Bedingte Wahrscheinlichkeit – Ein Schlüssel zur Stochastik

MANFRED BOROVCNIK, KLAGENFURT

Bedingte Wahrscheinlichkeit ist ein Schlüsselbegriff in der subjektivistischen Theorie. In der klassischen Theorie und Auffassung von Wahrscheinlichkeit – in Anlehnung an Kolmogorows Axiome jedoch ist bedingte Wahrscheinlichkeit lediglich ein untergeordneter Begriff. Dieser unterschiedliche Status führt naturgemäß zu Verwerfungen nicht nur in der theoretischen Konzeption sondern auch im persönlichen Verständnis stochastischer Begriffe. Der Aufsatz soll bedingte Wahrscheinlichkeit aus verschiedenen Perspektiven beleuchten:

- Nutzen *stochastischer* Information und Abgrenzung dieser gegenüber anderen Arten von Information.
- Mathematik und Philosophie.
- Didaktische Diskussion.
- Ideen und Strategien im Wettstreit.

Die Ausführungen belegen, dass die üblichen begrifflichen Reduktionen wenig hilfreich sind und erst ein weiter gefasster Begriff von Wahrscheinlichkeit die wohlbekannten Schwierigkeiten meistern lässt.

1. Perspektive des Nutzens: Wozu stochastische Information

Wenn eine Information andere Aussagen logisch ableiten lässt, so liegen die Dinge vergleichsweise einfach. Wenn kausale Zusammenhänge bekannt sind und man weiß, dass die genauen „ursächlichen" Bedingungen eingehalten sind, so kennt man die Folgen. In vielen Fällen hat die Information jedoch nur probabilistischen Charakter: sie ändert die Wahrscheinlichkeiten für andere Aussagen. Hartnäckig versuchen viele Menschen, die Situation so umzudeuten, dass sie gewohnte (logische oder kausale) Zusammenhänge erkennen oder vermischen logische Argumente mit vagen Wahrscheinlichkeitsbetrachtungen, um ihre Schlüsse zu ziehen und zu rechtfertigen. Anhand einfacher Beispiele soll gezeigt werden, wie wichtig eine stochastische Verwertung von Information ist, um entsprechende Schlüsse zu ziehen, und wie wichtig es ist, stochastische Schlüsse gegenüber anderen (logischen und kausalen) Denkmustern abzugrenzen.

1.1 Ursache und Wirkung

Hier soll anhand eines einfachen Urnenbeispiels demonstriert werden, dass stochastisches Denken und kausales Denken auf gänzlich verschiedenen Schienen verlaufen, wenngleich die beiden Ansätze intuitiv gerne vermischt werden.

Ein einfaches Beispiel: Die Falk-Urne

Wir ziehen zwei Mal ohne Zurücklegen aus einer Urne, in der sich zwei weiße und zwei schwarze Kugeln befinden (Falk & Konold 1992):

i. Man sieht die *erste (weiß)* und fragt, *ob die zweite weiß sein wird?*

ii. Die erste Kugel bleibt unbekannt.
Man sieht die *zweite (weiß)* und fragt, *ob die erst weiß war?*

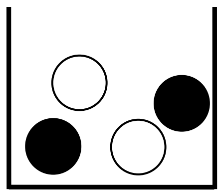
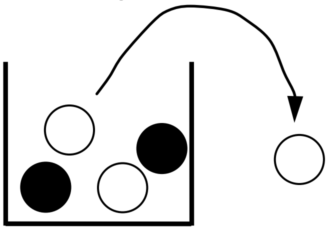


Abb. 1: Falk-Urne

Wahrscheinlichkeit ist für viele Menschen – und so wird es auch in der üblichen, klassischen (objektivistischen Theorie) vertreten – an das Objekt gebunden; Wahrscheinlichkeit ist eine Eigenschaft eines Objekts. Hier ist es eine Eigenschaft von bestimmten Urnen – eine Vorstellung, die nicht zuletzt durch unterrichtliche Bemühungen (die auch wichtige Einsichten vermitteln) nachhaltig verstärkt wird.

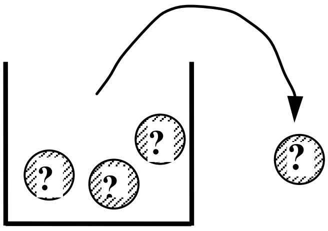
Zwei ganz unterschiedliche Aufgaben?

Aufgabe i.



1. Zug:

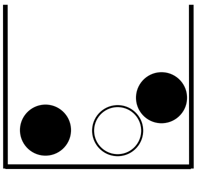
Kugel ist weiß.



2. Zug:

Wie groß ist die Wahrscheinlichkeit, dass auch die zweite Kugel weiß ist?

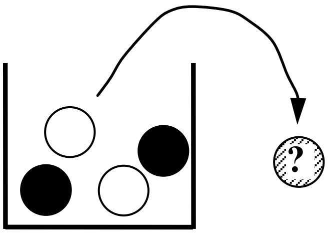
Wenn eine Urne da und in ihrer Zusammensetzung bekannt ist, so „verursacht“ sie die entsprechende Wahrscheinlichkeit. In der Tat, beim Vorwärts-Schluss ist die Urnenbelegung bekannt; das Beispiel macht üblicherweise keinerlei Schwierigkeiten.



Klar:
Die Urnenbelegung beim 2. Zug ist *bekannt*.

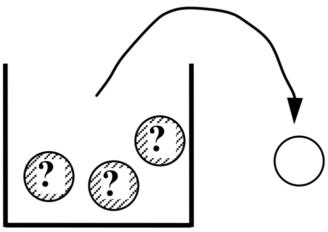
Abb. 2: Die bekannte Urne liefert die Lösung.

Aufgabe ii.



1. Zug:

Die Kugel bleibt verborgen.

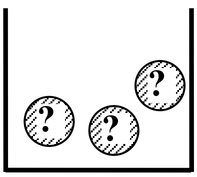


2. Zug:

Die Kugel ist weiß.

Wie groß ist die Wahrscheinlichkeit, dass auch die erste Kugel weiß ist?

Jetzt tun sich die Schwierigkeiten auf, denn man findet keine geeignete Urnenbelegung, welche die Aufgabe repräsentiert. Weil die Aufgabe stochastisch doch einigermaßen schwierig ist (und nicht nur deswegen), weichen viele auf Ersatzstrategien aus.



Oder:
Urnenbelegung ist *unbekannt*!
Wie sieht die Urnenbelegung beim zweiten Zug aus?

Abb. 3: Keine Urne – keine Lösung – oder?

Die Faszination von Ursache und Wirkung

i. In der Urne befinden sich nach dem ersten Zug weiße und schwarze Kugeln im Verhältnis 1:2; die Antwort lautet – ohne Probleme – 1/3. Die Begründungen sind meist kausaler Art: Jetzt hat man eben diese Urnenbelegung.

ii. Weil eine Repräsentation der Aufgabenstellung durch eine Urne fehlt, kommt eine Vielfalt persönlicher Strategien zur Wirkung; eine typische Reaktion lautet (siehe Borovcnik & Bentz 1990/2003): „Beim 1. Zug hatten wir 2 zu 2, daher 2/4 = ½; spätere Ereignisse können frühere nicht beeinflussen."

Zeitgebundenes Denken ist typisch für das kausale Paradigma. Die kausale Unabhängigkeit ist dann ein starkes Argument, das jetzt verhindert, anzuerkennen, dass später auftretende Ereignisse doch die Wahrscheinlichkeit früherer Ereignisse (deren Ausgang noch nicht bekannt ist) beeinflussen können. Sogar in extremeren Varianten des Beispiels haben hartnäckige Kausaldenker es abgelehnt, aus den späteren – bekannten – Ereignissen eine relevante Information zu gewinnen für die Wahrscheinlichkeitsbeurteilung des früheren (noch unbekannten) Ereignisses.

1.2 Medizinische Diagnose einer Krankheit *D*

Die medizinische Diagnose ist ein paradigmatischer Kontext, in dem genuin stochastische Information verwendet wird. Anhand von Indizien (den medizinischen Befunden) wird auf mögliche Krankheiten zurückgeschlossen. Immer besteht ein Risiko für Fehldiagnosen, der Zustand (welche Krankheit zutrifft) kann nur mit Wahrscheinlichkeiten evaluiert werden. Analoge Situationen treten vor Gericht auf, wenn kein Geständnis vorliegt (und sogar, wenn eines vorliegt). Die Schuldfrage wird anhand von Indizien zu klären sein. Belastende Information erhöht, entlastende verringert die Wahrscheinlichkeit der Täterschaft. Auch hier gibt es zwei unterschiedliche Fehlurteile, die verschiedene Auswirkungen haben. Immer sind *bedingte* Wahrscheinlichkeiten im Spiel, immer werden diese durch Auftreten (oder durch *Unterstellen*) neuer Information neu bewertet.

Das allgemeine Problem des Schließens von Befunden auf dahintersteckende Krankheiten

Nehmen wir an, die bedingte Wahrscheinlichkeit für einen positiven medizinischen Befund + (der ein Indiz für das Vorliegen einer Krankheit ist) sei bekannt: $P(+ \mid D) = 0{,}99$. Eine Person hat einen positiven Test. *Wie groß ist die Wahrscheinlichkeit, dass sie tatsächlich die Krankheit D hat?*

Gefragt ist die umgekehrte bedingte Wahrscheinlichkeit $P(D \mid +) = ?$ Ein häufig zu beobachtender Fehlschluss besteht in folgender *Gleichsetzung*: $P(D \mid +) = P(+ \mid D) = 0{,}99$. Auf der anderen Seite haben die beiden bedingten Wahrscheinlichkeiten eine völlig unterschiedliche Interpretation: Von Krankheit *D* zum Testergebnis + wird der Zusammenhang *kausal* interpretiert; diese Konnotation haftet auch an der entsprechenden bedingten Wahrscheinlichkeit. Vom Testresultat + zur Krankheit *D* aber sind Schlüsse durchaus zweifelhaft; viele Menschen können mit einer *diagnostisch* interpretierten bedingten Wahrscheinlichkeit wenig anfangen. Interessant ist die Spaltung in zwei „Lager": die einen glauben fest an die (hier) 0,99, während die anderen den gesamten Schluss in Zweifel ziehen.

Ein aktuelles Beispiel: Mammographie-Screening

Mammographie wird empfohlen, um Brustkrebs frühzeitig zu entdecken – zumindest für Frauen 40+. Prävalenz der Krankheit (Anteil an der Bevölkerung): 0,8% der Frauen zwischen 40 und 50 bekommen Brustkrebs. Über die Zuverlässigkeit der Diagnoseprozedur gibt es unterschiedliche Annahmen; häufig findet man folgende Angaben:

- Wenn eine Frau Brustkrebs hat, zeigt dies das Mammogramm mit einer Zuverlässigkeit von 90%.
- Wenn sie wirklich gesund ist, besteht ein Risiko von 7% für ein positives Mammogramm.

Wie groß ist die Wahrscheinlichkeit, dass eine Frau mit einem positiven Mammogramm wirklich Brustkrebs hat? Ärzte geben hier oft sehr hohe Wahrscheinlichkeiten an, meist über 90% (vielleicht in der naiven Umkehrung von oben). Patienten teilen sich wiederum in die zwei oben genannten Lager. Was die Einschätzung der Informationen durch Ärzte anbelangt, so findet man in Gigerenzer (2002) mehr; der Autor hat selbst in einer Gruppe von Primarärzten (auch Gynäkologen) nachgefragt mit ähnlich schlechten Ergebnissen (Borovcnik 2012).

1.3 Monty Hall – das Problem

Der Preis – ein Auto – ist hinter einem von 3 Vorhängen (Türen) versteckt, während hinter den anderen jeweils eine Ziege ist. Der Kandidat wählt ... Der Moderator ... Ziege ... bietet an ... wechseln ... Was ... tun? *Wie groß ist nun die Wahrscheinlichkeit, dass der Kandidat die richtige Tür hat?*

Abb. 4: Monty Hall – Hinter den Vorhängen.

Es ist sagenhaft, wie emotionalisiert die Diskussion um dieses einfach anmutende Beispiel geführt wird (siehe vos Savant, o.D.). Der Autor hat nun schon drei Wellen öffentlichen Interesses an diesem Problem miterlebt. Die Argumente gleichen einander bis in die genaue Formulierung. Viele Menschen lassen sich nicht davon überzeugen, dass Wechseln – gemäß den genau ausgeführten Spielregeln – die Situation bessert. Denn: die bedingte Wahrscheinlichkeit, die richtige Tür gewählt (und das Auto) zu haben, ist 1/3 (wie am Anfang!). Die Information, die der Moderator mit der geöffneten Tür anbietet, ist also wertlos hinsichtlich der gewählten Tür. Allerdings: Wenn diese nur 1/3 hat und nur mehr eine weitere Tür übrig bleibt, muss diese den Rest an Wahrscheinlichkeit auf sich vereinen, d.h., sie hat eine bedingte Wahrscheinlichkeit von 2/3. Daher verdoppelt Wechseln die Erfolgsaussichten.

Die mathematische Lösung (siehe Abschnitte 4 und 5) ist für viele Personen durchaus zugänglich: auch Mathematiker und Statistiker, etwa der berühmte Paul Erdös, befinden sich unter jenen, die sie kategorisch ablehnen. Die Simulation des Spieles viele Male ist noch am besten geeignet, die Zweifler zu überzeugen (?). Jedoch, wovon überzeugen; sie gehen lediglich vor der bitteren Realität der relativen Häufigkeiten in die Knie. Verstanden haben sie noch immer nicht, wieso die Information des Moderators diese Auswirkung hat. Der Umgang mit bedingten Wahrscheinlichkeiten, mit der Revision von Wahrscheinlichkeitsurteilen aufgrund neuer Informationen (Indizien) hat eigentlich nirgends im der – klassischen – Konzeption von Wahrscheinlichkeit einen rechten Platz und wird daher durch Beherrschen der Mathematik der Stochastik auch wenig gefördert.

2. Aus mathematischer und philosophischer Sicht

Was ist Zufall? Damit beschäftigt sich nicht nur die Stochastik als mathematische Disziplin (siehe Barnett 1973) sondern auch die Wissenschaftstheorie als Teildisziplin der Philosophie (siehe Stegmüller 1973). Jenseits der streng wissenschaftlichen Begründung von Begriffen kann man natürlich auch fragen, was Zufall eigentlich bedeuten soll – und wird zu keinem Ende kommen. So etwa hat Zufall durchaus etwas mit Divination (Gottesentscheid) zu tun und die Überlappung solch extrem unterschiedlicher Ansätze kann nur mit Verlust einer ganzheitlichen Perspektive ausgeblendet werden (Borovcnik 2011). Für uns essentiell ist, dass auch Antworten zu Zufall innerhalb der Mathematik von der jeweiligen Schule abhängen.

2.1 Objektivistische Fassung von Wahrscheinlichkeit

Der übliche Begriff von Wahrscheinlichkeit ist an die Deutung als relativer Häufigkeit in einer Serie von unabhängigen Wiederholungen desselben Zufallsexperiments gebunden. Diese Deutung gewinnt durch die Axiomatisierung von Kolmogorow fast Alleinstellungsrecht und lässt andere Deutungen von Wahrscheinlichkeit zurück. Entsprechend dominiert sie viele Lehrbücher von Stochastik und deren Anwendungen.

Zentrale Interpretation von Wahrscheinlichkeit in der objektivistischen Theorie

Bernoulli hat 1713 mit seinem Theorema aureum (heute als Bernoullisches Gesetz der großen Zahlen benannt) diese Deutung erstmals in den Vordergrund gebracht und gleichzeitig gerechtfertigt. Diese Deutung ermöglicht eine empirische Überprüfung einer im Modell unterstellten Wahrscheinlichkeit durch ein Experiment, was im naturwissenschaftlichen Paradigma seit der Renaissance ein grundsätzliches Erfordernis ist. Damit wird Wahrscheinlichkeit auch für die Physik tauglich, wo sie zunehmend zur Theoriebildung beiträgt (Entropie, Thermodynamik, um nur zwei Schlagworte zu nennen). Entsprechend befasst sich Kolmogorow in seiner bahnbrechenden Axiomatisierung des Wahrscheinlichkeitsbegriffs mit der Frage der Deutung. Kolmogorow (1933) sieht seine Theorie explizit als Rechtfertigung der Deutung von Wahrscheinlichkeit als relative Häufigkeit.

Zum ersten ermöglicht der axiomatische Begriff die Konstruktion von Wahrscheinlichkeitsmaßen auf Produkträumen. Dabei leistet der nicht in den Axiomen festgesetzte Begriff der Unabhängigkeit einen wesentlichen Beitrag. Mehr noch, es gelingt – über maßtheoretische Sätze (siehe auch Winkler 2012) – Wahrscheinlichkeit auf den unendlich-dimensionalen Raum aller Folgen zu erweitern. Auf diesen Folgen von möglichen Versuchsergebnissen kann man ein starkes Gesetz der großen Zahlen als eine Art Konvergenz beweisen. Dies sieht Kolmogorow als deutliche Absicherung, dass der damit gegründete Wahrscheinlichkeitsbegriff als (Idealisierung von) relativen Häufigkeiten gedeutet werden soll.

Interessant dabei, dass Abhängigkeit und bedingte Wahrscheinlichkeit in der Kolmogorow-Theorie selbst keinerlei Rolle spielen, wenngleich die Axiomatisierung zu einem Boom in der Untersuchung von stochastischen Prozessen, speziell auch von Markow-Prozessen führte. Bei diskreten Zustandsräumen in Markow-Ketten sind Übergangswahrscheinlichkeiten zwischen den Zuständen – das sind bedingte Wahrscheinlichkeiten – Basiselemente der mathematischen Modellbildung.

Bedingte Wahrscheinlichkeit für ein festes Ereignis E ordnet beliebigen Ereignissen A den Wert $P_E(A) := P(A \mid E) := \dfrac{P(A \cap E)}{P(E)}$ zu und ist nichts anderes als ein Wahrscheinlichkeitsmaß, d.h., sie erfüllt alle Axiome von Kolmogorow. Wie man von E und seiner Wahrscheinlichkeit zur bedingten Wahrscheinlichkeit von A kommt – welche Änderung im Vergleich zur (unbedingten) Wahrscheinlichkeit erfolgt (in Größe und Richtung), ist ohne Belang; weder für die Theoriebildung noch für Anwendungen. Es wird – in der Formel von Bayes – zu einem trivialen mathematischen Satz.

Allerdings führt der Weg zur Unabhängigkeit interessanterweise über die bedingte Wahrscheinlichkeit. Wenn nämlich die bedingte gleich der unbedingten Wahrscheinlichkeit ist, i.e. $P_E(A) = P(A)$, was auch als $P(A \cap E) = P(A) \cdot P(E)$ geschrieben werden kann, sagt man, die beiden Ereignisse sind *unabhängig*. Diese Produktformel bietet eine elegante Eliminierung der bedingten Wahrscheinlichkeit. Während also Unabhängigkeit zum zentralen Angelpunkt der Kolmogorow-Theorie wird, bleibt bedingte Wahrscheinlichkeit auf der Strecke.

Schlüssel und Mühlstein „Unabhängigkeit"

Der Begriff Unabhängigkeit wird damit zum Schlüssel für die Modellierung wiederholter Experimente – *das* Paradigma der üblichen Auffassung von Wahrscheinlichkeit. Gleichermaßen wird sie zur Voraussetzung zentraler Sätze der Wahrscheinlichkeitstheorie wie beispielsweise der Gesetze der großen Zahlen oder des zentralen Grenzverteilungssatzes (in der ursprünglichen Fassung); Winkler (2012) spricht von einem Fest der Unabhängigkeit. Trotz aller Bemühungen Kolmogorows, den Wahrscheinlichkeitsbegriff als (idealisierte) relative Häufigkeiten zu interpretieren, war man in der Mathematikergemeinschaft froh, dass eine zufriedenstellende mathematische Basis geschaffen war, um endlich die Vagheit von Intuitionen zu überwinden, weil Intuitionen gerade in diesem Bereich oft mehr Verwirrung stiften denn Klärung schaffen. Im Falle der Anwendung von Unabhängigkeit versucht man sich mit Fehlen „kausaler" Einflüsse zu behelfen. Ohnedies, so verweist man, gibt es die Möglichkeit, diese Modellannahme empirisch durch Experimente über einen statistischen Test zu validieren.

Der scheinbaren Rationalität des Ansatzes muss man aber einiges entgegen halten: So versagt der axiomatischer Zugang für den rationalen Aufbau statistischer Inferenz (siehe Stegmüller 1973). In der Wissenschaftstheorie wird nämlich gezeigt, dass statistische Tests (auf Basis jedweder objektivistischer Schule) einer rationalen Begründung entbehren. Sie arten mehr oder weniger zu einer Plausibilitätsprüfung aus. Besonders schwer trifft die Kritik die Möglichkeit der Prüfung der Annahme der Unabhängigkeit; diese ist experimentell keinem statistisch „sauberen" Test zugänglich, auch wenn ein Blick in die Statistikbücher einem anderes suggerieren will. Für das Lernen der Begriffe ist noch anzumerken, dass bedingte Wahrscheinlichkeit eine Quelle von Problemen darstellt; davon ist ja auch hier die Rede.

2.2 Subjektivistische Fassung der Theorie

Qualitative Information, durch Zahlen in ein Wahrscheinlichkeitsurteil umgemünzt, haben immer den Geruch von *Subjektivität*. Schon bei der Diskussion um die so genannte *moralische Wahrscheinlichkeit* für einen objektivistischen Begriff gehen die Auffassungen auseinander: unter welchem Schwellenwert sollten Wahrscheinlichkeiten einfach gleich null gesetzt werden? Reicht 10^{-10} oder muss man viel kleiner ansetzen? Im Gegensatz zur objektgebunden Zahl Wahrscheinlichkeit, die eine Eigenschaft eines Objekts (einer Versuchsanordnung; etwa ein Würfel) ist, hat es seit den frühen Phasen der mathematischen Entwicklung des Wahrscheinlichkeitsbegriffs auch eine Auffassung gegeben, wonach Wahrscheinlichkeit genuin ein Urteil des denkenden Subjekts über eine Aussage ist (siehe auch Borovcnik & Kapadia 2014).

Zentrale Interpretation der subjektivistischen Theorie

Bayes hat 1763 ein zum Bernoullischen Gesetz inverses Gesetz der großen Zahlen – in einem durch ein konkretes Modell abgesteckten Rahmen – bewiesen. Danach konvergieren die *Einschätzungen* einer Wahrscheinlichkeit gegen die relative Häufigkeit aus (unabhängigen) Experimenten. Dabei wird Wahrscheinlichkeit als der persönliche Grad des Vertrauens in eine Aussage aufgefasst. Ein solcher Grad gründet sich auf vielfältige qualitative und quantitative Information:

- Relative Häufigkeiten relevanter Experimente in der Vergangenheit;
- Expertenwissen (beschleunigte Lebensdauerversuche), Metaanalysen;
- persönliche Erwartungen.

Angelpunkt dieser Auffassung von Wahrscheinlichkeit und für das inverse Gesetz der großen Zahlen ist die Bayes-Formel (siehe 2.3). In dieser wird eine (allenfalls subjektive) a priori-Wahrscheinlichkeit für eine Aussage durch Daten (aus wiederholten Experimenten) in eine neue a posteriori-Wahrscheinlichkeit umgerechnet.

Auch diese Konzeption von Wahrscheinlichkeit ist durch einen axiomatischen Aufbau gerechtfertigt. de Finetti (1937) formuliert entsprechende Axiome für Präferenzen. Dazu zählt auch die so genannte *Konsistenz* der Präferenzen. Danach muss man Wetten vermeiden, die einen sicheren Verlust bringen.

Bedingte Wahrscheinlichkeit als Schlüssel und Mühlstein

Der subjektive Charakter von (priori) Wahrscheinlichkeiten ist – teilweise – behebbar: Wiederholtes Updaten durch neue Information mit der Bayes-Formel erlaubt eine neue Bewertung durch Wahrscheinlichkeiten, welche subjektives Vorwissen mit den harten Daten kombiniert. Hat man genügend Daten, so kann man den Einfluss subjektiven Inputs klein halten. Man kann auch untersuchen, wie sich (kleine oder größere) Änderungen der a priori-Wahrscheinlichkeiten auf die endgültige Entscheidung (oder Schätzung) auswirken (für ein einfaches Beispiel für einen Versicherungsvertrag siehe Borovcnik 2006).

Die Bayes-Formel (siehe weiter unten in 2.3) spielt eine zentrale Rolle in der subjektivistischen Theorie; sie ermöglicht ein wiederholten Updaten von Wahrscheinlichkeitsurteilen durch weitere Information: Eingangsurteile werden durch die Likelihood der neuen Information (i.e., der Daten) in neue Wahrscheinlichkeiten revidiert. Interessant ist, dass hier die Daten direkt eingebaut werden können, um eine Wahrscheinlichkeitsaussage von Aussagen (auch Hypothesen!) zu erhalten. Dagegen haben im objektivistischen Rahmen Hypothesen keine Wahrscheinlichkeiten – insbesondere weil diese sich einer empirischen Prüfung durch wiederholte Experimente entziehen.

Da sind wir auch schon im Kern der Kritik an der subjektivistischen Konzeption: In ihr ist Wahrscheinlichkeit genuin subjektiv, eine empirische Kontrolle fehlt. Dagegen wenden Subjektivisten nun ein: Mit genügend Daten sind die Schlüsse beider Schulen (fast) identisch. Der objektivistische Ansatz hat Objektivität und Rationalität keineswegs gepachtet, denn allein jede Anwendung von Mathematik ist schon subjektiv. Ferner ist die Rechtfertigung einer Deutung als relative Häufigkeit nicht besser als die axiomatische Rechtfertigung einer Deutung als Grad des (persönlichen) Vertrauens. Beide Ansätze genießen den Status einer axiomatisch fundierten Theorie. Überdies kann man die Bayes-Formel im Kolmogorow- oder im de Finetti-Ansatz ziemlich ähnlich formulieren; dieselben Worte und Zeichen haben aber natürlich eine ganz andere Interpretation. Insgesamt ist die Kontroverse in den Grundlagen der Stochastik eine Kontroverse über die Objektivität von Theorien – ein Streit also, der von vornherein unentscheidbar ist.

2.3 Odds und Bayes-Formel

Für die Messung des Grades des Vertrauens ist der *Vergleich* der Wahrscheinlichkeit einer Aussage und ihres Komplements hilfreich. Diese Chancenverhältnisse oder Odds bilden eine Basisinterpretation zur Kalibrierung einer persönlichen Wahrscheinlichkeit und lassen den Formalismus besser an die Situation des Revidierens der Wahrscheinlichkeit einer Aussage durch neue Daten anpassen.

Wahrscheinlichkeiten kann man auf zwei Arten ansprechen

Wahrscheinlichkeit kann man als Zahl zwischen 0 und 1 ansprechen; Kalibrieren auf einer Skala von 0 bis 1 war historisch ein großer Schritt. Wahrscheinlichkeit kann man auch als Verhältnis – Odds – ansehen: verglichen werden dabei die relativen Wahrscheinlichkeiten einer Aussage „einzutreten" *gegen* „auszubleiben". Es gibt eine wechselweise Beziehung zwischen *absoluten* und *relativen* Wahrscheinlichkeiten: $p = \frac{1}{6}$ führt zu Odds von $\frac{1}{6} : \frac{5}{6}$ oder $1:5$; allgemein führen Odds von $a:b$ zu $p = \frac{a}{a+b}$. Aus Odds liest man direkt die (fairen) Einsätze ab: Stakes (Netto-Wetteinsatzverhältnis) = Inverse Odds. Für den Umgang mit bedingten Wahrscheinlichkeiten sind Odds viel besser geeignet.

Bayes-Formel mit Odds und Vergleich mit der üblichen Darstellung

Seien H_1, H_2 ausschließende und erschöpfende Aussagen (Zerlegung des Raumes). Relativ zur (neuen) Evidenz E gilt:

$$\frac{P(H_1 \mid E)}{P(H_2 \mid E)} = \frac{P(H_1)}{P(H_2)} \times \frac{P(E \mid H_1)}{P(E \mid H_2)}$$

$$\text{Neue} - \text{posteriori} - \text{Odds} = \text{Alte} - \text{priori} - \text{Odds} \times \text{Likelihood} - \text{Verhältnis}$$

relativ zu empirischen Daten — ohne Daten — Macht der Daten, Zustände zu trennen

Zum Vergleich die übliche Notation: $P(H_1 \mid E) = \dfrac{P(H_1) \cdot P(E \mid H_1)}{P(H_1) \cdot P(E \mid H_1) + P(H_2) \cdot P(E \mid H_2)}$.

Diese zeigt weder den strukturellen Einfluss einzelner Faktoren noch die Richtung oder Größe der Änderung von priori- zu posteriori-Wahrscheinlichkeiten. Odds passen einfach besser zur *multiplikativen Struktur* des Problems: Wenn sich priori-Odds verdoppeln, so verdoppeln sich posteriori-Odds; wenn sich der Likelihoodquotient verdoppelt, so verdoppeln sich posteriori-Odds; etc. Die Likelihoodquotienten spiegeln die *Trennkraft der neuen Information* (Indiz) *E*. Je höher $\dfrac{P(E \mid H_1)}{P(E \mid H_2)}$, desto deutlicher zeigt *E* an, dass H_1 zutrifft. Die Größe der a posteriori-Wahrscheinlichkeit speist sich aus zwei ganz unterschiedlichen Quellen: die (allenfalls subjektive) a priori-Einschätzung *und* die Daten.

2.4 Das philosophische Dilemma

Für das objektivistische Paradigma ist Wahrscheinlichkeit strikt eine Eigenschaft von Objekten; Modellieren realer Probleme wird durch Häufigkeiten (und Tests) reguliert. Die Wahrscheinlichkeitstheorie kann auf den Axiomen aufgebaut in sich geschlossen dargestellt werden. Für den Weiterbau einer geschlossenen Theorie unter Einschluss auch des statistischen Testens braucht man aber zusätzliche Kriterien; wie man diese setzt, man erhält Widersprüche.

Das subjektivistische Paradigma verbindet mit Wahrscheinlichkeit ein Urteil einer Person, welche eine bestimmte Information hat. Auch diese Auffassung von Wahrscheinlichkeit ist durch eine axiomatische Theorie abgesichert. Auf ihrer Basis kann man die statistische Beurteilung von Hypothesen in die Wahrscheinlichkeitstheorie einbauen und zwar mit Hilfe der Bayes-Formel.

Das wissenschaftstheoretische Dilemma besteht nun darin, dass ein objektives Theoriengebäude für den statistischen Schluss versagt, auf der anderen Seite eine subjektivistische Konzeption subjektive Anteile hat, die als „nicht-wissenschaftlich“ abgetan werden. Stegmüller (1973) entscheidet sich an dieser Stelle für eine unbefriedigende Testtheorie, um eine Subjektivierung des Wahrscheinlichkeitsbegriffs zu vermeiden.

Die subjektivistische Position wird hierbei als subjektiv (!) kritisiert; daher wurde sie als Lösung für ein mathematisches Konzept, das auch für die Physik gelten soll, verworfen. Das hat die Theorie auf einen einseitig objektivistischen Begriff reduziert, mit all seinen Nachteilen: diese betreffen die Anwendungen und das Lernen der Begriffe. In den Anwendungen hat sich in der Zwischenzeit – nachdem man die Unlösbarkeit des Dilemmas im Grundlagenstreit mehr oder weniger anerkannt hat – eine gewisse „anything goes“-Haltung eingebürgert. Wenn man genügend Information hat, um a priori-Wahrscheinlichkeiten festzulegen, dann verwendet man sie; hat man zu wenig, versucht man, möglichst wenig informative a priori-Verteilungen zu verwenden, um ihren Einfluss gegenüber den Daten möglichst klein zu halten.

3 Aus mathematik-didaktischer Perspektive

Bedingte Wahrscheinlichkeit und Bayes-Formel sind Zwillinge, aber im objektivistischen Rahmen passen sie einfach nicht zusammen.

3.1 Philosophische Aspekte aus der Sicht der Mathematik

In bestimmten Anwendungen ist ein – zumindest formaler – Gebrauch Bayesianischer Methoden durchaus verbreitet. In mathematischen Darstellungen dominiert eindeutig die „Reduktion“ von Wahrscheinlichkeit auf etwas wie relative Häufigkeiten. Für die Unterweisung gibt es eigentlich nur eine primitive Häufigkeitsdeutung.

Ambivalenz der Mathematischen Statistiker zum Bayes-Ansatz

Je näher Statistiker zu bestimmten Anwendungen stehen, desto eher sind sie bereit, mindestens formal Bayesianische Methoden in ihr Werkzeug zu integrieren. Wenn es um philosophische Ansichten geht, so sehen mathematische Statistiker Wahrscheinlichkeit gerne als abstrakten Begriff, der eigentlich keiner Interpretation bedarf. Man beachte den Schwenk seit Kolmogorow; wenn sie dann eine Interpretation brauchen, greifen sie einfach auf eine – gelegentlich sehr einfache – Sicht als idealisierte relative Häufigkeit zurück.

Moore (1997) hat sich in einer angriffig geführten Diskussion über Bayes oder Nicht-Bayes in der Ausbildung an den Universitäten geäußert, dass sowieso ein frequentistischer Begriff (Wahrscheinlichkeit als idealisierte relative Häufigkeiten) für die Grundausbildung besser geeignet ist und dann, als inhaltliche Argumente fehlten, einfach festgestellt, dass Bayes-Methoden viel zu schwierig sind.

Ignorieren des Bayes-Ansatzes innerhalb der Didaktik

Der extensive Einsatz von Simulation in didaktischen Ansätzen „erklärt“ die Beziehung zwischen relativen Häufigkeiten und Wahrscheinlichkeit und „rechtfertigt“ die Konzeption von Wahrscheinlichkeit als idealisierte relative Häufigkeiten. Er verstärkt aber den Bias hin zur Häufigkeitsinterpretation und versagt dennoch bei bedingten Wahrscheinlichkeiten.

Carranza & Kuzniak (2008) studieren, wie sich die Vernachlässigung von Bayes-Sichtweisen in Curricula auswirkt. Sie erklären viele Verständnisschwierigkeiten im Unterricht dadurch, dass ein durch und durch objektivistischer Begriff aufgebaut wird, der sich dann bei der Lösung von Aufgaben (insbesondere von Aufgaben mit bedingten Wahrscheinlichkeiten als ungeeignet erweist. Die Beispiele erfordern einen weiter gefassten Wahrscheinlichkeitsbegriff, der subjektivistische Komponenten mit einschließt. Ein Katalog von dringlichen Aufgaben zur Wahrscheinlichkeit (ICME; siehe Borovcnik 2011) verweist nur in einem einzigen Punkt auf subjektive (!) Wahrscheinlichkeit:

- Prüfsteine eines Modelling-Zugangs und pseudo-konkrete Modelle;
- Probleme innerhalb eines frequentistischen Zugangs;
- Rolle subjektiver Wahrscheinlichkeit in frequentistischen Zugängen;
- Vorzüge von realen Anwendungen und artifiziellen Situationen;
- Eigenschaften des Stichprobenziehens (reale und Simulation); …

Die Schwierigkeit liegt u.a. auch darin, dass sich der Charakter der Aussagen in objektivistischer wie subjektivistischer Theorie unterscheidet. Die beiden Schulen sind auch einem jeweils anderen wissenschaftlichen Paradigma zuzuordnen. Ein „vermischter“ Ansatz kann daher auch nicht die Lösung sein. Vancsó (2009) versucht daher, beide Begriffe *parallel* zu entwickeln, sodass der Wechsel im Anwender vollzogen werden kann, wann welche Aspekte des Wahrscheinlichkeitsbegriffs dem Problem besser angepasst sind.

Für das Verständnis der Begriffe und Methoden kann man festhalten, dass der Bayesianische Ansatz dem, wie Leute denken, viel näher steht. Jede Anwendung ist genuin subjektiv. Die frequentistische Interpretation ist eine bloße *façon de parler*, die natürlich in speziellen Fällen sehr hilfreich ist. Aber sie wird überstrapaziert, wenn man damit einen rein objektiven Begriff aufbauen will, der natürlich etwas von seinem objektiven Gehalt auch in jede Anwendung hinüber retten soll.

3.2 Schwierigkeiten mit bedingter Wahrscheinlichkeit

Eine Reihe von Einflussfaktoren spielt nicht nur bei der Bearbeitung von Aufgaben eine wesentliche Rolle, sie verringern auch häufig die Akzeptanz der gelernten Begriffe. Bessere Mathematikkenntnisse allein helfen offenbar auch nicht wirklich.

Einfluss der Sprache und die Akzeptanz der Mathematik

Empirische Studien wie Borovcnik & Bentz (1991) Borovcnik & Bentz (1993/2003) oder Díaz & Batanero (2009) zeigen eine Reihe von Einflussfaktoren auf.

- Kontext und Sprache spielen eine kritische Rolle.
- Überlappung mit kausalem Denken.
- Komplizierte Mathematik und intuitive Gedanken spielen zusammen.

Darüber hinaus wird die Bearbeitung von einschlägigen Aufgaben wesentlich beeinflusst durch

- kleinste Variationen in der Formulierung,
- das Format der Daten (absolute Zahlen, Proportionen, Prozent, Wahrscheinlichkeiten),
- die Variationen im Kontext der Aufgaben,
- den Einbezug der Person in den Kontext und die Lösung.

Diese beeinflussen, *wie* das Problem und die gegebenen Daten wahrgenommen werden, *welche Methode* zur Lösung verwendet wird, und, *wie und ob* das Ergebnis interpretiert und *akzeptiert* wird.

Ein Dilemma der Ausbildung

Carranza & Kuzniak (2008) sprechen von einem Dilemma in der Ausbildung: Kurse behandeln nur Wahrscheinlichkeit interpretiert als Häufigkeit, die Beispiele dagegen benötigen die Bayes-Formel und Vertrautheit mit der subjektivistischen Konzeption. Studien wie Díaz & Batanero (2009) zeigen, dass mathematisches Wissen kaum hilft; dass kein Transfer allgemeiner Kenntnisse stattfindet; und dass wir jedes Beispiel von neuem „erlernen".

4 Aus der Sicht rivalisierender Ideen

Leute verwenden ihre privaten Kriterien. Gigerenzer (2007) zeigt, wie sehr sich Leute auf Bauchentscheidungen verlassen und analysiert, wie eine bewusste Mischung von rationalen Argumenten und Bauchentscheidungen besser oder zumindest praktikabler ist als sich gänzlich auf rationale Vorstellungen zurückzuziehen. Als Gründe für Ideen, die dem rationalen Paradigma fern stehen, kann man u.a. anführen:

- Es gibt kaum eine Möglichkeit zu prüfen, ob eine Modellierung passt.
- Erfolg oder Misserfolg erfährt man erst in der Zukunft.
- Erfolg kann ebenso gut mit „falschen Vorstellungen" erreicht werden.
- Erfolgskontrolle im Nachhinein ist schlecht (war ja nur Pech).

4.1 Konolds „Outcome-Orientierung"

Man kann Leute kaum von der Effizienz probabilistischer Begriffe überzeugen. Konold (1989) stellt fest, dass sie dazu neigen, jede Wahrscheinlichkeit umzuformulieren in eine direkte – absolut sichere – Vorhersage des Ergebnisses (outcome).

4.2 Die kausale Alternative zu Zufall

Wann immer kausale Assoziationen hervorgerufen werden, kann es dazu kommen dass probabilistische Ansätze zu ihren Gunsten verworfen werden. Das Verhältnis von Zufall und Kausalität ist schillernd und begleitet die Wissenschaft seit frühen Zeiten. Während Laplace noch Wahrscheinlichkeit als Instrument des Unwissenden sieht (sein Dämon, der über allem steht und alles weiß, kann auf diesen Hilfsbegriff verzichten), stehen wir heute an der anderen Seite der vollen Skala: in der Physik versucht man, den Kausalitätsbegriff loszuwerden und alles durch Zufall zu erklären (etwa auch den Urknall).

Kausale und andere Einflüsse auf die Interpretation von bedingten Wahrscheinlichkeiten

Angenommen, die Aussage C wird kausal und E als Wirkung aufgefasst. Zwei Fallstricke sind verbreitet mit der Aussage $P(C \mid E)$:

- die Ursache ist unabhängig von der Wirkung: $P(C \mid E) := P(C)$,
- bedingte Wahrscheinlichkeiten werden umgedreht: $P(C \mid E) := P(E \mid C)$.

Eine eigenartige Mischung halb verstandener Begriffe und Intuitionen kann zu Umformulierungen der Aufgabe führen und zusätzlich zur Identifikation mit einer falschen Auffassung der Situation verleiten. Die Situation kausal zu re-konstruieren ist eine attraktive Alternative. Das hatten wir im Beispiel mit der Falk-Urne schon thematisiert.

Das gilt in gleicher Weise auch für den Kontext medizinischer Diagnose. Eine Krankheit D kann als Ursache für einen positiven medizinischen Befund aufgefasst werden (Tversky & Kahneman 1980): $P(+ \mid D) = 0{,}99$. Viele drehen diese Aussage einfach um als $P(D \mid +) = 0{,}99$; andere wiederum sprechen der umgekehrten bedingten Wahrscheinlichkeit überhaupt jede Relevanz ab, weil sie ja nur assoziativ-diagnostisch aufgefasst werden kann und einer kausalen Deutung entbehrt.

Natürlich hängt der Wert von $P(D \mid +)$ zusätzlich von der Prävalenz ab, das ist die a priori-Wahrscheinlichkeit von D, umgemünzt als Anteil der Bevölkerung, der an D leidet. Diese Prävalenz ist jedoch trotz der üblichen Häufigkeitsinterpretation genuin subjektiv. Sie hängt von der relevanten Untergruppe ab, der eine Person angehört. Dementsprechend ist dann die a posteriori-Wahrscheinlichkeit, die Krankheit D nach einem positiven Befund zu haben, auch unterschiedlich, je nachdem, ob die Person einer besonderen Risikogruppe angehört oder nicht.

Hierarchie zwischen rivalisierenden Ideen

Die Ideen scheinen empirischen Befunden nach hierarchisch geordnet (siehe Borovcnik & Peard 1996); probabilistische Argumente haben am wenigsten Überzeugungskraft und werden sofort beiseitegeschoben, wenn sich kausale oder logische Argumente ergeben. Logische Zusammenhänge stechen kausale Argument aus. Ganz besonders hoch stehen private Gründe, die man durchaus verstehen kann; sie führen aber die Problembearbeitung auf ganz andere Ebenen (die müssen ja nicht unangepasst sein, aber sie haben mit einem rationalen Ansatz oft wenig gemein).

- **Private Gründe** (oft an starke Emotionen gebunden);
- logische Gründe;
- kausale Schemata;
- probabilistische Argumente.

Bedingte Wahrscheinlichkeiten haben nichts mit Zeit und Ursachen zu tun: sie sind ein weiter gefasstes Konzept, um Information einzubeziehen.

4.3 Mischung von privaten und mathematischen Vorstellungen

Private Vorstellungen führen zu großem Vertrauen in falsche Ansätze (siehe auch Lysø 2008). Besonders schwierig wird es, wenn sich private und mathematische Vorstellungen mischen. Dann erhält die Vorgangsweise subjektiv einen hohen Stellenwert.

Monty Hall – mit einer guten Notation beginnen

Die Aufgabe war ja schon festgehalten. Wir führen folgende Notationen ein: *A*, *B*, *C* Türen; c_A meint den *Zustand*, dass das Auto hinter *A* ist. Kandidat wählt *A*; Moderator öffnet *C* und zeigt eine Ziege.

Schattierte Zellen sind „verboten". Der Moderator hat keine Präferenz. Die Lösung liest man aus Zeile *C* (die offene Tür): $P(c_A \,|\, \text{“}C\text{”}) = \frac{1/6}{1/2} = 1/3$.

Meine Wahl = *A*

		Zustände			
		c_A	c_B	c_C	
Moderator öffnet	**A**	0	0	0	0
	B	1/6	0	1/3	1/2
	C	1/6	1/3	0	1/2
		1/3	1/3	1/3	1

Abb. 5: Kontingenztafel zur Lösung, wenn der Kandidat Tür *A* wählt.

Was ist paradox an Monty Hall?

Eine wilde Diskussion entspann sich nach der Lösung von vos Savant (o.D.). Merkwürdig ist, dass sich auch einschlägig Gebildete, Akademiker, Mathematiker und Statistiker daran beteiligten und keine wirklich rühmliche Rolle einnahmen.

Private Rekonstruktionen – Emotionen und Verantwortung

Viele Leute scheinen die Idee zu verwerfen, dass das Angebot, die erste Wahl ändern zu dürfen, ohne Hintergedanken und ohne besonderen Anlass *immer* gemacht wird; sie werden durch solch ein Angebot völlig verwirrt und beginnen nachzudenken, *warum* es ihnen gemacht wird. Es mag für sie dann besser sein, in die Augen des Moderators zu sehen, um zu erforschen, was eigentlich passiert, als ein abstraktes Modell zu suchen.

Auf der anderen Seite konnte der Autor immer wieder beobachten, dass Leute stur bei ihrer ersten Wahl bleiben. Das könnte auch damit zu tun haben, dass die erste Wahl ja wirklich *reiner Zufall* ist und da trägt man keinerlei Verantwortung. Eine Änderung aber entspricht dann ihrer *eigenen Entscheidung* und die müssten sie dann selbst *verantworten*. Wenn die falsch ist, sind sie „schuld", dass sie den Preis „verloren" haben (Borovcnik 2011). Historisch wurden kritische Entscheidungen oft an den Zufall überantwortet (Borovcnik & Kapadia 2014); noch heute wird etwa im Fußball die Seitenwahl mit einer Münze entschieden.

Illusion, die ganze Information genutzt zu haben

Die Information über die geöffnete Tür mit der Ziege wird verwendet, um die eine Tür zu eliminieren und Gleichwahrscheinlichkeit auf die beiden verbleibenden Türen zu übertragen. Es ist schon verwirrend, dass diese Information jenseits der Logik, eine Tür auszuschließen, auch genützt werden kann; d.h., sie kann als Indiz zwischen den beiden verbleibenden Türen fungieren. Dadurch, dass man die Information logisch schon genutzt hat, kann man der Illusion unterliegen, sämtliche Information schon ausgeschöpft zu haben und erhält damit ein noch tieferes Vertrauen in die Richtigkeit des eigenen Vorgehens.

Eine weitere Störquelle

Wenn der Kandidat Tür *A* gewählt hat, wird c_A häufig mit „der Kandidat *hat* das Auto" verwechselt. Die Lösung verwendet aber *mögliche und nicht aktuelle Zustände* (Hypothesen) und Wahrscheinlichkeiten dafür.

5 Aus der Sicht von Lösungsstrategien

Viele Leute formulieren die Aufgaben nach *ihrer* eigenen Sichtweise um und bestehen darauf, dass nur ihre Sicht zulässig ist. Verschiedene Lerntypen bevorzugen jeweils ihre Methoden. Lernende scheinen auch ihre Strategien zu ändern.

5.1 Naive Ansätze

Diese Ansätze sind durch eine eigenartige Prägung von vagen Intuitionen bestimmt, die weit über die übliche Modellierung hinausweisen. Sie führen zu Rekonstruktionen der Aufgabe oder zum Abrufen von Strategien, welche in anderen Zusammenhängen durchaus ihren Platz haben mögen. Besonders auffällig ist, dass naive Strategien auch nach entsprechenden Lernerfahrungen bestehen bleiben und eine Parallelwelt konstituieren.

Private Rekonstruktionen

Das Angebot zu wechseln ist ungewöhnlich; der Kandidat mag einen Grund dahinter suchen; wohin das führt, ist völlig offen. Der Kandidat vermeidet es, Verantwortung zu übernehmen – die erste Wahl ist Glück (Zufall), ein Wechsel der Türen ist eine Entscheidung und mag falsch sein. Hier verbleibt wenig Platz für mathematische oder rationale Argumente.

Argumente mit Gleichwahrscheinlichkeit

Das folgende Argument zeichnet sich durch eine Überlappung von mathematisch-logischen Argumenten und falschen Einschätzungen aus:

- Logisch gesehen wird durch die Information des Moderators eine Tür ausgeschlossen.
- Die zwei verbleibenden Türen sind am Beginn gleich wahrscheinlich. Das führt zu 1/2 für die beiden verbleibenden Türen.

Das erzeugt die Illusion, dass alle verfügbare Information ausgenützt wird.

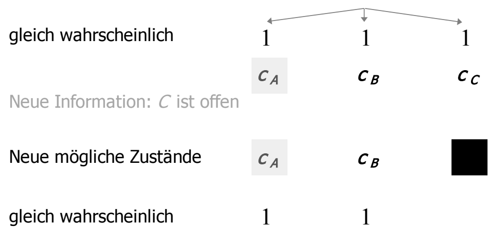


Abb. 6: Argumente auf der Basis von Gleichwahrscheinlichkeit (Kandidat wählt Tür *A*).

5.2 Probabilistische Strategien

In diesem Abschnitt werden korrekte Strategien zur Lösung des Monty Hall-Problems einander gegenübergestellt, die Wahrscheinlichkeitsargumente explizit aufnehmen.

Strategien, die auf der Bayes-Formel aufbauen

Alle Strategien hier fußen auf der Bayes-Formel, die durch unterschiedliche Darstellungsformen besser zugänglich gemacht werden soll.

a) Standardversion – mit Baum unterstützt

Ein Baumdiagramm hilft, aber hilft nur bedingt zur Aufgabe, den Lösungsprozess mit der Bayes-Formel zu veranschaulichen.

- Die Darstellung verschleiert die Struktur des Problems.
- Das Diagramm kann erst gezeichnet werden, wenn man die Wahl des Kandidaten kennt.
- c_A ist als *Möglichkeit* zu denken, die Wahrscheinlichkeit 1/3 hat.

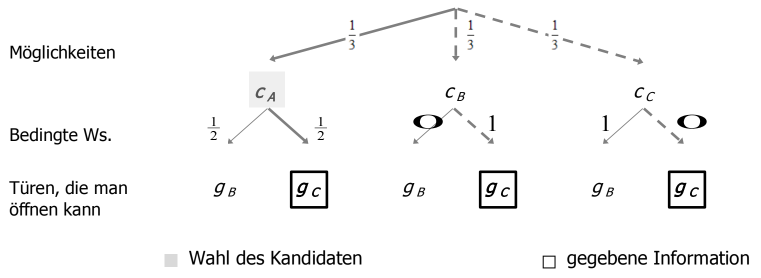


Abb. 7: Die übliche Bayes-Formel unterstützt durch ein Baumdiagramm (Kandidat wählt Tür *A*).

b) Bayes-Formel mit Odds

Die Darstellung zeigt klar die Funktion der zwei essentiellen Eingangsbausteine

- priori-Wahrscheinlichkeiten der Zustände (Hypothesen) von 1 : 1, und
- Likelihoodquotienten von 1/2 : 1 der Zustände unter der Information des Moderators

werden in posteriori-Odds von 1/2 : 1 oder 1 : 2 transformiert.

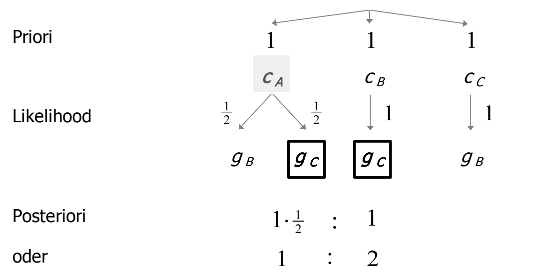


Abb. 8: Bayes-Formel mit Odds – durch ein Baumdiagramm unterstützt (Kandidat wählt Tür *A*).

c) Indikative Evidenz – jenseits der logischen Nutzung der Information

Tür *B* blieb zu und bietet eine *doppelte* Wahrscheinlichkeit für die bedingende Information im Vergleich zur Wahl des Kandidaten (Tür *A*); das verdoppelt ihre a posteriori-Odds auf 2 : 1 im Vergleich zu Tür *A*.

Diese Tür hat einen „Stress-Test“ bestanden – sie hätte ja geöffnet werden können; aber sie blieb verschlossen. Dies zeigt an, dass die Information des Moderators zu mehr taugt als nur zum Ausschließen einer Tür.

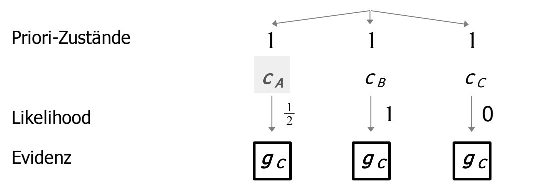


Abb. 9: Vergleich der Likelihoods zur Bewertung von Indizien, wenn der Kandidat Tür *A* wählt.

Strategien mit Kontingenztafeln

a) Tabelle mit Wahrscheinlichkeiten

Eine Kontingenztafel wird gemäß den Vorgaben des Beispiels aufgebaut. Die drei Zeilen stehen für die Möglichkeiten des Moderators, eine Tür zu öffnen; wir hatten zusätzlich unterstellt, dass der Kandidat Tür *A* wählt. Die Spalten stehen dann für die Zustände c_A, c_B, c_C und zeigen, wo sich das Auto befindet. Siehe Abb. 5 für das Ergebnis; gemäß den Regeln sind gewisse (schattierte) Zellen „verboten“. Die bedingte Zeilenverteilung, die der Wahl des Moderators entspricht, zeigt dann die neuen Chancen des Kandidaten: sie liegen unverändert bei 1/3.

b) Tabelle mit absoluten Zahlen: „Statistisches Dorf mit *N* Leuten“

Jetzt werden in die Kontingenztafel statt der Wahrscheinlichkeiten erwartete Anzahlen gemäß den Regeln eingetragen.

- Von N =300 Experiments „erwartet“ man 100 für jeden der Zustände.
- Von den 100 Fällen mit Zustand c_B, i.e., Auto hinter Tür *B*, muss der Moderator *immer* Tür *C* öffnen und die Ziege dort zeigen.

Meine Wahl = *A*

c_A ... Auto ist hinter Tür *A*

Moderator öffnet	Zustände c_A	c_B	c_C	
A	0	0	0	0
B	50	0	100	150
C	50	100	0	150
	100	100	100	300

Abb. 10: Kontingenztafel mit „Anzahlen“ anstelle der Wahrscheinlichkeiten, wenn der Kandidat Tür *A* wählt.

c) Methode mit dem Einheitsquadrat

Diese Methode veranschaulicht die Anteile in den Feldern der Kontingenztafel als (anteilige) Flächen im Einheitsquadrat. Wir identifizieren absolute oder relative Flächen im Einheitsquadrat:

- Horizontal: Streifen der Breite 1/3 entsprechend der Wahrscheinlichkeit, eine Tür zu wählen.
- Vertikal: Bedingte Wahrscheinlichkeit der Information (Evidenz; Ziege hinter Tür *C*) teilt die Streifen.
- Farbige Fläche entspricht der Information; 1/3 davon fällt auf c_A.

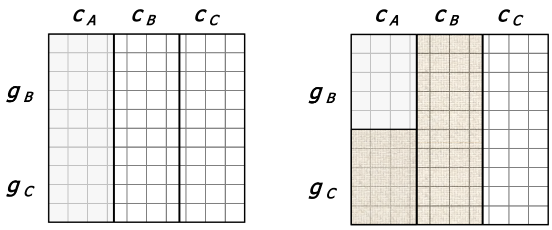


Abb. 11: Einheitsquadrat – zwei Stufen zur Aufteilung, wenn der Kandidat Tür *A* wählt. Streifen nach Auswahl des Kandidaten; dann Teilung der Streifen nach Information des Moderators.

5.3 Adäquate Strategien ohne Wahrscheinlichkeiten

Hier werden knapp einige Möglichkeiten illustriert, die ohne expliziten Gebrauch von Wahrscheinlichkeit auskommen.

Baumdiagramm mit absoluten Zahlen

Man zeichnet ein Baumdiagramm nach den Regeln des Spiels: Auf der 1. Stufe teilt man zwischen den 3 Zuständen; auf der 2. Stufe nach den Regeln. Anders als im gewöhnlichen Baumdiagramm werden aber nicht die Kanten mit den entsprechenden bedingten Wahrscheinlichkeiten beschriftet, es werden vielmehr an den Knoten natürliche Anzahlen „in Behälter“ geschrieben, die sich als verbleibende erwartete Anzahlen ergeben. Wenn das Auto hinter Tür *A* (welche der Kandidat gewählt hat) ist, dann wird zufällig eine der Türen (*B* oder *C*) geöffnet. Die „Töpfe“, welche der Information g_C entsprechen, haben insgesamt „150“; davon entfallen „50“ auf c_A; der Kandidat hat das Auto mit Wahrscheinlichkeit 50/150 = 1/3.

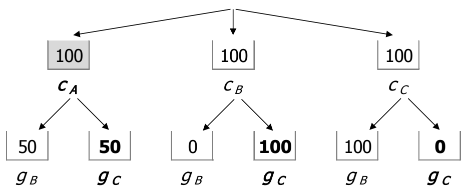
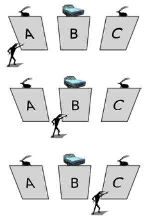


Abb. 12: Baumdiagramm mit absoluten Zahlen (erwarteten Häufigkeiten), wenn der Kandidat Tür *A* wählt.

Graphische Repräsentation der möglichen Fälle

Abb. 13 zeigt den Vorteil einer Wechselstrategie. Die Besetzung der Türen wird dabei festgehalten. Variiert wird nur die Wahl des Kandidaten; untersucht wird, was bei der Wechselstrategie passiert.



Moderator öffnet *C*
Wechsel der Türen gewinnt.

Moderator öffnet *A* oder *C*
Wechsel der Türen verliert.

Moderator öffnet *A*
Wechsel der Türen gewinnt. (Gigerenzer 2002).

Abb. 13: Schematische Zeichnung aller drei Fälle – Prüfung des Erfolgs der Wechselstrategie.

6 Schlussfolgerungen

Die Suche nach einer optimalen Methode des Unterrichts entspricht einer alten Vision, wird aber der Vielschichtigkeit des Begriffs „bedingter Wahrscheinlichkeit“ nicht gerecht. Probabilistische Strategien, die den Umgang mit Informationen und das Revidieren von Wahrscheinlichkeitsurteilen zum Kern haben, sind auf den ersten Blick schwieriger, bringen aber den hypothetischen Charakter von Wahrscheinlichkeit viel besser heraus. Umgehen mit bedingten Wahrscheinlichkeiten und Information zur Neubewertung von Aussagen ist genuiner Bestandteil *stochastischen Denkens* (Borovcnik 2006).

6.1 Suche nach einer optimalen Unterrichtsmethode

Gigerenzer verfolgt schon des Längeren ein Forschungsprogramm, das sich auch mit der Frage beschäftigt, welche Darstellungen effizienter sind im Hinblick auf ein Verstehen der Begriffe (Hoffrage et al 2002; Sedelmeier & Gigerenzer 2001). Danach „ist“ das Baumdiagramm mit absoluten Zahlen am besten. Es hilft, die richtige Lösung zu finden und zu akzeptieren, und, es vermeidet den Kalkül der Bayes-Formel. Sollte man daher bedingte Wahrscheinlichkeit nur mehr so unterrichten?

Jede der Methoden hat ihre relativen Vorzüge und Nachteile. Baumdiagramme zerstören die Symmetrie durch ihre Reihenfolge. Kontingenztafeln dagegen erlauben, Zeilen und Spalten zu vertauschen. Odds unterstützen den iterativen Prozess der Neubewertung. Als zusätzlich schwierig erweist sich etwa, dass Anzahlen die (bedingten) Wahrscheinlichkeiten ersetzen und damit den Charakter der Begriffe auf den Kopf stellen: sie wechseln von hypothetischen Aussagen hin zu faktischem Wissen. Das erleichtert nicht gerade einen rationalen Umgang mit (bedingten) Wahrscheinlichkeiten. Klarer Vorteil aber ist, dass man die Lösung unmittelbar – ohne weitere Zweifel – erhält.

6.2 Probabilistische Strategien für bedingte Wahrscheinlichkeiten

Es gibt durchaus spannende Strategien, bedingte Wahrscheinlichkeiten zu elementarisieren oder gar zu vermeiden (siehe oben). Allerdings geht das auf Kosten einer Vertrautheit mit (bedingten) Wahrscheinlichkeiten, die man später ja doch braucht. Wahrscheinlichkeiten und Häufigkeiten sind grundsätzlich verschieden. Wahrscheinlichkeiten durch Simulationen von wiederholten Experimenten zu einem Hilfskonzept zu reduzieren, verschiebt die Konnotation gänzlich in die objektivistische Ecke – mit den entsprechenden Nachteilen, wie sie Carranza & Kuzniak (2008) erörtern. Außerdem ist Wahrscheinlichkeit kein faktischer sondern ein hypothetischer Begriff. Auch die Elementarisierung mit Erwartungswerten (natürliche Häufigkeiten) verzerrt grundsätzliche Eigenschaften von Wahrscheinlichkeit, zumindest in der Perzeption der Lernenden. Natürliche Häufigkeiten als Anzahlen vermitteln die Illusion von (exakten) Fakten, die zudem bei kleinen Wahrscheinlichkeiten sehr unzuverlässig sind; siehe Borovcnik (2012).

Es bedarf daher didaktischer Ansätze, die jenseits von einer Vereinfachung (und Verzerrung) der Begriffe direkt auf den hypothetischen Charakter von Wahrscheinlichkeiten eingehen. Die auch in der Lage sind, den Umgang mit neuen Informationen und die Neubewertung von Aussagen in Form von bedingten Wahrscheinlichkeiten aufzugreifen. Als solche Ansätze sind zu nennen: Die *impliziten oder versteckten Lotterien* von Freudenthal (1973), die Bentz (1983) weiterentwickelt. Die Lotterie für den Moderator ist hierbei keinesfalls symmetrisch. Das Begünstigen-Konzept: Falk & Bar-Hillel (1983) thematisieren die Auswirkungen neuer Informationen direkt und erhalten damit qualitative Rückschlüsse über den Wert und die Richtung der Information. Unabhängig davon hat Borovcnik (1987) dieses Konzept ausgearbeitet und kann viele Trugschlüsse mit bedingten Wahrscheinlichkeiten erklären. In jüngerer Zeit hat Vancsó (2009) Begünstigen als Bestandteil eines offeneren Zugangs zu Wahrscheinlichkeit aufgenommen; er entwickelt parallel eine objektivistische und subjektivistische Sicht und lässt die Begriffe erst im Lernenden verschmelzen. Ein weiterer Zugang setzt auf Re-präsentation des Formalismus (Borovcnik & Peard 1996): Die Formeln mit Odds spiegeln die Ideen besser und geben über den Wert von Indizien Auskunft.

Die wiederkehrenden Puzzles mit bedingter Wahrscheinlichkeit zeigen, dass die Einschränkung auf eine Wahrscheinlichkeit als (Grenzwert) relative Häufigkeiten versagt, stabile Vorstellungen bei Lernenden aufzubauen. Warum sollte Wahrscheinlichkeit objektiv sein? Was macht eine objektivistische Konzeption objektiver? Abschließend sei auf zwei Zitate verwiesen: WAHRSCHEINLICHKEIT EXISTIERT NICHT (de Finetti, 1937). Wahrscheinlichkeit ist nur eine Form des Denkens über die Welt (Borovcnik, 2011). Und als Fazit: *Subjektivistische Deutungen machen den Ansatz erst sinnvoll.*

6.3 Epilog

In der Diskussion nach dem Vortrag hat sich ein Einwand aufgetan: „Die Didaktiker sollen doch nicht alles so verkomplizieren." Im Monty Hall-Beispiel kann man sich ja durch ein einfaches Argument orientieren. Wenn man immer bei der ersten Wahl bleibt, so hat man – klarerweise – eine Erfolgswahrscheinlichkeit von 1/3. Ergo muss man bei der Wechselstrategie eine Erfolgswahrscheinlichkeit von 2/3 haben. Die vielen Diagramme verwirren doch nur. Der Autor hat dann eine noch einfachere Sicht der Dinge ins Spiel gebracht: Der Kandidat wird gleich gefragt, ob er eine oder zwei Türen haben will. Klar, wird er sich für zwei Türen entscheiden. Beide Ansätze versagen, wenn die Information des Moderators abhängig von der getroffenen Wahl des Kandidaten gegeben wird. Also einmal nicht oder einmal doch (was zwar hier durch die Spielregeln ausgeschlossen ist, weil der Moderator das Angebot *immer* macht). Eine einfache Variante unter den gegebenen Spielregeln wäre: der Kandidat wählt eine außenstehende Tür. Der Moderator macht immer – so es möglich ist – die am weitesten entfernte Tür auf. In diesem Fall weiß der Kandidat (falls er diese Strategie des Moderators kennt oder errät) sicher, wo die Tür ist! Er hat dann die Erfolgswahrscheinlichkeit auf 1 erhöht und bleibt nicht wie der Wechsler auf 2/3 stehen.

Beide Ansätze gehen daher am wesentlichen der Situation vorbei: sie berücksichtigen in keinster Weise, dass neue Informationen (was uns der Moderator zeigt) die Wahrscheinlichkeit von Aussagen verändern oder gleich bleiben lassen. Mit bedingten Wahrscheinlichkeiten umgehen lernen heißt, flexibel auf Situationen eingehen. Dieses stochastische Denken gilt es auszubauen. Dies wird auch von Simulationen der Situation verpasst. Man wird wohl anerkennen müssen, dass die Wechselstrategie auf lange Sicht in 2/3 der Fälle Erfolg hat. Aber warum das so ist, erklärt sie ja nicht. Dieses Denken in Indizien ist genuin eine stochastische Tätigkeit und kann in vielen Situationen, auch des Alltags, mit Vorteil eingesetzt werden. Abschließend ein Zitat von David Spiegelhalter, einem führenden Spezialisten für Risiko, auf dem Klappentext eines kürzlich erschienenen Buches:

„I often get asked why people find probability so unintuitive and difficult. After years of research, I have concluded it's because probability really is unintuitive and difficult. This ground-breaking text acknowledges the full complexity of teaching this subject: the contributions face up to the competing interpretations of probability, emphasising the close connection to both human psychology and real-world problem-solving tasks. I am personally very pleased to see the subjective interpretation taken seriously, […]."

Literatur

Barnett, V. (1973): *Comparative statistical inference*. New York: Wiley.

Bentz, H.-J. (Ed.) (1983): Probleme im Umgang mit dem Zufall. *Der Mathematik-Unterricht*, *29*(1).

Borovcnik, M. (1987): Revising probabilities according to new information – A fundamental stochastic intuition. In R. Davidson, & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. Victoria: University of Victoria, 298-302. Online: www.stat.auckland.ac.nz/~iase/publications/icots2/Borovcnik.pdf (Zugriff: 23.11.2013).

Borovcnik, M. (2006): Probabilistic and statistical thinking. In M. Bosch (Ed.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education*. Barcelona: ERME, 484-506. Online: ermeweb.free.fr/CERME4/ (Zugriff: 23.11.2013).

Borovcnik, M. (2011): Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education: A joint ICMI/IASE Study*. New York: Springer, 71-83.

Borovcnik, M. (2012): Multiple Perspectives on the Concept of Conditional Probability. *Avances de Investigación en Didactica de la Mathemática*, 2, 5-27.

Borovcnik, M., & Bentz, H.-J. (1991): Empirical research in understanding probability. In R. Kapadia, & M. Borovcnik (Eds.), *Chance encounters*. Dordrecht: Kluwer, 73-105.

Borovcnik, M., & Bentz, H.-J. (1990/2003): *Intuitive Vorstellungen von Wahrscheinlichkeitskonzepten: Fragebögen und Tiefeninterviews* Technical Reports. Klagenfurt University.

Borovcnik, M., & Peard, R. (1996): Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International Handbook of Mathematics Education*. Dordrecht: Kluwer, 239-288.

Borovcnik M., & Kapadia, R. (2011): Modelling in probability and statistics – Key ideas and innovative examples. In J. Maaß, & J. O'Donoghue (Eds.), *Real-world problems for secondary school students – Case studies*. Rotterdam: Sense Publishers, 1-44.

Borovcnik, M., & Kapadia, R. (2014): A historical and philosophical perspective on probability. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education.* New York: Springer – to appear.

Carranza, P., & Kuzniak, A. (2008): Duality of probability and statistics teaching in french education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Proceedings of the Joint ICMI/IASE Study: Teaching Statistics in School Mathematics*. Granada: ICMI and IA-SE. Online: www.stat.auckland.ac.nz/~iase/publications/rt08/T1P2_Carranza.pdf (Zugriff: 23.11.2013).

de Finetti, B. (1937): La prévision: ses lois logiques, ses sources subjectives. *Annals of the Institute Henri Poincaré*, *7*, 1-68.

Díaz, C., & Batanero, C. (2009): University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, *4*(3), 131-162. Online: www.iejme.com/032009/main.htm (Zugriff: 23.11.2013).

Falk, R., & Bar-Hillel, M. (1983): Probabilistic dependence between events. *Two-Year-College-Mathematics Journal*, *14,* 240–247.

Falk, R., & Konold, C. (1992): The psychology of learning probability. In F. Sheldon & G. Sheldon (Eds.), *Statistics for the Twenty-First Century*, MAA Notes 26. Washington D. C.: The Mathematical Association of America, 151-164.

Freudenthal, H. (1973): *Mathematics as an educational task*. Dordrecht: Reidel.

Gigerenzer, G. (2002): *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2007): *Gut feelings: The intelligence of the unconscious*. New York: Viking.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002): Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition 84*, 343–352.

Kolmogorow, A.N (1933/1956): *Foundations of the theory of probability*. London: Chelsea.

Konold, C. (1989): Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59-98.

Lysø, K. O. (2008): Strengths and limitations of informal conceptions in introductory probability courses for future lower secondary teachers. In M. Borovcnik, D. Pratt, Y. Wu, & C. Batanero (Eds.), *Research and development in the teaching and learning of probability*. Monterrey: ICME 11. Online: www.stat.auckland.ac.nz/~iase/publications.php?show=icme11 (Zugriff: 23.11.2013).

Moore D. S. (1997): Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*(3), 254-261.

Sedelmeier, P., & Gigerenzer, G. (2001): Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General 130,* 380-400.

Stegmüller, W. (1973): *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, vol.4, 1. Teil: Personelle Wahrscheinlichkeit und Rationale Entscheidung, 2. Teil: Personelle und statistische Wahrscheinlichkeit. Berlin-New York: Springer.

Tversky, A., & Kahneman, D. (1980): Causal schemas in judgment under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, N.J.: L. Erlbaum.

Vancsó, Ö. (2009): Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, 4(3), 291-322. Online: www.iejme.com/032009/main.htm (Zugriff: 23.11.2013).

Winkler, R. (2012): Stochastik – ein Fest der Unabhängigkeit. *Schriftenreihe zur Didaktik der Mathematik der Österreichischen Mathematischen Gesellschaft (ÖMG)*, 44, 122-136.

vos Savant, M. (n.d.): *Game show problem*. Online: www.marilynvossavant.com/articles/gameshow.html (Zugriff: 23.11.2013)

Verfasser

Manfred Borovcnik
Alpen-Adria-Universität Klagenfurt, Institut für Mathematik, Universitätsstraße 65, 9020 Klagenfurt
manfred.borovcnik@uni-klu.ac.at