KARIN BINDER, STEFAN KRAUSS, CHRISTOPH WASSNER

Der Häufigkeitsdoppelbaum

Anteilswerte und bedingte Wahrscheinlichkeiten vorteilhaft visualisieren

LERNGRUPPE: 5.–13. Schuljahr

IDEE: Bedingte Wahrscheinlichkeiten und Anteilswerte werden anhand von Häufigkeitsdoppelbäumen besonders gut verstanden.

ONLINE-MATERIAL: Vorlage Häufigkeitsdoppelbaum

ZEITBEDARF: 2 Unterrichtsstunden

Der Umgang mit bedingten Wahrscheinlichkeiten führt bei Schülerinnen und Schülern immer wieder zu Problemen. Auch in realen Kontexten, wie beispielsweise der Medizin oder der Rechtsprechung, kommt es dabei wiederholt zu Fehlurteilen, die sogar zu unnötigen Operationen oder Gefängnisstrafen für unschuldige Personen führen können – vor allem bei sogenannten Bayesianischen Situationen (Borovcnik 2016, Binder/Vogel 2018). Dabei handelt es sich um Aufgaben, in denen auf eine bedingte Wahrscheinlichkeit $P_B(A)$ geschlossen werden soll und die mit der Formel von Bayes oder mithilfe von Pfadregeln gelöst werden können.

Bedingte Wahrscheinlichkeiten besser verstehen

Zwei Strategien sind bekannt, die das Verständnis bedingter Wahrscheinlichkeiten verbessern – zum einen die Verwendung absoluter Häufigkeiten anstelle von Wahrscheinlichkeiten und zum anderen die Visualisierung der statistischen Informationen.

Absolute Häufigkeiten werden besser verstanden als relative

Aufgaben mit bedingten Wahrscheinlichkeiten oder Anteilswerten werden nur selten richtig gelöst (dies zeigen viele Studien, vgl. Gigerenzer/Hoffrage 1995). Betrachten wir hierzu die folgende Aufgabe:

→ **Snapchat-Aufgabe (Wahrscheinlichkeiten)**
An deiner Schule sind 40 % Mädchen (M) und 60 % Jungen (J). Eine aktuelle Studie zur Mediennutzung von Schülerinnen und Schülern zeigt, dass 30 % der Mädchen, aber nur 10 % der Jungen den Messaging-Dienst Snapchat (S) nutzen. Du triffst in der Schule einen Snapchat-User. Wie wahrscheinlich ist es, dass es sich dabei um ein Mädchen handelt?

Die richtige Lösung lautet:

$$P_S(M) = \frac{P_M(S) \cdot P(M)}{P_M(S) \cdot P(M) + P_J(S) \cdot P(J)} = \frac{2}{3}$$

Die Aufgabe wird erfahrungsgemäß leichter lösbar, wenn alle Informationen mit absoluten Häufigkeiten dargeboten werden:

→ **Snapchat-Aufgabe (absolute Häufigkeiten)**
An deiner Schule sind von 500 Jugendlichen 200 Mädchen und 300 Jungen. Von den 200 Mädchen nutzen 60 Snapchat, von den 300 Jungen nutzen nur 30 Snapchat. Du triffst in der Schule einen Snapchat-User. Wie wahrscheinlich ist es, dass es sich dabei um ein Mädchen handelt? Oder anders gefragt: Wie viele der Snapchat-Nutzer deiner Schule sind Mädchen?

Nun sieht man die richtige Lösung deutlich schneller. Insgesamt gibt es 90 Snapchat-Nutzer an der Schule. Von diesen 90 Snapchat-Nutzern sind 60 Mädchen. Die richtige Antwort lautet also „Die Wahrscheinlichkeit, dass es sich bei dem Snapchat-User um ein Mädchen handelt, ist $\frac{2}{3}$ bzw. ca. 66,7%“.

Visualisierung der Informationen als zusätzliche Unterstützung

In der Schule werden Wahrscheinlichkeiten vor allem mit Vierfeldertafeln oder mit Baumdiagrammen visualisiert, die entweder mit absoluten oder mit relativen Häufigkeiten ausgefüllt werden (**Abb. 1**).

Erstaunlicherweise werden Baumdiagramme mit absoluten Häufigkeiten in den Knoten (Wassner u. a. 2004), die sich in Studien als besonders verständnisfördernd erwiesen haben (Binder u. a. 2018), in der Schule bislang kaum eingesetzt. **Abb. 1** (unten) zeigt zusätzlich auch noch Doppelbäume mit absoluten und mit relativen Häufigkeiten, die wir hier als neue Visualisierungsmöglichkeit vorstellen.

Es gibt in der didaktischen Literatur noch zahlreiche weitere Visualisierungsmöglichkeiten, wie beispielsweise Einheitsquadrate (Böcherer-Linder u. a. 2018) oder ikonische Darstellungen, die hier nicht weiter aufgegriffen werden sollen.

Welche der Visualisierungen aus **Abb. 1** unterstützen nun das Verständnis der Schülerinnen und Schüler mehr und welche weniger?

Empfehlungen für Visualisierungen

In Visualisierungen bedingter Wahrscheinlichkeiten sind unserer Ansicht nach vier zentrale Aspekte didaktisch hilfreich, die im Folgenden diskutiert werden sollen (**Tab. 1**).

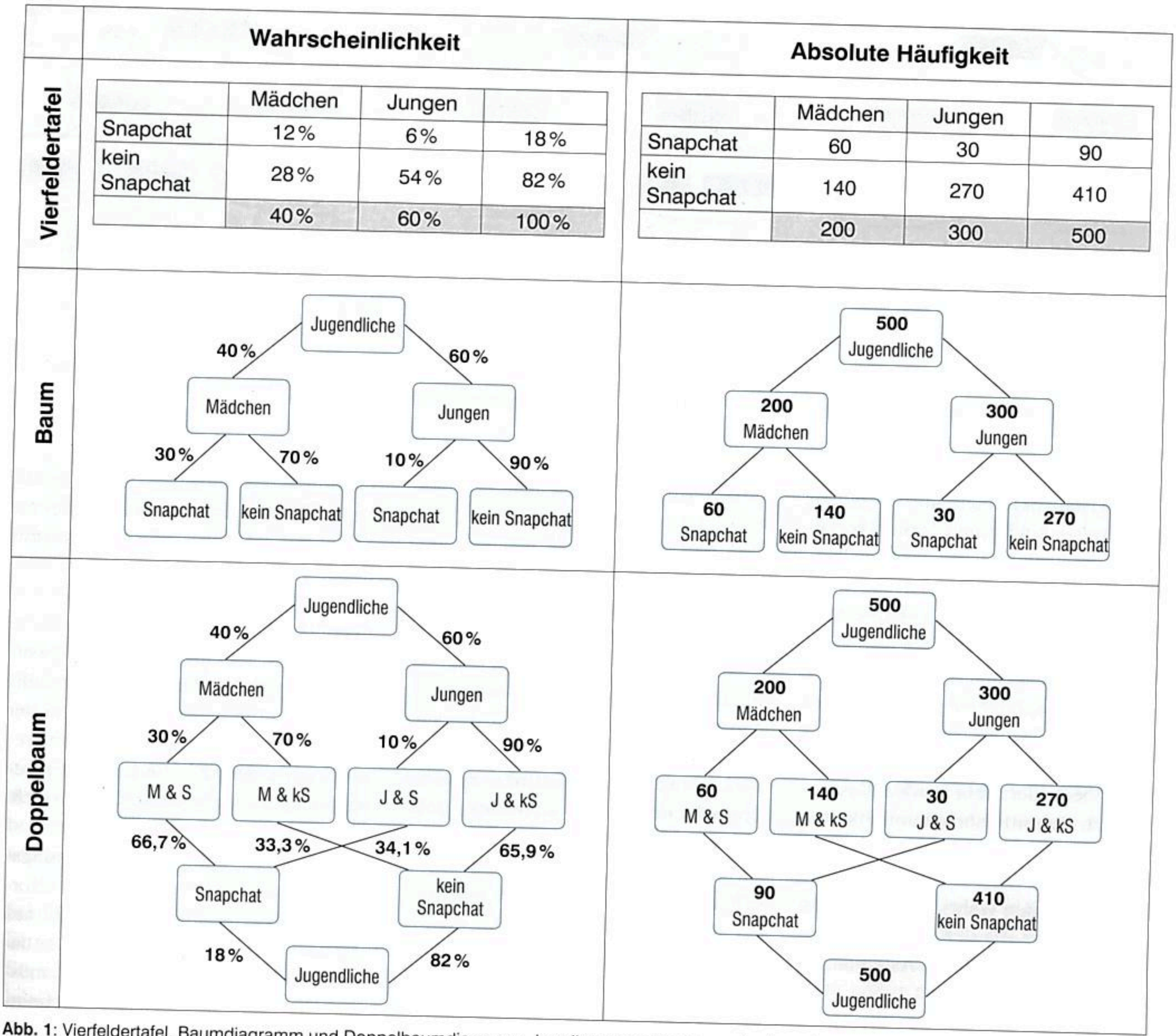


Abb. 1: Vierfeldertafel, Baumdiagramm und Doppelbaumdiagramm, jeweils mit relativen bzw. absoluten Häufigkeiten

	relative Häufigkeit			absolute Häufigkeit			beide (rel. + abs. Häufigkeit)	
Eigenschaft	Vierfelder-tafel	Baum	D-Baum	Vierfelder-tafel	Baum	D-Baum	Baum	D-Baum
enthält absolute Häufigkeiten				✔	✔	✔	✔	✔
bedingte Wahrscheinlichkeiten ablesbar		✔	✔				✔	✔
sequenzieller Charakter ist sichtbar		✔	✔		✔	✔	✔	✔
beide Leserichtungen sichtbar	✔		✔	✔		✔		✔

Tab. 1: Welche verständnisfördernden Eigenschaften besitzen die jeweiligen Visualisierungen?

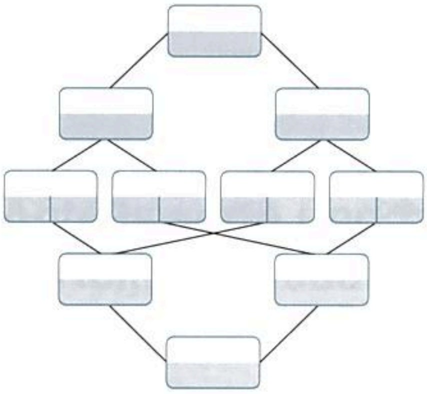


Abb. 2: Struktur eines Doppelbaums

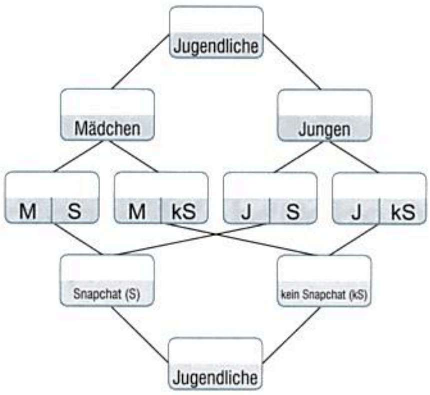


Abb. 3: Beschriftung des Baums

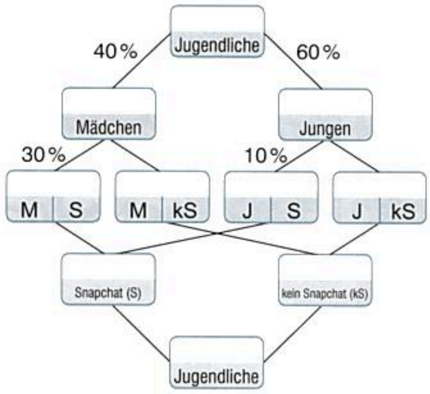


Abb. 4: Baumdiagramm mit allen Informationen aus der Aufgabe

1. Visualisierungen mit *absoluten Häufigkeiten* sind grundsätzlich vorzuziehen, da absolute Häufigkeiten zu konkreten Personengruppen (z. B. Jugendliche) gehören oder Objekten entsprechen, die man sich leichter vorstellen kann als Wahrscheinlichkeiten oder Anteile.

2. *Baumdiagramme* sind vor allem dann vorzuziehen, wenn *bedingte Wahrscheinlichkeiten* visualisiert werden sollen. In Vierfeldertafeln werden dagegen lediglich Schnittwahrscheinlichkeiten abgebildet. Man beachte, dass in der Vierfeldertafel (**Abb. 1**, oben) kein Platz für die bedingten Wahrscheinlichkeiten (z. B. 30 %) ist, die sich aber im Baumdiagramm bequem an die Äste platzieren lassen.

3. In Baumdiagrammen bleibt der sequenzielle Charakter der Aufgabe erhalten (1. Merkmal: Geschlecht, 2. Merkmal: nutzt Snapchat). In *Doppelbäumen* sind darüber hinaus beide Leserichtungen (erst Geschlecht, dann Snapchat vs. erst Snapchat, dann Geschlecht) gleichermaßen sichtbar. Die beiden Leserichtungen entsprechen den beiden möglichen sequentiellen Aufteilungen einer Personengruppe oder Objekten nach zwei Merkmalen.

4. Baumdiagramme haben darüber hinaus den Vorteil, dass sich Wahrscheinlichkeiten und absolute Häufigkeiten *gleichzeitig* darstellen lassen, da man einen Baum mit relativen Häufigkeiten und einen Baum mit absoluten Häufigkeiten „übereinander schieben" kann (z. B. Abb. 1 links und rechts), was mit den entsprechenden Vierfeldertafeln nicht ohne Weiteres möglich ist (in jeder Zelle steht immer nur eine Zahl).

Der Doppelbaum

Über die genannten Vorteile hinaus, hat der Häufigkeitsdoppelbaum vor allem drei entscheidende Merkmale (vgl. Binder/Krauss/Wassner 2018):

1. Zur Lösung von Aufgaben mit bedingten Wahrscheinlichkeiten benötigt man weder die Pfadregeln noch den Satz von Bayes.

2. Durch die Konstruktion eines Doppelbaumes lässt sich eine große Klasse von Aufgaben lösen, darunter fallen Aufgaben zu bedingten Wahrscheinlichkeiten, zu Schnittwahrscheinlichkeiten, Aufgaben, die bislang mit Pfadregeln gelöst werden usw.

3. Der Doppelbaum lässt sich bequem bereits in der Unterstufe bei Aufgaben zu Anteilswerten einsetzen und ist ein Werkzeug, das bis zum Abitur tragen kann („Spiralcurriculum"). Selbst viele Abituraufgaben zu bedingten Wahrscheinlichkeiten lassen sich so durch einfache Mittel (Prozentrechnung, Bruchrechnung) lösen (Binder u. a. 2018). Wird nur der Satz von Bayes angewendet, finden Schülerinnen und Schüler die Ergebnisse von Bayes-Aufgaben häufig nicht „einsichtig". Der kognitive Konflikt, der oft bei Bayesianischen Aufgaben entsteht, kann nur mit absoluten Häufigkeiten wirklich aufgelöst werden.

Der Häufigkeitsdoppelbaum

Im Folgenden soll die Konstruktion eines Häufigkeitsdoppelbaumes beschrieben werden, wie sie von der Unterstufe bis zum Abitur zur Anwendung kommen kann. Der Doppelbaum wird dabei als Struktur vorgegeben und gemeinsam mit den Lernenden gefüllt.

Schritt 1: Zeichnen der Struktur

Zunächst wird eine leere Struktur eines Doppelbaumdiagramms gezeichnet, wobei die untere graue Hälfte der Knoten jeweils für die Beschriftung reserviert ist (**Abb. 2**). Die genaue Fragestellung muss an dieser Stelle noch nicht beachtet werden.

Die Zweiteilung der Beschriftungsfelder in der mittleren Ebene ist erforderlich, weil dort die Informationen aus der Ebene darüber und aus der Ebene darunter zusammengeführt werden müssen. Hier werden dann also die Schnitt-Ereignisse dargestellt, beispielsweise: „Mädchen und Snapchat".

Natürlich kann der Doppelbaum auch um 90 Grad gedreht dargestellt werden, wie das auch bei normalen Baumdiagrammen manchmal gern gemacht wird.

Schritt 2: Beschriftung

Im nächsten Schritt erfolgt die Beschriftung jeweils in der unteren Hälfte der Knoten (**Abb. 3**).

Schritt 3: Eintragen der Daten

Im nächsten Schritt werden nach und nach alle statistischen Informationen an die Pfade des Baumes geschrieben, die in der Aufgabenstellung gegeben sind (**Abb. 4**). In der Snapchat-Aufgabe ist zunächst die Information gegeben, dass 40 % der Jugendlichen Mädchen sind und 60 % Jungen.

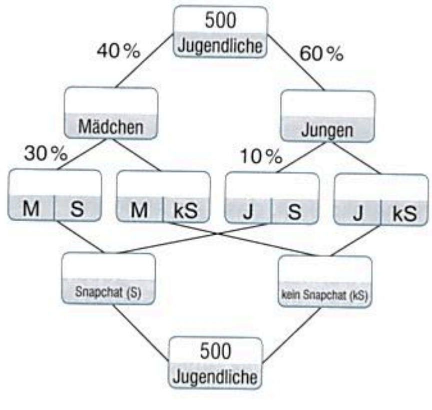


Abb. 5: Eintrag der (imaginären) Stichprobe

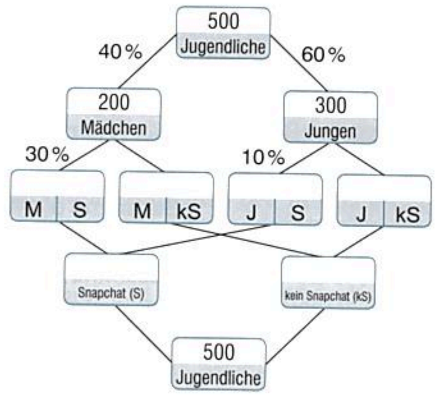


Abb. 6: Eintragen der Anzahlen der Mädchen und Jungen in die Knoten

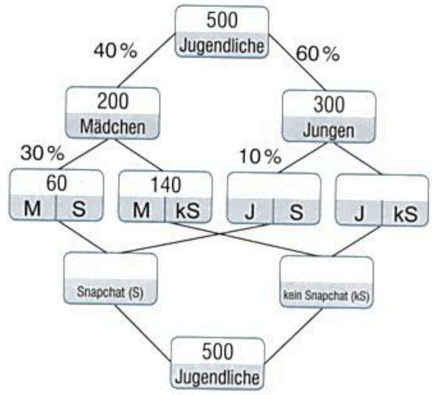


Abb. 7: Eintragen der Mädchen, die Snapchat nutzen bzw. nicht nutzen

Diese beiden Zahlen werden an die passenden Äste im Baumdiagramm eingetragen. Die nächste Information lautet, dass 30 % der Mädchen Snapchat nutzen. Auch für diese Zahl wird der passende Ast gesucht und die Information dort notiert. Schließlich wird noch in das Baumdiagramm geschrieben, dass 10 % der Jungen Snapchat nutzen.

Im nächsten Schritt erfolgt der Übergang zu absoluten Häufigkeiten. Ist in der Aufgabe keine konkrete Stichprobe gegeben, ist dies der didaktisch entscheidende Schritt.

Schritt 4: Wahl einer Stichprobe

Durch die Wahl einer „imaginären" Stichprobe wird es nun möglich, sich konkrete Personengruppen oder Objekte – in unserem Fall: Jugendliche – vorzustellen. Oft ist es ratsam, mit einer Stichprobe von 100 oder 1.000 zu beginnen. In unserem Beispiel betrachten wir 500 Jugendliche, die im obersten und im untersten Knoten eingetragen werden (**Abb. 5**).

Sind in der Aufgabenstellung sehr große oder sehr kleine Prozentzahlen gegeben (wie das z. B. oft bei medizinischen Kontexten der Fall ist), so muss eine entsprechend große fiktive Stichprobe gewählt werden, um in der mittleren Ebene ganze Zahlen zu erhalten. Durch ein einfaches „Anhängen von Nullen" kann das aber schnell korrigiert werden, falls man im ersten Schritt eine zu kleine Stichprobe ausgewählt hat (Man kann sich auch überlegen, wie man selbst bei Anteilswerten wie $\frac{1}{7}$ oder $\frac{1}{3}$ an den Ästen dennoch in der mittleren Ebene ganze Zahlen erhalten kann, vgl. Binder u. a. 2018).

Schritt 5: Was bedeuten die Informationen für die Stichprobe?

Im nächsten Schritt wird nun überlegt, was die gegebenen statistischen Informationen an den Ästen für die Stichprobe von 500 Jugendlichen bedeuten. Die 40 % bedeuten beispielsweise, dass von 500 Jugendlichen 200 Mädchen sind, während es sich bei den restlichen 300 Jugendlichen um Jungen handelt (**Abb. 6**).

Als Nächstes kann man die 30 % in ganze Zahlen übersetzen: Wenn 30 % der 200 Mädchen Snapchat nutzen, so bedeutet dies, dass 60 Mädchen Snapchat nutzen und 140 Mädchen Snapchat nicht nutzen. Auch diese absoluten Häufigkeiten können nun in den Doppelbaum eingetragen werden (**Abb. 7**).

Schließlich kann nun eingetragen werden, was es bedeutet, dass 10 % der Jungen Snapchat nutzen: Demnach nutzen 30 der 300 Jungen Snapchat, während 270 Snapchat nicht nutzen (**Abb. 8**).

Die letzten beiden noch leeren Knoten lassen sich nun aufgrund der einfachen additiven Strukturen bequem ergänzen: Insgesamt nutzen 90 Jugendliche Snapchat und 410 Jugendliche nutzen Snapchat nicht (**Abb. 9**).

Schritt 6: Ablesen der Lösung

Im sechsten und letzten Schritt können nun die Lösungen zu beliebigen Frage-

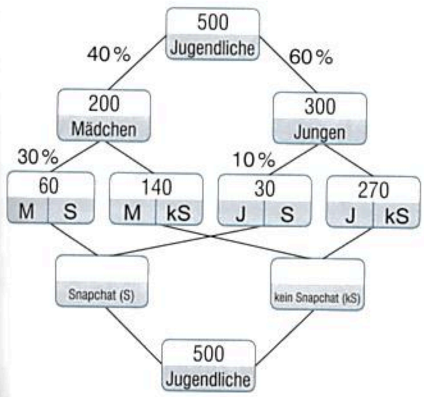


Abb. 8: Eintragen der Jungen, die Snapchat nutzen bzw. nicht nutzen

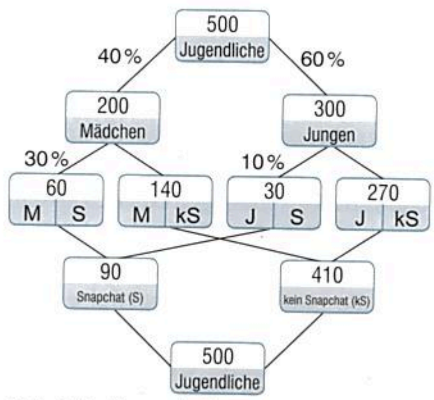


Abb. 9: Ergänzen der letzten beiden Knoten im Baum

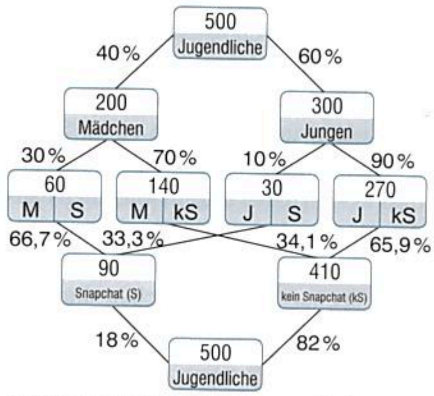
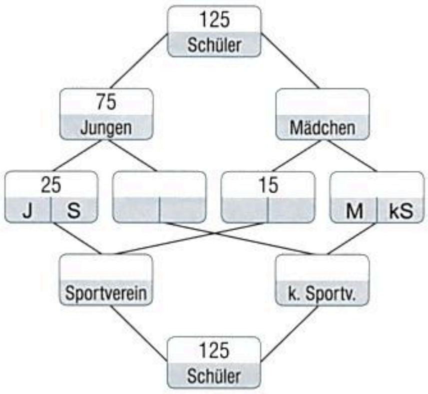


Abb. 10: Doppelbaumdiagramm mit allen absoluten und relativen Häufigkeiten

WISSENSWERT

Der Doppelbaum – eine Visualisierung für alle Klassenstufen



Aufgabe 1 – Klasse 6

Die 125 Schülerinnen und Schüler der 6. Klassen im Geschwister-Scholl-Gymnasium werden gefragt, ob sie einem Sportverein angehören oder nicht. 25 der 75 Jungen und 15 Mädchen gehören einem Sportverein an (siehe auch Baumdiagramm, oben).

- Ergänze die Beschriftungen in den grauen Feldern in der Mitte des Baumdiagramms.
- Fülle nun den Doppelbaum mit den noch fehlenden absoluten Häufigkeiten aus.
- Wie viele Mädchen sind nicht in einem Sportverein?
- Wie groß ist unter allen Sportvereinsmitgliedern der Anteil der Mädchen?

Aufgabe 2 – Klasse 11

Eine Untersuchung bei Neugeborenen soll eine Stoffwechselkrankheit erkennen. Dabei wird in 0,1 % aller Fälle eine vorliegende Stoffwechselerkrankung nicht entdeckt, während diese Untersuchungsmethode in 0,01 % aller Fälle irrtümlich eine Krankheit anzeigt. Durchschnittlich haben bei 1 100 000 Geburten 200 Neugeborene diese Krankheit.

- Zeichne einen Doppelbaum mit absoluten und mit relativen Häufigkeiten, der diesen Sachverhalt visualisiert.
- Bestimme die Wahrscheinlichkeit, dass ein als krank diagnostiziertes Baby diese Krankheit wirklich hat.
- Bestimme die Wahrscheinlichkeit, dass ein als gesund diagnostiziertes Baby auch nicht unter dieser Krankheit leidet.
- Zeige, dass die Seltenheit der Krankheit einen sehr großen Einfluss auf die Wahrscheinlichkeit hat, ob ein als krank diagnostizierter Säugling diese Krankheit wirklich hat oder nicht.

Aufgabe 3 – Abitur Bayern 2012

Nachdem die Verfilmung eines bekannten Romans erfolgreich in den Kinos gezeigt wurde, veröffentlicht eine Tageszeitung das Ergebnis einer repräsentativen Umfrage unter Jugendlichen. Der Umfrage zufolge hatten 88 % der befragten Jugendlichen den Roman zum Zeitpunkt des Kinostarts noch nicht gelesen, 18 % sahen die Verfilmung. Von den Befragten, die laut Umfrage den Roman zum Zeitpunkt des Kinostarts bereits gelesen hatten, gaben 60 % an, die Verfilmung gesehen zu haben.

Betrachtet werden folgende Ereignisse:
R: „Eine aus den Befragten zufällig ausgewählte Person hatte laut Umfrage den Roman zum Zeitpunkt des Kinostarts bereits gelesen."
V: „Eine aus den Befragten zufällig ausgewählte Person sah laut Umfrage die Verfilmung."

Bestimmen Sie die Wahrscheinlichkeit dafür, dass eine aus den Befragten zufällig ausgewählte Person, die laut Umfrage den Roman zum Zeitpunkt des Kinostarts noch nicht gelesen hatte, angab, die Verfilmung gesehen zu haben.

Löse die obige Abituraufgabe mit einem Häufigkeitsdoppelbaum.

stellungen abgelesen werden. In unserem Fall hatten wir die Frage gestellt: „Du triffst in der Schule einen Snapchat-User. Wie wahrscheinlich ist es, dass es sich dabei um ein Mädchen handelt?". Aus dem Baumdiagramm kann einfach abgelesen werden, dass 60 der 90 Snapchat-Nutzer Mädchen sind. Die Wahrscheinlichkeit $P_S(M)$ beträgt demnach $\frac{60}{90} = \frac{2}{3}$.

Falls es gewünscht ist, können auch die entsprechenden Wahrscheinlichkeiten an den Pfaden ergänzt werden (**Abb. 10**). Zur Beantwortung der Fragen ist dies aber nicht notwendig, da sich alle Wahrscheinlichkeiten nun durch einfache Verhältnisbildungen erhalten lassen.

Auch Schnittwahrscheinlichkeiten können so direkt abgelesen werden, indem die zweite beziehungsweise vierte Ebene übersprungen wird. Möchte man beispielsweise wissen, wie groß die Wahrscheinlichkeit dafür ist, dass es sich bei einem Jugendlichen um ein Mädchen handelt und dieser Jugendliche Snapchat nutzt, so entspricht dies 60 von 500 Jugendlichen, also einer Wahrscheinlichkeit von 12 %.

Hinweise für den Unterricht

Doppelbäume, die absolute und relative Häufigkeiten zugleich tragen können, haben viele hilfreiche Eigenschaften. Durch die Baumstruktur ist der sequenzielle Charakter der Aufgabenstellung deutlich erkennbar. Im Gegensatz zum einfachen Baumdiagramm erlaubt der Doppelbaum außerdem beide Leserichtungen (d. h. Teilmengenbildungen) gleichermaßen.

Die Komplettierung des Doppelbaumes gelingt bei einer großen Bandbreite von Aufgabenstellungen, beginnend von der Unterstufe bis hin zum Abitur – von absoluten Häufigkeiten über Anteilswerte bis hin zum Begriff der bedingten Wahrscheinlichkeit, und zwar unabhängig von der gegebenen Fragestellung.

Unserer Erfahrung nach kann der Häufigkeitsdoppelbaum bereits in der 5. Klasse (dann *nur* mit absoluten Häufigkeiten) eingeführt werden. Zu Beginn ist es möglicherweise etwas gewöhnungsbedürftig, dass sich in der unteren Hälfte des Doppelbaumes zwei Pfade überschneiden. Nach dem Lösen weniger Aufgaben können die Schülerinnen und Schüler aber gut mit dieser Visualisierung arbeiten (im Häufigkeitsdoppelbaum werden im Übrigen exakt dieselben Zahlen dargestellt, wie in einer Vierfeldertafel mit absoluten Häufigkeiten).

Wird der Doppelbaum sicher beherrscht, sollte man den flexiblen Einsatz der verschiedenen Visualisierungen (einfaches Baumdiagramm, Doppelbaum und Vierfeldertafeln) üben und zum Beispiel auch bewusst Vierfeldertafeln aus Doppelbäumen erstellen lassen und umgekehrt, um den Wechsel zwischen den verschiedenen Darstellungsarten zu üben. Spätestens bei der Einführung von relativen Häufigkeiten wird der Vorteil von (Doppel-)Baumdiagrammen bei Aufgaben mit bedingten Wahrscheinlichkeiten ersichtlich: Während für das Ausfüllen einer Vierfeldertafel Nebenrechnungen notwendig werden, können die gegebenen Informationen (egal, ob in Häufigkeiten oder Wahrscheinlichkeiten) einfach in den Doppelbaum eingetragen werden.

Literatur

Abitur 2012, Gymnasium Bayern, Mathematik, Bayerisches Staatsministerium für Unterricht und Kultus

Binder, K. / Krauss, S. / Wassner, C. (2018): Der Häufigkeitsdoppelbaum als didaktisch hilfreiches Werkzeug von der Unterstufe bis zum Abitur. – In: Stochastik in der Schule 38(1), S. 1–11.

Binder, K. / Vogel, M. (2018): Prä-Bayes'sche Verhältnisse. – In: mathematik lehren Heft 209, Friedrich Verlag S. 13–17.

Böcherer-Linder, K. / Eichler, A. / Vogel, M. (2018): Die Formel von Bayes. Kognitionspsychologische Grundlagen und empirische Untersuchung zur Bestimmung von Teilmenge-Grundmenge-Beziehungen. – In: Journal für Mathematikdidaktik, 39 (1). S. 127–146.

Borovcnik, M. (2016): „To Screen or not to screen." Dialoge zur medizinischen Diagnose. – In: mathematik lehren Heft 194, Friedrich Verlag S. 22–28.

Gigerenzer, G. / Hoffrage, U. (1995): How to improve Bayesian reasoning without instruction: frequency formats. – In: Psychological Review, 102(4), 684.

Wassner, C. / Martignon, L. / Biehler, R. (2004): Bayesianisches Denken in der Schule. – In: Unterrichtswissenschaft, 32 (1), S. 58–96.