Erwartungswert und Varianz bei der Binomial- und geometrischen Verteilung – mögliche Begründungen im Schulunterricht

HANS HUMENBERGER, WIEN

Zusammenfassung: *In diesem Aufsatz werden einige Möglichkeiten vorgestellt, schulrelevante Formeln für Erwartungswert und Varianz bei der Binomial- und geometrischen Verteilung mittels Plausibilitätsbetrachtungen zu begründen bzw. formal zu beweisen. Insbesondere soll gezeigt werden, dass es korrekte Beweise gibt, die sich nicht auf abstraktem Universitätsniveau befinden.*

1 Einleitung

Es dürfte unbestritten sein, dass die wichtigste diskrete Verteilung im Schulunterricht die Binomialverteilung ist. An zweiter (und letzter?) Stelle kommt die geometrische Verteilung, und in vielen Fällen kommen im Schulunterricht auch keine anderen diskreten Verteilungen vor (es gäbe da ja auch noch viele andere). Dieser Aufsatz soll kein Plädoyer dafür sein, doch so manche andere diskrete Verteilung in die Lehrpläne und -bücher aufzunehmen. Es geht uns hier vielmehr darum, eine Möglichkeit aufzuzeigen die entsprechenden Formeln für Erwartungswert und Varianz bei der Binomial- und bei der geometrischen Verteilung ohne höhere Mathematik zu beweisen. Denn oft hört man: „Ein wirklicher Beweis ist auf Schulniveau gar nicht *möglich*, weil den Lernenden *höhere* Konzepte fehlen." Das wird sich als falsch herausstellen. Eine andere Frage ist, ob in der Schule formale Beweise dafür *nötig* sind. Da wird man vermutlich in Grundkursen anders vorgehen als in Leistungskursen. Auch in Leistungskursen ist eine formale Argumentation nicht zwingend nötig (die wertvolle Unterrichtszeit kann vielleicht woanders besser eingesetzt werden?), aber bei Interesse ist sie *möglich* – das ist die zentrale Botschaft dieses Aufsatzes.

Stochastik-Veranstaltungen an Universitäten haben andere Möglichkeiten und Voraussetzungen als Stochastik-Kurse an Schulen. Das ist der Grund, warum manche *schulrelevante* Inhalte fachlicher Vorlesungen nicht auf dieselbe Art im Schulunterricht begründet werden können. Ein erstes Beispiel dafür ist die Additivität des Erwartungswertes: $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$ für beliebige Zufallsvariablen X, Y (mit existierenden Erwartungswerten)[1]. Wenn man Zufallsvariablen als *Funktionen* $\Omega \to \mathbb{R}$ auffasst, dann kann man für den Erwartungswert einer diskreten Zufallsvariable schreiben

$$\mathrm{E}(X)=\sum_{\omega\in\Omega} X(\omega)\cdot \mathrm{P}(\omega).$$

Mit dieser Schreibweise ist der Beweis von $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$ sofort klar:

$$\begin{aligned}\mathrm{E}(X+Y)&=\sum_{\omega\in\Omega}(X+Y)(\omega)\cdot \mathrm{P}(\omega)\\ &=\underbrace{\sum_{\omega\in\Omega} X(\omega)\cdot \mathrm{P}(\omega)}_{\mathrm{E}(X)}+\underbrace{\sum_{\omega\in\Omega} Y(\omega)\cdot \mathrm{P}(\omega)}_{\mathrm{E}(Y)}\end{aligned}$$

Aber diese Schreib- und Denkweise ist im Schulunterricht kaum üblich. Die dort verbreitete Erwartungswertdefinition

$$\mathrm{E}(X):=\sum_i x_i\cdot \mathrm{P}(X=x_i)$$

erlaubt leider keinen so kurzen Beweis des intuitiv naheliegenden Sachverhaltes $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$; es kämen Doppelsummen oder zumindest Doppelindices ins Spiel, und der Beweis würde vor allem zu einem Schreibproblem, das bei Schülern/innen kaum eine erhellende bzw. erklärende Wirkung erzielen dürfte. Eine sehr kompakte und vielleicht auch für Leistungskurs-Schüler/innen verdauliche Form wäre:

Die diskreten Zufallsvariablen X und Y seien gegeben durch die Verteilungen

$$\begin{matrix} x_1\ldots x_i\ldots \\ p_1\ldots p_i\ldots \end{matrix} \qquad \begin{matrix} y_1\ldots y_j\ldots \\ q_1\ldots q_j\ldots \end{matrix}.$$

Dann gilt mit $p_{ij}:=\mathrm{P}(X=x_i \text{ und } Y=y_j)$:

$$\begin{aligned}\mathrm{E}(X+Y)&=\sum_{i,j}(x_i+y_j)\cdot p_{ij}=\sum_{i,j}x_i\cdot p_{ij}+\sum_{i,j}y_j\cdot p_{ij}\\ &=\sum_i\left(x_i\cdot\underbrace{\sum_j p_{ij}}_{p_i}\right)+\sum_j\left(y_j\cdot\underbrace{\sum_i p_{ij}}_{q_j}\right)\\ &=\sum_i x_i\cdot p_i+\sum_j y_j\cdot q_j=\mathrm{E}(X)+\mathrm{E}(Y)\end{aligned}$$

Hier ist allerdings die Frage, ob diese sehr komprimierte Schreibweise für Schüler/innen im Regelunterricht (Grundkurs) geeignet ist. Die Konsequenz ist, dass man im Schulunterricht meist auf eine Begründung verzichtet.

In einem Gutachten zu diesem Aufsatz stand zu lesen:

$$„\mathrm{E}(X+Y)=\sum_{i=1}^{n}(x_i+y_i)\cdot p_i=\sum_{i=1}^{n}x_i\cdot p_i+\sum_{i=1}^{n}y_i\cdot p_i$$
$$=\mathrm{E}(X)+\mathrm{E}(Y)$$

Bisher konnte das als Begründung im Unterricht immer ausreichen."

Diese Aussage ist problematisch, denn die Zufallsvariablen X und Y müssen ja nicht gleich viele Ausprägungen (n) haben, und schon gar nicht müssen die Werte x_i bzw. y_i mit jeweils derselben Wahrscheinlichkeit p_i angenommen werden. Es wird dabei auch gar nicht darauf eingegangen, dass eine Ausprägung z_k der Zufallsvariable $Z=X+Y$ auf verschiedenste Weise zustande kommen kann. Durch die angegebene Formelzeile wird eine mathematische Exaktheit vorgespielt, die nicht vorhanden ist. Deswegen wurde oben geschrieben: „Die in der Schule verbreitete Erwartungswertdefinition […] erlaubt leider keinen so kurzen Beweis mehr". Das schreibe ich nicht gegen den/die mir unbekannte/n Gutachter/in, sondern für den Fall, dass diese Sichtweise weiter verbreitet ist.

2 Formeln für Erwartungswert und Varianz

2.1 Binomialverteilung

Wir beschäftigen uns zunächst mit möglichen Plausibilitätsbegründungen, die keine strengen Beweise sind. Solche sind wichtig und wertvoll im Unterrichtsgeschehen und sollten in keiner Weise abqualifiziert werden.

Wie kann man $\mathrm{E}(X)=n\cdot p$ bei der Binomialverteilung mit den Parametern n und p plausibel machen? Nur zu sagen oder zu schreiben, „das ist einleuchtend", ist etwas wenig.

Wegen der wichtigen Grundvorstellung „Wahrscheinlichkeit $\xrightarrow{\text{vorhersagen}}$ relative Häufigkeit" (in der umgekehrten Richtung steht *schätzen* statt *vorhersagen*), ist klar, dass mit $n\cdot p$ die *erwartete* absolute Häufigkeit des betrachteten Ereignisses in einem Bernoulli-Experiment vom Umfang n angegeben – anders ausgedrückt: die absolute Häufigkeit *gut vorhergesagt* – wird, also der mittlere (auf lange Sicht zu erwartende) Wert der binomialverteilten Zufallsvariable X.

Eine andere Möglichkeit wäre bei *konkreten* (kleinen) Werten von n (also z. B. $n=2$ oder $n=3$) den zugehörigen Erwartungswert laut Definition auszurechnen. Bei $n=2$ ergibt sich $2\cdot p$ und bei $n=3$ erhält man $3\cdot p$. Das legt dann die Vermutung nahe, dass sich im Allgemeinen $n\cdot p$ für den Erwartungswert ergeben wird, was mit dem Hinweis „man kann das auch allgemein zeigen" in einer gewissen Weise erledigt werden kann.[2]

In der Universitätsmathematik sind die zugehörigen Beweise sehr kurz, weil man sich dort meist eines in der Stochastik üblichen Tricks bedient: Man zerlegt die Zufallsvariable X in eine Summe anderer – sogar *unabhängiger* – Zufallsvariablen (sogenannter *Indikatorvariablen* des Ereignisses A): $X=X_1+X_2+\ldots+X_n$. Die einzelnen Indikatorvariablen X_k können dabei nur die Werte 0 oder 1 annehmen, je nachdem, ob A beim k-ten Versuch eingetreten ist oder nicht. Erwartungswerte und Varianzen dieser einzelnen X_k sind sehr leicht zu bestimmen, und somit (durch Summenbildung) auch Erwartungswert und Varianz von X – also insgesamt ein sehr kurzer und eleganter Weg. Im Schulunterricht kann man leider wieder kaum so vorgehen, weil die Zerlegung in Indikatorvariablen die meisten Lernenden überfordern würde. Außerdem steckt auch die Additivität der Varianz für unabhängige Zufallsvariablen dahinter, wieder etwas, was in der Schule kein Standardstoff ist. Außerdem wird im Schulunterricht das Thema der Unabhängigkeit von Zufallsvariablen in der Regel nicht tief genug behandelt.

Im Folgenden sollen zwei alternative Beweise für die Erwartungswertberechnung angegeben werden. Die zweite Methode funktioniert auch bei der Varianzberechnung. Diese Darstellung soll nicht bedeuten, dass der Autor vorschlägt, auf jeden Fall in der Schule formale Begründungen (Beweise) – statt möglicher Plausibilitätsbetrachtungen – zu favorisieren, sondern es sollen nur *Möglichkeiten* aufzeigt werden.

Die erste Möglichkeit hat mit *bedingten Erwartungswerten* zu tun, ein Konzept, das in seiner Allgemeinheit in der Schule kaum zur Verfügung stehen wird.[3] Das braucht es aber auch gar nicht, es genügt die Vorstellung, dass man einen Gesamterwartungswert als gewichtete Summe von „Erwartungswerten in Spezialfällen" berechnen kann. Das ist völlig analog zu: Gesamtwahrscheinlichkeiten $\left(\mathrm{P}(B)\right)$ können als gewichtete Summe von „Wahrscheinlichkeiten in Spezialfällen" $\left(\mathrm{P}(B|A),\mathrm{P}(B|\overline{A})\right)$ berechnet werden (die Gewichte sind genau die Eintrittswahrscheinlichkeiten der Spezialfälle, vgl. Abb. 1):

$$\mathrm{P}(B)=P(A)\cdot\mathrm{P}(B|A)+\mathrm{P}(\overline{A})\cdot P(B|\overline{A}).$$

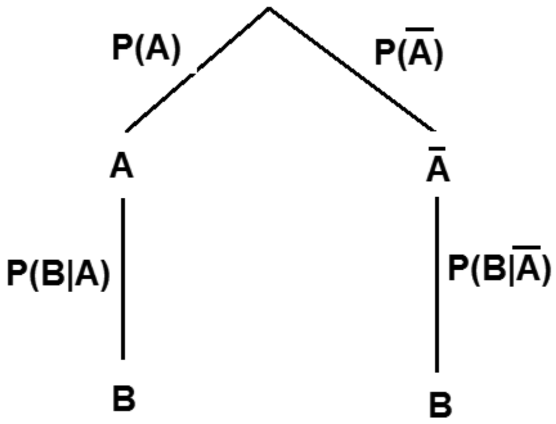


Abb. 1: Baumdiagramm für P(*B*) als gewichtetes Mittel in Spezialfällen

Ganz analog ist das bei Erwartungswerten (Abb. 2):

$$\mathrm{E}(X) = \mathrm{P}(A) \cdot \mathrm{E}(X|A) + \mathrm{P}(\overline{A}) \cdot \mathrm{E}(X|\overline{A}).$$

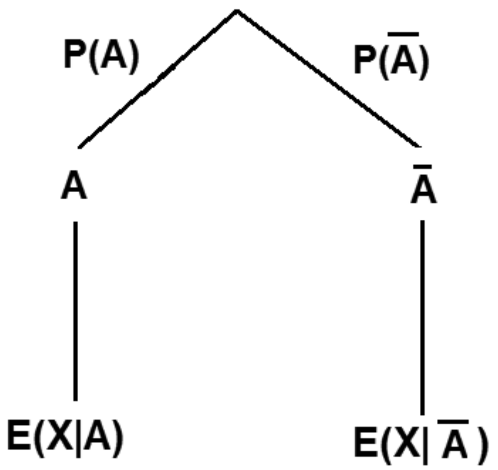


Abb. 2: Baumdiagramm für E(*X*) als gewichtetes Mittel in Spezialfällen

Wenn man weiß, dass $\mathrm{E}(X)$ in 20 % der Fälle den Wert E_1 hat und in 80 % der Fälle den Wert E_2, so ist wohl klar: $\mathrm{E}(X) = 0{,}2 \cdot E_1 + 0{,}8 \cdot E_2$ – mehr steckt hinter diesem Satz nicht.

Dabei ist nebensächlich, ob man diesen beiden Sätzen im Schulunterricht den fachlichen Namen gibt („Satz von der totalen Wahrscheinlichkeit", „Satz vom totalen Erwartungswert"); das zugehörige Phänomen sollte im Vordergrund stehen. Eine Veranschaulichung in Baumdiagrammen kann hier wertvolle Dienste leisten.

Mit diesem Prinzip können die Formeln für die Erwartungswerte der Binomial- und der geometrischen Verteilung begründet werden; auch darüber hinaus finden sich zahlreiche Anwendungen, die hier aber nicht thematisiert werden sollen (vgl. die Literatur in Anmerkung 3).

Weil bei diesem Zugang die Anzahl *n* der Versuchswiederholungen eine wichtige Rolle spielt, sei die Zufallsvariable *X* nun mit diesem Index versehen und heiße nun X_n.[4]

Mit A := „Erfolg beim 1. Versuch, d. h. das interessierende Ereignis tritt ein beim 1. Versuch" erhalten wir:

$$\mathrm{E}(X_n) = p \cdot \underbrace{\mathrm{E}(X_n|A)}_{1+\mathrm{E}(X_{n-1})} + (1-p) \cdot \underbrace{\mathrm{E}(X_n|\overline{A})}_{\mathrm{E}(X_{n-1})},$$

woraus sich unmittelbar

$$\mathrm{E}(X_n) = p + \mathrm{E}(X_{n-1})$$

ergibt. Man muss erkennen:

- $\mathrm{E}(X_n|A) = 1 + \mathrm{E}(X_{n-1})$:
  Wenn der 1. Versuch ein Erfolg war, dann hat man schon einen Erfolg („$1+$"), und es folgen im Mittel noch $\mathrm{E}(X_{n-1})$ Erfolge.
- $\mathrm{E}(X_n|\overline{A}) = \mathrm{E}(X_{n-1})$:
  Wenn der 1. Versuch kein Erfolg war, hat man einen Fehlversuch, und es folgen im Mittel noch $\mathrm{E}(X_{n-1})$ Erfolge.

Bei jeder Erhöhung von *n* um 1 kommt also beim Erwartungswert ein *p* dazu, m. a. W.: jede Versuchswiederholung liefert den gleichen Beitrag zum Erwartungswert. Es ist auch klar, dass dies auch schon für die erste Durchführung zutrifft (man kann bei Bedarf auch leicht $\mathrm{E}(X_1) = 1 \cdot p + 0 \cdot (1-p) = p$ explizit berechnen), so dass man schließlich hat:

$$\mathrm{E}(X_n) = n \cdot p.$$

Ein Vorteil dieser Methode ist, dass bei der Erwartungswertberechnung gar keine Summen mehr im Spiel sind, was für Schüler/innen in dieser Hinsicht einfacher ist. Außerdem kann dasselbe Prinzip auch bei der Berechnung des Erwartungswertes der geometrischen Verteilung angewandt werden (siehe unten). Der Nachteil ist, dass man den oben erläuterten Zusammenhang (Gesamterwartungswert als gewichtete Summe von Erwartungswerten in Spezialfällen) benötigt, und die gerade erwähnten Erkenntnisse $\mathrm{E}(X_n|A) = 1 + \mathrm{E}(X_{n-1})$ und $\mathrm{E}(X_n|\overline{A}) = \mathrm{E}(X_{n-1})$ weiterer Überlegungen bedürfen.

Die eben erarbeitete Beziehung $\mathrm{E}(X_n) = p + \mathrm{E}(X_{n-1})$ ist eine *Rekursion*sbeziehung, und dieses Prinzip wird in anderer Form auch bei der zweiten Möglichkeit angewandt.[5] Diese kann auch an mehreren Stellen angewandt werden, nämlich bei Erwartungswert und Varianz sowohl der Binomial- als auch der geometrischen Verteilung (siehe unten). Es ist so gesehen nicht nur ein einmaliger Trick, sondern eine effiziente „Methode", wichtige Parameter bei diskreten Verteilungen zu bestimmen. Sie würde z. B. auch bei der Poisson-Verteilung funktionieren; das ist kaum schulrelevant, zeigt aber die Mächtigkeit dieser „Methode". Auch scheint sie mir seitens der Schüler/innen gut nachvollziehbar zu sein.

Wenn man zur Abkürzung $p_k := \mathrm{P}(X = k)$ für $k = 0, 1, \ldots, n$ setzt, dann kann man bei der Binomialverteilung mit den Parametern n und p folgende Rekursionsbeziehung aufstellen:

$$p_{k+1} = \frac{n-k}{k+1} \cdot p_k \cdot \frac{p}{q} \qquad (*)$$

Eine Begründung dieser Rekursionsgleichung für die $B(n, p)$-Verteilung erhält man durch Einsetzen in die bekannte Formel

$$p_k = \mathrm{P}(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}.$$

Erwartungswert: $\mathrm{E}(X) = n \cdot p$

Aus (*) ergibt sich zunächst

$$q \cdot (k+1) \cdot p_{k+1} = n \cdot p \cdot p_k - k \cdot p \cdot p_k.$$

Summation über $\sum\limits_{k=0}^{n}$ auf beiden Seiten ergibt

$$q \cdot \underbrace{\sum_{k=0}^{n} (k+1) \cdot p_{k+1}}_{\mathrm{E}(X)} = n \cdot p \cdot \underbrace{\sum_{k=0}^{n} p_k}_{1} - p \cdot \underbrace{\sum_{k=0}^{n} k \cdot p_k}_{\mathrm{E}(X)}.$$

Man beachte, dass auf der linken Seite der Summand $0 \cdot p_0$ fehlt und der Summand $(n+1) \cdot \underbrace{p_{n+1}}_{0}$ zu viel ist, um $\mathrm{E}(X)$ darzustellen, aber beide Summanden sind ohnehin 0. Aus $q \cdot \mathrm{E}(X) = n \cdot p - p \cdot \mathrm{E}(X)$ folgt unmittelbar $\underbrace{(p+q)}_{1} \cdot \mathrm{E}(X) = n \cdot p$.[6]

Varianz: $\mathrm{V}(X) = n \cdot p \cdot q$

Das Prinzip ist dasselbe, nur die Durchführung ist ein wenig umfangreicher.

Multipliziert man (*) mit $(k+1)^2$, so erhält man:

$$q \cdot (k+1)^2 \cdot p_{k+1} = (k+1) \cdot (n \cdot p \cdot p_k - k \cdot p \cdot p_k)$$

Summation über $\sum\limits_{k=0}^{n}$ auf beiden Seiten ergibt:

$$q \cdot \underbrace{\sum_{k=0}^{n} (k+1)^2 \cdot p_{k+1}}_{\mathrm{E}(X^2)} = n \cdot p \cdot \underbrace{\sum_{k=0}^{n} k \cdot p_k}_{\mathrm{E}(X)} + n \cdot p \cdot \underbrace{\sum_{k=0}^{n} p_k}_{1}$$

$$- p \cdot \underbrace{\sum_{k=0}^{n} k^2 \cdot p_k}_{\mathrm{E}(X^2)} - p \cdot \underbrace{\sum_{k=0}^{n} k \cdot p_k}_{\mathrm{E}(X)}$$

$$\Leftrightarrow q \cdot \mathrm{E}(X^2) = \underbrace{n \cdot p}_{\mathrm{E}(X)} \cdot \mathrm{E}(X) + n \cdot p$$

$$- p \cdot \mathrm{E}(X^2) - p \cdot \underbrace{\mathrm{E}(X)}_{n \cdot p}$$

$$\Leftrightarrow \underbrace{(p+q)}_{1} \cdot \mathrm{E}(X^2) = (\mathrm{E}(X))^2 + n \cdot p - n \cdot p^2$$

Mit dem Verschiebungssatz erhält man daraus unmittelbar:

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = n \cdot p \cdot (1-p) = n \cdot p \cdot q$$

2.2 Geometrische Verteilung

Wie lange muss man beim Mensch-ärgere-dich-nicht im Durchschnitt auf die erste 6 warten? Diese Frage stellt sich wohl jedem einmal, auch wenn man nicht explizit darüber nachdenkt. Oft wird in Schulbüchern dazu nur mitgeteilt, dass die geometrische Verteilung mit Parameter p einen Erwartungswert von $\frac{1}{p}$ hat, vielleicht noch mit dem Hinweis, „das kann man auch beweisen", oder mit der Bemerkung, dass „dies ziemlich einleuchtend ist". Ist es auch, aber hier es ist u. E. sehr wichtig, trotzdem passende *Worte* zu finden. Es ist nicht schwierig, hier zumindest eine intuitive Erklärung, m. a. W. eine Plausibilitätsbegründung zu geben. Eine solche könnte folgendermaßen lauten:

Die Wahrscheinlichkeit für eine 6 ist $p = 1/6$. Eine besonders wichtige Grundvorstellung von Wahrscheinlichkeit ist: Sie ist eine gute Vorhersage für relative Häufigkeiten in einer langen Versuchsserie (Bernoulli-Experiment). Damit ist klar, dass die relative Häufigkeit der 6 auf lange Sicht wohl ca. 1/6 sein wird, das wiederum bedeutet, dass *durchschnittlich* jeder 6. Wurf eine 6 ist (bei gleichmäßiger Aufteilung wäre es genau jeder 6. Wurf, vgl. Abb. 3).

Abb. 3: Gleichmäßige Aufteilung der Sechsen

Wenn man sich die Sechsen nun so aufgereiht denkt, dann ist klar, dass 6 (allgemein: $1/p$) die durchschnittliche Wartezeit auf eine 6 (allgemein: auf ein Ereignis mit Wahrscheinlichkeit p) ist.

Ein bekannter *formaler* Beweis der Tatsache

$$\mathrm{E}(X) = \frac{1}{p}$$

verwendet gliedweises Differenzieren einer unendlichen geometrischen Reihe. Das ist bei Potenzreihen bekanntlich erlaubt innerhalb ihres Konvergenzradius, aber auf diese Problematik kann man in der Schule natürlich nicht eingehen. So gesehen liegt dieser Beweis technisch deutlich über Schulniveau.

Im Folgenden beschreiben wir drei verschiedene Möglichkeiten für einen Beweis der Erwartungswertformel der geometrischen Verteilung.

Mit „bedingten Erwartungswerten" (siehe oben: Binomialverteilung) angewandt auf die geometrische Verteilung und A := „Erfolg beim 1. Versuch" ergibt sich:

$$\mathrm{E}(X) = p \cdot \underbrace{\mathrm{E}(X|A)}_{1} + (1-p) \cdot \underbrace{\mathrm{E}(X|\overline{A})}_{1+\mathrm{E}(X)},$$

woraus sich unmittelbar

$$\mathrm{E}(X) = \frac{1}{p}$$

ergibt. Man muss erkennen:

- $\mathrm{E}(X|A) = 1$:

  Wenn der 1. Versuch ein Erfolg war, dann ist $X = 1$.
- $\mathrm{E}(X|\overline{A}) = 1 + \mathrm{E}(X)$:

  Wenn der 1. Versuch kein Erfolg war, hat man einen Fehlversuch („1"), und das Spiel beginnt von vorne („+ $\mathrm{E}(X)$").

Vor- und Nachteile dieser Methode wurden oben bei der Binomialverteilung schon genannt, sie gelten genauso auch hier.

Die zweite Methode der Erwartungswertberechnung involviert Summen, man braucht die Summenformel der unendlichen geometrischen Reihe ($q := 1 - p$):

$$\boxed{\mathrm{E}(X)} = \sum_{i=1}^{\infty} x_i \cdot \mathrm{P}(X = x_i) = \sum_{i=1}^{\infty} i \cdot q^{i-1} \cdot p \overset{p=1-q}{=}$$

$$= (1-q) \cdot \sum_{i=1}^{\infty} i \cdot q^{i-1} = (1-q) \cdot (q^0 + 2q^1 + 3q^2 + \ldots)$$

$$= (q^0 + 2q^1 + 3q^2 + \ldots) - (q^1 + 2q^2 + 3q^3 + \ldots)$$

$$= \underbrace{q^0}_{1} + q^1 + q^2 + \ldots = \frac{1}{1-q} = \boxed{\frac{1}{p}}$$

Diese Methode ist kurz und elegant, aber beim Ausmultiplizieren und Zusammenfassen (z. B. $3q^2 - 2q^2 = q^2$) werden Summanden (auch negative!) umgeordnet, und das ist nur bei absolut konvergenten Reihen erlaubt. Also wird auch hier etwas verwendet, dessen genaue Klärung im Rahmen des Schulunterrichts nicht möglich ist. Trotzdem ist dieser Beweis für den Schulunterricht m. E. geeignet.

Auch die Methode, die oben bei der Binomialverteilung gut funktioniert hat, kann hier verwendet werden (Rekursion); das ist hier dann die dritte Möglichkeit. Wir werden sehen: Auch dabei wird eigentlich umgeordnet in einer unendlichen Reihe, aber man hat ausschließlich positive Glieder, wodurch das Problem deutlich entschärft ist.

Wenn man zur Abkürzung $p_k := \mathrm{P}(X = k)$ für $k = 1, 2, \ldots$ setzt, dann kann man bei der geometrischen Verteilung mit Parameter p folgende besonders einfache Rekursionsbeziehung aufstellen (und leicht einsehen):

$$p_{k+1} = q \cdot p_k \qquad (**)$$

Wir multiplizieren auf beiden Seiten mit $k + 1$ und summieren auf beiden Seiten über $\sum\limits_{k=1}^{\infty}$:

$$\sum_{k=1}^{\infty} (k+1) \cdot p_{k+1} = q \cdot \sum_{k=1}^{\infty} (k+1) \cdot p_k$$

Auf der linken Seite steht hier die Summe $2 \cdot p_2 + 3 \cdot p_3 + \ldots$, es fehlt also genau der Wert $1 \cdot p_1 = p$ auf den Erwartungswert $\mathrm{E}(X)$, und auf der rechten Seite wird einfach ausmultipliziert. In diesem Ausmultiplizieren steckt natürlich ein Umordnen der Summanden, aber alle Summanden sind positiv.

$$\mathrm{E}(X) - p = q \cdot \underbrace{\sum_{k=1}^{\infty} k \cdot p_k}_{\mathrm{E}(X)} + q \cdot \underbrace{\sum_{k=1}^{\infty} p_k}_{1}$$

$$\Leftrightarrow \underbrace{(1-q)}_{p} \cdot \mathrm{E}(X) = \underbrace{p+q}_{1} \Leftrightarrow \mathrm{E}(X) = \frac{1}{p}$$

Man erkennt auch, dass hier im Prinzip dasselbe geschieht wie bei der zweiten Methode, nur etwas anders notiert; negative Summanden werden vermieden.

Nur um die Mächtigkeit dieser Methode zu zeigen, sei hier damit auch die Formel für die Varianz der geometrischen Verteilung begründet, auch wenn diese im Regelunterricht der Schule vermutlich selten eine Rolle spielt.

Wir multiplizieren (**) auf beiden Seiten mit $(k+1)^2$ und summieren über $\sum\limits_{k=1}^{\infty}$ auf beiden Seiten:

$$\sum_{k=1}^{\infty} (k+1)^2 \cdot p_{k+1} = q \cdot \sum_{k=1}^{\infty} (k+1)^2 \cdot p_k$$

Auf der linken Seite steht hier die Summe $2^2 \cdot p_2 + 3^2 \cdot p_3 + \ldots$, es fehlt also genau der Wert $1^2 \cdot p_1 = p$ auf den Erwartungswert $\mathrm{E}(X^2)$. Damit ergibt sich weiter (im Ausmultiplizieren auf der rechten Seite steckt wieder das „Umordnen“, aber wiederum nur positive Summanden!):

$$\mathrm{E}(X^2) - p = q \cdot \underbrace{\sum_{k=1}^{\infty} k^2 \cdot p_k}_{\mathrm{E}(X^2)} + 2q \cdot \underbrace{\sum_{k=1}^{\infty} k \cdot p_k}_{\mathrm{E}(X)} + q \cdot \underbrace{\sum_{k=1}^{\infty} p_k}_{1}$$

$$\Leftrightarrow \underbrace{(1-q)}_{p} \cdot \mathrm{E}(X^2) = (q+q) \cdot \underbrace{\mathrm{E}(X)}_{\frac{1}{p}} + \underbrace{p+q}_{1}$$

$$\Leftrightarrow p^2 \cdot \mathrm{E}(X^2) = q + \underbrace{q+p}_{1} \Leftrightarrow \mathrm{E}(X^2) = \frac{q+1}{p^2}$$

Dabei muss man in der Lage sein zu erkennen, dass $\sum_{k=1}^{\infty} k^2 \cdot p_k$ nichts anderes als $\mathrm{E}(X^2)$ darstellt.

Mit dem Verschiebungssatz erhält man weiter:

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

Für die Standardabweichung der nötigen Würfe bis zur ersten 6 beim Mensch-ärgere-dich-nicht-Spiel bedeutet dies

$$\sigma(X) = \sqrt{\mathrm{V}(X)} = \frac{\sqrt{q}}{p} = 6 \cdot \sqrt{\frac{5}{6}} = \sqrt{30} \approx 5{,}5.$$

Dies ist eine ziemlich starke Schwankung um den Erwartungswert 6, was viele in ihrer Kindheit (leidvoll?) erlebt haben.

3 Resümee

Es ist unbestritten, dass viele Begründungen in der Stochastik im Schulunterricht anders ablaufen müssen als in einem Mathematikstudium, denn das Abstraktionsniveau ist nicht dasselbe. Wichtige und oft angewandte Formeln wie z. B. für die Erwartungswerte der Binomial- und der geometrischen Verteilung sollten im Schulunterricht und in Schulbüchern aber nicht nur mitgeteilt werden mit dem Hinweis „der zugehörige Beweis ist zu kompliziert“. Nicht, dass eine solche Aussage im Schulunterricht prinzipiell abzulehnen wäre, aber in diesem speziellen Fall der Erwartungswerte gibt es viele Möglichkeiten, die das Schulniveau nicht übersteigen. Zumindest gut ausformulierte Plausibilitätsbetrachtungen sollten thematisiert werden. Denn sonst läuft der Unterricht Gefahr, zu einem Einsetzen in vorgesetzte Formeln zu degenerieren – und dieses Bild von Mathematik ist bekanntlich falsch und schädlich.

Im Aufsatz wurden mehrere Möglichkeiten aufgezeigt, formale Beweise beim in Rede stehenden Thema zu führen, die nicht Universitätsmathematik benutzen. Dies widerlegt die Ansicht, dass zugehörige formale Beweise notwendig das Schulniveau übersteigen würden. Jede Lehrkraft muss sorgfältig entscheiden, welche Begründungen hier gegeben werden (Plausibilitätsbegründungen oder doch formale Beweise) und es gibt auch gute Gründe, es im Schulunterricht bei Plausibilitätsbetrachtungen zu belassen (z. B. die begrenzte Unterrichtszeit lieber wo anders investieren). Aber immerhin besteht die *Möglichkeit*, bei Interesse auch formale Beweise zu thematisieren.

Einerseits haben Plausibilitätsbetrachtungen aus der Sicht des Faches den Nachteil, dass sie keine „strenge Gewissheit“ geben für Sachverhalte, sondern nur bestärkend wirken („das wird aus diesem oder jenem Grund wohl so sein“). *Die* Methode in der Wissenschaft Mathematik Gewissheit zu erlangen (formale Beweise) wird dabei nicht verwendet. Andererseits ist das Lernen von Mathematik nicht gleichzusetzen mit der Wissenschaft und Beweise haben ja nicht nur die Funktion Gewissheit zu schaffen (zu zeigen, *dass* etwas der Fall ist), sondern auch die Funktion zu klären, *warum* etwas der Fall ist. Bei diesem Aspekt können manchmal nicht formale Begründungen (evtl. auch „nur“ Plausibilitätsbetrachtungen) mehr leisten als formale Beweise, weil „das Wesentliche“ in formalen Beweisen (das das *Warum* für Mathematiker/innen klärt) für Lernende oft nicht sichtbar wird (zu technisch, zu abstrakt, etc.). Hier muss jede Lehrkraft immer wieder Abwägungen machen, die je nach Lerngruppe anders ausfallen können.

Anmerkungen

1 Wir beschränken uns hier auf *diskrete* Zufallsvariablen

2 Analog (kleine konkrete Werte von n nehmen und ausrechnen) kann man auch bei der Varianz vorgehen, um die Formel $\mathrm{V}(X) = n \cdot p \cdot q$ plausibel zu machen.

3 Zum Vorgehen vgl. Engel (1987, S. 65 f.), Henze (2001) und Humenberger (2000).

4 Hier übernimmt also X_n die Bedeutung der Zufallsvariablen X; bei den weiter oben erwähnten Indikatorvariablen hatte X_n eine andere Bedeutung. Ich danke Prof. Henze (Karlsruhe) für wichtige Hinweise zu dieser Methode.

5 Sie entstand aus einem Mailverkehr mit Dr. Resel (Wien). Vielleicht (sogar wahrscheinlich) gibt es in der Literatur irgendwo schon einen Hinweis auf die Methode *Rekursion* zum Zweck der Berechnung wichtiger Parameter bei diskreten Verteilungen; ich habe aber keinen gefunden, und kannte diese Methode bis zum besagten E-Mail-Verkehr auch nicht. Wenn jemand entsprechende Literaturstellen kennt, bitte ich um Bekanntgabe.

6 Beide Beweise sollten Lernende (zumindest in einem Leistungskurs) nicht prinzipiell überfordern. Man begreift dabei vielleicht auch etwas von dem sehr wichtigen *Warum* eines Sachverhaltes, nicht nur von seiner Gültigkeit. Wie Hanna (1990, S. 9) schreibt: „Ich bin der festen Überzeugung, dass es nützlich wäre, in die Diskussion eine explizite Unterscheidung einzuführen zwischen Beweisen, die beweisen, und Beweisen, die erklären.“

Literatur

Engel, A. (1987): *Stochastik.* Stuttgart: Klett.

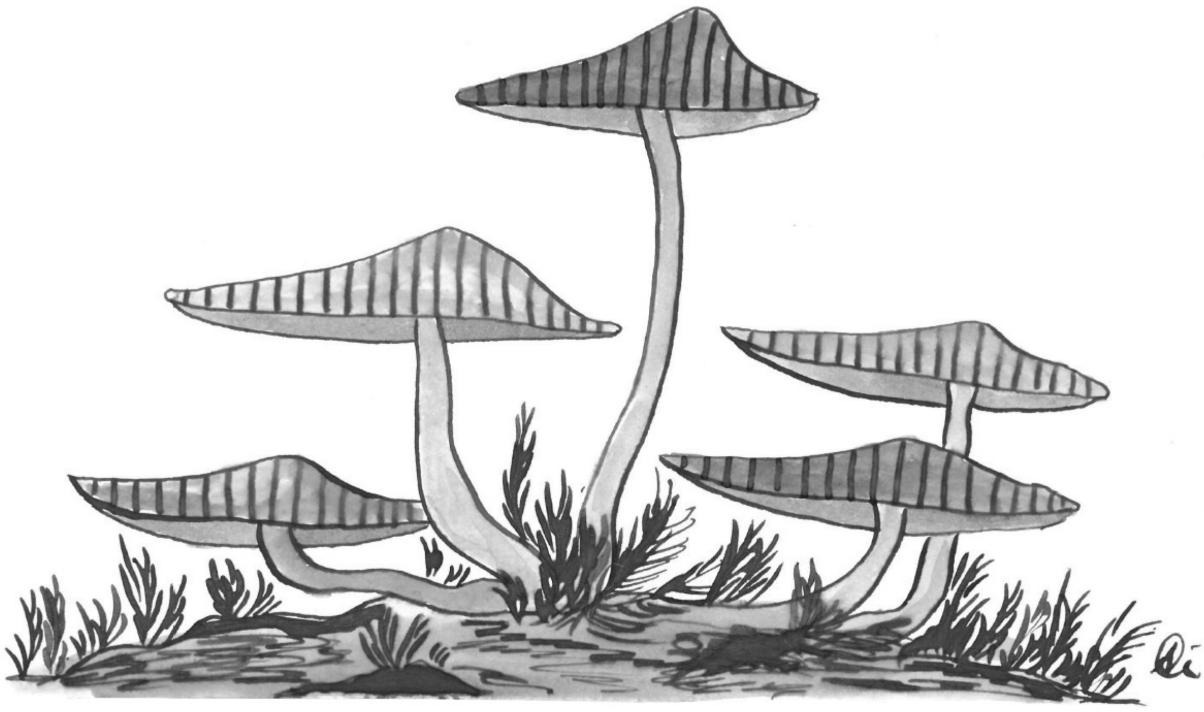
Hanna, G. (1990): Some Pedagogical Aspects of Proof. In: *Interchange* 21 (1), S 6–13.

Henze, N. (2001): Muster in Bernoulli-Ketten. In: *Stochastik in der Schule* 21 (2), S. 2–10.

Humenberger, H. (2000): Bedingte Erwartungswerte. In: *Stochastik in der Schule* 20 (1), S. 20–27.

Anschrift des Verfassers

Hans Humenberger
Fakultät für Mathematik
(und Zentrum für LehrerInnenbildung)
Universität Wien
Oskar-Morgenstern-Platz 1
A – 1090 Wien
hans.humenberger@univie.ac.at



Familie von Binomialpilzen
(fungus binomialis)
zum Parameter
$p \approx 2/3$