Der Häufigkeitsdoppelbaum als didaktisch hilfreiches Werkzeug von der Unterstufe bis zum Abitur

KARIN BINDER, STEFAN KRAUSS, REGENSBURG UND CHRISTOPH WASSNER, NÜRNBERG

Zusammenfassung: *Von der Unterstufe bis zum Abitur werden Schülerinnen und Schüler mit verschiedensten Aufgaben zu Personen (oder allgemein zu Objekten) mit zwei Merkmalen, die je zwei Ausprägungen haben, konfrontiert. Der vorliegende Beitrag zeigt, wie „Häufigkeitsdoppelbäume" als hilfreiches Werkzeug zum Verständnis von Anteilswerten (in der Unterstufe) und bedingten Wahrscheinlichkeiten (in der Oberstufe) eingesetzt werden können und wie eine schrittweise Einführung des Häufigkeitsdoppelbaumes gelingen kann. Die Ausführungen basieren auf einer Lehrerfortbildung, die im Februar 2017 an der Universität Regensburg stattgefunden hat.*

1 Einleitung

Im schulischen Stochastikunterricht sind Aufgaben weit verbreitet, in denen für eine Gruppe von Objekten (meist: Personen) zwei Merkmale spezifiziert werden, die jeweils zwei Ausprägungen haben (z. B. Jungen vs. Mädchen und Brillenträger vs. nicht Brillenträger). Abbildung 1 illustriert solche Aufgaben, die von der Unterstufe (mit Anteilen) bis zum Abitur (mit bedingten Wahrscheinlichkeiten) behandelt werden. Der vorliegende Beitrag verfolgt vier Ziele:

1. Die Rekapitulation bekannter und erfolgreicher Lösungsstrategien für Aufgaben dieser Art: a) Häufigkeitskonzept und b) Visualisierung.
2. Die Vorstellung des „Häufigkeitsbaumes" sowie dessen Ergänzung zum „Häufigkeitsdoppelbaum" (vgl. Wassner, Martignon & Biehler, 2004).
3. Eine intuitive Schritt-für-Schritt-Anleitung zur Erstellung und Komplettierung eines solchen Häufigkeitsdoppelbaumes als explizite Lösungsstrategie für solche Aufgaben (vgl. Abb. 1).
4. Die Illustration, dass zahlreiche Aufgaben aus der Unterstufe, der Mittelstufe und sogar aus dem Abitur mithilfe dieser Schritte auf einfache Weise gelöst werden können.

Die Innovation des vorliegenden Artikels liegt vor allem in der Aufsplittung in einzelne, für Schüler intuitiv verständliche Schritte sowie in der Diskussion der universellen Anwendbarkeit des vorgeschlagenen didaktischen Verfahrens für eine große Klasse an Aufgaben, die sich über die gesamte Sekundarstufe erstrecken.

Die Betrachtung von Situationen mit zwei Merkmalen, die jeweils zwei Ausprägungen haben können, erfolgt schulisch spätestens ab der Unterstufe und zieht sich über die Mittelstufe bis zur Einführung bedingter Wahrscheinlichkeiten und der Behandlung entsprechender Aufgaben im Abitur durch. So besteht beispielsweise eine Schulklasse aus Mädchen und Jungen (Merkmal 1), die entweder Mitglied in einem Sportverein sind oder nicht (Merkmal 2). Sogar in der Grundschule können bereits vergleichbare Situationen behandelt werden, beispielsweise enaktiv mithilfe von Steckwürfeln (Martignon & Kurz-Milcke, 2006). In der Unterstufe werden dann in der Regel weitere derartige Aufgaben gestellt, um Fragen nach Anteilswerten zu beantworten wie z. B. „Wie viel Prozent der Mädchen sind in einem Sportverein?". All diese Aufgaben bereiten bereits auf den Begriff der bedingten Wahrscheinlichkeit vor, der in der Regel in der gymnasialen Mittel- und Oberstufe unterrichtet wird. Eine Spielart bedingter Wahrscheinlichkeiten stellen sogenannte „Bayesianische Aufgaben" dar. Bayesianische Aufgabenstellungen können für verschiedene Berufsfelder relevant sein, wie beispielsweise Medizin oder Jura, und führen dort regelmäßig zu Trugschlüssen mit teils fatalen Konsequenzen (z. B. Stine, 1996). Aufgrund der Relevanz Bayesianischen Denkens in menschlichen Entscheidungsfindungsprozessen wurde diese Thematik in der Kognitionspsychologie bereits ausführlich untersucht. Dabei wurden auch Strategien vorgeschlagen, die richtiges Urteilen unterstützen (z. B. McDowell & Jacobs, im Druck).

Klassischerweise werden derartige Trugschlüsse anhand der berühmten „Mammographie-Aufgabe" illustriert:

Mammographie-Aufgabe (Wahrscheinlichkeitsversion)

Stellen Sie sich eine Frau vor, die am Routine-Screening für die Früherkennung von Brustkrebs teilnimmt. In ihrer Altersklasse liegt die Wahrscheinlichkeit einer Brustkrebserkrankung bei 1 %. Die Wahrscheinlichkeit, dass eine an Brustkrebs erkrankte Frau ein positives Mammogramm erhält, liegt bei 80 %. Die Wahrscheinlichkeit, dass eine völlig gesunde Frau ein positives Mammogramm erhält, liegt bei 9,6 %. Wie wahrscheinlich ist es nun, dass die

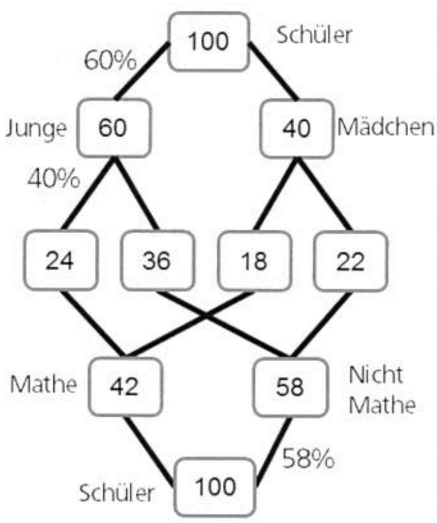
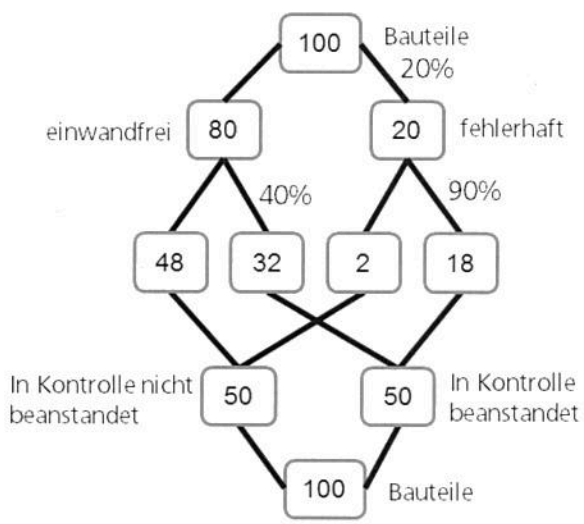
	Unterstufe	Abitur
Aufgabe	Bei einer Umfrage am Ortenburg-Gymnasium waren 60 % der Befragten Jungen. Von diesen gaben 40 % Mathematik als ihr Lieblingsfach an. Von allen Befragten nannten 58 % ein anderes Lieblingsfach als Mathematik.	Bei der Produktion von Halbleiterbauteilen eines bestimmten Typs ist im Mittel jedes fünfte Bauteil fehlerhaft. Jedes produzierte Bauteil wird abschließend einer Kontrolle unterzogen und dabei entweder als fehlerhaft oder als einwandfrei eingestuft. Im Rahmen der Kontrolle wird ein fehlerhaftes Bauteil mit einer Wahrscheinlichkeit von 90 % als fehlerhaft eingestuft. Die Wahrscheinlichkeit dafür, dass ein einwandfreies Bauteil als fehlerhaft eingestuft wird, beträgt 40 %.
Frage	Wie viele der Befragten waren Mädchen und gaben Mathematik als ihr Lieblingsfach an?	Ermitteln Sie die Wahrscheinlichkeit dafür, dass ein nach der Kontrolle zufällig ausgewähltes Bauteil fehlerhaft ist, wenn es im Rahmen der Kontrolle als einwandfrei eingestuft wurde.
Doppelbaum		
Antwort	18 von 100 Schülern sind Mädchen und haben Mathematik als Lieblingsfach. Das entspricht einem Anteil von 18 %.	2 von 50 Bauteilen, die im Rahmen der Kontrolle als fehlerfrei eingestuft wurden, sind in Wirklichkeit fehlerhaft. Das entspricht einem Anteil von 4 %.

Abb. 1: Typische Aufgaben zu zwei Merkmalen mit je zwei möglichen Ausprägungen aus der Unterstufe (aus Lambacher Schweizer 6 Mathematik für Gymnasien – Bayern) und dem Abitur (ISB Bayern, o. D.).

Frau tatsächlich an Brustkrebs erkrankt ist, wenn sie ein positives Mammogramm erhalten hat?

Selbst die meisten Ärzte sind nicht in der Lage, die statistischen Informationen adäquat miteinander zu verknüpfen (die richtige Antwort lautet 7,8 %), und scheitern bei der Beantwortung dieser Frage (Eddy, 1982).

Da der Umgang mit bedingten Wahrscheinlichkeiten für verschiedene Berufsgruppen eine wesentliche Grundlage darstellt, ist eine verständliche und frühe Beschäftigung mit solchen Situationen wichtig (für ein Beispiel aus dem juristischen Kontext siehe z. B. Krauss & Bruckmaier, 2014).

2 Absolute Häufigkeiten und Häufigkeitsbäume

Im vorliegenden Beitrag werden vor allem zwei Strategien fokussiert, die hilfreich im Umgang mit bedingten Wahrscheinlichkeiten sind: Absolute Häufigkeiten und die Visualisierung dieser absoluten Häufigkeiten in den Knoten eines „Häufigkeitsbaumes". Weitere denkbare Strategien sind z. B. die Visualisierung durch Einheitsquadrate (Böcherer-Linder & Eichler, 2017), ikonische Darstellungen (Brase, 2014), Mengendiagramme (Micallef et al., 2012) oder der Einsatz der Software Tinkerplots (Frischemeier, 2017). Nun soll erneut die Situation im Mammographie-Problem betrachtet werden. Die

Kognitionspsychologen Gigerenzer und Hoffrage (1995) konnten zeigen, dass Menschen statistische Informationen in Bayesianischen Aufgaben besser verstehen, wenn diese nicht als Wahrscheinlichkeiten oder in Prozent, sondern mithilfe von absoluten Häufigkeiten verdeutlicht werden (für die erste Adaption für die Didaktik siehe Krauss, 2003). Die Mammographie-Aufgabe lautet in diesem Fall:

Mammographie-Aufgabe (Häufigkeitsversion)

Von 10.000 Frauen sind 100 an Brustkrebs erkrankt. Von diesen 100 erkrankten Frauen erhalten 80 ein positives Mammogramm. Von den restlichen 9.900 völlig gesunden Frauen erhalten 950 ein positives Mammogramm. Wie viele der Frauen mit positivem Mammogramm sind tatsächlich an Brustkrebs erkrankt?

Deutlich mehr Menschen sind nun in der Lage, die Frage korrekt zu beantworten: Insgesamt erhalten 80 + 950 = 1.030 Frauen ein positives Mammogramm. Aber nur 80 von 1.030 Frauen mit positivem Mammogramm sind tatsächlich an Brustkrebs erkrankt. Ausdrücke der Art „3 von 5“ werden von Psychologen auch „natürliche Häufigkeiten“ genannt. Die Übersetzung statistischer Informationen von Wahrscheinlichkeiten in Häufigkeiten ist mittlerweile in der Didaktik der Stochastik weit verbreitet (Eichler & Vogel, 2009; Engel & Martignon, 2015; Wassner et al., 2004). Für weitere numerische Darstellungsarten statistischer Informationen siehe Bruckmaier et al. (2016).

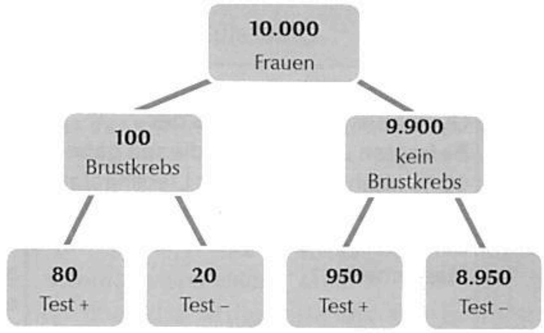
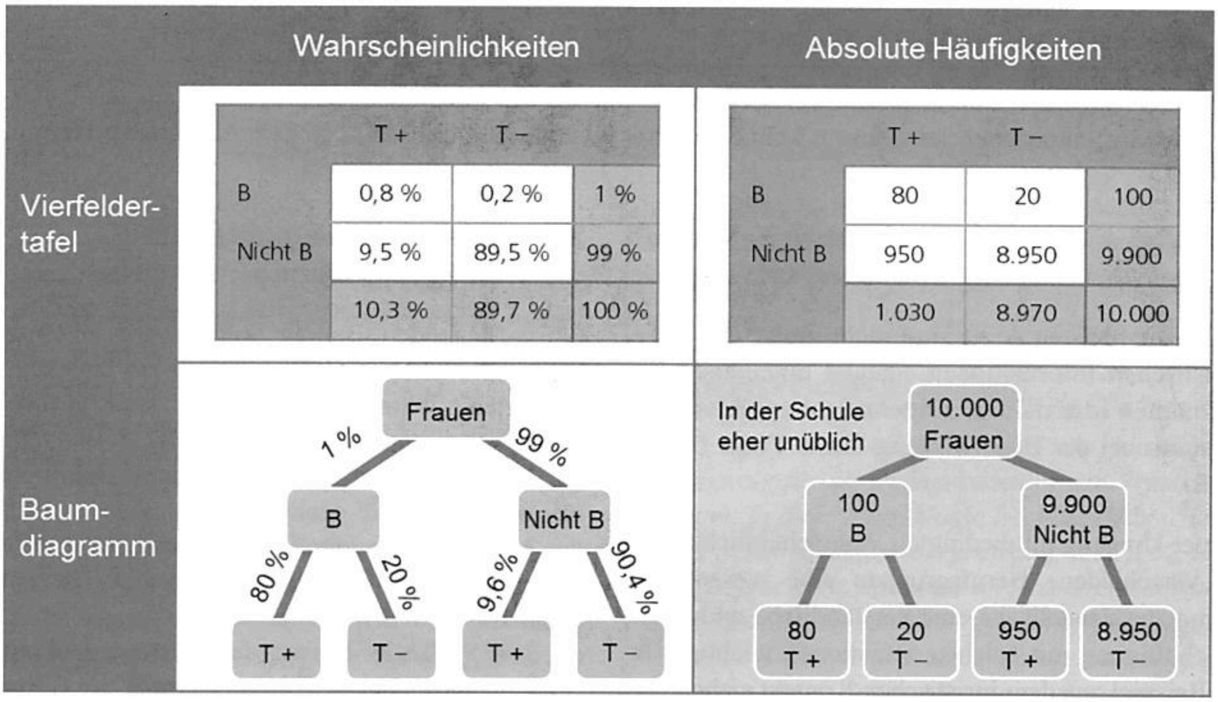


Abb. 2: Häufigkeitsbaum zur Mammographie-Aufgabe

Eine Visualisierung der Situation mithilfe eines Baumdiagramms mit absoluten Häufigkeiten (Abb. 2) kann die Lösungsfindung zusätzlich unterstützen. Allerdings ist nicht jede Visualisierung gleichermaßen hilfreich. Im schulischen Stochastikunterricht werden häufig Vierfeldertafeln und Baumdiagramme mit Wahrscheinlichkeiten an den Ästen eingesetzt, um Sachverhalte mit zwei dichotomen Merkmalen zu beleuchten. Binder et al. (2015) stellten in einer Untersuchung mit Schülerinnen und Schülern der 11. Klasse jedoch fest, dass diese Wahrscheinlichkeitsvisualisierungen die Lösungsfindung bei Bayesianischen Aufgaben kaum unterstützen, während Vierfeldertafeln und Baumdiagramme mit absoluten Häufigkeiten für die Schülerinnen und Schüler äußert hilfreich sind.

Grundsätzlich können sowohl Vierfeldertafeln als auch Baumdiagramme mit Wahrscheinlichkeiten oder mit absoluten Häufigkeiten ausgefüllt werden (siehe Abbildung 3). Erstaunlicherweise finden aus-



	T +	T -	
B	0,8 %	0,2 %	1 %
Nicht B	9,5 %	89,5 %	99 %
	10,3 %	89,7 %	100 %

	T +	T -	
B	80	20	100
Nicht B	950	8.950	9.900
	1.030	8.970	10.000

Abb. 3: Vier verschiedene Visualisierungen für das Mammographie-Problem.

gerechnet die Baumdiagramme mit absoluten Häufigkeiten in den Knoten (Abb. 3 rechts unten) schulisch bislang kaum Beachtung. Der Vorteil von Baumdiagrammen gegenüber Vierfeldertafeln ist, dass sogar gleichzeitig absolute und relative Häufigkeiten in einem Diagramm dargestellt werden können, wie in den späteren Ausführungen noch ersichtlich wird.

Überdies fällt auf, dass in den Wahrscheinlichkeitsvisualisierungen die Zahlen in der Vierfeldertafel nicht mit den Zahlen im Baumdiagramm übereinstimmen. Während in der Vierfeldertafel die Wahrscheinlichkeiten von *Konjunktionen* dargestellt werden, enthält das Baumdiagramm *bedingte Wahrscheinlichkeiten.* Anders verhält es sich aber bei den Häufigkeitsvisualisierungen: Die vier absoluten Zahlen im Inneren der Vierfeldertafel entsprechen tatsächlich den Zahlen in der untersten Ebene des Häufigkeitsbaumes.

Häufigkeitsbäume lassen sich noch zu „Häufigkeitsdoppelbäumen" erweitern, die beide Leserichtungen gleichermaßen darstellen (Wassner et al., 2007; Abb. 4). Wir behaupten, dass mithilfe dieser Struktur nicht nur Bayesianische Aufgaben, sondern eine große Klasse von Aufgaben zu zwei Merkmalen mit je zwei Merkmalsausprägungen dargestellt und einfach gelöst werden können, wie z. B. Anteilsfragen (Unterstufe) oder Fragen zu bedingten Wahrscheinlichkeiten oder Schnittwahrscheinlichkeiten (Mittel- und Oberstufe).

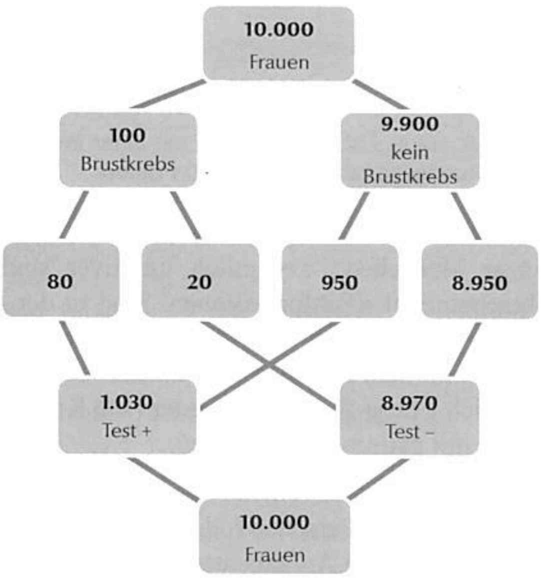


Abb. 4: Häufigkeitsdoppelbaum zum Mammographie-Problem

3 In sechs Schritten einen Häufigkeitsdoppelbaum erstellen

Nachfolgend sollen sechs Schritte vorgestellt werden, mit denen für Situationen zu zwei Merkmalen schnell und einfach ein Häufigkeitsdoppelbaum erstellt werden kann. Sobald der Häufigkeitsdoppelbaum gezeichnet wurde, lässt sich *jede beliebige* Fragestellung einfach aus dem Baumdiagramm ablesen. Haben Schülerinnen und Schüler zwei oder drei Aufgaben mithilfe eines Häufigkeitsdoppelbaums gelöst, können sie diese Struktur problemlos auf weitere Aufgaben übertragen.

Die sechs Schritte werden im Folgenden anhand der Wahrscheinlichkeitsversion des Mammographie-Problems illustriert. Ihr entscheidender Vorteil besteht darin, dass sie sich nicht nur für Bayesianische Aufgabenstellungen eignen, sondern auch zahlreiche weitere Aufgaben aus der Unter-, Mittel- und Oberstufe *in genau dieser Weise* gelöst werden können. Das bedeutet, dass der Häufigkeitsdoppelbaum also über die gesamte Schulzeit *nicht mehr modifiziert* werden muss.

Schritt 1 – Zeichnen der Struktur

Der erste Schritt besteht immer aus der Zeichnung der leeren Struktur eines Häufigkeitsdoppelbaumes (Abb. 5).

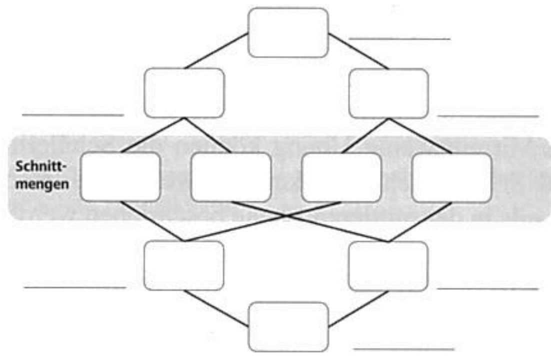


Abb.5: Leere Struktur eines Häufigkeitsdoppelbaumes

Schritt 2 – Beschriftung des Baumdiagramms

Im zweiten Schritt erfolgt eine Beschriftung des Baumdiagramms (Abb. 6). Hierzu ist noch keine Beachtung der Fragestellung erforderlich.

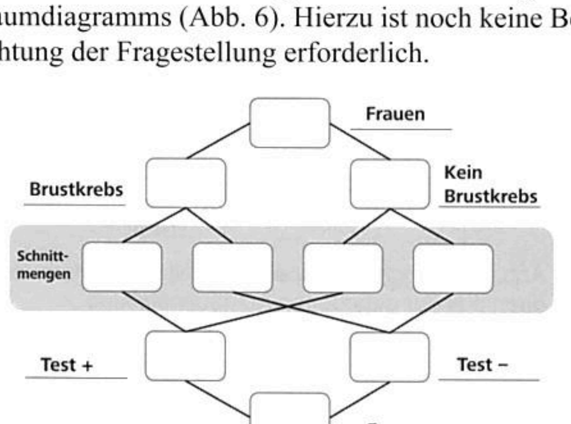


Abb. 6: Beschriftete Struktur eines Häufigkeitsdoppelbaumes

Neben dem obersten und untersten Knoten des Baumdiagramms wird der Begriff „Frauen" (allgemein: die Personen oder Objekte, die die beiden Merkmale aufweisen) notiert. In der zweiten und vierten Ebene finden nun die beiden betrachteten Merkmale Platz: In diesem Fall der Erkrankungsstand (Brustkrebs vs. kein Brustkrebs) und das Testergebnis der Mammographie (Test + vs. Test –).

Es ist hierbei nicht von Belang, in welche der beiden Ebenen 2 und 4 welches der beiden Merkmale eingetragen wird, da nach der Ausführung sämtlicher Schritte ein Häufigkeitsdoppelbaum entstanden sein wird, der beide Leserichtungen gleichermaßen ermöglicht. Ein grundlegendes Problem bei der eigenständigen Konstruktion von Baumdiagrammen ist häufig, dass Schülerinnen und Schüler nicht wissen, welche Ebene sie im Baumdiagramm zuerst zeichnen müssen. Dieses Problem gibt es bei der Zeichnung eines Häufigkeitsdoppelbaumes nicht mehr.

In der mittleren Ebene werden die vier Schnittmengen dargestellt. Der Knoten ganz links in der mittleren Ebene steht demnach für alle Frauen, die an Brustkrebs erkrankt sind *und* ein positives Testergebnis erhalten. Aus Platzgründen wird jedoch auf eine ausführliche Beschriftung der mittleren Ebene verzichtet. Mit ein wenig Übung können die Schülerinnen und Schüler schnell erkennen, welche Teilmengen jeweils in der mittleren Ebene beschrieben werden.

Schritt 3 – Eintragen aller Informationen aus der Aufgabe

Im nächsten Schritt werden alle statistischen Informationen, die in der Aufgabe gegeben sind, in das Baumdiagramm eingetragen (Abb. 7).

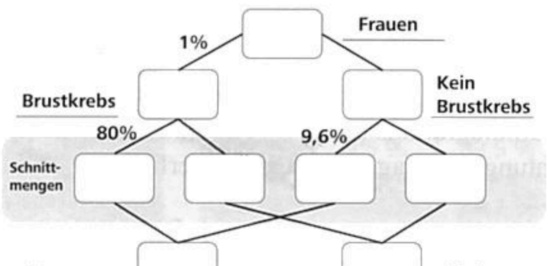


Abb. 7: Häufigkeitsdoppelbaum mit allen Informationen aus der gegebenen Aufgabenstellung

Die erste statistische Information aus der Aufgabenstellung lautet, dass 1 % der Frauen an Brustkrebs erkrankt ist. Diese Wahrscheinlichkeit wird nun an den passenden Ast geschrieben. Die zweite statistische Information wird in der Unterstufe z. B. als Anteil, in der Oberstufe oft als bedingte Wahrscheinlichkeit angegeben: Die Wahrscheinlichkeit, dass der Test positiv wird, wenn die Frau erkrankt ist, liegt bei 80 %. Auch für diese Information wird nun der passende Ast ausgesucht und die Wahrscheinlichkeitsinformation an diesen Ast geschrieben. Schließlich folgt die Information darüber, dass die Wahrscheinlichkeit eines positiven Tests bei einer gesunden Person bei 9,6 % liegt. Auch diese Information wird im Baumdiagramm an die passende Stelle geschrieben, bevor zum nächsten Schritt übergegangen wird.

Schritt 4 – Wahl einer (imaginären) Grundgesamtheit

Da empirische Studien vielfach zeigen konnten, dass statistische Informationen besser verstanden werden, wenn man sich dahinter konkrete Personen vorstellen kann, die jeweils als Merkmalsträger betrachtet werden können, wird nun eine „imaginäre Grundgesamtheit" gewählt, die oben und unten im Doppelbaum eingetragen wird (Abb. 8).

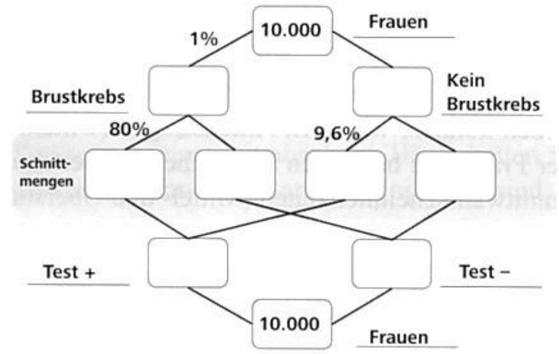


Abb. 8: Häufigkeitsdoppelbaum mit einer imaginären Grundgesamtheit von 10.000 Frauen

Dieser Schritt ist der eigentliche Kniff, da Anzahlen „ganzer Menschen" wesentlich intuitiver sind als Wahrscheinlichkeitsinformationen. Sind in der Aufgabe bereits absolute Häufigkeiten gegeben, kann man hier mit dem Lösungsalgorithmus beginnen. So lassen sich Ebene für Ebene bequem die Knoten des Baumes mit ganzen Zahlen füllen.

Oft ist es zielführend, von einer Grundgesamtheit von 1.000 oder 10.000 Personen (oder Objekten) auszugehen. Kommen in der Aufgabenstellung sehr große oder sehr kleine Prozentzahlen vor (die Wahrscheinlichkeit, dass eine Person HIV-infiziert ist, liegt z. B. bei 0,01 %), so müssen oftmals sehr große Grundgesamtheiten gewählt werden. Nach einer kurzen Einübungsphase entwickeln Schülerinnen und Schüler ein Gefühl dafür, wie eine geeignete Größe für die Grundgesamtheit gewählt werden kann. Wird die Grundgesamtheit im ersten Moment zu klein gewählt, kann diese einfach durch das Anhängen von Nullen vergrößert werden.

Sind die Wahrscheinlichkeiten als gewöhnliche (maximal gekürzte) Brüche gegeben, also beispielsweise P (Brustkrebs) = 1/3 und $P_{\text{Brustkrebs}}$ (Test positiv) = 4/5, so muss die Gesamtzahl an Personen durch das Produkt 3 · 5 teilbar sein (diese Bedingung muss für *beide* Seiten des Baumes erfüllt sein). Dabei ist zu beachten, dass eine Hauptnennerbildung nicht immer ausreicht, denn z. B. mit P (kein Brustkrebs) = 2/3 und $P_{\text{kein Brustkrebs}}$ (Test positiv) = 1/6 muss die Grundgesamtheit tatsächlich durch 18 (und nicht nur durch 6) geteilt werden. Somit kann es gelegentlich auch vorkommen, dass 300 (oder 7.000 etc.) Personen als Grundgesamtheit gewählt werden müssen, um in der mittleren Ebene ganze Zahlen zu erhalten.

Schritt 5 – Bestimmung der absoluten Häufigkeiten für die Teilmengen

Nun liegt eine konkrete Grundgesamtheit von Frauen vor, auf die alle statistischen Informationen aus der Aufgabenstellung bezogen werden können.

Wenn von 10.000 Frauen 1 % an Brustkrebs erkrankt sind, bedeutet das, dass in dieser Grundgesamtheit 100 Frauen Brustkrebs haben (Abb. 9).

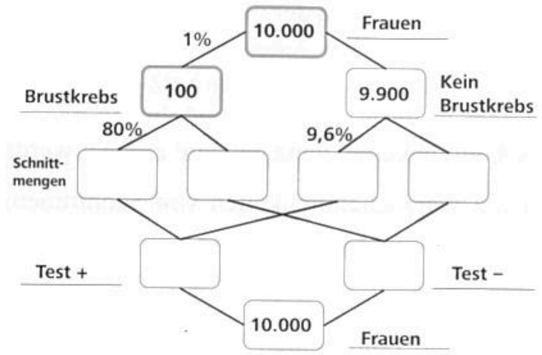


Abb. 9: Übersetzen der Wahrscheinlichkeiten in absolute Häufigkeiten

Daraus folgt durch eine einfache Subtraktion unmittelbar, dass 9.900 Frauen nicht an Brustkrebs erkrankt sind. Diese beiden Zahlen werden nun in die entsprechenden Knoten eingetragen.

Weiter gilt: Wenn nun 80 % der an Brustkrebs erkrankten Frauen ein positives Mammogramm erhalten, dann bedeutet dies, dass von den 100 Frauen mit Brustkrebs 80 ein positives Testergebnis erhalten (Abb. 10).

Auch hier kann aufgrund der einfachen additiven Strukturen bereits im Knoten daneben ergänzt werden, dass es dementsprechend 20 Frauen gibt, die trotz Brustkrebs ein negatives Testergebnis erhalten.

Wenn die Wahrscheinlichkeit eines positiven Testergebnisses für eine gesunde Frau bei 9,6 % liegt, dann heißt das weiterhin, dass in diesem Fall 950 von 9.900 gesunden Frauen ein positives Testergebnis erhalten (Abb. 11). Dementsprechend muss es 8.950 gesunde Frauen mit negativem Testergebnis geben.

Auch die nun noch fehlenden Werte lassen sich bequem im Baumdiagramm ergänzen. Die Gesamtanzahl der Frauen mit positivem Test ergibt sich aus den 80 kranken und den 950 gesunden Frauen, die ein positives Testergebnis erhalten haben: Dies sind insgesamt 1.030 Frauen mit positivem Test. Auf die gleiche Weise kann man sehen, dass es insgesamt 8.970 Frauen mit einem negativen Testergebnis gibt (Abb. 12).

Diese „Komplettierung" des Häufigkeitsdoppelbaumes gelingt bei einer großen Klasse von verschiedenen Aufgabenstellungen, auch wenn es sich gar nicht um Bayesianische Aufgaben handelt. Gelegentlich muss hierbei anstelle einer bestimmten Zahl auch

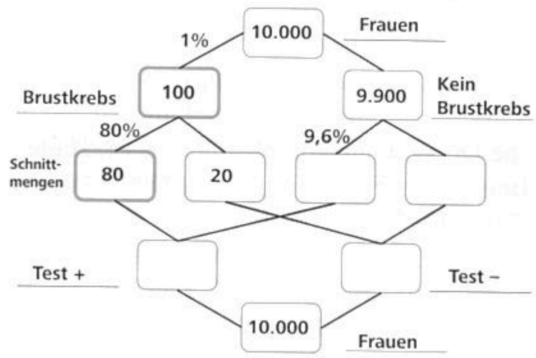


Abb. 10: Übersetzen der Wahrscheinlichkeiten in absolute Häufigkeiten

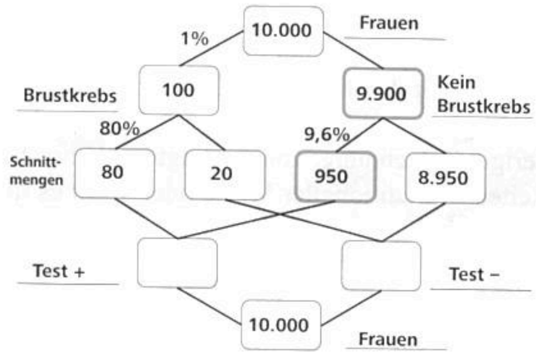


Abb. 11: Ergänzen der noch fehlenden Knoten über einfache additive Strukturen

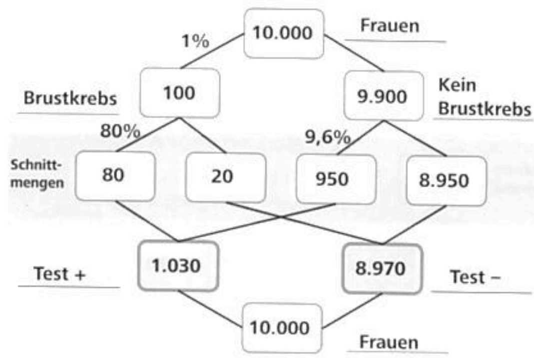


Abb. 12: Vollständig ausgefüllter Häufigkeitsdoppelbaum

eine Variable verwendet werden, aber selbst dann ist die Lösung auf Basis dieser Komplettierung oftmals mit Wissen aus der Unterstufe möglich (nämlich mit Bruchrechnung, der Grundgleichung der Prozentrechnung und dem Lösen einfacher Gleichungen).

Es sei angemerkt, dass die Aufteilung der Grundgesamtheit entsprechend der Wahrscheinlichkeiten dabei *idealisiert* (d. h. als Erwartungswerte) erfolgt. Unterrichtlich sollte hier an geeigneter Stelle die Variabilität der absoluten Zahlen in den Knoten beim Aufteilen einer anderen imaginären Grundgesamtheit thematisiert werden.

Bei Bedarf könnten nun durch einfache Verhältnisbildung zweier Knoten auch sämtliche an den Ästen noch fehlende bedingte Wahrscheinlichkeiten und Randwahrscheinlichkeiten ergänzt werden, die bislang noch nicht abgebildet sind.

Das Entscheidende dabei ist, dass sich ein Baumdiagramm mit Wahrscheinlichkeiten nicht durch so einfache Operationen komplettieren lässt: Hätte man zum Beispiel im Baumdiagramm (Abb. 12) – wie in der Schule üblich – nur die drei relativen Häufigkeiten an den Ästen gegeben, könnte man beispielsweise die Wahrscheinlichkeit P_{Test+}(Brustkrebs) (in absoluten Häufigkeiten einfach 80/1.030) nur mit hohem technischem Aufwand wie der Formel von Bayes oder den Pfadregeln erhalten. Dies gilt auch für alle anderen bedingten und unbedingten Wahrscheinlichkeiten im unteren Teil eines „Wahrscheinlichkeitsdoppelbaumes“. Ein Doppelbaum nur mit Wahrscheinlichkeiten lässt sich nicht durch einfache Additionen ergänzen, eine Komplettierung ist hier nur möglich durch eine vorherige Berechnung von bedingten Wahrscheinlichkeiten in traditioneller Weise oder – um es überspitzt zu sagen – wenn die Aufgabe bereits gelöst ist.

Schritt 6 – Antworten ablesen

Nachdem nun ein vollständiger Häufigkeitsdoppelbaum entstanden ist, können die Antworten auf alle möglichen Fragen bequem abgelesen werden (Abb. 13).

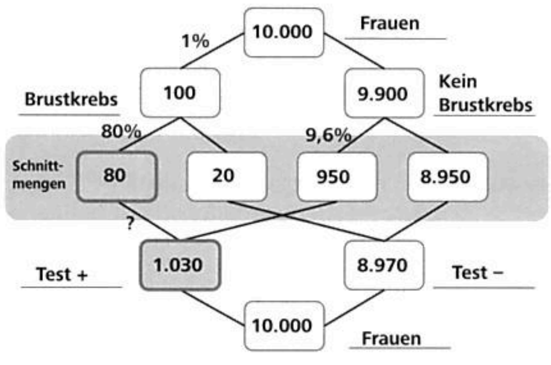


Abb. 13: Ablesen der korrekten Lösung aus dem Häufigkeitsdoppelbaum

Im klassischen Mammographie-Problem wird z. B. die Frage nach der bedingten Wahrscheinlichkeit gestellt, dass eine Frau tatsächlich erkrankt ist, wenn sie ein positives Testergebnis erhalten hat. Diese Frage ist aufgrund der erforderlichen „Inversion“ von $P_B(M+)$ auf $P_{M+}(B)$ für Schüler besonders herausfordernd. Um diese Frage zu beantworten, kann man die Schüler zunächst den passenden Ast und/oder die zugehörigen Knoten des Baumdiagramms markieren lassen. Die zugehörige Lösung kann nun einfach abgelesen werden: 80 von 1.030 Frauen mit positivem Test sind tatsächlich an Brustkrebs erkrankt.

Aber auch alle weiteren Fragen, die in der betrachteten Situation interessant sein könnten, lassen sich direkt ablesen. So sieht man beispielsweise, dass von 10.000 Frauen, die am Mammographie-Screening teilnehmen, 8.970 ein negatives Mammogramm erhalten. Eine interessante Frage ist hierbei z. B., wie wahrscheinlich es nun ist, dass die Frau tatsächlich gesund ist, wenn sie ein negatives Testergebnis erhält: 8.950 der 8.970 Frauen mit negativem Test sind tatsächlich gesund, das entspricht etwa 99,8 % der Frauen.

Nach Komplettierung des Häufigkeitsdoppelbaumes können an den acht von der mittleren Ebene (Schnittmengen) ausgehenden Ästen prinzipiell – und das ganz ohne Pfadregeln oder den Satz von Bayes (!) – alle acht in dieser Situation möglichen bedingten Wahrscheinlichkeiten ganz bequem ergänzt werden.

Selbst die Wahrscheinlichkeiten von Schnittmengen können nun schnell und einfach (und ebenfalls ohne Anwendung der Pfadregeln) aus dem Baumdiagramm entnommen werden. So ist z. B. die Wahrscheinlichkeit, dass eine Frau keinen Brustkrebs hat *und* ein positives Testergebnis erhält: 950 von 10.000 (oder 9,5 %). Wahrscheinlichkeiten von Schnittmengen können also durch „Überspringen“ einer Ebene ebenfalls als einfache Verhältnisse „abgelesen“ werden.

4 Beispiele aus der Unterstufe und der Mittelstufe

Die eben aufgeführten Schritte zeigen, dass zu zahlreichen Aufgaben zu bedingten Wahrscheinlichkeiten oder Schnittwahrscheinlichkeiten, die in der gymnasialen Oberstufe und im Abitur oft gestellt werden, ein Häufigkeitsdoppelbaum gezeichnet werden kann, mit dem alle anfallenden Fragen beantwortet werden können.

Abbildung 1 (linke Spalte) illustriert, dass der Häufigkeitsdoppelbaum aber bereits in der Unterstufe zum Einsatz kommen kann. Dabei ist bemerkenswert, dass zahlreiche Aufgaben aus der Unter-, Mittel- und Oberstufe nur sprachliche Unterschiede aufweisen. In der Unterstufe, wo den Schülerinnen und Schülern der Begriff der bedingten Wahrscheinlichkeit noch nicht bekannt ist, wird – oft im Zusammenhang mit dem Prozentrechnen – die „Anteilssprache" für ansonsten zu manchen Oberstufenaufgaben völlig strukturgleiche Aufgaben verwendet oder es werden direkt absolute Häufigkeiten kommuniziert.

Den Unterschied bei Aufgaben zu zwei Objekten (Menschen) mit jeweils zwei Merkmalen zwischen der Unter-, Mittel- und Oberstufe macht also oft nicht die Struktur der Information oder der benötigte Algorithmus, sondern die *Darstellung* der statistischen Informationen. Für alle statistischen Angaben in solchen Aufgaben, aber auch für die Fragestellung gibt es im Wesentlichen „drei isomorphe Gesichter":

1. Wie viele der A sind B?
2. Wie hoch ist der Anteil der B unter den A?
3. Wie hoch ist die Wahrscheinlichkeit für B unter der Bedingung A?

Alle beliebigen Fragen lassen sich jeweils auch entsprechend auf die beiden anderen Arten formulieren. Während Informationen und Fragestellungen der Art 3 erst nach der Einführung bedingter Wahrscheinlichkeiten zum Einsatz kommen können, sind 1 und 2 bereits in der Unterstufe möglich. Es kommen aber auch in der Mittel- und Oberstufe sowie in Abiturprüfungen Informationen der Art 1 und 2 vor.

Nachfolgende Aufgabe aus einem Schulbuch der 7. Klasse steht beispielhaft für Formulierungen in der „Anteilssprache" (Abb. 14, Fokus Mathematik 7 – Gymnasium Bayern):

Am Beethoven-Gymnasium beträgt der Anteil der Mädchen 45 %. 28 % der Mädchen und 36 % der Jungen kommen gewöhnlich mit dem Rad zur Schule. Welcher Prozentsatz aller Schüler nimmt für den Schulweg das Rad?

Folgendes Beispiel aus der 10. Klasse illustriert eine Aufgabe in der Sprache der Wahrscheinlichkeiten (Lambacher Schweizer 10, Mathematik für Gymnasien – Bayern):

Bei der Lehrkraft Müller kommt es in der Klasse 10 A in einer Unterrichtsstunde mit einer Wahrscheinlichkeit von 25 % zu einer Unterrichtsstörung. Im Falle einer Unterrichtsstörung erhöht sich der Blutdruck der Lehrkraft mit 80 % Wahrscheinlichkeit. Jedoch kommt es bei der Lehrkraft Müller auch ohne Unterrichtsstörung mit einer Wahrscheinlichkeit von 10 % zu erhöhtem Blutdruck. Mit welcher Wahrscheinlichkeit ist es in einer 10 A-Stunde laut und die Lehrkraft hat erhöhten Blutdruck?

5 Vorteile des Häufigkeitsdoppelbaumes

Man könnte nun einwenden, dass durch den Gebrauch von absoluten Häufigkeiten das Wahrscheinlichkeitskalkül vernachlässigt wird. Folgte man dieser Logik, müssten aber auch alle Aufgaben zu mit absoluten Häufigkeiten gefüllten Vierfeldertafeln aus dem Curriculum entfernt werden. Interessant ist demnach eher die Frage, welchen Vorteil der Häufigkeitsdoppelbaum gegenüber der *Vierfeldertafel mit absoluten Häufigkeiten* hat.

In Häufigkeitsdoppelbäumen ist die hierarchische Struktur der Daten besser abgebildet als in Vierfeldertafeln. Beide Richtungen der „sukzessiven Aufteilung" der Grundgesamtheit sind intuitiver sichtbar. Darüber hinaus lassen sich durch Lesen „von innen nach außen" alle (bedingten oder unbedingten) Wahrscheinlichkeiten ganz einfach „ablesen" und gegebenenfalls sogar an den Pfaden bequem ergänzen.

Zwar kann auch für eine Vierfeldertafel eine „imaginäre" Grundgesamtheit gewählt werden. Anders als im Baumdiagramm lassen sich in einer Vierfeldertafel allerdings nicht alle bedingten Wahrscheinlichkeiten bequem ergänzen. Wo beispielsweise sollte in Abbildung 15 die Wahrscheinlichkeit $P_{\text{Lehrkraft}}(\text{männlich}) = 20/50$ eingetragen werden? Beim Doppelbaum gibt

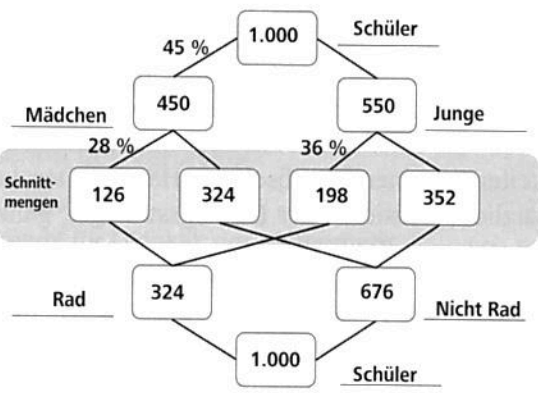


Abb. 14: Häufigkeitsdoppelbaum für eine typische Aufgabe aus der 7. Klasse

	Lehrkraft	Schüler	
männlich	20	570	590
weiblich	30	380	410
	50	950	1.000

Abb. 15: In einer Vierfeldertafel mit absoluten Häufigkeiten gibt es keinen „passenden Platz", um die acht bedingten Wahrscheinlichkeiten zu ergänzen.

es für jede dieser bedingten Wahrscheinlichkeiten einen entsprechenden Ast.

Im Häufigkeitsdoppelbaum lassen sich Lösungen zu „Bayesianischen" Aufgaben „auf der anderen Seite" des Baumes ablesen. Wahrscheinlichkeiten von Schnittmengen (z. B. eines Mädchens, das mit dem Rad zur Schule kommt) können durch „Überspringen einer Ebene" ablesen werden (in diesem Fall 126 von 1.000). Für die Bestimmung von beliebigen Wahrscheinlichkeiten sind hierbei weder das Konzept der bedingten Wahrscheinlichkeit noch die Pfadregeln erforderlich.

Man beachte darüber hinaus, dass in einer Vierfeldertafel mit relativen Häufigkeiten im Inneren ausschließlich Wahrscheinlichkeiten von Schnittmengen eingetragen sind, ganz im Gegensatz zum Baumdiagramm mit Wahrscheinlichkeiten, an dessen unteren Ästen gerade bedingte Wahrscheinlichkeiten stehen. Dies hat zur Folge, dass die Vierfeldertafel prinzipiell ein weniger mächtiges didaktisches Werkzeug zur Visualisierung bedingter Wahrscheinlichkeiten ist als ein Baumdiagramm. Die Strategie zur Lösung von Aufgaben ist nämlich jeweils identisch für beide Arten von Vierfeldertafeln: Die zwei zu einer Antwort gehörigen Zellen müssen in beiden Fällen auf die jeweils selbe Weise dividiert werden. Dies wird auch klar, wenn man die absoluten Häufigkeiten in obiger Vierfeldertafel (Abb. 15) durch relative Häufigkeiten ersetzt: Innen würden dann beispielsweise die Zahlen „0,02"; „0,57"; „0,03" bzw. „0,38" stehen (also abgesehen von einer Normierung auf 1 bzw. 100 % dieselben Zahlen). Somit eröffnet der Wechsel im Unterricht von einer Vierfeldertafel mit relativen Häufigkeiten zu einer mit absoluten Häufigkeiten keine zusätzliche Einsicht oder Lösungsstrategie, ganz im Gegensatz zum Baumdiagramm.

In einem Doppelbaum mit absoluten Häufigkeiten lassen sich darüber hinaus die entsprechenden relativen Häufigkeiten (und somit auch bedingten Wahrscheinlichkeiten) bequem ergänzen; alle Zahlen (relative *und* absolute Häufigkeiten) sowie ihre Beziehungen untereinander sind also auf einen Blick sichtbar (Abb. 16).

Gelegentlich werden etwas komplexere Abituraufgaben gestellt, in denen zusätzlich noch eine Variable in den Knoten des Häufigkeitsbaumes eingetragen werden muss. Aber selbst solche Aufgabenstellungen lassen sich sehr oft mit den hier vorgestellten Schritten lösen (siehe Fortbildung „Daten und Zufall: Einführung in den LehrplanPLUS und neue didaktische Ansätze", 2017).

6 Ausblick

Situationen zu Objekten mit zwei Merkmalen mit je zwei Ausprägungen lassen sich prinzipiell auf zahlreiche verschiedene Arten visualisieren. Im vorliegenden Beitrag wurden Vierfeldertafeln (mit Wahrscheinlichkeiten oder mit absoluten Häufigkeiten in den Zellen) und Baumdiagramme (mit Wahrscheinlichkeiten an den Ästen oder mit absoluten Häufigkeiten in den Knoten) fokussiert und ein Weg aufgezeigt, wie für viele Aufgabenstellungen ein Häufigkeitsdoppelbaum konstruiert werden kann.

Leider wird gerade das sehr verständnisfördernde Baumdiagramm mit absoluten Häufigkeiten im Unterricht oft „unterschlagen". Es gibt allerdings auch Anlass zur Hoffnung: Im neuen LehrplanPLUS (Bayern) für die gymnasiale Oberstufe werden (auf unsere Anregung hin) nun erstmals auch Baumdiagramme explizit erwähnt, in denen absolute Häufigkeiten in den Knoten eingetragen werden.

Der konkrete unterrichtliche Einsatz des Häufigkeitsdoppelbaumes wurde bereits von Wassner et al. (2007) mit Erfolg in der Unterrichtsreihe „Authentisches Bewerten und Urteilen unter Unsicherheit" für die gymnasiale Sekundarstufe I realisiert. Auch wenn sich die meisten bisherigen Vorschläge zum Einsatz von Häufigkeitsdoppelbäumen auf das Gymnasium beschränken, ist auch eine Implementierung an Real- und Mittelschulen möglich.

Anmerkung

Die Lehrerfortbildung „Daten und Zufall: Einführung in den LehrplanPLUS und neue didaktische Ansätze" mit zahlreichen weiteren Anwendungsbeispielen kann auf folgender Webseite heruntergeladen werden:

http://www.uni-regensburg.de/mathematik/didaktik-mathematik/lehrerfortbildungen/index.html

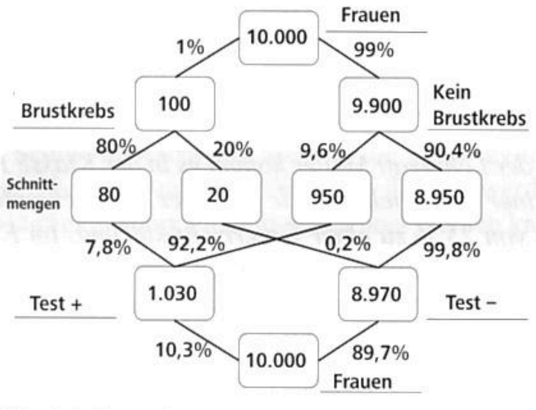


Abb. 16: Doppelbaum mit absoluten Häufigkeiten in den Knoten und relativen Häufigkeiten an den Ästen.

Literatur

Binder, K., Krauss, S. & Bruckmaier, G. (2015). Effects of visualizing statistical information – an empirical study on tree diagrams and 2 × 2 tables. *Frontiers in psychology, 6*(1186).

Böcherer-Linder, K. & Eichler, A. (2017). The impact of visualizing nested sets. An empirical study on tree diagrams and unit squares. *Frontiers in psychology, 7*(2026).

Brase, G. L. (2014). The power of representation and interpretation: Doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. *Journal of Cognitive Psychology, 26*(1), S. 81–97.

Bruckmaier, G., Binder, K. & Krauss, S. (2016). Numerische Darstellungsarten statistischer Informationen. In E.-M. Plackner & N. von Schroeders (Hrsg.), *Daten und Zufall. MaMut – Materialien für den Mathematikunterricht, 3* (S. 47–76). Hildesheim: Franzbecker.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In: Kahneman D., Slovic P., Tversky A. (Hrsg.): *Judgment under Uncertainty: Heuristics and Biases.* S. (249–267). Cambridge University Press, New York.

Eichler, A. & Vogel, M. (2009). *Leitidee Daten und Zufall: Von konkreten Beispielen zur Didaktik der Stochastik.* Springer-Verlag.

Engel, J. & Martignon, L. (2015). Dynamisch-interaktive Visualisierung elementarer Konzepte zu Daten und Wahrscheinlichkeiten. *Lernen und Lernstörungen, 4*(2), S. 139–145.

Freytag, C., Härtinger, R., Herz, A., Kammermeyer, F., Kurz, K., Sauermann, B., Sinzinger, M. & Zechel, J. (2005). *Fokus Mathematik 7 – Gymnasium Bayern.* Cornelsen, Berlin.

Frischemeier, D. (2017). Ein Statistikkurs für Mathematik-Studierende des Lehramts GHRGe „Statistisch denken und forschen lernen mit der Software TinkerPlots". In: R. Biehler (Hrsg.): *Statistisch denken und forschen lernen mit der Software TinkerPlots* (S. 189–320). Springer Fachmedien Wiesbaden.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), S. 684–704.

Götz, H., Schmid, A. & Weidig, I. (2010) *Lambacher Schweizer – Mathematik für Gymnasien Bayern 6.* Klett, Stuttgart.

ISB Bayern (o. J.). Übungsklausur 2013/2014 im Fach Mathematik – Länderübergreifender gemeinsamer Aufgabenpool. https://www.isb.bayern.de/download/14617/uebungskluebun_mathematik_2013_laendergemeinsamer_aufgabaufgab.pdf. Zugriff: 26.09.2017

Krauss, S. (2003). Wie man das Verständnis von Wahrscheinlichkeiten verbessern kann. *Stochastik in der Schule, 23*, S. 2–9.

Krauss, S. & Bruckmaier, G. (2014). Eignet sich die Formel von Bayes für Gerichtsverfahren? In U. Sproesser, S. Wessolowski & C. Wörn (Hrsg.). *Daten, Zufall und der Rest der Welt – Didaktische Perspektiven zur anwendungsbezogenen Mathematik* (S. 123–132). Wiesbaden: Springer.

Martignon, L. & Kurz-Milcke, E. (2006). *Educating children in stochastic modeling: Games with stochastic urns and colored tinker-cubes.* In A. Rossmann & B. Chance (Eds.), Working cooperatively in statistics education. ICOTS 7 – Salvador de Bahia. http://www.stat.auckland.ac.nz/~iase/publications/17/C443.pdf. Zugriff: 26.09.2017.

McDowell, M. & Jacobs, P. (im Druck). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological bulletin.*

Micallef, L., Dragicevic, P. & Fekete, J. D. (2012). Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics, 18*(12), S. 2536–2545.

Schmid, A. & Weidig, I. (2008). *Lambacher Schweizer – Mathematik für Gymnasien Bayern 10.* Klett, Stuttgart.

Stine, G. J. (1996). *Acquired immune deficiency syndrome: Biological, medical, social, and legal issues.* Englewood Cliff, NJ: Prentice Hall.

Wassner, C., Biehler, R., Schweynoch, S. & Martignon, L. (2007). Authentisches Bewerten und Urteilen unter Unsicherheit – Arbeitsmaterialien und didaktische Kommentare für den Themenbereich „Bayessche Regel" für den Stochastikunterricht der Sekundarstufe I. *Kasseler Online-Schriften zur Didaktik der Stochastik.*

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

Anschrift der Verfasser

Karin Binder
Lehrstuhl für Didaktik der Mathematik
Universität Regensburg
Universitätsstraße 31
93053 Regensburg
Karin.Binder@ur.de

Prof. Dr. Stefan Krauss
Lehrstuhl für Didaktik der Mathematik
Universität Regensburg
Universitätsstraße 31
93053 Regensburg
Stefan.Krauss@ur.de

Dr. Christoph Wassner
Martin-Behaim Gymnasium
Schultheißallee 1
90478 Nürnberg
wassner@martin-behaim-gymnasium.de