Eckpunkte einer Stochastikausbildung für Lehramtsstudierende

NORBERT HENZE, KARLSRUHE

Zusammenfassung: *Dieser Aufsatz stellt die Ausarbeitung eines gleichnamigen Vortrages dar, den ich anlässlich der Jubiläums-Herbsttagung des AK Stochastik der Gesellschaft für Didaktik der Mathematik am 10.12.2022 gehalten habe. Die Eckpunkte gehen von der Prämisse aus, dass universitäre Studienpläne für Studierende des* gymnasialen *Lehramts Mathematik nur eine einzige verpflichtende* fachwissenschaftliche *Vorlesung über Stochastik vorsehen. Die folgenden Ausführungen beziehen sich auf eine Lehrveranstaltung im Umfang von 4 Semesterwochenstunden mit zweistündigen Übungen und einem zweistündigen Tutorium. Eine solche, speziell für Studierende des gymnasialen Lehramts konzipierte Mathematik-Vorlesung gibt es am Karlsruher Institut für Technologie (KIT) seit dem Jahr 2013.*

1 Einführung

Unter der Leitidee Daten und Zufall nimmt die Stochastik mittlerweile einen festen Platz in den Mathematik-Curricula deutscher Gymnasien ein. Ein Teil der Lehrkräfte hat jedoch im Studium nie eine einführende Vorlesung in die Stochastik gehört. Für viele Lehrkräte war eine solche Veranstaltung zwar verpflichtend, jedoch hörten sie diese im Studium zusammen mit Kommilitoninnen und Kommilitonen, deren Studienziel letztlich ein Master in Mathematik war. Wenn – was manchmal der Fall ist – eine solche Lehrveranstaltung Sigma-Algebren und Maßtheorie zu breiten Raum gibt, besteht die Gefahr, dass sie im Hinblick auf einen lebendigen, nicht überwiegend an Rezepten ausgerichteten Stochastikunterricht von vergleichsweise beschränktem Nutzen ist. Im Folgenden führe ich aus, welche Inhalte meines Erachtens nach für eine einzige obligatorische Vorlesung als Einführung in die Stochastik für Studierende des gymnasialen Lehramts Mathematik unverzichtbar, wünschenswert und illusorisch sind.

Eine derartige Vorlesung findet nach einer Grundausbildung in Analysis und Linearer Algebra oft im dritten oder vierten Fachsemester statt. Wenn man eine solche Vorlesung hält, muss einem klar sein, dass von der Schule her oft nur Rezept-Kenntnisse in Stochastik vorhanden sind.

Fast jeder der folgenden Abschnitte trägt als Überschrift Grundbegriffe, die in einer einführenden Stochastikvorlesung hinreichend motiviert und diskutiert werden sollten. Unter bildungspolitischen Gesichtspunkten ist es bemerkenswert, dass substanzielle Teile der Vorlesung im Niveau nicht über das hinausgehen, was im Schulbuch Barth und Haller (1985) zu finden ist.

2 Grundräume, Ereignisse

Den Anfang machen *Grundräume* oder synonym *Ergebnisräume* sowie *Ereignisse* als Teilmengen dieser Grundräume. Grundräume werden – wie auch früher in Schulbüchern, s. z.B. Glaser et al. (1982) – üblicherweise mit dem Buchstaben Ω gekennzeichnet. Wenn das Ergebnis *eines* Würfelwurfs beschrieben werden soll, ist es selbstredend, $\Omega := \{1,2,3,4,5,6\}$ zu setzen, aber wie sollte man die Information, die das linke Bild in Abb. 1 vermittelt, mathematisch kompakt notieren?

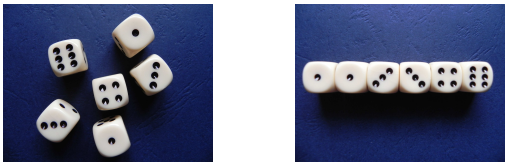

Abb. 1: Sechs Würfel wurden geworfen

Für dieses Bild wurden sechs nicht unterscheidbare Würfel gleichzeitig geworfen. Man sieht, dass zwei der Würfel eine Eins und zwei eine Drei zeigen. Einer der Würfel zeigt eine Vier und einer eine Sechs. *Eine* Möglichkeit, diese Information mathematisch zu notieren, besteht darin, die einzelnen Augenzahlen der Größe nach zu sortieren –was zum rechten Bild in Abb. 1 führt – und als Grundraum die Menge

$$\Omega := \{(b_1,\ldots,b_6) : 1 \le b_1 \le b_2 \le \ldots \le b_6 \le 6\}$$

aller 6-Tupel aus Zahlen von 1 bis 6 zu wählen, deren Komponenten nach aufsteigender Größe sortiert sind. Die durch das Bild vermittelte Information steckt dann im 6-Tupel $(1,1,3,3,4,6)$. Studierende sollten auch erfahren, wie wichtig Tupel als Darstellungsmittel für mehrstufige stochastische Vorgänge sind. So ist das kartesische Produkt $\Omega := \Omega_1 \times \ldots \times \Omega_n$, also die Menge aller n-Tupel $\omega = (a_1,\ldots,a_n)$ mit $a_j \in \Omega_j$ für jedes $j \in \{1,\ldots,n\}$, ein natürlicher Grundraum für einen n-stufigen stochastischen Vorgang, bei dem die Ergebnisse der j-ten Stufe durch den Grundraum Ω_j dargestellt werden.

Im Zusammenhang mit Ereignissen ist wichtig, mengentheoretische und damit korrespondierende logische Verknüpfungen zu thematisieren. So bedeutet etwa die *de Morgansche Regel*

$$(A \cup B)^c = A^c \cap B^c$$

für Ereignisse A und B in Worten, dass genau dann nicht mindestens eines der Ereignisse A und B eintritt, wenn keines dieser Ereignisse, also weder A noch B, eintritt. Dabei bezeichnet allgemein D^c das zu einem Ereignis D komplementäre Ereignis. Hier ist im Schulbereich die Schreibweise $\overline{D}$ üblich.

3 Zufallsvariablen

Der nächste Grundbegriff ist der einer *Zufallsvariable* oder synonym *Zufallsgröße*. Eine Zufallsvariable X ist eine Abbildung $X : \Omega \to \mathbb{R}$. Diese Definition findet man so in früheren Schulbüchern wie etwa auf S. 73 in Glaser et al. (1982). Ich sage den Studierenden aber auch, dass im Laufe der Entfachlichung der Schulmathematik jetzt Formulierungen wie etwa diese zu finden sind: „Ist die Größe, für die man sich interessiert, eine reelle Zahl, so nennt man diese Zufallsgröße“ (s. Biehler u.a. (2012), S. 108). Die einfachste Zufallsvariable, die zwei verschiedene Werte annehmen kann, ist die durch

$$\mathbf{1}_A(\omega) := \begin{cases} 1, \text{ falls } \omega \in A, \\ 0, \text{ sonst,} \end{cases}$$

definierte *Indikatorfunktion* eines Ereignisses $A \subset \Omega$. Eine Realisierung von $\mathbf{1}_A$ zeigt an, ob das Ereignis A eingetreten ist oder nicht. Sind $A_1, \ldots, A_n$ beliebige Ereignisse, so geben die Realisierungen der auch *Zählvariable* genannten *Indikatorsumme*

$$\mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n} \tag{1}$$

an, wie viele der Ereignisse $A_1, \ldots, A_n$ eintreten. Diese Situation hat man speziell beim Ziehen mit oder ohne Zurücklegen aus einer Urne mit roten und schwarzen Kugeln, wobei das Ereignise A_j für das Ziehen einer roten Kugel im j-ten Zug steht, vgl. Abschn. 8. Die große Bedeutung von Indikatorfunktionen und Indikatorsummen wurde schon vor 50 Jahren von Engel (1973) in einem Buch, das für „Lehrer, Schüler und Freunde der Mathematik“ gedacht war, betont. Ein wichtiger Aspekt von Zufallsvariablen besteht darin, dass sie ganz allgemein Ereignisse beschreiben. So ist etwa $\{X \leq t\} := \{\omega \in \Omega : X(\omega) \leq t\}$ eine Kurzschreibweise für die Menge aller Elemente im Grundraum, für die die Realisierung von X einen Wert kleiner oder gleich t ergibt.

4 Diskrete Wahrscheinlichkeitsräume

In meiner Vorlesung ist dann der nächste Grundbegriff der eines *diskreten Wahrscheinlichkeitsraums* (kurz: *diskreter W-Raum*). Dabei motiviere ich die Kolmogorovschen Axiome über das empirische Gesetz über die Stabilisierung relativer Häufigkeiten.

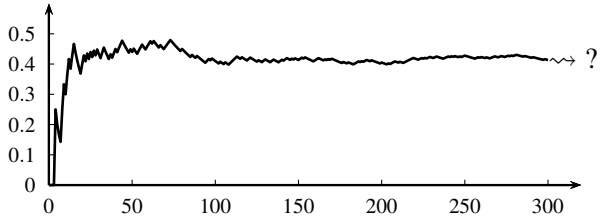

Abb. 2: Ergebnis eines Reißzweckenversuchs

So zeigt Abb. 2 das Ergebnis eines selbst durchgeführten Reißzweckenversuchs. Insgesamt wurde eine Reißzwecke 300-mal geworfen, und aufgetragen sind die fortlaufend notierten relativen Häufigkeiten der Ergebnisse, dass die Spitze der Reißzwecke oben lag. Man sieht, dass sich diese relativen Häufigkeiten bei wachsender Anzahl der Würfe stabilisieren, aber wogegen? Ich motiviere jetzt die Axiome

a) $\mathbb{P}(A) \geq 0$ (Nichtnegativität),

b) $\mathbb{P}(\Omega) = 1$ (Normierung),

c) $\mathbb{P}\left(\biguplus_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$ (σ-Additivität),

die an ein Wahrscheinlichkeitsmaß $\mathbb{P}$ als reellwertige Funktion auf dem System aller Teilmengen von Ω gestellt werden, anhand relativer Häufigkeiten. Dabei bedeutet das Pluszeichen innerhalb des Vereinigungszeichens, dass die Ereignisse $A_1, \ldots, A_n$ paarweise disjunkt sind. Für die in Abb. 2 dargestellten Fragezeichen sollten als Funktion der jeweils betrachteten Ereignisse die Eigenschaften a), b) und c) quasi als „Spielregeln“ gelten, denn sie gelten für relative Häufigkeiten, und zwar ganz egal, auf wie vielen Wiederholungen eines stochastischen Vorgangs unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen diese relativen Häufigkeiten fußen.

Natürlich sind für die Schule insbesondere *endliche W-Räume*, bei denen Ω eine endliche Menge ist, zentral, und hier sticht der *Laplacesche W-Raum* heraus. Diesen kennzeichnet, dass die Wahrscheinlichkeit eines Ereignisses A als der Quotient aus der Anzahl der für das Eintreten von A günstigen Fälle und der Anzahl aller möglichen Fälle *begrifflich interpretierbar* ist. Anhand der Frage, wie viele Versuche benötigt werden, um in einer Folge unabhängiger Bernoulli-Versuche den ersten Treffer zu erzielen, wird aber

auch klar, dass Grundräume mit abzählbar-unendlich vielen Elementen wie z.B. $\Omega = \{1, 01, 001, 0001, \dots\}$ auftreten.

In einem diskreten W-Raum ist Ω eine *abzählbare* (d.h. endliche oder abzählbar-unendliche) Menge. Nur im letzteren Fall, für den $\Omega =: \{\omega_1, \omega_2, \dots\}$ gesetzt sei, treten technische Probleme auf, wenn man mit einer Folge $(p_j)_{j\geq 1}$ nicht negativer reeller Zahlen, die die Bedingung $\sum_{j=1}^{\infty} p_j = 1$ erfüllen, startet und versucht, ein Wahrscheinlichkeitsmaß (kurz: *W-Maß*) $\mathbb{P}$ zu konstruieren, indem man

$$\mathbb{P}(A) := \sum_{j\geq 1 : \omega_j \in A} p_j, \qquad A \subset \Omega,$$

und damit insbesondere $p_j =: \mathbb{P}(\{\omega_j\})$, $j \geq 1$, setzt. Damit das so definierte $\mathbb{P}$ das Axiom der σ-Additivität erfüllt, benötigt man den sog. *großen Umordnungssatz für Reihen*, der in einer Analysis-Vorlesung oft nicht vorkommt. Ein Erklärvideo hierzu ist Henze (2019).

In der Vorlesung werden dann aus den Axiomen a)-c) diverse Folgerungen hergeleitet. Eine davon ist die Gleichung $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, eine andere die Formel des Ein- und Ausschließens für die Wahrscheinlichkeit der Vereinigung von Ereignissen.

5 Verteilung einer Zufallsvariablen

Der nächste Grundbegriff ist der einer *Verteilung* einer Zufallsvariablen X. Im bisherigen Rahmen kann X nur endlich oder abzählbar-unendlich viele Werte annehmen. Setzt man für jede reelle Zahl x

$$\mathbb{P}(X = x) := \mathbb{P}(\{X = x\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}),$$

so gilt $\mathbb{P}(X = x)$ nur für endlich viele oder abzählbar-unendlich viele Werte von x. An dieser Stelle der Vorlesung definiere ich das System solcher Paare $(x, \mathbb{P}(X = x))$ als die *Verteilung von X*, obwohl später in einem allgemeineren Rahmen die Verteilung von X als W-Maß auf der σ-Algebra der Borelmengen über $\mathbb{R}$ eingeführt wird (s. Abschn. 23). Die Funktion $\mathbb{R} \ni x \mapsto \mathbb{P}(X = x)$ wird meist *Wahrscheinlichkeitsfunktion von X* genannt.

Verteilungen von (diskret verteilten) Zufallsvariablen können mithilfe von Stabdiagrammen veranschaulicht werden. Abb. 3 zeigt ein solches Diagramm für eine Zufallsvariable, die fünf verschiedene Werte mit jeweils positiven Wahrscheinlichkeiten annehmen kann. Zusätzlich ist der Erwartungswert von X (s. Abschn. 7) eingezeichnet.

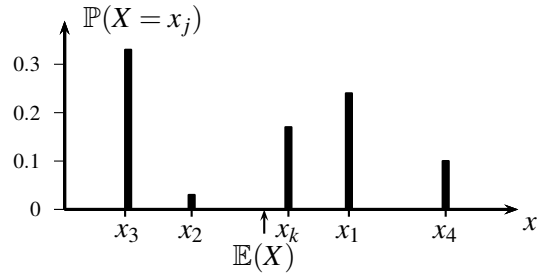

Abb. 3: Stabdiagramm der Verteilung von X

Allgemein ist das Stabdiagramm eine Veranschaulichung der Wahrscheinlichkeitsfunktion von X. Obwohl bei einer diskret verteilten Zufallsvariable nur einzelne Werte mit positiven Wahrscheinlichkeiten belegt werden und in diesem Zusammenhang keinerlei gedankliche Assoziation an *Flächen* entstehen sollte, findet man in Schulbüchern heute ausnahmslos *Histogramme* zur Darstellung der Verteilung von Zufallsvariablen, die ganzzahlige Werte annehmen. Andere Zufallsvariablen treten nicht (mehr) auf, s. hierzu Henze und Vehling (2019).

6 Kombinatorik, Urnen- und Fächer-Modelle

Im Hinblick auf Laplace-Modelle sind dann die nächsten Vorlesungsthemen Grundformeln der *Kombinatorik* sowie *Urnen-* und *Fächer-Modelle*. Zentral sind hier die *Multiplikationsregel der Kombinatorik* sowie *Permutationen* und *Kombinationen mit* und *ohne Wiederholung*. Dabei kann die grundlegende Bedeutung der Multiplikationsregel nicht hoch genug eingeschätzt werden. Nach dieser Regel gibt es $j_1 \cdot \ldots \cdot j_k$ k-Tupel der Form $(a_1, \dots, a_k)$, wenn man für jedes $i \in \{1, \dots, k\}$ die i-te Stelle a_i des Tupels auf genau j_i Weisen besetzen kann.

Der *Binomialkoeffizient* $\binom{n}{k}$ (den man „k aus n“ lesen sollte) ist begrifflich definiert als Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Ganz nah an dieser Definition bewegt man sich, wenn man im Hinblick auf die Schule $\binom{n}{k}$ als Anzahl binärer n-Tupel mit genau k Einsen einführt. So steht etwa das binäre 5-Tupel $(1,0,0,1,0)$ für die aus den Zahlen 1 und 4 bestehende Teilmenge der Zahlen von 1 bis 5. Schreiben wir über die Komponenten des Tupels fünf verschiedene, von 1 bis 5 nummerierte Objekte, so zeigt das obige Tupel an, dass die Objekte 1 und 4 ausgewählt wurden. Unverzichtbar für die Vorlesung sind auch begriffliche Beweise der Rekursionsformel für die Binomialkoeffizienten sowie für den allgemeinen binomischen Lehrsatz, der früher in Schulbüchern vorkam (s. hierzu Übungsaufgabe(!) 41 auf S. 115 in Barth und Haller

(1985) sowie Henze (2023). Um geistig flexibel zu bleiben, sollte auch die begriffliche Gleichwertigkeit von Urnen- und Fächer-Modellen thematisiert werden. Anstatt eine Kugel mit der Nummer j zu ziehen kann man auch ein Fach mit der Nummer j besetzen und umgekehrt. Auch das Paradoxon der überraschend frühen ersten Kollision beim rein zufälligen sequentiellen Besetzen von Fächern (Stichwort: Geburtstagsproblem) sollte man kennen.

7 Erwartungswert

Der nächste Grundbegriff der Stochastik ist der des *Erwartungswertes* einer Zufallsvariablen X. Ich gehe bei der Motivation dieses Begriffs auf dessen historische Wurzeln zurück. Ein stochastischer Vorgang habe die möglichen Ergebnisse $\omega_1,\ldots,\omega_s$, die mit den zugehörigen Wahrscheinlichkeiten $\mathbb{P}(\{\omega_1\}),\ldots,\mathbb{P}(\{\omega_s\})$ auftreten. Tritt ω_j auf, so erhält man den Geldbetrag $X(\omega_j)$ (in Euro) ausbezahlt, $j\in\{1,\ldots,s\}$. Wird der Vorgang n-mal unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen durchgeführt, und tritt dabei h_j-mal der Ausgang ω_j auf, so ist

$$\sum_{j=1}^{s} X(\omega_j)\cdot\frac{h_j}{n}$$

der *durchschnittliche Geldbetrag pro Vorgang*. Da sich die relativen Häufigkeiten h_j/n bei wachsendem n gegen die Wahrscheinlichkeiten $\mathbb{P}(\{\omega_j\})$ stabilisieren sollten, lässt sich der durch

$$\mathbb{E}(X) := \sum_{j=1}^{s} X(\omega_j)\cdot\mathbb{P}(\{\omega_j\})$$

definierte *Erwartungswert* von X als begründete Prognose für den auf die Dauer erhaltenen durchschnittlichen Wert von X pro stochastischem Vorgang deuten. Da die Studierenden die nötigen Kenntnisse aus der Analysis mitbringen, kann man diese Definition problemlos auf den Fall eines diskreten W-Raums mit abzählbar-unendlichem Grundraum Ω erweitern, indem

$$\mathbb{E}(X) := \sum_{j=1}^{\infty} X(\omega_j)\cdot\mathbb{P}(\{\omega_j\}) \qquad (2)$$

gesetzt und vereinbart wird, dass die hier auftretende unendliche Reihe *absolut* konvergieren soll. Ihr Wert hängt dann nicht von der konkreten Abzählung der Elemente von Ω ab. Gleichung (2) findet sich für den Fall eines endlichen W-Raums z.B. auf S. 172 in Barth und Haller (1985).

Aus (2) folgen unmittelbar die Linearität

$$\mathbb{E}(aX+bY) = a\mathbb{E}(X)+b\mathbb{E}(Y),\quad a,b\in\mathbb{R},$$

der Erwartungswertbildung sowie die Gleichung $\mathbb{E}(\mathbf{1}_A)=\mathbb{P}(A)$ für den Erwartungswert einer Indikatorfunktion. Damit ist der Erwartungswert einer Zählvariablen wie in (1) (ohne Kenntnis deren Verteilung!) gleich $\sum_{j=1}^{n}\mathbb{P}(A_j)$ und folglich gleich np, wenn die Ereignisse sämtlich die Wahrscheinlichkeit p besitzen. Nimmt X nur endlich viele Werte $x_1,\ldots,x_k$ an, so ergibt sich aus (2) durch Sortieren der Summanden nach gleichen Werten für $X(\omega)$ die momentan in der Schule als Definition des Erwartungswertes dienende *Darstellungsformel*

$$\mathbb{E}(X) = \sum_{j=1}^{s} x_j\cdot\mathbb{P}(X=x_j).$$

Was auf keinen Fall fehlen sollte, ist die Interpretation des Erwartungswertes als *physikalischer Schwerpunkt* beim Stabdiagramm der Verteilung von X, s. Abb. 3 und S. 172 in Barth und Haller (1985).

8 Binomial- und hypergeometrische Verteilung

Als Anwendung zeige ich dann auf, dass Zählvariablen im Zusammenhang mit dem rein zufälligen n-maligen Ziehen aus einer Urne, die r rote und s schwarze Kugeln enthält, je nach Ziehungsmodus eine Binomial- oder eine hypergeometrische Verteilung besitzen. Hier geht es um konkrete Modellierung, um fachlich auf sicherem Boden zu stehen. *Eine* Möglichkeit besteht darin, die Kugeln gedanklich von 1 bis $r+s$ durchzunummerieren und den roten Kugeln die Nummern 1 bis r und den schwarzen die Nummern $r+1$ bis $r+s$ zuzuweisen. Zieht man mit Zurücklegen, so ist die Menge $\mathrm{Per}_n^{r+s}(\mathrm{mW})$ aller n-Tupel $(a_1,\ldots,a_n)$ mit Komponenten aus $\{1,\ldots,r+s\}$ ein kanonischer Grundraum für diesen stochastischen Vorgang. In diesem Grundraum steht $A_j := \{(a_1,\ldots,a_n)\in\Omega : a_j\le r\}$ für das Ereignis, dass beim j-ten Zug eine rote Kugel gezogen wird. Bezeichnet $\mathbb{P}$ die Gleichverteilung auf Ω, so ergibt sich mithilfe der Multiplikationsregel der Kombinatorik

$$\mathbb{P}(A_j) = \frac{r}{r+s} =: p,\quad j\in\{1,\ldots,n\}, \qquad (3)$$

und somit besitzt die zufällige Anzahl $X := \sum_{j=1}^{n}\mathbf{1}_{A_j}$ der Male, bei denen eine rote Kugel gezogen wird, den Erwartungswert np. Die Binomialverteilung $\mathrm{Bin}(n,p)$ von X, also

$$\mathbb{P}(X=k) = \binom{n}{k}p^k(1-p)^{n-k},\quad k\in\{0,\ldots,n\},$$

wird erst danach hergeleitet. Zieht man *ohne* Zurücklegen, so ändert sich formal wenig. Der Grundraum Ω ist jetzt die Menge $\mathrm{Per}_n^{r+s}(\mathrm{oW})$ aller n-Permutationen der Zahlen $1,\ldots,r+s$ *ohne* Wiederholung und damit die Menge aller n-Tupel $(a_1,\ldots,a_n)$ wie oben, nur mit dem Unterschied, dass jetzt $a_1,\ldots,a_n$ *paarweise verschieden* sein müssen. Definiert man $A_1,\ldots,A_n$ sowie $X := \sum_{j=1}^n \mathbf{1}_{A_j}$ wie oben, und schreibt man (ebenfalls in unveränderter Notation) $\mathbb{P}$ für die Gleichverteilung auf Ω, so gilt unter Verwendung der Multiplikationsregel der Kombinatorik ebenfalls (3). Kombinatorische Überlegungen liefern weiter

$$\mathbb{P}(X=k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, \quad k \in \{0,1,\ldots,n\},$$

und damit eine hypergeometrische Verteilung für X.

9 Modellierung mehrstufiger stochastischer Vorgänge

Das nächste Vorlesungsthema ist die Modellierung mehrstufiger stochastischer Vorgänge mithilfe von *Start-* und *Übergangswahrscheinlichkeiten* sowie der sog. *ersten Pfadregel*. Der rote Faden ist hier das *Pólyasche Urnenschema*, welches von einer Urne mit r roten und s schwarzen Kugeln ausgeht. Aus dieser wird sukzessive rein zufällig gezogen. Dabei legt man nach jedem Zug die jeweils gezogene Kugel *zusammen mit c weiteren Kugeln derselben Farbe* in die Urne und mischt den Urneninhalt neu, bevor die nächste Ziehung erfolgt. Insgesamt wird n-mal gezogen, und die interessierende Zufallsvariable X ist die Anzahl der gezogenen roten Kugeln.

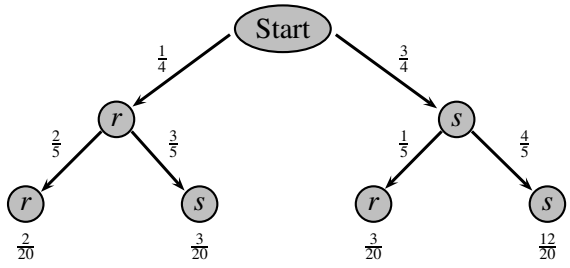

Abb. 4: Baumdiagramm zum Pólyaschen Urnenmodell

Im Hinblick auf die Schule darf das Baumdiagramm als wichtiges Strukturierungselement mehrstufiger Vorgänge nicht fehlen. Abbildung 4 zeigt ein solches Diagramm für das zweimalige Ziehen beim Pólyaschen Urnenschema mit $r=1$, $s=3$ und $c=1$.

Der Grundraum Ω für einen zweistufigen stochastischen Vorgang mit den Grundräumen Ω_1 und Ω_2 für die beiden Stufen ist die Menge $\Omega = \Omega_1 \times \Omega_2$ aller Paare $\omega := (a_1,a_2)$ mit $a_1 \in \Omega_1$ und $a_2 \in \Omega_2$. Die *Start-Verteilung* definiert Wahrscheinlichkeiten $p_1(a_1) \geq 0$ mit $\sum_{a_1\in\Omega_1} p_1(a_1) = 1$ für die Ergebnisse der ersten Stufe. Liefert die erste Stufe das Resultat a_1, so definiert die *Übergangswahrscheinlichkeit* $p_1(a_2|a_1)$ die Wahrscheinlichkeit dafür, dass *unter dieser Bedingung* bei der zweiten Stufe das Ergebnis a_2 auftritt. Mithilfe der aufgrund von Prozentrechnung motivierten sog. *ersten Pfadregel* definieren dann

$$\mathbb{P}(\{\omega\}) := p_1(a_1)\cdot p_2(a_2|a_1), \quad \omega = (a_1,a_2) \in \Omega,$$

sowie $\mathbb{P}(A) := \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})$ für $A \subset \Omega$ ein W-Maß $\mathbb{P}$ auf den Teilmengen von Ω. Die letzte Gleichung wird in der Schule *zweite Pfadregel* genannt. Die Modellierung von mindestens dreistufigen stochastischen Vorgängen erfolgt induktiv.

In der Vorlesung lernen die Studierenden dann die Pólya-Verteilung als Verallgemeinerung sowohl der Binomialverteilung als auch der hypergeometrischen Verteilung kennen. Der zu Beginn dieses Abschnitts eingeführte Parameter c darf nämlich auch gleich null bzw. negativ sein. Im ersten Fall wird die gezogene Kugel zurückgelegt, im zweiten werden nach jedem Zug Kugeln entnommen. Der Urneninhalt muss dann natürlich hinreichend groß sein. Ist c gleich minus eins, so erfolgt das Ziehen ohne Zurücklegen. Interessanterweise ist die Wahrscheinlichkeit, im k-ten Zug eine rote Kugel zu ziehen, unabhängig von k und von c gleich $\frac{r}{r+s}$. Diese Tatsache löst bei Studierenden stets einen großen Aha!-Effekt aus. Ein intuitives Verständnis hierfür liefert das Erklärvideo Henze (2020). Konkrete Unterrichtsvorschläge zum Pólyaschen Urnenmodell finden sich in Kap. 7 von Henze et al. (2021).

10 Bedingte Wahrscheinlichkeiten

Einem logischen Aufbau folgend geht es jetzt um bedingte Wahrscheinlichkeiten. Um die Definition der bedingten Wahrscheinlichkeit eines Ereignisses A unter der Bedingung, dass ein Ereignis B eintritt, zu motivieren, drängt es sich geradezu auf, gedanklich eine n-malige Durchführung eines stochastischen Vorgangs unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen zu betrachten. Bezeichnen $h_n(B)$ die Anzahl der Male, bei denen dabei das Ereignis B eintritt und $h_n(A\cap B)$ die Anzahl der Male, bei denen (zusätzlich zu B) auch noch A eintritt, so steht der Quotient

$$\frac{h_n(A\cap B)}{h_n(B)}$$

für den relativen Anteil unter allen Fällen, in denen B eintritt, in denen dann auch noch A eintritt. Teilt man hier Zähler und Nenner durch n, so führt das empirische Gesetz über die Stabilisierung relativer Häufigkeiten „geradezu zwangsläufig“ zur Definition

$$\mathbb{P}_B(A) := \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{falls } \mathbb{P} > 0, \quad (4)$$

der *bedingten Wahrscheinlichkeit von A unter der Bedingung B*.

Die Studierenden lernen zunächst kennen, dass Gleichung (4) mit Blick auf Anwendungen die Gestalt

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) \quad (5)$$

annimmt und die bedingte Wahrscheinlichkeit $\mathbb{P}(A|B)$ als *Modellbaustein* in Form einer Übergangswahrscheinlichkeit wie in Abschn. 9 *gegeben ist*. Im Fall einelementiger Mengen A und B ist (5) nichts anderes als die erste Pfadregel. Sie erfahren auch, dass die *Formel von der totalen Wahrscheinlichkeit*

$$\mathbb{P}(B) = \sum_{j=1}^{n} \mathbb{P}(A_j)\mathbb{P}(B|A_j)$$

aus einer Fallunterscheidung nach den sich paarweise ausschließenden Ereignissen $A_1, \ldots, A_n$ mit $\Omega = A_1 \cup \ldots \cup A_n$ herrührt und in der Schule *zweite Pfadregel* genannt wird, vgl. Abb. 5.

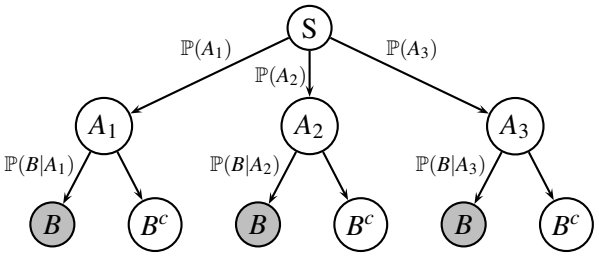

Abb. 5: Zur Formel von der totalen Wahrscheinlichkeit

Weitere wichtige Themen im Zusammenhang mit bedingten Wahrscheinlichkeiten sind die zwar mathematisch banale, aber im Hinblick auf Anwendungen wie etwa das maschinelle Lernen grundlegende *Bayes-Formel* sowie (unter dem Schlagwort *Datenkompetenz*) das Problem einer korrekten Interpretation positiver Resultate bei medizinischen Tests sowie das *Simpson-Paradoxon*. Ein aktuelles Beispiel zu diesem Paradoxon im Zusammenhang mit Todesfallraten bei Covid-19-Infektionen findet sich in Henze (2021a). Ausgiebig diskutiert werden sollte auch die prinzipielle Schwierigkeit, quasi „beiläufig erhaltene“ Teil-Informationen über den Ausgang eines stochastischen Vorgangs wie etwa beim *Drei-Türen-Problem* oder beim *Zwei-Jungen-Problem* in ein adäquates stochastisches Modell zu integrieren.

11 Stochastische Unabhängigkeit

Der nächste Grundbegriff ist der der *stochastischen Unabhängigkeit*. Ereignisse $A_1, \ldots, A_n$ heißen (*stochastisch*) *unabhängig*, falls für jede Auswahl von mindestens zwei dieser Ereignisse die Wahrscheinlichkeit von deren Durchschnitt gleich dem Produkt der Einzelwahrscheinlichkeiten dieser Ereignisse ist, wenn also für jede mindestens zweielementige Teilmenge T von $\{1, \ldots, n\}$ die Gleichung

$$\mathbb{P}\left(\bigcap_{j \in T} A_j\right) = \prod_{j \in T} \mathbb{P}(A_j)$$

erfüllt ist. Diese insgesamt $2^n - n - 1$ Gleichungen finden sich in anderer Notation in Übungsaufgabe 45 auf S. 163 in Barth und Haller (1985). Die Aufgabe besteht darin, die Äquivalenz dieser Gleichungen zu einer auf S. 156 gegebenen Definition der Unabhängigkeit nachzuweisen.

In der Vorlesung wird die Definition der Unabhängigkeit von Ereignissen ausgiebig motiviert und vor allem in Bezug auf bedingte Wahrscheinlichkeiten und reale Beeinflussung diskutiert. Instruktive Beispiele zeigen, dass aus der paarweisen stochastischen Unabhängigkeit von Ereignissen im Allgemeinen nicht auf deren Unabhängigkeit geschlossen werden kann, und dass etwa aus der einen Gleichung $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$ nicht notwendig die Unabhängigkeit von A_1, A_2 und A_3 folgt.

Nicht fehlen darf in diesem Zusammenhang die allgemeine Erzeugungsweise der Binomialverteilung: Sind $A_1, \ldots, A_n$ *irgendwelche* unabhängigen Ereignisse, die alle die gleiche Wahrscheinlichkeit p besitzen, so hat die Indikatorsumme $\mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_n}$ die Binomialverteilung $\mathrm{Bin}(n, p)$. Als Hintergrundwissen für den Unterricht wird gezeigt, dass diese Gegebenheiten für den Grundraum Ω aller 2^n binären n-Tupel $(a_1, \ldots, a_n)$ und $A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j = 1\}$ sowie $\mathbb{P}(\{(a_1, \ldots, a_n)\}) := p^k(1-p)^{n-k}$ mit $k := a_1 + \ldots + a_n$ vorliegen, s. auch Abschn. 12.4 von Henze et al. (2021) im Hinblick auf eine Konkretisierung für den Unterricht.

Was Datenkompetenz betrifft, sollten die Studierenden das Gruppen-Screening kennen. Außerdem konfrontiere ich sie mit dem traurigen Fall von Sally

Clark, die aufgrund einer falschen Unabhängigkeitsannahme im Zusammenhang mit doppeltem plötzlichen Kindstod ins Gefängnis kam.

12 Zufallsvektoren, gemeinsame Verteilung

Der nächste Grundbegriff ist der einer *gemeinsamen Verteilung* im Zusammenhang mit Zufallsvektoren. Ich beginne hier immer mit dem Fall von zwei Zufallsvariablen X und Y und wähle als Einstieg den zweifachen Würfelwurf, wobei die Zufallsvariablen X und Y die Augenzahl des ersten Wurfs bzw. das Maximum der Augenzahlen aus beiden Würfen angeben. Der Grundraum Ω ist also hier die Menge aller 36 gleichwahrscheinlichen Paare (i,j) mit $i,j \in \{1,\ldots,6\}$, und es gelten $X((i,j)) := i$ sowie $Y((i,j)) := \max(i,j)$. Abbildung 6 zeigt die Wahrscheinlichkeiten $\mathbb{P}(X=i,Y=j)$ in Form einer rechteckigen Tabelle. Durch Summation ergeben sich die in diesem Zusammenhang auch als *Marginalverteilungen* bezeichneten Verteilungen von X und von Y. Das Präfix „Marginal“ rührt daher, dass die Wahrscheinlichkeiten $\mathbb{P}(X=i)$ bzw. $\mathbb{P}(Y=j)$ an den Rändern dieses Rechteckschemas sichtbar werden (lat. *margo* = Rand), s. Abb. 6.

j

i	1	2	3	4	5	6	Σ	$\mathbb{P}(X=i)$
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$	
2	0	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$	
3	0	0	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$	
4	0	0	0	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$	
5	0	0	0	0	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{1}{6}$	
6	0	0	0	0	0	$\frac{6}{36}$	$\frac{1}{6}$	
Σ	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1	

$\mathbb{P}(Y=j)$

Abb. 6: Gemeinsame Verteilung und Marginalvert. von erster und größter Augenzahl

Allgemein definiere ich an dieser Stelle die *gemeinsame Verteilung* zweier Zufallsvariablen X und Y als das System der Schnitt-Wahrscheinlichkeiten $\mathbb{P}(X=x_i,Y=y_j)$, zusammen mit den abzählbar vielen Werte-Paaren (x_i,y_j), für die die Wahrscheinlichkeiten $\mathbb{P}(X=x_i)$ oder $\mathbb{P}(Y=y_j)$ positiv sind.

Die Studierenden lernen, dass die gemeinsame Verteilung zweier Zufallsvariablen X und Y im Allgemeinen nicht durch die Marginalverteilungen von X und Y festgelegt ist, und dass diese gemeinsame Verteilung die Verteilung jeder reellen Funktion $g(X,Y)$ von X und Y bestimmt. Nehmen X und Y die möglichen Werte $x_1,\ldots,x_r$ bzw. $y_1,\ldots,y_s$ an, so gilt

$$\mathbb{E}\big[g(X,Y)\big] = \sum_{i=1}^{r}\sum_{j=1}^{s} g(x_i,y_j)\mathbb{P}(X=x_i,Y=y_j).$$

Zwei Zufallsvariablen X und Y heißen (*stochastisch*) *unabhängig*, falls für alle x und y mit $\mathbb{P}(X=x)>0$ und $\mathbb{P}(Y=y)>0$ die Gleichung $\mathbb{P}(X=x,Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$ erfüllt ist. Für unabhängige Zufallsvariablen gilt die *Multiplikationsregel*

$$\mathbb{E}(X\cdot Y) = \mathbb{E}(X)\cdot\mathbb{E}(Y) \tag{6}$$

für Erwartungswerte. Dabei wird vorausgesetzt, dass die Erwartungswerte von X und von Y existieren. Erst jetzt erfolgt eine Verallgemeinerung der Theorie auf den Fall von mehr als zwei Zufallsvariablen.

13 Varianz, Kovarianz, Korrelation

An dieser Stelle der Vorlesung motiviere und thematisiere ich die Grundbegriffe *Varianz*, *Kovarianz* und *Korrelation*. Zentrale Botschaft ist, dass die durch

$$\mathbb{V}(X) := \mathbb{E}\big[\big(X-\mathbb{E}(X)\big)^2\big]$$

definierte *Varianz* einer Zufallsvariablen X auch ein Erwartungswert ist, nämlich der Erwartungswert der quadratischen Abweichung von X um den Erwartungswert $\mathbb{E}(X)$. Indem man die Linearität der Erwartungswertbildung ausnutzt, ergibt sich die Varianz einer Indikatorsumme $X=\sum_{j=1}^{n}\mathbf{1}_{A_j}$ zu

$$\begin{aligned}\mathbb{V}(X) &= \sum_{j=1}^{n}\mathbb{P}(A_j)\big(1-\mathbb{P}(A_j)\big)\\ &\quad +2\sum_{1\le i<j\le n}\big(\mathbb{P}(A_i\cap A_j)-\mathbb{P}(A_i)\mathbb{P}(A_j)\big)\end{aligned}$$

und damit insbesondere zu $np(1-p)$, falls $A_1,\ldots,A_n$ stochastisch unabhängig sind und die gleiche Wahrscheinlichkeit p besitzen. Im Zusammenhang mit der Varianz darf natürlich die aus der Ungleichung

$$\mathbf{1}_{\{|X(\omega)-\mathbb{E}X|\ge\varepsilon\}} \le \frac{1}{\varepsilon^2}\cdot\big(X(\omega)-\mathbb{E}X\big)^2, \quad \omega\in\Omega,$$

folgende und noch etwa in Glaser et al. (1982), S. 95, oder Barth und Haller (1985), S. 184, zu findende *Tschebyschow-Ungleichung*

$$\mathbb{P}(|X-\mathbb{E}(X)|\ge\varepsilon) \le \frac{\mathbb{V}(X)}{\varepsilon^2} \quad \text{für jedes } \varepsilon>0$$

nicht fehlen. Die durch $\mathrm{C}(X,Y) := \mathbb{E}\big[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\big]$ definierte *Kovarianz* zwischen zwei Zufallsvariablen X und Y motiviere ich über die aus der Linearität der Erwartungswertbildung folgende Gleichung

$$\mathbb{V}(X+Y) = \mathbb{V}(X)+\mathbb{V}(Y)+2\mathrm{C}(X,Y),$$

die auch die Namensgebung *Kovarianz* („mit der Varianz“) erklärt.

Aufgrund der Multiplikationsregel (6) für Erwartungswerte folgt, dass diese Kovarianz verschwindet, wenn X und Y unabhängig sind. Summe und Differenz der Augenzahlen beim zweifachen Würfelwurf sind ein Beispiel dafür, dass die Umkehrung dieser Implikation im Allgemeinen nicht gilt. Der *Pearsonsche Korrelationskoeffizient*

$$r(X,Y) := \frac{\mathrm{C}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

zwischen X und Y ist die nach Division durch das Produkt der Standardabweichungen normierte Kovarianz. Was man zum Verständnis dieses Korrelationskoeffizienten unbedingt wissen muss, ist, dass er im Zusammenhang mit einer Minimierungsaufgabe auftritt. Will man Y durch eine affine Funktion von X approximieren, so lässt sich die Güte dieser Approximation durch das Minimum der erwarteten quadratischen Abweichung zwischen Y und $a+bX$ quantifizieren, wobei über alle reellen Zahlen a und b minimiert wird. Diesbezüglich gilt

$$\min_{a,b\in\mathbb{R}} \mathbb{E}\big[(Y-a-bX)^2\big] = \mathbb{V}(Y)\big(1-r^2(X,Y)\big).$$

Da die rechte Seite nichtnegativ ist, folgt $-1 \le r(X,Y) \le 1$, was zur *Cauchy–Schwarz-Ungleichung* $\mathrm{C}^2(X,Y) \le \mathbb{V}(X)\mathbb{V}(Y)$ äquivalent ist. Studierende sollten auch gesehen haben, dass

$$\mathbb{E}\big[(Y-a-bX)^2\big] = \frac{1}{n}\sum_{j=1}^{n}\big(y_j-a-bx_j\big)^2$$

gilt, wenn der Zufallsvektor (X,Y) die Werte (x_j,y_j) für j von 1 bis n mit gleicher Wahrscheinlichkeit $\frac{1}{n}$ annimmt. Die obige Minimierungsaufgabe führt dann auf die *Methode der kleinsten Quadrate*.

14 Die Multinomialverteilung

Sind bei einem stochastischen Vorgang s verschiedene Ergebnisse möglich, die o.B.d.A. mit den Zahlen $1,2,\ldots,s$ codiert und *Treffer 1.Art*, …, *Treffer s-ter Art* genannt werden, und wird dieser Vorgang n-mal unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen ausgeführt, so interessiert, wie viele Treffer der einzelnen Arten auftreten. Besitzt der Treffer j-ter Art bei jedem dieser Vorgänge die gleiche Wahrscheinlichkeit p_j, $j\in\{1,\ldots,s\}$, so hat der Zufallsvektor $(X_1,\ldots,X_s)$ der Trefferzahlen nach n-maliger Ausführung die *Multinomialverteilung* $\mathrm{Mult}(n;p_1,\ldots,p_s)$, d.h., es gilt

$$\mathbb{P}(X_1=k_1,\ldots,X_s=k_s) = \frac{n!}{k_1!\cdot\ldots\cdot k_s!}p_1^{k_1}\cdot\ldots\cdot p_s^{k_s}$$

für jede Wahl nichtnegativer ganzer Zahlen $k_1,\ldots,k_s$ mit $\sum_{j=1}^{s}k_j=n$, s. hierzu z.B. Henze (2019a).

Als Hintergrundwissen für den Unterricht ist die Multinomialverteilung wichtig, denn sie ist nicht nur eine direkte Verallgemeinerung der Binomialverteilung, die sich für $s=2$ ergibt (und bei der man nur X_1 betrachten muss), sondern sie tritt schon bei der ganz banalen Frage auf, mit welcher Wahrscheinlichkeit man beim sechsfachen Würfelwurf jede Augenzahl genau einmal erhält. Außerdem ist sie die Grundlage für den Chi-Quadrat-Test, der etwa von Finanzämtern routinemäßig angewendet wird.

15 Wartezeitverteilungen

Im Zusammenhang mit dem Warten auf den ersten Treffer oder allgemeiner auf den k-ten Treffer in unabhängigen Bernoulli-Versuchen mit gleicher Trefferwahrscheinlichkeit p lernen die Studierenden die *geometrische Verteilung* und die *negative Binomialverteilung* kennen. So kann man auch Schülerinnen und Schülern die Frage stellen, mit welcher Wahrscheinlichkeit der dritte Treffer in unabhängigen gleichartigen Bernoulli-Versuchen im 16. Versuch auftritt. Eine clevere Schülerin könnte hier antworten: „In den ersten 15 Versuchen sollen genau zwei Treffer auftreten. Die beiden Versuche, die jeweils einen Treffer ergeben, kann ich auf $\binom{15}{2}$ Weisen auswählen. In 13 Versuchen soll dann kein Treffer auftreten, und die Wahrscheinlichkeit dafür ist $(1-p)^{13}$. Die Wahrscheinlichkeit dafür, dass in den beiden ausgewählten Versuchen sowie im 16. Versuch ein Treffer auftritt, ist gleich p^3. Die Antwort auf die gestellte Frage ist also $\binom{15}{2}p^3(1-p)^{13}$.“ Im Hinblick auf die Erfahrungswelt von Schülerinnen und Schülern behandle ich auch *Sammelbilderprobleme*, zu deren Lösung die Formel des Ein- und Ausschließens grundlegend ist. Nehmen wir an, zu einem vollständigen Set gehören sechs Bilder, die von eins bis sechs nummeriert sind, so können wir gedanklich auch einen Würfel werfen und nach der Verteilung der Anzahl X der Würfe fragen, bis jede Augenzahl mindestens einmal aufgetreten ist. Das Stabdiagramm der Verteilung von X hat die in Abb. 7 gezeigte Gestalt.

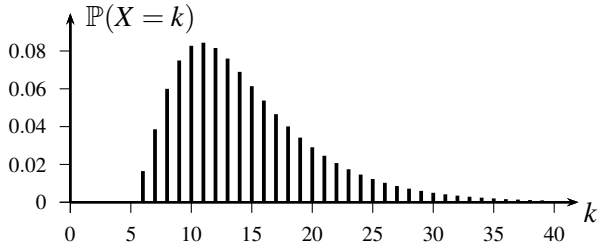

Abb. 7: Verteilung der Anzahl der Würfelwürfe, bis jede Augenzahl mindestens einmal auftritt

16 Die Poisson-Verteilung

Eine Verteilung, die früher in Schulbüchern auftrat (s. z.B. Glaser et al. (1982), S. 108) und als Hintergrundwissen für Lehramtsstudierende unverzichtbar ist, ist die *Poisson-Verteilung*. Diese entsteht unter anderem über das sog. *Gesetz seltener Ereignisse*

$$\lim_{n\to\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k}=\mathrm{e}^{-\lambda}\cdot\frac{\lambda^k}{k!},\quad k=0,1,2,\ldots$$

aus der Binomialverteilung $\mathrm{Bin}(n,p_n)$ beim Grenzübergang $np_n\to\lambda$, wobei $0<\lambda<\infty$. Die Poisson-Verteilung ist also insbesondere eine gute Approximation der Binomialverteilung $\mathrm{Bin}(n,p)$ bei großem n und kleinem p. Ein schönes historisches Beispiel für das Auftreten der Poisson-Verteilung beim radioaktiven Zerfall als spontanes Phänomen ist das Rutherford–Geiger-Experiment, s. z.B. Henze (2021), S. 197.

17 Schwaches Gesetz großer Zahlen, stochastische Konvergenz

Nicht fehlen darf auch das *schwache Gesetz großer Zahlen*. Sind $X_1,\ldots,X_n$ unabhängige Zufallsvariablen mit gleichem Erwartungswert μ und gleicher, existierender Varianz, so gilt für das arithmetische Mittel $\overline{X}_n:=\frac{1}{n}\sum_{j=1}^n X_j$:

$$\lim_{n\to\infty}\mathbb{P}\big(\big|\overline{X}_n-\mu\big|\geq\varepsilon\big)=0\quad\text{für jedes }\varepsilon>0.\qquad(7)$$

Abbildung 8 zeigt hierzu 300 simulierte Würfe eines fairen Würfels, also den Fall $\mathbb{P}(X_1=j)=\frac{1}{6}$ für $j\in\{1,\ldots,6\}$ und $\mu=3.5$.

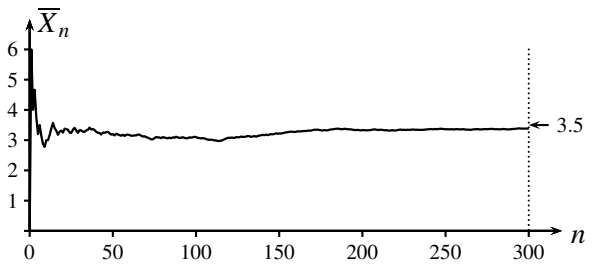

Abb. 8: Zum schwachen Gesetz großer Zahlen

Das schwache Gesetz großer Zahlen stellt innerhalb eines stochastischen Rahmens einen Zusammenhang zwischen arithmetischen Mitteln und Erwartungswerten her. Es folgt unmittelbar aus der Tschebyschow-Ungleichung, denn nach Rechenregeln über die Varianz besitzt die in (7) stehende Wahrscheinlichkeit die obere Schranke $\mathbb{V}(X_1)/(n\varepsilon^2)$. Ich sage den Studierenden an dieser Stelle auch, dass wir eigentlich das in (7) stehende W-Maß $\mathbb{P}$ mit einem Index n versehen müssten, weil im Rahmen diskreter W-Räume bislang nur Modelle für *endlich viele* unabhängige Zufallsvariablen zur Verfügung stehen. Im Spezialfall $X_j=\mathbf{1}_{A_j}$, $j\in\{1,\ldots,n\}$, mit unabhängigen Ereignissen $A_1,\ldots,A_n$, die alle die gleiche Wahrscheinlichkeit p besitzen, lässt das arithmetische Mittel $\overline{X}_n$ eine Deutung als zufällige relative Trefferhäufigkeit zu. Dieser als *Schwaches Gesetz großer Zahlen von Jacob Bernoulli* bekannte Fall kann auch in der Form $\overline{X}_n\xrightarrow{\mathbb{P}}p$ für $n\to\infty$ geschrieben werden. Dabei bezeichnet $\xrightarrow{\mathbb{P}}$ *stochastische Konvergenz*. Allgemein definiert man für eine Folge (Y_n) von Zufallsvariablen und eine reelle Zahl a die *stochastische Konvergenz von (Y_n) gegen a* durch

$$\lim_{n\to\infty}\mathbb{P}\big(\big|Y_n-a\big|\geq\varepsilon\big)=0\quad\text{für jedes }\varepsilon>0.$$

Dieser Konvergenzbegriff aus der Stochastik ist der einzige, der thematisiert und mit Rechenregeln wie etwa der Vererbung von stochastischer Konvergenz unter stetigen Abbildungen unterfüttert wird.

18 Zentraler Grenzwertsatz

Obligatorischer Bestandteil einer einführenden Stochastikvorlesung ist auch der *zentrale Grenzwertsatz von de Moivre–Laplace*. Ist S_n eine Zufallsvariable mit der Binomialverteilung $\mathrm{Bin}(n,p)$, wobei $0<p<1$, und bezeichnet $S_n^*:=(S_n-np)/\sqrt{np(1-p)}$ die zugehörige standardisierte Zufallsvariable, so gilt für jede Wahl von $a,b\in\mathbb{R}$ mit $a<b$:

$$\lim_{n\to\infty}\mathbb{P}(a\leq S_n^*\leq b)=\int_a^b\varphi(t)\,\mathrm{d}t.\qquad(8)$$

Dabei bezeichnet $\varphi(t)=\mathrm{e}^{-t^2/2}/\sqrt{2\pi}$, $t\in\mathbb{R}$, die sog. *Gaußsche Glockenkurve*. Der Grenzübergang in (8) ist in Abb. 9 veranschaulicht. Für den Fall $n=100$ und $p=0.3$ zeigt Abb. 9 das Histogramm der standardisierten Binomialverteilung. Dabei sind die Mittelpunkte der Basislinien der Rechtecke auf der horizontalen Achse gleich $x_{n,k}:=(k-np)/\sqrt{np(1-p)}$, $k\in\{0,\ldots,n\}$, und somit gleich den Werten, die S_n^* annehmen kann. Breite und Höhe des Rechtecks mit Mittelpunkt $X_{n,k}$ sind gleich $1/\sqrt{np(1-p)}$ bzw.

gleich $\sqrt{np(1-p)}\binom{n}{k}p^k(1-p)^{n-k}$. Die Fläche des Rechtecks ist also gleich $\mathbb{P}(S_n^* = x_{n,k})$. Zusätzlich ist noch das Schaubild der Funktion φ eingezeichnet.

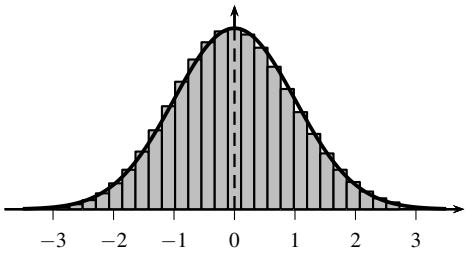

Abb. 9: Histogramm der Verteilung von S_n^* für $n = 100$, $p = 0.3$ mit Schaubild von φ

Die Studierenden sollten verinnerlicht haben, was beim Übergang von der Binomialverteilung zur Normalverteilung passiert und idealerweise auch eine Beweisidee der Grenzwertaussage (8) gesehen haben, s. z.B. Henze (2022a). In der Vorlesung bleibt jedoch keine Zeit, einen formalen Beweis zu führen. Ich gebe stets auch den zentralen Grenzwertsatz von Lindeberg–Lévy als Verallgemeinerung des Satzes von de Moivre–Laplace an.

19 Deskriptive Statistik

Im Rahmen einer einführenden Vorlesung muss es auch einen gewissen Anteil an Statistik geben. Ich beginne hier immer mit der deskriptiven Statistik, die auch ganz am Anfang der Vorlesung stehen könnte. Nach einer Klärung von *Merkmaltypen* (*quantitativ/ qualitativ*, *nominal/ordinal*, *diskret/stetig*) geht es unter anderem um *empirische Häufigkeitsverteilungen* sowie *Balken-* und *Kreisdiagramme*. Weitere grafische Darstellungsmittel sind das *Histogramm* sowie das *Stamm- und Blatt-Diagramm*. Als Lagemaße treten das *arithmetische Mittel* sowie der sich aus der *geordneten Stichprobe* ergebende *empirische Median* auf. Wohingegen das arithmetische Mittel von $x_1, \ldots, x_n$ die Summe $\sum_{j=1}^n (x_j - t)^2$ als Funktion von t minimiert, minimiert der empirische Median die Summe $\sum_{j=1}^n |x_j - t|$ der absoluten Abweichungen als Funktion von t, s. z.B. Henze (2020a).

Weitere Lagemaße sind *empirische p-Quantile* sowie *α-getrimmte Mittel* als Kompromiss zwischen arithmetischem Mittel und empirischem Median. Was Streuungsmaße betrifft, gehe ich auf *Stichprobenvarianz* und *Stichprobenstandardabweichung* sowie auf die *mittlere absolute Abweichung*, die *Stichprobenspannweite*, den *Quartilsabstand* sowie die *Median-Abweichung* als robustes Streuungsmaß ein. Das Kapitel schließt mit dem Thema *Boxplot*. Die *empirische Regressionsgerade* wird schon im Zusammenhang mit der Methode der kleinsten Quadrate (vgl. Abschn. 13) behandelt.

20 Punktschätzung

Was die Grundprobleme der induktiven Statistik betrifft, reicht es m.E. aus, sich auf den Fall einer unbekannten Trefferwahrscheinlichkeit bei Bernoulli-Versuchen zu beschränken. Ich konfrontiere die Studierenden immer mit der Frage: „Bei 100 unabhängigen Bernoulli-Versuchen mit unbekannter Trefferwahrscheinlichkeit p haben sich 38 Treffer ergeben. Wie groß ist p?“ Man sieht schnell ein, dass die einzig richtige Antwort auf diese Frage „Es gilt $0 < p < 1$“ lautet; man ist aber auch bereit, Unsicherheit in Kauf zu nehmen, um eine gewisse Eingrenzung der für möglich erachteten Werte von p vornehmen zu können. Im Folgenden sei S_n eine Zufallsvariable mit der Binomialverteilung $\mathrm{Bin}(n,p)$, wobei n bekannt und p unbekannt sei. Sollen Wahrscheinlichkeiten berechnet werden, muss also immer p spezifiziert werden, was durch Indizierung von $\mathbb{P}$ mit p kenntlich gemacht wird. Man schreibt also $\mathbb{P}_p$, und auch die Berechnung von Erwartungswerten und Varianzen wird durch Indizierung mit p notiert.

Ich starte mit dem *Maximum-Likelihood-Schätzprinzip*, wobei *Maximum Likelihood* im Folgenden stets mit ML abgekürzt wird. Nach diesem Prinzip studiert man für gegebenes n und k die Wahrscheinlichkeit $\mathbb{P}_p(S_n = k)$ *als Funktion von* p. Die durch

$$L_k(p) := \mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k},\ 0 \le p \le 1,$$

definierte Funktion $L_k : [0,1] \to \mathbb{R}$ heißt *Likelihoodfunktion* zu(r Beobachtung) k. Das ML-Schätzprinzip besteht darin, denjenigen Wert von p für den glaubwürdigsten zu halten, welcher der beobachteten Trefferanzahl k die größte Wahrscheinlichkeit verleiht.

Die Funktion L_k nimmt ihr Maximum für $p = \frac{k}{n} =: \widehat{p}$ an. Abbildung 10 zeigt Likelihoodfunktionen für den Fall $n = 10$ und $k = 2$, $k = 6$ sowie $k = 7$.

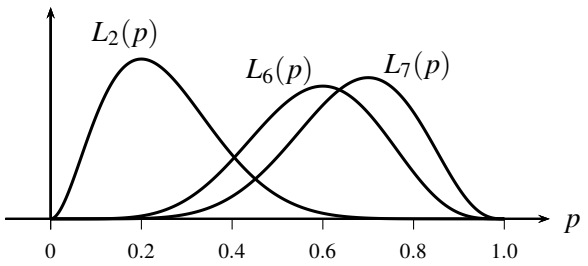

Abb. 10: Likelihoodfunktionen ($n = 10$)

Der als relative Trefferhäufigkeit interpretierbare Wert $\widehat{p}$ heißt *ML-Schätzwert* für p zu k. Dieser Schätzwert ist eine Realisierung der Zufallsvariablen

$$T_n := \frac{S_n}{n}, \tag{9}$$

und diese Zufallsvariable heißt *ML-Schätzer* für p. Die Güte des *Schätzverfahrens* wird durch Eigenschaften des Schätzers T_n beurteilt. Es gelten

$$\mathbb{E}(T_n) = p, \quad \mathbb{V}_p(T_n) = \frac{p(1-p)}{n}, \quad 0 \le p \le 1.$$

Die erste Eigenschaft heißt *Erwartungstreue* von T_n. Zusammen mit der zweiten Eigenschaft folgt, dass sich die Verteilung des Schätzers T_n bei wachsendem n immer mehr um den unbekannten Wert p konzentriert, und zwar ganz gleich, welches dieser Wert ist. Je nach zeitlichem Spielraum kann man vielleicht noch weitere Schätzprobleme behandeln und auf die Begriffe *Verzerrung* (*bias*) und *mittlere quadratische Abweichung* eingehen.

21 Konfidenzbereiche

Konfidenzbereiche wie auch statistische Tests (s. nächster Abschnitt) stellen eine große intellektuelle Herausforderung für Studierende dar, und es bleibt abzuwarten, ob sie – nachdem einige Bundesländer bereits einen Paradigmenwechsel von statistischen Tests hin zu Konfidenzbereichen vollzogen haben – im gymnasialen Stochastikunterricht wirklich *verstanden* werden. Ich beschränke mich im Folgenden auf das Problem, eine begründete Aussage über die als unbekannt angenommene Wahrscheinlichkeit p zu treffen, wenn eine Realisierung einer Zufallsvariablen S_n mit der Binomialverteilung $\mathrm{Bin}(n,p)$ vorliegt.

Ein *Konfidenzbereich* (synonym: *Vertrauensbereich*) für p ist formal eine *Zufallsvariable*, *deren Realisierungen Teilmengen von* $[0,1]$ *sind*. Da es im Folgenden nur um Intervalle geht, wird von jetzt an von einem *Konfidenzintervall* (synonym: *Vertrauensintervall*) die Rede sein. Das Vertrauen, welches man in ein Konfidenzintervall legt, drückt sich in einer *Konfidenzwahrscheinlichkeit* (synonym: *Vertrauenswahrscheinlichkeit*) aus. Diese Wahrscheinlichkeit wird üblicherweise mit $1-\alpha$ bezeichnet. Dabei ist α eine kleine positive Zahl wie z.B. 0.05 oder 0.01.

Ein *Konfidenzintervall für p zur Konfidenzwahrscheinlichkeit* $1-\alpha$ ist ein Intervall $I_n := [U_n, O_n] \subset [0,1]$, dessen (zufällige) Grenzen U_n und O_n von α und von S_n abhängen, wobei gilt:

$$\mathbb{P}_p(I_n \ni p) \ge 1-\alpha \quad \text{für jedes } p \in [0,1].$$

Hier ist es wichtig, $I_n \ni p$ und nicht $p \in I_n$ zu schreiben, weil nicht p zufällig ist, sondern das „Zufalls-Intervall“ I_n das unbekannte p „überdeckt“. Die Realisierung $S_n = k$ liefert eine Realisierung von I_n in Form eines *konkreten Konfidenzintervalls* wie in Abb. 11 (für $n = 50$ und $\alpha = 0.05$). Für diese Abbildung wurde für $p = 0.35$ 30-mal ein nach der Methode von Clopper und Pearson (s. Henze (2021), S. 251ff.) gewonnenes konkretes Konfidenzintervall erstellt. Nur eines dieser Intervalle enthält p nicht.

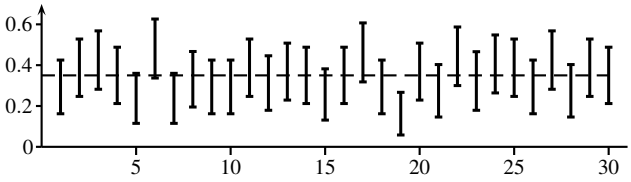

Abb. 11: Konkrete Konfidenzintervalle für p

In der Vorlesung stelle ich ein Konstruktionsprinzip für Konfidenzintervalle sowie das Verfahren von Clopper und Pearson vor. Mithilfe des zentralen Grenzwertsatzes ergeben sich approximative Konfidenzintervalle für p bei großem Stichprobenumfang n (s. Henze (2021), S. 256ff). Liegt ein konkretes Konfidenzintervall wie etwa $[0.38, 0.54]$ vor, so hat man großes Vertrauen in die Aussage, dass für p die Ungleichungen $0.38 \le p \le 0.54$ gelten. Eine Wahrscheinlichkeitsaussage ist aber dann nicht mehr möglich, weil p unbekannt, aber fest ist.

22 Statistische Tests

Auch statistische Tests dürfen nicht fehlen, da sie in vielen Bundesländern immer noch fester Bestandteil der schulischen Curricula sind. Damit man eingesehen hat, dass Fehler bei solchen Tests unvermeidlich sind, reicht es m.E. wieder aus, sich auf den in der Schule auftretenden Fall von Tests bezüglich der unbekannten Wahrscheinlichkeit p der Binomialverteilung $\mathrm{Bin}(n,p)$ zu beschränken (obwohl ich immer auch den allgemeineren Fall eines statistischen Modells wie in Abschn. 30.2 von Henze (2021) behandle). Wenn sich die Studierenden klar gemacht haben, dass 38 erzielte Treffer aus insgesamt 100 unabhängigen Bernoulli-Versuchen mit unbekannter Trefferwahrscheinlichkeit p von jedem $p \in (0,1)$ in dem Sinne „herrühren können“, dass $\binom{100}{38} p^{38}(1-p)^{62}$ für jedes solche p positiv ist, folgt zwangsläufig, dass jede Hypothese über p, die ja p auf eine echte Teilmenge von $(0,1)$ einschränkt, abgelehnt werden kann, obwohl sie stimmt. Umgekehrt ist es auch möglich, dass eine Hypothese nicht abgelehnt wird, obwohl sie in Wirklichkeit nicht zutrifft, denn die

Testentscheidung fußt ja auf der vor Durchführung des Tests zufälligen Trefferanzahl.

Gegenüber Konfidenzbereichen hat sich beim Testen herausgestellt, dass den Studierenden allein die Fülle der Begriffe (*Hypothese*, *Alternative*, *kritischer Bereich*, *Annahmebereich*, *Prüfgröße* (bzw. *Testgröße*), *Fehler erster und zweiter Art*, *Gütefunktion*, *Test zum Niveau* α) Schwierigkeiten bereitet. In der Vorlesung werden ausführlich insbesondere *ein-* und *zweiseitige Binomialtests* und Fragen der Konsistenz sowie der Planung des nötigen Stichprobenumfangs, um relevante Alternativen mit einer vorgegebenen Mindestwahrscheinlichkeit aufzudecken, behandelt. Auch der klassische *Chi-Quadrat-Anpassungstest* ist fester Bestandteil der Vorlesung.

Nicht nur für die Studierenden besteht ein allgemeines Problem mit statistischen Tests darin, dass deren Ergebnisse meist falsch interpretiert werden, weil man sowohl der Gültigkeit der Hypothese als auch der Alternative „Wahrscheinlichkeiten“ zuschreiben möchte, s. z.B. Mossburger (2014).

23 Allgemeine Modelle

Bislang ging es um *diskrete* W-Räume und *diskrete* Verteilungen. Da in der Schule auch *stetige* Verteilungen vorkommen, sollten Lehramtsstudierende das Konzept eines *allgemeinen W-Raumes* und damit das *Kolmogorovsche Axiomensystem* kennen. Ich beweise zunächst, dass eine „Längen-Funktion“, die naheliegende Eigenschaften besitzen sollte, nicht auf allen Teilmengen der reellen Zahlen definiert werden kann, s. Henze (2019b). Diese Erkenntnis fördert die Einsicht, dass es auch keine stetige Gleichverteilung auf allen Teilmengen des Einheitsintervalles gibt, und dass man bei einem überabzählbaren Grundraum Ω Wahrscheinlichkeiten von Teilmengen von Ω, die die Eigenschaften a)-c) aus Abschn. 4 erfüllen, nur noch für *gewisse* (*Ereignisse* genannte) Teilmengen fordern kann. Das mit $\mathcal{A}$ bezeichnete System dieser Teilmengen sollte den Grundraum Ω und mit jeder Menge auch deren Komplement enthalten. Fordert man noch, dass mit je abzählbar-unendlich vielen Mengen auch deren Vereinigung zu $\mathcal{A}$ gehören soll, so ist $\mathcal{A}$ definitionsgemäß eine σ-*Algebra* über Ω. Ein *allgemeiner W-Raum* ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$, in dem Ω eine *beliebige* nichtleere Menge und $\mathcal{A}$ eine σ-Algebra über Ω sind. Weiter ist (das W-Maß) $\mathbb{P}$ eine auf $\mathcal{A}$ definierte reelle Funktion, die die Eigenschaften a)-c) aus Abschn. 4 erfüllt. Im Fall $\Omega = \mathbb{R}$ setzt man $\mathcal{A} := \mathcal{B}$. Dabei bezeichnet $\mathcal{B}$ die kleinste σ-Algebra über $\mathbb{R}$, die alle offenen Mengen enthält (sog. *Borelsche* σ-*Algebra*).

Geradezu zwangsläufig ergibt sich, dass sich nicht mehr jede Funktion $X : \Omega \to \mathbb{R}$ mit dem Attribut *Zufallsvariable* schmücken darf. Damit etwa

$$F(t) := \mathbb{P}(X \le t) := \mathbb{P}\big(\{\omega \in \Omega : X(\omega) \le t\}\big) \quad (10)$$

wohldefiniert ist, muss die rechts stehende Menge zu $\mathcal{A}$ gehören. Diese Bedingung wird für jedes $t \in \mathbb{R}$ gefordert und bedeutet die sog. *Messbarkeit* von X. Aus ihr folgt, dass für jede Borelmenge B

$$\mathbb{P}^X(B) := \mathbb{P}\big(X^{-1}(B)\big) = \mathbb{P}\big(\{\omega \in \Omega : X(\omega) \in B\}\big)$$

wohldefiniert ist. Das auf $\mathcal{B}$ definierte W-Maß $\mathbb{P}^X$ heißt *Verteilung von* X. Es ist durch die in (10) stehende Funktion F, die sog. *Verteilungsfunktion von* X, eindeutig bestimmt.

24 Stetige Verteilungen

Eine Zufallsvariable X heißt (*absolut*) *stetig* (*verteilt*), falls es eine nichtnegative messbare Funktion $f : \mathbb{R} \to \mathbb{R}$ gibt, so dass gilt:

$$F(t) := \mathbb{P}(X \le t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x, \quad t \in \mathbb{R}. \quad (11)$$

Die Funktion f heißt (*Wahrscheinlichkeits-*)*Dichte* von X (bzw. von F). Ich sage den Studierenden an dieser Stelle, dass hier der Lebesguesche Integralbegriff zugrundeliegt, damit die Verteilung von X ein W-Maß auf der Borelschen σ-Algebra $\mathcal{B}$ ist. Da die auftretenden Dichten stückweise stetig sind, kann aber für konkrete Berechnungen der Art

$$\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}x, \quad -\infty < a < b < \infty,$$

mit dem Riemannschen Integralbegriff gearbeitet werden. Am Beispiel der Cantorschen Verteilungsfunktion (s. Henze (2019c)) lernen die Studierenden, dass nicht jede stetige Verteilungsfunktion F in der Form (11) geschrieben werden kann.

Was konkrete stetige Verteilungen betrifft, sind die *stetige Gleichverteilung* auf einem Intervall, die (gedächtnislose) *Exponentialverteilung* sowie die *Normalverteilung* obligatorisch. Dabei sollten falls möglich Querverbindungen zwischen Verteilungen aufgezeigt werden. Besitzt etwa X eine Gleichverteilung in $(0,1)$, so hat $-\log(1-X)/\lambda$ eine Exponentialverteilung mit Parameter λ. Bei der Normalverteilung ist eine der entscheidenden Botschaften, dass es im Wesentlichen nur die Standardnormalverteilung gibt, denn ist X standardnormalverteilt, so hat $\sigma X + \mu$ eine Normalverteilung mit Erwartungswert μ und Varianz σ^2. Weitere stetige Verteilungen sind die

durch Anwendung der Exponentialfunktion auf eine normalverteilte Zufallsvariable entstehende *Lognormalverteilung*, die *Chi-Quadrat- Verteilung* sowie die (keinen Erwartungswert besitzende) *Cauchy-Verteilung* als Verteilung des Tangens eines gleichverteilten Winkels. Im Zusammenhang mit stetigen Verteilungen halte ich auch die aus der Verteilungsfunktion F über die Festsetzung

$$F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad 0 < p < 1,$$

hervorgehende *Quantilfunktion* für wichtig. Sie dient nicht nur der Definition des p-Quantils $F^{-1}(p)$, sondern bewirkt auch, dass über die *Quantiltransformation* $U \mapsto X := F^{-1}(U)$ aus einer im Intervall $(0,1)$ gleichverteilten Zufallsvariablen U eine Zufallsvariable X mit der Verteilungsfunktion F entsteht. Dieser Sachverhalt ist für Simulationen von Bedeutung.

25 Abschließende Bemerkungen

Man kann sich wünschen, dass auch Kenntnisse über erzeugende Funktionen, über Prinzipien zur Erzeugung von Pseudozufallszahlen und über Simulation sowie über bedingte Erwartungswerte und bedingte Verteilungen vorhanden sind. Auch Schätzer und Tests bei stetigen Verteilungen wie etwa der t-Test sind sicherlich wünschenswert. Man sollte sich aber keinerlei Illusionen hingeben. Die Vorlesungszeit ist beschränkt, und die Stochastik ist aufgrund ihrer vielen Konzepte und spezifischen Denkweisen nicht einfach zu vermitteln. Was eine solche Vorlesung mit Sicherheit nicht auch noch leisten kann, ist ein umfangreicher Umgang mit Daten unter Einbezug eines Statistikpaketes.

Danksagung: Ich danke den beiden Gutachtern bzw. Gutachterinnen für wertvolle Hinweise.

Literatur

Barth, F.; Haller, R. (1985): Stochastik. Leistungskurs. München: Ehrenwirth.

Biehler, R., u.a. (2012): Mathematik Neue Wege. Arbeitsbuhc für Gymnasien, Stochastik. Braunschweig: Bildungshaus Schulbuchverlage.

Engel, A. (1973): Wahrscheinlichkeitsrechnung und Statstik. Band 1. Klett Studienbücher. Stuttgart: Ernst Klett Verlag.

Glaser et al. (Hrsg.) (1982): Sigma: Grundkurs Stochastik. Stuttgart: Ernst Klett Verlag.

Henze, N. (2019): Der große Umordnungssatz für Reihen, Erklärvideo. https://doi.org/10.5445/DIVA/2019-261

Henze, N. (2019a): Die Multinomialverteilung. Erklärvideo. https://doi.org/10.5445/DIVA/2019-260

Henze, N. (2019b): Die Unlösbarkeit des Maßproblems. Erklärvideo. https://doi.org/10.5445/DIVA/2019-189

Henze, N. (2019c): Die Cantorsche Verteilungsfunktion. Erklärvideo. https://doi.org/10.5445/DIVA/2019-176

Henze, N. (2020): Die Pólya-Verteilung. Erklärvideo. https://doi.org/10.5445/IR/1000119434

Henze, N. (2020a): Arithmetisches Mittel und Median: Minimaleigenschaften. Erklärvideo. https://doi.org/10.5445/IR/1000122606

Henze, N. (2021): Stochastik für Einsteiger. 13. Auflage. Heidelberg: Springer Spektrum.

Henze, N. (2021a): Ein Simpson-Paradoxon bei Covid-19-Todesfallraten. Stochastik in der Schule **41**(3), 33–35.

Henze, N. (2022a): Zentrale Grenzwertsätze für die Binomialverteilung. Erklärvideo. https://doi.org/10.5445/IR/1000152134

Henze, N. (2023): Binomialkoeffizienten – verstehen oder rechnen? Stochastik in der Schule **43**(1), 13–18.

Henze, N., Vehling, R. (2019): Der verwirrende Siegeszug des Histogramms in deutsche Klassenzimmer: Sind Stabdiagramme tot? MU **65**(1), 33–41.

Henze, N.; Müller, K.; Schilling, J. (2021). Stochastik rezeptfrei unterrichten – Anregungen für spannende Lehre über den Zufall: Heidelberg: Springer Spektrum.

Mossburger, M. (2014): Unklare Begriffe und Wunschdenken bei Signifikanztests. Stochastik in der Schule **34**(1), 2–8.

Anschrift des Verfassers:
Prof. i.R. Dr. Norbert Henze
KIT Distinguished Senior Fellow
Institut für Stochastik
Karlsruher Institut für Technologie (KIT)
Englerstr. 2
76131 Karlsruhe
Henze@kit.edu