Random Walk mit Nachwuchs

Projektarbeit zur Vorlesung
Computergestütztes Wissenschaftliches Rechnen
27. August 2021

Bearbeitet von: Mohamad Al Farhan
Betreuer: Tristan Grosskopf
Prof. Heidrich-Meisner
SoSe 2021

Inhaltsverzeichnis

1 Einleitung

Random Walker bzw. zufällige Irrfahrer sind wichtige mathematische Modelle, welche bei der statistischen Physik von großer Bedeutung sind. In dieser Arbeit wird eine Variante des Random-Walker-Modells untersucht, welche Vermehrung der Walker auf einem eindimensionalen periodischen Gitter erlaubt. Ziel ist es ein besseres Verständnis über die Dichte der Walker sowie das Konvergenzverhalten dieser Dichte anhand einer Simulation zu gewinnen.

2 Theorie

Betrachtet wird folgendes Modell: Das eindimensionale periodische Gitter, auf dem sich die Walker befinden, besteht aus L möglichen Aufenthaltsfeldern mit den Indizes zwischen $x = 0$ und $x = L - 1$. Dabei bedeuten die periodischen Randbedingungen, dass das Gitter wie ein Kreis behandelt wird. Befindet sich ein Walker auf dem Feld $x = -1$, so ist dies äquivalent dazu, dass sich der Walker auf Feld $x = (L - 1)$ befindet. Analog sind für einen Walker die Felder $x = L$ und $x = 1$ nicht unterscheidbar. Ein Gitterfeld kann außerdem nur zwischen zwei Zuständen unterscheiden: besetzt oder leer. D.h. mehrere Walker auf einem Feld sind wie ein einziger Walker zu behandeln.
Die Simulation der Walker kann als eine Kette zufälliger Ereignisse, welche t-Schritte lang ist, aufgefasst werden. Zu jedem Zeitschritt wird ein Walker zufällig ausgewählt. Dieser durchläuft dann eine der folgenden Entwicklungen:

- Der Walker verschwindet mit einer Wahrscheinlichkeit p.
- Der Walker bewegt sich um d Felder nach links mit Wahrscheinlichkeit $q/2$ oder um d Felder nach rechts (wieder mit Wahrscheinlichkeit $q/2$). d ist eine ganze Zahl und wird als Gitterkonstante bezeichnet.
- Mit einer Wahrscheinlichkeit von $1 - p - q$ verschwindet der Walker von seinem Platz x und wird durch zwei neue bei den Feldern $x - d$ und $x + d$ ersetzt.

Die Dichte der Walker ρ wird als die Anzahl der Felder, welche von den Walkern besetzt sind (N_{Walker}), dividiert durch die Länge des Gitters L definiert:

$$\rho = \frac{N_{\text{Walker}}}{L} \tag{1}$$

Der arithmetische Mittelwert $\langle \rho \rangle$ einer Datenmenge $M = \{\rho_i | i \in I\}$ welche N Elementen enthält wird nach folgender Gleichung berechnet:

$$\langle \rho \rangle = \sum_{i \in I} \frac{\rho_i}{N} \tag{2}$$

Von der Datenmenge M ist die Standardabweichung[1] in Gleichung 3 definiert.

$$\sigma_\rho = \sqrt{\frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{N}} \tag{3}$$

Nach dem Gesetz der großen Zahlen[2] konvergiert die Summe in Gleichung 2 für $n \to \infty$ fast sicher gegen eine Zahl ρ. Diese wird als Mittelwert bezeichnet.

[1] Diese Definition wurde aus der Vorlesung entnommen.
[2] [1] Satz 5.17 S. 114

3 Methodik

3.1 Erzeugung der zufälligen Zahlen

Mit folgender Gleichung wird eine Zahlenfolge $z_j \in \mathbb{N}$ generiert:

$$z_{j+1} = (a \cdot z_j + c) \bmod m. \tag{4}$$

Dabei sind die Werte der Parameter dieser Folge in Tabelle 1 aufgetragen. Diese Folge wird als ein linearer konstanter Zufallszahlengenerator verwendet. Sie startet bei z_0 und durchläuft zuerst j Iterationen, um potentielle Verzerrungen wegen der Auswahl des Startwerts zu mindern. Danach wird die Zahl z_{j+1} als Quelle der gebrauchten Zufallszahlen verwendet. Wird eine Zufallszahl r im Wertebereich zwischen 0 und 1 benötigt, so wird z_{j+1} durch $(m-1)$ dividiert. Wird eine andere Zufallszahl i mit k möglichen ganzzahligen Werten zwischen 0 und k gebraucht, so wird sie durch

$$i = z_{j+1} mod(k) \tag{5}$$

generiert. Anzumerken ist, dass sowohl bei i als auch bei r der Wertebereich beide Randwerte enthält. Außerdem verwenden i und r gemeinsam den gleichen Generator, wobei jeder Aufruf den Generator um einen Schritt iteriert. D.h. i generieren und danach r generieren liefert unterschiedliche Werte im Vergleich zu der umgekehrten Reihenfolge, da im ersten Fall $i \propto z_{j+1}$ und $r \propto z_{j+2}$ sind, während im zweiten Fall die Indizes umgekehrt liegen.

Tabelle 1: Werte der verwendeten Parameter des Zufallszahlengenerators

Parameter	Wert
z_0	42
j	4367222
a	1664525
c	1013904223
m	2^{32}

3.2 Durchlauf der Simulation

Die Simulationsdauer beträgt 500 Schritte. Es wird mit einem eindimensionalen Array der Länge $L = 10$ gearbeitet. Die Stellen des Arrays können die Werte 0 $\widehat{=}$ Stelle ist frei und 1 $\widehat{=}$ Stelle ist besetzt beinhalten. Die periodischen Randbedingungen werden wie in der Theorie beschrieben implementiert. Zu Anfang befindet sich ein Walker auf der ersten Stelle. Zu jedem Zeitschritt wird eine Zufallszahl i zur Auswahl des Walkers generiert. Der Wertebereich für i ist ganzzahlig zwischen 1 und Anzahl der Walker k. Dann wird eine Zufallszahl $r \in [0, 1]$ generiert und es wird wie in der Theorie beschrieben einer der folgenden Schritte durchgeführt:

- Falls $r < p$, verschwindet der Walker
- Falls $p < 1 - q$, wird der Walker verdoppelt
- Falls $1 - q < r < 1 - \frac{q}{2}$, springt der Walker nach rechts
- Falls $1 - \frac{q}{2} < r < 1$, springt er nach links

Anschließend wird die Anzahl der Walker k aktualisiert. Beträgt $k = 0$, so wird die Simulation abgebrochen. Sonst durchläuft die Simulation die weiteren Schritte mit den aktualisierten Werten. Für die Simulation wird q auf $0,5$ festgesetzt.

3.3 Berechnung der Dichte

Die Simulation wird N mal durchgeführt. Dabei wird nach Ende jedes Durchlaufs die Dichte nach Gleichung 1 berechnet. Von den berechneten Dichten wird der arithmetische Mittelwert nach Gleichung 2 und die Standardabweichung nach Gleichung 3 berechnet. Die Dichte wird als der Mittelwert angegeben.

3.4 Untersuchung der Dichte und deren Konvergenz

Im ersten Teil wird die mittlere Dichte in Abhängigkeit der Wahrscheinlichkeit p untersucht. Dafür wird die mittlere Dichte sowie die deren Standardabweichung über $N = 1000$ Simulationen für verschiedene Gitterkonstanten $d \in \{1, 3, 2\}$ berechnet. Für jede Gitterkonstante wird die mittlere Dichte sowie deren Standardabweichung für 500 verschiedene äquidistante Werte von $p \in [0; 0,5]$ ermittelt.
Im zweiten Teil wird das Konvergenzverhalten der mittleren Dichte untersucht. Hierfür wird ein Wert p ausgewählt, bei dem die mittlere Dichte nicht verschwindet (z.B. $p = 0.05$). Dann wird die Anzahl der verwendeten Simulationen bei der Berechnung der mittleren Dichte variiert. Die mittlere Dichte sowie deren Standardabweichung wird also für N Simulationen zwischen 1 und 3000 berechnet.

4 Implementation

Die main()-Funktion befindet sich in der Datei `main.c`. Alle anderen Funktionen des Programmes sind in der Datei `code.c` definiert. Das Programm kann mit dem Befehl `make` compiliert werden. Dabei ist sicherzustellen, dass im gleichen Verzeichnis die Datei `Makefile` existiert. Zum Starten des Programmes wird die Datei `Loesung` ausgeführt. Die **Laufzeit** kann bis zu 5 Minuten betragen. Die Bibliothek `math` wird für die Modular-Arithmetik sowie die Berechnung der quadratischen Wurzel verwendet. Für die Ausgabe in der Konsole sowie in den Text-Dateien muss die Header-Datei `stdio` aus der Standardbibliothek eingebunden werden.

4.1 Flow des Programmes

In der `main()`-Funktion wird zuerst mit Hilfe der Funktion `generate_r()` die Zahlenfolge z_j j-mal iteriert so wie in Teil 3.1 beschrieben wurde. Die Funktion `generate_r()` kann außerdem später verwendet werden, um die Zufallsvariable r zu erzeugen. Dann wird mit einer Schleife und der Funktion `average_density()` die mittlere Dichte und Standardabweichung gemäß Teil I aus 3.4 berechnet und in der Textdatei `A1_values` abgespeichert. Analog wird Teil II aus 3.4 gelöst und das Ergebnis in der Datei `A2_values` abgespeichert.
`average_density()` berechnet den Mittelwert und Standardabweichung nach 3.3, wofür sie die Funktion `density()` N-mal aufruft. In `density()` wird eine ganze Simulation gemäß 3.2 durchgeführt, wonach die Dichte der Walker übergeben wird. Sie erstellt das Array der Walker. In einer Schleife mit t-Schritten verwendet sie `pick_a_walker()`, um die Zufallszahl i zu generieren und danach den Index des i-ten ausgewählten Walkers zu erhalten, und wendet die zufällige Änderung darauf mit der Funktion `update_state()` an. Um Sprünge der Walker über den Rand des Arrays hinweg zu behandeln, wird die Funktion `find_index()` verwendet. Sie sorgt außerdem für die periodischen Randbedingungen.
Eine genauere Erläuterung der Funktionen und deren Parameter und Ausgaben kann in der Dokumentation in der Datei `code.c` gefunden werden.

5 Ergebnisse und Diskussion

5.1 Mittlere Dichte der Walker

In Abbildungen 1, 5 und 3 sind die in Teil I aus 3.4 berechneten mittleren Dichten in Abhängigkeit der Wahrscheinlichkeit p für die Gitterkonstanten $d \in \{1,2,3\}$ dargestellt. Die Standardabweichungen der Dichte für die Gitterkonstanten $d \in \{1,2,3\}$ sind in Abhängigkeit von p in den Abbildungen 2, 6 und 4 aufgetragen. Die Verläufe bei $d=1$ und $d=3$ sind fast identisch. Für die Beschreibung der Verläufe wird die Gitterkonstante $d=1$ betrachtet. In Abbildung 1 wird eine abnehmende mittlere Dichte ρ bei wachsendem p beobachtet. Ab einer Grenzwahrscheinlichkeit $p \approx 0,16$ verschwindet die mittlere Dichte und bleibt bei $\rho = 0$ konstant. Dies ist zu erwarten, da p der Wahrscheinlichkeit entspricht, bei der ein ausgewählter Walker verschwindet. und je größer diese ist, desto niedriger ist der Erwartungswert der Anzahl der Walker nach einer Simulation. Ergo ergibt sich daraus der beobachtete Verlauf. Die mittlere Dichte ρ bei der Gitterkonstante $d=2$ folgt einem ähnlichen Verlauf (Abbildung 5). Auffällig ist allerdings, dass bei $d=2$ die mittlere Dichte und die Standardabweichung verglichen mit den beiden anderen Gitterkonstanten insgesamt niedriger liegen. Die Grenzwahrscheinlichkeit hat auch einen kleineren Wert. Dies kann damit erklärt werden, dass die Länge $L=10$ und $d=2$ nicht teilerfremd zueinander sind. Dadurch ist es unmöglich, mehr als die Hälfte der Felder mit Walkern zu besetzen. Mit dieser oberen Schranke der Anzahl der Walker kann der schnellere Abfall der mittleren Dichte mit zunehmender Wahrscheinlichkeit p erklärt werden: Weniger Walker verschwinden schneller. Dies erklärt auch, weshalb die Verläufe bei $d=1$ und $d=3$ so ähnlich sind. Beide Werte sind teilerfremd zu $L=10$. Es ist möglich alle Felder mit Walkern zu besetzen.

Für die Verläufe der Standardabweichung der Dichte σ ist die Entwicklung in Abhängigkeit von p für die Gitterkonstante $d=1$ (Abbildung 2) und $d=3$ (Abbildung 4) ebenfalls ähnlich. Für $d=2$ folgt die Standardabweichung einem ähnlichen Verlauf, welcher schneller abfällt und insgesamt niedrigere Werte hat. Für kleine p dominieren zwei Ereignisse bei den Walkern: Springen und Duplizieren. Im Bereich 0 bis $0,6$ tritt noch die dritte Möglichkeit, dass die Walker verschwinden, auf, wodurch die Varianz und damit die Standardabweichung der Dichte zunimmt. Danach wird der Einfluss der beiden anderen Ereignisse schwächer bis das Ereignis ’"Verschwinden‘"dominiert, wodurch die Varianz abnimmt bis sie 0 erreicht. In diesem Fall enden fast alle Simulationen mit 0 Walkern.

5.2 Konvergenzverhalten der mittleren Dichte

Zur zweiten Teil aus 3.4 wird die mittlere Dichte ρ in Abhängigkeit der Anzahl der verwendeten Simulationen N in Abbildung 7 aufgetragen. Die mittlere Dichte liegt ca. konstant bei $0,35$. Bei kleinem N wird eine Streuung um diesen Wert im Größenordnung 10^{-1} beobachtet. Diese nimmt ab je größer N Wird, verschwindet jedoch nicht komplett für die Stichprobe von N. Vermutlich konvergiert die Dichte für $N \to \infty$. Der beobachtete Trend ist im Einklang mit dem Gesetz der großen Zahlen.
Die Standardabweichung der Dichte σ wird in Abhängigkeit von N doppellogarithmisch in Abbildung 8 dargestellt. In dieser Darstellung ist eine lineare Abnahme von σ zu beobachten. Dies ist wiederum im Einklang mit dem Gesetz der großen Zahlen ($\sigma \propto \frac{1}{\sqrt{N}}$) sowie der Definition der Standardabweichung in Gleichung 3. Die Beobachtung stimmt also mit den Erwartungen überein.

5.3 Fehlerquellen und Bewertung der Methodik

Der Zufallszahlengenerator aus 3.1 erzeugt keine echten zufälligen Zahlen. Der Startwert z_0 ist auch im Programm festgelegt. Dies hat zur Folge, dass das Programm deterministisch abläuft, d.h bei jedem Aufruf liefert das Programm die gleichen Ergebnisse. Problematisch wäre dies, wenn die Anzahl der generierten Zufallszahlen innerhalb einer Auswertung größer als m liegt, da z_j periodisch mit Periode m ist. Damit

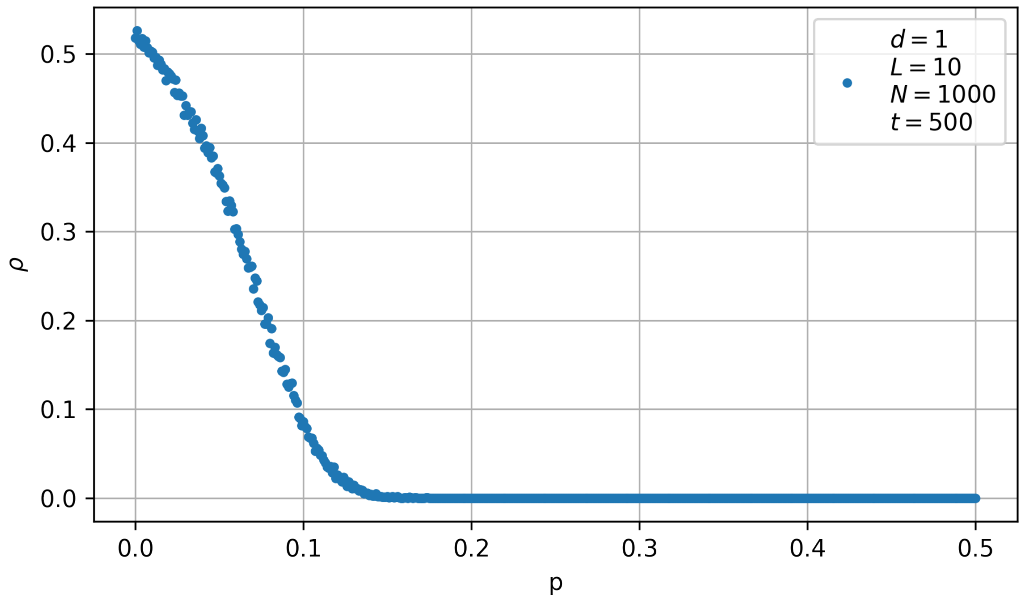

Abbildung 1: Mittlere Dichte der Walker ρ, berechnet als Mittelwert der Dichte über N Simulationen, welche jeweils für t-Schritte laufen, in Abhängigkeit der Wahrscheinlichkeit p. Die Gitterkonstante d beträgt 1.

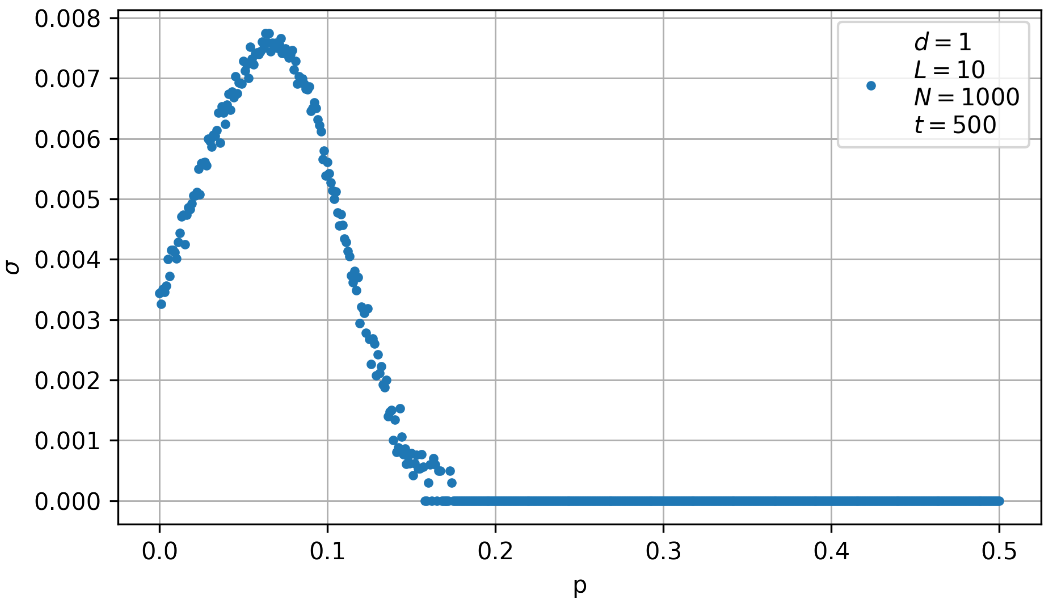

Abbildung 2: Standardabweichung der Mittlere Dichte der Walker ρ, berechnet über N Simulationen, welche jeweils für t-Schritte laufen, in Abhängigkeit der Wahrscheinlichkeit p. Die Gitterkonstante d beträgt 1.

wird die Simulation verzerrt. Dies ist jedoch nicht der Fall, da Teil II die größere Menge an generierten Pseudozufallszahlen braucht und dennoch weniger Zahlen als m braucht:

$$\sum_{N=1}^{3000} 500 \cdot N = \frac{500 \cdot 3000 \cdot 3001}{2} = 2250750000 < 2^{32} = m$$

Für die Zwecke des Programmes ist die Folge der Pseudozufallszahlen gut genug. Das Problem des deterministischen Ablaufs kann z.b. durch Auswahl verschiedenen Startwerten z_0 beispielsweise die Uhrzeit

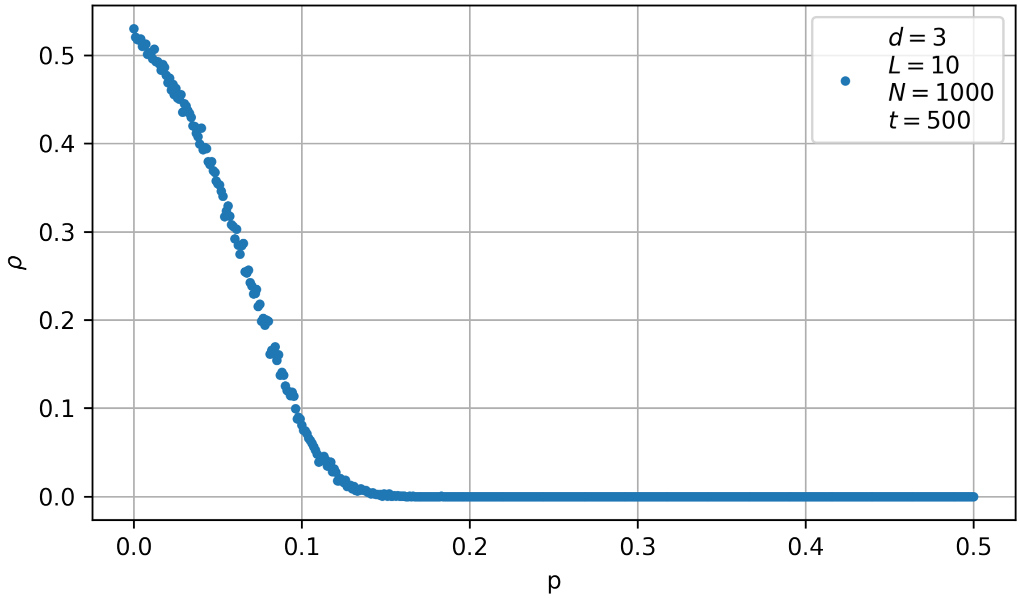

Abbildung 3: Mittlere Dichte der Walker ρ, berechnet als Mittelwert der Dichte über N Simulationen, welche jeweils für t-Schritte laufen, in Abhängigkeit der Wahrscheinlichkeit p. Die Gitterkonstante d beträgt 3.

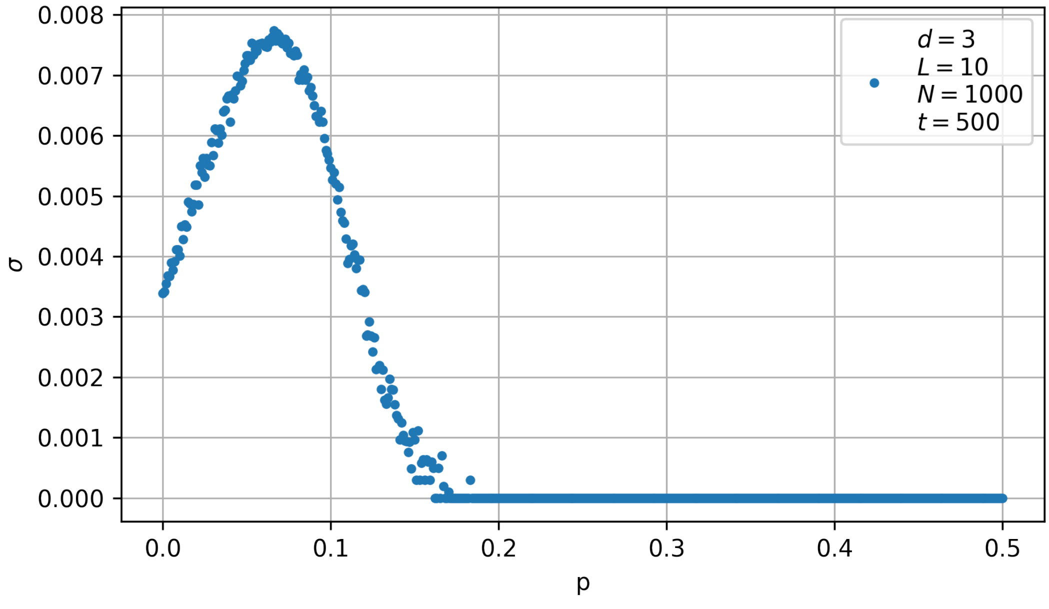

Abbildung 4: Standardabweichung der Mittlere Dichte der Walker ρ, berechnet über N Simulationen, welche jeweils für t-Schritte laufen, in Abhängigkeit der Wahrscheinlichkeit p. Die Gitterkonstante d beträgt 3.

gelöst werden. Durch die Mittelung über mehrere Simulationen sind potentielle Verzerrung durch die Auswahl des Startwerts gemindert.

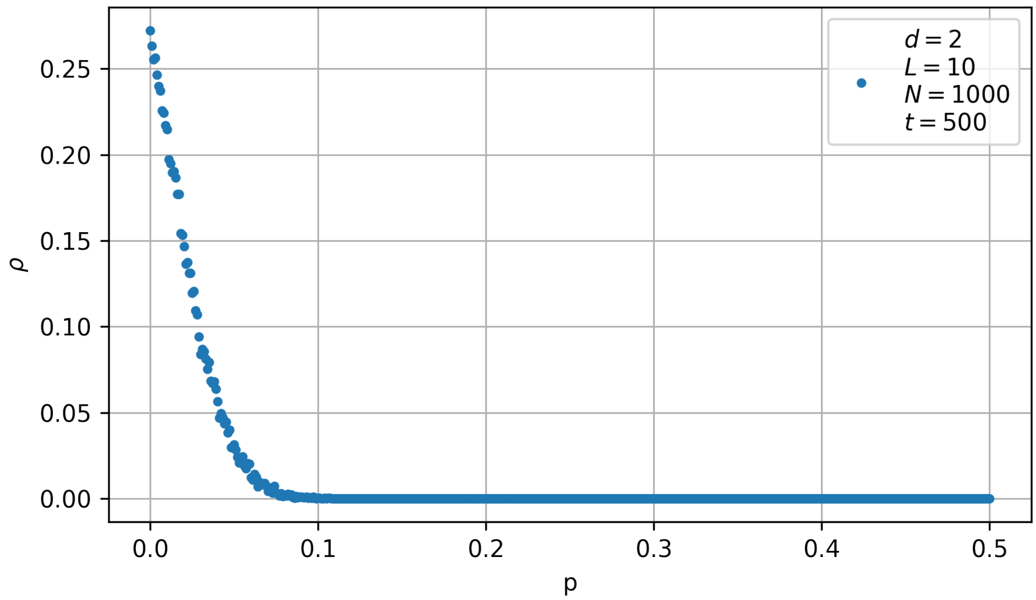

Abbildung 5: Mittlere Dichte der Walker ρ, berechnet als Mittelwert der Dichte über N Simulationen, welche jeweils für t-Schritte laufen, in Abhängigkeit der Wahrscheinlichkeit p. Die Gitterkonstante d beträgt 2.

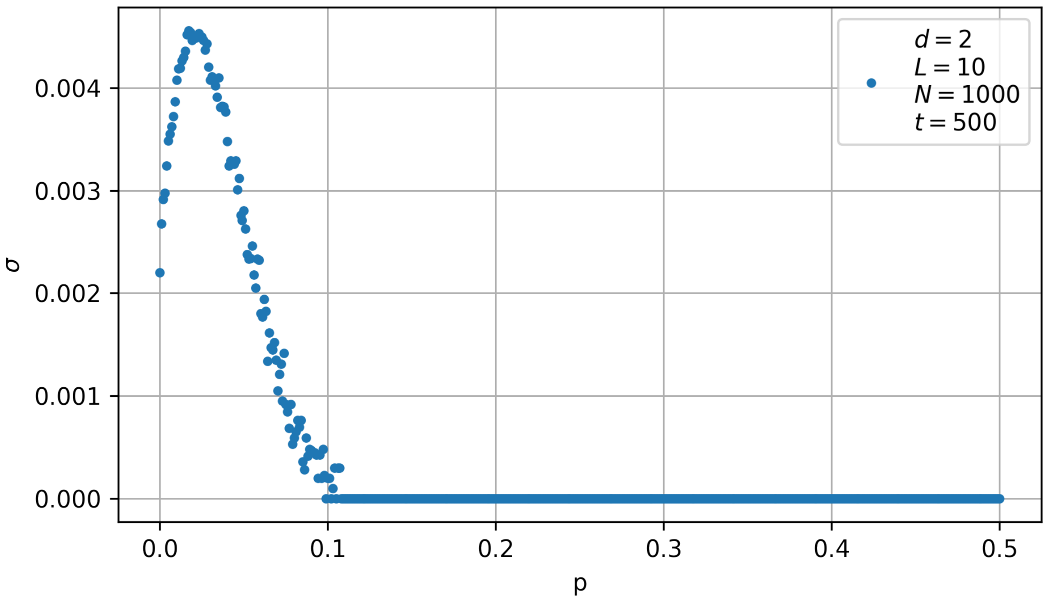

Abbildung 6: Standardabweichung der Mittlere Dichte der Walker ρ, berechnet über N Simulationen, welche jeweils für t-Schritte laufen, in Abhängigkeit der Wahrscheinlichkeit p. Die Gitterkonstante d beträgt 2.

Literatur

[1] A. Klenke: *Wahrscheinlichkeitstheorie* 2. Auflage, Springer-Verlag GmbH Deutschland

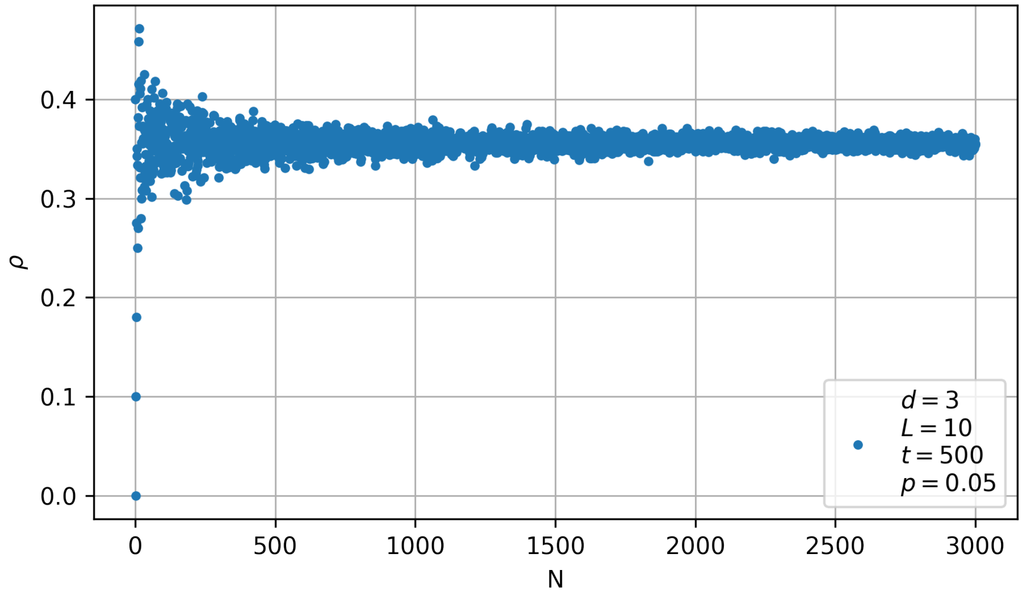

Abbildung 7: Mittlere Dichte ρ in Abhängigkeit der Anzahl der verwendeten Simulationen N. Jede Simulation durchläuft t-Schritte auf einem Gitter mit $L = 10$ Felder und einer Gitterkonstante $d = 3$ sowie die Wahrscheinlichkeit für das Verschwinden eines ausgewählten Walkers $p = 0,05$.

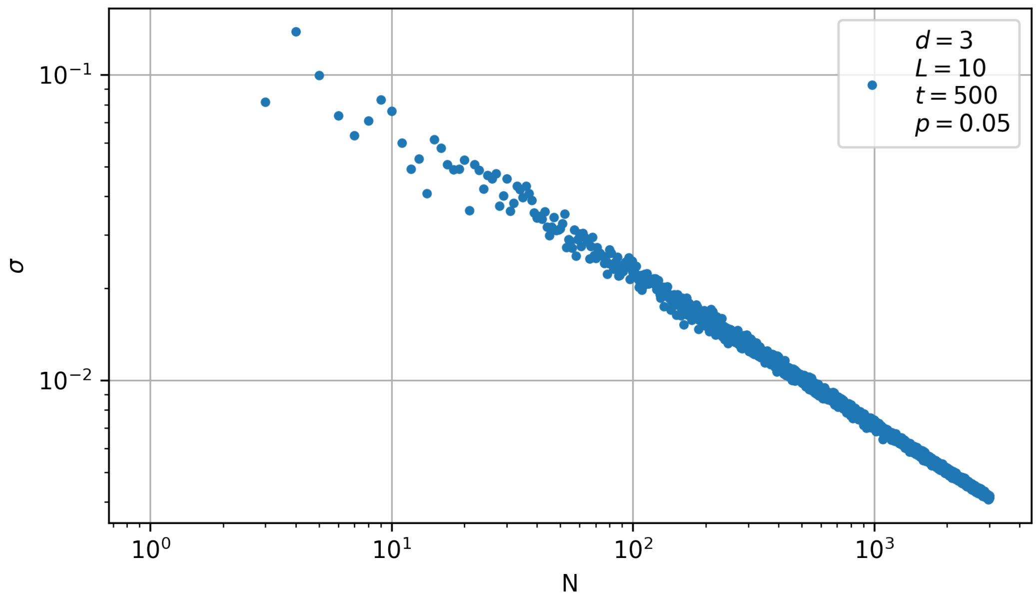

Abbildung 8: Doppellogarithmische Darstellung der Standardabweichung σ der Dichte in Abhängigkeit der Anzahl der verwendeten Simulationen N. Jede Simulation durchläuft t-Schritte auf einem Gitter mit $L = 10$ Felder und einer Gitterkonstante $d = 3$ sowie die Wahrscheinlichkeit für das Verschwinden eines ausgewählten Walkers $p = 0,05$.