Erzeugen und Testen von Zufallszahlen

Jürgen Zumdick

1. Einleitung

Eine Lerngruppe wird aufgefordert n = 100 Zufallszahlen (0 oder 1) nach folgenden Methoden zu erzeugen: Die Hälfte der Gruppe benutzt

a) eine Münze
   oder
b) die Zufallszahlenfunktion eines Taschenrechners
   oder
c) die Zufallszahlenfunktion einer Software
   oder
d) einen der im 6. Abschnitt angegebenen Zufallsgeneratoren.

Die andere Hälfte versucht sich selbst als Zufallsgenerator und notiert nach eigenem Gutdünken eine Zufallszahlenfolge.
Ziel des Unterrichtsvorhabens ist es, Testverfahren zu entwickeln, mit deren Hilfe die erdachten Zufallszahlen herausgefunden werden können. Auch sollen die Testverfahren helfen, die Qualität von Zufallsgeneratoren zu beurteilen.

Beispiel für eine derartige Folge:
00100 10001 10100 01101 01100 01010 10101 10001 10110 00011 01011 01010 11100 11010 10110 11001 10111 00001 11001 01011.

2. Der Test (1. Teil – Test auf Häufigkeit)

Da die Wahrscheinlichkeit von 0 bzw. 1 p = 0,5 beträgt, sollte etwa 50-mal die Null auftreten.
Die folgende Tabelle gibt einen Überblick, über die Wahrscheinlichkeiten der Abweichungen vom Erwartungswert:

	Wahrscheinlichkeit, dass die Anzahl der Nullen/Einsen außerhalb des links stehenden Intervalls liegt
[40; 60] $p(X \le 39 \vee X \ge 61) = 2 \cdot \sum_{k=0}^{39} \binom{100}{k} \cdot \left(\frac{1}{2}\right)^{100} = 2 \cdot 0{,}018 = 0{,}036$	0,036
[42; 58]	0,088
[44; 56]	0,194
[46; 54]	0,368
[47; 53]	0,484
[48; 52]	0,618

Die obige Zahlenfolge enthält 46 Nullen und folglich 54 Einsen. Eine derartige Abweichung vom Erwartungswert tritt sehr häufig auf – in mehr als 48% aller Fälle. Daher ist die obige Ziffernfolge nicht verdächtig.

Dieser Test ist natürlich nicht ausreichend, da er keinerlei Wechsel zwischen 0 und 1 berücksichtigt. Dies geschieht dann mit dem folgenden Test.

3. Der Test (2. Teil – Test auf Wechsel)

Die obige Folge enthält W = 61 Wechsel. Um die Frage zu beantworten, ob diese Anzahl im Rahmen der Erwartung liegt, kann man zunächst induktiv vorgehen.

i (n = 3)	0	1	2
Ausfälle	000, 111	001, 011, 110, 100	010, 101
p(W= i)	$\frac{2}{2^3}$	$\frac{4}{2^3}$	$\frac{2}{2^3}$

E(W) = 1, $v(W) = E(W^2) - E^2(W) = 0 \cdot \frac{2}{2^3} + 1^2 \cdot \frac{4}{2^3} + 2^2 \cdot \frac{2}{2^3} - 1^2 = 1{,}5 - 1 = 0{,}5$

i (n = 4)	0	1	2	3
p(W = i)	$\frac{2}{2^4}$	$\frac{6}{2^4}$	$\frac{6}{2^4}$	$\frac{2}{2^4}$

E(W) = 1,5 V(W) = E(W²) – E²(W) = 3 – 2,25 = 0,75

i (n = 5)	0	1	2	3	4
p(W = i)	$\frac{2}{2^5}$	$\frac{8}{2^5}$	$\frac{12}{2^5}$	$\frac{8}{2^5}$	$\frac{2}{2^5}$

E(W) = 2 V(W) = E(W²) – E²(W) = 5 – 4 = 1

Vermutung: $E(W) = (n-1) \cdot \frac{1}{2}$ und $V(W) = (n-1) \cdot \frac{1}{4}$

Dies entspräche dem Erwartungswert und der Varianz einer binomialverteilten Zufallsgröße $B\left(n-1; \frac{1}{2}\right)$. Kürzt man in den obigen Tabellen die Wahrscheinlichkeiten mit 2, so erhält man in der Tat die entsprechenden Verteilungen. Dass es sich um eine Binomialverteilung handelt, lässt sich auch wie folgt begründen: Ab dem 2. Wurf treten die Ausfälle Wechsel/kein Wechsel jeweils mit der Wahrscheinlichkeit 0,5 ein. Bei 100 Würfen werden also $(100-1) \cdot \frac{1}{2} = 49{,}5$ Wechsel erwartet.

Die folgende Tabelle gibt einen Überblick, über die Wahrscheinlichkeiten der Abweichungen vom Erwartungswert:

	Wahrscheinlichkeit, dass die Anzahl der Wechsel außerhalb des links stehenden Intervalls liegt
[40; 60]	0,044

$p(X \le 39 \vee X \ge 61) = 2 \cdot \sum_{k=0}^{39} \binom{99}{k} \cdot \left(\frac{1}{2}\right)^{99} = 2 \cdot 0{,}022 = 0{,}044$	
[42; 58]	0,108
[44; 56]	0,228
[46; 54]	0,422
[48; 52]	0,688

Die beobachtete Abweichung tritt in weniger als 5% aller Fälle auf. Daher ist die Zahlenfolge verdächtig.
Aber auch diese beiden Tests sind zusammen noch nicht scharf genug. Deshalb sollte zumindest noch der folgende Test durchgeführt werden.

4. Der Test (3. Teil – Test auf k-Blöcke)

Die Zufallszahlenfolgen wird in Blöcke der Länge k eingeteilt (hier k = 4).

0010 0100 0110 1000 1101 0110 0010 1010 1011 0001 1011 0000 1101 0110 1010 1110 0110 1010 1101 1001 1011 1000 0111 0010 1011

X sei die Anzahl der Blöcke mit vier gleichen Ziffern
(X = 1 im Beispiel)
Y sei die Anzahl der Blöcke mit drei gleichen Ziffern
(Y = 15 im Beispiel)
Z sei die Anzahl der Blöcke mit zwei gleichen Ziffern
(Z = 8 im Beispiel)

Die Wahrscheinlichkeit für einen Block mit vier gleichen Ziffern beträgt $\frac{2}{2^4} = \frac{1}{8}$.

Da es 25 Blöcke gibt, ist X nach B(25; $\frac{1}{8}$) verteilt. Folglich gilt E(X) = $25 \cdot \frac{1}{8} = 3{,}125$.

	Wahrscheinlichkeit, dass die Anzahl der Blöcke mit vier gleichen Ziffern außerhalb des links stehenden Intervalls liegt
[0; 6]	0,03
[0; 5]	0,083
[1 ;5]	0,119
[1; 4]	0,231
[2; 4]	0,358
[3; 4]	0,575

Die beobachtete Abweichung tritt in etwa 36% aller Fälle auf.

Die Wahrscheinlichkeit für einen Block mit drei gleichen Ziffern beträgt $\frac{8}{2^4} = \frac{1}{2}$.

Y ist folglich nach B(25; $\frac{1}{2}$) verteilt mit dem Erwartungswert E(Y) = 12,5.

	Wahrscheinlichkeit, dass die Anzahl der Blöcke mit drei gleichen Ziffern außerhalb des links stehenden Intervalls liegt
[7; 18]	0,017

[8; 17]	0,043
[9; 16]	0,108
[10;15]	0,23
[11; 14]	0,424

Die beobachtete Abweichung tritt in etwa 42% aller Fälle auf.

Die Wahrscheinlichkeit für einen Block mit zwei gleichen Ziffern beträgt $\frac{6}{2^4} = \frac{3}{8}$.

Z ist folglich nach B(25; $\frac{3}{8}$) verteilt mit dem Erwartungswert E(Y) = 9,375.

	Wahrscheinlichkeit, dass die Anzahl der Blöcke mit zwei gleichen Ziffern außerhalb des links stehenden Intervalls liegt
[5; 14]	0,036
[5; 13]	0,064
[6; 13]	0,097
[6; 12]	0,15
[7; 12]	0,215
[7; 11]	0,305
[8; 11]	0,411
[8; 10]	0,538
[9; 10]	0,682

Die beobachtete Abweichung tritt in etwa 68% aller Fälle auf.
Insgesamt liefert dieser Test somit keinen Hinweis auf eine verdächtige Zahlenfolge.

5. Verallgemeinerung der Tests

Die Tests auf Häufigkeit und auf Blöcke lassen sich auf beliebige Wahrscheinlichkeiten verallgemeinern, da jeweils Binomialverteilungen vorliegen.

Beim Test auf Wechsel liegt jedoch für $p \neq \frac{1}{2}$ keine Binomialverteilung vor, da dann ein Wechsel nicht immer mit derselben Wahrscheinlichkeit auftritt. In diesem Fall ist wie folgt zu verfahren:

Es sei X_i der Ausfall im i-ten Wurf, $Y_i = \begin{cases} 1 & falls\ X_{i-1} \neq X_i \\ 0 & falls\ X_{i-1} = X_i \end{cases}$ $i = 2,...,n$ und p(0) = 1.

Dann gilt für die Anzahl der Wechsel $W = \sum_{i=2}^{n} Y_i$.

$$E(W) = \sum_{i=2}^{n} E(Y_i)$$

$$E(Y_i) = 1 \cdot p(Y_i = 1) + 0 \cdot p(Y_i = 0) = p(Y_i = 1) = p(\{01,10\}) = pq + qp = 2pq.$$

Also:
E(W) = 2·(n-1)·p·q.

$$V(W) = V\left(\sum_{i=2}^{n} Y_i\right) = E\left(\left(\sum_{i=2}^{n} Y_i\right)^2\right) - E^2\left(\sum_{i=2}^{n} Y_i\right) = E\left(\left(\sum_{i=2}^{n} Y_i\right)^2\right) - (2 \cdot (n-1) \cdot p \cdot q)^2$$

$$E\left(\left(\sum_{i=2}^{n} Y_i\right)^2\right) = E\left(Y_2{}^2 + Y_3{}^2 + \ldots + Y_n{}^2 + 2\cdot Y_2\cdot Y_3 + 2\cdot Y_2\cdot Y_4 + \ldots + 2\cdot Y_2\cdot Y_n + \ldots 2\cdot Y_{n-1}\cdot Y_n\right)$$

$$= \sum_{i=2}^{n} E\left(Y_i{}^2\right) + 2\cdot\sum_{i=2}^{n-1} Y_i\cdot Y_{i+1} + 2\cdot\left(E\left(Y_2\cdot Y_4\right) + E\left(Y_2\cdot Y_5\right) + .. + E\left(Y_{n-2}\cdot Y_n\right)\right)$$

In dieser Summe treten 3 Typen von Zufallsgrößen auf: Y_i^2 (Typ 1), $Y_i\cdot Y_{i+1}$ (Typ 2) und $Y_i\cdot Y_{i+k}$ mit k ≥ 2 (Typ 3).

Berechnung des Erwartungswertes für Typ 1:

$$E(Y_i^2) = 1^2\cdot p(Y_i = 1^2) + 0^2\cdot p(Y_i = 0^2) = p(Y_i = 1^2) = p(Y_i = 1) = p(\{01,10\}) = pq + qp = 2pq$$

Berechnung des Erwartungswertes für Typ 2:

Dazu wird folgende Tabelle betrachtet:

(i-1)-ter Wurf	i-ter Wurf	(i+1)-ter Wurf	$Y_i\cdot Y_{i+1}$	Wahrscheinlichkeit
0	0	0	0·0 = 0	p^3
0	0	1	0·1 = 0	p^2q
0	1	0	1·1 = 1	p^2q
0	1	1	1·0 = 0	pq^2
1	0	0	1·0 = 0	p^2q
1	0	1	1·1 = 1	pq^2
1	1	0	0·1 = 0	pq^2
1	1	1	0·0 = 0	q^3

$$E(Y_i\cdot Y_{i+1}) = p^2\cdot q + pq^2 = p\cdot q\cdot(p+q) = p\cdot q$$

Hinweis:

Da $E(Y_i\cdot Y_{i+1}) \neq E(Y_i)\cdot E(Y_{i+1})$, sind Y_i und Y_{i+1} nicht unabhängig voneinander.

Berechnung des Erwartungswertes für Typ 3:

Da Y_i und Y_{i+k} mit k ≥ 2 unabhängig voneinander sind, gilt

$E(Y_i\cdot Y_{i+k}) = E(Y_i)\cdot E(Y_{i+k}) = 2pq\cdot 2pq = 4p^2q^2$.

Alternative Herleitung:

(i-1)-ter Wurf	i-ter Wurf		(i+k-1)-ter Wurf	(i+k)-ter Wurf	$Y_i\cdot Y_{i+k}$	Wahrscheinlichkeit
0	0		0	0	0·0 = 0	p^4
0	0		0	1	0·1 = 0	p^3q
0	0		1	0	0·1 = 0	p^3q
0	0		1	1	0·0 = 0	p^2q^2
0	1		0	0	1·0 = 0	p^3q
0	1		0	1	1·1 = 1	p^2q^2
0	1		1	0	1·1 = 1	p^2q^2
0	1		1	1	1·0 = 0	pq^3
1	0		0	0	1·0 = 0	p^3q
1	0		0	1	1·1 = 1	p^2q^2
1	0		1	0	1·1 = 1	p^2q^2

1	0		1	1	1·0 = 0	pq^3
1	1		0	0	0·0 = 0	p^2q^2
1	1		0	1	0·1 = 0	pq^3
1	1		1	0	0·1 = 0	pq^3
1	1		1	1	0·0 = 0	q^4

$E(Y_i \cdot Y_{i+k}) = p^2q^2 + p^2q^2 + p^2q^2 + p^2q^2 = 4p^2q^2$.

Der Typ 1 tritt (n-1)-mal auf, der Typ 2 (n-2)-mal . Die folgende Tabelle zeigt, wie oft der Typ 3 auftritt:

Indexpaare	Anzahl
(2;4), (2;5), ...,(2;n)	n-3
(3;5), (3;6), ..., (3;n)	n-4
....	....
(n-3;n-1), (n-3;n)	2
(n-2;n)	1

Die Gesamtzahl beträgt somit $1+2+..+n-3=\frac{(n-3)\cdot(n-2)}{2}$

Zusammenfassung:

$$V(W) = (n-1)\cdot 2pq + 2\cdot(n-2)\cdot pq + 2\cdot\frac{(n-3)\cdot(n-2)}{2}\cdot 4p^2q^2 - \left(2(n-1)pq\right)^2 =$$
$$2pq\cdot(n-1+n-2) + (n^2-5n+6)\cdot 4p^2q^2 - 4\cdot(n^2-2n+1)p^2q^2 =$$
$$2pq\cdot(2n-3) + 4p^2q^2\cdot(5-3n)$$

Für $p=\frac{1}{2}$ folgt $E(W) = 2\cdot(n-1)\cdot\frac{1}{4} = \frac{n-1}{2}$ und

$V(W) = 2\cdot\frac{1}{4}(2n-3) + \frac{1}{4}\cdot(5-3n) = \frac{n-1}{4}$ und damit die Ergebnisse aus 3.

Um die Wahrscheinlichkeit der Abweichungen vom Mittelwert zu bestimmen, wird die Ungleichung von Tschebyschew herangezogen: $p(|X - E(X)| \geq t\cdot\sqrt{V(X)}) \leq \frac{1}{t^2}$

Für t = 2, 3, 4, 5, 6 und n = 100 sowie p = 0,5 ergibt sich folgende Tabelle:

	Wahrscheinlichkeit, dass die Anzahl der Wechsel außerhalb des links stehenden Intervalls liegen
[39;60]	≤ 0,25
[36;65]	≤ 0,12
[31;70]	≤ 0,063
[26;75]	≤ 0,04
[21;80]	≤ 0,03

Die Tabelle sagt aus, dass Abweichungen mit 61 Wechseln in höchstens 25% aller Fälle auftreten. Diese Aussage ist natürlich deutlich unschärfer als die Aussage im 3. Abschnitt, da in der Tschebyschew-Ungleichung nur Erwartungswert und Varianz und keine weiteren Informationen über die Verteilung eingehen.

Herleitung der Ungleichung von Tschebyschew:

$$V(X)=\sum_{i=1}^{n}\left(x_i-E(X)\right)^2\cdot p(X=x_i)\geq \sum_{|x_i-E(X)|\geq c}\left(x_i-E(X)\right)^2\cdot p(X=x_i)$$

$$\geq \sum_{|x_i-E(X)|\geq c} c^2\cdot p(X=x_i)=c^2\cdot \sum_{|x_i-E(X)|\geq c} p(X=x_i)=c^2\cdot p(|\,X-E(X)\,|\geq c)$$

Es folgt:

$$p(|\,X-E(X)\,|\geq c)\leq \frac{V(X)}{c^2}$$

Für $c=t\cdot\sqrt{V(X)}$ folgt das Gewünschte.

Ergänzung:

X sei die Anzahl der Erfolge eines Bernoulli-Versuches. X ist dann binomialverteilt mit E(X) = np und V(X) = npq. Y=$\frac{X}{n}$ ist dann die relative Häufigkeit mit E(Y) = p und V(Y) = $\frac{pq}{n}$. Wendet man auf Y die Tschebyschew-Ungleichung an, so erhält man:

$$p\left(\left|\frac{X}{n}-p\right|\geq c\right)\leq \frac{pq}{n\cdot c^2} \text{ bzw. } p\left(\left|\frac{X}{n}-p\right|<c\right)>1-\frac{pq}{n\cdot c^2}$$

Dies ist das schwache Gesetz der großen Zahl: Die Wahrscheinlichkeit, dass sich die relative Häufigkeit von der theoretischen Wahrscheinlichkeit um weniger als c unterscheidet, strebt gegen 1.

Es gilt auch das starke Gesetz der großen Zahl: Mit Wahrscheinlichkeit 1 ist $\lim\limits_{n\to\infty}\frac{X}{n}=p$

6. Zufallsgeneratoren

1. $x_n=e^{x_{n-1}+\pi}-\left[e^{x_{n-1}+\pi}\right]$ erzeugt Zufallszahlen mit $0\leq x_n<1$, wobei [ ] die Gaußklammer bedeutet. Die Zufallszahlen 0 und 1 erhält man z.B. durch $z_n=\left\lfloor x_n\cdot 10^k\right\rfloor \bmod 2$, wobei k eine beliebige natürliche Zahl ist. Beträgt die aktuelle Uhrzeit z.B. 15.37, so könnte man $x_0=\frac{15\cdot 60+37}{24\cdot 60}$ wählen.
2. Mit der linearen Kongruenzmethode werden gleichmäßig verteilte ganze Zufallszahlen im Intervall [0; m-1] erzeugt: $x_n=\left(a\cdot x_{n-1}+c\right)\bmod m$ Die Zufallszahlen 0 und 1 erhält man durch $z_n=x_n \bmod 2$. Für den Startwert x_0 muss gelten: $0\leq x_0\leq m$.
   Beispiel: a = 24298, c = 99991, m= 199017
   Beträgt die aktuelle Uhrzeit z.B. 15.37, so könnte man x_0 = 1537 wählen.
   Soll das Intervall [0; m-1] auf das Intervall [a; b] abgebildet werden, so ist folgende Transformation durchzuführen: $x_n'=\frac{x_n}{m-1}\cdot(b-a)+a$
3. Normalverteilte Zufallszahlen werden wie folgt erzeugt: Zunächst wird ein Paar (p_n; q_n) gleichmäßig verteilter Zufallszahlen im Intervall [0;1] erzeugt. Die normalverteilten Zufallszahlen ergeben sich dann wie folgt:

$$x_n = \sqrt{-2 \cdot \ln(p_n)} \cdot \cos(2 \cdot \pi \cdot q_n)$$

4. Quadratischer Restgenerator

   $$x_n = (x_{n-1})^2 \bmod m$$

   Dabei ist m = p·q und p, q sind Primzahlen mit p mod 4 = 3 und q mod 4 = 3. Der Startwert x_0 ist teilerfremd zu m.

   Einige Beispiele:

p	q	m	x
103	419	43157	120
811	619	502009	367
2027	2099	4254673	1251
5011	1187	5948057	6183

   Es werden ganzzahlige Zufallszahlen im Intervall [0; m] erzeugt.

5. Cliff Random Number Generator

   $x_n = \left|100 \cdot \ln\left(x_{n-1}\right)\right| - \left\lfloor\left|100 \cdot \ln\left(x_{n-1}\right)\right|\right\rfloor$, Startwert x_0 mit $0 < x_0 < 1$.

   Es werden Zufallszahlen zwischen 0 und 1 erzeugt.