Numerik mit Zufall: Monte Carlo-Simulationen

Christian Hettlage

12. Januar 2000

Inhaltsverzeichnis

1 Einführung

Mit über 2100 einarmigen Banditen und 95 Spieltischen [1] bietet das bekannte Casino von Monte Carlo die ideale Gelegenheit, sich über Zufälligkeit und Nichtvorhersehbarkeit Gedanken zu machen. Es ist daher keine allzu große Überraschung, daß gerade dieser Ort Pate für die Benennung von numerischen Methoden stand, die auf der Nutzung von Zufallszahlen basieren. Ein gewisser Hauch von Glanz und Gloria ist der Bezeichnung *Monte Carlo-Methoden* wohl kaum abzustreiten.

Auf den ersten Blick ist die Verwendung von Zufallszahlen erstaunlich: Sinn einer Berechnung ist es ja, ein reproduzierbares Ergebnis, möglichst mit

einer Abschätzung seiner Genauigkeit, zu erhalten. Diesem Anspruch kann ein zufallsbedingtes Verfahren per definitionem nicht gerecht werden. Wir werden jedoch sehen, daß man dennoch auch bei Monte Carlo-Methoden quasi-reproduzierbare Ergebnisse mit einer verläßlichen Fehlerabschätzung erzielen kann.

Die Anwendung von Monte Carlo-Methoden ist (wie ja nicht anders zu erwarten) breitgefächert. Im Rahmen dieser Darstellung wollen wir uns hauptsächlich auf zwei einfache Beispiele beschränken, die im folgenden kurz dargestellt werden.

Zweidimensionales Ising-Modell des Ferromagneten

Wir betrachten ein zweidimensionales quadratisches Gitter von N Atomen i mit einem Spin s_i, der die Werte $-\hbar/2$ und $+\hbar/2$ annehmen kann. Im allgemeinen wird man davon ausgehen, daß zwischen allen Spins dieses Gitters eine Wechselwirkung besteht. Diese fällt jedoch mit wachsender Entfernung stark ab, so daß man näherungsweise davon ausgehen darf, daß nur benachbarte Spins miteinander wechselwirken. Die Energie H des gesamten Gitters läßt sich dann schreiben als

$$H = -\sum_{(i,j)\in\mathrm{NN}} J s_i s_j,$$

wobei NN die Menge der nächsten Nachbarn bezeichnet (vgl. Abb. 1). Wir gehen dabei von periodischen Randbedingungen aus, d. h. wir identifizieren den linken mit dem rechten und den oberen mit dem unteren Gitterrand. Die Konstante J ist materialabhängig. Im Falle eines ferromagnetischen Materials hat sie einen positiven Wert. Liegt nun zusätzlich ein äußeres Magnetfeld B vor, tritt ein zusätzlicher Energieterm auf, und man erhält

$$H = -\sum_{(i,j)\in\mathrm{NN}} J s_j s_k - \sum_i K s_i \tag{1}$$

mit einer positiven, magnetfeldabhängigen Konstanten K. Im folgenden ziehen wir den Faktor $\hbar/2$ in die Konstanten J und K, so daß wir von $s_i \in \{\pm 1\}$ ausgehen dürfen.

Problem des Handelsreisenden

Gegeben seien N beliebig in einer Ebene verteilte Städte S_i und eine Metrik $d(S_i, S_j)$, die den Abstand zwischen zwei Städten S_i und S_j beschreibt.

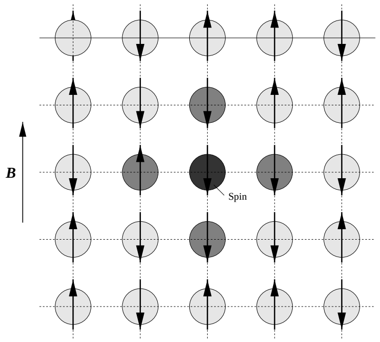

Abbildung 1: Zweidimensionales Ising-Modell mit äußerem Magnetfeld. Die mittelgrau gefärbten Atome sind die nächsten Nachbarn des dunkelgrau gefärbten Atoms.

Das Problem des Handelsreisenden (*traveling salesman problem*, kurz TSP) besteht dann darin, den kürzesten Weg zu finden, auf dem man jede Stadt genau einmal besucht und zum Schluß zum Ausgangspunkt zurückkehrt (vgl. Abb. 2). Anders ausgedrückt, besteht das Problem darin, die Funktion

$$f(W) = \sum_{i=1}^{N} d(W_i, W_{i+1})$$

zu minimieren, wobei W_1 bis W_N alle Städte umfassen und $W_{N+1} = W_1$ gilt. Das so formulierte Optimierungsproblem ist NP-hart. Ein möglicher Lösungsansatz besteht darin, das diskrete in ein kontinuierliches Problem zu überführen, das mit den Methoden der linearen Optimierung behandelt werden kann. Durch Hinzufügen geeigneter Nebenbedingungen wird dabei gewährleistet, daß man die Lösung des ursprünglichen Problems erhält [7]. Im Rahmen dieser Darstellung werden wir eine einfachere Behandlung des Problems diskutieren, die in der Regel zwar einen sehr guten, nicht aber den optimalen Weg ergibt.

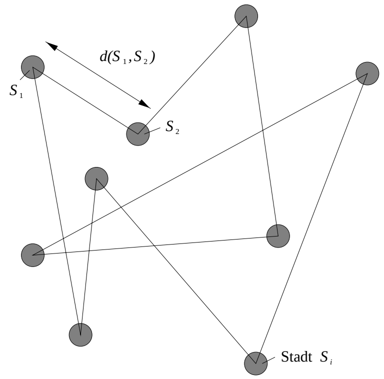

Abbildung 2: Problem des Handelsreisenden

2 Statistische Mechanik

Monte Carlo-Methoden werden häufig zur Lösung von Problemen der statistischem Mechanik verwendet. Zudem sind sie selber inhärent statistisch. Wir wollen daher zunächst kurz die relevanten Grundzüge der statistischen Mechanik beschreiben.

Ausgangspunkt der statistischen Mechanik ist die Überlegung, daß ein makroskopisches System aus sehr vielen einzelnen Teilchen besteht, deren Zustand zum einen unbekannt und zum anderen zeitlich variabel ist. Wir numerieren die (der Einfachheit halber als abzählbar angenommenen) Einteilchenzustände durch und bezeichnen im folgenden mit n_i die Anzahl der Teilchen im i-ten Zustanbd. Ferner handelt es sich beim Mikrozustand eines Systems um seine genaue mikroskopische Beschreibung, d. h. er gibt den Einteilchenzustand eines jeden einzelnen Teilchens explizit an. Das grundlegende Postulat der statistischen Mechanik läßt sich nun wie folgt formulieren:

Bei gegebener Gesamtenergie treten alle erlaubten Mikrozustände eines Systems mit gleicher Wahrscheinlichkeit auf.

Der Rest ist Rechnen. Um zu zeigen, wie dieses Rechnen vor sich geht, betrachten wir ein System von nichtwechselwirkenden Teilchen[1]. Die statistische Behandlung dieses Systems kann auf die folgende Weise erfolgen [5]:

1. Bestimmung der Einteilchenzustände i und der zugehörigen Energien ϵ_i
2. Angabe der möglichen Verteilungen n_i
3. Abzählen der Anzahl verschiedener Mikrozustände, die einer Verteilung n_i entsprechen
4. Berechnung der mittleren Verteilung

Als Beispiel diskutieren wir das im vorangegangenen Abschnitt vorgestellte Ising-Modell für $J = 0$. In diesem Fall ergibt sich:

1. Jedes Atom kann die Zustände „Spin up“ und „Spin down“ annehmen; die zugehörigen Energien betragen $-K$ bzw. $+K$.
2. Alle möglichen Verteilungen n_i müssen die gleiche Teilchenzahl N und Energie U ergeben:
$$n_1 + n_2 = N \tag{2}$$
$$n_1 K - n_2 K = U \tag{3}$$
3. Die Anzahl $t(\{n_i\})$, die einer bestimmten Verteilung n_i entsprechen, ist gerade die Anzahl von Möglichkeiten N Spins so anzuordnen, daß n_1 Spins nach oben und n_2 Spins nach unten weisen. Daher gilt
$$t = \frac{N!}{n_1! n_2!} \tag{4}$$
4. Für sehr viele Teilchen kann man praktisch davon ausgehen, daß nur eine einzige Verteilung n_i vorliegt. Die mittlere entspricht daher der wahrscheinlichsten Verteilung. Man muß also das Maximum von Gl. 4 unter den in Form von Gln. 2 und 3 gegebenen Nebenbedingungen bestimmen. Auf diese Weise ergibt sich $n_2/n_1 = \mathrm{e}^{-2\beta K}$. β läßt sich mit dem Kehrwert des Produkts aus Boltzmann-Konstante und Temperatur identifizieren.

[1] Es sei auf das hier auftretende Dilemma hingewiesen: Auf der einen Seite wird angenommen, daß zwischen den Teilchen keine Wechselwirkung auftritt. Auf der anderen Seite kann sich aber natürlich nur dann eine Gleichgewichtsverteilung einstellen, wenn die Teilchen miteinander wechselwirken…

Diese Überlegungen lassen sich sofort auf das allgemeine System unterscheidbarer, nichtwechselwirkender Teilchen übertragen. Hier ergibt sich für die mittlere Verteilung die Beziehung
$$n_i = \frac{N}{Z} \exp\left(-\frac{\epsilon_i}{k_\mathrm{B} T}\right) \tag{5}$$
mit der durch die Summe
$$Z = \sum_i \exp\left(-\frac{\epsilon_i}{k_\mathrm{B} T}\right)$$
definierten Zustandssumme Z. Geht man nun zum kontinuierlichen Fall über, ist die Verteilung n_i durch eine spektrale Energiedichte $f(\epsilon) \equiv dn/d\epsilon$ zu ersetzen. Für diese gilt völlig analog
$$f(\epsilon) = \frac{1}{Z} \exp\left(\frac{-\epsilon}{k_\mathrm{B} T}\right),$$
wobei Z durch
$$Z = \int_0^\infty g(\epsilon) \exp\left(\frac{-\epsilon}{k_\mathrm{B} T}\right) d\epsilon$$
mit der Zustandsdichte (Zustände pro Energieintervall) $g(\epsilon)$ gegeben ist. Im Experiment interessiert man sich im allgemeinen nicht für die Eigenschaft A_i eines einzelnen Atoms, sondern vielmehr für die entsprechende über alle Atome gemittelte Größe A:[2]
$$A \equiv \langle A_i \rangle = \frac{1}{N} \sum_i A_i n_i = \frac{1}{Z} \sum_i A_i \exp\left(-\frac{\epsilon_i}{k_\mathrm{B} T}\right)$$
Auf diese Weise lassen sich die bekannten thermodynamischen Größen wie Temperatur, Druck und Entropie aus der Beschreibung im Rahmen der statistischen Mechanik gewinnen.

Bei der oben skizzierten Herleitung der Boltzmannverteilung (Gl. 5) wurde angenommen, daß das betrachtete System eine bestimmte Energie besitzt. Es ist instruktiv, diese Annahme etwas näher zu begründen. Es könnte ja beispielsweise sein, daß die innere Energie eines Systems in einem Wärmebad großen Schwankungen unterworfen ist. Immerhin ist ja die Schwankung der Energie eines einzelnen Atoms von derselben Größenordnung wie die Energie selbst!

[2] Wir beschränken uns im folgenden auf den Fall diskreter Zustände.

Wenn wir davon ausgehen, daß sich die Atome des gegebenen Systems nicht gegenseitig beeinflussen, stellen die Energien der einzelnen Atome unabhängige Zufallsgrößen dar. Für diese gilt der zentrale Grenzwertsatz:

Satz 1 (Zentraler Grenzwertsatz) *Es seien $s_1, s_2, \ldots, s_N$ N unabhängige Zufallsgrößen, die den Wahrscheinlichkeitsverteilungen $w_1(s_1)$, $w_2(s_2)$, $\ldots$, $w_N(s_N)$ unterliegen. Ferner sei $x = \sum_{i=1}^{N} s_i$ die Summe der Zufallsgrößen,*

$$\bar{x} = \int_{-\infty}^{\infty} ds_1 w_1(s_1) \int_{-\infty}^{\infty} ds_2 w_2(s_2) \ldots \int_{-\infty}^{\infty} ds_N w_N(s_N) x$$

der zugehörige Erwartungswert und

$$(\Delta x)^2 = \int_{-\infty}^{\infty} ds_1 w_1(s_1) \int_{-\infty}^{\infty} ds_2 w_2(s_2) \ldots \int_{-\infty}^{\infty} ds_N w_N(s_N) (x - \bar{x})^2$$

das entsprechende Schwankungsquadrat. Dann ist die Wahrscheinlichkeitsverteilung $P(x)$ von x durch die Gaußkurve

$$P(x) = \frac{1}{\sqrt{2\pi}\Delta x} \exp\left(-\frac{(x-\bar{x})^2}{2(\Delta x)^2}\right)$$

gegeben.

Der Beweis sei hier übergangen. Er findet sich beispielsweise in [4]. Der Wert d, für den $P(\bar{x} \pm d)/P(\bar{x}) = 1/e$ gilt, ist offensichtlich durch $d^2 = 2(\Delta x)^2$ bestimmt. Die Breite b von $P(x)$ läßt sich daher angeben als $b = \sqrt{2}\Delta x$. Nun kann man leicht zeigen, daß

$$\bar{x} = \sum_{i=1}^{N} \bar{s}_i \sim O(N)$$

und

$$(\Delta x)^2 = \sum_{i=1}^{N} (\Delta s_i)^2 \sim O(N)$$

gilt. Für die relative Breite $b/\bar{x}$ ergibt sich daher

$$\frac{b}{\bar{x}} = \frac{\sqrt{2}\Delta x}{\bar{x}} \sim O\left(\frac{1}{\sqrt{N}}\right).$$

Da die innere Energie eines Systems eine Summe von N unabhängigen Zufallsgrößen darstellt, ist ihre relative Breite also von der Ordnung $1/\sqrt{N}$. Für alle praktischen Belange kann man daher davon ausgehen, daß nur *ein* Energiewert angenommen wird. Allgemein formuliert, ergibt diese Überlegung den

Satz 2 *In einem System von N (nichtwechselwirkenden) Atomen gilt für die relative Breite $\Delta A/A$ einer makroskopischen, über die Atome des Systems gemittelten Größe A*

$$\frac{\Delta A}{A} \sim O\left(\frac{1}{\sqrt{N}}\right).$$

Im Normalfall kann man daher davon ausgehen, daß das System nur den Wert A annimmt.

Abschließend werfen wir noch einen Blick auf das Ising-Modell des vorangegangenen Abschnitts für den Fall $J \neq 0$. Hier wechselwirken benachbarte Atome miteinander, so daß die oben skizzierte Herleitung der Boltzmannverteilung nicht verwendet werden kann. Wir behelfen uns mit einem Perspektivenwechsel: Bislang haben wir ein System immer als Ensemble einzelner Atome aufgefaßt. Angesichts der immensen Zahl von Atomen kann man ein makroskopisches System in der Regel aber ebenfalls als ein Ensemble von makroskopischen Systemen auffassen. (So könnte man beispielsweise ein System mit 10^{24} Atomen in 10^{12} Systeme mit jeweils 10^{12} Atomen unterteilen.) Diese Systeme sind für den Fall des Ising-Modells unterscheidbar und wechselwirken praktisch nicht miteinander, da ja nur für die Atome am Rand der Systeme eine Wechselwirkung vorliegt. Zudem spricht nichts dagegen, die einzelnen Systeme als einzelne Teilchen aufzufassen, deren Einteilchenenergien durch Gl. 1 gegeben sind. Damit ist die obige Herleitung wieder anwendbar, und für das Ensemble von Systemen ergibt sich eine Boltzmannverteilung. Andererseits haben wir jedoch angenommen, daß jedes einzelne System seinerseits makroskopisch ist. Damit müssen die Verteilungsfunktionen für ein einzelnes System und das Ensemble von Systemen die gleiche Form besitzen. Man erhält also für das Ising-Modell mit N Atomen

$$n_i = \frac{N}{Z} \exp\left(-\frac{1}{k_\mathrm{B}T}\left(-\sum_{(i,j)\in\mathrm{NN}} J s_i s_j - \sum_i K s_i\right)\right),$$

wobei Z analog zum Fall $J = 0$ definiert ist.

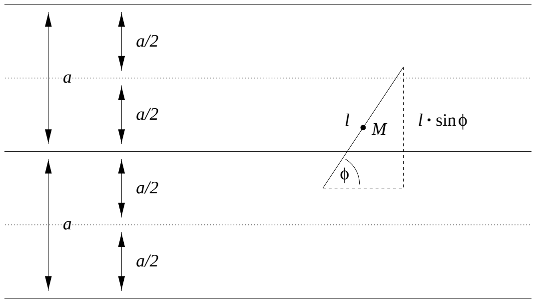

Abbildung 3: Buffonsches Nadelproblem

3 Monte Carlo-Integration

Eine etwas unorthodoxe (und nicht besonders effektive) näherungsweise Bestimmung von π geht auf G. I. Leclerc, Graf von Buffon, zurück (Buffonsches Nadelproblem [8]): Gegeben seien eine Nadel der Länge l und drei parallele Linien, deren Abstand zueinander den Wert $a \geq l$ besitze (vgl. Abb. 3). Durch N-maliges zufälliges Werfen der Nadel werde die Wahrscheinlichkeit P dafür bestimmt, daß die Nadel die mittlere Linie schneidet. „Zufälliges Werfen“ soll hierbei bedeuten, daß der Winkel ϕ zu den Linien im Intervall $[0, \pi]$und der Abstand x des Mittelpunkts M der Nadel von der mittleren Linie im Intervall $[0, \frac{a}{2}]$ gleichverteilt sind.

Die Nadel schneidet die mittlere Linie offensichtlich genau dann, wenn die Bedingung

$$x \leq \frac{l}{2} \sin \phi$$

erfüllt ist. Da zudem die Wahrscheinlichkeitsverteilungen von ϕ und x einfach durch $1/\pi$ bzw. $2/a$ gegeben sind, erhält man für P somit

$$P = \int_0^\pi d\phi \frac{1}{\pi} \int_0^{\frac{l}{2}\sin\phi} dx \frac{2}{a} = \frac{2l}{\pi a}.$$

Man erhält π aus P also gemäß der Formel

$$\pi = \frac{2l}{aP}.$$

Es läßt sich nun problemlos ein Algorithmus formulieren, der dieses Verfahren simuliert:[3]

algorithm buffons-needle(N, l, a)
begin
  $P := 0$;
  for $i = 1, 2, \ldots, N$ **do**
  begin
    $x :=$ random value from $[0, a/2]$;
    $\phi :=$ random value from $[0, \pi]$;
    if $x \leq (l/2) * \sin(\phi)$ **then**
      $P := P + 1$;
  end
  $P := P/N$;
  $p := (2 * l)/(a * P)$;
end

Dieser Algorithmus stellt ein einfaches Beispiel für eine Monte Carlo-Integration dar. Fassen wir das wesentliche etwas allgemeiner: Gegeben sei ein bestimmtes Integrationsgebiet $\Omega \subset \mathrm{R}^n$ mit dem Volumen $|\Omega|$ und eine reellwertige Funktion f auf Ω. Wir wollen das Integral $\int_\Omega f(x) d^n x$ näherungsweise berechnen. Dazu seien für N zufällig aus Ω herausgegriffene Punkte die Funktionswerte $f_i \equiv f(x_i)$ berechnet. Die f_i stellen dann unabhängige Zufallsgrößen dar. Man macht sich leicht klar, daß bei der oben diskutierten Berechnung von π $\Omega = [0, \frac{a}{2}] \times [0, \pi]$ und $f(x, \phi) = \Theta(\frac{l}{2} \sin \phi - x)$ (mit der Heaviside-Funktion Θ) ist.

Die Wahrscheinlichkeitsverteilung $w(f)$ der Funktionswerte von f ist im allgemeinen nicht bekannt. Sie muß jedoch die Beziehung

$$\int_{-\infty}^{\infty} f w(f) df = \frac{1}{|\Omega|} \int_\Omega f(x) d^n x \tag{6}$$

erfüllen. Nun sei die Summe y der f_i betrachtet. Für sie gilt offensichtlich der zentrale Grenzwertsatz, d. h. die Wahrscheinlichkeitsverteilung $P(x)$ stellt eine Gaußverteilung mit dem Maximum bei

$$\bar{y} = \sum_{i=1}^{N} \bar{f}_i = N \bar{f} = N \int_{-\infty}^{\infty} f w(f) df \tag{7}$$

[3]Der Algorithmus ist wirklich nur eine Simulation; insbesondere, da er den Wert von π benötigt…

dar. Aus Gln. 6 und 7 folgt unmittelbar, daß das Produkt aus N, $|\Omega|$ und $\langle f_i \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} f_i$ im Limes $N \longrightarrow \infty$ gegen das zu berechnende Integral konvergiert. Da zudem die die Ungenauigkeit gemäß dem zentralen Grenzwertsatz mit $1/\sqrt{N}$ abnehmen sollte, ist der folgende Satz plausibel:

Satz 3 (Fundamentalsatz der Monte Carlo-Integration) *Es sei f eine auf $\Omega \subset R^n$ definierte reellwertige Funktion. Ferner seien $x_1, x_2, \ldots, x_N$ zufällige Punkte, die gleichverteilt in Ω liegen. Für die zugehörigen Funktionswerte $f_i \equiv f(x_i)$ seien die Größen $\langle f \rangle$ und $\langle f^2 \rangle$ definiert durch*

$$\langle f \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} f_i \qquad \langle f^2 \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} f_i^2.$$

Dann gilt

$$\int_\Omega f(x) d^n x \approx |\Omega|\langle f \rangle \pm |\Omega|\sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}}. \tag{8}$$

Der in Gl. 8 auftretende Fehlerterm ist hierbei lediglich als heuristische Abschätzung zu verstehen. Die Forderung, daß die zufälligen x_i gleichverteilt in Ω sind, erweist sich als nachteilhaft, falls f einen großen Wertebereich umfaßt. Verschwindet f beispielsweise auf einer Teilmenge Ω' von Ω, sollten möglichst keine $x_i \in \Omega'$ gewählt werden, da diese nicht zum zu berechnenden Integral beitragen. Es muß also Ziel sein, statt der Gleichverteilung eine andere geeignete Verteilung $p(x)$ (mit $\int_\Omega p d\Omega = 1$) zu verwenden. In diesem Fall ergibt sich analog zur oben durchgeführten Diskussion

$$\left\langle \frac{f}{p} \right\rangle \stackrel{N \longrightarrow \infty}{\longrightarrow} \overline{\left(\frac{f}{p}\right)} = \int_{-\infty}^{\infty} w\left(\frac{f}{p}\right) d\left(\frac{f}{p}\right) = \int_\Omega \frac{f}{p} p dx$$

und damit die folgende Verallgemeinerung des Hauptsatzes:

Satz 4 (verallgemeinerter Fundamentalsatz) *Es sei f eine auf $\Omega \subset R^n$ definierte reellwertige Funktion. Ferner seien $x_1, x_2, \ldots, x_N$ zufällige Punkte, die gemäß der normierten Verteilung $p(x)$ verteilt sind. Dann gilt:*

$$\int_\Omega f(x) d^n x \approx \left\langle \frac{f}{p} \right\rangle \pm \sqrt{\frac{\langle (f/p)^2 \rangle - \langle f/p \rangle^2}{N}}.$$

Wählt man eine Verteilung $p(x)$ anstelle der Gleichverteilung, spricht man *importance sampling*. Intuitiv wird man davon ausgehen, daß $f(x)/p(x)$ möglichst konstant sein sollte. Für eine nähere Diskussion sei auf [9] verwiesen.

Für ein Gebiet Ω mit leicht zu parametrisierendem Rand und eine praktisch konstante Funktion f wird man für die Berechnung des Integrals $\int_\Omega f d^n x$ kaum auf Monte Carlo-Methoden zurückgreifen. Anders sieht es jedoch aus, falls die Integralgrenzen für einen gegebenen Rand nur sehr schwer aufzustellen sind. Insbesondere aber erweist sich die Monte Carlo-Integration von Vorteil, wenn f auf einer Teilmenge von Ω dominiert, die a priori nicht bekannt ist.

Um zu zeigen, wie man sich in einem solchen Fall behelfen kann, betrachten wir wieder das zweidimensionale Ising-Modell aus Abschnitt 1: Umfaßt das gegebene Gitter N Atome, so gibt es für die Spins insgesamt 2^N verschiedene Zustände. Andererseits wurde in Abschnitt 2 erwähnt, daß praktisch nur eine einzige Verteilungsfunktion realisiert wird. Daß bedeutet zwangsläufig, daß der überwiegende Großteil der möglichen Zustände nicht zu den makroskopischen Größen beiträgt.

Dementsprechend ist es für eine Monte Carlo-Integration von entscheidender Bedeutung, daß man sich auf die relevanten Zustände beschränkt. Allerdings ist zunächst völlig unklar, welche Zustände überhaupt von Bedeutung sind. Wir müssen also einen Algorithmus formulieren, der zielsicher die relevanten Zustände aufspürt.

Um dies zu bewerkstelligen, gehen wir allgemein von einem System aus, das in diskreten Zuständen S_i vorliegen kann, und betrachten ein (hinreichend großes) Ensemble solcher Systeme. $\{n_i\}$ bezeichne die Verteilung der Zustände in diesem Ensemble, d. h. n_i sei die Anzahl der Systeme im Zustand S_i. Ferner sei n_i^* die im thermischen Gleichgewicht zu erwartende Verteilung. Wir führen nun eine Übergangswahrscheinlichkeit $W_{r\to s}$ für den Übergang zwischen den Zuständen r und s ein, die für die Gleichgewichtsverteilung das Prinzip der detaillierten Bilanz erfüllt:

$$n_r^* W_{r\to s} = n_s^* W_{s\to r}. \tag{9}$$

Wendet man diese Übergangswahrscheinlichkeiten auf die Verteilung $\{n_i\}$ an, so kommt es zum einen zu $n_r W_{r\to s}$ Übergangen von Zustand S_r nach Zustand S_s, zum anderen zu $n_s W_{s\to r}$ Übergängen von S_s nach S_r. Für die Differenz $\Delta n_{rs} = n_r W_{r\to s} - n_s W_{s\to r}$ dieser Übergänge gilt daher

$$\Delta n_{rs} = n_r W_{s\to r}\left(\frac{W_{r\to s}}{W_{s\to r}} - \frac{n_s}{n_r}\right) = n_r W_{s\to r}\left(\frac{n_s^*}{n_r^*} - \frac{n_s}{n_r}\right),$$

wobei die letzte Umformung aus Gl. 9 folgt. Man sieht sofort, daß Δn_{rs} positiv ist, falls das Verhältnis n_s/n_r kleiner als im thermischen Gleichgewicht ist.

In diesem Fall wächst n_r also auf Kosten von n_s, was zu einer Vergrößerung des Verhältnisses führt. Ist dagegen das Verhältnis größer als im Gleichgewichtsfall, wird es verkleinert. In beiden Fällen führt also die Anwendung der Übergangswahrscheinlichkeiten zu einer Verschiebung in Richtung Gleichgewichtsverteilung.[4] Eine Folge $(\{n_i\}_k)$ von Verteilungen, die sich durch sukzessive Anwendung der Übergangswahrscheinlichkeiten ergibt, konvergiert somit gegen $\{n_i^*\}$.

Die bisherigen Überlegungen bleiben grundsätzlich gültig, wenn man vom Ensemble zu einem einzelnen System übergeht. Wir erhalten daher [3, 7]

Satz 5 (Metropolis-Algorithmus) *Für ein System mit den diskreten Zuständen S_i seien Übergangswahrscheinlichkeiten $W_{r\to s}$ zwischen S_r und S_s eingeführt, so daß für die Verteilung $\{n_i^*\}$ im thermischen Gleichgewicht das Prinzip der detaillierten Bilanz,*

$$n_r^* W_{r\to s} = n_s^* W_{s\to r},$$

erfüllt ist. Ferner seien T_0 ein beliebiger Anfangszustand und (T_k) eine Folge von Zuständen, die sich durch sukzessive Anwendung der Übergangswahrscheinlichkeiten auf T_0 ergibt. Dann konvergiert (T_k) gegen das thermische Gleichgewicht, d. h. für hinreichend große k stellen die T_k praktisch Gleichgewichtszustände dar. Handelt es sich bei der Gleichgewichtsverteilung des betrachteten Systems um die Boltzmannverteilung, gilt diese Aussage insbesondere für die durch

$$W_{r\to s} = \begin{cases} 1 & \textit{für } H(S_s) < H(S_r) \\ \exp\left(\frac{H(S_r)-H(S_s)}{k_\mathrm{B}T}\right) & \textit{sonst} \end{cases} \tag{10}$$

definierten Übergangswahrscheinlichkeiten, wobei $H(S_i)$ die Energie des Zustands S_i bezeichnet.

Daß die letzte Aussage des Satzes gilt, zeigt man, indem man einfach die in Gl. 10 gegebenen Übergangswahrscheinlichkeiten in Gl. 9 einsetzt. Es ist hierbei noch auf einen subtilen Punkt hinzuweisen: Es ist offenbar möglich, daß die so definierten Übergangswahrscheinlichkeiten $W_{r\to s}$ die Ungleichung

$$\sum_{s=1,\, s\neq r}^{N} W_{r\to s} > 1,$$

[4] Dies ist natürlich nur eine Plausibilitätsbetrachtung. Tatsächlich kann sich das Verhältnis aufgrund anderer Übergänge vom Gleichgewichtszustand weg verschieben. Für alle Zustände als ganzes bleibt die Überlegung jedoch richtig.

erfüllen und sie somit nicht normiert sind. Die obige Argumentation bleibt hiervon jedoch völlig unbeeinflußt. Die praktische Umsetzung des Metropolis-Algorithmus kann nun beispielsweise dadurch erfolgen, daß man, von einem beliebigen Anfangszustand ausgehend, nacheinander für alle Spins überprüft, ob die Energieänderung ΔH bei Umkehr des Spins positiv oder negativ ist. Erweist sie sich als negativ, wird der Spin auf jeden Fall, erweist sie sich als positiv, wird er mit der Wahrscheinlichkeit $\mathrm{e}^{-\Delta H/k_\mathrm{B}T}$ umgekehrt. Für die bei diesem Verfahren verwendete Übergangswahrscheinlichkeit sind die detaillierten Bilanzgleichungen offensichtlich erfüllt. Strenggenommen ändert sie sich dabei von Schritt zu Schritt, was jedoch die Gültigkeit des Metropolis-Algorithmus nicht beeinflußt.

Betrachten wir als Beispiel die Magnetisierung M im Ising-Modell. Diese ist für einen bestimmten Zustand S_i gegeben durch

$$M = \sum_{k=1}^{N} \mu s_k,$$

wobei $s_k \in \{\pm 1\}$ den Spin und μ das Bohrsche Magneton bezeichnen. Für ihre Berechnung ergibt sich aus dem bislang Gesagten prinzipiell der folgende Algorithmus:

```
algorithm magnetisation-by-metropolis(L,T,m)
begin
  for i = 0, 1, ..., L+1 do for j = 0, 1, ..., L+1 do
    s[i][j] := 1;
  for k = 1, 2, ..., m do
  begin
    for i = 1, 2, ..., L do for j = 1, 2, ..., L do
    begin
      d := 2 * J * s[i][j] * (s[i-1][j] + s[i+1][j] + s[i][j-1] + s[i][j+1])
              +2 * K * s[i][j];
      x := random number from [0, 1];
      if d < 0 or x ≤ exp(-d/(k_B * T)) then
        s[i][j] := -s[i][j];
    end
    for k = 1, 2, ..., L do
    begin
      s[k][0] := s[k][L];
      s[k][L+1] := s[k][1];
```

```
      s[0][k] := s[L][k];
      s[L + 1][k] := s[1][k];
    end
    M := 0;
    for i = 1, 2, ..., L do for j = 1, 2, ..., L do
      M := M + s[i][j];
    M := M/(L * L);
    print the magnetization M;
  end
end
```

Strenggenommen müßte man natürlich über die so gewonnenen Magnetisierungen mitteln, anstatt sie einfach auszudrucken. Da der hier gegebene Algorithmus jedoch nur zur Illustration dient, übergehen wir diese Mittelung.

Die praktische Umsetzung des so gewonnenen Algorithmus kann beispielsweise mittels eines C-Programmes erfolgen (vgl. [7]). Abb. 4 zeigt Ergebnisse für ein 100×100 Gitter, die mit einem solchen Programm gewonnen wurden. Man beachte hierbei, daß es sich beim vierten Zustand *nicht* um eine Gleichgewichtskonfiguration handelt. Vielmehr weist die Existenz von Gebieten verschiedener Spins darauf hin, daß die Relaxation ins thermische Gleichgewicht noch nicht abgeschlossen ist.

Auch auf das Problem des Handelsreisenden läßt sich der Metropolis-Algorithmus anwenden. Dies führt zum Prinzip des Simulated Annealing. Bevor wir hierauf näher eingehen, wenden wir uns jedoch zunächst kurz der Problematik der Zufallszahlen zu.

4 Zufallszahlen

Ob die Naturgesetze deterministisch oder von Zufall geprägt sind, ist eine im Rahmen der heutigen Physik nicht zu beantwortende Frage.[5] Sicher ist jedoch, daß es sich bei Rechnern (und insbesondere bei Programmen) um streng deterministische Systeme handelt, so daß man mit ihnen schon per definitionem keine Zufallszahlen erzeugen kann. Nichtsdestoweniger sind in verschiedenen Programmiersprachen „Zufallszahlengeneratoren“ enthalten.

[5] Quantenmechanisch gesehen, hängt dies davon ab, ob die Reduktion der Wellenfunktion ein Zufallsprozeß ist oder nicht (vgl. [2]).

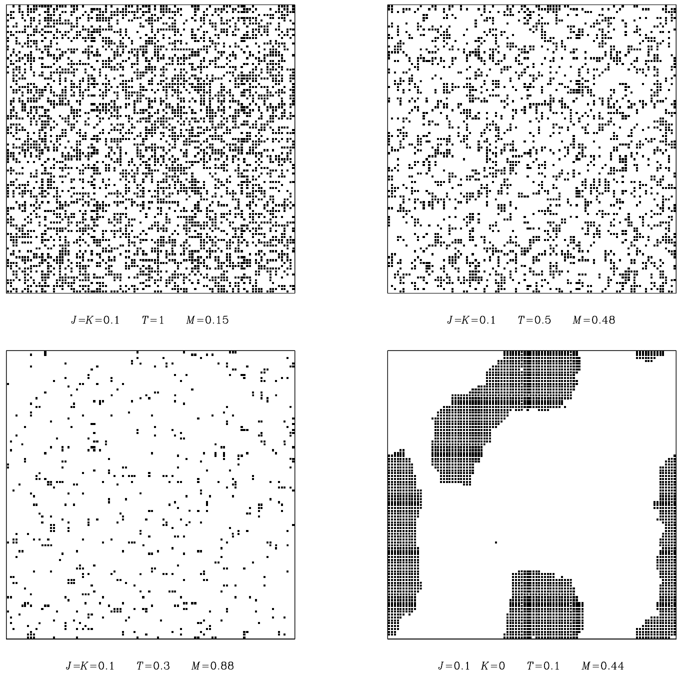

Abbildung 4: Beispiele für die Zustände eines 100×100 Gitters von Elektronenspins und die entsprechenden Magnetisierungen M (in Einheiten des Bohrschen Magnetons) für die jeweils angegebenen Werte für die Konstanten J, K aus Gl. 1 und die Temperatur T. k_{B} ist auf 1 normiert. Der gezeigte Zustand für $K = 0$ ergab sich, indem das Gitter von $T = 1$ ausgehend in Schritten von 0.1 möglichst schnell abgekühlt wurde. Wie die Existenz von Gebieten verschiedener Spinstellung zeigt, ist das System noch nicht ins thermische Gleichgewicht relaxiert.

Im einfachsten Fall berechnen diese die $(i+1)$-te Zufallszahl I_{i+1} gemäß der Formel

$$I_{i+1} = aI_i + c \pmod m,$$

aus der i-ten Zufallszahl, wobei a, c und m geeignete natürliche Zahlen darstellen. Offensichtlich muß sich eine auf diese Weise erzeugte Sequenz von Zufallszahlen spätestens nach m Zahlen wiederholen. Ein für viele Anwendungen akzeptabler (aber keinesfalls perfekter!) Generator ergibt sich für die Wahl $a = 7^5 = 16807$, $m = 2^{31} - 1 = 2147483647$ und c=0 [9].

Selbstverständlich existieren erheblich ausgefeiltere Algorithmen für die Berechnung von Zufallszahlen (vgl. [9]). Es ist jedoch zu beachten, daß auch diese Algorithmen letztlich deterministisch sind. Dazu kommt, daß verschiedene Probleme verschieden auf Korrelationen innerhalb der verwendeten Zufallszahlen reagieren [7]. Die Problematik der Zufallszahlen läßt sich daher zusammenfassen in folgendem

Ratschlag 1 *Traue keinem Zufallszahlengenerator!*

Vielmehr sollte man mit Zufallszahlen gewonnene Ergebnisse stets kritisch überprüfen.

5 Simulated Annealing

In Abschnitt 3 wurde gezeigt, wie man mit Hilfe des Metropolis-Algorithmus Gleichgewichtszustände konstruieren kann. Das Prinzip des Simulated Annealing besteht nun darin, daß man, um den Grundzustand eines Systems bei der Temperatur $T = 0$ zu ermitteln, von einer von Null verschiedenen Temperatur ausgeht und das System langsam auf $T = 0$ abkühlt, wobei man bei den durchlaufenen Temperaturen jeweils mit Hilfe des Metropolis-Algorithmus für die thermische Relaxation ins Gleichgewicht sorgt.

Um zu sehen, worin der Sinn dieses Verfahrens liegt, betrachten wir das Problem des Handelsreisenden. Dieses Problem hat zunächst offensichtlich nichts mit der statistischen Mechanik zu tun. Man kann ihm jedoch einen „statistischen Anstrich“ geben, indem man die Länge L eines bestimmten Weges, d. h.

$$L = \sum_{i=1}^{N} d(W_i, W_{i+1}),$$

formal als Energie auffaßt und zugleich fordert, daß die möglichen Wege des Handelsreisenden der Boltzmannverteilung genügen. (Im folgenden schreiben wir den Exponentialterm in dieser Verteilung einfach als $e^{-L/T}$.) Dann ist klar, daß die Lösung des ursprünglichen Problems gerade dem Gleichgewichtszustand für die Temperatur $T = 0$ entspricht. (Bildhaft gesprochen, läßt sich eine hohe Temperatur vielleicht mit einem energischen Handelsreisenden vergleichen, der es nicht für nötig erachtet, einen kurzen Weg zu suchen.) Der Gewinn dieser Umformulierung besteht darin, daß das neue (statistische) Problem mit dem Metropolis-Algorithmus behandelt werden kann.

Da der Exponentialterm $e^{-L/T}$ für $T = 0$ verschwindet, besagt der Metropolis-Algorithmus in diesem Fall gerade, daß ein neuer Weg genau dann anstelle des alten genommen werden sollte, wenn er kürzer ist. Dies legt den folgenden einfachen Algorithmus zur Lösung des Problems des Handelsreisenden nahe:

algorithm very-heuristic-solution-of-TSP(m)
begin
  oldpath := some randomly chosen allowed path;
  for $i = 1, 2, \ldots, m$ **do**
  begin
    newpath := *oldpath* with some allowed random change;
    if length of *newpath* < length of *oldpath* **then**
      oldpath := *newpath*;
  end
  print the solution *oldpath* of the traveling salesman problem;
end

Die linke Hälfte von Abb. 5 verdeutlicht schematisch die diesem Algorithmus zugrunde liegende Idee. In ihr ist die Länge als Funktion der erlaubten Wege aufgetragen. Von dem Weg n ausgehend, mögen als neue Wege nur Weg $n-1$ und $n+1$ in Frage kommen. Für den gezeigten Längenverlauf führt der Algorithmus daher zwangsläufig zum Minimum, d. h. zur gesuchten Lösung. Anders sieht der Fall jedoch aus, wenn der Längenverlauf die in der rechten Hälfte von Abb. 5 gezeigte Form hat. Auch hier führt der Algorithmus zunächst zu Wegen immer kürzerer Länge. Sobald jedoch das lokale Minimum erreicht ist, ändert sich nichts mehr, da die Länge des neuen ja stets länger als die des alten Weges ist. Das lokale Minimum erweist sich also als

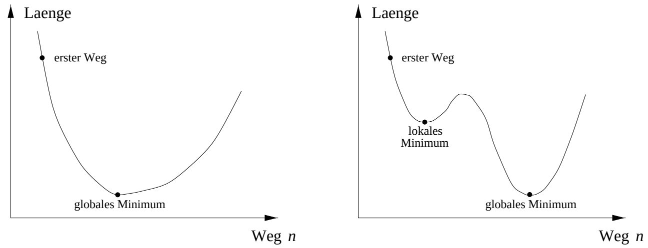

Abbildung 5: Schematische Darstellung möglicher Längenverläufe als Funktion des Weges. Es wird angenommen, daß von Weg n nur Übergänge nach $n-1$ und $n+1$ erfolgen können. *Links:* Der Metropolis-Algorithmus für $T = 0$ führt zum optimalen Weg. *Rechts:* Der optimale Weg wird vom Metropolis-Algorithmus für $T = 0$ nicht erreicht, da das lokale Minimum nicht verlassen werden kann.

Falle, aus der es kein Entrinnen gibt. Man erhält somit einen Weg, der noch erheblich länger als der optimale Weg ist.

Eine Möglichkeit, diesem Problem zu entgehen, bestünde darin, den neuen Weg jeweils rein zufällig zu wählen. Anschaulich würde dies das „Tunneln“ durch den „Potentialwall“ zwischen dem lokalen und dem globalen Minimum ermöglichen. Wenngleich dieses Verfahren zwangsläufig das globale Minimum liefern muß, hat sie einen entscheidenden Nachteil: Sie dauert zu lang! Der Grund hierfür besteht darin, daß bei der Wahl eines neuen Weges keinerlei Information verwendet wird. Man muß also zufällig den optimalen Weg wählen. Das bedeutet aber gerade, daß die Anzahl der notwendigen Versuche von derselben Ordnung wie die Anzahl der erlaubten Wege ist und somit einer astronomisch hohen Zahl entspricht.

Alternativ kann man zulassen, daß mit einer bestimmten Wahrscheinlichkeit ein neuer Weg auch dann genommen wird, wenn er länger als der alte ist. Eine solche Wahrscheinlichkeit ist ja aber sowieso gegeben, falls die Temperatur von Null verschieden ist. Man muß demnach einfach eine hinreichend hohe Temperatur verwenden, um das lokale Maximum zu durchqueren. Andererseits hat eine von Null verschiedene Temperatur natürlich den Nachteil, daß der Gleichgewichtszustand nicht der Lösung des Problems des Handelsreisenden entspricht.

Dies legt nahe, zwar von einer hohen Temperatur auszugehen, diese jedoch langsam auf Null abzusenken. Dies ist offensichtlich gerade das Prinzip des Simulated Annealing. Die entscheidende Frage ist nun, wie schnell die Abkühlung erfolgen darf. Offensichtlich hängt die Antwort hierauf von der Höhe ΔL der zu überquerenden Potentialwälle ab. Da diese unbekannt sind, wird deutlich, daß es sich auch bei der Lösung des Problems des Handelsreisenden mittels Simulated Annealing lediglich um ein heuristisches Verfahren handelt. Insbesondere darf man nicht erwarten, auf diese Weise tatsächlich den optimalen Weg zu erhalten. Allerdings stehen die Chancen gut, daß sich auf diese Weise zumindest ein Weg ergibt, der nur unwesentlich länger ist. Abschließend geben wir noch ein sehr einfaches Beispiel für einen Algorithmus, der das Problem des Handelsreisenden mittels Simulated Annealing behandelt [7]:

algorithm TSP-by-simulated-annealing(N,T_1,T_2,ΔT,m)
begin
  path := some arbitrary allowed path;
  for $T = T_2,\ T_2 - \Delta T,\ T_2 - 2\Delta T,\ \ldots,\ T_1$ **do**
  begin
    for $i = 1,\ 2,\ \ldots,\ m$
    begin
      pos := random integer from [1,N];
      l := random integer from [2,N];
      newpath := *path* with towns $pos,\ pos+1,\ \ldots,\ pos+l-1 \pmod N$
        in reverse order;
      comment For example, let *path* = 3-5-6-2-1-4. Then if $pos = 2$,
        $l = 3$, one gets *newpath* = 3-2-6-5-1-4, and if $pos = 5$,
        $l = 4$, one gets *newpath* = 4-1-6-2-5-3.;
      x := random number from [0,1];
      d := length of *newpath*;
      if $d <$ length of *path* **or** $x \leq \exp(-d/T)$
        path := *newpath*;
    end
    print length of *path*;
  end
end

Die praktische Realisierung dieses Algorithmus in Form eines C-Programmes liefert beispielsweise das in Abb. 6 gezeigte Ergebnis.

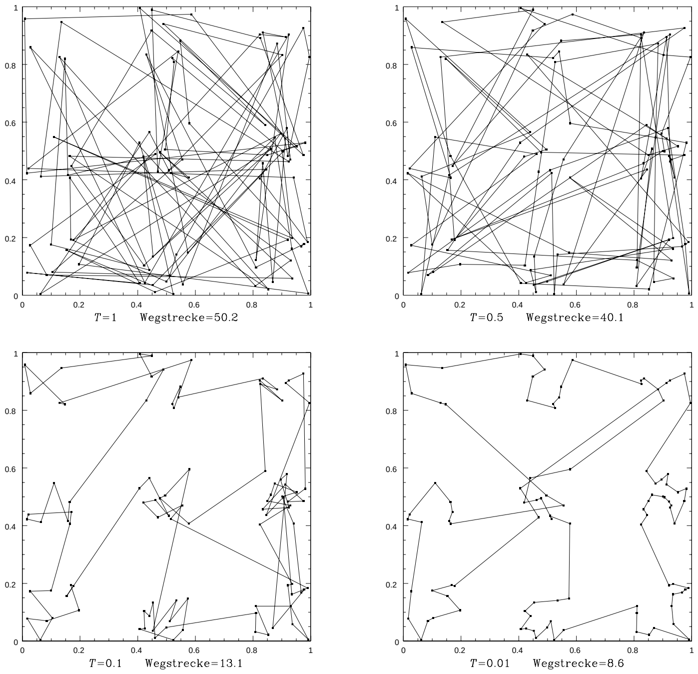

Abbildung 6: Lösung des Problems des Handelsreisenden mittels Simulated Annealing. Betrachtet werden 100 Städte in neun räumlich voneinander getrennten Quadraten, und es wird eine euklidische Metrik vorausgesetzt. Von T=1 ausgehend, erfolgt eine langsame Abkühlung des Systems auf $T = 0.01$. Man erkennt deutlich, daß die Zahl der Übergänge zwischen den neun mit Städten besetzten Gebieten mit sinkender Temperatur abnimmt. Für die verschiedenen Temperaturen ist jeweils die Wegstrecke angegeben, die der Handelsreisende auf dem gezeigten Rundweg zurücklegen muß.

6 Schlußbemerkung

Der interessierte Leser wird sich denken, daß das in dieser Darstellung zu Monte Carlo-Methoden Gesagte doch nicht alles gewesen sein kann. Recht hat er! So findet man beispielsweise in [7] weitere physikalische Probleme, die mittels Zufallszahlen behandelt werden können. Speziell mit der Anwendung von Monte Carlo-Simulationen auf die statistische Mechanik beschäftigt sich [3]. Wie man Monte Carlo-Algorithmen konstruiert und dokumentiert, zeigt [6]. Und natürlich haben auch die *Numerical Recipes* [9] einiges zum Thema Zufallszahlen und Monte Carlo beizusteuern.

Vielleicht fragt sich der Leser zudem noch, ob die Benutzung von Zufallszahlen nun ein Gewinn ist oder nicht. Er wird die Antwort ahnen: Ob Würfeln von Nutzen ist, hängt entscheidend vom konkret zu behandelnden Problem ab. Aber eine Ausnahme von dieser vielleicht etwas unbefriedigenden Aussage ist zumindest denkbar:

> If God played dice...
> ...He'd win! [10]

Literatur

[1] http://www.monte-carlo.com/gaming.html.

[2] Jim Baggott. *The Meaning of Quantum Theory.* Oxford University Press, 1992.

[3] Kurt Binder and Dieter W. Heermann. *Monte Carlo Simulation in Statistical Physics.* Springer Series in Solid-State Sciences. Springer-Verlag, third edition, 1997.

[4] Torsten Fließbach. *Statistische Physik.* Spektrum Akademischer Verlag, zweite Auflage, 1995.

[5] Tony Guénault. *Statistical Physics.* Chapman & Hall, second edition, 1995.

[6] Stanislaw Jadach. *Practical Guide to Monte Carlo.* physics/9906056, 1999.

[7] Wolfgang Kinzel and Georg Reents. *Physics by Computer.* Springer-Verlag, 1998.

[8] Meyers Lexikonredaktion (Hrsg.). *Schülerduden Die Mathematik.* Dudenverlag, dritte Auflage, 1991.

[9] William H. Press et al. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, second edition, 1992.

[10] Ian Stewart. *Does God Play Dice?* Basil Blackwell, 1989.