Der Weg eines Betrunkenen

2

Hätte es damals schon Computer gegeben, wäre es für unseren Mathematiker um einiges leichter gewesen, den Weg des Betrunkenen zu bestimmen. Er hätte nicht nur eine beliebige Anzahl Schritte simulieren, sondern gleichzeitig auch eine Vielzahl von Betrunkenen auf den Weg schicken können. Holen wir dies zum allgemeinen Verständnis der Methode heute nach.

2.1 Der lineare Weg eines Betrunkenen

Zur Programmierung wird ein Tool benutzt, das vielen Anwendern zur Verfügung steht, die meisten aber kaum kennen, *Visual Basic for Application* (kurz *VBA*). Der Entwicklungseditor ist in jeder Microsoft Office Anwendung, wie beispielsweise Word, PowerPoint und Excel verfügbar. Aufgerufen wird er mit der Tastenkombination ALT+F11. Wer sich mit VBA näher befassen will, findet in meinem Buch *Excel + VBA* im Kap. 1 eine Einführung.

Unser erstes Modell besteht aus einem Betrunkenen, einer Laterne (dem Nullpunkt) und aus den Bewegungen Vorwärts und Rückwärts. Außerdem wird eine feste Schrittlänge vorausgesetzt. Die Darstellung der Bewegung des Betrunkenen, auf einer Zeitachse aufgetragen, könnte in etwa wie in Abb. 2.1 aussehen.

Die nachfolgende Prozedur im VBA-Code simuliert das 1D-Modell mit eintausend Betrunkenen, die sich jeweils einhundert Schritte von der Laterne entfernen. Der Zufallszahlengenerator erzeugt gleichverteilte Pseudozufallszahlen im Intervall von Null bis Eins.

Für die Simulation werden die Zufallszahlen in der unteren Hälfte des Intervalls als Rückschritt gewertet, während die Zufallszahlen in der oberen Hälfte des Intervalls für das Vorwärtskommen sorgen. Gleichzeitig wird der Weg des ersten

H. Nahrstedt, *Die Monte-Carlo-Methode,* essentials,
DOI 10.1007/978-3-658-10149-7_2

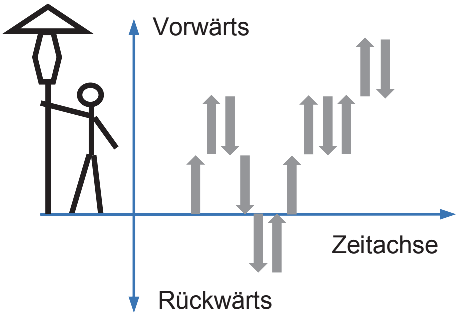
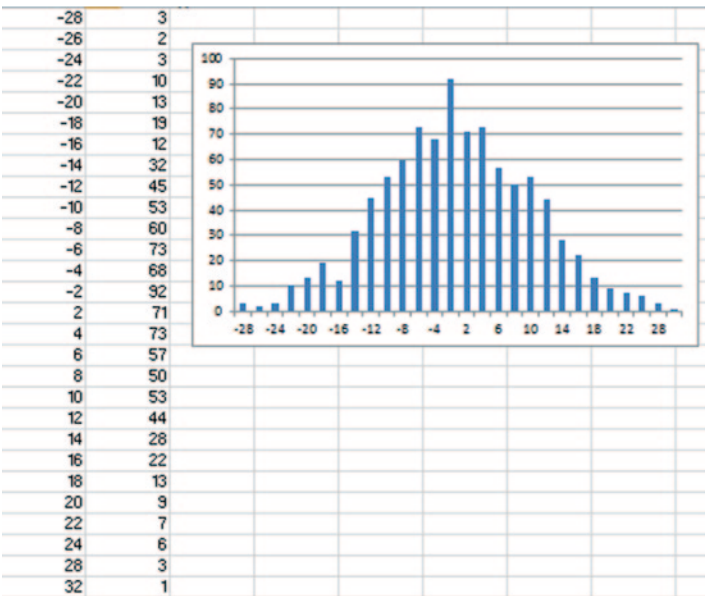


Abb. 2.1 1D-Modell zur Bewegung eines Betrunkenen

Betrunkenen aufgezeichnet, so dass er anschließend anschaulich wiedergegeben werden kann.

Ein mögliches Ergebnis dieser Simulation zeigt Abb. 2.2. Darin ist ein Excel-Tabellenblatt zu sehen, mit der Häufigkeit der erzeugten Positionen der eintausend Betrunkenen nach einhundert Schritten. Das daraus resultierende Balkendiagramm, auch als Histogramm bezeichnet, zeigt eine Häufigkeitsverteilung um den

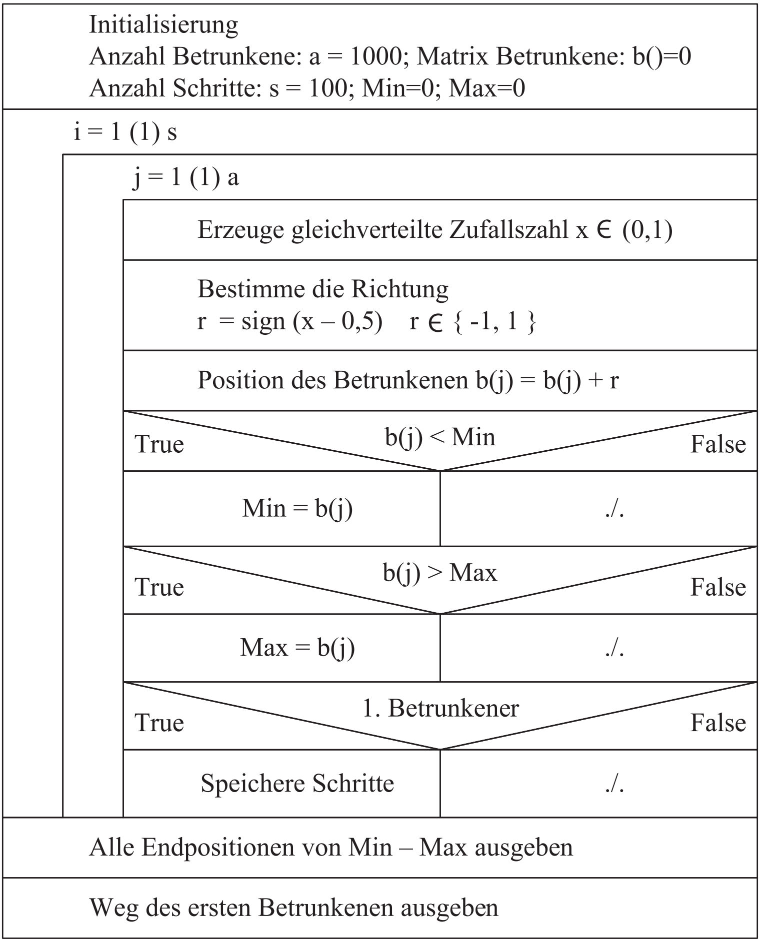


-28	3
-26	2
-24	3
-22	10
-20	13
-18	19
-16	12
-14	32
-12	45
-10	53
-8	60
-6	73
-4	68
-2	92
2	71
4	73
6	57
8	50
10	53
12	44
14	28
16	22
18	13
20	9
22	7
24	6
28	3
32	1

Abb. 2.2 Ergebnis einer Simulation mit dem 1D-Modell

Nullpunkt. Jede erneute Simulation würde andere Werte liefern, aber immer würde sich wahrscheinlich eine solche Verteilung ergeben, doch garantiert ist sie nicht. Es wäre auch absurd hier von einer Garantie zu sprechen, denn die ist durch den Zufallscharakter der Stichprobe ja bereits ausgeschlossen.

Algorithmus 1 1D-Simulationsmodell



Download 1: mc_01_Zufallsweg

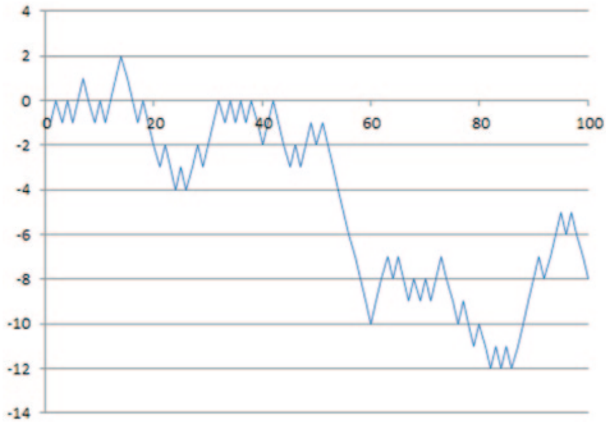


Abb. 2.3 Simulierter Weg des ersten Betrunkenen

Abbildung 2.3 zeigt noch die simulierten Schritte des ersten Betrunkenen. Während auf der x-Achse die Zeit in Schritten aufgetragen ist, zeigt die y-Achse die Abweichung von der gewollten Richtung.

2.2 Die Normalverteilung

Mit jeder Simulation, auch mit mehr Betrunkenen und Schritten, ergeben sich immer andere Werte. Es ist das Aufgabengebiet der induktiven Statistik, aus Stichproben und Teilbeobachtungen Gesetzmäßigkeiten der Grundgesamtheit abzuleiten. Dazu bedient sie sich unterschiedlicher Methoden.

Ein erster Schritt ist, die Daten mit Hilfe von *Verteilungsparametern* zusammenzufassen. Sie beschreiben die Lage der Daten und ihre Ausbreitung um einen mittleren Wert, den so genannten *Mittelwert*. Er bestimmt sich aus n Merkmalen x_i nach folgender Formel

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

Doch es gibt auch viele Betrunkene, die nicht wieder bei der Laterne landen. Ein gebräuchliches Maß für diese Abweichungen ist die *Varianz*. Aus einer Menge von n Merkmalswerten x_i bestimmt sich die Varianz σ^2 aus der Gleichung

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 \tag{2.2}$$

Um die Varianz bestimmen zu können, muss also zuvor der Mittelwert bekannt sein. Die Quadratwurzel aus der Varianz wird als *Standardabweichung* σ bezeichnet.

Das Simulationsmodell erhält nach der Auswertung noch folgende Ergänzung im Code, die diese Kenngrößen auswertet und sie mittels der Anweisung `debug.print` in das Direktfenster der Auswertung schreibt (Algorithmus 2).

Algorithmus 2 Auswertung der Häufigkeitsverteilung

```
i = 1; Mittelwert m = 0; Anzahl a = 0
Für jede Position p
    m = m + p(i)
    a = a +1
    i = i +1
Mittelwert m = m / a
i = 1; Varianz v = 0
Für jede Position p
    v = v + (p(i) – m)²
    i = i +1
Varianz v  = v / (a – 1)
Standardabweichung σ  = √(v)
```

Die Betrachtung des Diagramms lässt vermuten, dass es sich bei der Darstellung um eine *Glockenkurve* handelt oder genauer gesagt, dass hier eine Normalverteilung vorliegt. Und in der Tat haben Stichproben mit einer hinreichenden Menge

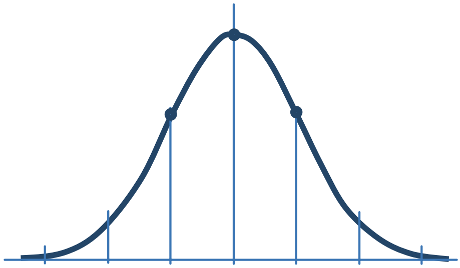


Abb. 2.4 Die Kurve der Normalverteilung

aus einer Grundgesamtheit die Eigenschaft einer Normalverteilung. Ja mehr noch, sie sind durch Mittelwert μ (Gl. 2.1) und Standardabweichung σ (Gl. 2.2) ausreichend beschrieben. Dies ist der Inhalt des *Zentralen Grenzwertsatzes*. Er sagt weiterhin aus, dass Erwartungswert und Standardabweichung der Stichprobe auch die der Grundgesamtheit entsprechen. Denn mitunter sind die Daten der Grundgesamtheit zu groß und nur Stichproben verwendbar. Mit den Werten von $\mu=0$ und $\sigma=1$ spricht man von einer *Standardnormalverteilung*. Abbildung 2.4 zeigt die Kurve der Normalverteilung mit ihren Eigenschaften.

So ist sie symmetrisch zum Mittelwert, der auch mit Modus, Median oder Erwartungswert bezeichnet wird. Die Kurve ist unimodal, da sie nur ein Maximum besitzt. Die Wendepunkte der stetigen Kurve sind genau eine Standardabweichung vom Erwartungswert entfernt.

Die Aussagen aus einer Simulation sind immer nur so gut wie das benutzte Modell. Und hier liegen auch die größten Fehlerpotentiale. So wird sich in der Realität kein Betrunkener auf einer Linie vor und zurück bewegen. Um den wirklichen Prozess im Modell abzubilden, muss das Modell verbessert werden.

2.3 Der Weg eines Betrunkenen in der Ebene

Wir kommen der Realität schon etwas näher, wenn wir zu Vorwärts und Rückwärts auch noch Links und Rechts zulassen. Unser Modell erhält nun zwei Koordinaten und der Weg eines Betrunkenen könnte wie in Abb. 2.5 dargestellt aussehen. Der nachfolgende Algorithmus 3 simuliert ein 2D-Modell mit eintausend Betrunkenen, die sich jeweils einhundert Schritte von der Laterne entfernen. Der Zufallszahlengenerator erzeugt zwei Pseudozufallszahlen für einen Schritt im Intervall von Null bis Eins. Für die Simulation werden die Zufallszahlen in der unteren Hälfte

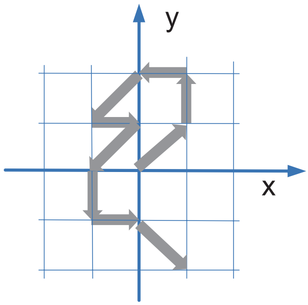


Abb. 2.5 Möglicher Weg eines Betrunkenen in der Ebene nach dem 2D-Modell

des Intervalls als negative Bewegung im Sinne der Koordinaten x und y gewertet, während die Zufallszahlen in der oberen Hälfte des Intervalls für eine positive Bewegung stehen. Das Ergebnis zeigt ebenfalls eine Verteilung um den Nullpunkt (Laterne), aber nun in der Ebene (Abb. 2.6).

Im Algorithmus werden auch die Positionen des ersten Betrunkenen in einer Tabelle ausgegeben. Abbildung 2.7 zeigt diesen Weg. Auch wenn die Simulation einen großen Weg in y-Richtung vorwärts und rückwärts zeigt, so befindet sich die Endposition schließlich wieder nahe der Nulllinie (roter Punkt).

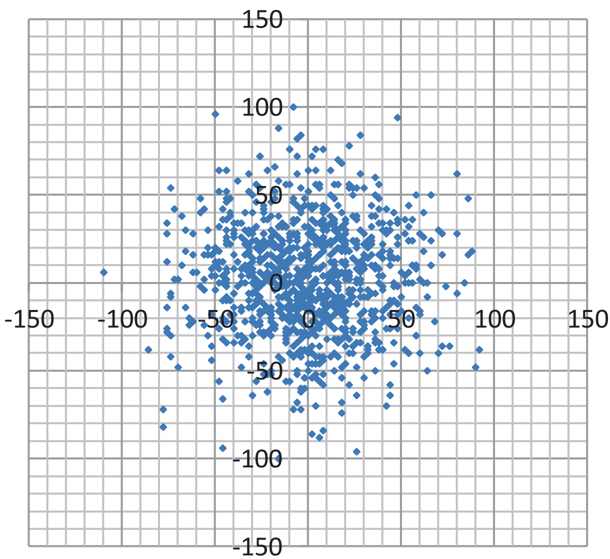


Abb. 2.6 Endpositionen nach einer Simulation

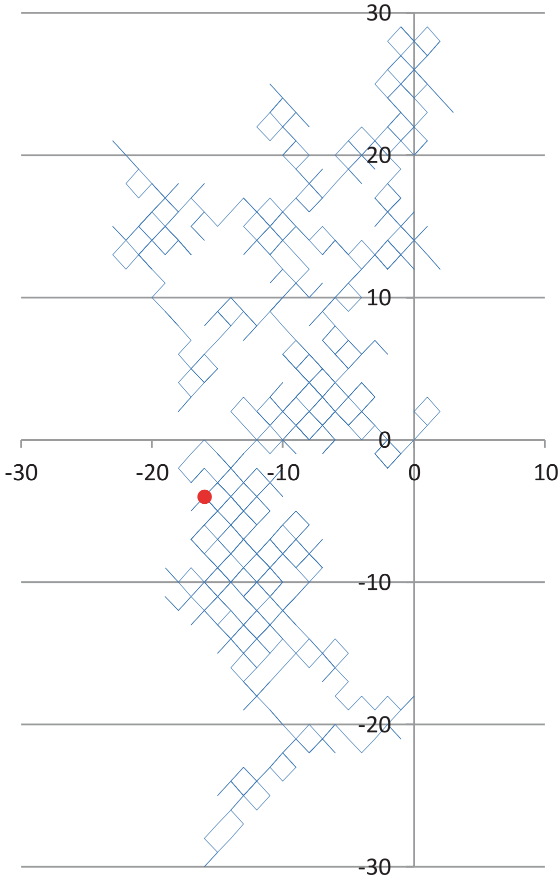
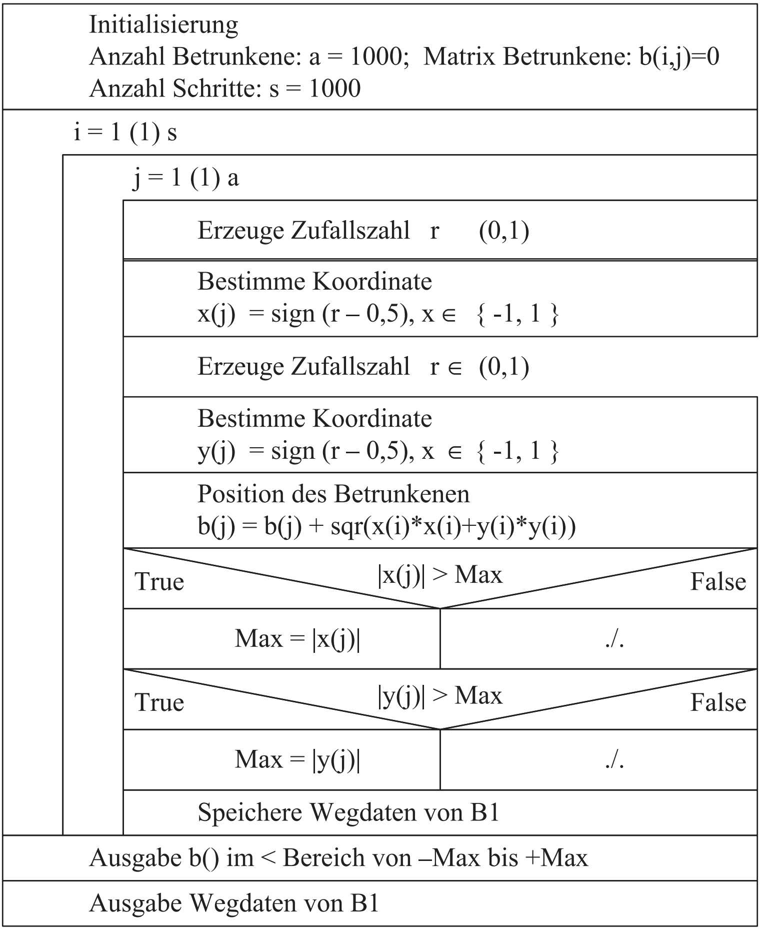


Abb. 2.7 Simulierter Weg des ersten Betrunkenen

Algorithmus 3 2D-Simulationsmodell



Download 3: mc_03_Zufallsweg

Aus den Positionen in der Ebene lässt sich nur eine ungefähre Anhäufung erkennen. Aus den Koordinaten wird in einer weiteren Spalte der vektorielle Abstand jeder Position zur Laterne bestimmt.

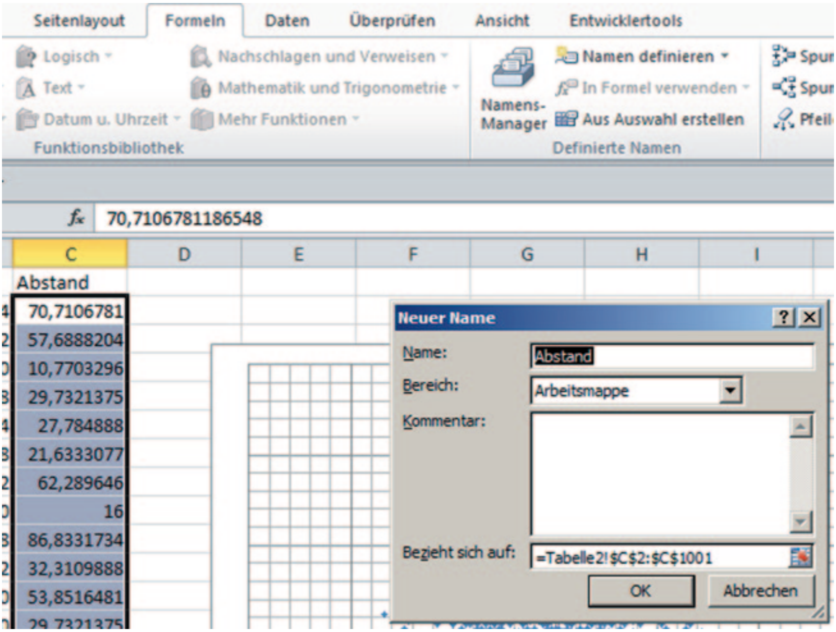


Abb. 2.8 Bereichsnamen für alle Abstände vergeben

Eine Häufigkeitsverteilung erhalten wir, wenn wir die berechneten Abstände in der Spalte C in Klassen einteilen. Dazu bekommt der Datenbereich C2:C1001 den Bereichsnamen *Abstand* (Abb. 2.8).

Eine Matrixfunktion wird in Excel mit den Tasten STRG+UMSCH+ENTER abgeschlossen. Die Funktion Häufigkeit ist eine solche und erstellt eine Häufigkeitsverteilung als einspaltige Matrix in genau dem markierten Bereich (Abb. 2.9).

Eine Darstellung der Häufigkeitsverteilung in einem Säulendiagramm liefert, leicht erkennbar, wieder eine Normalverteilung, was ja zu erwarten war. Doch die Symmetrieachse liegt diesmal nicht bei null wie in dem 1D Modell. Das 2D-Modell liefert eine Erkenntnis, die das 1D-Modell nicht liefern konnte.

An diesem einfachen Beispiel zeigt sich anschaulich die Problematik der Simulation. Nur ein Modell mit allen Einflussparametern kann auch realistische Aussagen erbringen. Das 2D-Modell zeigt uns, dass es eher unwahrscheinlich ist, dass ein Betrunkener zur Laterne zurückfindet.

Auf einem weiteren Tabellenblatt nehmen wir eine Klasseneinteilung in 10-ner Stufen vor. Auch dieser Bereich A2:A17 erhält den Bereichsnamen *Klassen*. Dann

Springer

http://www.springer.com/978-3-658-10148-0

Die Monte-Carlo-Methode
Beispiele unter Excel VBA
Harald, N.
2015, IX, 45 S. 34 Abb., Softcover
ISBN: 978-3-658-10148-0