Manfred Borovcnik und Erich Neuwirth
Klagenfurt und Wien

Rekursive Zugänge zu Wahrscheinlichkeitsproblemen und ihr Potential zur Modellbildung

Zusammenfassung:

Rekursive Zugänge zu Wahrscheinlichkeitsproblemen – so genannte Irrfahrten oder Markoff-Ketten – bieten die Möglichkeit, lokal (einsichtig) zu modellieren. Die globalen funktionalen Zusammenhänge sind bei derartigen Modellen meist nicht unmittelbar zu sehen. Das trifft etwa auf die Herleitung der Binomialverteilung, oder auf Wartezeitprobleme beim Sammeln der Elemente einer Serie zu. Die Schwierigkeiten können aber umgangen werden: Man lässt etwa eine Tabellenkalkulation die Rekursion einfach durchrechnen.

Jetzt könnte man einwenden, dass dabei ein wichtiges Element mathematischer Modellierung verloren geht, nämlich zu studieren, wie Parameter der Lösung (etwa der Erwartungswert) von den Eingangsparametern abhängen. In vielen Fällen kann man auch Parameter unmittelbar in das computerbasierte Rechenmodell einbauen, es gibt aber noch einen weiteren Weg: Man kann die funktionale oder statistische Abhängigkeit durch Simulation darstellen und studieren. Man kann ferner die lokale Rechnung für einen Wert eines Parameters variabel gestalten. Durch gezielte Veränderung dieses Parameters und gleichzeitige Berechnung der Zielgrößen kann man dynamisch dessen Einfluss analysieren und graphisch darstellen.

1. Einleitung

Das Problem der vollständigen Serie

Die Wahrscheinlichkeitsrechnung ist u. a. deswegen so erfolgreich, weil nur wenige Typen von (diskreten und stetigen) Verteilungen sehr viele Anwendungssituationen angemessen modellieren lassen. Darüber darf man nicht vergessen, dass in vielen weiteren Situationen die Verteilungen der zugrunde liegenden Zufallsvariablen nicht mit den Standardmodellen zu erfassen sind und daher – in der Situation – abgeleitet werden müssen. Das würde natürlich den Unterricht überfrachten: Es geht um mitunter relativ einfach formulierbare Probleme, deren kombinatorische Lösung jedoch außer Reichweite liegen kann oder zumindest zu viel an Zeit erfordert.

Ein beliebtes Beispiel ist das so genannte Sammelbildproblem: Ein Sammelalbum hat Platz für N Bilder. Man sammelt diese durch wiederholte Auswahl (Kauf etc.). Dabei modelliert man die Auswahl durch eine (diskrete) Gleichverteilung für alle Bilder; die einzelnen Auswahlen werden als (stochastisch) unabhängig voneinander „gedacht". Dabei ergeben sich mehrere Fragen:

- Wie viele „Käufe" sind erforderlich, bis man das Album voll hat? – Problem der vollständigen Serie.
- Wie viele – verschiedene – Bilder hat man im Durchschnitt, wenn man n Bilder kauft?

Im Unterricht umgeht man üblicherweise die Schwierigkeiten mit der Kombinatorik, indem man die Bedingungen realisiert und Daten durch Simulation beschafft. Aus der deskriptiven Analyse dieser Daten erhält man dann Schätzungen für die gefragten Größen. Das ist insoweit unbefriedigend, weil man die Lösung als vollständige Umgehung einer genaueren Analyse des Problems verstehen kann. Nirgends wird aus dem Lösungsansatz klar, wie die Lösung das Problem erfasst. Außerdem werden ja nur Schätzungen angebbar, die dann von Simulation zu Simulation schwanken. Ein anderer Ansatz besteht in der rekursiven Umformulierung der Aufgabe.

Bei rekursiven Ansätzen wird manchmal kritisiert, dass aus ihnen nicht klar hervorgeht, wie sie das gestellte Problem lösen. Die rekursive Sichtweise bietet aber eine deutliche Struktur für das Problem und damit für die in seine Modellierung gesteckten Annahmen. Sie stellt ein effizientes Werkzeug für die Lösung vielfältiger Aufgaben dar, das erstreckt sich bis hin zu Aufgaben zu Markoff-Ketten. Ein weiterer Vorteil des hier vorgestellten Ansatzes liegt darin, dass die vielen Vorteile, die eine Tabellenkalkulation wie EXCEL bietet, ausgenützt werden, und dass die Tabelle selbst die rekursive Struktur (z. B. Bezug auf die vorhergehende Zeile) klar ersichtlich macht.

Rekursive Algorithmen: Die Türme von Hanoi

Wie viele Züge y_n braucht man um den Stapel von n Ringen von einem Stab auf einen anderen hinüberzubringen, wenn man folgende Regeln einhalten muss:

> „Das Spiel besteht aus drei Stäben A, B und C, auf die mehrere gelochte Scheiben gelegt werden, alle verschieden groß. Zu Beginn liegen alle Scheiben auf Stab A, der Größe nach geordnet, mit der größten Scheibe unten und der kleinsten oben. Ziel des Spiels ist es, den kompletten Scheiben-Stapel von A nach C zu versetzen."

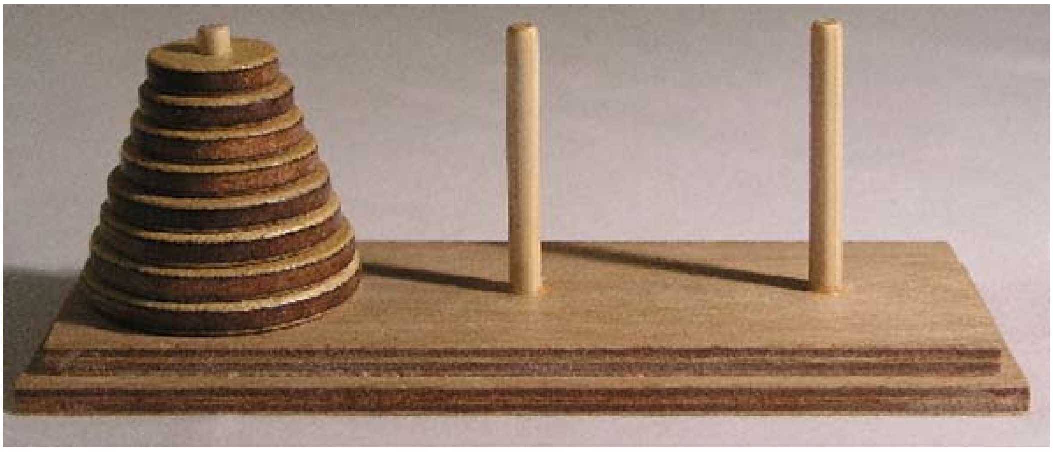

Abb. 1a: Die Türme von Hanoi (Wikipedia)

> „Bei jedem Zug darf die oberste Scheibe eines beliebigen Stabes auf einen der beiden anderen Stäbe gelegt werden, vorausgesetzt, dort liegt nicht schon eine kleinere Scheibe. Folglich sind zu jedem Zeitpunkt des Spieles die Scheiben auf jedem Feld der Größe nach geordnet.“ (aus: Wikipedia).

Wenn Sie das Problem schon kennen, mussten Sie vielleicht intuitive Hemmschwellen überwinden, weil Sie dem rekursiven Ansatz

$$y_n = 2 \cdot y_{n-1} + 1$$

zur Bestimmung der Anzahl der nötigen Schritte y_n nichts abgewinnen konnten.

Dabei ist es ganz einfach. Wenn man beispielsweise weiß, dass man einen Turm der Höhe 3 von einem Stab auf einen anderen transferieren kann, dann kann man einen Turm der Höhe 4 von A nach C transferieren, indem man die oberen 3 Scheiben von A nach B transferiert, dann die vierte Scheibe von A nach C transferiert und danach die 3 kleineren Scheiben von B nach C transferiert. Die Zahl der Züge, die wir für den Viererturm benötigen, ist also $2 \cdot x$, wobei x die Zahl der Züge für den Dreierturm ist, plus 1 Zug für die größte Scheibe.

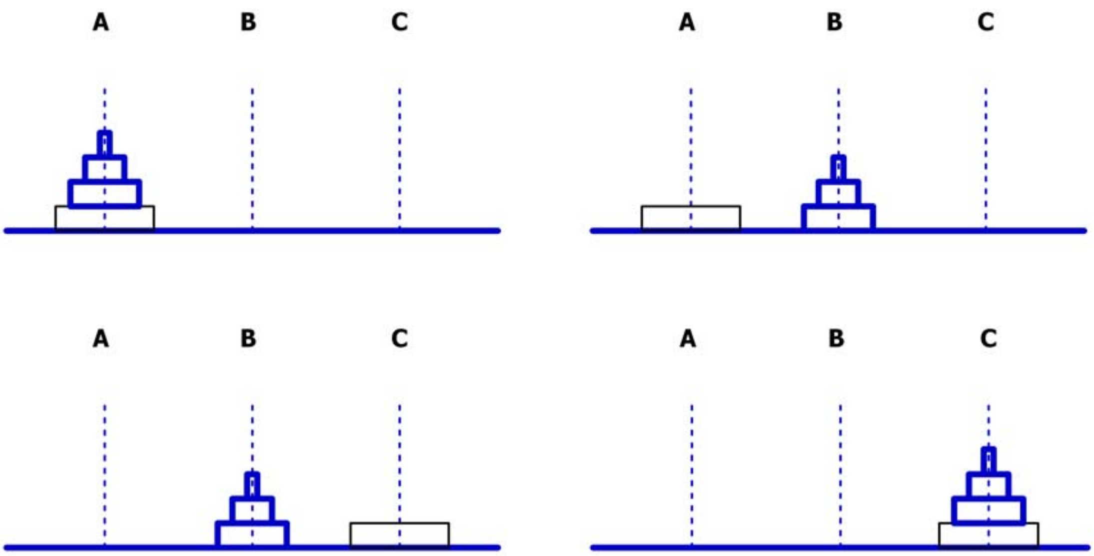

Abb. 1b: Der Algorithmus mit 4 Bausteinen

Das Rezept gilt allgemein: Die Zugzahl für eine Scheibe ist 1, und die Zugzahl für „eine Scheibe mehr“ beträgt $2 \cdot x + 1$, wobei x die Ausgangszugzahl ist.

Damit kann man ausgehend von $y_1 = 1$ alle Werte y_n berechnen. Man kennt damit zwar noch keine optimale Strategie, man weiß aber bereits, dass es mit den so berechneten Zugzahlen auf jeden Fall eine Lösung gibt. Es braucht allerdings oft mehr als „eine Schrecksekunde“, bis man einsieht, dass man – nachdem man die Struktur der Abhängigkeit erkannt hat – die Lösung durch primitives Einsetzen von Anfang her bekommt.

Bildungsziele

Einer der Autoren hat unterrichtliche Aufgaben in der Stochastik durch ein Wechselspiel zwischen Intuitionen und Mathematik zu „entwickeln“ versucht (Borovcnik 1992). Eine intuitive Vorstellung wird in Problemstellungen aufgegriffen; durch mathematische Tätigkeiten werden einerseits Vorformen der späteren mathematischen Begriffe aufgebaut, andererseits werden die Vorstellungen verändert. Großes Augenmerk ist dabei auf ein wechselweise entstehendes „Spiel“ zwischen Intuitionen und den Zwischenstufen der mathematischen Begriffe zu legen. Diese „Rückkoppelung“ soll nach und nach die „fundamentalen“ Ideen der Stochastik erschließen.

Mit der Verbreitung von Tabellenkalkulationen wie EXCEL gibt es eine Reihe leicht durchzuführender Tätigkeiten, die zum individuellen Begriffserwerb beitragen können. In Zeiten von Bildungsstandards beschreibt man die Unterrichtsziele nicht mehr mit Begriffen sondern mit Bildungszielen, die man prägnant mit Leitideen – einer Abwandlung von fundamentalen Ideen – beschreibt. Die im Folgenden vorgeschlagenen Tätigkeiten tragen zu den Leitideen des Unterrichts wesentlich bei, wie die folgende Abbildung zeigen mag.

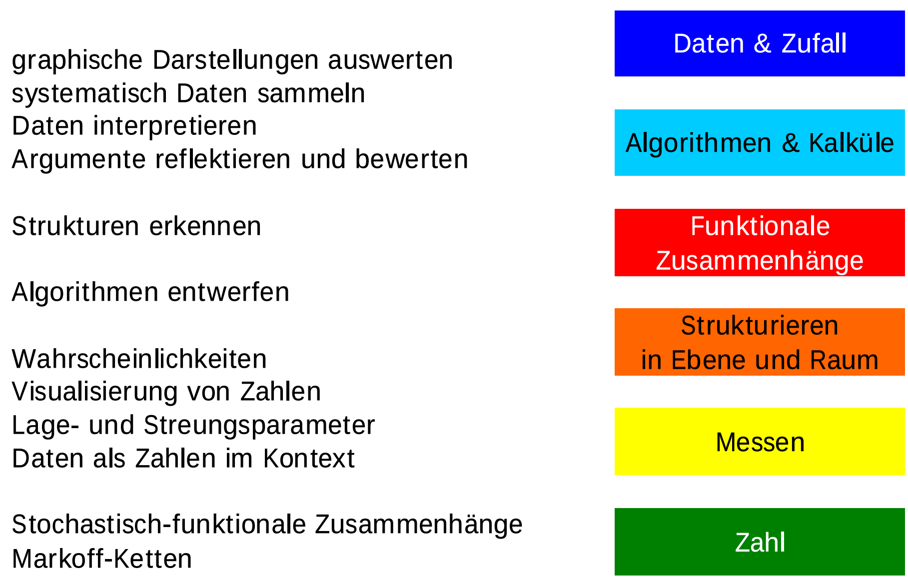

Abb.. 2: Leitideen aus den deutschen Bildungsstandards 2003 und Aktivitäten im Aufsatz – im endgültigen Beschluss fehlte dann leider die Idee „Algorithmen und Kalküle“

Speziell wird ein Algorithmus entworfen, der die Situation strukturiert und zu Einsichten verhelfen kann. Als Ergebnis hat man ein Tableau, in welchem man vielerlei Strukturen erkennen kann, welche dann auch Antworten auf die gestellten Fragen liefern. Damit wird ein Überbau über eine Reihe von kombinatorischen Problemstellungen ermöglicht, in welchen die einzelnen Probleme eingeordnet werden können. Weitere Aktivitäten sind schematisch in Abb. 2 angeführt, es wird weiter unten klar werden, dass die Überlegungen im Aufsatz Beiträge dazu leisten. Als Besonderheit sei darauf verwiesen, dass der angesprochene Algorithmus stochastisch-funktionale Zusammenhänge aufgreift; diese kann man bis zu Markoff-Ketten verallgemeinern und anwenden.

2. Verschiedene Versionen des Problems der vollständigen Serie

Das Sammelbildproblem ist schon lange vor Althoff (2000) diskutiert worden. Hier werden einige damit verwandte Aufgabenstellungen ausformuliert:

1. Aus einem Teig formt man $N = 6$ Semmeln. Wie viele Rosinen muss man beimengen, damit man mit einer Sicherheit (Wahrscheinlichkeit) von mindestens 99% in allen Semmeln wenigstens eine Rosine findet?
2. Man gibt n (≤ 6) Rosinen zum Teig. Wie groß ist die „Sicherheit“, dass keine Semmel zwei Rosinen hat?
   Nimmt man für die Semmeln Tage und für die Rosinen Personen, so erhält man das Geburtstagsproblem – keine zwei Personen haben denselben Geburtstag bzw. wenigstens zwei haben denselben Geburtstag.
3. Man mischt n (≥ 6) Rosinen zum Teig. Wie groß ist die „Sicherheit“, dass sich in allen Semmeln wenigstens eine Rosine findet?
   Nimmt man für die Semmeln die Bilder im Album und für die Zuteilung einer Rosine den „Kauf“ (die Auswahl) eines Bildes, so landet man beim Sammelbildproblem und der Wahrscheinlichkeit, mit n Auswahlen die vollständige Serie zu erhalten.
4. Man gibt zum Teig $n = N$ Rosinen, also gleich viele Rosinen wie Semmeln. Wie groß ist die Sicherheit, dass man nun in allen Semmeln eine Rosine findet? Anmerkung: Nur dieser Sonderfall ist leicht zu lösen.
5. Wie viele Rosinen benötigt man im Durchschnitt – mit anderen Worten: wie groß ist erwartete, benötigte Anzahl von Rosinen, bis alle Semmeln (wenigstens) eine Rosine zugeteilt bekommen haben? Im Sammelbildproblem ist das die Frage nach der Zahl der nötigen Ankäufe von Bildern, bis man das Album voll hat. Genauso gut kann fragen, wie groß die Streuung dieser Zahl ist.

Das Rosinenproblem ist mithin gleichwertig zum Sammelbildproblem. Alle genannten Aufgaben entsprechen den beiden Grundtypen von kombinatorischen Aufgaben – siehe die folgende Tabelle. Bei der deren Modellierung kombinatorischer Probleme stellt man (siehe etwa Krengel 2003) mit Vorteil die folgenden Zusatzfragen.

Ziehen von *n* Kugeln aus einer Urne mit *N* Kugeln – Sammelbildproblem	**Verteilen von *n* Murmeln auf *N* Zellen – Rosinenproblem**
• Reihenfolge wesentlich?	• Sind die Murmeln unterscheidbar?
• Mehrfach ziehen erlaubt? Ohne bzw. mit Zurücklegen.	• Mehrfach belegen erlaubt?

Die beiden Fragestellungen und ihre Lösungen entsprechen einander vollständig: Je nach den Antworten auf die Fragen erhält man Variationen (Reihenfolge ist zu berücksichtigen), oder Kombinationen (Reihenfolge spielt keine Rolle) – ohne bzw. mit Wiederholungen. Der Spezialfall $n = N$ ist leicht gelöst; hier helfen auch Baumdiagramme. Das allgemeine Sammelbildproblem gilt aber als schwierig.

Das ist auch darauf zurückzuführen, dass man die Rosinen künstlich unterscheiden muss, damit man die Gleichwahrscheinlichkeit von Belegungen sichert. Das wiederum ist eine Voraussetzung für die Anwendung der Laplace-Wahrscheinlichkeit. Beim Erwartungswert (Problem 5) kommt noch hinzu, dass man mit den Ausdrücken (es handelt sich um Reihen) geschickt operieren muss, um zum Ergebnis zu kommen.

Daher umgeht man gerne die Kombinatorik und löst die Aufgaben mittels der Simulationsmethode (wie etwa Kühleitner 2007). Dabei werden die Modellvoraussetzungen nachgestellt, wodurch man ein Szenario von Wiederholungen der Situation und somit Daten erhält. Aus der deskriptiven Analyse dieser Daten erhält man einen *Schätzwert* für die mittlere Anzahl von Rosinen, bis alle Semmeln wenigstens eine Rosine enthalten – in der Sammelbild-Variante: die mittlere Anzahl von „Käufen“ von Einzelbildern, bis das Album voll ist.

Simulation stellt eine wertvolle Methode dar: Man erhält meist brauchbare Schätzwerte für die Lösung und man muss die Voraussetzungen des Modells offen auf den Tisch legen, damit man die Simulation durchführen kann. Dadurch gewinnt man möglicherweise einen tieferen Einblick in das Geschehen und darein, warum die Lösung nun wirklich eine Lösung des anstehenden Problems ist. Bei Lernenden entsteht aber gelegentlich doch der Bedarf nach weiterer mathematischer Rechtfertigung der Lösung.

Der rekursive Ansatz zur Lösung der gestellten Aufgaben bietet eine weitere Struktur für die Situation und hat überdies den Vorteil, dass man damit alle Aufgabenstellungen unter einem gemeinsamen Dach vereinen kann. Dabei stellt man den Vorgang der Verteilung von Rosinen (Murmeln) auf die Semmeln (Zellen) als Irrfahrt in der Ebene dar.

3. Umformulierung der Situation als Irrfahrt

Die Rosinen werden eine nach der anderen – schrittweise – den Semmeln „zugeteilt“

Die Zuordnung erfolgt etwa so: Die erste Rosine „belegt“ eine Semmel. Die zweite eine weitere, die dritte "erwischt" eine schon belegte, ... Es wird unterstellt, als ob die Rosinen in einem zeitlichen Takt auf die Semmeln aufgeteilt werden. Der zeitliche Verlauf dieser Aufteilung wird dann durch einen Pfad im quadratischen Zahlengitter dargestellt – siehe Abbildung 3. Die senkrechte Achse entspricht der Zeit bzw. der Nummer der gerade verteilten Rosine. Auf der waagrechten Achse werden die „Zustände“ aufgetragen – hier sind das die Anzahl schon „belegter“ Semmeln. In der Tabellenkalkulation stellt eine Zelle einen Knoten im Zahlengitter.

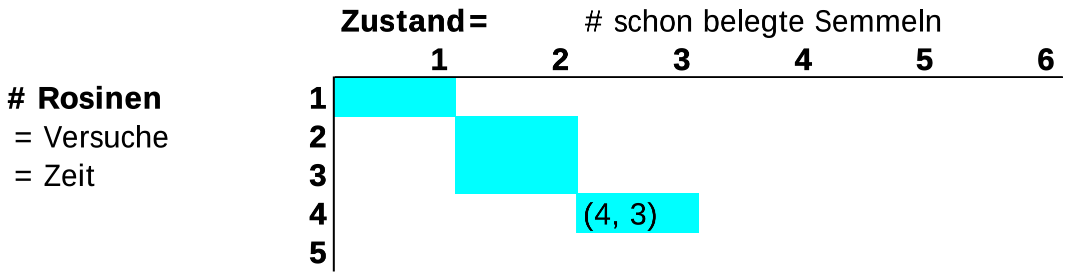

Abb. 3: Irrfahrt – zeitlicher Verlauf der Zuteilung der Rosinen zu den Semmeln

Der Knoten (4, 3) bedeutet: nach der Zuteilung von 4 Rosinen wurden 3 verschiedene Semmeln belegt. Die Irrfahrt startet natürlich beim Knoten (1, 1).

Verzweigung im Vorausblick

Vom Knoten (4, 3), d. h., vier Rosinen sind verteilt, 3 Semmeln schon "besetzt", kann man mit der 5. Rosine eine von den 4 bereits besetzten Semmeln erwischen oder eine noch nicht besetzte (es gibt davon $N - 4$); d. h., man erreicht nur die Knoten (5, 3) und (5, 4). Bei der Zuteilung wird aus allen Semmeln zufällig – mit gleicher Wahrscheinlichkeit – gewählt, also gilt allgemein (wenn man von Randzellen absieht):

Die Verzweigung erfolgt nach folgenden stochastischen Regeln:

- nach unten mit Wahrscheinlichkeit $\boxed{\textbf{belegt} / N}$
- oder nach unten und rechts mit $\boxed{\textbf{frei} / N}$.

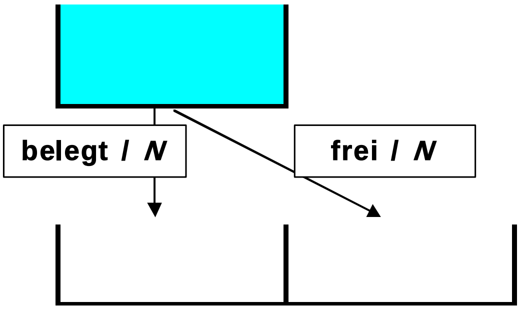

Abb. 4: Vorausblick

Diese Vorausberechnung hilft aber nichts, wenn man berechnen will, wie groß die Wahrscheinlichkeit ist, einen bestimmten Knoten, etwa (5, 4), zu „besuchen".

Verzweigung im Rückblick

Hier hilft der rekursive Ansatz: das Problem wird auf ein „einfacheres" zurückgeführt, dieses wiederum – nach demselben Ansatz – auf ein noch einfacheres usw. Wenn man zum „Zeitpunkt" 5 die Besuchswahrscheinlichkeit von (5, 4) zu bestimmen hat, so kann man diese auf die Besuchswahrscheinlichkeiten der einzelnen Knoten zum Zeitpunkt 4 zurückführen. Man kann zu (5, 4) nur vom Knoten links darüber (4, 3) oder vom Knoten direkt darüber (4, 4) gelangen. „Kennt" man die Besuchswahrscheinlichkeiten dieser Knoten und berücksichtigt man die – im Vorausblick bekannten – Verzweigungswahrscheinlichkeiten von diesen Knoten zum "Auffangknoten", so erhält man die gesuchte Besuchswahrscheinlichkeit zum Zeitpunkt $t = 5$. Natürlich kennt man auch die benötigten Besuchswahrscheinlichkeiten zum Zeitpunkt $t = 4$ nicht. Aber diese führt man rekursiv auf jene zu $t = 3$ usw. zurück, bis man zu $t = 1$ angelangt ist. Da kennt man die Besuchswahrscheinlichkeit, weil die Irrfahrt ganz sicher – mit Wahrscheinlichkeit 1 – im Knoten (1, 1) beginnt.

Für den Auffangknoten (i, k), d. h., i Rosinen sind verteilt, k Semmeln belegt, gilt mit Ausnahme von Randzellen: Man kann nur vom Knoten $(i-1, k-1)$ links darüber oder vom Knoten $(i-1, k)$ direkt darüber kommen (Abb. 3), und zwar mit den Wahrscheinlichkeiten

$$p_{\mathrm{L}} = \frac{N-(k-1)}{N} \qquad \text{bzw.} \qquad p_{\mathrm{D}} = \frac{k}{N}$$

Bezeichnet man die Besuchswahrscheinlichkeit des Knotens (i, k) mit $\mathrm{P}(i, k)$, so gilt:

$$\mathrm{P}(i, k) \;=\; \frac{N-(k-1)}{N} \cdot \mathrm{P}(i-1, k-1) + \frac{k}{N} \cdot \mathrm{P}(i-1, k)$$

Auf diese Weise kann man die gefragten Besuchswahrscheinlichkeiten rekursiv von oben nach unten berechnen. Mit den Bezeichnungen

$$\mathrm{L} := \mathrm{P}(i-1, k-1) \quad \text{bzw.} \quad \mathrm{D} := \mathrm{P}(i-1, k),$$

ergibt sich plakativ (siehe Abb. 5):

$$\mathrm{P}(i, k) \;=\; p_{\mathrm{L}} \cdot \mathrm{L} + p_{\mathrm{D}} \cdot \mathrm{D}$$

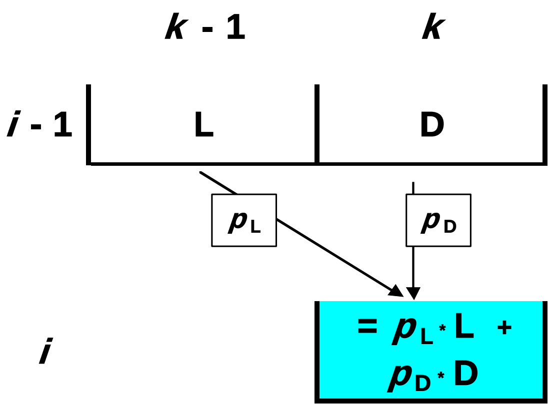

Abb. 5: Rückblick – Die Rekursion

In algebraischer Schreibweise liegt hier folgende Funktion zweier ganzzahliger positiver Argumente vor (siehe Neuwirth 2001):

$$\begin{array}{lcll}
\mathrm{P}(1,1) & = & 1 & \\
\mathrm{P}(1,k) & = & 0 & k>1 \\
\mathrm{P}(n,1) & = & \frac{1}{N}\cdot \mathrm{P}(n-1,1) & n>1 \\
\mathrm{P}(n,k) & = & \frac{k}{N}\cdot \mathrm{P}(n-1,k)+\frac{N-(k-1)}{N}\cdot \mathrm{P}(n-1,k-1) & n>1,\ k>1
\end{array}$$

In Neuwirth (2001) werden auch allgemeine Beweistechniken für eine große Zahl von Theoremen über derartige Funktionen gezeigt. Unsere speziellen Rekursionskoeffizienten hängen von k, aber nicht von n ab und sind daher innerhalb einer Spalte (fester Index k) konstant. Mit einem derartigen Modell lässt sich jede Irrfahrt beschreiben, wo Übergänge nur zum selben Ort oder zum rechten Nachbarn möglich sind und die Übergangswahrscheinlichkeiten zeitlich konstant sind.

Die rekursive Formulierung der Verteilung von n Rosinen auf N Semmeln wird durch eine Irrfahrt dargestellt. Dabei wird ein Zahlengitter von oben nach unten durchwandert. Dadurch kommt in natürlicher Weise eine „Zeitkomponente“ hinzu.

Der Zeittakt der Irrfahrt wird durch Zuordnung einer weiteren Rosine zu einer der Semmeln bestimmt. Statt von der Anzahl der Rosinen und den bereits belegten Semmeln zu sprechen, kann man auch formulieren, wie lange (Zeit) es braucht, um eine bestimmte Anzahl von Semmeln mit einer gewissen Wahrscheinlichkeit zu belegen.

Man kann auch von der „Wartezeit“ sprechen, wie lange es braucht, bis alle Semmeln mit einer gewissen Wahrscheinlichkeit belegt sind, und meint dabei, wie viele Rosinen benötigt werden, damit alle Semmeln mit dieser gewissen Wahrscheinlichkeit wenigstens eine Rosine haben.

In der Sprache des verwandten Sammelbildproblems bedeutet das, wie viele Bilder zu kaufen sind, damit man mit einer bestimmten Sicherheit eine vorgegebene Anzahl von verschiedenen Bildern im Sammelalbum eingeklebt hat, bzw., wie viele Bilder man kaufen muss, damit das Album mit einer bestimmten Sicherheit voll ist. Hier wird die Bezeichnung „Wartezeit“ noch suggestiver, es ist die Wartezeit, bis die Serie vollständig ist.

Das Tableau der Irrfahrt ist in Abb. 6 festgehalten. Dabei stehen in den einzelnen Knoten die jeweiligen Besuchswahrscheinlichkeiten. In Zeile **10** stehen die Wahrscheinlichkeiten der Zustände (=Anzahl belegter Semmeln) nach der Zuordnung von 10 Rosinen. In Spalte **6** stehen die Wahrscheinlichkeiten, mit der jeweiligen Anzahl von Rosinen bereits alle sechs Semmeln belegt zu haben.

Unmittelbar einsichtig ist, dass die Irrfahrt das rechte obere Dreieck nicht erreichen kann, weil etwa mit 5 Rosinen nicht schon 6 Semmeln belegt sein können. Spalten **7** und höher haben in der Basisversion des Beispiels keinen Sinn, weil ja nur 6 Semmeln da sind; sie stehen für die Verallgemeinerung des Beispiels bereit.

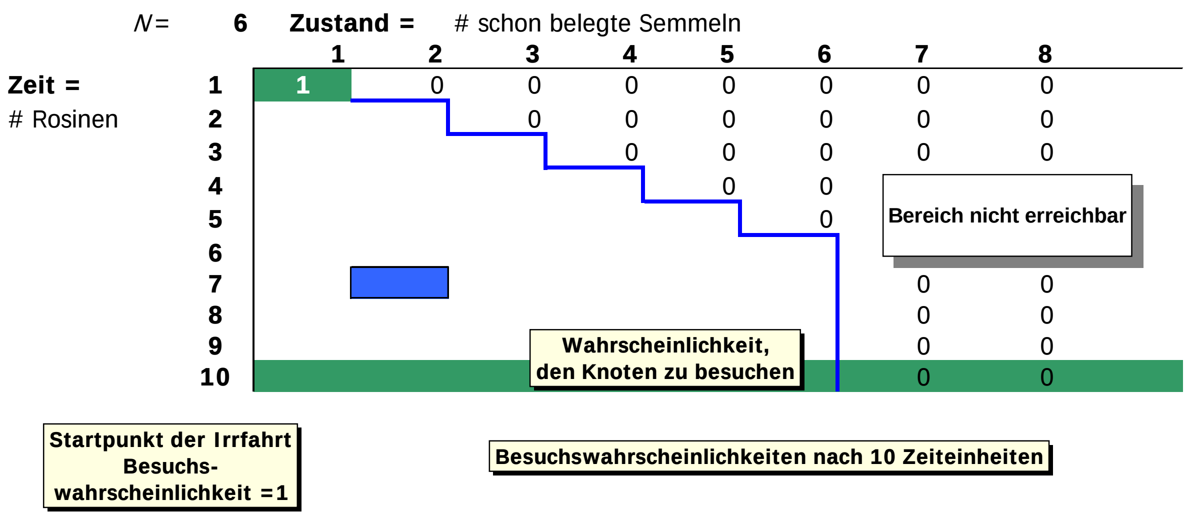

Abb. 6: Das Tableau mit den Besuchswahrscheinlichkeiten

4. Umsetzung der Irrfahrt in Excel

Die kleine Anleitung, wie man das Tableau mit den Besuchswahrscheinlichkeiten in Excel erstellt (siehe Abb. 7), zeichnet sich durch effizienten Einsatz von relativer und absoluter Adressierung aus. Dadurch benötigt man nur wenige Angaben: in einer Zelle wird die rekursive Formel für Besuchswahrscheinlichkeit in Abhängigkeit von der Zeile davor eingetragen. Den Rest kann man durch Kopieren, vorzugsweise durch Markieren und Ziehen des so genannten Ausfüllkästchens, erledigen. Bleiben noch die Randwahrscheinlichkeiten.

1. Zeile: kennen wir sowieso

A	B	C	D	E	F	G	H	I
	N=	6	# schon belegte Semmeln					
			1	2	3	4	5	6
		1	1	0	0	0	0	0
		2						

2. Zeile: *1. Eintrag*: Zustand wird nur von direkt darüber erreicht

A	B	C	D	E	F	G	H	I
	N=	6	# schon belegte Semmeln					
			1	2	3	4	5	6
		1	1	0	0	0	0	0
		2	0,167					
		3						

=D$2*D3/$C$1
D2*pD

2. Zeile: *2. Eintrag*: Jetzt die Rekursion von Links darüber - Direkt darüber

A	B	C	D	E	F	G	H	I
	N=	6	# schon belegte Semmeln					
			1	2	3	4	5	6
		1	1	0	0	0	0	0
		2	0,167					
		3						
		4						
	inen	5						

=D3*($C$1-E$2+1)/$C$1 + E3*E$2/$C$1
$D3*p_L + E3*p_D$

Rest von Zeile 2: 2. Eintrag nach rechts kopieren

A	B	C	D	E	F	G	H	I
	N=	6	# schon belegte Semmeln					
			1	2	3	4	5	6
		1	1	0	0	0	0	0
		2	0,167	0,833				

Rest der Tabelle: Zeile 2 nach unten kopieren

A	B	C	D	E	F	G	H	I
	N=	6	# schon belegte Semmeln					
			1	2	3	4	5	6
		1	1	0	0	0	0	0
		2	0,167	0,833	0,000	0,000	0,000	0,000

Ergebnis: Die Nullen besser dargestellt

A	B	C	D	E	F	G	H	I
	N=	6	# schon belegte Semmeln					
			1	2	3	4	5	6
	Rosinen	1	1	0	0	0	0	0
		2	0,167	0,833	0	0	0	0
		3	0,028	0,417	0,556	0	0	0
		4	0,005	0,162	0,556	0,278	0	0
		5	0,001	0,058	0,386	0,463	0,093	0
		6	0,000	0,020	0,231	0,502	0,231	0,015
		7	0,000	0,007	0,129	0,450	0,360	0,054
		8	0,000	0,002	0,069	0,365	0,450	0,114
		9	0,000	0,001	0,036	0,278	0,497	0,189
		10	0,000	0,000	0,019	0,203	0,506	0,272

Abb. 7: Illustrationen zur Implementierung in EXCEL

5. Bearbeitung der Aufgaben in Excel

Einige der Lösungen erhält man direkt aus der Tabelle in Abb. 8.

Die Spalte **6** ist rechte – absorbierende – Schranke für die Irrfahrt. Mehr als 6 Semmeln haben wir in der Basisversion des Beispiels nicht. Wenn man diese Spalte zum ersten Mal kreuzt, hat man (frühestens) alle Semmeln belegt – Sammelbildproblem: das Album ist mit den eben durchgeführten Kauf eines Bildes voll geworden. Danach geht die Irrfahrt nur mehr nach unten weiter.

Aufgabe 1: Kreuzt man bei $n = 34$ Rosinen das erste Mal die Spalte **6**, so gibt es dafür eine Wahrscheinlichkeit von 0,988. Die Sicherheit, ist noch knapp unter den geforderten 99%, dass alle Semmeln bereits eine Rosine haben. Bei $n = 36$ ist es dann so weit; das ist die Lösung von Aufgabe 1 (im Tableau von Abb. 8 könnte dies auch schon bei $n = 35$ erfüllt sein, weil das Ergebnis von 0,9899 nur auf drei Stellen mit 0,990 angegeben ist; hier ergibt sich ein kleiner Nachteil aus dem Bestreben heraus, einen guten visuellen Überblick zu ermöglichen).

Aufgabe 2: Keine Semmel soll 2 und mehr Rosinen haben, wenn $n \le 6$ Rosinen zugeteilt sind. Das bedeutet gleich viele „beteilte" Semmeln wie Rosinen. Die Lösung steht in der „Diagonale". Die Gegenwahrscheinlichkeit ist dann die Wahrscheinlichkeit, dass mindestens eine Semmel 2 oder mehr Rosinen enthält. In der Sprache des Geburtstagsproblems: Wenigstens zwei haben denselben „Wochentag" als Geburtstag (wenn man unterstellt, es gäbe nur 6 Tage in der Woche).

Aufgabe 3: Alle Semmeln sollen schon wenigstens eine Rosine haben. Wir befinden uns also im Schnittpunkt von Zeile $\boldsymbol{n}$ und Spalte **6**. In Abhängigkeit von der Zahl der ausgeteilten Rosinen n steht dort die Wahrscheinlichkeit, die Bedingung zu erfüllen.

Aufgabe 4: Der Spezialfall, dass bereits $n = N$ (in der Basisversion sind das $n = 6$) Rosinen reichen, damit alle Semmeln auch bereits eine zugeteilt bekommen haben, steht am Schnittpunkt der „Diagonale" mit Spalte **6**.

Aufgabe 5: Für die erwartete Zahl der Rosinen, die man benötigt, damit alle N = 6 Semmeln wenigstens mit einer Rosine belegt sind (Aufgabe 5), betrachten wir noch ein Mal die Spalte **6** in Abb. 8. Diese gibt an, mit welcher Wahrscheinlichkeit der Knoten $(n, 6)$ besucht wird. Die Nullen in den ersten fünf Zeilen sind einsichtig, weil man mit weniger als 6 Rosinen nicht alle 6 Semmeln belegen kann. Das Ereignis, dass bei n = 6 Rosinen alle Semmeln belegt sind, hat eine Wahrscheinlichkeit von 0,015 (aus Spalte **6**). Bei n = 7 Rosinen schon alle Semmeln belegt zu haben, hat eine Wahrscheinlichkeit von 0,054 (der direkt darunter liegende Wert in Spalte **6**). Zieht man den vorhergehenden Tabellenwert ab, d. h.,

$$0{,}054 - 0{,}015 = 0{,}039,$$

so erhält man gerade die Wahrscheinlichkeit, bei Zuteilung der 7. Rosine auf die Semmeln erstmalig alle Semmeln zu belegen. In der gleichen Weise kann man alle Differenzen übereinander liegender Werte in Spalte **6** interpretieren. Diese Wahrscheinlichkeiten werden in Spalte **P** festgehalten.

Nun definiert man dazu eine Zufallsvariable X wie folgt: Die Zufallsvariable erhält als Wert die Nummer jener Rosine, bei deren Zuteilung man in der Irrfahrt erstmalig die Spalte **6** erreicht. Man kann bei dieser Zufallsvariablen von der Anzahl X benötigter Rosinen sprechen, damit alle Semmeln belegt sind. Interpretiert man die Aufteilung der Rosinen auf die Semmeln hintereinander als Zeittakt, so kann man auch von der Wartezeit X sprechen, bis alle Semmeln belegt sind.

Die erwartete Anzahl der Rosinen, die benötigt werden, damit man alle Semmeln belegt hat, ist daher der Erwartungswert der eben festgelegten Zufallsvariablen X. In der Sprache der Sammelbilder ist diese Zufallsvariable gerade die Wartezeit auf die vollständige Serie, das ist die Anzahl der Bilder, die erforderlich ist, damit man alle verschiedenen Bilder der Serie in seinem Sammelalbum hat.

Diese Verteilung der Wartezeit kann man in Abb. 9 betrachten. Aus Spalte **P** in Abb. 8 erhält man mit Hilfe von Spalte **E(X)** den Erwartungswert als Summe, mit Hilfe von Spalte **E(X²)** den entsprechenden Erwartungswert des Quadrats, und mit dem Verschiebungssatz für Varianzen erhält man die Varianz und schließlich die Standardabweichung:

$$\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - \{\mathrm{E}(\mathrm{X})\}^2 \qquad \text{bzw.} \qquad \sigma = \sqrt{\mathrm{V}(\mathrm{X})}\,.$$

6. Zu viele Semmeln und Rosinen

Das mathematisch gleiche Problem in einer anderen konkreten Einbettung ist Folgendes: Beim Lotto in Österreich gibt es 8.145.060 Tipps. Wenn in einer Runde 10 Millionen Tipps abgegeben wurden, wie viele verschiedene Tipps wurden dann im Durchschnitt gesetzt? Wie viele Tipps sind im Schnitt notwendig, dass alle möglichen Tipps tatsächlich gesetzt sind? Die Anzahl der nicht gesetzten Tipps der ersten Frage ist auch die Frage nach der Wahrscheinlichkeit eines Jackpots.

In dieser Größenordnung ist das Problem mit Excel nicht mehr behandelbar. Allerdings kann man Approximationslösungen finden, siehe auch Knessl und Keller (1991). Hierzu wird ist ein weiterer Aufsatz des Autors gemeinsam mit E. Neuwirth geplant.

7. Auswirkung einiger Parameter

In den gestellten Aufgaben erhebt sich ganz natürlich die Frage nach dem Einfluss von Eingangsgrößen auf die Lösung, insbesondere kann man sich fragen, wie sich die Zahl der vorhandenen Zellen auswirkt.

> Aufgabe 1:
> Wie ändert sich der 99%-Punkt in Aufgabe 1, wenn sich die Zahl der Zellen ändert? Wie ändert sich der Schwellenwert, wenn sich die geforderte Mindestwahrscheinlichkeit (Sicherheit) ändert?
>
> Aufgabe 2:
> Wie ändern sich die Wahrscheinlichkeiten, dass keine zwei Rosinen in einer Zelle sind in Abhängigkeit von den verteilten Rosinen? Im Geburtstagsproblem entspricht dem die Zahl der Personen und die davon abhängige Wahrscheinlichkeit für denselben Geburtstag von mindestens zweien (die Gegenwahrscheinlichkeit zu zuvor).

n N = Zahl Semmeln

6

Anzahl Ros	Anzahl schon belegter Semmeln 1	2	3	4	5	6	Wartezeit P $P(X=n)$	E(X) $n \cdot P(X=n)$	$E(X^2)$ $n^2 \cdot P(X=n)$	$\sqrt{V(X)}$
1	1	0	0			0	0,000	0,000	0,000	
2	0,167	0,833	0			0	0,000	0,000	0,000	
3	0,028	0,417	0,556	0	0	0	0,000	0,000	0,000	
4	0,005	0,162	0,556	0,278	0	0	0,000	0,000	0,000	
5	0,001	0,058	0,386	0,463	0,093	0	0,000	0,000	0,000	
6	0,000	0,020	0,231	0,502	0,231	0,015	0,015	0,093	0,556	
7	0,000	0,007	0,129	0,450	0,360	0,054	0,039	0,270	1,890	
8	0,000	0,002	0,069	0,365	0,450	0,114	0,060	0,480	3,841	
9	0,000	0,001	0,036	0,278	0,497	0,189	0,075	0,675	6,076	
34	0,000	0,000	0,000	0,000	0,012	0,988	0,002	0,083	2,809	
35	0,000	0,000	0,000	0,000	0,010	0,990	0,002	0,071	2,482	
36	0,000	0,000	0,000	0,000	0,008	0,992	0,002	0,061	2,190	
37	0,000	0,000	0,000	0,000	0,007	0,993	0,001	0,052	1,928	
38	0,000	0,000	0,000	0,000	0,006	0,994	0,001	0,045	1,695	
65	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,001	0,036	
66	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,031	
							1,000	14,70	254,89	6,24

Diagonale: Lösung #2

n=N Lösung #4

mindestens 99% Lösung #1

Einige Zeilen sind ausgeblendet

Wartezeit Lösung #3

Erwartungswert Lösung #5

Standardabweichung zu Lösung #5

Abb. 8: Ergebnis der rekursiven Berechnung: Tableau mit Zusatzspalten für den Erwartungswert

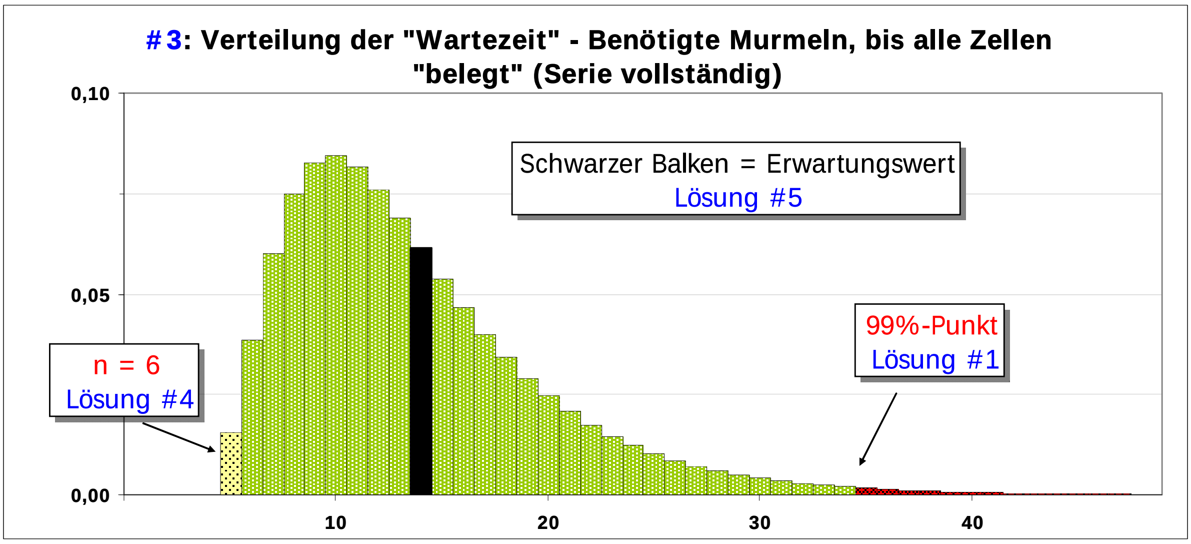

Abb. 9: Die Verteilung der Wartezeit mit ihren Parametern als Lösung der Aufgaben

Wie ändern sich diese Wahrscheinlichkeiten für keine zwei Rosinen in Abhängigkeit von den vorhandenen Zellen? Im Geburtstagsproblem entspricht dem etwa der Übergang von Tagen im Jahr zum Geburtsmonat, Jahreszeit oder zu Wochentagen.

Aufgabe 3:
Man kann fragen, wie viele Zellen man mit *n* verteilten Rosinen mit welcher Wahrscheinlichkeit schon belegt hat (die *n*-te Zeile im Tableau in Abb. 8)? Wie hängt diese Verteilung von der Zahl der verfügbaren Zellen ab?

Gefragt ist eigentlich nach der Wartezeit, bis man alle Zellen mit wenigstens einer Rosine belegt hat (die *N*-te Spalte im Tableau). Wie hängt diese Wartezeit von der Zahl der Zellen ab?

Aufgabe 4:
Gefragt ist die Wahrscheinlichkeit, dass die Mindestzahl $n = N$ von Rosinen auch schon reicht, damit sich in allen Zellen schon eine Rosine befindet. Wie hängt diese Wahrscheinlichkeit von der Zahl der Zellen ab?

Aufgabe 5:
Wie hängt die Verteilung der Wartezeit, bis alle Zellen „befüllt" sind, von der Zahl der verfügbaren Zellen ab? Insbesondere kann man fragen, wie der Erwartungswert bzw. die Standardabweichung dieser (sehr schiefen) Verteilung von der Zahl der Zellen abhängen? Fragt man nach einem Quantil, das mit einer Sicherheit von *p* unterschritten wird, landet man beim *p*-Quantil dieser Verteilung und wiederum bei Frage 1.

All diese Fragen kann man leicht mit Hilfe von EXCEL untersuchen, indem man die Eingangszellen für die Parameter *N* und der Sicherheit mit einem Schieberegler variabel macht.

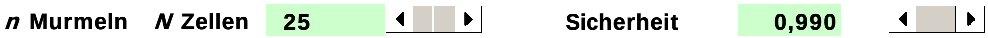

Abb. 10: Schieberegler für Einflussgrößen

Bei der Berechnung des Erwartungswerts und der Standardabweichung der Wartezeit stößt man allerdings doch wieder an Grenzen in EXCEL, weil die unendlichen Reihen mit endlichen Summen approximiert werden müssen. Man muss das Tableau doch ziemlich großzügig erweitern, weil die Schiefe der Verteilung die unendlichen Reihen doch recht langsam konvergieren lässt und erst bei 1000 und mehr Summanden eine ausreichende Genauigkeit erreicht wird.

Der kleine Nachteil dabei: Eine schöne Animation der Verteilung der Wartezeit mit Einschluss des Erwartungswerts und des Quantils in Abhängigkeit von der Zahl der Zellen N bzw. der Sicherheit p – in der Graphik in Abb. 9 erlebt man die Veränderungen wie in einem Film, wenn man die Regler „zieht“ und damit die Einflussgrößen steuert – wird dann ziemlich langsam.

Die folgenden Schaubilder (Abb. 11 und 12) zeigen einige Abhängigkeiten und geben Anlass zu weiteren Fragen: Kann man die gefundenen Abhängigkeiten auch mathematisch beweisen? Gerade weil sich die angesprochenen Konvergenzprobleme leicht demonstrieren lassen, sind „Zweifel“ durchaus berechtigt. Wir können vieles berechnen, aber kommen durch die Hintertüre nicht auch Artefakte (wie Rechenfehler, Approximationsfehler etc.) herein? Natürlich kann man diesen Fragen nachgehen, aber das spricht dann eher nur eine Minderheit von Lernenden an.

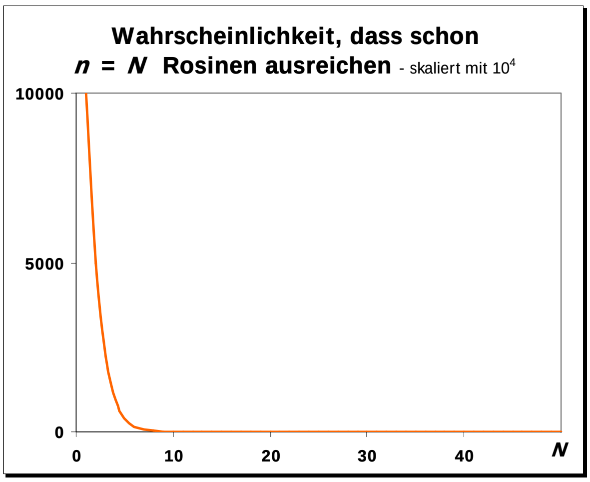

Abb. 11: Der starke Rückgang der Lösungswahrscheinlichkeit in Aufgabe 4

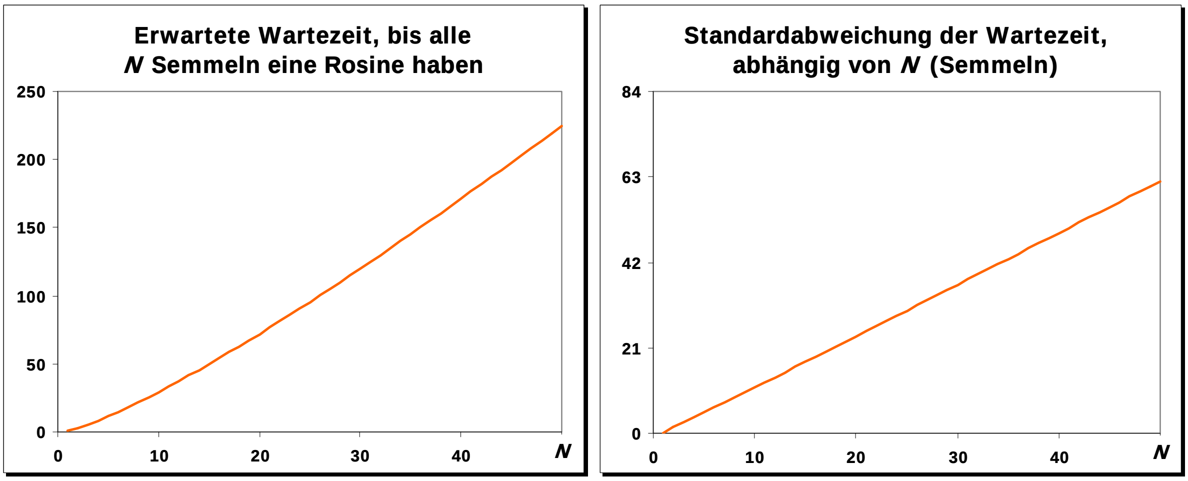

Abb. 12: Ordnung des Wachstums von Erwartungswert und Standardabweichung in Aufgabe 5

8. Zusammenfassung und Ausblick

Henze (2003, 193ff) diskutiert das Thema sehr ausführlich und in einer noch allgemeineren Form, wobei er explizite Formeln für die Verteilung und die Momente der „Wartezeit“ mit doch noch relativ elementaren Mitteln (der Siebformel) mathematisch herleitet.

Der in der vorliegenden Arbeit vorgestellte rekursive Ansatz ist allgemeiner als man zunächst vielleicht vermuten wird. Mit derselben Methode kann man die Binomialverteilung als Irrfahrt (im Pascalschen Dreieck) herleiten. Sie hat ganz spezielle Eigenschaften, sie ist die einzige Irrfahrt, die räumlich und zeitlich homogen ist.

Viele andere kombinatorische Probleme haben eine rekursive Variante, die viel leichter einer Lösung zugänglich ist als die statisch-algebraische. Der rekursive Ansatz lässt viele Aufgabenstellungen lösen, ohne dass man auf das eigentliche Problem eingeht. Verwiesen sei hier nur auf die "Türme von Hanoi" oder auf die rekursive Bestimmung des größten gemeinsamen Teilers von zwei ganzen Zahlen (siehe etwa Hartmann 2006, S. 63f). Für die Implementierung des Lösungswegs in einer Software ergeben sich nicht nur Vereinfachungen, sondern auch Einsparung von Ressourcen, insbesondere auch von Rechenzeit.

In den Tabellen werden Strukturen sichtbar: Im Tableau von Abb. 8 findet man Aufgabe 2 in der Diagonale wieder usw. Die graphische Darstellung der Wartezeit (Abb. 9) erhält man durch eine weitere Spalte mit Berechnungen. Die „Verteilung“ der Wartezeit lässt das Problem visuell noch einmal ganz anders – aus einer globalen Perspektive heraus – erscheinen. Jetzt versteht man aufgrund der Schiefe dieser Verteilung, dass die Wartezeiten auf die volle Serie so stark streuen und erheblich größere Wartezeiten durchaus realistisch sind.

Das Besondere an einer Tabellenkalkulation liegt darin, dass man die Graphiken zu Animationen umgestalten kann, wenn man Einflussgrößen über Schieberegler steuert. Jetzt werden weitere Einsichten möglich. Die Ergebnisse des „Filmes“ kann man wiederum als Tabellen und Graphiken speichern und weiter bearbeiten, wie das in den Abb. 11 und 12 erfolgt ist.

Jetzt kann man die Entwicklung der erwarteten Wartezeit in ihrer Qualität studieren und kommt darauf, dass sie mehr als linear wächst, aber vielleicht weniger als quadratisch. Mit mehr Mathematik könnte man die Ordnung des Wachstums mit $N \cdot \ln(N)$ erweisen. Die Ergebnisse der Berechnungen und graphischen Darstellungen regen zu weiterer mathematischer Tätigkeit an, auch im Sinne der Begriffsbildungen und Beweistechniken. Darüber hinaus sind numerische Probleme der Approximation auf der „Tagesordnung".

Neben diesen Meriten kann man EXCEL (oder eine Tabellenkalkulation allgemein) heute zu den üblichen Werkzeugen zählen. Wenn irgendwo in Anwendungen gerechnet wird, so ist das ein erster Ansatz. Know-how, welches frühzeitig hier erworben wird, gibt den Lernenden dann gute Karten in die Hand.

Ein Excel-File sowohl mit der Lösung für $N = 6$ Zellen (Semmeln) als auch für allgemeines N für alle angesprochenen Aufgaben finden Sie im Verweis von Borovcnik auf eine Internetquelle.

Das Sammelbildproblem – in der Variante mit dem Erwartungswert (Aufgabe 5) – ist für den Unterricht mathematisch zu anspruchsvoll. Die Fragestellung liegt aber nahe, wenn man Aufgabe 3 behandelt. Ja, sie ist eigentlich noch viel nahe liegender als die anderen Fragen: Wie lange braucht man, um das Album voll zu bekommen? Die übliche didaktische Antwort auf solche Schwierigkeiten in der Stochastik ist die Simulation – ein Ansatz, den z. B. Kühleitner (2007) in seinem Unterrichtsvorschlag zum Sammelproblem verfolgt.

Ganz allgemein hat sich der Ansatz der Simulation in den letzten Dekaden als vielleicht der erfolgreichste erwiesen. Auf zwei jüngere Quellen sei noch verwiesen (Christie 2004, Borovcnik 2006), die einen – ganz anderen – Zugang zum statistischen Testen ergeben, das so genannte Resampling. Dabei wird die Genauigkeit der Schätzung eines Parameters der Grundgesamtheit durch wiederholte Stichproben aus der einen schon vorhandenen Stichprobe evaluiert. So paradox das klingen mag, es funktioniert und ist viel einsichtiger als das eigentliche, traditionelle Testen.

Die Pointe im Vorschlag hier ist, dass sich eine Reihe mathematischer Tätigkeiten durch die rekursive Umformulierung des Problems eröffnen. Eine weitere Folge ist die vereinheitlichende Darstellung so verschiedener kombinatorischer Probleme durch die *Anordnung* im Tabellenblatt. Klarerweise dient der Vorschlag hier auch, um die Effektivität rekursiver Algorithmen i. A. zu belegen und den algorithmischen Zugang zur Mathematik gegenüber dem statisch-begrifflichen aufzuwerten.

Der rekursive Ansatz ist typischerweise in der Behandlung von Markoff-Ketten wieder zu finden. Dass Markoff-Ketten nicht ganz weit weg vom Unterricht sind, hat G. Schröpfer (Lehrer in Graz) schon vor Jahren durch seine diesbezüglichen unterrichtlichen Bemühungen gezeigt (leider fehlt ein Verweis darauf); in jüngerer Zeit hat man die Ideen in Mathe-Prisma (o. J.) aufgegriffen. Es ergeben sich die bekannten Anknüpfungen zur Linearen Algebra.

Wer sich angeregt fühlt, mehr über die Möglichkeiten von EXCEL für die Stochastik und den Unterricht in Mathematik insgesamt zu erfahren, kann sich etwa in Borovcnik (o. J.), Neuwirth und Arganbright (2004), Neuwirth (o. J.) oder auf der Internet-Seite *Spreadsheets in Education* orientieren. J. Meyer hat „versprochen“, auf EXCEL-Anfragen über e-mail zu antworten – er hat einen großen Fundus von EXCEL-Lösungen für den Unterricht der Sekundarstufen.

Die vorliegenden Ausführungen stellen eine Weiterentwicklung der Gedanken in Borovcnik (2007) dar. Den rekursiven Zugang haben die Autoren auch genützt, um die Binomialverteilung herzuleiten (Borovcnik und Neuwirth, 2006). Die Rekursion wird im Vergleich zu hier noch einfacher.

Literatur

Althoff, H. (2000): Zur Berechnung der Wahrscheinlichkeiten für das Vorliegen einer vollständigen Serie (Sammelbildproblem). *Stochastik in der Schule* 20(1), 18–20.

Bartz, S. (2007): Excelblatt vereinfacht Stochastik. *Stochastik in der Schule* 27(2), 25–29.

Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss. Deutscher Bildungsserver.
http://www.bildungsserver.de/zeigen.html?seite=2145 oder Beschluss (4.12.2003)
http://www.kmk.org/schul/Bildungsstandards/Mathematik_MSA_BS_04-12-2003.pdf
oder Entwurf (4.7.2003)
http://www.kmk.org/aktuell/Bildungsstandards/Mathematik04072003.pdf (Datum der Abfrage: 11.12.2008)

Borovcnik, M. (1992): *Stochastik im Wechselspiel von Intuitionen und Mathematik.* BI, Mannheim.

Borovcnik, M. (o. J.): *EXCEL für den Unterricht in Stochastik.* Externe Links „Educational Statistics an der Universität Klagenfurt“.
www.mathematik.uni-kassel.de/stochastik.schule/ (Datum der Abfrage 11.12.2008)

Borovcnik, M. (2006): Daten – Zufall – Resampling. J. Meyer (Hg.): *Anregungen zum Stochastikunterricht*, Bd. 3. Franzbecker, Hildesheim, 143–158.

Borovcnik, M. (2007): Das Sammelbildproblem – Rosinen und Semmeln und Verwandtes: Eine rekursive Lösung mit Irrfahrten. *Stochastik in der Schule* 27(2), 19–24.

Borovcnik, M. und Neuwirth, E. (2006): *Ein rekursiver Zugang zur Binomialverteilung.*
http://www.osg.or.at/main.asp?VID=1&kat1=60&kat2=401&kat3=312&Text=&GenLiPage=2&NID=39 (Datum der Abfrage 11.12.2008)

Christie, D. (2004): Resampling mit Excel. *Stochastik in der Schule* 24(3), 22–27.

Hartmann, P. (2006): *Mathematik für Informatiker. Ein praxisbezogenes Lehrbuch*, 3. Aufl. Vieweg, Wiesbaden.

Knessl, C. und Keller, J. B. (1991): Stirling number asymptotics from recursion equations using the ray method. *Stud. Appl. Math.* 84 no. 1, 43–56.

Krengel, U. (2003): *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* Vieweg, Wiesbaden.

Kühleitner, M. (2007): Das Sammelproblem – eine Simulation mit Excel. *Stochastik in der Schule* 27(1), 24–26.

Mathe-Prisma: *Craps – Die Würfel sind gefallen* (Ansatz mit Markoff-Ketten). http://www.matheprisma.uni-wuppertal.de/Module/Craps/ (Datum der Abfrage 11.12.2008).

Meyer, J.: *EXCEL-Files in der Lehrerfortbildung.* Auf Anfrage beim Autor: J.M.Meyer@t-online.de .

Neuwirth, E. und Arganbright, D. (2004): *The Active Modeler: Mathematical Modeling with Microsoft Excel.* Brooks, Cole.

Neuwirth, E. (o. J.): *Spreadsheets, Mathematics, Science, and Statistics Education.* http://sunsite.univie.ac.at/Spreadsite/ (Datum der Abfrage 11.12.2008).

Neuwirth, E. (2001): Recursive Combinatorial Functions: Extending Galton's Board. *Discrete Mathematics* 239, 33–51.

Spreadsheets in Education: www.sie.bond.edu.au/ (Datum der Abfrage 11.12.2008).

Wikipedia: *Türme von Hanoi.* http://de.wikipedia.org/wiki/T%C3%BCrme_von_Hanoi (Datum der Abfrage 11.12.2008).