Wann kann eine Binomialverteilung überhaupt entstehen?

NORBERT HENZE, KARLSRUHE

Zusammenfassung: *Anlass für diesen Aufsatz ist die Frage einer Lehrkraft, ob die Binomialverteilung als Verteilung der Anzahl der Erfolge bei unabhängigen Bernoulli-Versuchen auch entstehen kann, wenn die Erfolgswahrscheinlichkeiten verschieden sind. Die Antwort ist Nein, und die Begründung dafür ist erhellend.*

1 Allgemeines zur Binomialverteilung

Die Binomialverteilung gilt momentan als Schlüsselkonzept für den gymnasialen Stochastikunterricht. Ausgangspunkt sind n „unabhängige Bernoulli-Versuche“ mit den möglichen Ausgängen *Erfolg* und *Misserfolg*. Dabei wird angenommen, dass die mit p bezeichnete Erfolgswahrscheinlichkeit über alle Versuche hinweg konstant ist. Modelliert die Zufallsgröße X die Anzahl der Erfolge in den n Versuchen, so besitzt X die Binomialverteilung $\mathrm{Bin}(n,p)$, d.h., es gilt

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

was üblicherweise als *Formel von Bernoulli* bezeichnet wird.

Ein nachhaltiges Wissen über die Binomialverteilung entsteht, wenn man sowohl deren additive Struktur als auch die aus der Unabhängigkeit der einzelnen Versuche resultierenden Konsequenzen verinnerlicht hat: Die Zufallsgröße X *summiert etwas auf*. Die Anzahl der Erfolge ist nichts anderes als die Anzahl der Einsen, wenn man jeden Erfolg als 1 und jeden Misserfolg als 0 codiert. Dieser konzeptionellen Einsicht ist es wenig förderlich, wenn in Schulbüchern je nach Situation Erfolg und Misserfolg als K/Z (Kopf/Zahl), L/R (Links/Rechts) oder T/F (Treffer/Fehlschlag) notiert werden, siehe z.B. Lergenmüller u.a. (2012). Auch die im Zusammenhang mit der Binomialverteilung langweiligen Baumdiagramme (siehe Henze (2023b)) unterstützen nicht diese Einsicht in die additive Struktur. Ein Verständnis für die stochastische Unabhängigkeit von Ereignissen, die sich auf unterschiedliche Bernoulli-Versuche beziehen, wird etwa dadurch erreicht, dass man die Ergebnisse der Bernoulli-Versuche als binäres n-Tupel $(a_1,\ldots,a_n)$ notiert. Dabei bezeichnet $a_j=1$ und $a_j=0$ einen Erfolg bzw. einen Misserfolg im j-ten Versuch. Das konkrete 10-Tupel $(0,1,1,0,0,0,1,0,0,1)$ ist also selbstredend. Es macht auch klar, dass bei unabhängigen Bernoulli-Versuchen keine Verzweigungen betrachtet werden müssen. Die 10 Versuche, die dieses Tupel ergeben, könnten von 10 Personen zeitlich und räumlich getrennt durchgeführt worden sein! Binäre n-Tupel ermöglichen auch unmittelbar eine begriffliche Einführung des Binomialkoeffizienten $\binom{n}{k}$: Dieser ist die Anzahl der binären n-Tupel mit genau k Einsen, siehe z.B. Henze (2023a). Ein Zusammenhang mit einer k-Auswahl von insgesamt n Dingen ohne Beachtung der Reihenfolge ist unmittelbar gegeben, wenn man sich die n Objekte über die einzelnen Plätze des Tupels notiert denkt. Unter den Objekten, die man ausgewählt hat, schreibt man eine Eins, unter die anderen eine Null. Da das obige 10-Tupel nach der ersten Pfadregel und wegen der Kommutativität der Multiplikation die Wahrscheinlichkeit $p^4(1-p)^6$ besitzt, wird auch sofort klar, warum hier noch mit $\binom{10}{4}$ multipliziert werden muss, damit die Wahrscheinlichkeit entsteht, dass in den 10 Versuchen insgesamt 4 Treffer auftreten.

Ein weiterer Vorteil der Tupel besteht darin, dass man bei Bernoulli-Versuchen geradezu gedrängt wird, eine Zufallsgröße als das zu begreifen, was sie wirklich ist, nämlich keine Variable (siehe z.B. Lergenmüller u.a. (2012), S. 109), sondern als eine Abbildung (Funktion, Zuordnung, Algorithmus), die (in unserem Fall) jedem n-Tupel $(a_1,\ldots,a_n)$ die Summe $a_1+\ldots+a_n$ zuordnet, denn X modelliert ja die Anzahl der Erfolge. Die formale Definition ist also

$$X\big((a_1,\ldots,a_n)\big):=a_1+\ldots+a_n.$$

Diese Einsicht liefert auch unmittelbar einen begrifflichen Beweis des Additionsgesetzes für die Binomialverteilung. Dieses als Satz formulierte Gesetz lautet wie folgt:

Satz: Sind X und Y stochastisch unabhängige Zufallsvariablen mit den Binomialverteilungen $\mathrm{Bin}(m,p)$ bzw. $\mathrm{Bin}(n,p)$, so besitzt die Summe $X+Y$ die Binomialverteilung $\mathrm{Bin}(m+n,p)$.

Um diesen Sachverhalt auf Schulniveau einzusehen, benötigt man nur die oben eingeführte Tupel-Darstellung. Wir denken uns insgesamt $m+n$ unabhängige Bernoulli-Versuche mit gleicher Erfolgswahrscheinlichkeit durchgeführt. Die zufällige Anzahl der Erfolge in den ersten m Versuchen sei X, und Y stehe für die Anzahl der Erfolge in den weiteren n Versuchen. Dann haben X und Y die Binomialverteilungen $\mathrm{Bin}(m,p)$ bzw. $\mathrm{Bin}(n,p)$, und sie sind stochastisch unabhängig, da sie sich auf getrennte Blöcke unabhängiger Bernoulli-Versuche beziehen (hier muss man nicht weiter formalisieren). Die Summe $X+Y$ zählt aber die Treffer aus allen $m+n$ Versuchen und hat somit die Binomialverteilung $\mathrm{Bin}(m+n,p)$.

Ein formaler Beweis ist auch nicht schwierig, denn er beinhaltet nur eine Fallunterscheidung. Das Ereignis $\{X+Y=k\}$ tritt genau dann ein, wenn für ein $j \in \{0,\ldots,k\}$ die Zufallsgrößen X und Y die Werte j bzw. $k-j$ annehmen. Wir erhalten also

$$\mathbb{P}(X+Y=k) = \sum_{j=0}^{k} \mathbb{P}(X=j, Y=k-j).$$

Wegen der stochastischen Unabhängigkeit von X und Y gilt $\mathbb{P}(X=j,Y=k-j) = \mathbb{P}(X=j)\mathbb{P}(Y=k-j)$, und die vorausgesetzten Binomialverteilungen von X und Y sowie die Festsetzung $\binom{\ell}{r} := 0$, falls $r > \ell$, ergeben

$$\begin{aligned}
&\mathbb{P}(X+Y=k)\\
&= \sum_{j=0}^{k} \binom{m}{j} p^j (1-p)^{m-j} \binom{n}{k-j} p^{k-j} (1-p)^{n-k+j}\\
&= p^k (1-p)^{m+n-k} \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}\\
&= \binom{m+n}{k} p^k (1-p)^{m+n-k}.
\end{aligned}$$

Das letzte Gleichheitszeichen folgt durch eine Fallunterscheidung aller $\binom{m+n}{k}$ binären $(m+n)$-Tupel mit genau k Einsen danach, wie viele dieser Einsen auf den ersten m Plätzen dieses Tupels stehen.

Nach diesen verständnisorientierten Vorüberlegungen zur Binomialverteilung wenden wir uns nun der in der Zusammenfassung aufgeworfenen Frage zu. Dabei sei vereinbart, dass alle Erfolgswahrscheinlichkeiten größer als null und kleiner als eins seien.

2 Beantwortung der Eingangsfrage

Es seien $X_1,\ldots,X_n$ stochastisch unabhängige Zufallsgrößen, wobei X_j die Binomialverteilung $\mathrm{Bin}(1,p_j)$ besitze $(j=1,\ldots,n)$. Die Zufallsgröße modelliert also das Ergebnis eines j-ten Bernoulli-Versuchs mit der Erfolgswahrscheinlichkeit p_j, und

$$X := X_1 + \ldots + X_n \tag{1}$$

steht für die Anzahl der Erfolge aus allen n Versuchen. Die Eingangsfrage lautet, ob X für eine spezielle Wahl von $n \geq 2$ und $p_1,\ldots,p_n$ binomialverteilt sein kann. Auf jeden Fall ist klar, dass X jeden der Werte $0,1,\ldots,n$ annehmen kann und somit für X nur eine Binomialverteilung der Gestalt $\mathrm{Bin}(n,p)$ mit einem geeigneten p infrage kommt.

2.1 Maximaler Wert und Erwartungswert

Gilt $X \sim \mathrm{Bin}(n,p)$, so folgt $\mathbb{P}(X=n) = p^n$. Wegen (1) und der Unabhängigkeit von $X_1,\ldots,X_n$ gilt andererseits

$$\begin{aligned}
\mathbb{P}(X=n) &= \mathbb{P}(X_1=1)\cdot\ldots\cdot\mathbb{P}(X_n=1)\\
&= p_1\cdot\ldots\cdot p_n.
\end{aligned}$$

Somit erhalten wir

$$p = \sqrt[n]{p_1\cdot\ldots\cdot p_n}.$$

Die Erfolgswahrscheinlichkeit p ist also notwendigerweise das *geometrische Mittel* von $p_1,\ldots,p_n$. Auf der anderen Seite gelten sowohl $\mathbb{E}(X) = np$ (wegen $X \sim \mathrm{Bin}(n,p)$) als auch $\mathbb{E}(X) = p_1+\ldots+p_n$ (wegen (1) und $X_j \sim \mathrm{Bin}(1,p_j)$ für jedes j). Hieraus folgt

$$p = \overline{p} := \frac{1}{n}(p_1+\ldots+p_n), \tag{2}$$

und damit ist p zugleich das *arithmetische Mittel* von $p_1,\ldots,p_n$; es gilt also

$$\sqrt[n]{p_1\cdot\ldots\cdot p_n} = \overline{p}.$$

Wie können Schülerinnen und Schüler einsehen, dass in dieser Gleichung im Allgemeinen das Kleiner-Gleich-Zeichen „$\leq$“ steht und Gleichheit nur im Fall $p_1 = \ldots = p_n$ gilt? Hierzu benötigen wir nur, dass die Exponentialfunktion die Ungleichung

$$\mathrm{e}^x \geq 1+x, \quad x \in \mathbb{R}, \tag{3}$$

erfüllt, wobei das Gleichheitszeichen nur im Fall $x=0$ eintritt. Setzt man hier für jedes $j \in \{1,\ldots,n\}$

$$x := \frac{p_j}{\overline{p}} - 1 \tag{4}$$

und multipliziert die sich ergebenden Ungleichungen, so folgt wegen $\mathrm{e}^{a+b} = \mathrm{e}^a\mathrm{e}^b$

$$\exp\left(\sum_{j=1}^{n}\left(\frac{p_j}{\overline{p}}-1\right)\right) \geq \prod_{j=1}^{n}\frac{p_j}{\overline{p}}$$

Wegen $p_1+\ldots+p_n = n\overline{p}$ ist die Summe gleich null, und $\mathrm{e}^0 = 1$ ergibt $p_1\cdot\ldots\cdot p_n \leq \overline{p}^n$ und damit

$$\sqrt[n]{p_1\cdot\ldots\cdot p_n} \leq \overline{p}. \tag{5}$$

Da das Gleichheitszeichen in der verwendeten Ungleichung (3) nur für $x = 0$ eintritt, tritt „=“ in (5) nur ein, wenn für jedes $j \in \{1, \ldots, n\}$ die Gleichung $p_j/\overline{p} - 1 = 0$ gilt (vgl. (4)), und das ist nur im Fall $p_1 = \ldots = p_n = \overline{p}$ möglich (siehe auch Henze und Vehling (2024) für eine weitere Anwendung der Ungleichung zwischen arithmetischem und geometrischem Mittel in der Stochastik).

2.2 Erwartungswert und Varianz

Wir stellen jetzt ein zweites Argument zur Begründung der Gleichung $p_1 = \ldots = p_n = \overline{p}$ vor. Durch das Gleichsetzen der Erwartungswerte von $X_1 + \ldots + X_n$ und einer $\mathrm{Bin}(n, p)$-verteilten Zufallsgröße wissen wir, dass (2) gilt. Wegen $X \sim \mathrm{Bin}(n, p)$ ergibt sich

$$\mathbb{V}(X) = n\,\overline{p}\,(1 - \overline{p}). \tag{6}$$

Da die Varianz einer Summe unabhängiger Zufallsvariablen gleich der Summe der Varianzen der einzelnen Summanden ist, gilt andererseits

$$\begin{aligned} \mathbb{V}(X) &= p_1(1-p_1) + \ldots + p_n(1-p_n) \\ &= n\overline{p} - \left(p_1^2 + \ldots + p_n^2\right). \end{aligned}$$

Der Vergleich mit (6) zeigt, dass

$$n\overline{p}^2 = p_1^2 + \ldots + p_n^2 \tag{7}$$

gelten muss. Wir werden jetzt sehen, dass die linke Seite von (7) immer kleiner oder gleich der rechten Seite ist, und dass Gleichung (7) nur für den Fall $p_1 = \ldots = p_n$ gilt.

Hierzu setzen wir kurz $Q := p_1^2 + \ldots + p_n^2$ und definieren für jedes reelle x

$$\begin{aligned} f(x) &:= (p_1 x + 1)^2 + \ldots + (p_n x + 1)^2 \qquad (8) \\ &= Qx^2 + 2n\overline{p}x + n. \end{aligned}$$

Wegen (8) gilt $f(x) \geq 0$ für jedes x, und Nullsetzen der Ableitung

$$f'(x) = 2Qx + 2n\overline{p}$$

liefert die Minimalstelle $x_0 = -n\overline{p}/Q$. Es gilt

$$0 \leq f(x_0) = n - \frac{(n\overline{p})^2}{Q}$$

und damit $n\overline{p}^2 \leq Q$. Die linke Seite von (7) ist also höchstens gleich der rechten Seite (diese Aussage ist nichts anderes als die Cauchy–Schwarz-Ungleichung für die Vektoren $(1, \ldots, 1)$ und $(p_1, \ldots, p_n)$). Damit ist die Varianz einer Zufallsgröße mit der Binomialverteilung $\mathrm{Bin}(n, p)$ mindestens so groß wie die Varianz der in (1) definierten Zufallsvariablen X, die den gleichen Erwartungswert besitzt. Weil aber in (7) das Gleichheitszeichen steht, folgt $f(x_0) = 0$. Da $f(x_0)$ nach (8) eine Summe von Quadraten ist, muss $(p_j x_0 + 1)^2 = 0$ und somit $p_j x_0 + 1 = 0$ für jedes $j = 1, \ldots, n$ gelten. Dies ist nur im Fall $p_1 = \ldots = p_n =: p$ möglich, und dann ist $x_0 = -1/p$.

3 Fazit

Die Frage einer Lehrkraft, ob die Anzahl X der Erfolge aus n unabhängigen Bernoulli-Versuchen mit Trefferwahrscheinlichkeiten $p_1, \ldots, p_n$ binomialverteilt sein kann, lässt sich wie folgt beantworten: Ja, aber nur dann, wenn in der Ungleichung

$$\sqrt[n]{p_1 \cdot \ldots \cdot p_n} \leq \frac{1}{n}(p_1 + \ldots + p_n)$$

zwischen dem geometrischen und dem arithmetischen Mittel oder in der Cauchy–Schwarz-Ungleichung

$$(p_1 \cdot 1 + \ldots + p_n \cdot 1)^2 \leq (p_1^2 + \ldots + p_n^2) \cdot n$$

für die Vektoren $(p_1, \ldots, p_n)$ und $(1, 1, \ldots, 1)$ das Gleichheitszeichen eintritt. In jedem dieser Fälle ist das nur für $p_1 = \ldots = p_n$ möglich.

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

Literatur

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

Henze, N. (2023b): Weg mit der Bernoulli-„Kette“! *Stochastik in der Schule* 43(1), 19–23.

Henze, N., und Vehling, R. (2024): Abrücken von der Gleichverteilung bei Fächerbesetzungen: Stochastik trifft Analysis. *Stochastik in der Schule* 44(1), 19–26.

Lergenmüller u.a. (Hrsg.) (2012): Mathematik Neue Wege. Arbeitsbuch für Gymnasien. Stochastik. Bildungshaus Schulbuchverlage, Braunschweig.

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