozufallsgenerator aber nicht unbedingt zutreffen muss, zumal dem hier verwendeten Generatortyp auch einige statische Nachteile bescheinigt werden (Zielinski 1972). Hinzu kommt noch, dass die Dauer bis zum nächsten Abruf einer Zufallszahl von dem letzten Wert B_i abhängt. Im Fall $B_i = +1$ ist dieses Zeitintervall größer als im Fall $B_i = -1$, da bei $B_i = +1$ zweimal umgepolt wird, was zusätzliche Zeit beansprucht. Das Zeitintervall (gemessen als Anzahl der in dieser Zeit erzeugten Zufallszahlen) sei L für den Innenkreis bzw. $L + D$ für den Außenkreis. Wurde nach einem Zufallszahlenaufruf für den Innenkreis entschieden, dann erfolgt der nächste Aufruf nach der Zeit L. Ist $R(L)$ positiv, so bedeutet das, dass bei der nächsten Entscheidung der Innenkreis wieder etwas bevorzugt wird. Wird gerade der Außenkreis durchlaufen und ist $R(L + D)$ negativ, dann wird die nächste Entscheidung bevorzugt nicht wieder für den Außenkreis ausfallen, sondern für den Innenkreis. In jedem Fall wird also bei dem nachfolgenden Durchlauf der Innenkreis etwas bevorzugt, wodurch eine scheinbare akausale Korrelation verursacht werden könnte. Man beachte, dass diese Eigenschaft dann auch nicht durch einen Wechsel der Kodierung ($B_i = -1$ für Außenkreis und $B_i = +1$ für Innenkreis) verändert wird. Wäre durch einen anderen Aufbau der Gleise die Dauer der Durchläufe gleich, so wäre $D = 0$ und $R(L + D) = R(L)$, womit dieser Effekt verschwinden würde.

Es sei nun W_j die Wahrscheinlichkeit, dass im j-ten Schienendurchlauf die Weiche in Richtung Außenkreis gestellt war, d.h. der entsprechende Wert von B war 1. Dann erhält man für den nächsten Durchlauf

$$W_{j+1} = f(W_j) = P_{+/-}(L)(1 - W_j) + P_{+/+}(L + D)W_j \tag{11}$$

Man kann leicht zeigen, dass sich die mit (11) berechnete Folge der W_j auf einen konstanten Wert W einpegelt, der durch die Rekursionsformel (11) nicht mehr verändert wird. Dieser Wert ergibt sich durch Gleichsetzen von W_{j+1} mit W_j in (11) unter Beachtung von (9) und (10) zu

$$W = 0{,}5 + \frac{P_{+/+}(L+D) + P_{+/-}(L) - 1}{2(1 + P_{+/-}(L) - P_{+/-}(L+D)} = 0{,}5 + \frac{R(L+D) - R(L)}{2[2 - R(L+D) - R(L)]}. \qquad (12)$$

Man sieht also, dass im Fall $R(L+D) < R(L)$ die Wahrscheinlichkeit $W = P(B=1)$ kleiner als 0,5 sein wird. Hierbei ist der Zeitunterschied D besonders wichtig, da mit $D=0$ nach (11) und (10) $P_{+/+}(L+D) + P_{+/-}(L) - 1 = P_{+/+}(L) + P_{+/-}(L) - 1 = 0$ gelten würde und damit gemäß (12) $W = 0{,}5$.

Ob der hier geäußerte Verdacht bestätigt werden kann, hängt davon ab, ob die Autokorrelation $R(L)$ sehr stark von dem Wert L abhängt. Dieses soll im folgenden Abschnitt untersucht werden.

Unregelmäßigkeiten bezüglich der Autokorrelation $R(L)$

Für die nach (8) berechnete Autokorrelation $R(L)$ erhalten wir folgende Eigenschaften. Wegen der Periodizität gilt $B_{i+M} = B_i$. Damit folgt

$$R(M) = 1 \qquad (13)$$

und

$$R(L+M) = R(L) \qquad (14)$$

Weiterhin folgt aus der Symmetrieeigenschaft $\rho(B_i, B_{i+L}) = \rho(B_{i+L}, B_i)$ von Korrelationen die Beziehung

$$R(-L) = R(L) \qquad (15)$$

und mit (14) folgt

$$R(L) = R(M-L). \qquad (16)$$

Eine weitere erstaunliche Besonderheit des Generatortyps (3) wurde im Anhang 2 bewiesen, nämlich

$$R(\tfrac{M}{2}) = -1. \qquad (17)$$

Damit ist also stets $B_i + B_{i+\frac{M}{2}} = 0$ und

$$\sum_i B_i B_{i+L} = \sum_i B_i (B_{i+l} - B_{i+L} - B_{i+L+\frac{M}{2}}) = -\sum_i B_i B_{i+L+\frac{M}{2}} = -R(L + \tfrac{M}{2}).$$

Mit (15) und (14) folgt

$$R(L) = -R(\tfrac{M}{2} - L) \qquad (18)$$

Für $L = \frac{M}{4}$ gilt damit

$$R(\tfrac{M}{4}) = 0. \qquad (19)$$

Für das weitere Studium der Autokorrelationen genügt es also, diese nur für $0 < L < \frac{M}{4}$ zu berechnen, da man mit (18) hiermit auch die $R(L)$ für $\frac{M}{4} < L < \frac{M}{2}$ und anschließend mit (16) alle $R(L)$ für $\frac{M}{2} < L < M$ erhält. Der Bereich $0 < L < \frac{M}{4}$ enthält aber immerhin noch 4194303 L-Werte, so dass es kaum sinnvoll erscheint, für alle L $R(L)$ zu berechnen. Außerdem ist die Umlaufdauer der Lokomotive nicht konstant, sondern hat eine gewisse

Schwankungsbreite. Wir zerlegen deswegen den Bereich $0 < L < \frac{M}{4}$ in mehrere Intervalle und berechnen für jedes Intervall den Mittelwert der entsprechenden $R(L)$. Da die Anzahl $\frac{M}{4} - 1 = 4194303$ in die Primfaktoren $3 \cdot 23 \cdot 89 \cdot 683$ zerlegt werden kann, bietet es sich an, diesen Bereich in 683 Intervalle zu je $3 \cdot 23 \cdot 89 = 6141$ L-Werte einzuteilen. Zur Berechnung der mittleren Autokorrelation je Intervall kann die in Anhang 3 beschriebene Rekursionsformel vorteilhaft genutzt werden, womit es also nicht erforderlich ist, jeweils alle 6141 $R(L)$-Werte zu berechnen. Die dabei erhaltenen Intervalle mit extremen mittleren Autokorrelationen wurden in Tabelle 2 aufgelistet. Anschließend wurden von diesen beiden Intervallen die einzelnen $R(L)$ berechnet. Die extremen $R(L)$ und die entsprechenden L-Werte wurden ebenfalls in Tabelle 2 aufgeführt.

Tabelle 2: Intervalle mit der minimalen bzw. maximalen mittleren Autokorrelation.

Intervall-Nr.	L von - bis	mittlere Autokorrelation	Extremalwert von $R(L)$	L-Wert mit extremen $R(L)$
398	2437978 - 2444118	- 0,0000201	-0,16312	2438913
286	1750186 - 1756326	0,0000342	0,17021	1755391

Wäre die Umlaufdauer des Innenkreises $L = 1755391$ und die des Außenkreises $L + D = 2438913$, so würde man also eine scheinbare akausale Korrelation beobachten können. Wir müssen jedoch berücksichtigen, dass die Umlaufzeiten gewissen Schwankungen unterliegen, womit die mittleren Autokorrelationen in den Schwankungsbereichen um die oben genannten L-Werte nicht mehr so extrem ausfallen würden, da die Umlaufzeit nicht in jedem Falle exakt den oben genannten Werten entspricht. Außerdem werden die konkreten mittleren Umlaufzeiten in dem Modelleisenbahn-Experiment auch nicht unbedingt mit den obigen Werten identisch sein. Ob der Pseudozufallszahlengenerator auch für die konkreten Zeitwerte dieses Experimentes eine akausale Korrelation vortäuscht, soll im nächsten Abschnitt untersucht werden.

Das Modelleisenbahn-Experiment im Zeitraffer

Wie im Anhang 4 gezeigt wurde, ist es möglich, zu einer Zufallszahl c_i unmittelbar die Zahl c_{i+L} zu bestimmen, ohne alle dazwischen liegenden Werte c_j berechnen zu müssen. Mit dieser Technik kann das Modelleisenbahn-Experiment, zumindest was die Zufallszahlenfolgen betrifft, quasi im Zeitraffer durchgeführt werden. Ausgehend von der Abbildung auf S. 67 bei Hagel und Tschapke (2002) nehmen wir die Durchlaufdauer des Innenkreises als normalverteilt an mit dem Erwartungswert μ_I und der Standardabweichung $\mu_I \cdot v\,\%/100$, wobei $v\,\%$ den Variationskoeffizienten ($= 100 \cdot \sigma / \mu$) bezeichnet. Aus der besagten Abbildung erhält man – nach Augenmaß – für $v\,\%$ ungefähr den Wert 3 %. Der Erwartungswert μ_I wurde von Hagel und Tschapke (S. 13) für das Erstexperiment mit 60000 angegeben. Für die mittlere Dauer μ_A im Außenkreis ist etwa der doppelte Wert $(\mu_A = 2\mu_I)$ anzusetzen (Hagel, persönliche Mitteilung). Es ist sinnvoll, für den Außenkreis den gleichen Variationskoeffizienten $v\,\%$ und damit die doppelte Standardabweichung anzunehmen.

Wie bei Hagel und Tschapke (2002) wurden in einem Experiment jeweils 10000 Kreisdurchläufe simuliert. Um genauere Schätzwerte für die Verteilungsparameter zu erhalten, wurden jedoch nicht nur 10, sondern jeweils 1000 Experimente durchgeführt. Die Simulation der jeweiligen Umlaufzeiten erfolgte gemäß der Normalverteilung (Programm RANNOR aus dem SAS-Programmsystem) mit dem Erwartungswert μ und der Standardabweichung $\mu \cdot v\%/100$. μ wurde jeweils in Abhängigkeit von dem im letzten Umlauf erhaltenen Wert B gewählt, und zwar $\mu = \mu_A$ im Fall $B = 1$ und $\mu = \mu_I$ im Fall $B = -1$. Die Berechnung der Zufallswerte B erfolgte mit dem QBASIC-Generator (2).

In Tabelle 3 werden die Ergebnisse von jeweils zwei Varianten gezeigt. Bei Variante 1 wurden vor jedem der 1000 Experimente die beiden Zufallsgeneratoren neu initiiert, bei Variante 2 jedoch nur vor dem ersten Experiment.

Wegen $V(B) = 1$ erhält man für das Mittel $\overline{B}_{ex}$ aus 10000 Umläufen eines Experimentes die Standardabweichung $\sqrt{V(\overline{B}_{ex})} = 0{,}01$. Die aus den jeweils 1000 Experimenten ermittelten Schätzwerte dieser Standardabweichung lagen zwischen 0,0096 und 0,01167, was durchaus als normal angesehen werden kann. Tabelle 3 enthält das Gesamtmittel $\overline{B}$ aller 1000 Experimente und den entsprechenden *z*-Wert $z = \overline{B}/\sqrt{V(\overline{B})}$, wobei die theoretische Standardabweichung $\sqrt{V(\overline{B})} = 0{,}000316$ des Gesamtmittels $\overline{B}$ verwendet wurde.

Für die Parameter aus dem Eisenbahnexperiment, also $\mu_A = 120000$, $\mu_I = 60000$ und $v\% = 3$, zeigt die Tabelle 3 keine Auffälligkeiten. Verwenden wir jedoch $\mu_A = 1{,}5 \cdot M$ $(R(1{,}5M) = -1)$ und $\mu_I = M$ $(R(M) = 1)$, so kann man bei extrem kleinen Variationskoeffizienten $v\%$ eine deutlich signifikante Bevorzugung des Innenkreises feststellen. Dass mit kleineren Variationskoeffizienten $v\%$ eher Signifikanz auftritt, ist verständlich, da dann die Umlaufzeiten öfter mit ihren Erwartungswerten μ_I bzw. μ_A identisch sind, womit dann das Gesamtergebnis stärker durch die extremen Autokorrelationswerte $R(\mu_A) = -1$ und $R(\mu_I) = 1$ bestimmt wird.

Weiterhin wird in Tabelle 3 auch der Fall $\mu_I = 1755391$ und $\mu_A = 2438913$ demonstriert, da man für diese Werte gemäß Tabelle 2 auch sehr extreme Autokorrelationen erhält. Obwohl hier die Autokorrelationen aber nicht so stark von Null abweichen, wie im vorhergehenden Fall, sind bei kleinen $v\%$ die *z*-Werte bedeutend extremer. Dieses resultiert aus der Tatsache, dass bei gleicher relativer Standardabweichung $v\%$ die absolute Standardabweichung wegen der kleineren μ-Werte wesentlich kleiner ist, womit hier die Umlaufzeiten bedeutend häufiger mit den Erwartungswerten $\mu_I = 1755391$ bzw. $\mu_A = 2438913$ übereinstimmen und damit die entsprechenden extremen Autokorrelationen aus Tabelle 2 sich stärker durchsetzen.

Tabelle 3: Ergebnisse des Simulationsexperiments (Erläuterung: siehe Haupttext).

$\sigma\%$	Lauf-variante	$\mu_I = 60000$ $\mu_A = 120000$		$\mu_I = 1755391$ $\mu_A = 2438913$		$\mu_I = M = 16777216$ $\mu_A = 1{,}5M = 25165824$	
		$\overline{B}$	z	$\overline{B}$	z	$\overline{B}$	z
3.0000	1	-0.00002	-0.066	-0.00066	-2.100	-0.00014	-0.457
	2	-0.00040	-1.254	-0.00008	-0.254	-0.00001	-0.020
0.1000	1	-0.00013	-0.424	0.00035	1.099	0.00017	0.530
	2	0.00032	1.022	0.00035	1.107	-0.00021	-0.672
0.0100	1	-0.00005	-0.169	-0.00016	-0.496	-0.00017	-0.545
	2	-0.00026	-0.819	-0.00015	-0.481	-0.00075	-2.372
0.0010	1	0.00006	0.182	-0.00394	-12.446	-0.00148	-4.694
	2	0.00008	0.252	-0.00353	-11.146	-0.00153	-4.851
0.0001	1	0.00031	0.978	-0.03444	-108.905	-0.02129	-67.338
	2	0.00008	0.252	-0.03435	-108.607	-0.02196	-69.442

Fazit

Tabelle 3 zeigt, dass es prinzipiell durchaus möglich ist, dass auf Grund spezifischer Eigenschaften der Autokorrelationen des verwendeten Zufallszahlengenerators akausale Korrelationen vorgetäuscht werden. Dieses ist jedoch nur für ganz spezielle mittlere Umlaufzeiten der Fall, wobei die Varianz der Umlaufzeiten auch noch sehr klein sein muss. Für die Umlaufzeiten des Eisenbahn-Experiments von Hagel und Tschapke (2002) waren diese Bedingungen jedoch nicht erfüllt, womit also durch die Spezifika des QBASIC-Zufallsgenerators der von Hagel und Tschapke (2002) beobachtete Effekt nicht erklärt werden kann.

Die hier durchgeführte Untersuchung ging von dem Verdacht aus, dass spezielle Besonderheiten der Autokorrelation der Zufallszahlen zur scheinbaren akausalen Korrelation führen könnten. Mit einiger Fantasie könnten anderen Forschern weitere Verdachtsmomente einfallen, die als Ausgangspunkt für weitere Untersuchungen dienen könnten, zu denen hiermit ausdrücklich ermuntert sei. Einige Hilfsmittel dafür wurden mit diesem Artikel bereitgestellt. Auch der QBASIC-Befehl RANDOMIZE, der zur Initiierung der Zufallszahlen dient, wurde auf Bitte von Johannes Hagel vom Autor rekonstruiert. Das Programm kann auf Wunsch zur Verfügung gestellt werden.

Noch wichtiger sind aber Wiederholungen des Originalexperiments, wobei auch andere technische Systeme verwendet werden sollten, die in vertretbarer Zeit eine bedeutend größere Umlaufanzahl zulassen. Weiterhin wäre auch ein schnellerer PC zu empfehlen, mit dem eine zeitlich dichtere Zufallszahlenfolge erzeugt werden kann, wodurch der Lokomotive (bzw. dem technischen Analogon) eine differenziertere „Entscheidung“ zwischen benachbarten Zufallszahlen ermöglicht wird.

Literatur

Hagel, J.; Tschapke, M. (2002): Zum experimentellen Nachweis akausaler Korrelationseffekte in unbelebten Systemen. *Zeitschrift für Anomalistik* 2, 6-31.

Hull, T. E.; Dobell, A. R. (1964): Mixed congruential random number generators for binary machines. *J. Assoc. Comput. Mach.* 11, 31-40.

Knuth, D. E. (1997): The Art of Computer Programming, Vol. II. Addison-Wesley Publishing Company.

Siebenborn, T. (2002): Zwei grundlegende Probleme. *Zeitschrift für Anomalistik* 2, 61-62.

Zielinski, R. (1972): Erzeugung von Zufallszahlen. Fachbuchverlag Leipzig.

Anhang

Anhang 1: Bestimmung des Pseudozufallsgenerators

Nachdem festgestellt wurde, dass erst nach $M = 2^{24}$ Zufallszahlen die erste Zahl sich wiederholt, und damit die Zykluslänge M festlag, wurde angenommen, dass ein gemischter Generator der Form $c_{i+1} = \text{mod}\,(a \cdot c_i + b, M)$ vorliegt. Mit drei aufeinander folgenden Zahlen z_1, z_2 und z_3 des QBASIC-Generators wurden zunächst die Werte $c_i = z_1 \cdot M \quad (i = 1,2,3)$ berechnet und ganzzahlig gerundet. Anschließend wurde für alle $a = 0,1,\ldots,M-1$ folgendes berechnet:

$b = c_2 - a \cdot c_1$.

Falls $b < 0$, dann $b \Leftarrow b + M$.

$c = \text{mod}\,(a \cdot c_2 + b, M)$.

Ist $c = c_3$, so ist das Zahlenpaar (a,b) ein Kandidat des gesuchten Generators. Nach Untersuchung aller möglichen a fand sich nur der in (2) genannte Kandidat.
Zur Programmierung dieses Generators war ein Programm erforderlich, welches die Multiplikation von Zahlen a und c der Größenordnung M zulässt, d.h. es müssen Zahlendarstellungen mit mindestens $2 \cdot 24$ bit möglich sein. Diese Bedingung ist im Programmsystem SAS erfüllt (60 bit), weswegen dieses System für alle weiteren Rechnungen verwendet wurde.
Zur Kontrolle wurde nun ein voller Zyklus der mit dem QBASIC-Generator erzeugten Zufallszahlen in SAS übernommen. Diese Folge konnte komplett mit dem gefundenen Generator (1) rekonstruiert werden.

Anhang 2: $R(M/2) = -1$

Im Folgenden verwenden wir oft anstelle der Gleichheitsbeziehung $x = y$ die Äquivalenzbeziehung $x \equiv y$, was hier bedeuten soll, dass x und y nach ganzzahliger Division durch M den gleichen Rest liefern. Man sieht leicht, dass mit (1) auch

$$c_{n+L} \equiv a^L \cdot c_n + b \cdot \sum_{j=0}^{L-1} a^j \tag{A1}$$

gilt (Hull und Dobell 1964). Mit der Summenformel der geometrischen Reihe folgt

$$c_{n+L} \equiv a^L c_n + b\frac{a^L - 1}{a-1} = \frac{a^L - 1}{a-1}(c_n(a-1)+b) + c_n\,. \tag{A2}$$

Wegen (3) ist a in der Form

$$a = 4 \cdot k + 1$$

darstellbar, wobei k ganzzahlig und positiv ist. Setzen wir $T(L) = (a^L - 1)/(a-1)$, dann gilt:

$$T(L) = \frac{a^L - 1}{a-1} = \frac{(4k+1)^L - 1}{4k} = \frac{\left[\sum_{i=0}^{L}(4k)^i \binom{L}{i}\right] - 1}{4k} = \sum_{i=1}^{L}(4k)^{i-1}\binom{L}{i}. \tag{A3}$$

Insbesondere für $L = M$ und $c_n = 0$ folgt wegen der Periodizität $c_{n+M} = c_n = 0$ und wegen (A2)

$$0 = c_{n+M} - c_n \equiv \frac{a^M - 1}{a-1}b = \frac{(a^{M/2}+1)(a^{M/2}-1)}{(a-1)}b =$$

$$\left[\frac{4k}{a-1}(a^{M/2}-1)+2\right]T(\frac{M}{2})b = 2\left[2kT(\frac{M}{2})+1\right]\cdot T(\frac{M}{2}) \equiv 0\,.$$

Das heißt, dass der letzte Ausdruck durch $M = 2^{24}$ teilbar sein muss. Da der Wert in den eckigen Klammern nicht durch 2 teilbar ist, muss $T(\frac{M}{2})$ mindestens durch 2^{23} teilbar sein. Wäre $T(\frac{M}{2})$ durch $M = 2^{24}$ teilbar, so wäre wegen (A2) $c_{n+M/2} = c_n$. Dann wäre der Zyklus aber nicht erst nach M, sondern nach $M/2$ Zufallszahlen vollendet, was nicht möglich ist. Daher ist $T(\frac{M}{2})$ nur durch $\frac{M}{2} = 2^{23}$ teilbar bzw. in der Form $T(\frac{M}{2}) = \frac{M}{2} \cdot u$ mit einer ungeraden Zahl u darstellbar. Da u und auch b ungerade sind, gibt es ganze Zahlen q und r mit $u = 2q+1$ und $b = 2r+1$. Damit folgt aus (A2)

$$c_{n+M/2} \equiv \frac{M}{2}(2q+1)\left[c_n \cdot 4k + 2r + 1\right] + c_n$$

$$= M(2q+1)\left[c_n 2k + r\right] + Mq + \frac{M}{2} + c_n \equiv \frac{M}{2} + c_n\,.$$

Für $c_n < \frac{M}{2}$ folgt also $c_{n+M/2} = c_n + \frac{M}{2} \geq \frac{M}{2}$. Für $c_n \geq \frac{M}{2}$ folgt $c_{n+M/2} \equiv c_n + \frac{M}{2}$ $\equiv c_n + \frac{M}{2} - M \;\; = c_n - \frac{M}{2} < \frac{M}{2}$. Für das gemäß (4) definierte B gilt also $B_{n+M/2} = -B_n$, d.h. $\mathrm{R}(\frac{m}{2}) = -1$.

Anhang 3: Berechnung der mittleren Autokorrelation

Zur Ermittlung der mittleren Autokorrelation $\overline{\mathrm{R}} = \frac{1}{A-E+1}\sum_{L=A}^{E}\mathrm{R}(L)$ über alle $L = A,...,E$ ist die Summe $S = \sum_{L=A}^{E}\mathrm{R}(L) = \sum_{L=A}^{E}\sum_{i=1}^{M} B_i \cdot B_{i+L} = \sum_{i=1}^{M} B_i C_i$ mit $C_i = \sum_{L=A}^{E} B_{i+L}$ zu berechnen. Für $i=1$ berechne man die Werte B_1 bis B_{1+E} und die Summe C_1 der Werte von B_{1+A} bis B_{1+E} und damit den Summand $B_1 \cdot C_1$ der Summe S. Den nächsten Summanden $B_2 \cdot C_2$ erhält man durch Berechnung der Nachfolger von B_1, B_{1+A} und B_{1+E}, also B_2, B_{2+A} bzw. B_{2+E}. Hiermit erhält man $C_2 = C_1 + B_{2+E} - B_{1+A}$. Analog erhält man alle weiteren C_i aus $C_i = C_{i-1} + B_{i+E} - B_{i-1+A}$. Mit dieser Rekursionsformel ist es also möglich, $\overline{\mathrm{R}}$ mit nur drei Folgen von Zufallszahlen zu berechnen, nämlich mit B_i, B_{i+A} und B_{i+E} für $i = 1,...,M$.

Anhang 4: Berechnung von c_{n+L} aus c_n

Gemäß (A2) erhält man c_{n+L} aus

$$c_{n+L} \equiv T(L)\cdot(c_n(a-1)+b)+c_n \qquad \text{(A4)}$$

mit $T(L)$ aus (A3). Da hier von $T(L)$ nur der Rest $\bmod\,(T(L),M)$ nach Division durch M interessiert, genügt es die Summe in (A3) nicht bis $i = L$, sondern nur bis $i = 12$ zu berechnen, da alle weiteren Summanden durch $4^{12} = 2^{24} = M$ teilbar sind (Hull und Dobell 1964).

Sind S_i $(i = 1,...,12)$ diese 12 Summanden, so gilt $S_1 = L$ und $S_i = S_{i-1} \cdot \frac{L-i+1}{i} \cdot 4k$. Um auch von den S_i den nur interessierenden Wert $\bmod\,(S_i, M)$ berechnen zu können, ist von $4S_{i-1}\cdot(L-i+1)$ jeweils ein Faktor abzuspalten, der durch i teilbar ist. Wir verwenden deswegen für die Werte $L-i+1$ die Faktordarstellung $L-i+1 = f_i \cdot p_i$. Zur Berechnung der p_i werden zunächst alle $p_i = 1$ gesetzt und anschließend folgendermaßen verändert:

Bestimme $j_3 = \bmod(L,3)$, $j_5 = \bmod(L,5)$, $j_7 = \bmod(L,7)$, $j_9 = \bmod(L,9)$ und $j_{11} = \bmod(L,11)$ und multipliziere p_{j_3+1}, p_{j_3+4}, p_{j_3+7} und p_{j_3+10} mit 3, p_{j_5+1} und p_{j_5+6} mit 5, p_{j_7+1} mit 7, p_{j_9+1} mit 3 und $p_{j_{11}+1}$ mit 11. Anschließend sind die f_i aus $f_i = (L-i+1)/p_i \quad (i=1,...,12)$ zu berechnen. Nun ist $F_1 = f_1$ und $P_1 = p_1$ zu setzen und anschließend für $i=2,...,12$ Folgendes zu berechnen:

$$F_i = \bmod\left[\mathrm{F}_{\mathrm{i}\text{-}1} \cdot 2k \cdot f_i, M\right], \qquad P_i = P_{i-1} \cdot 2p_i / i$$

Wegen der speziellen Berechnung der p_i ist P_i stets ganzzahlig und kleiner als M. Die Summanden S_i erhält man nun aus $S_i = \bmod\left[F_i \cdot P_i, M\right]$. Die entsprechende Summe wird in (A4) anstelle von $T(L)$ verwendet. Bei den weiteren Multiplikationen in (A4) sind von dem Produkt stets weiterhin nur der Rest $\bmod\left[\text{Produkt}, M\right]$ zu verwenden, um im Bereich der darstellbaren Zahlen zu bleiben.

Korrespondenzanschrift:

PD Dr. Volker Guiard
Zum Laakkanal 14, D-18109 Rostock
E-Mail: guiard@anomalistik.de

Kommentare zu Guiard: Kann ein Pseudozufallsgenerator eine akausale Korrelation vortäuschen?

BERND DÜRRER

Weitere Vorschläge zur Verbesserung des Versuchsaufbaus

Einige im Anschluss an den Artikel von Hagel und Tschapke (2002) in der *Zeitschrift für Anomalistik* veröffentlichte Kommentare kritisierten, dass ein Pseudozufallszahlengenerator verwendet wurde. Es wurde die Vermutung geäußert, dass die beobachtete akausale Korrelation nur auf unzureichende statistische Eigenschaften dieses Generators zurückzuführen sei. Volker Guiard beschreibt in seinem vorliegenden Artikel zunächst den Algorithmus dieses Generators und diskutiert anschließend die Eigenschaften der Autokorrelationsfunktion des Generators. Er kommt aufgrund dieser Betrachtungen zu dem Schluss, dass die von Hagel und Tschapke beobachtete akausale Korrelation kein durch den Generator bedingtes Artefakt sein könne.