Bsp. 3.15 *In der BRD gab es im Zeitraum 1970-1990 insgesamt 25 171 123 registrierte Lebendgeburten, davon waren 12 241 392 Mädchen.*

Berechnen Sie die ein 95% *Vertrauensintervall für die Wahrscheinlichkeit einer Mädchengeburt!*

Das zufällige Ereignis einer Mädchengeburt wird dargestellt durch eine Bernoulli-verteilte Zufallsvariable, $X_i \sim B(1,p)$. Sei $n = 25171123$ und

$$S_n = \sum_{i=1}^{n} X_i$$

die zufällige Anzahl der Mädchengeburten.

Wir wissen, $\mathbf{E}S_n = n \cdot p$ und $\mathrm{Var}\, S_n = n \cdot p \cdot (1-p)$.

Weiter sei $u_{0.975}$ das 0.975-Quantil der Standardnormalverteilung, d.h

$$\Phi(u_{0.975}) = 0.975.$$

Nachsehen in der Tabelle liefert $u_{0.975} = 1.96$.

Aus dem ZGWS folgt

$$P(\frac{|S_n - np|}{\sqrt{VarS_n}} \leq u_{0.975}) = 0.95.$$

Die folgenden Ungleichungen gelten jeweils mit Wkt. 0.95:

$$
\begin{aligned}
|S_n - np| &\leq 1.96 \cdot \sqrt{np(1-p)} \\
(S_n - np)^2 &\leq 1.96^2 np(1-p) \\
n^2p^2 - 2S_nnp + S_n^2 &\leq 1.96^2np - 1.96^2np^2
\end{aligned}
$$

$$(n^2 + 1.96^2n)p^2 - (1.96^2n + 2nS_n)p + S_n^2 \leq 0$$

bzw. wenn wir die Schätzung

$$\hat{p} = \frac{S_n}{n}$$

für die Anzahl der Mädchengeburten einsetzen,

für die Randpunkte des Vertrauensintervalls

$$p_{1,2} = \frac{1}{n + 1.96^2}\Big(n\hat{p} + \frac{1.96^2}{2} \pm 1.96\sqrt{n\hat{p}(1-\hat{p}) + \frac{1.96^2}{4}}.$$

Hier haben wir

$$\hat{p} = \frac{S_n}{n} = \frac{12241392}{25171123} = 0.48633$$

95%-Vertrauensintervall: $[0.48613, 0.48652]$.

Bsp. 3.16 (Fortsetzung des vorigen Beispiels) *Angenommen, es würde gelten $p = \frac{1}{2}$. Mit welcher Wkt. würden dann höchstens 12 241 392 auftreten?*

$$
\begin{aligned}
P(S_n \leq 12241392) &= P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{12241392 - np}{\sqrt{np(1-p)}}\Big) \\
&= \Phi(\frac{12241392 - np}{\sqrt{np(1-p)}}) \\
&= \Phi(-137.2) \leq 3 \cdot 10^{-4091}.
\end{aligned}
$$

Bsp. 3.17 (Roulette) *Beim Roulette gibt es 36 Zahlen, die geraden sind schwarz, die ungeraden rot, dazu die 37, die ist grün. Bei Setzen der richtigen Farbe gibt es den doppelten Einsatz, bei Setzen der richtigen Zahl den 36 fachen Einsatz. Zwei Spieler A und B spielen folgende Strategie: A setzt auf Farbe, B auf Zahl. Beide spielen 100 mal, und jetzen jeweils 10 Euro.*

Wie groß ist die Wkt., dass sie nach 100 Spielen mindestens 40 Euro gewonnen haben?

Wir beschreiben die Gewinne/Verluste im i-ten Spiel durch Bernoulli-Zufallsvariablen,

$$X_i: \quad \begin{pmatrix} 10 & -10 \\ \frac{18}{37} & \frac{19}{37} \end{pmatrix}, \qquad Y_i: \quad \begin{pmatrix} 350 & -10 \\ \frac{1}{37} & \frac{36}{37} \end{pmatrix}$$

$$
\begin{aligned}
\mathbf{E}X_i &= 10 \cdot \frac{18}{37} - 10 \cdot \frac{19}{37} = -\frac{10}{37} =: \mu_A \\
Var X_i &= \mathbf{E}X_i^2 - (\mathbf{E}X_i)^2 = 100 - (\frac{10}{37})^2 =: \sigma_A^2 \\
\mathbf{E}Y_i &= 350 \cdot \frac{1}{37} - 10 \cdot \frac{36}{37} = -\frac{10}{37} =: \mu_B \\
Var Y_i &= \mathbf{E}Y_i^2 - (\mathbf{E}Y_i)^2 = 350^2\frac{1}{37} + (-10)^2\frac{36}{37} - (\frac{10}{37})^2 =: \sigma_B^2
\end{aligned}
$$

$$
\begin{aligned}
P\Big(\sum_{i=1}^{100} X_i \geq 40\Big) &= P\Big(\frac{\sum_{i=1}^{100} X_i - n\mu_A}{\sqrt{n}\sqrt{Var X_i}} \geq \frac{40 - n\mu_A}{\sqrt{n}\sqrt{Var X_i}}\Big) \\
&= 1 - \Phi\Big(\frac{40 - n\mu_A}{\sqrt{n}\sqrt{Var X_i}}\Big) \\
&= 1 - \Phi(0.067) = 0.47 \\
P\Big(\sum_{i=1}^{100} Y_i \geq 40\Big) &= P\Big(\frac{\sum_{i=1}^{100} Y_i - n\mu_B}{\sqrt{n}\sqrt{Var X_i}} \geq \frac{40 - n\mu_B}{\sqrt{n}\sqrt{Var X_i}}\Big) \\
&= 1 - \Phi\Big(\frac{40 - n\mu_B}{\sqrt{n}\sqrt{Var X_i}}\Big) \\
&= 1 - \Phi(0.12) = 0.45
\end{aligned}
$$

Wahrscheinlichkeitsverteilungen (Zus.)

Binomial $X \sim B(n, p)$

X : Anzahl von “Erfolgen” bei

n Versuchen und Erfolgswkt. p.

Poisson $X \sim Poi(\lambda)$

X : Anzahl von “Erfolgen” bei

n Versuchen und Erfolgswkt. p,

n groß und p klein, $n \cdot p = \lambda$.

X : $\#$ Ankünfte in einem Zeitintervall

Geometrische Verteilung $X \sim Geo(\lambda)$

X : Anzahl der Versuche bis zum ersten “Erfolg”

Gleichverteilung $X \sim R(a, b)$

Zufallszahlen

Exponential $X \sim Exp(\lambda)$

“gedächtnislose” stetige Verteilung.

Normal $X \sim N(\mu, \sigma^2)$

Zentraler Grenzwertsatz

Fehlergesetz (viele kleine unabh. Fehler)

Kapitel 4

Grundlagen der Simulation

Contents

1 Einführung

Komplexe Problemstellungenen, die einer anlytischen Behandlung nur sehr schwer oder gar nicht zugänglich sind

- Lösung von diskreten (oder analytischen) Optimierungsaufgaben, z.B. Travelling Salesman Problem
- Berechnung von Integralen
- Untersuchung des Verhaltens von Algorithmen, z.B. Sortier- und Suchverfahren
- Theorie oft nur asymptotisch. Verhalten im Endlichen?
- “Wer nix kapiert, der simuliert”.

Stochastische Optimierungsverfahren

- Mutation und Selektion
- Simulated Annealing
- Genetische Algorithmen

Allen diesen Verfahren ist gemeinsam, daß Zustandsübergänge zufällig geschehen und zwischenzeitlich auch mit gewissen (kleinen) Wahrscheinlichkeiten auch schlechtere Lösungen akzeptiert werden.

Vorteil: “Optimum” wird in Polynomialzeit gefunden.

Nachteil: “Optimum” wird nur mit hoher Wkt. gefunden.

Grundlage aller Simulationverfahren sind gleichverteilte Zufallsgrößen $X \sim \mathbb{R}(0,1)$.

$$P(X < x) = \int_0^x dt,$$

d.h. X hat die Dichtefunktion:

$$f(x) = \begin{cases} 1 & \text{, falls } 0 \le x < 1 \\ 0 & \text{, sonst.} \end{cases}$$

Das Kernproblem der Simulation ist deshalb die Erzeugung von Folgen unabhängiger gleichverteilter Zufallsgrößen X_i.

Bez.: Zufallszahlen.

2 Erzeugung von Zufallszahlen

2.1 Exakte Methoden von Hand

Methode 1: Es werden zufällig , gleichverteilt, die Zahlen $0, 1, \ldots, 9$ erzeugt.

$$X : \begin{pmatrix} 0 & 1 & \ldots & 8 & 9 \\ \frac{1}{10} & \frac{1}{10} & \ldots & \frac{1}{10} & \frac{1}{10} \end{pmatrix}.$$

Realisierung:

1. Es werden Karten mit den Zahlen 0 bis 9 beschriftet. Für jede Zahl ist dabei die Anzahl der Karten gleich. Nun zieht man zufällig Karten und legt sie wieder zurück. Die sich ergebende Folge von Ziffern schreibt man auf.
2. Es können die bereits bekannten Urnen– bzw. Glücksradmethoden verwendet werden.

Wir erhalten in jedem Falle eine Folge von Ziffern. Wir schreiben sie in eine Tabelle der folgenden Form:

$$\left| \begin{array}{l} 3\ \underline{8\ 7\ 0\ 9\ 1}\ \ldots \\ 2\ 4\ 9\ 1\ 3\ 2\ \ldots \\ \vdots\ \vdots\ \vdots\ \vdots\ \vdots\ \vdots \end{array} \right|$$

Nun wählen wir zufällig Fünferblocks (es können auch Blocks von mehr Zahlen sein) aus und betrachten diese als Dezimalstellen, d.h. wir erhalten beispielsweise die Zahl $0,87091$. Auf diese Weise erhalten wir Zufallszahlen auf dem Intervall $[0,1[$.

Methode 2: Wir erzeugen zufällig die Ziffern 0 und 1, beispielsweise mittels Münzwurf. Dann erhalten wir eine Zufallsgröße

$$X: \quad \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Wir erhalten eine Folge $d_1 d_2 d_3 \ldots d_n \ldots$ von Nullen und Einsen. Dann ermitteln wir:

$$z := \sum_{i=1}^{n} d_i \cdot 2^{-i} \leq 1 - \Big(\frac{1}{2}\Big)^n$$

Für die so erhaltene Zahl z gilt: $0 \leq z < 1$.

Methode 3: (4–Würfel–Spezialwürfeln)

Wir beschriften vier Würfel nach folgender Vorschrift:

1. Würfel: 0, 1, 2, 3, 4, 5

2. Würfel: 0, 6, 12, 18, 24, 30

3. Würfel: 0, 36, 72, 108, 144, 180

4. Würfel: 0, 216, 432, 648, 864, 1080

Wir werfen diese Würfel gleichzeitig und bilden die Summe der Augen. Das ergibt eine Zahl k, für die gilt: $0 \leq k \leq 1295$. Die Zufallsgröße $X := \frac{k}{1295}$ ist dann annähernd gleichverteilt über dem Intervall $[0, 1[$.

2.2 Elektronische Erzeugung

In elektronischen Geräten (Bauelementen) fließen auch im Ruhezustand Ströme (weißes Rauschen bzw. „white noise“), deren Spannungen zeitlich zufällig schwanken. Nun kann man einen bestimmten Schwellwert der Spannung festlegen und innerhalb von Zeitintervallen gleicher Länge zählen, wie oft dieser kritische Spannungswert überschritten wird. Beispielsweise läßt sich bei jedem Überschreiten des Wertes ein Impuls auslösen. Diese Impulse können dann gezählt werden. Im Falle einer geraden Anzahl von Impulsen wird als Zufallsziffer eine 1 realisiert, andernfalls eine 0. Aus der resultierenden 0–1–Folge erhält man nach obigem Muster eine Zufallszahl.

2.3 Mathematische Erzeugung

Die bisher betrachteten Verfahren sind alle sehr aufwendig und deshalb praktisch schwer anwendbar. Aus diesem Grunde spielen in der Simulation nur die mathematischen Methoden (Algorithmen) zur Erzeugung von Zufallszahlen eine Rolle. Die mit diesen Methoden generierten Zufallszahlen (gewissermaßen ein Ersatz für Zufallszahlen) werden auch als Pseudozufallszahlen bezeichnet. Algorithmen, die Pseudozufallszahlen erzeugen, werden auch Zufallszahlengeneratoren genannt.

Die multiplikative Kongruenzmethode

Wir geben folgende Startwerte vor: $m, a, z_0 \in \mathbb{Z}^+$. Wir definieren die Folge

$$z_{i+1} := a \cdot z_i \pmod m.$$

Offenbar:

$$a \cdot z_i = k \cdot m + z_{i+1}; \qquad 0 \leq z_{i+1} < m \quad (k \in \mathbb{N}, i = 1, 2, \ldots).$$

$$u_i = \frac{z_i}{m}, \quad (i = 1, 2, \ldots)$$

ist eine Folge von Pseudozufallszahlen zwischen 0 und 1.

Frage: Sind diese u_i ($i \in \mathbb{N}$) annähernd eine Folge unabhängiger, auf dem Intervall $[0, 1[$ gleichverteilter Zufallszahlen?

Frage: Geeignete Wahl der Zahlen a, m und z_0.

Bsp. 4.1

- *RANDU (IBM):* $m = 2^{31}$, $a = 2^{16} + 3$;
- *RANDA (PRIME):* $m = 2^{31} - 1$, $a = 16807$;
- *SIMULA (CDC):* $m = 2^{59}$, $a = 5^{11}$.
- *SAS:* $m = 2^{31} - 1$, $a = 397204094$.

Verallgemeinerung: Die lineare Kongruenzmethode

Wir geben wieder Werte vor: $m, a, r, z_0 \in \mathbb{Z}^+$ und definieren die Folge

$$z_{i+1} = (a \cdot z_i + r) \pmod m$$

und die Folge von Zufallszahlen ist

$$u_i := \frac{z_i}{m} \quad (i \in \mathbb{N}).$$

Bsp. 4.2 *Turbo-Pascal:* $z_{n+1} = 134775813 z_n + 1 (\text{mod } 2^{32})$

Die mehrfache lineare Kongruenzmethode

Diese Methode stellt eine Erweiterung der linearen Kongruenzmethode dar. Als Startwerte geben wir hier vor: $m, a_1, \ldots, a_k, r, z_0, \ldots, z_{(k-1)} \in \mathbb{Z}^+$. Wir definieren die Folge für $n > (k-1)$

$$z_n = \left(\sum_{l=1}^{k} a_l \cdot z_{n-l} + r \right) \pmod m.$$

Die Zufallszahlenfolge ist dann wieder

$$u_n := \frac{z_n}{m}.$$

2.4 Eigenschaften von Pseudozufallszahlen

Wünschenswerte Eigenschaften von Zufallszahlen

- Einfacher Algorithmus, wenig Rechenzeit.
- möglichst viele verschiedene Zufallszahlen sollen erzeugbar sein
  $\Rightarrow$ lange Periode.
  $\Rightarrow M$ möglichst groß (etwa in der Nähe der oberen Grenze des INTEGER–Bereichs)
- k-Tupel $(U_1, \ldots, U_k) \sim R(0,1)^k$, $k \leq 10$
  $\Rightarrow$ Test auf Gleichverteilung.
- “Unabhängigkeit”
  Test auf Autokorrelation
  Plot der Punkte (U_i, U_{i+k}), $k = 1, 2...$
  es sollten keine Muster zu erkennen sein.

Bsp. 4.3 *Wir wählen* $m = 2^4$, $a = 11, z_0 = 3$. *Wir berechnen die ersten Werte der Folge:*

$$\begin{aligned} z_1 &= 11 \cdot 3 \pmod{16} = 1 \\ z_2 &= 11 \cdot 1 \pmod{16} = 11 \\ z_3 &= 11 \cdot 11 \pmod{16} = 9 \\ z_4 &= 11 \cdot 9 \pmod{16} = 3 \end{aligned}$$

Dann gilt natürlich: $z_5 = z_1$ *und die Folge wiederholt sich.*

In diesem Beispiel ist also die Periodenlänge statt gleich 16 (wie theoretisch möglich) nur gleich 4. Das zeigt deutlich, wie die Wahl der Parameter die Qualität der ermittelten Pseudozufallszahlen beeinflußt.

Satz 4.1 *Wenn* $m = 2^k$, $a \mod 8 \in \{3, 5\}$, z_0 *ungerade und* $r = 0$ *sind, so hat die multiplikative Kongruenzmethode die maximal mögliche Periodenlänge* 2^{k-2}.

In allen anderen Fällen gilt, daß die Periodenlänge kleiner als 2^{k-2} *ist.*

Satz 4.2 *Die lineare Kongruenzmethode besitzt genau dann die volle Periodenlänge m, falls die folgenden Bedingungen erfüllt sind:*

1. $\text{ggT}(r, m) = 1$ *(*$\text{ggT}(0, m) := m$*);*
2. $a \bmod p = 1$*, für alle Primfaktoren* p *von* m*;*
3. $a \bmod 4 = 1$*, falls* m *ein Vielfaches von* 4 *ist.*

Ziel: Gleichverteilung dieser Pseudozufallszahlen in $[0, 1[$.

Aber: Bilden wir Paare (u_1, u_2), (u_3, u_4), (u_5, u_6), usw. aufeinanderfolgender Zufallszahlen und tragen sie in das Einheitsquadrat ein. Es entsteht ein (zweidimensionales) Scatterplot von Punkten. Die Pseudozufallszahlen sind dann akzeptabel, wenn sich hier eine gleichmäßige Verteilung ergibt und keine Struktur erkennbar ist. Entstehen dagegen (Linien)muster, so ist der Zufallszahlengenerator schlecht.

Diese Darstellung kann auch für k–Tupel definiert werden. Dann haben wir entsprechend Punkte im k–dimensionalen Raum.

Es sei $\{z_i\}_{i\in\mathbb{N}}$ eine Folge von Werten, die mit der multiplikativen Kongruenzmethode mit
$m = 2^t$, $a = 5 \pmod 8$ und $z_0 = 1 \pmod 4$
ermittelt wurden, d.h.:

$$z_{i+1} = a \cdot z_i \pmod{2^t}.$$

$$u_i = \frac{z_i}{2^t}.$$

Wir bilden nun k–Tupel von aufeinanderfolgenden Pseudozufallszahlen:

$$\mathbf{u}_{(k)} = (u_l, \ldots, u_{l+k-1}) = \left(\tfrac{z_l}{2^t}, \ldots, \tfrac{z_{l+k-1}}{2^t}\right).$$

Für diese k–Tupel von Pseudozufallszahlen gilt:

$$\mathbf{u}_{(k)} \in \left(\tfrac{1}{4} \cdot b_1 + G\right) \cap [0,1[^k.$$

Dabei ist:

$$G = \left\{ \sum_{i=1}^{k} q_i \cdot \mathbf{b}_i \colon\, q_1, \ldots, q_k \in \mathbb{Z} \right\}$$

$$\mathbf{b}_1 = \frac{1}{2^{t-2}} \cdot \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{k-1} \end{pmatrix}, \mathbf{b}_2 = \mathbf{e}_2, \ldots, \mathbf{b}_k = \mathbf{e}_k.$$

Das ist ein k–dimensionales Gitter von Zufallszahlen.

Bem: Sei u_0 die erste Zufallszahl. Die ersten k Zufallszahlen haben die Form

$$u_0 \cdot ((1, a, \ldots, a^{k-1})(\text{mod}\, m))/m = u_0 \cdot \frac{\mathbf{b}_1}{4} + g,$$

wobei $g \in G$ ein geeigneter Vektor ist, so daß die $u_l, l = 1, \ldots, k$, auch im Intervall $(0, 1)$ liegen.

Anstelle der ersten kann mit einer beliebigen Zufallszahl begonnen werden.

Bsp. 4.4 (RANDU)

$$M = 2^{31}, \qquad a = 2^{16} + 3, \qquad c = 0$$

$$\begin{aligned} X_{i+2} &= (2^{16}+3)X_{i+1} + c_1 2^{31} \\ &= (2^{16}+3)^2 X_i + c_1 2^{31}(2^{16}+3) + c_2 2^{31} \\ &= (6 \cdot 2^{16} + 9)X_i + 2^{31}\big(2X_i + (2^{16}+3)c_1 + c_2\big) \\ &= 6(2^{16}+3)X_i - 9X_i + c_3 2^{31} \\ &= 6X_{i+1} - 9X_i + c_4 2^{31} \end{aligned}$$

$c_i \in \mathbb{Z}, i = 1, \ldots, 4.$

Daraus folgt:

$$U_{i+2} - 6U_{i+1} + 9U_i \in \mathbb{Z}.$$

Beispielmuster (SAS)

`/sasuser/Stochastik/ZufallszahlenMuster.sa`