

Schüler-SimuLab
Kursreihe stochastische
Simulationen
Kurs 4

Weitere stochastische Simulationen

Stefan Hartmann

Forschungszentrum caesar

25. Januar 2008

Ein paar einleitende Worte

Nach den Reaktionen auf meinen ersten Kurs zu stochastischen Simulationen, der sehr gut bei allen Teilnehmerinnen und Teilnehmern ankam, habe ich mich entschieden -auch wenn dies eigentlich nicht geplant war- einen weiteren Kurs zu verschiedenen stochastischen Simulationen zu konzipieren und diesen dem eigentlich im Anschluss an den letzten Kurs geplanten Kurs zur stochastischen Finanzmathematik, der nun wohl bis März oder April 2006 fertiggestellt wird, noch voranzustellen.

Im Gegensatz zum einführenden Kurs, der einen eher spielerischen Einstieg in das Thema mit einer Nutzung bereits im Vorfeld fertiggestellter Programme darstellte, wird das Hauptaugenmerk im vorliegenden Kurs darin liegen, die Schülerinnen und Schüler vermehrt in die Lage zu versetzen, stochastische Probleme geeignet zu modellieren, stochastische Simulationen am Computer zu programmieren und die Ergebnisse geeignet zu interpretieren.

Dem Kurs geht daher in der Praxis eine kleine Einführung in das Arbeiten mit VBA in Excel voraus. Der Grund liegt nicht darin, dass ich diese Sprache für besonders geeignet halte oder ähnliches (eher im Gegenteil), sondern ich habe mich dazu entschieden, weil man relativ sicher davon ausgehen kann, dass jede Teilnehmerin und jeder Teilnehmer Excel inklusive VBA zu Hause besitzt und daher von dort aus alles noch einmal in Ruhe nachvollziehen kann. Auch möchte ich die Schülerinnen und Schüler nicht mit besonderem Informatikfachwissen oder Programmiertechniken vertraut machen. Das Programmieren dient hier nur als Mittel zum Zweck und soll in erster Linie die grundsätzliche Implementierbarkeit von Problemen aus dem Bereich der Stochastik auf einen Computer in Form von Simulationsprogrammen aufzeigen und näherbringen. Zu einem Transfer in andere Programmiersprachen und eine effizientere Programmierung mögen die Schülerinnen und Schüler in Eigenarbeit zu Hause ermutigt werden, hoffentlich mit genügend Anreizen aus dem Kurs versehen. Diese Einführung in die VBA-Programmierung wird hier im Skript nicht illustriert; ich beschränke mich hier auf die mathemati-

schen Fragestellungen und Lösungen.

Inhaltlich ist der Kurs breit gefächert: Simulationen mit geometrischem Hintergrund wie das bekannte Stockproblem oder das nicht minder berühmte Nadelexperiment von Buffon stehen ebenso im Fokus wie eine elementare Einführung in die Theorie der Random Walks und der Ruinwahrscheinlichkeiten. Am zweiten Tag des Kurses geht es um spielerische Fragestellungen, etwa um Gewinnstrategien beim (Mini-)Kniffel sowie Gewinnaussichten beim (Mini-)Monopoly und Glücksrad. Bei letzterem wird „nebenbei“ mathematisch die hypergeometrische Verteilung diskutiert. Die Stichprobenvarianz wird als Mittel zur Messung der „Güte“ einer stochastischen Simulation eingeführt. Auf diese Weise werden zwei verschiedene Möglichkeiten zur Schätzung der Kreiszahl π verglichen.

Die im Kurs verwendeten Programme sind aus Schülerpraktika von Leif Wilberts, Maike Hansen und Ilona Faßbender entstanden. Ich danke allen dreien von Herzen für ihre riesige Mühe und großartige Arbeit!

Ich hoffe, dass der Kurs gut ankommt, allen Teilnehmerinnen und Teilnehmern viel Freude bereitet und zu einer Vermehrung des mathematischen Wissens und möglicher Simulationstechniken führt, die auch bei der Lösung von konkreten Aufgaben aus klassischen naturwissenschaftlichen Gebieten hilfreich sein können.

Bonn, den 19.01.2006

Stefan Hartmann

Kapitel 1

Simulationen mit geometrischem Hintergrund

Simulation 1: Das Stockproblem

Wir möchten uns zunächst mit dem folgenden, aus meiner Sicht sehr witzigen und interessanten Problem beschäftigen:

Man zerbricht einen Stock zufällig in drei Teile. Wie groß ist die Wahrscheinlichkeit, dass man damit ein Dreieck legen kann?

Zunächst einmal müssen wir klären, ob wir die Aufgabenstellung überhaupt einheitlich verstanden haben:

Gruppenaufgabe:

Wie verstehst du diese Aufgabe? Gibt es eventuell auch noch andere Möglichkeiten, wie die Aufgabenstellung interpretiert werden könnte? Welche möglichen Interpretationen fallen euch ein?

Man kann die Aufgabenstellung (mindestens) auf die folgenden vier Weisen verstehen:

Interpretation 1

Man wählt die beiden Bruchstellen **gleichzeitig** und zufällig aus.

Interpretation 2

Man wählt die beiden Bruchstellen **nacheinander** und zufällig aus. Hierbei bricht man zunächst ein Stück zufällig ab, lässt es zu Boden fallen und bricht dann **vom dem noch in der Hand befindlichen Reststück** zufällig noch ein Stück ab.

Interpretation 3

Man wählt die beiden Bruchstellen **nacheinander** und zufällig aus. Hierbei bricht man zunächst ein Stück zufällig ab und dann **von einem der beiden entstehenden Stücke** noch ein Stück (die Auswahl des Teils, von dem man das zweite Stück abbricht, ist hierbei zufällig).

Interpretation 4

Man wählt die beiden Bruchstellen **nacheinander** und zufällig aus. Hierbei bricht man zunächst ein Stück zufällig ab und dann **von dem längeren der beiden Stücke** noch ein Stück.

Gruppenaufgabe:

Wie hoch schätzt du die verschiedenen Wahrscheinlichkeiten ein? Gib einmal Tipps ab! Sind sie alle gleich hoch oder gibt es Unterschiede? Wann kann man überhaupt aus drei Stöcken ein Dreieck zusammenlegen? Hängt das vielleicht von dem Verhältnis der Längen der Seiten ab? Wie genau?

Aufgabe 1:

Versuche für alle vier Interpretationen mittels stochastischen Simulationen eine näherungsweise Wahrscheinlichkeit dafür zu ermitteln, dass man die drei Stücke zu einem Dreieck zusammensetzen kann. Stimmen die Ergebnisse mit deiner Vermutung überein?

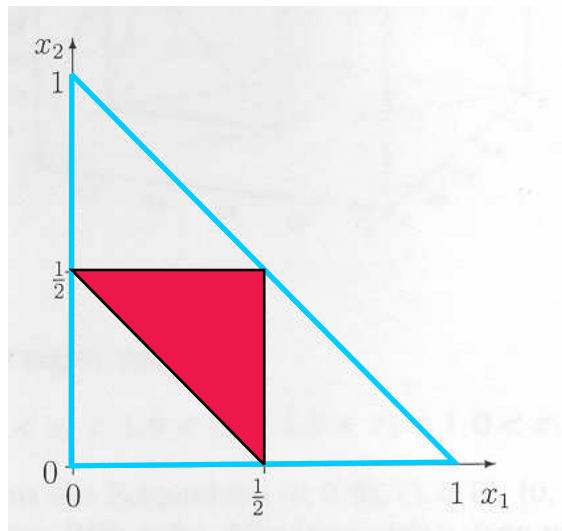
Wir möchten das Problem nun mathematisch auflösen. Zum Teil seid ihr, je nach Alter und vor allem Vorwissen, noch nicht in der Lage, die Lösungen zu verstehen; in diesem Fall könnt ihr euch ja einfach das exakte Ergebnis anschauen, ohne die Rechnung nachzuvollziehen.

Zunächst Allgemeines: Die Stocklänge sei für die folgenden Überlegungen ohne Einschränkung gleich 1. Man kann genau dann ein Dreieck aus drei Stücken zusammenlegen, wenn jedes der drei Teile kürzer ist als die beiden anderen zusammen (es muss also $x < y + z$ für alle Wahlen von Seiten x , y und z gelten; diese Ungleichung nennt man auch **Dreiecksungleichung**). Dies ist offenbar genau dann der Fall, wenn alle drei Stücke kleiner als $\frac{1}{2}$ sind.

zur Interpretation 1

Es sei x_1 die Länge des ersten Stücks und x_2 die Länge des zweiten Stücks. Damit ist die Länge des dritten Stücks $1 - x_1 - x_2$. Es werden x_1 und x_2 aus dem Intervall $(0, 1)$ mit der Randbedingung $x_1 + x_2 < 1$ zufällig gewählt. (Wir machen das im Programm so, dass wir die beiden „Schnittpunkte“ zufällig und unabhängig voneinander auf dem Intervall $(0, 1)$ wählen.) Die Frage ist, wie groß die Wahrscheinlichkeit dafür ist, dass $x_1 < \frac{1}{2}$, $x_2 < \frac{1}{2}$ und $x_1 + x_2 > \frac{1}{2}$ ist.

Dazu schaue man sich das folgende Bild an:



der günstige Bereich für das Zusammenlegen der Stöcke zu einem Dreieck

Der Punkt (x_1, x_2) wird aus dem Dreieck mit den Eckpunkte $(0, 0)$, $(0, 1)$ und $(1, 0)$ gewählt. Alle Punkte, die in dem rot markierten Teildreieck liegen, erfüllen die Bedingung. Die Fläche des Teildreiecks entspricht genau $\frac{1}{4}$ des großen Dreiecks. Daher ist die Wahrscheinlichkeit, in dieser Interpretation aus den drei Stücken ein Dreieck zusammenlegen zu können, gerade $p = \frac{1}{4} = 0.25 = 25\%$.

Bemerkung

Allgemein gilt (in dieser Interpretation): Die Wahrscheinlichkeit, dass man ein n -Eck legen kann, wenn man den Stock zufällig in n Teile zerlegt, beträgt:

$$p_n = 1 - \frac{n}{2^{n-1}}.$$

zur Interpretation 2

Es sei x_1 die Länge des ersten Stücks und x_2 die Länge des zweiten Stücks. Damit ist die Länge des dritten Stücks $1 - x_1 - x_2$. Zuerst wird x_1 zufällig aus dem Intervall $(0, 1)$ ausgewählt und dann x_2 zufällig aus dem Intervall $(0, 1 - x_1)$. Die Frage ist wiederum, wie groß die Wahrscheinlichkeit dafür ist,

dass $x_1 < \frac{1}{2}$, $x_2 < \frac{1}{2}$ und $x_1 + x_2 > \frac{1}{2}$ ist.

Für fortgeschrittene (!) Leser folgt noch die genaue Berechnung:

Offenbar ist

$$f_{X_1=x_1}(x_2) = \frac{1}{1-x_1}$$

die bedingte Dichte von X_2 gegeben $X_1 = x_1$. Für festes x_1 liegt (x_1, x_2) genau dann im günstigen Bereich, wenn

$$\frac{1}{2} - x_1 < x_2 < \frac{1}{2}$$

gilt (vergleiche die obige Abbildung). Wir erhalten daher für die Wahrscheinlichkeit, dass ein Dreieck zusammengelegt werden kann:

$$\begin{aligned} p &= \int_0^{\frac{1}{2}} \int_{\frac{1}{2}-x_1}^{\frac{1}{2}} f_{X_1=x_1}(x_2) dx_2 dx_1 \\ &= \int_0^{\frac{1}{2}} \int_{\frac{1}{2}-x_1}^{\frac{1}{2}} \frac{1}{1-x_1} dx_2 dx_1 \\ &= \int_0^{\frac{1}{2}} \frac{x_1}{1-x_1} dx_1 \\ &= \int_0^{\frac{1}{2}} -\frac{1-x_1}{1-x_1} dx_1 + \int_0^{\frac{1}{2}} \frac{1}{1-x_1} dx_1 \\ &= -\frac{1}{2} + [-\ln(1-x_1)]_0^{\frac{1}{2}} \\ &= -\frac{1}{2} - \ln\left(\frac{1}{2}\right) \\ &\approx 0.193. \end{aligned}$$

In diesem Fall ist die Wahrscheinlichkeit, ein Dreieck legen zu können, also deutlich kleiner, nämlich ungefähr 19.3%!

Bemerkung

Auch in dieser Interpretation interessiert uns natürlich die Wahrscheinlichkeit p_n des verallgemeinerten Problems, dass man ein n -Eck legen kann, wenn man den Stock zufällig in n Teile zerlegt. Dieses Problem wurde jedoch bis heute noch nicht analytisch gelöst. Man weiß nur, dass p_n mit n monoton steigt und dass für $n \geq 5$ folgendes gilt:

$$0.28 < p_n < 0.31$$

mit

$$p_n \approx 0.31 \quad \text{für große } n.$$

zur Interpretation 3

Dieses Problem ist genau das gleiche wie bei der Interpretation 2 und liefert daher auch die gleiche Wahrscheinlichkeit wie dort. Der Grund ist der Folgende: Die Fälle, wo wir bei der Interpretation 2 zufällig ein Stück abbrechen, das kleiner als $\frac{1}{2}$ ist, entsprechen genau den Fällen, wo wir uns bei der Interpretation 3 zufällig für die längere Seite entscheiden, von der wir anschließend ein weiteres Stück abbrechen. Andererseits entsprechen die Fälle, wo wir bei der Interpretation 2 zufällig ein Stück abbrechen, das größer als $\frac{1}{2}$ ist, genau den Fällen, wo wir uns bei der Interpretation 3 zufällig für die kürzere Seite entscheiden, von der wir dann anschließend ein weiteres Stück abbrechen.

zur Interpretation 4

Da wir die Hälfte der Fälle „rausschmeißen“ und diese nun nicht mehr zu betrachtenden Fälle alle **garantiert zu keinem Dreieck** führen, verdoppelt sich natürlich die Wahrscheinlichkeit im Vergleich zu Interpretation 3. Diese beträgt daher ungefähr 0.386, also 38.6%.

Simulation 2: Das Buffon'sche Nadelproblem

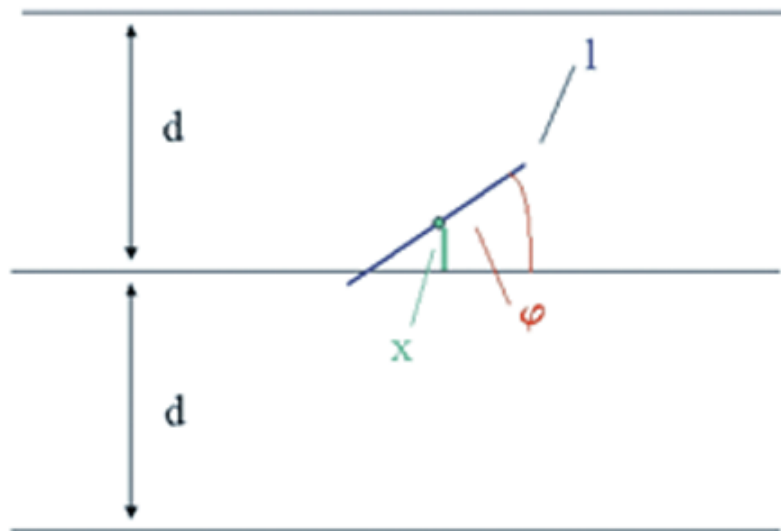
In der Ebene seien parallele Geraden gezogen, die den Abstand d haben. Auf diese Ebene werde zufällig eine Nadel der Länge l geworfen, wobei $l < d$ sei. Mit welcher Wahrscheinlichkeit schneidet die Nadel eine der eingezeichneten Geraden?

Dieses Experiment wurde von dem französischen Naturforscher Georges Louis Leclerc de Buffon (1707-1788) im Jahre 1777 erdacht. Die Antwort auf diese Frage wird uns eine weitere Möglichkeit liefern, die Kreiszahl π simulativ anzunähern. Wir werden sie mit der klassischen MC-Simulation aus dem Kurs „Erste Stochastische Simulationen“ vergleichen. Im Internet gibt es zahlreiche Seiten, wo dieses Verfahren implementiert worden ist und mittels JAVA-Applets auch selbst durchgeführt werden kann, beispielsweise auf <http://www.mste.uiuc.edu/reese/buffon/buffon.html>. Wir verwenden aber unsere eigene Simulation (wenn auch ohne Applet), weil wir darin auch die Ermittlung gewisser statistischer Kenngrößen (siehe unten) einfügen können.

Zur Modellierung und damit zum mathematischen Teil macht man den folgenden Ansatz:

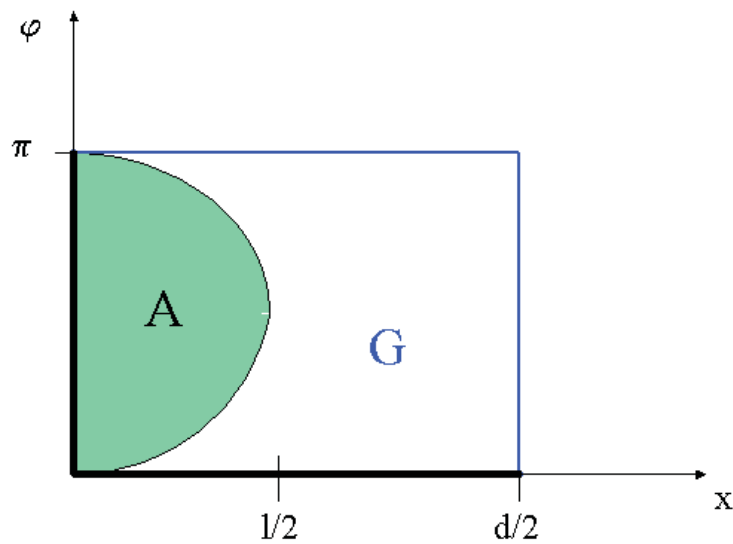
Wir ziehen in einer Ebene parallele Geraden im Abstand d und lassen auf die Ebene Nadeln der Länge l fallen. Der Einfachheit halber nehmen wir $l < d$ an. Dadurch kann eine Nadel höchstens eine der Geraden schneiden. Wie groß ist nun die Wahrscheinlichkeit, dass sie genau eine Gerade schneidet?

Die Lage der Nadel kann durch zwei Daten festgelegt werden, die Einfluss auf die Frage haben, ob eine Gerade geschnitten wird oder nicht: durch die Lage des Mittelpunktes in der Ebene und durch den Winkel, den die Nadel mit den Geraden einschließt. Offenbar spielt aber nur die Koordinate eine Rolle, die senkrecht zu den Geraden liegt, also der „Abstand des Mittelpunktes der Nadel von der nächstgelegenen Geraden“. Diesen bezeichnen wir mit x . Weiterhin sei φ der Winkel, den die Nadel mit dieser Geraden einschließt. Deutlich mögen die Bezeichnungen durch die folgende Skizze werden:



Lage der Nadeln zu den parallelen Geraden

Für jedes Versuchsergebnis (x, φ) eines Abstandes x und eines Winkels φ gilt damit: $0 \leq x \leq \frac{d}{2}$ (denn so groß kann der Abstand zur nächstgelegenen Geraden höchstens sein) und $0 \leq \varphi < \pi$. Alle möglichen Punkte (x, φ) liegen daher in dem eingezeichneten Rechteck G :



der günstige Bereich für das Nadelexperiment von Buffon

Die geworfene Nadel schneidet genau dann eine der gezeichneten Geraden, wenn die Bedingung

$$x \leq \frac{l}{2} \sin(\varphi)$$

erfüllt ist, wenn also der Punkt (x, φ) in A liegt (siehe obige Skizze). Die Fläche A hat den Inhalt

$$|A| = \int_0^\pi \frac{l}{2} \sin(\varphi) d\varphi = \frac{l}{2} (-\cos(\varphi)) \Big|_0^\pi = l.$$

Wird das Experiment so durchgeführt, dass kein Punkt (x, φ) des Rechtecks bevorzugt auftritt (dann liegt eine Gleichverteilung vor), so erhält man für die gesuchte Wahrscheinlichkeit:

$$p = \frac{|A|}{|G|} = \frac{l}{\pi \cdot \frac{d}{2}} = \frac{2l}{\pi d}.$$

Führt man dieses Experiment n -mal durch (n hinreichend groß), so gilt für die relative Häufigkeit $r_n(A)$ der Versuche, bei denen die Nadel eine Gerade schneidet, dass sie (wie immer, nach dem Gesetz der großen Zahlen) näherungsweise gleich der theoretischen Wahrscheinlichkeit ist:

$$r_n(A) \approx \frac{2l}{\pi d}.$$

Damit kann die Zahl π gemäß

$$\pi \approx \frac{2l}{dr_n(A)}$$

näherungsweise bestimmt werden. Auch in der Praxis wurde das Experiment real (also nicht mit einer Computersimulation, sondern mit echten Nadeln) schon häufiger durchgeführt. Hier die daraus von einigen Experimentalforschern erzielten Näherungswerte für π :

- 3.1596 (Wolf, im Jahre 1850, 5000 Würfe)
- 3.1553 (Smith, im Jahre 1855, 3204 Würfe)
- 3.1419 (Fox, im Jahre 1894, 1120 Würfe)
- 3.1415929 (Lazzarini, im Jahre 1901, 3408 Würfe)

Bemerkung:

Leider benötigt man für das Programm, wenn man es in VBA so programmiert wie oben beschrieben, selber einen Näherungswert für π , da man ständig den Sinus im Bogenmaß (explizit oder programmintern) berechnen muss. Man hat also eine Art „Teufelskreis“. Das ist natürlich sehr unschön. Vielleicht hat ja jemand von euch Lust, zu Hause eine Simulation zu programmieren, die ohne die Kenntnis von π auskommt...

Wir kennen ja aus dem Kurs „Erste stochastische Simulationen“ bereits eine andere Methode, mit Hilfe einer stochastischen Simulation die Kreiszahl π anzunähern. Wir wollen die beiden Methoden jetzt einmal miteinander vergleichen. Wichtig ist natürlich, dass der Mittelwert von sehr vielen Simulationsschritten tatsächlich in der Nähe von π liegt. Was aber noch viel wichtiger ist, ist die Frage, wie sehr das Verfahren „stret“, d.h. wie groß die Schwankungen sind. Als Maß dafür bietet sich die sogenannte Stichprobenvarianz an. Etwas ähnliches haben wir im Kurs zum Linearen Kongruenz-Generator beim χ^2 -Test bereits kennengelernt:

Gegeben seien Werte x_1, \dots, x_n mit dem Mittelwert

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Dann ist die **Stichprobenvarianz**

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]$$

ein Maß für die Streuung der Messreihe.

Warum teilt man durch $n-1$ und nicht durch n ? Das hat mathematische Gründe, kann aber auch ohne tieferes Wissen über sogenannte erwartungstreue Schätzer plausibel gemacht werden: Wegen

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = n \cdot \underbrace{\left(\frac{1}{n} \sum_{i=1}^n x_i - \bar{x} \right)}_{=0} = 0$$

ist die letzte Abweichung $x_n - \bar{x}$ bereits durch die ersten $n - 1$ Abweichungen festgelegt. Somit variieren nur $n - 1$ Abweichungen frei. Man sagt: Es gibt nur $n - 1$ sogenannte **Freiheitsgrade**.

Gruppenaufgabe

Führe beide stochastischen Simulationsmethoden zur Näherung der Kreiszahl π jetzt 100 mal mit jeweils 50.000 Simulationsschritten durch. Dadurch erhältst du in beiden Fällen 100 Näherungswerte x_1, x_2, \dots, x_{100} für π . Berechne jeweils den Mittelwert und die Stichprobenvarianz und vergleiche die beiden Verfahren. Wie interpretierst du das Ergebnis?

Bemerkung

Normalerweise werden Approximationen über π natürlich nicht über stochastische Simulationen durchgeführt, sondern etwa durch die Auswertung von Reihendarstellungen mit Hochleistungsrechnern. Es ist klar, dass stochastische Verfahren im Gegensatz dazu ineffizient und nur noch von historischem Interesse sind (insbesondere weil der Fehler der Schätzung auch mit steigender Simulationsanzahl nicht kontrolliert werden kann). Interessant ist hingegen die prinzipielle Approximationsmöglichkeit analytischer Probleme durch stochastische Simulationen, die seit Jahren etwa in der modernen stochastischen Finanzmathematik eine sehr wichtige Rolle spielt. Von daher lohnt es sich sehr wohl, über solche Phänomene anhand solch einfacher Beispiele nachzudenken.

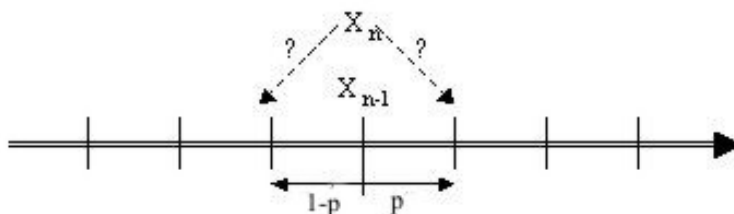
Kapitel 2

Irrfahrten (Random Walks)

Rekurrenz- und Transienzfragen

Wir wollen uns nun mit sogenannten **Irrfahrten** (auch in der deutschen Literatur eher unter dem englischen Ausdruck **Random Walks** bekannt) beschäftigen. Zunächst betrachten wir den einfachsten (eindimensionalen) Fall und bewegen uns dabei auf den ganzen Zahlen der reellen Achse.

Wir starten bei irgendeiner ganzen Zahl x . Von dort aus gehen wir mit Wahrscheinlichkeit p um eine ganze Einheit nach rechts und mit Wahrscheinlichkeit $1 - p$ um eine ganze Einheit nach links. Vom neuen Punkt aus beginnt dann das gleiche Spiel von vorn, d.h. wir gehen wieder mit Wahrscheinlichkeit p um eine ganze Einheit nach rechts und mit Wahrscheinlichkeit $1 - p$ um eine ganze Einheit nach links usw:



Irrfahrt im Eindimensionalen

Mit X_n bezeichnen wir unsere Position zur Zeit n , für $n = 0, 1, 2, \dots$. Demzufolge gilt, da wir mit Wahrscheinlichkeit 1, also ganz sicher, in x starten:

$$P(X_0 = x) = 1$$

sowie

$$P(X_n = X_{n-1} + 1) = p \quad \text{und} \quad P(X_n = X_{n-1} - 1) = 1 - p$$

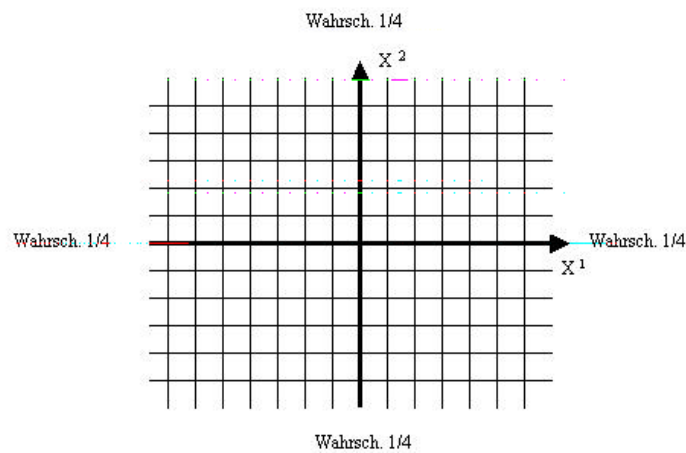
für alle $n \in \mathbb{N}$.

Im Falle $p = \frac{1}{2}$ spricht man auch von einer **symmetrischen Irrfahrt**. Im Folgenden werden wir zumeist solche symmetrischen Irrfahrten betrachten.

Das ganze Konzept können wir natürlich auf höherdimensionale Räume erweitern. Auf dem Gitter

$$\mathbb{Z} \times \mathbb{Z} = \{(x, y) : x, y \in \mathbb{Z}\}$$

in der reellen Ebene $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ sähe das dann im symmetrischen Fall so aus:



symmetrische Irrfahrt im Zweidimensionalen

Wir haben einen Prozess $X_n = \begin{pmatrix} X_n^1 \\ X_n^2 \end{pmatrix}$ mit

$$P\left(\begin{pmatrix} X_0^1 \\ X_0^2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = 1$$

und

$$\begin{aligned}P\left(\begin{pmatrix} X_n^1 \\ X_n^2 \end{pmatrix} = \begin{pmatrix} X_{n-1}^1 \\ X_{n-1}^2 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) &= \frac{1}{4}, \\P\left(\begin{pmatrix} X_n^1 \\ X_n^2 \end{pmatrix} = \begin{pmatrix} X_{n-1}^1 \\ X_{n-1}^2 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right) &= \frac{1}{4}, \\P\left(\begin{pmatrix} X_n^1 \\ X_n^2 \end{pmatrix} = \begin{pmatrix} X_{n-1}^1 \\ X_{n-1}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) &= \frac{1}{4}, \\P\left(\begin{pmatrix} X_n^1 \\ X_n^2 \end{pmatrix} = \begin{pmatrix} X_{n-1}^1 \\ X_{n-1}^2 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right) &= \frac{1}{4}.\end{aligned}$$

Wir gehen also, ausgehend von einem Punkt $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{Z} \times \mathbb{Z}$, mit jeweils einer Wahrscheinlichkeit von 25% zu allen vier benachbarten Gitterpunkten.

Entsprechend kann man auch die Bedingungen für eine dreidimensionale symmetrische Irrfahrt aufstellen (mit Wahrscheinlichkeit $\frac{1}{8}$ für jede Richtung).

Nun interessieren wir uns für die Wahrscheinlichkeit, dass wir uns, beginnend in $x = 0$, im Raum verlaufen oder (und damit zugleich präziser ausgedrückt) für die Wahrscheinlichkeit des Gegenereignisses, nämlich dass wir immer wieder (unendlich oft) zum Ausgangspunkt zurückkehren. Wie groß ist also

$$P(X_n = 0 \text{ für unendlich viele } n),$$

wenn wir in $X_0 = 0$ starten? Ändert sich das vielleicht mit zunehmender Dimension?

Gruppenaufgabe:

Was tippt ihr: Wie hoch ist die Wahrscheinlichkeit dafür, im ein-, zwei- und dreidimensionalen Gitter unendlich oft zum Ausgangspunkt zurückzukehren?

Wenn wir dieses Experiment nun simulieren wollen, stoßen wir natürlich an die Grenzen der Simulationstechnik. Schließlich können wir ein Expe-

riment, dessen „Erfolg“ von „Messungen“ an **unendlich vielen Zeitpunkten** abhängt, nicht durch **endlich viele Durchläufe eines Algorithmus** simulativ nachahmen. Daher können wir das Experiment nur mit Hilfe einer Simulation annähern und auf Grund des Ergebnisses der Simulation Vermutungen über die zu ermittelnden Wahrscheinlichkeiten anstellen, die dann anschließend mathematisch noch sauber nachzuweisen sind. In der Theorie der Wahrscheinlichkeitsrechnung stellt die Unendlichkeit ein wesentlich geringeres Problem dar, das man mathematisch gut in den Griff kriegt.

Zurück zu unserem Problem: Die Wahrscheinlichkeit in 1.000.000.000 möglichen Zeitpunkten mindestens 10 mal zum Nullpunkt zurückzukehren, sollte eine erste Annäherung an die Wahrscheinlichkeit dafür sein, in unendlich vielen Zeitpunkten unendlich oft zum Nullpunkt zurückzukehren.

Gruppenaufgabe:

Mache im fertigen Programm für die Dimensionen 1, 2 und 3 jeweils 10 Simulationsschritte des folgenden Experimentes:

Führe, beginnend im Nullpunkt, 1.000.000.000 Schritte der symmetrischen Irrfahrt durch und breche ab, sobald du 10 mal wieder am Nullpunkt angekommen bist. Bricht das Experiment vor dem 1.000.000.000ten Simulationsschritt ab, so nennen wir das Experiment „erfolgreich“ (wir haben uns nicht verlaufen und sind stattdessen „immer wieder“ (hier 10 mal) zum Ursprung zurückgekehrt). Berechne aus diesen 10 Simulationsschritten für alle drei Dimensionen die relative Häufigkeit des genannten Experiments. Gib nun Schätzungen für die tatsächlichen Wahrscheinlichkeiten in allen drei Dimensionen ab. (Im Dreidimensionalen dauert die Simulation eine halbe Ewigkeit. Breche sie ab, wenn dir zu langweilig wird. ;-))

Man sieht hier ganz deutlich die Grenze der Simulationsmethode: Derartige Probleme mit unendlichem Zeithorizont können mit Hilfe von Simulationen gar nicht oder nur höchst unbefriedigend bearbeitet werden!

Jetzt zur Theorie:

Man nennt einen Zustand **rekurrent**, wenn man ihn mit Sicherheit (also mit Wahrscheinlichkeit 1) unendlich oft wieder erreicht, nachdem man bei ihm

gestartet ist. Andernfalls heißt er **transient**. Witzigerweise gilt dann automatisch, dass die Wahrscheinlichkeit für das unendlich häufige Wiederkehren gleich 0 ist. Die Wahrscheinlichkeit kann also niemals dazwischen liegen, also etwa gleich $\frac{1}{2}$ sein. Dies ist eine Folge des sogenannten **0-1-Gesetzes**, nach dem unter bestimmten Voraussetzungen Ereignisse, deren Eintreffen von unendlich vielen Zeitpunkten abhängt, immer die Wahrscheinlichkeit 0 oder 1 haben. Dies kann man mathematisch präzisieren und auch nachweisen. Genauer kann man zudem zeigen:

Ein Zustand x ist genau dann rekurrent, wenn

$$\sum_{n=1}^{\infty} P(\text{man erreicht } x \text{ in genau } n \text{ Zeiteinheiten}) = +\infty$$

gilt. Dies bedeutet intuitiv: Für große n ist

$$P(\text{man erreicht } x \text{ in genau } n \text{ Zeiteinheiten})$$

immer noch so groß, dass die folgende Reihe nichtnegativer Glieder

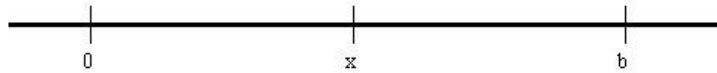
$$\begin{aligned} & P(\text{man erreicht } x \text{ in genau } 1 \text{ Zeiteinheiten}) \\ + & P(\text{man erreicht } x \text{ in genau } 2 \text{ Zeiteinheiten}) \\ + & P(\text{man erreicht } x \text{ in genau } 3 \text{ Zeiteinheiten}) \\ + & \dots \\ & \vdots \end{aligned}$$

über alle Schranken wächst, also gleich $+\infty$ ist. Man kann nun zeigen, dass diese Reihe im ein- und zweidimensionalen Fall divergiert (also gleich $+\infty$ ist) und im drei- und höherdimensionalen Fall konvergiert. Daher kehrt man im ein- und zweidimensionalen Fall mit Sicherheit immer wieder zum Ausgangspunkt zurück, während man sich im drei- und höherdimensionalen Fall mit ebenso großer Sicherheit irgendwann für alle Zeiten verläuft. Offenbar hängt diese Wahrscheinlichkeit nicht vom Ausgangspunkt ab, von dem aus man startet.

Die ein- und zweidimensionale Irrfahrt sind also rekurrent, während alle höherdimensionalen Irrfahrten transient sind.

Ruinwahrscheinlichkeiten

Wir versehen nun im eindimensionalen Fall die Irrfahrt mit sogenannten **Adsorptionspunkten**, also Punkten, aus denen man „nicht mehr herauskommt“, wenn man sie einmal erreicht hat. Ein solcher Punkt kann etwa bei einem Glücksspiel der Kontostand sein, bei dem man ruiniert ist (das sei bei uns der Punkt $x = 0$) oder, im günstigen Fall, der Kontostand $x = b$, bei dem man seinen vorher angestrebten Gewinn b erzielt hat und aufhört. Wir wollen nun, zunächst wieder simulativ, zu Aussagen über die **Ruinwahrscheinlichkeit** kommen, also die Wahrscheinlichkeit, dass man den Punkt $x = 0$ erreicht, bevor man den Punkt $x = b$ erreicht hat:



Skizze zur Erläuterung der Ruinwahrscheinlichkeit

Genauer: Der Prozess $(X_n)_{n \in \mathbb{N}}$ messe den Kontostand eines Spielers zum Zeitpunkt n . Zu jedem Zeitpunkt n wird ein Spiel durchgeführt, das der Spieler mit Wahrscheinlichkeit p gewinnt und mit Wahrscheinlichkeit $1 - p$ verliert. Im Falle des Gewinns wächst der Kontostand um einen Euro, im Falle des Verlustes reduziert er sich um den gleichen Betrag. Es handelt sich bei $(X_n)_{n \in \mathbb{N}}$ um eine Irrfahrt mit der Erfolgswahrscheinlichkeit p . Wir starten diese Irrfahrt mit $x > 0$ Euro und geben uns ein Gewinnziel b Euro mit $b > x$ vor. Es sei T_0 der Zeitpunkt des erstens Eintreffens in $x = 0$ (also des Ruins) und T_b der Zeitpunkt des ersten Eintreffens in $x = b$ (also des Erreichens des Gewinnziels). Man nennt T_0 und T_b auch **Stoppzeiten**. Jetzt interessieren wir uns, wie gesagt, für die Ruinwahrscheinlichkeit, also für die Wahrscheinlichkeit

$$P(T_0 < T_b).$$

Aufgabe 2:

Versuche im symmetrischen Fall (also dem Fall $p = \frac{1}{2}$) die Ruinwahrscheinlichkeit für verschiedene Startwerte $x > 0$ mit Gewinnziel $b = 100$ simulativ zu bestimmen. Hast du eine Vermutung für eine allgemeine Formel? Mache das Gleiche auch noch für verschiedene Einzelgewinnwahrscheinlichkeiten p (mit $p > \frac{1}{2}$ und $p < \frac{1}{2}$) und trage speziell für den Startwert $x = 50$ die simulativ bestimmten Ruinwahrscheinlichkeiten für verschiedene Erfolgswahrscheinlichkeiten p in eine Tabelle ein.

Bemerkung: Im Falle $p \neq \frac{1}{2}$ wirst du es allerdings mit an Sicherheit grenzender Wahrscheinlichkeit nicht schaffen, eine explizite Formel für die Ruinwahrscheinlichkeit zu finden.

Jetzt wieder zur Theorie:

Es sei für $x \in \{0, 1, \dots, b\}$ p_x die Ruinwahrscheinlichkeit bei Start in x , also:

$$p_x = P(T_0 < T_b | X_0 = x).$$

Dann gilt für alle $x \in \{1, 2, \dots, b-1\}$

$$p_x = p \cdot p_{x+1} + (1-p) \cdot p_{x-1}$$

(denn ich komme mit Wahrscheinlichkeit p in den Zustand $x+1$ und dort ist die Ruinwahrscheinlichkeit dann p_{x+1} und ich komme mit Wahrscheinlichkeit $p-1$ in den Zustand $x-1$ und dort ist die Ruinwahrscheinlichkeit dann p_{x-1}) sowie:

$$p_0 = 1 \quad \text{und} \quad p_b = 0.$$

(Denn beginne ich (hypothetisch) mit 0 Euro, bin ich sofort ruiniert; beginne ich (ebenso hypothetisch) mit meinem Gewinnziel, höre ich sofort auf und kann nicht ruiniert werden).

Dieses Gleichungssystem besitzt für $p = \frac{1}{2}$ eine eindeutige Lösung, und zwar eine von sehr einfacher Gestalt:

$$p_x = P(T = T_0 | X_0 = x) = 1 - \frac{x}{b}.$$

Die Wahrscheinlichkeit in den Ruin zu gelangen, ist also im symmetrischen Fall $1 - \frac{x}{b}$, wenn man mit x Euro startet, d.h. linear in x .

Im nicht-symmetrischen Fall (für $p \neq \frac{1}{2}$) erhält man die (ebenfalls eindeutige) Lösung

$$p_x = \frac{\left(\frac{1-p}{p}\right)^b - \left(\frac{1-p}{p}\right)^x}{\left(\frac{1-p}{p}\right)^b - 1}.$$

Zusatzaufgabe:

Vergleiche deine Simulationsergebnisse aus Aufgabe 2 nun mit den theoretischen Ruinwahrscheinlichkeiten. Wie stark weichst du ab?

Kapitel 3

Würfelspiele

Mini-Kniffel

Wir wollen nun eine Mini-Version des bekannten Würfelspieles „Kniffel“ mit zudem stark vereinfachten Regeln betrachten und dabei drei unterschiedliche Spielstrategien vergleichen. Mit Hilfe einer stochastischen Simulation versuchen wir herauszufinden, welche Strategie die beste ist, d.h. mit welcher Strategie man langfristig die meisten Spiele gewinnt.

Zunächst ein paar Worte zu den

Regeln:

Gewürfelt wird gleichzeitig mit drei vierseitigen Würfeln (also Tetraedern), die jeweils die Augenzahlen „1“, „2“, „3“ und „4“ haben. Nun kann man in einem zweiten Versuch mit einer beliebigen Anzahl der drei bereits geworfenen Würfel noch einmal würfeln und den Rest der Würfel stehenlassen.

Danach wird ausgewertet, wie viele Punkte man erzielt hat, und zwar gemäß der folgenden Punkteverteilung:

1-er-Paar	:	1 Punkt
2-er-Paar	:	2 Punkte
3-er-Paar	:	3 Punkte

4-er-Paar	:	4 Punkte
Straße	:	5 Punkte
1-er-Dreierpasch	:	6 Punkte
2-er-Dreierpasch	:	7 Punkte
3-er-Dreierpasch	:	8 Punkte
4-er-Dreierpasch	:	9 Punkte .

Hierbei ist eine „Straße“ eine der beiden Würfelkombinationen $(1, 2, 3)$ oder $(2, 3, 4)$.

In der Praxis lassen wir nun drei virtuelle Spieler mit unterschiedlichen Spielstrategien gegeneinander antreten. Jeder versucht in 10 aufeinanderfolgenden Spielrunden unter Einhaltung seiner Strategie sein Glück, und am Schluss werden die Punkte zusammengezählt. Wer in den 10 Spielen die meisten Punkte erzielt hat, hat das Spiel gewonnen.

Nun zu den drei Strategien:

1. Strategie (Priorität: hohe Zahlen)

Würfelt man beim ersten Wurf einen Dreierpasch, so hört man natürlich auf. Andernfalls lässt man (auch wenn man eine Straße gewürfelt hat) den/die Würfel mit der höchsten Augenzahl stehen und würfelt den/die anderen Würfel ein zweites Mal.

Beispiele:

$(2, 2, 2)$ → alles stehenlassen

$(2, 3, 4)$ → 4 stehenlassen; 2, 3 neu würfeln

$(1, 3, 3)$ → 3, 3 stehenlassen; 1 neu würfeln

$(1, 1, 3)$ → 3 stehenlassen; 1, 1 neu würfeln

2. Strategie (Priorität: Straße)

Würfelt man beim ersten Wurf einen Dreierpasch, so hört man natürlich auf. Das Gleiche gilt für das Erreichen einer Straße. Bei allen anderen Konstellationen setzt man (auch wenn man im ersten Versuch ein Paar gewürfelt hat) bedingungslos auf eine Straße, wenn man diese durch das Werfen eines einzigen Würfels erreichen kann. Nur wenn eine Straße „sehr unwahrscheinlich“ ist, d.h. wenn man für das Erreichen zwei Würfel neu werfen müsste, setzt man auf einen Dreierpasch. Gibt es zwei Würfel mit unterschiedlicher Augenzahl, durch deren erneutes Werfen man eine Straße erzielen könnte, so wähle man den Würfel mit der niedrigeren Augenzahl.

Beispiele:

$(2, 2, 2) \rightarrow$ alles stehenlassen

$(1, 2, 3) \rightarrow$ alles stehenlassen

$(1, 2, 4) \rightarrow$ 2, 4 stehenlassen; 1 neu würfeln

$(2, 2, 4) \rightarrow$ 2, 4 stehenlassen; 2 neu würfeln

$(1, 1, 4) \rightarrow$ 1, 1 stehenlassen; 4 neu würfeln

3. Strategie (Priorität: Dreierpasch)

Würfelt man beim ersten Wurf einen Dreierpasch, so hört man natürlich auf. In allen anderen Fällen lässt man, falls vorhanden, ein Paar stehen und würfelt den dritten Würfel neu. Hat man aber kein Paar im ersten Wurf erreicht, so lässt man (auch dann, wenn man im ersten Versuch eine Straße gewürfelt hat!) den Würfel mit der höchsten Augenzahl stehen und wirft die beiden anderen Würfel neu.

Beispiele:

$(2, 2, 2) \rightarrow$ alles stehenlassen

$(1, 2, 4) \rightarrow$ 4 stehenlassen; 1, 2 neu würfeln

$(1, 2, 3) \rightarrow$ 3 stehenlassen; 1, 2 neu würfeln

$(2, 2, 4) \rightarrow$ 2, 2 stehenlassen; 4 neu würfeln

Gruppenaufgabe:

Vergleiche die drei Strategien mit Hilfe des fertigen Simulationsprogramms. Das Vorgehen ist dabei wie folgt: In jedem Simulationsschritt werden 10 Spielrunden durchgeführt. Es wird also 10 mal (für alle drei Spieler zugleich, nicht für alle einzeln) gewürfelt und dann anhand der unterschiedlichen Strategien weitergewürfelt. Die Punkte aus diesen 10 Spielrunden werden in einem temporären Konto für jeden Spieler zusammengezählt. Der (alleinige) Sieger der 10 Spielrunden bekommt 1 Punkt auf ein gesondertes dauerhaftes Konto gutgeschrieben. Sollte es mehrere punktgleiche Sieger geben, so bekommt jeder Sieger 0.5 Punkte auf dieses dauerhafte Konto gutgeschrieben. Danach beginnt der neue Simulationsschritt. Das temporäre Konto wird wieder auf 0 Punkte zurückgesetzt, während das dauerhafte Konto in jedem Simulationsschritt weiter hochgezählt wird.

Wir führen nun 10.000 Simulationsschritte durch und schauen, welcher Spieler (d.h. welche Strategie) insgesamt die meisten Punkte erzielt hat, also den höchsten Stand auf dem dauerhaften Konto hat.

Zusatzaufgabe:

Versuche nun, die Simulation selber zu programmieren. Überlege dir zudem eine eigene Strategie (vielleicht eine Kombination aus den drei vorhandenen Strategien) und trete damit gegen die drei virtuellen Spieler als vierter virtueller Spieler an.

Mini-Monopoly

Wir wollen uns nun einer Mini-Version des bekannten Monopoly-Spieles zuwenden und interessieren uns dabei dafür, welchen Gewinn oder Verlust wir langfristig zu erwarten haben (das Vorhandensein von negativem Kapital, also Schulden, sei erlaubt). Das Spielfeld sieht hierbei wie folgt aus:

Nur zu Besuch	3 Gefängnis	4 SimuLab- Straße (Miete 200 €)	5 Frei Parken
	2 caesar- Straße (Miete 100 €)	Mini- Monopoly	6 Bonnplatz (Miete 300 €)
	1 Los	8 Kennedyallee (Miete 400 €)	7 Gehe in das Gefängnis

Mini-Monopoly

Man startet auf dem Feld 1 („Los“) mit einem Startkapital von 2000 Euro. Gewürfelt wird mit einem normalen Würfel. Die auf den Straßen aufgedruckten Mietpreise sind beim Erreichen des jeweiligen Feldes an die „Bank“ zu zahlen. Beim Erreichen des Losfeldes (und nicht etwa beim Überschreiten, im Gegensatz zum richtigen Monopoly!) erhält man 2000 Euro. Kommt man auf das Feld „Gehe in das Gefängnis“, so muss man unmittelbar auf Feld 3 rücken. Dort hat man drei Versuche eine „6“ zu würfeln. Gelingt einem dies, so darf man erneut würfeln und mit der gewünschten Augenzahl ohne Zahlung eines Strafbetrages weiterücken. Würfelt man hingegen dreimal

hintereinander keine „6“, so muss man 1000 Euro Strafe zahlen, rückt auf das Feld „Nur zu Besuch“ und würfelt anschließend normal weiter. Beim Betreten der Felder „Nur zu Besuch“ und „Frei Parken“ passiert gar nichts.

Gruppenaufgabe:

Was tippst du: Wird man im Laufe des Spiels eher Geld gewinnen oder verlieren? Wie hoch erwartest du den Kontostand nach 100.000-maligem Würfeln?

Zur Erinnerung: Nimmt eine Zufallsvariable X die Werte x_1, \dots, x_m an, so gilt für ihren Erwartungswert:

$$E[X] = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots + x_m \cdot P(X = x_m).$$

Es werde nun für alle $n \in \mathbb{N}$ mit X_n der Kontostand nach n -maligem Würfeln bezeichnet. Dann gilt für den Erwartungswert von X_1 (ohne Beachtung der möglichen Strafzahlung im Gefängnis beim Würfeln einer 6):

$$\begin{aligned} E[X_1] &= 1900 \cdot \frac{1}{6} + 2000 \cdot \frac{1}{6} + 1800 \cdot \frac{1}{6} + 2000 \cdot \frac{1}{6} + 1700 \cdot \frac{1}{6} + 2000 \cdot \frac{1}{6} \\ &= 1900, \end{aligned}$$

d.h. man verliert im Durchschnitt 100 Euro nach dem ersten Würfeln (man war ja mit 2000 Euro gestartet).

Zusatzaufgabe:

Wie lautet der exakte Erwartungswert von X_1 , wenn man die mögliche Strafzahlung im Gefängnis einberechnet?

Tipp: Die Wahrscheinlichkeit, in drei Versuchen keine 6 zu werden, beträgt $\left(\frac{5}{6}\right)^3$.

Wie du dir vorstellen kannst, wird es **sehr schwierig** nun den Erwartungswert etwa von $X_{100.000}$ in ähnlicher Weise zu berechnen. Daher berechnen wir

ihn doch einfach näherungsweise durch eine stochastische Simulation!

Gruppenaufgabe:

Bestimme mit Hilfe des Simulationsprogramms den Erwartungswert von $X_{100.000}$ näherungsweise. Führe hierzu 1000 Simulationsschritte durch. Entspricht das Ergebnis deinen Erwartungen? Ist es wenigstens im Nachhinein plausibel?

Zusatzaufgabe:

Versuche die Simulation nun selber zu programmieren.

(TIPP. Wenn du zum Beispiel auf Feld „7“ stehst und eine „5“ würfelst, so bist du bei fortlaufender Zählung „eigentlich“ auf Feld „12“. Dieses Feld gibt es aber nicht. Du erhältst aber, wenn du „modulo 8“ rechnest, das richtige Feld. Teile hierzu einfach immer das aktuelle Feld durch 8 und betrachte den Rest, der bei dieser Division entsteht (hier 4). Wenn m die gewürfelte Augenzahl ist, so bekommst du die neue Position im VBA-Code wie folgt:

```
Position = (Position + m) Mod 8 .
```

Hierbei gibt es nur ein Problem: Stehst du nach dem Würfeln auf der Position 8, so wird durch diese Operation deine Position zur Position „0“, die es nicht gibt. Also muss du noch Folgendes einfügen:

```
If Position = 0 Then  
Position = Position + 8  
Endif ,
```

um diesen Fehler im genannten Spezialfall wieder gutzumachen.)

Kapitel 4

Ein Lotterie-Problem: Ziehen ohne Zurücklegen (hypergeometrische Verteilung)

Zur Erinnerung: Mit Hilfe des Galton-Brettes hatten wir im Kurs „Erste Stochastische Simulationen“ spielerisch die Binomialverteilung kennengelernt: Ist p die Wahrscheinlichkeit, dass die Kugel nach rechts fällt, so gilt für die Wahrscheinlichkeit, dass die Kugel im $(k + 1)$ -ten Feld von links ankommt:

$$P(\text{„}(12 - k) \text{ mal nach links, } k \text{ mal nach rechts“}) = \binom{12}{k} \cdot p^k \cdot (1 - p)^{12-k}.$$

Hierbei war

$$\binom{12}{k} = \frac{12!}{k! \cdot (12 - k)!} = \frac{12 \cdot 11 \cdot \dots \cdot (12 - k + 1)}{k \cdot (k - 1) \cdot \dots \cdot 1}$$

ein Binomialkoeffizient (sprich: „12 über k “) und stellte die Anzahl der Möglichkeiten dar, aus einer 12-elementigen Menge (hier die 12 Positionen, an denen sich „die Kugel entscheiden muss, ob sie nach rechts oder links rollt“) k Elemente auszuwählen (nämlich diejenigen Positionen, an denen die Kugel nach rechts fällt).

Es handelt sich hierbei um ein klassisches „Ziehen mit Zurücklegen“-Experiment: Wir ziehen 12 mal aus einer Urne, die 2 Schilder enthält: Auf einem der beiden Schilder steht „rechts“, auf dem anderen „links“. Nach jeder Ziehung legen wir das gezogene Schild zurück und ziehen erneut. Nach 12 solcher Ziehungen zählen wir, wie oft wir das Schild mit dem Schriftzug „rechts“ (und

wie oft wir entsprechend das Schild mit dem Schriftzug „links“ gezogen haben. Die Reihenfolge spielt hierbei keine Rolle. Haben wir (egal wann!) k -mal das Schild mit dem Schriftzug „rechts“ (und entsprechend $(12 - k)$ -mal das Schild mit dem Schriftzug „links“) gezogen, so landet die Kugel am Ende immer auf dem $(k + 1)$ -ten Feld von links.

Nun wollen wir durch ein ähnliches spielerisches Experiment eine neue Verteilung kennenlernen, nämlich die sogenannte **hypergeometrische Verteilung**, die analog die Wahrscheinlichkeit beim **Ziehen ohne Zurücklegen** beschreibt.

Doch zunächst wollen wir mal etwas spielen, bevor wir uns der trockenen Mathematik zuwenden:

Lotterie mit 100 Losen

Bei einer Lotterie werden 100 Lose angeboten. Als ersten Preis gibt es einen Urlaub, als zweiten ein Herrenfahrrad, als dritten ein Damenfahrrad sowie als Trostpreise 20 Flaschen Most, 10 Gläser Honig, 30 Äpfel und 25 Nüsse. Jemand kauft 3 Lose (als erster).

Wie groß ist die Wahrscheinlichkeit ...

- (a) ... alle 3 Haupttreffer zu gewinnen?
- (b) ... keinen Haupttreffer zu gewinnen?
- (c) ... mindestens einen Haupttreffer zu gewinnen?
- (d) ... das Herrenfahrrad zu gewinnen?
- (e) ... nichts zu gewinnen?
- (f) ... 3 Trostpreise zu gewinnen?
- (g) ... 1 Most, 1 Honig und 1 Apfel zu gewinnen?
- (h) ... 3 Äpfel oder 3 Nüsse zu gewinnen?
- (i) ... nur den Urlaub zu gewinnen ?

Aufgabe 3:

Schreibe ein Simulationsprogramm, mit dem du die gesuchten Wahrscheinlichkeiten näherungsweise bestimmst (etwa mit 10.000 Simulationsschritten). Ordne der Zahl 1 den Urlaub zu, der Zahl 2 das Herrenfahrrad, der Zahl 3 das Damenfahrrad, etc. Erzeuge dann pro Simulationsschritt drei Zufallszahlen zwischen 1 und 100. Beachte bitte, dass in jedem Simulationsschritt keine Zahl mehr als einmal gezogen werden darf (da die Lose ja nicht zurückgelegt werden und man so zum Beispiel nicht zwei Herrenfahrräder gewinnen kann). Überlege dir also zunächst einmal, wie du Doppelziehungen am besten vermeidest. Am einfachsten könnte man es so machen:

```
zufallszahl1 = Int(Rnd() * 100) + 1
zufallszahl2 = zufallszahl1
Do While (zufallszahl2 = zufallszahl1)
zufallszahl2 = Int(Rnd() * 100) + 1
Loop
: ,
```

aber vielleicht fallen dir ja auch andere (bessere) Möglichkeiten ein.

Nun wollen wir uns klarmachen, wie man Wahrscheinlichkeiten von solchen Ereignissen berechnet, die aus einem „Ziehen ohne Zurücklegen“ resultieren:

Wir wissen ja bereits:

Aus n verschiedenen Elementen kann man ohne Berücksichtigung der Reihenfolge und ohne Zurücklegen k Stück auf

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k \cdot (k-1) \cdot \dots \cdot 1}$$

verschiedene Weisen auswählen, wobei $1 \leq k \leq n$ vorausgesetzt ist und $0! := 1$ gesetzt wird.

Nun kommen wir zur hypergeometrischen Verteilung, die die Wahrscheinlichkeiten beim Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge modelliert:

Eine Urne enthalte N Kugeln, darunter M schwarze und $N - M$ weiße. Es werden ohne Zurücklegen n Kugeln mit einem Griff (oder: ohne Beachtung der Reihenfolge) gezogen, mit $1 \leq n \leq N$. Die Wahrscheinlichkeit, dass darunter genau k schwarze und $n - k$ weiße Kugeln sind, beträgt:

$$p_k = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Beweis:

Es gibt $\binom{N}{n}$ mögliche Fälle, aus N Kugeln n Stück ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge zu ziehen. Aus den M schwarzen Kugeln lassen sich k Stück auf $\binom{M}{k}$ verschiedene Arten auswählen. Zu jeder festen Auswahl der k schwarzen Kugeln gibt es $\binom{N-M}{n-k}$ Möglichkeiten, die restlichen $n - k$ weißen Kugeln aus den $N - M$ weißen Kugeln auszuwählen. Es gibt also insgesamt $\binom{M}{k} \cdot \binom{N-M}{n-k}$ günstige Fällen, bei denen sich unter den n gezogenen Kugeln genau k schwarze und $n - k$ weiße Kugeln befinden. Daraus folgt die Behauptung.

Wir möchten nun mit Hilfe der hypergeometrischen Verteilung bzw. dazu analogen Überlegungen die Wahrscheinlichkeiten für die oben genannten Ereignisse berechnen:

zu (a): Aus den 3 Haupttreffern sind 3 Elemente und aus den übrigen 97 Nicht-Haupttreffern 0 Elemente zu ziehen. Daher gilt:

$$p_a = \frac{\binom{3}{3} \cdot \binom{97}{0}}{\binom{100}{3}} = \frac{1}{161700} \approx 6.184 \cdot 10^{-6}.$$

zu (b): Aus den 97 Nicht-Haupttreffern sind 3 Elemente und aus den 3 Haupttreffern 0 Elemente zu ziehen. Daher gilt:

$$p_b = \frac{\binom{97}{3} \cdot \binom{3}{0}}{\binom{100}{3}} \approx 0.9118.$$

zu (c): Offenbar ist dieses Ereignis das Gegenereignis des Ereignisses von Aufgabenteil (b). Daher gilt:

$$p_c = 1 - p_b \approx 0.0882.$$

zu (d): Aus dem 1 Herrenfahrrad ist 1 Element und aus den 99 Nicht-Herrenfahrrädern sind 2 Elemente zu ziehen. Daher gilt:

$$p_d = \frac{\binom{1}{1} \cdot \binom{99}{2}}{\binom{100}{3}} = 0.03.$$

Gruppenaufgabe:

Versuche die übrigen Aufgaben (e)-(i) nun selber zu lösen. Vergleiche dabei zur Kontrolle deiner Rechnungen die exakten Ergebnisse mit den Näherungswerten aus deiner Simulation.

Hier zu Kontrolle noch die Ergebnisse:

$$p_e \approx 0.0014 ,$$

$$p_f \approx 0.6108 ,$$

$$p_g \approx 0.0371 ,$$

$$p_h \approx 0.0393 ,$$

$$p_i \approx 0.00004 .$$