Wartezeitprobleme in Bernoulli-Ketten – ein verständnisorientierter Zugang

NORBERT HENZE, KARLSRUHE

Zusammenfassung: *Wir nehmen den aktuellen Aufsatz Dorner (2017) zum Anlass, das inhaltliche Verständnis einer Bernoulli-Kette anhand einiger elementarer Wartezeitprobleme zu fördern.*

1 Einleitung

Bernoulli-Ketten bilden ein zentrales Thema im gymnasialen Stochastikunterricht. Sie modellieren in unabhängiger Folge durchgeführte Versuche mit jeweils zwei möglichen Ausgängen, die üblicherweise als Treffer bzw. Niete bezeichnet und mit 1 bzw. 0 codiert werden. Charakteristisch für Bernoulli-Ketten ist nicht nur die stochastische Unabhängigkeit von Ereignissen, die sich auf unterschiedliche Versuche beziehen, sondern auch eine von Versuch zu Versuch gleich bleibende Trefferwahrscheinlichkeit, die im Folgenden mit p bezeichnet werde. Dabei nehmen wir stets $0 < p < 1$ an. Der aktuelle Aufsatz Dorner (2017) behandelt elementare Wartezeitprobleme bei Bernoulli-Ketten, wobei die zentrale Bedeutung der geometrischen Reihe und deren Ableitungen zur Bestimmung von Erwartungswerten und Varianzen betont wird. Wir nehmen diesen Aufsatz zum Anlass, einerseits einen das entdeckende Lernen fördernden Zugang zu diesem technischen Hilfsmittel wiederzubeleben; zum anderen zeigen wir auf, wie ein größeres inhaltliches Verständnis der Bernoulli-Kette über Wartezeitprobleme erreicht werden kann.

2 Warten auf den ersten Treffer

Nach der ersten Pfadregel hat die mit W_1 bezeichnete Anzahl der bis zur Erzielung des ersten Treffers benötigten Versuche die Verteilung

$$\mathbb{P}(W_1 = i) = q^{i-1} p, \qquad i = 1, 2, \ldots .$$

Dabei haben wir $q := 1 - p$ gesetzt und werden auch im Folgenden so verfahren. Um nachzuweisen, dass es sich hierbei um eine (um 1 verschobene) geometrische Verteilung handelt, und um den Erwartungswert und die Varianz von W_1 zu bestimmen, kann man die geometrische Reihe

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad 0 < x < 1, \tag{1}$$

sowie deren erste und zweite Ableitung

$$\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}, \quad 0 < x < 1, \tag{2}$$

$$\sum_{k=2}^{\infty} k(k-1) x^{k-2} = \frac{2}{(1-x)^3}, \quad 0 < x < 1, \tag{3}$$

verwenden. Beim Warten auf den zweiten Treffer kann auch noch die dritte Ableitung

$$\sum_{k=3}^{\infty} k(k-1)(k-2) x^{k-3} = \frac{6}{(1-x)^4}, \quad 0 < x < 1, \tag{4}$$

nützlich sein. Natürlich gelten diese Identitäten für jedes x mit $|x| < 1$, für unsere Zwecke reicht aber die Erkenntnis der jeweiligen Gültigkeit für positive x völlig aus, was technische, mit dem Betrag zusammenhängende Fragen ausblendet.

Obwohl Beziehung (1) momentan nicht Standardstoff an deutschen Gymnasien ist, können Schülerinnen und Schüler diese Formel sowie (2), (3) und (4) relativ leicht selbst entdecken und begreifen. Letztlich läuft alles darauf hinaus, dass für jedes y mit $0 < y < 1$ die Folge y^n für $n \to \infty$ gegen Null konvergiert. Da die damit zusammenhängenden Überlegungen nicht zu einem tieferen *stochastischen* Verständnis der behandelten Wartezeitprobleme beitragen, haben wir sie in einen eigenen Abschnitt 8 ausgelagert.

Natürlich kann man den Erwartungswert

$$\mathbb{E}(W_1) = \frac{1}{p} \tag{5}$$

über die Formel (1) aus*rechnen*, aber was ist damit – insbesondere, wenn man die Identität (2) nicht *begriffen* hat – gewonnen? Kann man nicht *einsehen*, dass der Erwartungswert von W_1 gleich $1/p$ „sein muss“? Hier bieten sich verschiedene, in Henze (2000) vorgestellte Methoden an.

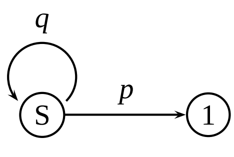

Abb. 1: Zustandsgraph „Warten auf den ersten Treffer“

Zunächst kann man anhand des in Abbildung 1 dargestellten Zustandsgraphen und einer Betrachtung der sich nach dem ersten Versuch einstellenden Situation wie folgt argumentieren: Hat man einen Treffer erhalten, so ist die Realisierung $W_1 = 1$ eingetreten. Unter dieser Bedingung ist also der Erwartungswert von W_1 gleich 1. Ergibt der Versuch jedoch eine Niete, so ist man in der Ausgangssituation, nur mit dem Unterschied, dass ein zusätzlicher Versuch stattfand. Unter der Bedingung, dass der erste Versuch eine Niete ergibt, ist also der Erwartungswert von W_1 gleich $1 + \mathbb{E}(W_1)$. Nach diesem manchmal auch als *Mittelwertregel* oder *Formel vom totalen Erwartungswert* (vgl. Henze (2017), S. 214) bezeichneten Vorgehen folgt

$$\mathbb{E}(W_1) = p \cdot 1 + q \cdot (1 + \mathbb{E}(W_1))$$

und damit (5).

Eine heuristische Vorgehensweise, die an die Deutung des Erwartungswertes als „Durchschnittswert auf lange Sicht" anknüpft, betrachtet den relativen Anteil H_n/n der erzielten Treffer in einer Bernoulli-Kette der Länge n. Bei wachsendem n sollte sich H_n/n gegen die Trefferwahrscheinlichkeit p stabilisieren und damit der als *pro Treffer durchschnittlich benötigte Versuchsanzahl* interpretierbare Kehrwert n/H_n gegen $1/p$. Somit „muss" $\mathbb{E}(W_1) = 1/p$ gelten! Eine weitere, auf Martingalmethoden (siehe Li (1980)) fußende und auf Schulniveau überzeugende heuristische Überlegung betrachtet das Warten auf den ersten Treffer in der Bernoulli-Kette als ein Glücksspiel zwischen Spielern und einer Bank. Vor jedem Versuch setzt ein Spieler einen Euro ein und wettet auf *Treffer*. Im Erfolgsfall erhält er $1/p$ Euro, und das Spiel ist beendet. Andernfalls geht er leer aus, und er oder ein anderer Spieler setzt auf *Treffer*. Das Spiel ist fair, denn die erwartete Auszahlung ist für jeden Spieler gleich $p \cdot 1/p = 1$ Euro. Die Bank muss bei diesem Spiel unabhängig von dessen in der Anzahl der Versuche gemessenen *zufälligen* Dauer W_1 den *festen* Betrag von $1/p$ Euro auszahlen. Da das Spiel fair ist, müssen sich auf die Dauer Einnahmen und Ausgaben ausgleichen, weshalb der Erwartungswert von W_1 gleich $1/p$ sein muss.

Auch die Gleichung

$$\mathbb{V}(W_1) = \frac{1-p}{p^2} \tag{6}$$

für die Varianz kann mithilfe der Formel vom totalen Erwartungswert wie folgt erhalten werden: Es gilt

$$\mathbb{E}(W_1^2) = p \cdot \mathbb{E}\left[W_1^2|1\right] + q \cdot \mathbb{E}\left[W_1^2|0\right]. \tag{7}$$

Dabei bezeichnen $\mathbb{E}\left[W_1^2|1\right]$ und $\mathbb{E}\left[W_1^2|0\right]$ die Erwartungswerte von W_1^2 unter der Bedingung, dass der erste Versuch einen Treffer bzw. eine Niete ergibt. Der erste dieser bedingten Erwartungswerte ist $1 = 1^2$, da sich die Realisierung $W_1 = 1$ ergeben hat. Im zweiten Fall hat man einen vergeblichen Versuch gemacht, was sich dahingehend auswirkt, dass W_1 um Eins zu vergrößern ist. Es gilt also

$$\begin{aligned}\mathbb{E}\left[W_1^2|0\right] &= \mathbb{E}\left[(1+W_1)^2\right] \\ &= 1 + 2 \cdot \mathbb{E}(W_1) + \mathbb{E}(W_1^2).\end{aligned}$$

Zusammen mit (7) und $\mathbb{E}(W_1) = 1/p$ folgt dann mithilfe direkter Rechnung

$$\mathbb{E}(W_1^2) = (2-p)/p^2, \tag{8}$$

woraus sich wegen $\mathbb{V}(W_1) = \mathbb{E}(W_1^2) - (\mathbb{E}(W_1))^2$ Darstellung (6) ergibt.

3 Gedächtnislosigkeit: Ein Charakteristikum der geometrischen Verteilung

Ein tieferes Verständnis für das im nächsten Abschnitt behandelte Warten auf den zweiten oder weitere Treffer stellt sich ein, wenn man sich Folgendes (in einer Rückschau auf ein Spiel wie *Mensch! ärgere dich nicht!* vielleicht sogar mit dem Adverb *schmerzlich* versehen) klar macht: Die Wahrscheinlichkeit für einen Treffer erhöht sich nicht, auch wenn eine noch so lange Serie von Nieten aufgetreten ist. Mathematisch präzisiert betrachten wir die bedingte Wahrscheinlichkeit

$$\mathbb{P}(W_1 = i+m | W_1 \geq i+1), \quad i \geq 1, m \geq 1,$$

im $(i+m)$-ten Versuch den ersten Treffer zu erzielen, wenn man nacheinander i Nieten beobachtet hat, was gleichbedeutend mit $W_1 \geq i+1$ ist. Nach Definition der bedingten Wahrscheinlichkeit $\mathbb{P}(A|B)$ eines Ereignisses A unter der Bedingung, dass ein Ereignis B eintritt, als Quotient $\mathbb{P}(A \cap B)/\mathbb{P}(B)$ und der Tatsache, dass aus dem Ereignis $W_1 = i+m$ das Ereignis $W_1 \geq i+1$ folgt, ergibt sich

$$\begin{aligned}\mathbb{P}(W_1 = i+m | W_1 \geq i+1) &= \frac{\mathbb{P}(W_i = i+m, W_1 \geq i+1)}{\mathbb{P}(W_1 \geq i+1)} \\ &= \frac{\mathbb{P}(W_1 = i+m)}{\mathbb{P}(W_1 \geq i+1)} \\ &= \frac{q^{i+m-1}p}{q^i} = q^{m-1}p \\ &= \mathbb{P}(W_1 = m).\end{aligned}$$

Die Verteilung von W_1 hat also in der Tat „kein Gedächtnis“ . Die Bernoulli-Kette „merkt sich insbesondere nicht“, wie viele Nieten sie „in der Vergangenheit produziert hat“. Vielmehr kann sie in Bezug auf das Warten auf den ersten Treffer „wie neu geboren“ angesehen werden. Die Beziehung

$$\mathbb{P}(W_1 = i+m|W_1 \geq i+1)) = \mathbb{P}(W_1 = m), \ \ i,m \geq 1,$$

ist aber auch in folgendem Sinn *charakteristisch* für die geometrische Verteilung: Ist X eine Zufallsvariable, die die möglichen Werte $1,2,\ldots$ annimmt, und gelten $0 < \mathbb{P}(X = 1) < 1$ sowie

$$\mathbb{P}(X = i+m|X \geq i+1) = \mathbb{P}(X = m), \ \ i,m \geq 1, \quad (9)$$

so gibt es – wie wir gleich sehen werden – ein $p \in (0,1)$ mit

$$\mathbb{P}(X = j) = q^{j-1}p, \quad j \geq 1. \qquad (10)$$

Die Zufallsvariable X hat also die gleiche Verteilung wie W_1. Zum Beweis setzen wir $p := \mathbb{P}(X = 1)$ sowie $i = 1$ in (9). Da aus $X = m+1$ das Ereignis $X \geq 2$ folgt, nimmt (9) nach Definition der bedingten Wahrscheinlichkeit sowie $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 1) = 1 - p$ die Gestalt

$$\frac{\mathbb{P}(X = m+1)}{1-p} = \mathbb{P}(X = m), \quad m \geq 1,$$

an, woraus sich unmittelbar (10) ergibt.

4 Warten auf den zweiten Treffer

Wie in Dorner (2017) kann man schnell die Verteilung der mit W_2 bezeichneten Anzahl der Versuche bis zum Auftreten des zweiten Treffers herleiten. Soll dieser im j-ten Versuch erfolgen, so müssen in den vorangegangenen $j-1$ Versuchen genau ein Treffer und $j-2$ Nieten auftreten. Dabei führt jede der $j-1$ möglichen Auswahlen der Versuchsnummer für den ersten Treffer zur gleichen Wahrscheinlichkeit $q^{j-2}p^2$ für eine konkrete Reihenfolge aus Treffern und Nieten bis zum zweiten Treffer. Es folgt

$$\mathbb{P}(W_2 = j) = (j-1)q^{j-2}p^2, \quad j \geq 2. \qquad (11)$$

Hieraus kann man wie in Dorner (2017) mithilfe von (3) und (4) Erwartungswert und Varianz von W_2 zu

$$\mathbb{E}(W_2) = \frac{2}{p}, \quad \mathbb{V}(W_2) = \frac{2q}{p^2}$$

berechnen (man beachte, dass gegenüber dem Aufsatz Dorner (2017) p und $1-p$ durchweg vertauscht sind). Tiefere *Einsichten* gewinnt man hierdurch sicherlich nicht. Ganz nah liegt natürlich die Frage, *warum* Erwartungswert und Varianz verglichen mit dem Warten auf den ersten Treffer jeweils doppelt so groß sind. Die Antwort erschließt sich unmittelbar begrifflich, wenn man sich klar macht, dass die Anzahl der Versuche bis zum zweiten Treffer die Summe der beiden Zufallsvariablen W_1 und $W_2 - W_1$ ist. Dabei modelliert $W_2 - W_1$ gerade die Anzahl der Versuche, die nach dem ersten Treffer nötig sind, um den zweiten Treffer zu erhalten.

Wir werden jetzt zeigen, dass die Zufallsvariablen W_1 und $W_2 - W_1$ stochastisch unabhängig sind, und dass $W_2 - W_1$ die gleiche Verteilung wie W_1 besitzt. Seien hierzu i, j beliebige natürliche Zahlen. Da die beiden Ereignisse $W_1 = i$ und $W_2 = i + j$ gemeinsam genau dann eintreten, wenn in der Bernoulli-Kette auf $i-1$ Nieten ein Treffer folgt und sich danach $j-1$ Nieten, gefolgt von einem Treffer, einstellen, ergibt sich

$$\begin{aligned} \mathbb{P}(W_1 = i, W_2 - W_1 = j) &= \mathbb{P}(W_1 = i, W_2 = i+j) \\ &= q^{i-1}p \cdot q^{j-1}p \\ &= \mathbb{P}(W_1 = i) \cdot q^{j-1}p. \end{aligned}$$

Summiert man hier über i, so folgt

$$\mathbb{P}(W_2 - W_1 = j) = q^{j-1}p, \quad j \geq 1.$$

Wegen

$$\mathbb{P}(W_1 = i, W_2 - W_1 = j) = \mathbb{P}(W_1 = i)\mathbb{P}(W_2 - W_1 = j)$$

für jede Wahl von i und j sind W_1 und $W_2 - W_1$ in der Tat stochastisch unabhängig, und $W_2 - W_1$ ist verteilt wie W_1. Da die Erwartungswertbildung ganz allgemein additiv ist und Gleiches für die Varianzbildung von *unabhängigen* Zufallsvariablen zutrifft (siehe z. B. Henze (2017), S. 77, S, 167), folgt

$$\begin{aligned} \mathbb{E}(W_2) &= \mathbb{E}(W_1) + \mathbb{E}(W_2 - W_1) = 2\mathbb{E}(W_1) = \frac{2}{p}, \\ \mathbb{V}(W_2) &= \mathbb{V}(W_1) + \mathbb{V}(W_2 - W_1) = 2\mathbb{V}(W_1) = \frac{2q}{p^2}. \end{aligned}$$

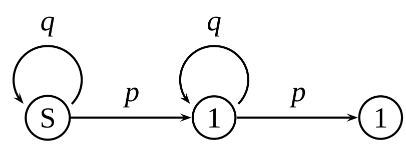

Abb. 2: Zustandsgraph „Warten auf den zweiten Treffer“

Schülerinnen und Schüler gewinnen hier strukturelle Einsichten auch mithilfe des in Abbildung 2 dargestellten Zustandsgraphen, der das Warten auf den zweiten Treffer illustriert. Im Startzustand S verbleibt man solange, wie sich ausschließlich Nieten einstellen, was jeweils mit Wahrscheinlichkeit q zutrifft. Hat man den ersten Treffer erzielt, so erreicht man den mit 1 bezeichneten Nachbarknoten. Da dieser jetzt die Funktion des Startknotens S einnimmt und man sich nur merken muss, wie viele Versuche es gedauert hat, bis man hierhin gelangt ist, wird klar, dass unabhängig von der Anzahl dieser Versuche die Ausgangssituation des Wartens auf den ersten Treffer vorliegt. Am Ende wird die Anzahl der jetzt noch nötigen Versuche bis zum Erreichen des rechts stehenden „End-Knotens" zur bisherigen Versuchsanzahl addiert.

Wir möchten zum Schluss dieses Abschnitts noch einer nahe liegenden Frage nachgehen, die auch zum Verständnis einer Bernoulli-Kette beiträgt. Nehmen wir an, der zweite Treffer sei im j-ten Versuch erzielt worden; es sei also das Ereignis $\{W_2 = j\}$ eingetreten. Wie ist *unter dieser Bedingung* die Nummer des Versuchs für den ersten Treffer verteilt? Schülerinnen und Schüler könnten hier vermuten, dass unter dieser Bedingung die Zufallsvariable W_1 mit gleicher Wahrscheinlichkeit $1/(j-1)$ jeden der Werte $1, 2, \ldots, j-1$ annimmt. Macht man sich klar, dass die Ereignisse $\{W_1 = i\}$ und $\{W_2 = j\}$ genau dann beide eintreten, wenn die Bernoulli-Kette (in dieser Reihenfolge) mit $i-1$ Nieten, einem Treffer, $j-i-1$ Nieten und einem Treffer startet, so bestätigt man obige Vermutung durch direktes Nachrechnen, denn mit (11) folgt für jedes $i = 1, \ldots, j-1$

$$\begin{aligned}\mathbb{P}(W_1 = i | W_2 = j) &= \frac{\mathbb{P}(W_1 = i, W_2 = j)}{\mathbb{P}(W_2 = j)} \\ &= \frac{q^{i-1} p q^{j-i-1} p}{(j-1) q^{j-2} p^2} \\ &= \frac{1}{j-1},\end{aligned}$$

was zu zeigen war. Man beachte, dass diese Wahrscheinlichkeit nicht von p abhängt.

5 Warten auf den r-ten Treffer

Der Binomialkoeffizient $\binom{n}{k}$ beschreibt die Anzahl der Möglichkeiten, aus n Objekten k auszuwählen, wobei die Reihenfolge nicht berücksichtigt wird. Mit dieser Bedeutung von $\binom{n}{k}$ kann man die Verteilung der mit W_r bezeichneten Anzahl der Versuche bis zum r-ten Treffer leicht erhalten. Offenbar nimmt W_r die möglichen Werte $r, r+1, r+2, \ldots$ an. Da das Ereignis $\{W_r = j\}$ für ein $j \geq r$ genau dann eintritt, wenn der j-te Versuch einen Treffer ergibt und unter den davor stattfindenden $j-1$ Versuchen genau $r-1$ Treffer und $j-r$ Nieten auftreten, folgt

$$\mathbb{P}(W_r = j) = \binom{j-1}{r-1} q^{j-r} p^r, \quad j \geq r. \qquad (12)$$

Der Binomialkoeffizient beschreibt dabei die Anzahl der Möglichkeiten, aus den $j-1$ vor dem r-ten Treffer stattfindenden Versuchen diejenigen $r-1$ Versuche für den ersten, zweiten, ..., $(r-1)$-ten Treffer auszuwählen.

Die Verteilung von W_r ist eine um r verschobene negative Binomialverteilung, siehe z. B. Henze (2017), S. 188 ff. Obwohl sie mit relativ elementaren Überlegungen erhalten werden kann, taugt Darstellung (12) auf Schulniveau kaum, um die Normierungbedingung $\sum_{j=r}^{\infty} \mathbb{P}(W_r = j) = 1$ nachzuweisen oder Erwartungswert und Varianz von W_r zu bestimmen. Schon im Fall $r = 3$ zeigt sich, dass man zur Berechnung des Erwartungswertes über die Darstellung

$$\mathbb{E}(W_r) = \sum_{j=3}^{\infty} j \, \mathbb{P}(W_3 = j)$$

wissen muss, was mit der Summe

$$\sum_{j=3}^{k} j(j-1)(j-2) q^{j-3}$$

passiert, wenn k über alle Grenzen wächst. Dass pfiffige Schülerinnen und Schüler nach einem Hinweis durch eine Lehrkraft hier eine Antwort geben können, wird in Abschnitt 8 gezeigt.

Der Schlüssel zum *Verständnis, warum*

$$\mathbb{E}(W_r) = \frac{r}{p}, \qquad \mathbb{V}(W_r) = \frac{rq}{p^2} \qquad (13)$$

gelten, liegt in einer nach den Überlegungen des vorigen Abschnitts offensichtlichen *stochastischen Struktur* des Wartens auf den r-ten Treffer. Bezeichnet allgemein W_j die Anzahl der Versuche bis zum j-ten Treffer, $j = 1, 2, \ldots, r$, so stellt sich W_r in der Form

$$W_r = W_1 + (W_2 - W_1) + \ldots + (W_r - W_{r-1}) \qquad (14)$$

als Summe von Wartezeiten dar. Wie früher steht dabei W_1 für die Zahl der Versuche bis zum ersten Treffer, und $W_i - W_{i-1}$ ist die Zahl der Versuche, die man nach dem $(i-1)$-ten Treffer benötigt, um den i-ten Treffer zu erzielen $(i = 2, \ldots, r)$. Die Zufallsvariablen $W_1, W_2 - W_1, \ldots, W_r - W_{r-1}$ sind stochastisch unabhängig, und sie besitzen alle die gleiche Verteilung wie W_1. Um dieses nach den bisherigen Überlegungen einsichtige Resultat zu beweisen, beschränken wir uns auf den Fall $r = 3$, da der allgemeine Fall nur einen höheren Schreibaufwand bedeutet, aber keine zusätzlichen Erkenntnisse liefert. Für beliebige natürliche Zahlen i,j und k bedeutet das Ereignis $\{W_1 = i, W_2 - W_1 = j, W_3 - W_2 = k\}$, dass die Bernoulli-Kette mit $i-1$ Nieten, einem Treffer, $j-1$ Nieten, einem Treffer, $k-1$ Nieten und wiederum einem Treffer beginnt. Es gilt also

$$\begin{aligned} & \mathbb{P}(W_1 = i, W_2 - W_1 = j, W_3 - W_2 = k) \\ = \; & q^{i-1} p q^{j-1} p q^{k-1} p. \end{aligned} \tag{15}$$

Summiert man hier über i und k, so ergibt sich

$$\begin{aligned} \mathbb{P}(W_2 - W_1 = j) &= q^{j-1} p \\ &= \mathbb{P}(W_1 = j), \qquad j \geq 1, \end{aligned}$$

und Summation über i und j liefert

$$\begin{aligned} \mathbb{P}(W_3 - W_2 = k) &= q^{k-1} p \\ &= \mathbb{P}(W_1 = k), \qquad k \geq 1. \end{aligned}$$

Hieraus folgt die stochastische Unabhängigkeit von W_1, $W_2 - W_1$ und $W_3 - W_2$ sowie die identische Verteilung dieser Zufallsvariablen. Mit dieser Einsicht und der Additivität der Erwartungswertbildung sowie (wegen der stochastischen Unabhängigkeit) auch der Varianzbildung erschließen sich die in (13) stehenden Resultate angesichts von (14) sowie (5) und (6) unmittelbar.

Für Lehrerinnen und Lehrer sei angemerkt, dass aufgrund der Summenstruktur (14) der zentrale Grenzwertsatz von Lindeberg und Lévy (siehe z. B. Brokate u.a. (2016), S.887) Anwendung findet. Mit $\mu := 1/p$ und $\sigma^2 := q/p^2$ gilt für jedes reelle x die Grenzwertaussage

$$\lim_{r\to\infty} \mathbb{P}\left(\frac{W_r - r\cdot\mu}{\sigma\cdot\sqrt{n}} \leq x\right) = \int_{-\infty}^{x} \frac{\mathrm{e}^{-t^2/2}}{\sqrt{2\pi}}\, \mathrm{d}t.$$

Die Wartezeit W_r ist also bei großem r nach Standardisierung approximativ standardnormalverteilt.

Wie am Ende des vorigen Abschnitts möchten wir auch hier die Frage stellen, welche gemeinsame Verteilung die Wartezeiten $W_1, W_2, \ldots, W_{r-1}$ haben, wenn bekannt ist, dass der r-te Treffer im k-ten Versuch aufgetreten und damit das Ereignis $\{W_r = k\}$ eingetreten ist. Intuitiv ist zu erwarten, dass alle $\binom{k-1}{r-1}$ Auswahlen von $r-1$ aller Versuche mit den Nummern $1,2,\ldots,k-1$ hierfür die gleiche Wahrscheinlichkeit besitzen. Die Verteilung wäre also identisch mit der Verteilung der Gewinnzahlen eines „Lottos $r-1$ aus $k-1$“. Auch hier beschränken wir uns beim Nachweis auf den Fall $r = 3$. Seien hierzu i,j,k natürliche Zahlen mit $1 \leq i < j < k$. Da die Ereignisse $\{W_1 = i, W_2 = j, W_3 = k\}$ und $\{W_1 = i, W_2 - W_1 = j - i, W_3 - W_2 = k - j\}$ identisch sind, liefern (15) (mit $j-i$ anstelle von j und $k-j$ anstelle von k) sowie (12) mit $r = 3$ und k anstelle von j

$$\begin{aligned} \mathbb{P}(W_1 = i, W_2 = j | W_3 = k) &= \frac{\mathbb{P}(W_1 = i, W_2 = j, W_3 = k)}{\mathbb{P}(W_3 = k)} \\ &= \frac{q^{i-1} p q^{j-i-1} p q^{k-j-1} p}{\binom{k-1}{2} q^{k-3} p^3} \\ &= \frac{1}{\binom{k-1}{2}}. \end{aligned}$$

Unter der Bedingung $W_3 = k$ ist also die gemeinsame Verteilung von W_1 und W_2 in der Tat identisch mit der Verteilung der beiden Gewinnzahlen bei einem „Lotto 2 aus $k-1$“.

6 Warten auf 01

Dorner (2017) fragt in Abschnitt 3.2 seiner Arbeit nach dem Erwartungswert der mit W_{01} bezeichneten Anzahl der Versuche, bis zum ersten Mal das Muster 01, also eine Niete, gefolgt von einem Treffer, aufgetreten ist. Da das Ereignis $\{W_{01} = k\}$ genau dann eintritt, wenn für ein $i \in \{0,1,\ldots,k-2\}$ die Bernoulli-Kette mit i Treffern, gefolgt von $k-i-1$ Nieten und einem weiteren Treffer, beginnt, kann man die Verteilung von W_{01} relativ leicht gewinnen. Mithilfe von (2) und (3) leitet dann Dorner (2017) Erwartungswert und Varianz von W_{01} zu

$$\mathbb{E}(W_{01}) = \frac{1}{pq}, \quad \mathbb{V}(W_{01}) = \frac{1-3pq}{(pq)^2} \tag{16}$$

her. Auch in diesem Fall wird nicht klar, *warum* der Erwartungswert von W_{01} gleich $1/(pq)$ sein muss. Auf Schulniveau überzeugend ist hier wieder die allgemeiner in Henze (2000) beschriebene und schon beim Warten auf den ersten Treffer vorgestellte Martingalmethode von Li (1980): Das Warten auf das Muster 01 wird als Glücksspiel zwischen Spielern und einer Bank aufgefasst. Vor jedem Versuch setzt ein Spieler einen Euro ein und wettet darauf, dass in den beiden nächsten Versuchen der Bernoulli-Kette die Sequenz 01 auftritt. Kommt im nächsten Versuch eine 1, so ist das Spiel und damit der Einsatz verloren. Andernfalls erhält der Spieler $1/q$ Euro, muss diesen Betrag aber im nächsten Spiel auf 1 setzen. Tritt eine 0 auf, so ist das Spiel ebenfalls verloren. Andernfalls erhält der Spieler $1/(pq)$ Euro. Das Spiel ist fair, denn der Spieler hat einen Euro eingesetzt, und der Erwartungswert seiner Auszahlung beträgt $pq \cdot 1/(pq) = 1$ Euro. Die Bank zahlt bei diesem Spiel den festen Betrag $1/(pq)$ Euro aus, und sie nimmt einen zufälligen Betrag in Höhe von W_{01} Euro ein, denn vor dem letzten, für den vorletzten Spieler zum Gewinn führenden Spiel wird auch ein Euro eingesetzt. Da das Spiel fair ist, müssen sich auf die Dauer Einnahmen und Ausgaben die Waage halten, und somit muss $\mathbb{E}(W_{01}) = 1/(pq)$ gelten.

Eine vielleicht noch einfachere begriffliche Herleitung der Gleichung $\mathbb{E}(W_{01}) = 1/(pq)$ kann über die Mittelwertregel (Formel vom totalen Erwartungswert) erfolgen. Hierzu betrachten wir die beiden möglichen Anfänge 1 und 0 der Bernoulli-Kette. Ist der erste Versuch ein Treffer, so haben wir im Hinblick auf das Erreichen des Musters 01 einen vergeblichen Versuch gemacht, was zur Gleichung

$$\mathbb{E}(W_{01}|1) = 1 + \mathbb{E}(W_{01})$$

führt. Ergibt der erste Versuch jedoch eine Niete, so warten wir von da ab auf den ersten Treffer, und die Wartezeit dafür besitzt den bereits bekannten Erwartungswert $1/p$. Die Mittelwertregel liefert also

$$\begin{aligned} \mathbb{E}(W_{01}) &= p \cdot \mathbb{E}(W_{01}|1) + q \cdot \mathbb{E}(W_{01}|0) \\ &= p \cdot (1 + \mathbb{E}(W_{01})) + q \cdot \frac{1}{p}, \end{aligned}$$

was unmittelbar zum gewünschten Ergebnis führt.

Wendet man die Mittelwertregel auf die Zufallsvariable W_{01}^2 an, so folgt

$$\mathbb{E}(W_{01}^2) = p \cdot \mathbb{E}(W_{01}^2|1) + q \cdot \mathbb{E}(W_{01}^2|0). \tag{17}$$

Auch hier bedeutet ein Treffer zu Beginn, dass die Zufallsvariable W_{01} um Eins vergrößert wird, was zu

$$\begin{aligned} \mathbb{E}(W_{01}^2|1) &= \mathbb{E}\left[(1+W_{01})^2\right] = 1+2\mathbb{E}(W_{01})+\mathbb{E}(W_{01}^2) \\ &= 1 + \frac{2}{pq} + \mathbb{E}(W_{01}^2) \end{aligned} \tag{18}$$

führt. Ergibt der erste Versuch eine Niete, so folgt

$$\mathbb{E}(W_{01}^2|0) = \mathbb{E}(1+W_1)^2 = 1 + 2\mathbb{E}(W_1) + \mathbb{E}(W_1^2),$$

wobei W_1 wie in Abschnitt 2 die Wartezeit auf den ersten Treffer angibt. Mit (5) und (8) ergibt sich

$$\mathbb{E}(W_{01}^2|0) = \frac{p^2+p+2}{p^2}.$$

Setzt man dieses Resultat sowie (18) in (17) ein, so erhält man nach kurzer Rechnung

$$\mathbb{E}(W_{01}^2) = \frac{2-3pq}{(pq)^2}$$

und damit wegen $\mathbb{V}(W_{01}) = \mathbb{E}(W_{01}^2) - (\mathbb{E}(W_{01}))^2$ Darstellung (16) für die Varianz von W_{01}.

Man beachte, dass man im Spezialfall $p = 1/2$ im Mittel vier Versuche auf das Auftreten des Musters 01 wartet. Obwohl in diesem Fall das Muster 11 die gleiche Eintrittswahrscheinlichkeit $1/4$ besitzt, tritt dieses Muster „im Mittel später auf“als 01, denn der Erwartungswert seiner mit W_{11} bezeichneten Wartezeit beträgt 6. Auch hier kann man den Erwartungswert von W_{11} selbst für eine allgemeine Trefferwahrscheinlichkeit p leicht bestimmen, indem man die Fälle, dass die Bernoulli-Kette mit 0 bzw. 10 bzw. 11 beginnt, getrennt betrachtet. Im ersten bzw. zweiten Fall hat man im Hinblick auf das Erreichen von 11 einen bzw. zwei vergebliche Versuche gemacht, und im dritten Fall ist man nach zwei Versuchen am Ziel. Die Mittelwertregel liefert somit

$$\mathbb{E}(W_{11}) = q \cdot (1 + \mathbb{E}(W_{11})) + pq \cdot (2 + \mathbb{E}(W_{11})) + 2p^2,$$

was nach elementarer Rechnung zu

$$\mathbb{E}(W_{11}) = \frac{1+p}{p^2}$$

und damit im Spezialfall $p = 1/2$ zu $\mathbb{E}(W_{11}) = 6$ führt, s. auch Burch (2014). Bezüglich allgemeiner Muster sei z. B. auf Feller (1968), S. 326 ff., Henze (2000) oder Humenberger (2000) verwiesen.

7 Warten auf 11 oder 00

Dorner (2017) thematisiert in Abschnitt 3.1 seiner Arbeit das Warten auf entweder 11 oder 00, also zweimal direkt hintereinander Niete oder Treffer in einer Bernoulli-Kette. Bezeichnet X die Anzahl der Versuche, bis einer dieser beiden Fälle erstmalig eintritt, so ist die Verteilung von X relativ leicht bestimmt, da sich für das Eintreten des Ereignisses $\{X = n\}$ Nullen und Einsen $n-1$ mal abwechseln müssen und sich dann das letzte Symbol wiederholt. Man muss also nur die beiden Fälle unterscheiden, dass der erste Versuch einen Treffer oder eine Niete liefert. Mithilfe der geometrischen Reihe und deren Ableitungen werden dann Erwartungswert und Varianz von X hergeleitet, wobei sich

$$\mathbb{E}(X) = \frac{-p^2+p+2}{p^2-p+1} \tag{19}$$

und

$$\mathbb{V}(X) = \frac{-p \cdot (2p^3 - 4p^2 + 7p - 5)}{(p^2-p+1)^2} \tag{20}$$

ergeben. Angesichts der Symmetrie der Fragestellung hinsichtlich Treffern und Nieten würde man hier

Formeln erwarten, die auch augenscheinlich symmetrisch in p und q sind, also nur vom Produkt pq abhängen. In der Tat gelten

$$\mathbb{E}(X) = \frac{2+pq}{1-pq}, \quad \mathbb{V}(X) = \frac{pq(5-2pq)}{(1-pq)^2}. \qquad (21)$$

In Beantwortung einer Frage von Dorner (2017) zeigen wir jetzt, wie obige Gleichungen auch ohne Verwendung der geometrischen Reihe gewonnen werden können. Der Einfachheit halber beschränken wir uns dabei auf den Erwartungswert von X. Zerlegt man alle Möglichkeiten nach den beiden ersten Versuchsausgängen der Bernoulli-Kette, also 11, 01, 10 und 00, so liefert die Mittelwertregel zunächst

$$\begin{aligned} \mathbb{E}(X) \;&=\; p^2 \cdot \mathbb{E}(X|11) + pq \cdot \mathbb{E}(X|01) \qquad (22) \\ &\quad + pq \cdot \mathbb{E}(X|10) + q^2 \cdot \mathbb{E}(X|00). \end{aligned}$$

Hier gelten

$$\mathbb{E}(X|11) = \mathbb{E}(X|00) = 2,$$

denn im Fall zweier Treffer oder Nieten gleich zu Beginn nimmt X auf jeden Fall den Wert 2 an. Wie kommt man aber an die im Folgenden mit $x := \mathbb{E}(X|01)$ bzw. $y := \mathbb{E}(X|10)$ abgekürzten verbleibenden bedingten Erwartungswerte? Hier führt ein einfaches stochastisches Argument zum Ziel: In der Situation 01, also zuerst eine Niete und dann ein Treffer betrachten wir gedanklich den Ausgang des dritten Versuchs. Ist dieser ein Treffer (was mit Wahrscheinlichkeit p geschieht), so hat sich nach genau drei Versuchen erstmalig 11 oder 00 eingestellt, sodass $\mathbb{E}(X|01)$ mit Wahrscheinlichkeit p den Wert 3 annimmt. Ergibt sich hingegen im dritten Versuch eine Niete, so haben wir nach drei Versuchen die Sequenz 010 beobachtet. Jetzt – und das ist der springende Punkt! – befinden wir uns aber nach einem mitzuzählenden (vergeblichen) Versuch in der Situation, dass die Kette (nach Vergessen des ersten Versuchs!) mit 10 gestartet ist. Diese Überlegung führt zur Gleichung

$$x = p \cdot 3 + q \cdot (1+y).$$

Mit dem gleichen Argument ergibt sich

$$y = q \cdot 3 + p \cdot (1+x).$$

Diese beiden Gleichungen in zwei Unbekannten lassen sich schnell zu

$$x = \frac{1+2p+3q^2+pq}{1-pq}, \; y = \frac{1+2q+3p^2+pq}{1-pq}$$

auflösen. Einsetzen von $x = \mathbb{E}(X|01)$ und $y = \mathbb{E}(X|10)$ in (22) liefert dann nach kurzer Rechnung die in (21) angegebene Darstellung für $\mathbb{E}(X)$. Für die obige stochastische Argumentation kann der in Abbildung 3 dargestelle Zustandsgraph hilfreich sein.

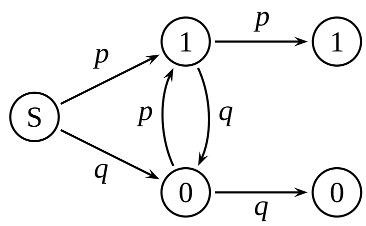

Abb. 3: Zustandsgraph „Warten auf 11 oder 00“

8 Die geometrische Reihe und ihre Ableitungen

Wir zeigen jetzt, dass (1) – (4) mit überschaubarem Aufwand aus nachstehendem Resultat folgen.

Hilfssatz: Es sei (b_n) eine Folge positiver reeller Zahlen. Gibt es ein q mit $0 < q < 1$ und

$$\frac{b_{n+1}}{b_n} \le q$$

für jedes hinreichend große n, so gilt $\lim_{n\to\infty} b_n = 0$.

Gilt nämlich $b_{n+1}/b_n \le q$ für jedes $n \ge N$, so folgt $b_k \le b_N q^{k-N}$ für jedes $k \ge N$ und damit die Behauptung, da q^{k-N} für $k \to \infty$ gegen Null konvergiert.

Schülerinnen und Schüler können (1) leicht begreifen, wenn sie durch Ausmultiplizieren die Gleichung

$$(1+x+x^2+\ldots+x^n)(1-x) = 1-x^{n+1} \qquad (23)$$

herleiten und hierbei das Aha!-Erlebnis eines Teleskop-Effektes erfahren. Da x^{n+1} im Fall $0 < x < 1$ bei wachsendem n gegen Null strebt, folgt (1) unmittelbar. Um (2) einzusehen, könnte man Schülerinnen und Schülern die Frage stellen, was passiert, wenn man die Regeln der Klammerrechnung auf

$$\left(1+2x+3x^2+\ldots+(n+1)x^n\right)(1-x)$$

anwendet. Auch hier gibt es partielle Auslöschungseffekte, und man erhält

$$1+x+x^2+\ldots+x^n - (n+1)x^{n+1},$$

also für $x \neq 1$ zusammen mit (23) das Resultat

$$\sum_{k=1}^{n+1} kx^{k-1} = \frac{1-x^{n+1}}{(1-x)^2} - \frac{(n+1)x^{n+1}}{1-x}. \qquad (24)$$

Setzen wir $b_n = (n+1)x^{n+1}$, so können wir wegen $b_{n+1}/b_n = (1+1/n)x$ im Fall $0 < x < 1$ den Hilfssatz mit $q = (1+x)/2$ anwenden, sodass (2) folgt. Hier müssen Schülerinnen und Schüler nur einsehen, dass die Ungleichung $(1+1/n)x \le (1+x)/2$ gleichbedeutend mit $n \ge 2/(1/x-1)$ ist.

Was passiert nun mit einem Anfangsstück

$$2 \cdot 1 + 3 \cdot 2x + 4 \cdot 3x^2 + \ldots + (n+1)nx^{n-1}$$

der in (3) stehenden Reihe bei Multiplikation mit $1-x$? Elementares Rechnen und Zusammenfassen der jeweiligen Koeffizienten vor gleichen Potenzen von x liefert das Ergebnis

$$2\sum_{k=1}^{n} kx^{k-1} - (n+1)nx^n.$$

Zusammen mit (24) ergibt sich also für jedes x mit $x \neq 1$ die Identität

$$\begin{aligned}\sum_{k=2}^{n+1} k(k-1)x^{k-2} &= \frac{2(1-x^n)}{(1-x)^3} \\ &\quad - \frac{2nx^n}{(1-x)^2} - \frac{(n+1)nx^n}{1-x}.\end{aligned}$$

Nach den bisherigen Überlegungen muss man nur noch den Hilfssatz auf die durch $b_n = (n+1)nx^n$ definierte Folge (b_n) anwenden, sodass auch (3) bewiesen ist. Die Herleitung von (4) macht jetzt ebenfalls keine größeren Schwierigkeiten, denn es gilt

$$\begin{aligned}&\sum_{k=3}^{n} k(k-1)(k-2)x^{k-3}(1-x) \\ = \;& 3\cdot 2\cdot 1 + \sum_{k=4}^{n} k(k-1)(k-2)x^{k-3} \\ &- \sum_{k=3}^{n-1} k(k-1)(k-2)x^{k-2} - n(n-1)(n-2)x^{n-2}.\end{aligned}$$

Mit einer Indexverschiebung in der zu subtrahierenden Summe folgt wegen

$$k(k-1)(k-2) - (k-1)(k-2)(k-3) = 3k(k-1)$$

das Ergebnis

$$\begin{aligned}&\sum_{k=3}^{n} k(k-1)(k-2)x^{k-3}(1-x) \\ = \;& 3\sum_{k=2}^{n-1} k(k-1)x^{k-2} - n(n-1)(n-2)x^{n-2}.\end{aligned}$$

Da man schon weiß, dass die letzte Summe gegen $2/(1-x)^3$ konvergiert, folgt (4) durch Anwendung des Hilfssatzes auf die Folge $b_n = n(n-1)(n-2)x^{n-2}$.

Einer der Gutacher wies darauf hin, dass man (1) und (2) auch mit Mitteln der Stochastik erhalten kann. Da der erste Treffer mit Wahrscheinlichkeit Eins irgendwann auftritt, folgt durch Zerlegung danach, in *welchem* Versuch dies passiert, die Gleichung

$$1 = p + qp + q^2p + q^3p + q^4p + \ldots.$$

Klammert man hier p aus und teilt dann beide Seiten der Gleichung durch p, so folgt (1) mit $x = q$. Geht man des Weiteren von (5) aus, so folgt

$$\frac{1}{p} = 1\cdot p + 2\cdot qp + 3\cdot q^2p + 4\cdot q^3p + \ldots.$$

Klammert man wiederum p aus und dividiert beide Seiten durch p, so hat man auch einen „stochastischen Beweis“ für (2) (mit $x := q$) erhalten.

Danksagung: Ich danke den Gutachtern sowie Reimund Vehling für wertvolle Hinweise.

Literatur

Brokate, M., Henze, N., Hettlich, F., Meister, A., Schranz-Kirlinger, G., und Sonar, Th. (2016): Grundwissen Mathematikstudium. Höhere Analysis, Numerik und Stochastik. Berlin u.a.: Springer Spektrum.

Burch, J. (2014): Erfolge in Serie: Ein rekursiver Ansatz. *Stochastik in der Schule* 34 (2), S. 19.

Dorner, Ch. (2017): Einfache Wartezeitprobleme. *Stochastik in der Schule* 37 (3), S. 8 – 15.

Feller, W, (1968): An Introduction to Probability Theory and Its Applications. Volume 1. Neudruck der 3. Auflage. New York. John Wiley.

Götz, St. (2016): Ein einfaches Wartezeitproblem. *Stochastik in der Schule* 36 (3), S. 16 – 20.

Henze, N. (2000): Muster in Bernoulli-Ketten. *Stochastik in der Schule* 21 (2), S. 2–10.

Henze, N. (2017): Stochastik für Einsteiger. 11. Auflage: Heidelberg. Springer Spektrum.

Humenberger, H. (2000): Kopf-Adler-Muster in Münzwurfserien, unendliche Reihen und Fibonacci-Zahlen. *Stochastik in der Schule* 20, S. 15-22.

Li, S.-Y. R. (1980): Martingale approach to the study of occurrence of sequence patterns in repeated experiments. *Ann. Prob.* 8, S. 1171-1176.

Anschrift des Verfassers:

Prof. Dr. Norbert Henze
Institut für Stochastik
Karlsruher Institut für Technologie (KIT)
Englerstr. 2
76131 Karlsruhe
Henze@kit.edu