Dunkelfeldforschung

Antonius Warmeling, Hagen
Illustrationen von Dr. Wolfgang Zettlmeier, Barbing

© Peter Dazeley/The Image Bank/Getty Images Plus

Ladendiebstahl, Drogenkonsum oder auch Gewalt in Beziehungen sind sogenannte „Dunkelfelder". Das bedeutet, dass man zum Beispiel auf die Frage „Haben Sie schon einmal geklaut?" mit hoher Wahrscheinlichkeit keine ehrliche Antwort bekommt. Deshalb wird man bei solchen Befragungen z. B. den Anteil der Diebe in unserer Gesellschaft stark unterschätzen. Die Dunkelfeldforschung ist eine praktische Anwendung für folgenden stochastischen Verfahren und Sätze: bedingte Wahrscheinlichkeit, Pfadregeln und Satz von Bayes. In dieser Unterrichtseinheit üben Ihre Schüler diese Inhalte anwendungsorientiert und testen anschließend ihr Wissen in einer Lernerfolgskontrolle.

Dunkelfeldforschung

Oberstufe (Niveau)

Antonius Warmeling, Hagen
Illustrationen von Antonius Warmeling

Die Schüler lernen:

den Begriff der bedingten Wahrscheinlichkeit sicher anzuwenden. Sie bestimmen die Wahrscheinlichkeit verschiedener Ereignisse mithilfe von Baumdiagrammen. Auch den Satz von der totalen Wahrscheinlichkeit und den Satz von Bayes lernen sie kennen. Die Aufgaben sind in den Kontext „Dunkelfeldforschung" eingebettet, der die Schüler ansprechen wird. Eine Lernerfolgskontrolle rundet den Beitrag ab.

Überblick:

Legende der Abkürzungen:
Ab = Arbeitsblatt **LEK** = Lernerfolgskontrolle

Thema	Material	Methode
Haben Sie schon einmal geklaut?	M1	Ab
Ja (?!) – Karten für die Einstiegssimulation	M2	Ab
Varianten in der Dunkelfeldforschung – Übung 1	M3	Ab
Aufgaben zur Excel-Simulation	M4	Ab
Wovon ist die Güte der Schätzung abhängig?	M5	Ab
Lernerfolgskontrolle	M6	Ab, LEK

Erklärung zu Differenzierungssymbolen

einfaches Niveau	mittleres Niveau	schwieriges Niveau
	Dieses Symbol markiert Zusatzaufgaben.	

Kompetenzprofil:

Inhalt: Bedingte Wahrscheinlichkeit, Baumdiagramm, Pfadregeln, totale Wahrscheinlichkeit, Satz von Bayes, Ereignis, Gegenereignis, Dunkelziffer
Medien: GTR/CAS, GeoGebra
Kompetenzen: Mathematisch argumentieren und beweisen (K1), Probleme mathematisch lösen (K2), mathematisch modellieren (K3), mathematische Darstellungen verwenden (K4), Kommunizieren (K6)

Hinweise

Wenn man Menschen zu sozial unerwünschten Verhaltensweisen oder Einstellungen befragt, kann man davon ausgehen, dass viele nicht wahrheitsgemäß antworten und deshalb der Anteil der Menschen mit diesen Eigenschaften mindestens stark unterschätzt wird. Davon sind viele klassische Dunkelfelder betroffen wie zum Beispiel Drogenkonsum, Gewalt in Beziehungen oder auch der hier thematisierte Ladendiebstahl. Die Dunkelfeldforschung ist eine praktische Anwendung für die oben genannten stochastischen Verfahren und Sätze.

Voraussetzungen

Wichtige Grundlagen für die Behandlung der Dunkelfeldforschung sind der **Begriff der bedingten Wahrscheinlichkeit** und der Umgang mit **Baumdiagrammen**. Der **Satz von der totalen Wahrscheinlichkeit** und der **Satz von Bayes** sind hilfreich.

Bedingte Wahrscheinlichkeiten

A und B seien zwei Ereignisse und $\overline{A}$ bzw. $\overline{B}$ die entsprechenden Gegenereignisse. Dann bezeichnet $P(B|A)$ die Wahrscheinlichkeit des Ereignisses B, wenn vorher das Ereignis A eingetreten ist. Man liest kurz „P von B unter der Bedingung A". In einigen Schulbüchern wird die entsprechende Schreibweise $P_A(B)$ verwendet.

Mithilfe der Pfadmultiplikations- und der Pfadadditionsregel ergibt sich daraus der **Satz von der totalen Wahrscheinlichkeit** (vgl. Baumdiagramm):

$$P(B) = P(A) \cdot P(B \mid A) + P(\overline{A}) \cdot P(B \mid \overline{A})$$

Wenn $x = P(A)$ gesucht wird, erhält man aus der obigen Gleichung:

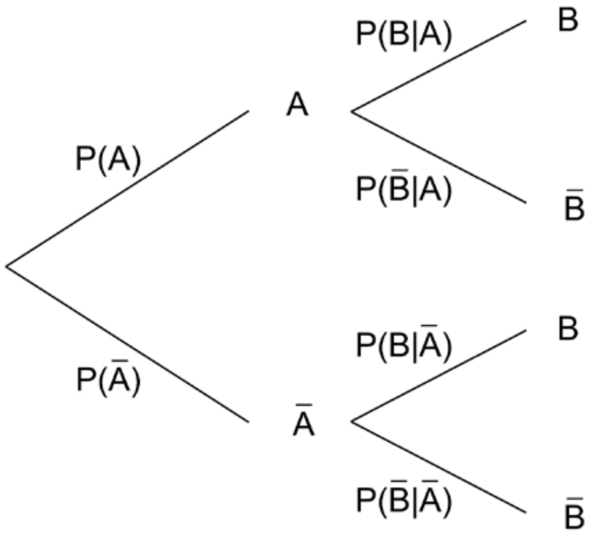
$$P(B) = x \cdot P(B \mid A) + (1 - x) \cdot P(B \mid \overline{A})$$

$$\Rightarrow P(B) - P(B \mid \overline{A}) = x \cdot (P(B \mid A) - P(B \mid \overline{A}))$$

$$\Rightarrow \frac{P(B) - P(B \mid \overline{A})}{P(B \mid A) - P(B \mid \overline{A})} = x = P(A)$$

Nach dem **Satz von Bayes** gilt:

$$P(A|B) = \frac{P(A) \cdot P(B \mid A)}{P(A) \cdot P(B \mid A) + P(\overline{A}) \cdot P(B \mid \overline{A})}$$



Grafik: Dr. W. Zettlmeier

Dunkelfeldforschung mit der Randomized-Response-Technik

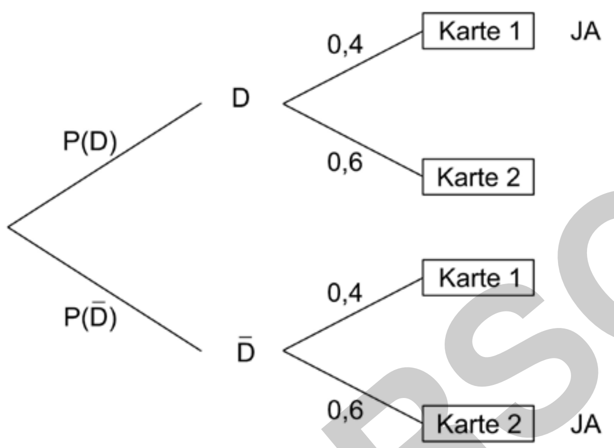
Die Dunkelfeldforschung, die aus den USA kommt, versucht durch Befragungen den Anteil der nicht erfassten Kriminalfälle (**Dunkelziffer**) zu ermitteln. Das ist aber sehr aufwendig, sodass es in Deutschland bis jetzt tatsächlich nur wenige Studien dazu gibt. Wenn man Menschen zu sozial unerwünschten Verhaltensweisen oder Einstellungen befragt, kann man davon ausgehen, dass viele nicht wahrheitsgemäß antworten und deshalb der Anteil der Menschen mit diesen Eigenschaften mindestens stark unterschätzt wird. Davon sind viele klassische Dunkelfelder betroffen wie zum Beispiel **Drogenkonsum, Gewalt in Beziehungen** oder auch der hier thematisierte **Ladendiebstahl**.

Bei der **Randomized-Response-Technik** sorgt eine Zufallsverschlüsselung dafür, dass die Einstellung des Einzelnen geschützt ist, eine Gruppenauswertung der befragten Stichprobe aber dennoch möglich ist. Dies soll hier am Beispiel einer Befragung zum Ladendiebstahl verdeutlicht werden. Der Kriminologe fragt nicht: „Haben Sie schon einmal einen Ladendiebstahl begangen?“, sondern lässt die Befragten aus einem Stapel von z. B. **4 Karten mit Frage 1 und 6 Karten mit Frage 2** eine Karte ziehen. Dann bittet er das Gegenüber um eine ehrliche Antwort auf die gezogene Frage.

Auf Karte 1 steht „Ist es richtig, dass Sie schon einmal einen Ladendiebstahl begangen haben?“, was ein Ladendieb wahrheitsgemäß mit JA beantworten müsste. Die zweite Frage „Ist es richtig, dass Sie noch nie einen Ladendiebstahl begangen haben?“ müsste dagegen der Nicht-Dieb mit JA beantworten. Der Kriminologe kann also aus der einzelnen Antwort keine Rückschlüsse ziehen, weil er nicht weiß, welche Karte der Befragte bezogen hat.

Es sei X(JA) der Anteil der JA-Sager in der Stichprobe. Die Abschätzung von $P(D) = x$ erfolgt nun über $X(JA) = x \cdot 0{,}4 + (1 - x) \cdot 0{,}6$.

Durch Umformung erhält man $x = \dfrac{X(JA) - 0{,}6}{-0{,}2}$.



Grafik: Dr. W. Zettlmeier

Man erkennt zwei **Probleme**, beide werden in der Unterrichtsreihe thematisiert:

1. Ist $X(JA) \geq 0{,}6$, antworten also mehr als mindestens 60 % der Befragten mit JA, so ist die Berechnung nicht möglich, weil sich für P(D) ein negativer Wert ergibt. Diese Gefahr ist umso geringer, je größer die Stichprobe ist.
2. Die Wahrscheinlichkeit P(JA) kann durch den Anteil der JA-Sager in der Stichprobe (X(JA)) abgeschätzt werden. Dass dieser Schätzer erwartungstreu ist, soll hier nicht thematisiert werden, kann aber bei Meyer [2] nachgelesen werden. Große Zufallsschwankungen können durch relativ große Stichproben vermieden werden.

Durch die Wahl geeigneter Parameter (z. B. das Verhältnis der Kartenanzahlen 1 und 2) kann man die Streuung der Schätzwerte verkleinern. Dies kann aber zulasten der Anonymisierung gehen, wie man mit P(D|JA) und/oder P($\bar{D}$ |NEIN) zeigen kann.

Die Simulationsdateien

Zur Simulation größerer Stichproben gibt es im Archiv mehrere Dateien.

Die Datei **diebe1_einfach.xls** ist schon ab Excel 97 und auch mit OpenOffice.CALC lauffähig. In die gelben Felder gibt man zunächst den Stichprobenumfang n, die Anzahl der Karten 1 und 2 sowie den tatsächlichen Anteil der Diebe p ein. Danach wird durch die Programmierung für jede Person aus der Stichprobe festgelegt, ob sie ein Dieb ist oder nicht (in Spalte D), und die Ziehung einer Karte (Spalte E) zusammen mit der entsprechenden Antwort (Spalte F) simuliert. Im grünen Feld A7 wird zusätzlich noch der Anteil der Diebe in der Stichprobe angegeben, weil $n \cdot p$ nicht unbedingt eine natürliche Zahl ergibt. In G8 und G9 finden Sie die Anzahl der JA-Stimmen und die Schätzung für P(D). Die Simulation lässt sich durch Drücken von F9 beliebig oft wiederholen.

Die Datei **diebe1_einzeln.xlsm** ist z. T. mit Visual Basic realisiert und enthält Funktionen, die erst mit Excel 2010 lauffähig sind. Da sie Makros enthält, müssen Sie unter Datei → Optionen → Sicherheitscenter → Einstellungen für Sicherheitscenter → Einstellungen für Makros „alle Makros mit Benachrichtigung deaktivieren".

Dann können Sie nach Start der Datei die Makros per Klick aktivieren. Die Alternative „alle Makros aktivieren" ist auch möglich, wird aber aus Sicherheitsgründen von Microsoft nicht empfohlen.

Die Eingaben in der Datei sind genauso wie in **diebe_einfach.xls**. Zusätzlich werden in einem Häufigkeitsdiagramm die klassierten Daten für P(D) (Klassenbreite 0,01) dargestellt. Zu Beginn der Simulation drücken Sie einmal den Button **Reset**, der die Anzahl der Simulationen auf Null setzt und die alten Ergebnisse löscht. Dann fügen Sie mit Klick auf **neue Simulation** jeweils das Ergebnis einer Simulation hinzu. Unter dem Diagramm werden noch das arithmetische Mittel und die Standardabweichung zur besseren Orientierung der Streuung der Daten angegeben. Wenn die Schüler das nicht sehen sollen, setzen Sie die Schriftfarbe einfach auf Weiß.

Die Datei **diebe1_hundert.xlsm** enthält zusätzlich einen Button, der 100 Simulationen durchrechnet und auf einen Schlag im Häufigkeitsdiagramm darstellt.

Diebe2_einfach.xls und **diebe2_einzeln.xlsm** sind dagegen für die Hand des Lehrers gedacht, um z. B. auch für ein alternatives Verfahren (siehe **M 3**, Aufgabe 2 und **M 6**) Schätzwerte oder Häufigkeitsverteilungen erstellen zu können.

Zur Durchführung der Unterrichtsreihe

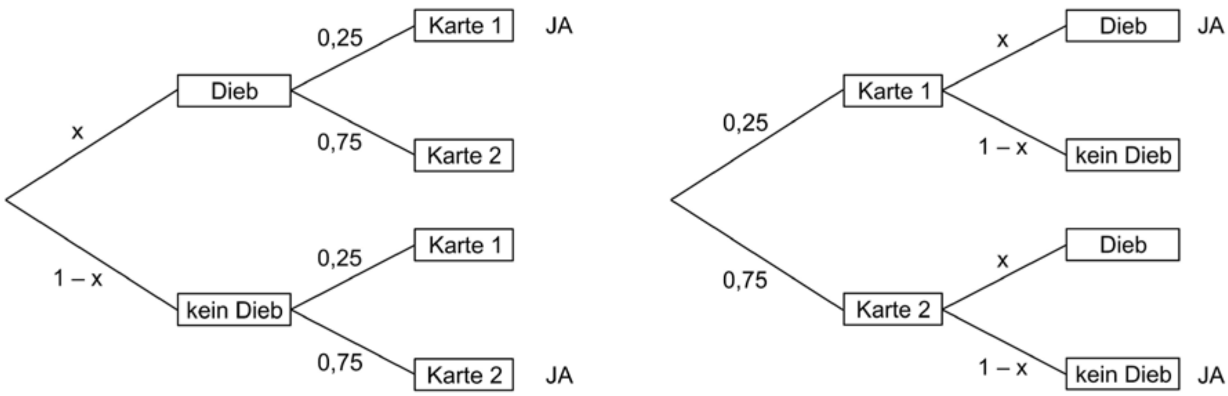
Sie sollten dieses Verfahren der Dunkelfeldforschung nicht zur Bestimmung tatsächlicher Wahrscheinlichkeiten für einen Ladendiebstahl, Drogenkonsum o. Ä. nutzen. Das könnte nicht nur zu Irritationen bei den Lernenden und ihren Eltern führen. Die Stichproben sind dafür zu klein und Ihnen fehlt – zum Vergleich und für die Simulationen – die tatsächliche Wahrscheinlichkeit P(Dieb).

Daher markiert der Autor die Schüler zu Beginn der Einstiegsstunde mithilfe von kleinen Zetteln, die sie aus einer Socke o. Ä. ziehen und keinem anderen zeigen, als DIEB oder KEIN DIEB. Damit kann er P(Dieb) bezogen auf den Kurs auch dann feststellen, wenn mal ein oder zwei Schüler fehlen. Wenn Sie z. B. von 24 Zetteln 8 mit DIEB gekennzeichnet haben, so ist P(D) = 1/3. Fehlt ein Schüler, so ist – je nach übrig gebliebenem Zettel – p(D) = 7/23 oder p(D) = 8/23.

Auf der Basis dieser Markierung führen die Schüler nun die erste Simulation in Partnerarbeit durch. Das Verfahren ist auf dem Arbeitsblatt **M 1** beschrieben. Es ist so gestaltet, dass Sie bei Bedarf auch andere Kartenverhältnisse realisieren können. Eigentlich müsste man vor jeder weiteren Simulation die Schüler neu markieren.

Da dabei aber immer der gleiche Anteil und damit die gleiche Wahrscheinlichkeit herauskommt und die zweite Stufe der Simulation ein Zufallsexperiment ist, kann man darauf verzichten. Nachdem die Ergebnisse von 10 Simulationen an der Tafel stehen, eröffnen Sie die Auswertung zum Beispiel mit der Frage: „Können Sie mir nun die Wahrscheinlichkeiten angeben, dass ein beliebiger Schüler schon mal einen Ladendiebstahl begangen hat?"

In der Regel werden die Ergebnisse sehr stark streuen. X(JA) lässt sich schnell als Quotient aus Anzahl der Ja-Sager und der Größe der Stichprobe bestimmen. Aber wie können die Lernenden nun P(D) abschätzen? Bei den Schülern des Autors, die um die Wichtigkeit dieses Veranschaulichungsmittels wissen, kam sofort der Vorschlag, die Befragung mithilfe eines Baumdiagramms darzustellen. Gegebenenfalls hilft der Hinweis, dass es sich um einen zweistufigen Zufallsversuch handelt. Dabei können die Schüler – zum Beispiel bei 1 x Karte 1 und 3 x Karte 2 – zwei verschiedene Baumdiagramme zeichnen:



Grafik: Dr. W. Zettlmeier

Beide Diagramme führen zu folgender Gleichung, wobei X(JA) der Anteil der JA-Sager in der Stichprobe ist.

$$X(JA) = 0{,}25x + 0{,}75(1 - x) = 0{,}75 - 0{,}5x \quad \Rightarrow \quad x = \frac{0{,}75 - X(JA)}{0{,}5}$$

Da die Wahrscheinlichkeiten für Karte 1 bzw. Karte 2 und damit auch für das Ereignis „JA" aber nicht gesichert sind, sondern lediglich in langen Serien angenähert erreicht werden, kann über die Bestimmungsgleichung nur eine Schätzung von P(D) erfolgen.

Die Ergebnisse X(D) werden in der Regel ebenfalls stark differieren, evtl. sind sogar eine Null oder ein negatives Ergebnis dabei (wenn $X(JA) \geq 0{,}75$). Eine erste Auswertung könnte die Berechnung der Lageparameter $\bar{x}$ (arithmetisches Mittel) und s (Standardabweichung) sein. Notfalls kann man sich statt s auch mit der Spannweite begnügen. Die Streuung der Werte bietet Gelegenheit, über eine Verbesserung der Schätzung nachzudenken. Der erste Vorschlag geht immer in die Richtung, den Stichprobenumfang zu vergrößern. Meistens schlagen die Schüler dazu vor, die bisherigen Stichproben zu einer zusammenzufassen. Das ist zwar in der Demoskopie streng genommen nicht erlaubt, weist aber die Verbesserungen bei größeren Stichproben klar nach. Nachdem auch für die Gesamtstichprobe das X(D) berechnet ist, verraten Sie den Schülern zum Vergleich den wahren Anteil der Diebe. Auch wenn sich die Schätzung zufälligerweise als ziemlich gut erweist, sollten Sie herausarbeiten, dass das nicht so sein muss. Vielmehr gilt es darüber nachzudenken, wie man die Güte der Schätzung durch geeignete Veränderungen verbessern kann.

Für diese erste Erarbeitungsphase brauchen Sie zwei Stunden. Wichtig ist, dass die Schüler den Unterschied zwischen der Wahrscheinlichkeit P(D) und der relativen Häufigkeit X(D) als Ergebnis einer Stichprobe – und damit Schätzwert für P(D) – auseinanderhalten. Im Anschluss daran sollten Sie Zeit geben für die Übung 1 (**M 3**) zur Festigung des bisher Gelernten. Aufgabe 1 können Sie durchaus als Hausaufgabe vorbereiten lassen. Eine ausführliche Besprechung ist aber nötig, insbesondere um den letzten Fall herauszuarbeiten, dass bei $p(\text{Karte 1}) = p(\text{Karte 2}) = 0{,}5$ das Verfahren nicht funktioniert, d. h. keine Schätzung möglich ist.

Mit Material **M 4** steigen Sie in die Diskussion um die Verbesserung der Schätzung ein. Der Vorschlag, den Stichprobenumfang zu vergrößern, sollte schon am Ende der 2. Stunde gekommen sein. Setzen Sie daher die Simulation **diebe1_einzeln.xlsm** (notfalls auch **diebe1_einfach.xls**) ein, um den Stichprobenumfang Schritt für Schritt zu erhöhen. Wichtig ist, dass die Schüler merken, dass eine einmalige Simulation irreführend sein kann, weil auch bei kleinem Stichprobenumfang die richtige Wahrscheinlichkeit zufällig getroffen werden kann. Erkenntnisse bringt also nur eine Serie von Simulationen, deren Streuung entweder über die Standardabweichung oder notfalls auch über die Spannweite erfasst werden kann.

Danach kommt die Phase, in der die Schüler in Gruppen bestimmte Thesen durch gezielt ausgewählte Simulationen (siehe **M 5**) untersuchen. Dazu ist es wichtig, dass jeder Gruppe mindestens ein **PC** oder **Laptop** zur Verfügung steht, damit Ihre Schüler ihre Ideen ausprobieren und umsetzen können. Zum Beispiel durch Kopieren der Häufigkeitsverteilungen in ein Worddokument können sie ihre Ergebnisse anschließend den Mitschülern präsentieren. Falls Sie diese Untersuchungen aufgrund der Rahmenbedingungen auch im Plenum machen müssen, sollten Sie sich auf die Aussagen b) oder c) konzentrieren. Material **M 6** können Sie zur Leistungsüberprüfung einsetzen. Aufgabe 2 erfordert allerdings eine ausführliche Bearbeitung und Besprechung von **M 4**.

Minimalplan

Wenn Sie wenig Zeit haben, können Sie das Arbeitsblatt zur Verbesserung der Schätzungen (**M 5**) weglassen.