Logik II

Vorlesung, Universität Bielefeld, Sommersemester 2015 (v.19.3)

Cornelis Menke

Inhalt

1. Fehlschlüsse

1.1. Wiederholung: Grundbegriffe der Logik

Die **Logik** (engl. **logic**) untersucht (nach modernem Verständnis) die Gültigkeit von Argumenten.

Ein **Argument** (Schluß; engl. **argument**) ist eine Reihe von Aussagen, den **Prämissen** (engl. **premises**, Brit. Engl. auch: premisses), die eine andere Aussage, die **Konklusion** (engl. **conclusion**), stützen sollen.

Aussagen (Propositionen, wahrheitsdefinite Sätze; engl. **propositions**) sind **wahr** oder **falsch** (nicht aber gültig oder ungültig).

Argumente (Schlüsse, logische Folgerungen) sind **gültig** (engl. **valid**) oder **ungültig** (engl. **invalid**) (nicht aber wahr oder falsch).

Ein ungültiges Argument kann eine wahre Konklusion haben. Ein gültiges Argument kann eine falsche Konklusion haben.

Gültige Argumente sind **wahrheitserhaltend** (engl. **truth preserving**). Alternative Formulierungen: die Konklusion folgt (logisch) aus den Prämissen; wenn die Prämissen wahr sind, muß auch die Konklusion wahr sein; das Argument ist risiko-frei.

Gültigkeit (engl. **validity**) ist eine Beziehung zwischen Prämissen und Konklusion.

Ein gültiges Argument, dessen Prämissen wahr sind, wird **schlüssig** (engl. **sound**) genannt. – Entsprechend läßt sich die Schlüssigkeit von Argumente auf genau zwei Weisen bestreiten: indem man zeigt, daß wenigstens eine der Prämissen falsch ist, oder, indem man zeigt, daß das Argument ungültig ist. Nur das zweite ist Gegenstand der Logik.[1]

[1] Anmerkung: Bitte beachten Sie die Formulierung: "… läßt sich die *Schlüssigkeit* von Argumenten auf genau zwei Weisen bestreiten": Es gibt, wie wir sehen werden, durchaus andere Arten, ein Argument zu kritisieren/anzugreifen.

Ein **Konditional** (engl. **conditional**) ist eine Aussage der Form "Wenn A, dann B". Dabei nennt man A **Andezedens** (engl. **antecedent**), B **Konsequens** (engl. **consequent**) des Konditionals, und schreibt: $A \supset B$.

Als **modus (ponendo) ponens** bezeichnet man Argumente mit der logischen Form:

A, $A \supset B$, also: B.

Als **modus (tollendo) tollens** bezeichnet man Argumente mit der logischen Form:

$\neg A$, $B \supset A$, also: $\neg B$.

Als **Syllogismen** werden gültige Argumente bezeichnet mit (genau) zwei Prämissen und einer Konklusion, die jeweils eine der folgenden Formen haben: "Alle S sind P", "Einige S sind P", "Alle S sind $\neg P$" oder "Einige S sind $\neg P$". Ein Standard-Beispiel ist der sog. **Modus Barbara**[2]:

Alle S sind M.	Obersatz oder praemissa major
Alle M sind P.	Untersatz oder praemissa minor
Alle S sind P.	Konklusion oder conclusio

1.2. Fehlschlüsse

Der Ausdruck "**Fehlschluß**" (engl. **fallacy**) ist mehrdeutig und bezeichnet:

(i) alle nicht-gültigen Schlüsse, besonders aber solche, die (aufgrund der Argumentform) möglicherweise leicht für gültig gehalten werden können (sog. **formaler Fehlschluß** oder **non sequitur**);

(ii) andere typische oder häufige Argumentationsfehler, ohne daß diese notwendig einen Anspruch auf Gültigkeit erheben (sog. nicht-formale Fehlschlüsse; eine Reihe nicht-formaler Fehlschlüsse werden wir in dieser Vorlesung behandeln);

(iii) auch bestimmte (und gerade) *gültige* Schlüsse, nämlich zirkuläre Schlüsse (die behandeln wir in Kürze). – Wenn Sie die Bedeutungen (i), (ii) und (iii) zusammen betrachten, erkennen Sie, daß "Fehlschluß" ein problematischer Ausdruck ist.

Zur **Auswahl und Klassifikation**: Ein zwingendes Kriterium der Klassifikation von Fehlschlüssen gibt es nicht, wie schon Augustus De Morgan in seiner *Logic* feststellt: "There *is* no such thing as a classification of the ways in which men may arrive at an error: it is much to be doubted whether there ever *can* be."[3] Dennoch gibt es eine

[2] Die Bezeichnung ist ursprünglich eine Merkhilfe: Prämissen bzw. Konklusionen der Form "Alle S sind P" wurden mit dem Buchstaben **a** abgekürzt; der Modus B**a**rb**a**r**a** bezeichnet dann einen Syllogismus mit drei Aussagen dieser Form.

[3] Augustus De Morgan, *Formal Logic: or, The Calculus of Inference, Necessary and Probable*,

weitgehende Übereinstimmung über das ‘Material’, d. h. eine kanonische Sammlung von Fehlschlüssen.[4] Diese geht v.a. auf die Sammlung von Fehlschlüssen zurück, die sich in den *Sophistici Elenchi* (Περὶ τῶν Σοφιστικῶν Ἐλέγχων) Aristoteles’ findet, dem letzten Buch des sog. **Organon** (“Werkzeug”: der Sammlung von Aristoteles’ Schriften zur Logik im weiten Sinne). Aristoteles’ Fehlschluß-Katalog wurde meist mit kleinen Änderungen beibehalten.

Die meisten dieser kanonischen Fehlschlüsse werden Ihnen aber selten oder nie begegnen. Unser Kriterium, um zu entschieden, welche Fehlschlüsse für uns interessant sind, ist: Wir behandeln einige der Fehlschlüsse, die tatsächlich in der Philosophie oder den Wissenschaften vorkommen (in diesem Abschnitt), und dann weiter solche, die im Feld der Wahrscheinlichkeitstheorie liegen, einem weiteren Thema der Vorlesung (in späteren Abschnitten).[5]

1.3. Bejahung des Konsequens

Als Fehlschluß der **Bejahung des Konsequens** (engl. fallacy of **affirming the consequent**)[6] bezeichnet man einen Fehlschuß der Form:

$A \supset B,\ B$, also: A.

Dies ist ein Beispiel für einen formalen Fehlschluß.

> Beispiel: If Bacon wrote *Hamlet*, then Bacon was a great writer. Bacon was a great writer. Therefore, Bacon wrote *Hamlet*.[7]

Daß der Schluß ungültig ist, ist klar. Weniger klar ist, warum man ihm einen eigenen Namen gegeben hat, wie Hamblin bemerkt: “Every invalid inference-schema of the propositional calculus [...] could, in theory, be dignified with a special name and treated similarly, yet we do not hear of any others.”[8]

Ein möglicher Grund erschließt sich nicht von der Logik her, sondern aus der Bedeutung des Schluß-Schemas besonders in der Wissenschaft: Das Modell der sog. hypothetisch-deduktiven Bestätigung von Theorien scheint eben diese Argument-Form zu haben:

London: Taylor and Walton, 1847; 2nd. ed., London 1926, S. 276; zitiert nach Hamblin, S. 13.

[4] Hamblin, S. 13.

[5] Literatur zu Fehlschlüssen: Arthur Schopenhauer, *Eristische Dialektik oder Die Kunst Recht zu behalten*, 1830/1831. C. L. Hamblin, *Fallacies*, London 1970.

[6] Lat. “fallacia consequentis”, gr. “παρὰ τὸ ἑπόμενον”.

[7] Aus: Hamblin, S. 35.

[8] Hamblin, S. 37; Aristoteles führt ihn interessanterweise nicht unter den formalen Fehlschlüssen auf; vgl. ib., S. 36.

> Wenn Fresnels Theorie des Lichts richtig ist, dann findet sich im Zentrum des Schattens einer kleinen opaken Scheibe ein heller Fleck (der sog. Poissonsche Fleck), der gerade so hell ist, als wäre die Scheibe nicht da (gegeben freilich ein ‘geeigneter’ Aufbau des Experiments: kohärentes Licht usw.). Man findet experimentell den Poissonschen Fleck. *Also:* Fresnels Theorie des Lichts ist richtig.

Logisch ist dies sicher nicht gültig – aber ist dies ein schlechtes Argument? (Und wenn ja: Gäbe es bessere?)

1.4. Bestätigungsfehler (confirmation bias)

Der **Bestätigungsfehler** (engl. **confirmation bias**) besteht in der Neigung, eher nach Bestätigungen einmal akzeptierter Annahmen zu suchen als nach widersprechenden Befunden, und widersprechende Befunde auszublenden oder zu gering zu gewichten. Ein Sonderfall des Bestätigungsfehlers ist die **voreingenommene Statistik** (engl. **biased statistics**). Der Fehler findet sich schon bei Cicero beschrieben:

> Doch als einmal jener Diagoras, der Atheist genannt wird, nach Samothrake kam und ihm ein Freund sagte: ‘Du, der du meinst, daß die Götter die menschlichen Angelegenheiten vernachlässigen, siehst du nicht an so vielen Votivtafeln[9], wie viele Menschen dank einem Gelübde einem Sturme entronnen sind und heil in den Hafen gelangen konnten?’ – ‘Gewiss’, erwiderte jener, ‘es gibt eben keine Tafeln von denjenigen, die Schiffbruch erlitten haben und im Meere ertrunken sind.’ [10]

Im ***Novum Organum*** **Francis Bacons** [11] (1620) findet der Bestätigungsfehler sich als eine von vier Klassen von Idola und falschen Begriffen (“Idola et notiones falsae”, a 38 [12]). Der Ausdruck “Idola” (von gr. εἴδωλον) meint eigentlich ‘falsche Erscheinung’; Bacon verwendet den Ausdruck zur Bezeichnung von ‘Verzerrungen’, die den Fortschritt der Wissenschaften hindern. Den Bestätigungsfehler (a 41, 45–52) diskutiert er unter den **Idola Tribus** (Idole des Stammes: der menschlichen Natur eigene Hinderungsgründe):

> Der menschliche Geist setzt [...] in den Dingen eine größere Ordnung und Gleichförmigkeit voraus, als er darin findet. [...] Der menschliche Verstand zieht in das, was einmal sein Wohlgefallen erregt hat – sei es, weil es so überliefert und geglaubt worden ist, sei es,

[9] Votivgaben sind Gaben, die in einem Heiligtum aufgrund eines Gelübdes für eine Rettung aus einer Notlage (Seenot, Krankheit) dargebracht werden.

[10] Marcus Tullius Cicero, *De natura deorum*, III, 89; zitiert nach: id., *Vom Wesen der Götter/De natura deorum* (übers. v. Olof Gigon u. Laila Straume-Zimmermann). Berlin: De Gruyter, 2011, S. 293; 295.

[11] Francis Bacon war Lord Chancellor of England unter James I; sein *Novum Organum* – geplant als Teil eines größeren Werks, der *Instauratio Magna*, der ‘Großen Erneuerung’ [der Wissenschaften] – lehnt sich mit dem Titel an Aristoteles’ *Organon* an. Es gibt 6 überlappende Behandlungen der Idola in den Schriften Bacons (vgl. Hamblin, S. 143–148) – wir beschränken uns hier auf die Behandlung im *Novum Organum*.

[12] Das *Novum Organum* wird hier nach Buch (a/b) und Paragraph zitiert nach der Ausgabe: Francis Bacon, *Neues Organon*, Teilband I (hrsg. v. Wolfgang Krohn). Hamburg: Meiner, 1990.

> weil es anziehend ist –, auch alles andere mit hinein, damit es jenes bestätige und mit ihm übereinstimme. Und wenn auch die Bedeutung und Anzahl der entgegengesetzten Fälle größer ist, so beachtet er sie nicht, oder verachtet sie, schafft sie durch Haarspalterei beiseite […].[13]

Bei Bacon findet sich eine Betrachtung von Fehlschlüssen, die nicht nur an der Logik orientiert ist, sondern auch kognitive Fehlerquellen (und auch sprachliche und soziale) einbezieht. Die kognitive Dimension ist inzwischen in der empirischen Psychologie gut untersucht: Etwa durch Experimente, die zeigen, daß viele Leute zu positiven Teststrategien neigen (etwa keine Kontrollgruppen bei Experimenten anlegen) oder sich nach Diskussionen eher an die die eigenen Position bestätigenden Beiträge erinnern als an Einwände.

Der Ausdruck 'confirmation bias' geht auf den englischen Denkpsychologen **Peter Wason** zurück. Dieser führte in den 1960er Jahren eine Reihe von Experimenten durch, die zeigen sollten, daß Menschen eher versuchten, eigene Hypothesen zu bestätigen als zu prüfen. Einer der Versuche, die er durchführte, ist der sog. **Wason Four Cards Test** (oder Wason Selection Task). Vor den Probanden, zumeist Studenten, lagen auf einem Tisch vier Karten mit Aufschriften, z. B.

$$\mathbf{E} - \mathbf{F} - \mathbf{7} - \mathbf{4}\,;$$

die Probanden sollten – durch Umdrehen einzelner Karten – die folgende Aussage prüfen:

> **Wenn auf einer Seite der Karte ein Vokal ist, dann ist auf der anderen Seite eine gerade Zahl.**

Die meisten Probanden drehten die Karte mit der Aufschrift "E" um; dann aber auch häufig die "4" (was die Behauptung des Versuchsleiters nicht prüfen kann), aber nur selten die "7". Nicht einmal 10 % der Probanden lösten den Test richtig.[14] Wason interpretierte dieses Ergebnis als Resultat eines confirmation bias. (Ob dies die einzige denkbare Interpretation ist, sei dahingestellt.[15])

Von dem "Bestätigungsfehler" als Fehlschluß spricht man nicht in dem selben Sinne wie bei dem der Bejahung des Konsequent: der Vorwurf ist nicht, der Schluß sei nicht gültig (das ist offenkundig, und der Anspruch wird nicht erhoben). Er ist ein *kognitiver* oder *psychologischer Fehlschluß*, der sich sowohl auf die Beurteilung deduktiver Schlüsse (wie beim Wason Four Cards Test) als auch auf induktive Schlüsse (wie im Fall der Votivtafeln) beziehen kann (zu induktiven Schlüssen s.u.).

[13] Ibid., a 45–46.

[14] Peter Wason, Self-Contradictions, in: P. Johnson-Laird, P. Wason, *Thinking: Readings in Cognitive Science*. Cambridge 1977.

[15] Denken Sie an: Drückt sich in der Bevorzugung der "4" zwingend eine Bevorzugung von Bestätigungen gegenüber Prüfungen aus? Oder sollte man nicht sagen, die "4" sei keine Bestätigung, sondern irrelevant? Oder vielleicht ist die Aufgabe auch einfach zu abstrakt für viele Leute?

2. Die Petitio principii

2.1. Die Petitio principii

Von einer **Petitio principii** oder kurz Petitio (engl. **begging the question**) spricht man, wenn bei einer Argumentation (s. u.) das Argumentationsziel (die Konklusion) – der strittige Punkt – in gewisser Weise in den Argumentationsgründen (Prämissen) einfach angenommen oder vorausgesetzt wird. Das englische "begging the question" ("to beg" im Sinne von "fragen", also wörtlich: 'nach der Frage fragen') ist eine Übersetzung des lateinischen Petitio principii; der Sinn der Formulierung erschließt sich aber nicht leicht und führt im Englischen zu Mehrdeutigkeiten. Mit "petitio principii" gemeint ist: 'Die Annahme/Voraussetzung der ursprünglichen (strittigen) Aussage.' [16] (Man spricht bisweilen auch von einem Zirkelschluß.[17])

Hier wird es jetzt interessant: Denn einerseits wird in einem gewissen Sinn **bei allen deduktiv gültigen Argumenten im Antezedens das Konsequens vorausgesetzt** – sonst wären diese ja nicht gültig. Andererseits können gültige Argumente – und in einem gewissen Sinne: gerade gültige Argumente – durchaus geeignet sein, jemanden aufgrund der Prämissen von der Konklusion zu überzeugen. Es stellen sich zwei Fragen: (i) Worin genau soll der Fehler bei einer Petitio principii liegen? (ii) Trifft dieser Fehler auf alle gültigen Schlüsse zu?

[16] Der Ausdruck geht zurück auf Aristoteles, *Analytica Priora* 64 b: "τὸ ἐν ἀρχῆι αἰτεῖσθαι", wörtlich: 'fragen nach dem, was am Anfang (τὸ ἐν ἀρχῆι) steht'. Die Deutung des Altphilologien Ernst Kapp: Mit "das, was am Anfang steht" bezeichnet Aristoteles die Konklusion in dialektischen Streitgespräch (denn von der gehe man, ganz wörtlich, aus)'; und "fragen" bedeutet "nach der Akzeptanz/Zustimmung zu einer Prämisse fragen" (Kapp, S. 14–15; vgl. Hamblin, S. 31–32).

[17] Lat. "circulus in probando" oder "circulus vitiosus". Teils unterscheidet man auch zwischen Petitio und Zirkelschluß und nennt nur solche Schlüsse zirkulär, in denen sich die Konklusion unter den Prämissen befindet (s. u.); aber der Sprachgebrauch schwankt.

2.2. Exkurs: Definition und Explikation

Unter **Definition** soll ein Vorschlag zur Festsetzung der Bedeutung/des Gebrauchs eines Ausdrucks verstanden werden.[18] Der zu bestimmende Ausdruck heißt **Definiendum**, der Bestimmungsvorschlag **Definiens**. Definitionen (in diesem Sinn) sind nicht wahr-oder-falsch, sondern Setzungen. Besonders deutlich ist dies bei sog. stipulativen Definitionen wie in der Mathematik ("Sei $\mathcal{P}(A)$ die Wahrscheinlichkeit von A.").

Eine Form der Definition ist die **Explikation** (engl. **explication**). Explikationen dienen der Präzisierung eines bestehenden Sprachgebrauchs. Der gegebene Ausdruck heißt **Explikandum**, der präzisere, der diesen ersetzen soll, das **Explikat**. Die Ausdrücke gehen auf Rudolf Carnap (1950) zurück, der vier Adäquatheitskriterien nennt: **Ähnlichkeit**, **Exaktheit**, **Fruchtbarkeit** und **Einfachheit** – in Carnaps Worten:

> 1. Das Explikat muß dem Explikandum so weit *ähnlich* sein, daß in den meisten Fällen, in denen bisher das Explikandum benutzt wurde, statt dessen das Explikat verwendet werden kann. [...] 2. Die Regeln für den Gebrauch des Explikats müssen in *exakter* Weise gegeben werden. [...] 3. Das Explikat soll *fruchtbar* sein, d. h. die Formulierung möglichst vieler genereller Aussagen gestatten. [...] 4. Das Explikat soll so *einfach* wie möglich sein.[19]

2.3. Explikationen von Petitio principii

Es gibt mehrere Ansätze zur Explikation von 'Petitio principii':

(i) Die Petitio principii als logischer Zirkelschluß

> Explikation I: Unter den Prämissen eines Arguments befindet sich eine, die identisch oder äquivalent mit der Konklusion ist (man könnte dies den '**logischen Explikationsansatz**' nennen).

Vorteile der logischen Explikation: **(i)** Die Explikation fällt in das Gebiet der Logik und läßt sich eindeutig anwenden (**Exaktheit**, **Einfachheit**). **(ii)** Sie paßt gut zu einem intuitiven (d. h. unreflektierten) Vorverständnis von 'Zirkelschluß' als einem Schluß, der sich 'im Kreis dreht' (**Ähnlichkeit?**).

Nachteile der logischen Explikation: **(i)** Identität ist eigentlich zu eng (**Ähnlichkeit!**) – es macht die Petitio trivial, denn dann wäre eine Petitio eigentlich einfach ein Argument mit überflüssigen Prämissen:

[18] Dies ist selbst eine Definition (im vorgeschlagenen Sinn); den Ausdruck "Definition" könnte man auch weiter fassen und z. B. sog. lexikalische (analytische) Definitionen einschließen, d. h. Bestimmungen des tatsächlichen Sprachgebrauchs.

[19] Rudolf Carnap, *Induktive Logik und Wahrscheinlichkeit.* Wien 1959, S. 12–18; hier: S. 15; Hervorhebungen im Original. Ursprünglich in: Rudolf Carnap, *Logical Foundations of Probability.* Chicago u. a. 1950, S. 3–8. Terminologie im Englischen: Das 'explicatum' solle '*similar to the explicandum*' sein, '*exact*', '*fruitful*' und 'as *simple* as possible' (hier: S. 7; Hervorhebungen im Original).

A, B, C; also: B.

Hier könnte man die Prämissen A, C und D einfach streichen (es bliebe: 'B, also: B') – das Argument ist nutzlos, aber gültig und eigentlich nicht einmal verwirrend. Dies wäre es vielleicht im Fall der Äquivalenz – aber nur dann, wenn diese nicht offenkundig ist (man müßte die Äquivalenz erst zeigen), und dann 'fehlt' im Argument zugleich etwas: es hätte (logisch) zu viele und zugleich (pragmatisch) zu wenige Prämissen. **(ii)** Es ist gar nicht so leicht, interessante Beispiele für diese Art von 'Argumenten' zu finden (**Ähnlichkeit**, **Fruchtbarkeit**).

(ii) Die Petitio principii als epistemischer Zirkelschluß

> Explikation II: Die Akzeptanz einer der Prämissen setzt *in irgendeiner Weise* die Akzeptanz der Konklusion voraus (man könnte die den '**epistemischen Explikationsansatz**' nennen).

Ein Variante dieser Explikation findet sich in Simon Blackburns *Oxford Dictionary of Philosophy*:

> The best definition is that an argument begs the question [= ist eine Petitio] if it contains a definite premise or move that would not be accepted by any reasonable person who is initially prone to deny the conclusion." [20]

Vorzüge: Das klingt erst einmal präzise. Und die Idee dieser Explikation scheint Sinn zu ergeben: Ein Argument ist dann zirkulär, wenn unter den Annahmen eine ist, die nur der akzeptiert, der auch die Konklusion akzeptiert. (Denn wenn er die Konklusion akzeptiert, ist das Argument überflüssig; wenn nicht, wäre es gültig, aber (in seiner Einschätzung) nicht schlüssig.)

Aber: Ist "premise *or move*" (meine Hervorhebung) eigentlich (wohl)bestimmt? (wir sind immer noch in der Logik, und "move" wurde, jedenfalls bislang, noch nicht definiert). Ist "prone" ('geneigt') präzise? Was soll "initially" bedeuten? (Ein logisches Argument ist eine Beziehung zwischen Propositionen – eine solche Beziehung hat keinen Anfang.) Man müßte auch bestimmen, was "reasonable persons" sind. Sind also die Ausdrücke im Explikans überhaupt **präzise**? Und ist die Formulierung "prone to deny" nicht **zu eng**, also **unähnlich/unfruchtbar**? Was ist mit Konklusionen, über die niemand eine (Vor-) Meinung hat? Weiterhin: Wer sind 'reasonable persons'? Wir könnten 'reasonable persons' definieren, etwa als diejenigen, die gültige Schlüsse erkennen. Dann wäre der Definitionsversuch aber möglicherweise eine **zirkuläre Definition**. (Um 'zirkuläres Schließen' zu definieren, beziehen wir uns auf 'reasonable persons' – aber sind dies nicht gerade diejenigen, denen Fehlschlüsse nicht unterlaufen?) Andererseits: Wir befassen uns mit Fehlschlüssen (und beschreiben, benennen und klassifizieren diese), weil Leuten – auch vernünftigen Leuten (?) – diese Fehl-

[20] Simon Blackburn, *The Oxford Dictionary of Philosophy*, Oxford 1994, s. v. begging the question.

schlüsse unterlaufen; sonst gäbe es (unter 'vernünftigen Leuten') ja gar kein Problem. Ist der Definitionsversuch also überhaupt **plausibel/fruchtbar**? Oder *entweder* unplausibel *oder* zirkulär? Schließlich: Ist das Explikat **einfach**?

Eine andere Variante des epistemischen Explikationsansatzes:

> Ein Argument ist zirkulär genau dann, wenn man nur feststellen kann, ob die Prämissen wahr sind, indem man zuvor feststellt, ob [daß] die Konklusion wahr ist.

Vorzüge: Diese Explikation klingt etwas weniger problematisch. Sie hat weniger unklare Ausdrücke im Explikat (der Begriff der 'Wahrheit' einer Aussage z. B. ist eingeführt, anders als der der 'Akzeptanz'). Sie ist auch **einfach** und macht die Kernidee der Explikation auch klarer: Eine Petitio wäre ein Argument mit einem *Schlüssigkeits*problem: das Argument ist offensichtlich gültig, aber ob es schlüssig ist (die Prämissen auch wahr sind), läßt sich nur sagen, wenn man den Wahrheitswert der Konklusion kennt; insofern ist es als Argument nutzlos. Die Explikation erfüllt auch in einem bestimmten Sinn die Adäquatheitskriterien **Ähnlichkeit** und **Fruchtbarkeit**, denn wenigstens einige gültige Argumente wären danach keine Petitio (man könnte also 'gute' und 'schlechte' gültige Schlüsse unterscheiden):

> Jeder, der volljährig ist, darf wählen.
> Emmy ist volljährig.
> ———————————————
> Emmy darf wählen.

(Hier liegt Pointe darin, daß eine Prämisse eine Aussage über ein Recht ist und damit epistemisch unabhängig von der zweiten Prämisse ist; analoges gilt für sog. praktische Syllogismen. Wir klammern diese Fälle im weiteren aus.)

Nachteile: Ist die Explikation **exakt**? Sie setzt voraus, daß es allgemeine Maßstäbe/ Standards dafür gibt, die Wahrheit einer Aussage zu beurteilen (wie soll man sie sonst anwenden?). Dies führt erstens aus der Logik heraus, und zweitens stimmt es (gegenwärtig) nicht: Im Gegenteil sind (gerade in der Philosophie) viele Kontroversen auch Kontroversen gerade über die Frage, wie sich die Wahrheit von Aussagen feststellen läßt. Die Explikation ist also **unvollständig**: Ob ein Argument eine Petitio ist, ist nicht allein abhängig vom Argument, sondern auch von weiteren Annahmen (über korrekte Wahrheitskriterien).

(iii) Die Petitio als Fehler der Argumentation

> Explikation III. Eine Argumentation – der Gebrauch eines Arguments – ist in einer Kontroverse zwischen zwei Parteien genau dann zirkulär, wenn unter den Prämissen des Arguments eine ist, deren Wahrheit die andere Partei nicht feststellen kann, ohne festzustellen, daß Konklusion wahr ist.

Diese Explikation sieht in der Petitio ist nicht eine Eigenschaft eines Arguments (als

Verhältnis zwischen Aussagen), sondern einer **Argumentation**, d. h. des Gebrauchs eines Arguments in einer Kontroverse. Und um festzustellen, ob eine Argumentation zirkulär ist, muß man wissen, worum sich die Kontroverse dreht und und zwischen wem sie stattfindet. (Und oft versteht man den Vorwurf, ein Argument sei eine Petitio, besser, wenn man diese nicht als Fehler des Arguments, sondern der Argumentation auffaßt.) In diesem Sinne bestimmt das *Oxford Companion to Philosophy* kurz und knapp:

> **beggin the question,** or ***Petitio principii.*** Literally, requesting what is sought, or at issue. So, requesting an opponent to grant what the opponent seeks a proof of.[21]

2.4. Sind alle gültigen Argumente eine Petitio principii?

Begreift man die Petitio als Fehler der Argumentation, fragt man danach, ob ein Argument *andere* überzeugt (von der Wahrheit der Konklusion, wegen der Wahrheit der Prämissen). Eine andere Frage ist, ob *man selbst* durch gültige Schlüsse zu einer Erkenntnis (der Konklusion) aufgrund der Prämissen kommen kann. Kann man durch gültige Schlüsse etwas lernen, was man nicht schon wußte?

Der locus classicus ist John Stuart Mills *System of Logic Ratiocinative and Inductive* von 1843. Statt über gültige Schlüsse spricht Mill über Syllogismen. Sein Punkt ist: Man kann im Syllogismus die Prämissen nur wissen, wenn man die Konklusion schon weiß.

> We have now to inquire, whether the syllogistic process, that of reasoning from generals to particulars, is, or is not, a process of inference; a progress from the known to the unknown; a means of coming to a knowledge of something which we did not know before. [...] It must be granted that in every syllogism, considered as an argument to prove the conclusion, there is a *Petitio principii.* When we say,
>
> All men are mortal, Socrates is a man, therefore Socrates is mortal;
>
> it is unanswerably urged by the adversaries of the syllogistic theory, that the proposition, Socrates is mortal, is presupposed in the more general assumption, All men are mortal: that we cannot be assured of the mortality of all men, unless we were previously certain of the mortality of every individual man.[22]

(Mills weitere Überlegung: Die Aussage "The Duke of Wellington is mortal" sei (damals noch eine Vermutung und also) erschlossen; aber der *Schluß* bestehe nicht in dem Übergang von "All men are mortal" zu "The Duke of Wellington is mortal" (modern:

[21] Ted Honderich (ed.), *The Oxford Companion to Philosophy*, Oxford 1995, s. v. begging the question.

[22] John Stuart Mill: *System of Logic Ratiocinative and Inductive: Being A Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, II iii (= *Collected Works*, Vol. 7, 1974, S. 183–184).

im Verhältnis zwischen diesen Aussagen und "The Duke of Wellington is a man"), sondern im Übergang von den Aussagen: John ist gestorben, Thomas ist gestorben usw. zu "All men are mortal".[23] Mit anderen Worten: Wenn die Logik die Lehre von den Schlüssen ist, ist sie nicht die Lehre von den *gültigen* Schlüssen, weil es keine gültigen *Schlüsse* gibt.[24])

Wenn Mill recht hat, scheint es, als wäre ein gültiges Argument tatsächlich (von Ausnahmen abgesehen, s.o.) eine Petitio, fehlschlüssig und nutzlos. **Sind gültige Argumente und 'gute' Argumente nicht dasselbe?** – Wir unterscheiden zwei Fragen:

(i) Lernt man durch gültige Argumente gar nichts?

Daß man durch gültige Argumente nichts Neues lernt, stimmt bei einfachen Argumenten wie Syllogismen vielleicht sogar:

All men must die.
All kings are men.
―――――――――
All kings must die.

Aber sobald die Argumente komplexer werden, gilt dies nicht mehr: [25]

Definition $\cos(\alpha)$	… bekannt: Ankathete/Hypothenuse
Definition $\sin(\alpha)$	… bekannt: Gegenkathete/Hypothenuse
Satz von Pythagoras	… bekannt: $a^2 + b^2 = c^2$
$\cos^2(\alpha) + \sin^2(\alpha) = 1$	… könnte Ihnen neu sein [26]

Man lernt hier freilich nur, was man implizit schon wußte – keinen neuen 'Gehalt'. (Dieser Nutzen ist in der Mathematik ungleich größer als in der Logik.)

(ii) Was nützen einem gültige Argumente für die Begründung von Aussagen?

Mills Kritik hat eine weitere Pointe: "[…] that we cannot be assured of the mortality of all men, unless we were previously certain of the mortality of every individual man." Hier geht es um den Nutzen gültiger Schlüsse für die *Begründung* von Aussagen.

[23] Ibid., S. 186–187.
[24] Zu Mills Gedankengang siehe Geoffrey Scarre, *Logic and Reality in the Philosophy of John Stuart Mill*. Dordrecht u. a. 1989, chap. II, bes. S. 46 ff. Der Einwand hat eine lange Tradition (René Descartes; Dougald Steward; George Campbell; Richard Whately u.a.) und Nachleben (Alexander Bain) – um nur einige zu nennen.
[25] Ein Punkt, den im Kern Richard Whately (*Elements of Logic*, London [8] 1844, S. 239 ff.) gegen Lockes Zweifel an der Nützlichkeit der Logik einwendet; vgl. Scarry 1989, bes. S. 21.
[26] Zeichnen Sie ein rechtwinkliges Dreieck $\triangle(a, b, c)$ mit den Seiten a, b und c. Die Länge der Hypothenuse sei $c = 1$ (der Einfachheit halber). Dann einfach α einzeichnen, Definitionen anwenden, Pythagoras anwenden.

(a) Man bekommt tatsächlich ein Problem, wenn man *Begründungen* und *gültige Argumente* für ein- und dasselbe hält. Denn dann hört man mit dem Begründen entweder irgendwann auf (da fehlt dann eine Begründung), oder man macht immer weiter (dann wird man nie fertig), oder man dreht sich im Kreise (sog. **Münchhausen-Trilemma** oder **Agrippas Trilemma**[27]).

(b) Aber die Logik lehrt überhaupt nicht, daß man gültige Argumente stets 'von vorne nach hinten' lesen muß – daß die Prämissen einen von der Konklusion überzeugen sollen. Man sagt von gültigen Argumenten, diese seien **wahrheitserhaltend**, und zu Recht:

Bacon wrote *Hamlet*.	Wenn die Prämissen wahr sind …
If Bacon wrote *Hamlet*, then Bacon was a great writer.	↓
Bacon was a great writer.	… dann ist die Konklusion wahr

Aber Gültigkeit ist ein Verhältnis zwischen Aussagen, und man kann gültige Argumente durchaus auch rückwärts lesen:

Bacon wrote *Hamlet*.	… dann ist auch mind. eine der Prämissen falsch.
If Bacon wrote *Hamlet*, then Bacon was a great writer.	↑
Bacon was a great writer.	Wenn die Konklusion falsch ist …

Liest man gültige Argumente rückwärts, sind sie **'falschheitserhaltend'** – dies kann sehr nützlich sein! (Ins Unreine gesprochen, könnte man sagen: Wenn man etwas weiß (daß die Konklusion falsch ist), kann man schließen, daß man etwas anderes nicht weiß!)

(c) Aber nichts zwingt einen, Begründungen und gültige Argumente für ein- und dasselbe zu halten. Das meiste, was wir zu wissen glauben, ist *nicht* durch gültige Argumente begründet, sondern von anderen gelernt (sog. **testimoniales Wissen**: fast das gesamte Wissen) oder **induktiv gestützt/bestätigt** (alle Wissenschaften) oder beides. Vielfach sind allgemeine Aussagen viel besser gestützt als die Spezialfälle, die aus ihnen folgen. (denken Sie an wissenschaftliche Theorien).

Zusammenfassung: Die Logik, als Lehre von der Gültigkeit von Argumenten, ist nicht die Lehre der Begründung von Aussagen. Gültige Argumente sind nicht dasselbe wie 'gute' Argumente. Ob sie wenigstens *eine* (sinnvolle) Art der Begründung sind neben anderen, ist strittig;[28] indirekt nützlich sind sie sicher.

[27] Der Ausdruck 'Münchhausen-Trilemma' stammt von Hans Albert (*Traktat über kritische Vernunft*, Tübingen 1968, S. 11); er spielt auf den Versuch des Barons von Münchhausen an, sich an den eigenen Haaren aus einem Sumpf zu ziehen; die drei Möglichkeiten nennt Albert Dogmatismus, infiniten Regress und Psychologismus. Der Ausdruck 'Agrippas Trilemma' bezieht sich auf dem Skeptiker Agrippa (dem der Doxograph Diogenes Laertios (3. Jh.) 5 sog. Tropen zuschreibt, unter diesen die 3 genannten).

[28] Vgl. Sie z. B. bei Ernst Tugendhat, Ursula Wolf, *Logisch-semantische Propädeutik* (Stuttgart: Reclam, 1983, S. 13–15), die die formale Logik (anders als Mill) als eine Form der Wahrheitsbegrün-

3. Gehaltserweiternde Schlüsse

3.1. Das Duhem'sche Problem

Die Kehrseite davon, daß gültige Argumente wahrheitserhaltend sind, war, daß sie (rückwärts gelesen) 'falschheitserhaltend' sind (vgl. den modus tollens). Auch wenn also (deduktiv) gültige Argumente für die Begründung von Aussagen nicht dienlich sind: Erlauben Sie es wenigstens zu erkennen, welche Aussagen *nicht* zutreffen? Genügt insbesonere etwa eine falsche Implikation (Vorhersage) einer Theorie, um zu erkennen, daß die Theorie falsch ist?

Nein. Der Grund findet sich im 10. Kapitel von Pierre Duhems *Ziel und Struktur der physikalischen Theorien*[29]; es behandelt "Die physikalische Theorie und das Experiment". Duhem führt dort aus: Aus einer isolierten Hypothese allein folge nichts, sondern nur aus einem *System* (einer Menge) von Hypothesen. Daher seien gescheiterte Vorhersagen kein Beweis, daß die geprüfte Hypothese falsch sei: "All dies zusammengefaßt ergibt sich, daß der Physiker niemals eine isolierte Hypothese, sondern immer nur eine ganze Gruppe von Hypothesen der Kontrolle des Experiments unterwerfen kann. Wenn das Experiment mit seinen Voraussagen in Widerspruch steht, lehrt es ihn, daß wenigstens eine der Hypothesen, die diese Gruppe bilden, unzulässig ist und modifiziert werden muß."[30] Nur wisse man eben nicht, welche: Ein Uhrmacher könne bei einer kaputten Uhr die Uhr zerlegen und die Teile einzeln prüfen, der Physiker hingegen sei in der Lage eines Arztes, der einen kranken Patienten nicht aufschneiden kann.

dung auffassen.

[29] Pierre M. M. Duhem: *La Théorie Physique: Son Objet, Sa Structure*, Paris 1906. Zitiert nach: id., *Ziel und Struktur der physikalischen Theorien*, Hamburg: Meiner [1908] 1998.

[30] Ibid., S. 248.

Daß man einen Fehler allein logisch nicht isolieren kann, nennt man das **Duhem'sche Problem** (**Duhem's problem**)[31]:

$$\begin{array}{l} \text{Theorie} \wedge \text{Hilfsannahmen} \wedge (\ldots) \supset \text{Vorhersage} \\ \neg\,\text{Vorhersage} \\ \hline \neg\,\text{Theorie} \vee \neg\,\text{Hilfsannahmen} \vee \neg\,(\ldots) \end{array}$$

Man spricht von der **Unterbestimmtheit** (**underdetermination**) des Systems von Theorie und Hilfsannahmen durch Logik und Erfahrung (die Vorhersage). Das Duhem'sche Problem spielt in vielen Bereichen bes. der theoretischen Philosophie (Sprachphilosophie, Wissenschaftstheorie usw.) eine wichtige Rolle; es begegnet uns später noch einmal in anderer Form (bei der Logik der Signifikanztests).

3.2. Gehaltserweiternde Argumente

Für gültige Argumente spricht, daß sie **wahrheitserhaltend** (**truth preserving**) sind: Sind die Prämissen wahr, so ist es auch die Konklusion. Die Kehrseite davon, daß gültige Argumente wahrheitserhaltend sind, ist, daß sie nicht gehaltserweiternd sind: sie sagen einem 'über die Sache' nichts Neues.

Als **gehaltserweiternd** (**ampliative**[32]) werden wir Argumente bezeichnen, bei denen man 'inhaltlich' etwas lernt. Die Kehrseite ist, daß sie nicht wahrheitserhaltend sind: Es ist möglich, daß alle Prämissen wahr, die Konklusion solcher Argumente aber falsch ist. – Statt von gehaltserweiternden Argumenten spricht man oft auch von induktiven (i.G.z. deduktiven), wahrscheinlichen (i.G.z. notwendigen/zwingenden) oder riskanten (i.G.z. sicheren) Argumenten.[33]

In diesen verschiedenen Benennungen spiegeln sich die beiden **Ziele** wider, die mit nicht-deduktiven Argumenten verfolgt werden: **Gehaltserweiterung** und **Verläßlichkeit** (in Charles Peirce' Terminologie: *uberty* und *security*[34]). Zwischen diesen

[31] Man spricht auch von der 'Duhem-Quine-These', und unter dieser Benennung findet man in der Literatur auch andere Formulierungen des Problems: Es gebe stets die Möglichkeit, eine Theorie bei nicht-zutreffenden Vorhersagen durch Modifikation anderer Annahmen zu 'retten'; es gebe stets mehrere Theorien, die die selben Vorhersagen machten bzw. die Phänomene (gleich gut) 'erklärten'.

[32] Der Ausdruck 'ampliative' geht auf den Pragmatisten Charles Peirce zurück (*A Theory of Probable Inference* (= *Studies in Logic. By Members of the Johns Hopkins University*, Vol. 1), Boston 1883, S. 143 = id., *Writings*, Vol. 4 (ed. W. Kloesel), Bloomington/Indianapolis 1989, S. 420). [**CSP bezieht sich wiederum vermutlich auf Kant.**]

[33] Der klassische Ausdruck wäre 'induktiv' (s.u.). Dieser Ausdruck bezeichnet aber erstens auch besonderes Schlüsse 'vom Besonderen auf Allgemeines' und zweitens auch besonders enumerative Induktionen (s.u.). Nicht allen Schlüssen wiederum lassen sich Wahrscheinlichkeiten in einem klaren Sinn zuordnen, wenn man dies überhaupt für sinnvoll hält (s.u.). 'Risiko' wiederum bezeichnet technisch nicht die Wahrscheinlichkeit eines Mißerfolgs, sondern spezifisch auch das Produkt aus dieser Wahrscheinlichkeit und einem möglichen Schaden.

[34] "I think logicians should have two principal aims: 1st, to bring out the amount and kind of *security* (approach to certainty) of each kind of reasoning, and 2nd, to bring out the possible and esperable *uberty*, or value in productiveness, of each kind." Brief an Frederic A. Woods, 1913, in:

beiden Zielen besteht ein **Zielkonflikt**: Verläßlichkeit kann man sich durch Verzicht auf Gehaltserweiterung erkaufen (Beschränkung auf gültige Argumente), und Gehaltserweiterung durch Verzicht auf Verläßlichkeit (alle Argumente gleichermaßen zulassen). Wie man beide Ziele gewichten soll, ist (wiederum) keine Frage der Logik; aber was man hoffen kann, ist, die Zusammenhänge zu studieren, um eine Gewichtung mit Verstand wählen zu können.

3.3. Inductio per enumerationem simplicem

Es gibt (ebenso wie bei Fehlschlüssen) keine zwingende Art, gehaltserweiternde Argumente zu klassifizieren.

Die klassische Entgegensetzung ist einfach die zwischen **deduktiven** und **induktiven Argumenten**. Induktion, lat. inductio, ist die Übersetzung von Aristoteles' epagogé (ἐπαγογή; bei Aristoteles noch nicht terminologisch), wörtlich: das Heranführen (an eine allgemeine Aussage, über Beispiele [35]). Daher stammen die **traditionellen Definitionen von Deduktion und Induktion**: Die Deduktion führe *vom Allgemeinen zum Besonderen* (gedacht wird meist an Syllogismen). Die Induktion führe entsprechend *vom Besonderen zum Allgemeinen.*[36] Diese Definitionen sind problematisch (es gibt Gegenbeispiele) und eigentlich nur noch von historischem Interesse.[37]

Das Paradigma für Schlüsse vom Besonderen auf das Allgemeine ist die **Inductio per enumerationem simpicem**:

Alle bislang beobachteten Schwäne sind weiß.

Alle Schwäne sind weiß.

Die Benennung als **Inductio per enumerationem simplicem** geht auf Francis Bacon zurück, der die Induktion 'der Alten' (= alten Griechen) eben als 'Induktion durch

Charles S. Peirce, *The Essential Peirce: Selected Philosophical Writings, Vol. I (1867–1893)* (edd. Nathan Houser, Christian Kloesel), Bloomington and Indianapolis: Indiana University Press, 1992, 553 n. 7 (= CP 8.384); Hervorhebungen im Original. Vgl. id., An Essay toward Improving Our Reasoning in Security and Uberty [1913], in: Charles S. Peirce, *The Essential Peirce: Selected Philosophical Writings, Vol. II (1893–1913)* (ed. Peirce Editions Project), Bloomington and Indianapolis: Indiana University Press, 1998, S. 463–474. (*Uberty* (von frz. *uberté*): Rich growth, fruitfulness, fertility; copiousness, abundance, so OED online [2.5.2015]).

[35] Ist, wer langsam liest, ein guter Leser? Eher nicht. Ist, wer langsam läuft, ein guter Läufer? Auch nicht. Ist, wer langsam lernt, ein guter Lerner? Ist er nicht. – Also: Langsamkeit ist wohl keine Tugend.

[36] Vgl. etwa Mill über die klassische Unterscheidung: "Reasoning [...] is popularly said to be of two kinds: reasoning from particulars to generals, and reasoning from generals to particulars; the former being called Induction, the latter Ratiocination or Syllogism." Mill 1843, II.i.3 (= *Collected Works* Vol. VII, S. 162). Vgl. auch die sich anschließenden Ausführungen Mills, warum diese Bestimmung unpräzise sei.

[37] Einige gültigen Modi des Syllogismus führen vom Allgemeinen auf *Allgemeines* (denken Sie an den Modus Barbara, s.o. 1.1). Unter den induktiven Schlüssen wiederum finden sich solche, die vom Besonderen auf *Besonderes* führen (s.u.; vielleicht kann man die freilich wegdefinieren, John Mill aber etwa hielt diese sogar für grundlegend).

bloßes Aufzählen' beschreibt. Bacon betont aber auch gleich die Schwäche enumerativer Induktionen: Induktion durch simples Aufzählen sei kindisch ("Inductio enim quae procedit per enumerationem simplicem res puerilis est"), denn Gegenbefunde ließen sich nicht ausschließen, und geurteilt werde meist aufgrund von **zu wenigen Fällen** und **solchen, die sich gerade anbieten**.[38] (Heute würde man sagen: Die empirische Grundlage ist oft zu klein und die Auswahl fragwürdig; vgl. den Fehlschluß der voreingenommenen Statistik.) Bacon (und nach ihm andere, etwa Mill) *kritisiert* also diese Art der Induktion, und eine gute Übersetzung ist also 'Induktion durch *schlichtes/bloßes* Aufzählen' (nicht: *einfaches* Aufzählen).[39] Wir bleiben hier bei der lateinischen Benennung, um die 'Inductio per enumerationem simplicem' von der modernen 'enumerativen Induktion' unterscheiden zu können.

(Bacon und Mill empfehlen anstelle der Inductio per enumerationem simplicem Formen der sog. **eliminativen Induktion**, d. h. eines Schlusses auf die wahre Generalisierung durch Ausschluß falscher Generalisierungen; wir lassen die eliminative Induktion hier aus. **[= Sherlock Holmes' Schluß]**)

Die Kritik an der Inductio per enumerationem simplicem betrifft die Verläßlichkeit des Schlusses; im 20. Jh. gewinnt dieser Schluß in verbesserter Form als **enumerative Induktion** wieder Anhänger: Viele empirische Generalisierungen ('Alle Raben sind schwarz'; 'Alle Salze sind wasserlöslich', usw.) lassen sich nämlich als enumerative Induktionen beschreiben.[40]

3.4. Gehaltserweiternde Argumente: Peirce' Klassifikation

Immer noch hilfreich für das Verständnis auch der gegenwärtigen Sicht ist eine Unterteilung von Argumenten, die sich bei dem amerikanischen Pragmatisten Charles S. Peirce findet **[Aussprache: Pörs, nicht Piers!, engl. purse, nicht peers!]**. Peirce unterscheidet zwischen zwei Arten gehaltserweiternder Argumente: **Induktionen** (im engeren Sinn) und **Abduktionen**.[41]

[38] *Novum Organum*, a 105: "Inductio enim quae procedit per enumerationem simplicem res puerilis est, et precario concludit, et periculo exponitur ab intantia contradictoria, et plerumque secundum pauciora quam par est, et ex his tantummodo quae praesto sunt, pronunciat." Vgl. ib., S. 44/45 (Krohn) und a 69.

[39] Mill pries dann auch Bacon für diese Kritik: "It was, above all, by pointing out the insufficiency of this rude and loose conception of Induction, that Bacon merited the title ... of Founder of the Inductive Philosophy"; *Logic*, S. 392; vgl. 312–313; 404; letztlich gehe es aber nicht ohne diese: III.xxi.1, S. 566; 583; 609.

[40] Hans Reichenbach schreibt dies in *Experience and Prediction* (1935, S. 389) Mill zu: "It is the great merit of John Stuart Mill to have pointed out that all empirical inferences are reducible to the inductio per enumerationem simplicem."

[41] Erstmalig in: Charles S. Peirce, A Theory of Probable Inference, in: Id., Writings, Vol. 4 (ed. W. Kloesel), Bloomington/Indianapolis 1989, S. 408–450.

$$\textbf{Inference}\begin{cases}\text{Explicative (Deductive)}\\[1em] \textbf{Ampliative}\begin{cases}\textbf{Induction}\\[1em] \textbf{Abduction}\text{ (Hypothesis, Retroduction)}\end{cases}\end{cases}$$

Abb. 3.1. Charles Peirce' Klassifikation von Argumenten. Die Übersicht stellt eine Kompilation dar (Peirce' hat seine Auffassungen immer wieder modifiziert); sie ist also historisch nicht genau.[42] – Biographische Notiz: Charles Sanders Peirce spielt für die Thematik dieser Vorlesung eine besondere Rolle: Auf ihn gehen nicht nur die moderne Klassifikation und zentrale Begriffe zurück; er ist auch Begründer des modernen statistischen Schließens (s. Vorlesung 9).[43]

Peirce' Paradigma für **Induktionen** sind Schlüsse von einzelnen Instanzen auf eine allgemeine Regel:[44]

Case. – These beans are from this bag.
Result. – These beans are white.

Rule. – All the beans from this bag are white.

Die Formulierung scheint auf den ersten Blick ein wenig umständlich (eine Prämisse mehr als oben bei den Schwänen). Vier Punkte sind bemerkenswert: (i) Die Ausdrücke in Prämissen und Konklusion (beans, bag, white) sind dieselben. (ii) Das Paradigma ist weniger eine Generalisierung über vorgefundene Instanzen (beobachtete weiße Schwäne) als vielmehr eine Entnahme (Stichprobe) wie für eine Qualitätskontrolle. (iii) Man kann Bohnen prinzipiell zählen (sowohl die Stichprobe als auch die Bohnen im Sack).[45] (iv) Die Konklusion ist nicht qualifiziert. (Die letzten beiden Punkte hängen zusammen, wie wir sehen werden.)

Die zweite Art von gehaltserweiternden Schlüssen bezeichnet Peirce als **Abduktionen**[46]. Sein Paradigma ist folgendes:[47]

[42] In seiner frühen Phase unterscheidet Peirce zwischen Induction und Hypothesis (Abduction) als verschiedenen Formen von Schlüssen; später zwischen den drei Formen der Induktion (wobei Crude Induction wenig Bedeutung zukomme), während er die drei Arten – Abduction, Deduction, Induction – als drei *Phasen* des wissenschaftlichen Forschens auffaßt.

[43] Die moderne Notation in der Logik geht übrigens auch auf ihn zurück ($\prod_x(\text{ball}_x \multimap\!\!\prec \text{red}_x)$), heute $\forall x(bx \to rx)$; Freges Notation ist (zurecht) ungebräuchlich).

[44] Charles Peirce, Deduction, Induction, and Hypothesis, in: Popular Scientific Monthly 1878 (= EP 1, S. 188). Ein späterer Ansatz findet sich in: A Neglected Argument for the Reality of God, 1908 (= EP 2, S. 434 ff.; bes. S. 440–445).

[45] Es dürfte beispielsweise einen Unterschied machen, ob die Stichprobe aus 3 oder aus 300 Bohnen besteht, oder ob 3 % oder 30 % der Bohnen aus dem Sack umfaßt.

[46] In diesem Sinn erst bei Peirce; vgl. OED Online s.v. 3b [2.5.2015].

[47] Charles S. Peirce, Pragmatism as the Logic of Abduction (= 7th Harvard lecture) 1903, in: EP 2, S. 226–241, hier S. 231 (CP 5.189).

The surprising fact, C, is observed;
But if A were true, C would be a matter of course.

Hence, there is reason to suspect that A is true.

Bemerkenswert ist hier: (i) Die Ausdrücke in Prämissen und Konklusion (A, C) sind nur teilweise identisch. (ii) Es gibt keine Angabe über die Herkunft von A gibt – nur, daß es 'überraschend' ist. (iii) Es gibt nur *eine* überraschende Tatsche, und zählen kann man also wenig.[48] (iv) Die Konklusion ist qualifiziert ('reason to suspect'). (Die letzten beiden Punkte hängen zusammen, wie wir sehen werden.)

3.5. Stichproben als Modell gehaltserweiternder Argumente

Es ist es ein fruchtbarer Gedanke, bestimmte gehaltserweiternde Argumente nach dem Vorbild von **Stichprobenentnahmen** zu interpretieren. Viele gehaltserweiterende Argumente lassen sich mittels Aussagen über **Stichproben** (**sample**)[49] aus einer gegebenen **Grundgesamtheit** (**population**) gut fassen. Und es könnte auf diese Weise gelingen, **Aussagen über die Stärke (Verläßlichkeit, security) eines gehaltserweiternden Arguments** treffen zu können – denn Stichprobenentnahmen sind formal gut zu beschreiben.

Dies gilt v.a. für zwei der drei Formen: Man kann aus Aussagen über eine Stichprobe auf Aussagen über die Grundgesamtheit schließen, oder umgekehrt von Aussagen über die Grundgesamtheit auf Aussagen über eine Stichprobe. (Im dritten Fall, dem der Abduktion, schließt man von Merkmalen der 'Grundgesamtheit' und Merkmalen einer 'Stichprobe' darauf, daß es sich bei diesem um eine Stichprobe aus jenem handelt – dieser Fall macht mehr Probleme.)

3.6. Statistische Syllogismen

Argumente, in denen aus Aussagen über die Grundgesamtheit auf Aussagen über eine Stichprobe geschlossen wird, bezeichnet man als **statistische Syllogismen** (**statistical syllogism**)[50]. Sie haben die Form:

Aussage über Grundgesamtheit
Aussage über Stichprobenentnahme

——————————————— [wahrscheinlich]

Aussage über Stichprobe

[48] Dies stimmt nicht notwendig, aber führt hier zu weit; vgl. Deduction, Induction, Hypothesis, EP 1, S. 192.

[49] Mit 'Stichproben' sind hier stets sog. Zufallsstichproben (random sample) gemeint, im Gegensatz zu sog. Quotenstichproben. (Bei Umfragen etwa, in denen Ausschöpfungsquote/Rücklaufquote gering ist, kann es sinnvoll sein, eine hinsichtlich bestimmter Merkmale bewußte Auswahl zu treffen.)

[50] Den Namen trägt der Schluß wegen seiner Ähnlichkeit mit dem üblichen, 'kategorischen' Syllogismus.

Ein Beispiel wäre:

Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.
Diese Bohne ist dem Sack entnommen.
——————————————— [wahrscheinlich]
Diese Bohne ist eine Arabica-Bohne.

Dieses Argument ist gehaltserweiternd bzw. riskant – man kann sich nicht sicher sein, daß die Konklusion wahr ist. Aber je höher der Anteil von Arabica-Bohnen im Sack ist, um so 'häufiger' wird man richtig liegen (s.u.). Man kann auch auf Aussagen über eine größere Stichprobe schließen:

Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.
Diese 10 Bohnen sind dem Sack entnommen.
——————————————— [wahrscheinlich]
Etwa 9 der 10 Bohnen sind voraussichtlich Arabica-Bohnen.

Nota bene: In diesem Fall wird man nicht unbedingt 'häufiger' richtig liegen, wenn der Anteil an Arabica-Bohnen hoch ist: Denken Sie an den Fall, daß 50 von 100 Bohnen Arabica wären, und man vermutete entsprechend, 5 von 10 Bohnen der Stichprobe wären ebenfalls Arabica-Bohnen.

In beiden Fällen hängt aber viel daran, *wie* die Bohnen der Stichprobe dem Sack entnommen wurden (wir werden später sagen: an dem Wahrscheinlichkeitsmodell).

Man muß diese Argumente von **statistischen Deduktionen** unterscheiden; diese sind gültige Argumente mit Wahrscheinlichkeitsaussagen, die die mathematische Wahrscheinlichkeitstheorie behandelt. (Diese hatte Peirce oben im Sinn.) Vergleichen Sie:

Dieser Würfel hat 6 Seiten: 1, 2, 3, 4, 5, 6
Der Würfel ist fair.
———————————————
Die Wahrscheinlichkeit, eine "4" zu werfen, ist 1/6.

3.7. Enumerative Induktion und statistische Generalisierung

Argumente von Aussagen über eine Stichprobe auf Aussagen über die Grundgesamtheit bezeichnet man als **enumerative Induktion** (**enumerative induction**) bzw. (bei statistischen Aussagen) als **statistische Generalisierung**:

Aussage über Stichprobe
Aussage über Stichprobenentnahme
——————————————— [wahrscheinlich]
Aussage über Grundgesamtheit

(Mit den formalen Eigenschaften von Schlüssen dieser Art befaßt sich die mathematische Theorie der Schätzung (estimate) (= induktive Statistik).) Beispiel:

Diese 10 Bohnen sind dem Sack entnommen.
9 der 10 Bohnen sind Arabica-Bohnen.
——————————————————— [wahrscheinlich]
Im Sack befinden sich wohl ca. 90 Arabica-Bohnen und 10 Robusta-Bohnen.

Wiederum hängt einiges daran, *wie* die Bohnen der Stichprobe dem Sack entnommen wurden (an dem Wahrscheinlichkeitsmodell) – Sie erinnern sich an die voreingenommene Statistik –, aber auch an dem Umfang der Stichprobe (mehr ist besser).

4. Gehaltserweiternde Schlüsse (cont'd)

4.1. Wahrscheinlichkeitsschlüsse

Stichproben als Modell für (bestimmte) gehaltserweiternde Argumente zu nehmen hat einen augenscheinlichen Vorteil: Man die 'Wahrscheinlichkeit', richtig zu liegen (daß also die Konklusion wahr ist), mathematisch behandeln. Mit anderen Worten: Man hat ein 'Maß' für die Verläßlichkeit (security) des Arguments. Wir werden Argumente, bei denen sich eine solche Behandlung anbietet – also besonders statistische Syllogismen und statistische Generalisierungen – als **Wahrscheinlichkeitsschlüsse (probable inference)** bezeichnen.

Vorüberlegung: Nimmt man Mills Kritik an gültigen Argumenten ernst, sind statistische Generalisierungen erkennbar interessanter als statistische Syllogismen: Nicht unbedingt, weil man mehr lernt, wenn man denn richtig liegt (*uberty*)[51], sondern v.a., weil man Informationen über eine Grundgesamtheit braucht, um statistische Syllogismen anzuwenden – und in vielen Fällen erhält man diese nur über eine statistische Generalisierung. Wir werden dennoch zunächst statistische Syllogismen betrachten, da diese einfacher zu behandeln sind.

Zuerst werden wir die folgende Frage betrachten: **Was heißt hier 'wahrscheinlich'?**

Betrachten Sie den folgenden Fall eines **statistischen Syllogismus**: In einem Sack mit Kaffeebohnen (Bohnensack = Grundgesamtheit) befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen (Aussage über die Grundgesamtheit bzw. über die Häufigkeit eines 'Merkmals' (Arabica = Merkmal) in der Grundgesamtheit). Aus dem Sack entnehmen wir eine Handvoll Bohnen (Bohnen = Stichprobe), etwa:

[51] Dies muß nicht zwingend so sein: Zu wissen, welches Los in einer Lotterie gewinnt, ist so informativ wie nützlich.

Im Sack befinden sich 90 Arabica- und 10 Robusta-Bohnen. (Grundges.)
10 Bohnen werden ‘geeignet’ aus dem Sack entnommen.

Die Bohnen sind ca. 9 Arabica-Bohnen und 1 Robusta-Bohne. (Stichprobe)

Voraussetzen müssen wir noch, daß die 10 Bohnen der Probe auf eine bestimmte, ‘geeignete’ Weise (nämlich: ‘zufällig’) entnommen wurden (was immer das heißt).

Der Schluß ist gehaltserweiternd: Man kann sich nicht sicher sein, daß es wirklich 9 Arabica-Bohnen und 1 Robusta-Bohne sein werden. Es könnten sogar 10 Robusta-Bohnen sein – dies ist nur nicht sehr wahrscheinlich (aber die Wahrscheinlichkeit ließe sich einfach berechnen). In diesem Fall bleibt die Aussage über die Häufigkeit (den Anteil) von Arabica-Bohnen erhalten, wir verlieren nur die **Information über die absoluten Häufigkeiten** in der Grundgesamtheit (diese ist ja endlich groß). Dies ist aber ein Spezialfall.

Vergleichen Sie den Fall, daß nur **eine einzige Bohne** entnommen wird:

Im Sack befinden sich 90 Arabica- und 10 Robusta-Bohnen. (Grundges.)
Eine Bohne wurde ‘geeignet’ aus dem Sack entnommen.

Die Bohne ist eine Arabica-Bohne. (Stichprobe)

Vermutlich ist die Konklusion richtig. Die **Informationen über die relativen Häufigkeiten** der Bohnen im Sack sind nun aber ebenfalls verloren. Wir wären auch ausgehend von einer anderen Grundgesamtheit (z. B. einem Sack mit 99 Arabica-Bohnen und 1 Robusta-Bohne) zum selben Ergebnis gekommen, können aber nicht darauf zurückschließen. Dies ist meistens der Fall: Bei einer einzigen Bohne ist die Information über die relativen Häufigkeiten ganz verloren; bei den meisten Stichproben ist die Information wenigstens verzerrt (z. B. beim Entnehmen von 9 Bohnen aus dem betrachteten Sack Bohnen – weil man nur ganze Bohnen entnehmen kann).

4.2. Wahrscheinlichkeit als Eigenschaft des Schlusses

Die Wahrscheinlichkeit läßt sich (nicht notwendig, aber gut) als eine **Eigenschaft des Arguments** verstehen; dies ist die (auf Peirce zurückgehende) kanonische Auffassung:

Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.
Diese Bohne ist dem Sack entnommen.

——————————— [wahrscheinlich]

Diese Bohne ist eine Arabica-Bohne.

Die Wahrscheinlichkeit sagt aus, wie häufig man richtig und wie häufig man falsch liegt, wenn man Argumente der Art verwendet. – Im Beispiel läge man in 9 von 10 Fällen richtig, wenn man **wiederholt** Bohnen aus dem Sack entnähme (mit Zurücklegen). – Die Informationen über die relativen Häufigkeiten in der Grundgesamtheit gehen in dieser Auffassung nicht verloren, sondern finden sich als Häufigkeit (Wahrscheinlichkeit) wieder, mit der ein Argument von wahren Prämissen zu wahren Konklusionen führt.

Die Form eines statistischen Syllogismus wäre folgende:

$$\frac{\begin{array}{l} X \text{ Prozent der } B\text{s sind } A. \\ \text{Dieses } x \text{ ist aus } B \text{ entnommen.} \end{array}}{\text{Dieses } x \text{ ist } A.} \; [X \text{ Prozent}]$$

Dieses Argument ist gehaltserweiternd und nicht-gültig. Die Stärke eines statistischen Syllogismus hängt wohl von der Anzahl der *A*s unter *B*s ab, also von X: Liegt X nahe bei 100, liegt ein starkes Argument vor (die Prämissen stützen die Konklusion stark: Wer so schließt, liegt meistens richtig); ist $X = 50$, stützen die Prämissen die Konklusion nicht, sondern sind irrelevant; gilt $X < 50$, sprechen die Prämissen eher gegen die Konklusion: Wer so schließt, liegt öfter flasch als richtig.[52]

Argumente nach dem Vorbild von Stichproben aufzufassen, ist eine fruchtbare Metapher; aber sie ist nicht ohne **Probleme:**

(i) "... in the long run"?

In jedem einzelnen Fall kann man daneben liegen (die Konklusion kann falsch sein). Einzig, wenn man Wahrscheinlichkeitsschlüsse wiederholt verwendet – "**in the long run**" – darf man erwarten, daß sich das Verhältnis wahrer zu falschen Konklusionen dem der Häufigkeit X der Ausprägung eines Merkmals in der Grundgesamtheit annähert. Aber was hilft dies, wenn man Argumente nur einmal 'verwendet'?

(ii) "... wiederholte Schlüsse"?

Man findet immer wieder die folgende Formulierung: Die Wahrscheinlichkeit eines Arguments entspreche "der Wahrscheinlichkeit, mit der ein Argument von wahren Prämissen zu wahren Konklusionen führt". Dies ist nicht falsch, aber was bedeutet es? – Es gibt eine klare Vorstellung davon, was es bedeutet, mehrfach (wiederum: mit Zurücklegen) eine Bohne aus einem Sack zu entnehmen. Es gibt eine klare Vorstellung davon, was es bedeutet, wiederholt Lotto zu spielen. (Beides sind Handlungen.) Es gibt eine leidlich klare Vorstellung davon, was es bedeutet, ein Experiment zu wiederholen. Aber gibt es eine klare Vorstellung davon, was es bedeutet, "ein Argument mehrfach zu verwenden" bzw. "**einen Schluß mehrfach zu ziehen**"? (Wenn ich zweimal einen Schluß 'denke': habe ich ihn dann zweimal 'gezogen'?) – Die Logik bestimmt Argumente/Schlüsse als Beziehungen zwischen Aussagen; wie diese Bestimmung sich auf die kanonische Auffassung von Wahrscheinlichkeitsschlüssen übertragen läßt, ist nicht offensichtlich.

[52] Vgl. Salmon, S. 87–89.

(iii) "… Entnahme von Stichproben"?

Es gibt einige Eigentümlichkeiten von Wahrscheinlichkeitsschlüssen, die keine Entsprechung bei deduktiven/gültigen Schlüssen haben. Die zweite Prämisse etwa haben wir jeweils in einer ungewohnten Form formuliert:

Dieses x ist aus B geeignet/zufällig entnommen.

In vielen Logik-Lehrbüchern findet sich stattdessen eine Formulierung, die aus der deduktiven Logik vertrauter ist:

x ist B.[53]

Die hier gewählte Formulierung soll sicherstellen, daß die Aussage über die Häufigkeit, mit der das Argument zu richtigen bzw. falschen Konklusionen führt, richtig ist. (In welcher Weise, werden wir später diskutieren.) Aber während wiederum im betrachteten Fall eines Sacks mit Bohnen eine klare Vorstellung besteht, was es bedeutet, eine Bohne aus dem Sack zu entnehmen, ist nicht unbedingt klar, was es etwa bedeutet, ein Exemplar eines Schwans aus der Menge der Schwäne zu 'entnehmen' (wenn man sich beispielsweise für die Aussage interessiert, die meisten Schwäne seien weiß).[54]

4.3. Das Lotterie-Paradox

Eine naheliegende Idee ist, die Wahrscheinlichkeit des Arguments wegzulassen, wenn der Schluß "stark genug" ist (die Wahrscheinlichkeit "hoch genug"[55]). Dies scheint eine realistische Annahme: Viele Annahmen gelten als etabliert, werden "akzeptiert", ohne weiter hinterfragt zu werden; und die Vorstellung, jeder würde die Gründe (und wie gut die Gründe sind) stets im Hinterkopf haben, ist weltfremd.[56] Dies bietet sich besonders in extremen Fällen an:

In einer (fairen) Lotterie befinden sich 1 000 Lose; nur eins ist ein Gewinn.
Dieses Los wurde gezogen.

Dieses Los ist eine Niete.

Die Kehrseite dieses Ansatzes bilden eine Reihe von (teils sehr technischen) Problemen, zu denen dieser Ansatz führt. Ein Beispiel dafür ist das **Lotterie-Paradox**

[53] Vgl. etwa Salmon, S. 88.

[54] Ein anderes Beispiel für eine Eigentümlichkeiten von Wahrscheinlichkeitsschlüssen, die keine Entsprechung bei deduktiven/gültigen Schlüssen haben, ist die Forderung nach "Prädesignation": Man muß wesentliche Elemente des Schlusses spezifizieren, bevor man "den Schluß zieht", damit die Aussage über die Häufigkeiten stimmt.

[55] Die Frage, welche Wahrscheinlichkeit "hoch genug" ist, kann die Logik nicht beantworten. Ein oft verfolgter Ansatz ist, von den Folgen einer solchen "Entscheidung" auszugehen: Ist ein Irrtum folgenschwer, sollte man strengere Maßstäbe anlegen. Dies setzt aber, wiederum, voraus, daß *Handlungen* an die Entscheidung geknüpft sind. Vgl. Richard Rudner, The Scientist *Qua* Scientist Makes Value Judgments. *Philosophy of Science* 20 (1), 1953, 1–6.

[56] Vgl. Henry Kyburg, *Science and Reason*, Oxford 1990, ch. 4, bes. S. 64.

(**lottery paradox**; Henry Kyburg, 1961).[57] Wir betrachten die (faire) Lotterie mit 1 000 Losen, von denen genau eines ein Gewinn ist, näher. Zum einen weiß man dann, daß **genau ein Los gewinnt**. Zum anderen aber läßt sich im Fall jedes einzelnen Loses ein starkes Argument dafür geben, daß dieses Los nicht gewinnen werde. Daraus aber folgt, daß man meinen sollte, **keines der Lose werde gewinnen**. Ein und dasselbe Argument scheint also zu unvereinbaren Konklusionen zu führen.[58]

Das Lotterie-Paradox hat zu immer neuen Versuchen von 'Lösungen' eingeladen und tut dies immer noch. Aber vielleicht ist es gar nicht plausibel, es 'lösen' zu wollen: Wenn es einen Informationsverlust gibt, dann sollte es Probleme geben – und das Lotterie-Paradox zeigt positiv, *wo* solche Probleme auftreten. Verliert man die Informationen über Häufigkeiten (trennt die Konklusionen von deren Prämissen), dann kann man mit den Konklusionen nicht einfach weiterarbeiten. Stellt man eine bestimmte Frage (Gewinnt Los Nr. 210? Gewinnt eines der Lose außer Los Nr. 210?), erhält man ggf. eine gute Antwort; aber die Antworten auf verschiedene Fragen können einender widersprechen.

4.4. Gegen "wahrscheinliche Konklusionen"

Man könnte die Wahrscheinlichkeit auch auf die Konklusion übertragen. Eine Möglichkeit wäre, sie zum Teil der Konklusion zu machen:

> Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.
> Diese Bohne ist dem Sack entnommen.
> ___
> Diese Bohne ist [*mit der Wahrscheinlichkeit X*] eine Arabica-Bohne.

Eine andere, die Konklusion als eine qualifizierte Aussage zu betrachten:

> Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.
> Diese Bohne ist dem Sack entnommen.
> ___
> [*Mit der Wahrscheinlichkeit X*:] Diese Bohne ist eine Arabica-Bohne.

Viele Logiker haben sich entschieden gegen diesen Ansatz gewehrt. **Carl Hempel** hat sich dagegen ausgesprochen, Qualifizierungen in die Konklusion hineinzuschreiben, mit folgendem Argument: Die Konklusion sei entweder wahr oder falsch, aber nicht wahrscheinlich oder unwahrscheinlich. Dies könne man von einer Aussage nur

[57] Henry E. Kyburg, *Probability and the Logic of Rational Belief*, Middletown, CT, 1961, S. 197.

[58] Kyburg wollte eigentlich gerade darauf hinaus, das dies nicht paradox sei. Man könne, so Kyburg, die Konklusionen durchaus 'akzeptieren', nur eben nicht zusammen: die logische Konjunktion gelte nicht mehr – was immer dies bedeutet (vgl. Hacking 1980, S. 155: "To state A and to state B, on the same occasion that one declines to state $A\&B$, can only invite an inquiry as to what on earth that means."). Das Paradox ist ursprünglich als Paradox von Regeln für die 'rationale Akzeptierbarkeit' von Aussagen formuliert (unter Akzeptanz versteht Kyburg "admit into my corpus of belief". Diesen Urspungskontext werden wir beiseitelassen. Eine interessante Antwort auf das Lotterie-Paradoxon findet sich bei Hacking 1980, S. 155–156.

relativ zu anderen Aussagen (den Prämissen) sagen, und es könne der Fall sein, daß ein- und dieselbe Aussage zugleich wahrscheinlich und unwahrscheinlich ist, bezogen auf verschiedene Prämissenmengen. (Wie auch – analog – ein gültiges Argument nicht die Wahrheit der Konklusion zeigt, sondern nur deren Wahrheit *gegeben* die der Prämissen.) [59] Die Idee ist: Insofern die Bohne zufällig aus diesem Sack entnommen wurde (also: gegeben die Prämissen), vermute man (mit Recht), mit 90-prozentiger Wahrscheinlichkeit werde es eine Arabica sein – aber die Prämissen dürfe man nicht weglassen.

Man könnte dem ersten Einwand entgegenhalten, daß die Analogie zur gültigen Argumenten nicht hergestellt werden müsse – warum sollten sich gültige und gehaltserweiternde Argumente nicht gerade in diesem Punkt unterscheiden? Daß die Konklusion entweder wahr oder falsch sein muß, könnte man ebenfalls bestreiten: Es gibt ja statistische Deduktionen, also gültige Argumente, bei denen sowohl die Prämissen als auch Konklusion Wahrscheinlichkeitsaussagen sind.

Der zweite Einwand findet sich u.a. bei **Wesley Salmon**[60] detaillierter ausgeführt. Salmon argumentiert ebenfalls dafür, daß in einem statistischen Syllogismus die Aussage über die Wahrscheinlichkeit auf den Schluß bezogen werden müsse. Er betrachtet das folgende Argument:

> 75 percent of the beans in the barrel are grade A.
> The next bean to be drawn from the barrel is a bean in the barrel.
> ___
> The next bean to be drawn from the barrel ist grade A.[61]

Die Konklusion dürfe nicht heißen: "The next bean to be drawn from the barrel is *probably* grade A." Auch er argumentiert mit einer Analogie zur Deduktion: Seien *alle* Bohnen von der Güteklasse A, lasse sich der Schluß frei folgendermaßen ausdrücken: "Since all of the beans in the barrel are grade A, the next bean to be drawn from the barrel *must be* of grade A." Hierbei verweise "must be" nicht auf eine Notwendigkeit, sondern diene als Hinweis darauf, daß die Aussage die Konklusion eines deduktiven Arguments sei. Analog könne das Argument durchaus folgendermaßen ausgedrückt werden: "Since 75 percent of the beans in the barrel are grade A, the best bean to be drawn from the barrel *probably* of grade A." Hier verweise "probably" aber analog einzig darauf, daß ein induktiver Zusammenhand zwischen Prämissen und Konklusion bestehe: "Just as 'must be' is not part of the conclusion itself, so is 'probably' not part of the conclusion itself." [62]

[59] Carl G. Hempel, Aspects of Scientific Explanation, in: id., *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: The Free Press, 1965, S. 382–383. Vgl. Salmon, S. 89. Vgl. die Behandlung in Hacking 1980, S. 147–149, sowie Christopher Hookway, *Peirce*, London/New York 1985, S. 211–213.
[60] Salmon, S. 89.
[61] Ibid., S. 88.
[62] Ibid., S. 89.

Auch hier ist die Analogie zur Deduktion nicht zwingend (dies bedeutet aber auch nicht, daß sie zwingend verfehlt wäre). Es ist freilich kein Wunder, daß keine Abhandlung über Logik die Konklusion deduktiv gültiger Argumente mit der Qualifizierung "notwendig" versieht: Denn eine solche Konklusion wird – offensichtlich (und wie Salmon auch sogleich einräumt [63]) – *durch die Prämissen nicht gestützt.* Dies ist im induktiven Fall mit "wahrscheinlich", wenigstens auf den ersten Blick, nicht offensichtlich.

Daß man schließlich von der Konklusion des induktiven Arguments nur *aufgrund der Prämissen* aussagt, daß diese "wahrscheinlich" der Fall sei, und man die Prämissen also erstens nicht weglassen kann und zweitens aufgrund von anderen Prämissen durchaus auch zu einer anderen Bewertung kommen könnte, ist völlig analog zu deduktiv gültigen Schlüssen: Auch hier fußt die Behauptung der Konklusion auf deren Verhältnis zu den Prämissen, und wären diese anders, könnte es auch die (Bewertung der) Konklusion sein.

Man könnte ebensogut andersherum argumentieren: Auf der Grundlage der Prämissen auf eine "wahrscheinliche" Konklusion zu schließen, ist kein gehaltserweiternder Schluß – denn es kann nicht der Fall sein, daß die Prämissen wahr, die Konklusion aber falsch ist. Argumente dieser Art lassen sich auch als statistische Deduktionen auffassen.

4.5. Abduktion und Inference to the Best Explanation (IBE)

Peirce' Abduktions-Konzeption steht in der Tradition der sog. **hypothetischen (hypothetisch-deduktiven) Methode** (**method of hypotheses**): Man schließt auf die Wahrheit einer Annahme (der Hypothese) aufgrund der Tatsache, daß diese wahre Implikationen hat. Der logischen Form nach ist dies eine Bejahung des Konsequens. Bei Peirce tritt aber die Forderung hinzu, die Hypothese müsse eine **Erklärung** des "überraschenden Phänomens" geben: "If A were true, C would be a matter of course."

Eine **Abduktion** (im ursprünglichen Sinn) ist ein **Schluß auf eine plausible/mögliche Erklärung**; daher die Qualifizierung in der Konklusion ("reason to suspect"). Es wird nur darauf geschlossen, daß die Hypothese A eine fruchtbare Annahme ist, die weiterzuverfolgen sich lohnt. (Die Frage, welchen Kriterien eine gute Abduktion genügen muß, ist immer noch strittig; sie wird v.a. in der Wissenschaftstheorie diskutiert.)

Beispiel: Wir vermuten, daß es in der Sonne die Elemente Eisen, Titan und weitere Metalle gibt; direkt nachschauen kann man aber nicht. Die Annahme stützt sich dar-

[63] Ibid., S. 89.

auf, daß diese Metalle jeweils das Licht einer bestimmter Wellenlängen absorbieren (und als Licht anderer Wellenlänge wieder abgeben). Im Licht der Sonne 'fehlen' nun bestimmte Wellenlängen weitgehend: Es finden sich schwarze Teile im sonst kontinuierlichen Lichtspektrum, die sog. **Fraunhofer-Linien** – und dies gerade bei den Wellenlängen, die für bestimmte Metalle charakteristisch sind. Das Vorkommen dieser Elemente würde dieses überraschende Phänomen also auf natürliche Weise erklären, und es erscheint sinnvoll, sie weiter zu untersuchen.[64]

Davon unterscheiden werden wir den **Schluß auf die beste Erklärung** (**Inference to the Best Explanation**, kurz **IBE**).[65] Hier ist die Idee, daß diejenige Annahme, die bestimmte Phänomene am besten erklärt, damit auch die vermutlich richtige Erklärung ist. Hier ist die Konklusion nicht qualifiziert (irren kann man sich natürlich dennoch). Kennen muß man nicht allein das Phänomen und dessen mögliche Erklärung (wie bei der Abduktion), sondern mehrere (alle?) mögliche Erklärungen; der Ansatz ist also komparativ (man vergleicht verschedene Erklärungen), idealerweise eliminativ (man verwirft alle Möglichkeiten bis auf eine). Zudem muß man sicherstellen, daß diese Erklärungen 'hinreichend' gut sind bzw. daß sich die richtige Erklärung (hoffentlich!) unter den untersuchten befindet.

Auch bei der IBE muß man ausbuchstabieren, was eine gute Erklärung auszeichnet. Harman merkt zur IBE an: "Presumably such a judgment will be based on considerations such as which hypothesis is simpler, which is more plausible, which explains more, which is less *ad hoc*, and so forth.[66] (Es gibt zahlreiche andere Kataloge sog. epistemischer bzw. methodologischer Werte.)

Beispiel: Die **Kontinentaldrifttheorie** Alfred Wegeners (1912, 1915) erklärt nicht allein die genaue Passung der Küstenlinien Südamerikas und Afrikas – eine Passung, die noch besser übereinstimmt, wenn man die Schelfränder der Kontinente betrachtet; die Südkontinente (also auch die Antarktika, Indien und Australien) passen lassen sich insgesamt gut zu einer größeren Landmasse zusammenfügen (Gondwana). Auf allen Südkontinenten finden sich Nachweise der Verbreitung des Glossopteris-Farns sowie weitere Flora und Fauna. Faltengebirge und Scherungen in Südamerika und Afrika stimmen überein. Auf allen Südkontinenten finden sich Hinweise auf die Eiszeit in

[64] Vgl. Charles Peirce, Deduction, Induction, and Hypothesis, 1878 (s.o.), EP 1, S. 192. (Die Entdeckung machten Bunsen und Kirchhoff 1860; bei Peirce wird nicht ganz klar, daß es sich um Absorptionslinien handelt (die jeweiligen Wellenlängen also weitgehend *fehlen*).

[65] Der Ausdruck wurde von Gilbert Harman geprägt (The Inference to the Best Explanation. *The Philosophical Review* 74 (1), 1965, S. 88–95); Harman vertrat besonders die Auffassung, die enumerative Induktion lasse sich als Form von IBE verstehen (soweit enumerative Induktionen gültig seien, stellten sie einen Spezialfall von IBE dar. (Vgl.: id., Enumerative Induction as Inference to the Best Explanation. *The Journal of Philosophy* 65 (18), 1968, S. 529–533). Der Ausdruck wurde durch Peter Lipton (*Inference to the Best Explanation*, London/New York [1991] ²2004) wiederbelebt.

[66] Gilbert Harman, The Inference to the Best Explanation. *The Philosophical Review* 74 (1), 1965, S. 89.

Carbon und Perm, u.a. Gletscherspuren. – Einzelne Phänomene werden auch von anderen Theorien erklärt (etwa die Verbreitung von Flora und Fauna durch postulierte Landbrücken zwischen den Kontinenten), aber keine andere erklärt alle diese Befunde aufgrund einer Annahme.[67]

Abb. 3.1. Einige von Alfred Wegener angeführte Belege für die 'Verschiebungstheorie'. $\mathbf{a_{1,2}}$ Übereinstimmung von Küstenlinien; **b** Ähnlichkeiten von Gesteinsformationen; **c** Auffaltung von Gebirgen in Verschiebungsrichtung; **d** 'zurückbleibende' Inselgruppen und Kaps ('Knick'); **e** ähnliche Faltungen von Strata; **f** Übereinstimmungen von Gesteinsarten; **g** *Glossopteris*-Flora (Gondwana; Perm: gepunktet).

Begrifflichkeit: Die Begrifflichkeit schwankt wieder einmal; schon Peirce selbst spricht von Schlüssen auf plausible Erklärungen mal von Abduktionen, teils aber auch von

[67] Alfred Wegener, *Die Entstehung der Kontinente und Ozeane*, Braunschweig 1915, ^{4}1929. Eine Rekonstruktion (von mehreren) der Argumente Wegeners und seiner Gegner geben Paul Thagard and Gregory Nowak (The Explanatory Coherence of Continental Drift. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1988*, Volume One: Contributed Papers, 1988, S. 118–126). – Eine der frühesten Verweise auf die Ähnlichkeiten der Küstenlinien findet sich im *Novum Organum* Bacons (II.27; möglicherweise bezieht sich Bacon dort aber auf Ähnlichkeiten der Westküsten beider Kontinente).

Hypothesen oder auch von ‘Retroduktionen’. Der Ausdruck ‘Abduktion’ wird heute oft auch gleichbedeutend mit ‘Schluß auf die *beste* Erklärung’ verwendet.

5. Zufall

5.1. Der Spielerfehlschluß

Der **Spielerfehlschluß** (**gambler's fallacy**) oder Monte-Carlo-Fehlschluß besteht darin, ein **zufälliges Ereignis** für wahrscheinlicher zu halten, wenn es zuvor seltener als erwartet aufgetreten ist, bzw. für unwahrscheinlicher, wenn es zuvor öfter als erwartet aufgetreten ist.

Ein **Erklärungsansatz**: Man erwartet von zufälligen Ereignissen auch eine erkennbar zufällige Verteilung. Kennt man einen einen Teil der Verteilung und sieht darin (zu) viel Regelmäßigkeit [68], vermutet man, diese müsse sich bald wieder ausgleichen. (Bei seltenen Ereignissen kann auch ein einziger Fall genügen.[69]) **Beispiele:**

1913 soll sich im **Casino de Monte-Carlo** in Monaco der folgende Vorfall ereignet haben: Bei einem Roulette-Spiel soll 26 Mal hintereinander Schwarz gefallen sein – und viele Spieler haben mit dem Setzen auf Rot viel Geld verloren (während das Haus viel Geld gewann). (Auf diesen Vorfall geht die Benennung "Monte-Carlo-Fehlschluß" zurück.)

Pierre-Simon Laplace, 1812: "Ich habe Männer gesehen, welcher [sic] sehnlich wünschten, einen Sohn zu haben, und denen es deshalb unangenehm war, wenn in dem Monathe, in welchem sie Väter werden sollten, Knaben geboren wurden. Sie bildeten sich ein, das Verhältniß oder [sic] männlichen Geburten zu den weiblichen müßte am Ende jedes Monathes dasselbe seyn, und glaubten daher, die schon geborenen

[68] Die meisten Menschen erwarten von einer zufälligen Verteilung interessanter Weise *zu wenig* Regelmäßigkeit: Bittet man Leute, eine 'zufällige' Abfolge von 'Kopf' und 'Zahl' aufzuschreiben, notiert kaum jemand siebenmal 'Kopf' bzw. 'Zahl' in Folge – dabei ist dies durchaus nicht unwahrscheinlich; vgl. Hacking, S. 30–31.

[69] Hat etwa ein Schüler einer Klasse am 29. Februar Geburtstag, könnte man auf den Gedanken kommen, es wäre damit noch unwahrscheinlicher, daß auch ein zweiter Schüler am gleichen Tag Geburtstag hat – weil man es für besonders unwahrscheinlich hält, daß dies in einer Klasse zweimal vorkommt. – Was scheint Ihnen als Strategie beim Lotto intuitiv (ohne langes Nachdenken) aussichtsreicher: Immer Ihren Geburtstag zu tippen oder immer die Gewinnzahlen der vorherigen Woche?

Knaben machten es wahrscheinlicher, daß nächsten [sic] Geburten Mädchen bringen würden.“ [70]

Portfolio-Theorie: Viele Privatanleger investieren in sog. Exchange-Traded Funds (ETFs), d. h. börsengehandelte Fonds; die meisten davon sind passiv gemanagte Fonds, die einen Index (etwa den DAX) ‘nachbilden’ (und nicht, wie aktiv gemangte Fonds, gezielt in bestimmte Firmen investieren). Die Überlegung hinter dieser Entscheidung ist einmal, daß für diese Fonds nur geringe Gebühren anfallen (Management und Handel sind billig), vor allem aber die Auffassung, daß niemand langfristig ‘den Markt schlägt’, d. h. einen Gewinn macht, der über der durchschnittlichen Marktentwicklung liegt: Aktienkurse seien nicht vorhersagbar, sondern zufällig.[71] – Viele Ratgeber und Zeitungsartikel empfehlen mit diesem Argument, in ETFs zu investieren. Weiterhin raten sie davon ab, bei einer länger andauernden Hausse (guten Entwicklung der Märkte) zu investieren – damit man sich nicht ärgert, wenn die ‘Blase platzt’. Ist man gut beraten, beide Ratschläge zu befolgen?

Wo liegt der Fehler beim Spielerfehlschluß? Dem Spielerfehlschluß könnte man die folgende Form geben:

> Ob man einen Jungen oder ein Mädchen bekommt, ist Zufall.
> In einem bestimmten Monat werden *erwartbar* gleich viele Jungen wie Mädchen geboren.
> Es wurden in diesem Monat schon viele Jungen geboren.
> ———
> Es werden im verbleibenden Teil des Monats mehr Mädchen geboren.

Dieses Argument ist **inkonsistent**. Um dies genauer zu fassen, betrachten wir zunächst Zufallsexperimente.

(Es gibt Möglichkeiten, sich gleich klarzumachen, daß hier etwas nicht stimmen kann: (i) Woher sollen die Kinder denn wissen, als was sie geboren werden sollen (wie viele andere Kinder schon geboren wurden)? (ii) Die Monatseinteilung ist ja nur eine Konvention, und nichts hindert uns, den Tag der Geburt als ersten oder letzten Tag des Monats eines anderen Kalenders aufzufassen. Auch in diesem Kalender würden in einem bestimmten Monat annähernd gleich viele Jungen wie Mädchen geboren – aber wäre es der erste Tag, wäre dann noch alles offen? und wäre es der letzte, wäre das Geschlecht des Kindes dann festgelegt?)

[70] Pierre-Simon Laplace, *Philosophischer Versuch über Wahrscheinlichkeiten* (übers. v. F. W. Tönnies). Heidelberg 1819, S. 175–176.

[71] Mit anderen Worten, man hält den Markt für (ziemlich) ‘effizient’. Die Idee hinter dieser sog. Effizienzmarkthypothese ist, daß alle wesentlichen Informationen über die zukünftige Entwicklung schon in den Preisen enthalten sind – hätte jemand Informationen über die zukünftige Entwicklung (und Geld übrig), so würde er oder sie es investieren und sich so der Preis anpassen. Die Theorie effizienter Märkte geht auf Eugene Fama zurück; er erhielt 2013 den Alfred-Nobel-Gedächtnispreis für Wirtschaftswissenschaften der Schwedischen Reichsbank.

5.2. Zufallsexperimente: Unabhängigkeit

Wir betrachten **Zufallsexperimente** (**statistical experiment**), konkreter eine 'Zufallsapparatur' (chance setup). Die Bezeichnung 'Zufallsapparatur' stammt von Ian Hacking:[72]

> A *chance setup* is a device or part of the world on which might be conducted one or more *trials*, experiments, or observations; each trial must have a unique *result* which is a member of a *class of possible results*.[73]

Ein Beispiel wäre etwa ein Münze, die geworfen wird (Würfel, ein Roulette oder eine Lotterie täten es auch). Werfen wir die Münze, sprechen wir von einem **Experiment** (**trial**); die **möglichen Ergebnisse oder Ausfälle** (**possible results**) sind KOPF und ZAHL.

Man hat eine intuitive (undurchdachte) Idee, was es bedeutet, die Münze sei **fair**.

An eines denkt man sofort: Die Ausfälle KOPF und ZAHL müssen **gleichwahrscheinlich** sein; das bedeutet, daß langfristig die Häufigkeit beider Ausfälle gleich groß sein sollte. Man sagt, die Münze habe **keinen Bias/keinen systematischen Fehler**[74]. Experimente mit (endlich vielen) gleichwahrscheinlichen Ausfällen bezeichnet man als **Laplace-Experimente**.

Aber fehlender Bias ist nur notwendig für Fairness, nicht hinreichend. **Fairness hat zwei Seiten**! Betrachten Sie die folgende Sequenz von KOPF (●) und ZAHL (○):

● ● ○ ○ ● ● ○ ○ ● ● ○ ○ ● ● ○ ○ ● ● ○ ○ ...

Die Ausfälle scheinen gleichwahrscheinlich zu sein, und sind (wenn es so weitergeht) langfristig gleichhäufig. Aber sie sind vorhersagbar, und also nicht zufällig.

Ausfälle zufälliger Würfe sind voneinander **unabhängig** (**independent**). Dies kann man unterschiedlich ausbuchstabieren: Die Ausfälle sollen **zufällig** (**random**) sein, d. h. unbeeinflußt von einander/von vorhergehenden Ausfällen. (Beachten Sie bitte: Im Englischen unterscheidet man oft zwischen 'chance' und 'randomness'). Man könnte auch so formulieren: Der Zufallsgenerator soll "**kein Gedächtnis**" haben; oder: Es gibt **keine Gewinnstrategie**.[75]

Eine (Laplace'sche) Zufallsapparatur ist fair genau dann wenn (i) die **Apparatur ohne Bias** ist und (ii) die **Ausfälle unabhängig** sind. Für alles Folgende ist die Unabhängigkeit das weitaus wichtigere Merkmal.

[72] Die Bezeichnung hat sich allerdings nicht durchgesetzt.
[73] Ian Hacking, *Logic of Statistical Inference*. Cambridge: Cambridge University Press, 1965, S. 13.
[74] Die Übersetzungen von 'Bias' schwanken: systematischer Fehler, (Meß-)Abweichung, Voreingenommenheit, Verzerrung.
[75] Vgl. Hacking, S. 25–29.

5.3. Der Spielerfehlschluß (cont'd)

Die Ausfälle eines fairen Zufalls-Experiments sind also gleichwahrscheinlich und unabhängig. Damit läßt sich nun die Inkonsistenz beim Spielerfehlschluß besser beschreiben.

Ob man einen Jungen oder ein Mädchen bekommt, ist Zufall.
In einem bestimmten Monat werden *erwartbar* gleich viele Jungen wie Mädchen geboren.
Es wurden in diesem Monat schon viele Jungen geboren.

Es werden im verbleibenden Teil des Monats mehr Mädchen geboren.

Man könnte sagen, es sind zwei Argumente; das **erste Argument** ist das folgende:

Ob man einen Jungen oder ein Mädchen bekommt, ist Zufall.

In einem bestimmten Monat werden *erwartbar* [76] gleich viele Jungen wie Mädchen geboren.

Das Argument setzt voraus (Prämisse), daß die Geburten *unabhängig* sind (wären sie es nicht, läge kein Zufall vor).

Das **zweite Argument** wiederum ist das folgende:

In einem bestimmten Monat werden [*erwartbar*] gleich viele Jungen wie Mädchen geboren.
Es wurden in diesem Monat schon viele Jungen geboren.

Es werden im verbleibenden Teil des Monats mehr Mädchen geboren.

Dieses Argument ist, wenn man das "erwartbar" ignoriert, eine gültige statistische Deduktion – es setzt dann aber voraus, daß die Geburten *abhängig* sind (daß also gerade kein Zufall vorliegt). Stützt sich aber erste Prämisse in diesem Argument (die Gleichverteilung) auf die Unabhhängigkeit der Geburten (ist also nur 'erwartbar'), so kann man diese im folgenden nicht als abhängig betrachten. Die Inkonsistenz beim Argument liegt also darin, die Geburten (Ausfälle des Experiments) zugleich als unabhängig und als abhängig zu betrachten.

Ob Ausfälle eines Versuchs wirklich zufällig sind, kann die Logik nicht sagen. Wenn man annimmt, daß die Ausfälle zufällig und also unabhängig sind, dann liegen dem Spielerfehlschuß unvereinbare Annahmen zugrunde. Aber wenn beim Roulette 25 mal hintereinander ROT fällt, dann könnte man dies auch als Beleg dafür sehen, daß das Roulette nicht fair und die Ausfälle nicht zufällig sind oder nicht ohne Bias. Dies spräche aber eher *dafür*, künftig auf ROT zu setzen. Mit anderen Worten: Das **Zufallsmodell** kann auch falsch sein.

[76] Hier steht 'erwartbar' (und nicht 'wahrscheinlich'), da es sich um ein Argument von Einzelwahrscheilichkeiten auf eine erwartbare Wahrscheinlichkeit handelt (eine statistische Deduktion, die sich wahrscheinlichkeitstheoretisch begründen ließe).

6. Grundlagen der Wahrscheinlichkeitstheorie

6.1. Wahrscheinlichkeiten von Ereignissen und Aussagen

Wir betrachten wiederum einen Sack mit Kaffeebohnen und eine Stichprobe daraus:

> Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.
> Diese Bohne ist dem Sack entnommen.
> ———————————————————— [wahrscheinlich]
> Diese Bohne ist eine Arabica-Bohne.

Wenn wir uns hier fragen, wie wahrscheinlich es ist, daß wir eine Arabica-Bohne erhalten, wenn wir eine Probe-Bohne entnehmen – mit anderen Worten: Wir fragen nach der **Wahrscheinlichkeit eines Ereignisses** (nämlich, eine Bohne in der Hand zu halten, wenn wir etwas bestimmtes tun).

Wir haben uns entschlossen, Wahrscheinlichkeitsargumente nach dem Vorbild solcher Stichprobenentnahmen aufzufassen, um diese besser behandeln zu können. Ein Argument ist eine Beziehung zwischen Aussagen, und entsprechend sollte auch dieses gehaltserweiternde Argument eine Beziehung zwischen Aussagen sein.

Wir können also das Argument auch anders verstehen: Wir uns fragen, wie wahrscheinlich es ist, daß es wahr ist, daß eine Arabica-Bohne erhalten, wenn wir eine Probe-Bohne entnehmen – mit anderen Worten: Wir fragen nach der **Wahrscheinlichkeit, daß eine Aussage/Proposition wahr ist** (nämlich: "Diese Bohne ist eine Arabica-Bohne." – wenn andere Aussagen wahr sind).

Ganz unproblematisch war diese Auffassung von Wahrscheinlichkeitsargumenten nicht (s.o., 3.9), aber eine fruchtbare Metapher. Wir werden im folgenden strenger unterscheiden zwischen der Wahrscheinlichkeit von Ereignissen einerseits und der von Aussagen (die von Ereignissen, Mengen und Häufigkeiten handeln) andererseits.

Es sind eigentlich nur zwei verschiedene Weisen, über Wahrscheinlichkeiten zu sprechen (die “Sprache der Logiker” und die “Sprache der Statistiker”, wie Hacking sie nennt). Wir werden beide “Sprachen” parallel verwenden – sie lassen sich weitgehend ineinander übersetzen.

Eigentlich wäre aus Sicht der Logik naheliegend, allein die logische “Sprache” zu verwenden; aber für viele Anwendungsfälle bietet sich die statistische Betrachtungsweise an. Diese hat zudem einen **wesentlichen Vorzug**: Betrachtet man Mengen, lassen sich in vielen Fällen **Aussagen über Wahrscheinlichkeiten von Ereignissen** als **Aussagen über Häufigkeiten**, d. h. über endliche Mengen, verstehen – und während vielen Menschen der Umgang mit Wahrscheinlichkeiten Probleme bereitet, ist der Umgang mit Aussagen über Häufigkeiten oft deutlich leichter.

Die Aussagen in Wahrscheinlichkeits-Argumenten waren Aussagen über zwei verschiedene Arten von Gegenständen: Aussagen über Einzel-Ereignisse, Mengen und Häufigkeiten (“Es ist eine Arabica!”; “Im Sack befinden sich 90 Arabica-Bohnen und 10 Robusta-Bohnen.”) sowie **Aussagen über Handlungen** (“Ich entnehme vier Bohnen.”). Diesen Handlungen bzw. Aussagen über diese Handlungen werden wir keine Wahrscheinlichkeiten zuordnen – ihnen entspricht in der Wahrscheinlichkeitstheorie einfach die Idee der “Zufallsauswahl”.

Die Wahrscheinlichkeitstheorie ist die Theorie **statistischer Deduktionen**; d. h. wir werden im folgenden allein gültige Argumente (bzw. Rechnungen) untersuchen. Auf den Nutzen der Wahrscheinlichkeitstheorie für gehaltserweiternde Argumente kommen wir später zurück.[77]

6.2. Die Begrifflichkeit der Wahrscheinlichkeitstheorie

Seien A und B **Aussagen**;

dann bezeichne:

$\mathcal{P}(A)$ die Wahrscheinlichkeit der Aussage A.

Weiterhin bezeichne:

$A \vee B$ die Adjunktion,

$A \wedge B$ $(A \,.\, B)$ die Konjunktion,

$\neg A$ $(\bar{A})$ die Negation

der jeweiligen Aussagen.

Seien A und B **Ereignismengen**;

dann bezeichne:

$\mathcal{P}(A)$ die Wahrscheinlichkeit des Ereignisses A.

Weiterhin bezeichne:

$A \cup B$ die Vereinigungsmenge,

$A \cap B$ die Schnittmenge,

$\neg A$ (A') das (absolute) Komplement

der jeweiligen Ereignisse.

[77] Vgl. zum folgenden Hacking, ch. 4.

Konventionen für $\mathcal{P}(A)$:

$0 \leq \mathcal{P}(A) \leq 1$

$\mathcal{P}(\Omega) = 1$ (dabei bezeichnet Ω die Tautologie bzw. ein sicheres Ereignis)

6.3. Die Additionsregel (einzelner Versuch)

Wir betrachten ein (einzelnes) Experiment mit einer Zufallsmaschine. – Die Zufallsmaschine hat u.a. die möglichen Ergebnisse A und B, und wir fragen uns, wie wahrscheinlich es ist, daß dabei das Ergebnis A oder das Ergebnis B herauskommt. Sind A und B **wechselseitig ausschließend**, dann kann man ihre Wahrscheinlichkeiten **addieren**:

$A \wedge B$ falsch, dann $\mathcal{P}(A \vee B) = \mathcal{P}(A) + \mathcal{P}(B)$; entsprechend für Mengen:

$A \cap B$ leer, dann $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B)$.

Beispiel. Frage: Unsere Zufallsmaschine ist ein Sack Kaffeebohnen mit 6 Robusta-Bohnen, 2 Arabica-Bohnen und 2 Bohnen, die verdorben sind, so daß man nicht mehr sagen kann, welche Sorte es ist. Alle Bohnen sind gleichgroß und gleichschwer und auch sonst nicht zu unterscheiden. Wie wahrscheinlich ist es, eine Arabica *oder* eine verdorbene Bohne zu ziehen?

Antwort: Da man nur *entweder* eine Arabica oder eine verdorbene (oder eine anderen) Bohne ziehen kann, und es jeweils 2 davon gibt (und insgesamt 10 Bohnen, so daß die Wahrscheinlichkeit, eine bestimmte Bohne zu ziehen, also jeweils $\mathcal{P}(\text{Bohne}) = \frac{1}{10}$ ist), ergibt sich:

$\mathcal{P}(\text{Arabica} \vee \text{verdorben}) = \mathcal{P}(\text{Arabica}) + \mathcal{P}(\text{verdorben}) = \frac{2}{10} + \frac{2}{10} = \frac{4}{10}$.

Merken: **A-usschließend? dann A-ddieren.**

6.4. Die Multiplikationsregel (mehrere Versuche)

Wir betrachten mehrere Versuche mit der Zufallsmaschine. – Die Zufallsmaschine hat u.a. die möglichen Ergebnisse A und B, und wir fragen uns, wie wahrscheinlich es ist, daß im ersten Versuch das Ergebnis A, im zweiten das Ergebnis B herauskommt (A **und** B)?

Sind A und B **unabhängig**, dann kann man Wahrscheinlichkeiten **multiplizieren**:

A und B unabhängig, dann $\mathcal{P}(A \wedge B) = \mathcal{P}(A) \cdot \mathcal{P}(B)$; entsprechend für Mengen:

A und B unabhängig, dann $\mathcal{P}(A \cap B) = \mathcal{P}(A) \cdot \mathcal{P}(B)$.

Beispiel. Frage: Unsere Zufallsmaschine ist wieder der Sack Kaffeebohnen mit 6 Robusta-Bohnen, 2 Arabica-Bohnen und 2 verdorbenen Bohnen. Wie wahrscheinlich ist es, erst eine Robusta und dann einen verdorbene Bohne zu ziehen (mit **Zurücklegen**[78])?

Antwort: Da die beiden Versuche (Zurücklegen!) unabhängig sind, und es 6 Robusta und 2 verdorbene Bohnen gibt (und insgesamt 10 Bohnen, so daß die Wahrscheinlichkeit, eine bestimmte Bohne zu ziehen, also jeweils $\mathcal{P}(\text{Bohne}) = \frac{1}{10}$ ist), ergibt sich:

$$\mathcal{P}(\text{erst Robusta} \wedge \text{dann verdorben}) = \mathcal{P}(\text{Robusta}) \cdot \mathcal{P}(\text{verdorben}) = \tfrac{6}{10} \cdot \tfrac{2}{10} = \tfrac{12}{100}.$$

Merken: **U-nabhängig? dann m-U-ltiplizieren.**

6.5. Zusammengesetzte Ereignisse (mehrere Zufallsmaschinen)

Zusammengesetzte Ereignisse sind Ausfälle von Versuchen an verschiedenen Zufallsmaschinen. (Von ‘verschiedenen’ Zufallsmaschinen können wir auch bei Fällen an einer Machine ‘ohne Zurücklegen’ sprechen, da sich dabei ja die Wahrscheinlichkeiten einer Zufallsmaschine ändert.) Interessiert sind wir an einem Ergebnis, das von mehreren Ausfällen zugleich abhängig ist, ohne aber (wie bei der Multiplikationsregel) zugleich die einzelnen Ausfälle festzulegen.

Ansatz/Strategie: Wir können bisher nur addieren und multiplizieren, also müssen wir nach ausschließenden und nach unabhängigen Ausfällen suchen, und die Versuchsabfolge entsprechend strukturieren. (Gelingt dies nicht, wird es kompliziert – aber uns geht es um die Ideen hinter der Wahrscheinlichkeit, nicht das Rechnen – wer sucht, wird also finden.)

Beispiel: In zwei verschiedenen Räumen (H12 und H7) lagern je ein Sack mit Kaffeebohnen; in H12 einer mit 300 Arabica und 100 Robusta-Bohnen, in H7 einer mit 100 Arabica und 300 Robusta.

Wir werfen zunächst die Münze; bei KOPF ziehen wir einen Bohne aus H12, bei ZAHL ziehen wir einen Bohne aus H7.

Frage: Wie groß ist die Wahrscheinlichkeit, einen Arabica-Bohne zu ziehen?

Antwort: Wir können eine Arabica entweder aus H12 oder aus H7 ziehen – nicht aber beides zugleich – dies schließt sich aus (Additionsregel):

$$\mathcal{P}(\text{Arabica}) = \mathcal{P}(\text{erst H12, dann Arabica}) + \mathcal{P}(\text{erst H7, dann Arabica}).$$

[78] ‘Zurücklegen’ ist immer ein Hinweis auf Unabhängigkeit (legt man nicht zurück, verändern sich die Wahrscheinlichkeiten)!

Aus H12 ziehen wir eine Arabica, wenn wir erst KOPF werfen und dann etwas Glück haben – beides ist unabhängig (Multiplikationsregel):

$$\mathcal{P}(\text{erst H12, dann Arabica}) = \mathcal{P}(\text{KOPF}) \cdot \mathcal{P}(\text{Arabica aus H12}) = \tfrac{1}{2} \cdot \tfrac{300}{400} = \tfrac{3}{8}.$$

Analog:

$$\mathcal{P}(\text{erst H7, dann Arabica}) = \mathcal{P}(\text{ZAHL}) \cdot \mathcal{P}(\text{Arabica aus H7}) = \tfrac{1}{2} \cdot \tfrac{100}{400} = \tfrac{1}{8}.$$

Somit:

$$\begin{aligned}\mathcal{P}(\text{Arabica}) &= \mathcal{P}(\text{erst H12, dann Arabica aus H12}) \\ &\quad + \mathcal{P}(\text{erst H7, dann Arabica aus H7}) \\ &= \tfrac{3}{8} + \tfrac{1}{8} = \tfrac{1}{2}.\end{aligned}$$

Die Wahrscheinlichkeit, eine Arabica-Bohne zu ziehen, ist also 1/2.

7. Bedingte Wahrscheinlichkeit

7.1. Kategorische und konditionale (bedingte) Aussagen

Wir unterscheiden kategorische Aussagen von konditionalen Aussagen. **Kategorische Aussagen** (categorical statements) stellen fest, daß etwas der Fall ist; dies können auch Aussagen über Möglichkeiten [79] oder Wahrscheinlichkeiten sein:

> Die Bahn kommt in einer Minute. – Die Niederschlagswahrscheinlichkeit beträgt 30 %.

Als **konditionale** oder **bedingte Aussagen** (conditional statements) bezeichnen wir alle Aussagen, die sich als Wenn-dann-Gefüge verstehen lassen (‘wenn’ und ‘dann’ müssen aber nicht verwendet werden). Dies können wiederum auch Aussagen über Wahrscheinlichkeiten sein:

> Wenn ich mir jetzt eine anzünde, kommt die Bahn in einer Minute. –
> Wenn die Großwetterlage gleich bleibt, beträgt die Niederschlagswahrscheinlichkeit 30 %.
> – Bei gleichbleibender Großwetterlage beträgt die Niederschlagswahrscheinlichkeit 30 %.

Anmerkung: Man kann es der Form einer Aussage nicht unbedingt ansehen, ob es sich um eine kategorische oder eine konditionale Aussage handelt. Aussagen über Dispositionen etwa könnte man ebensogut als kategorische Aussagen (‘Kochsalz ist wasserlöslich.’) wie auch als konditionale Aussagen (‘Wenn man Kochsalz in Wasser gibt, löst es sich.’) auffassen. Wir ignorieren die damit verbunden Fragen hier.

Schreibweisen für kategorische und konditionale Wahrscheinlichkeiten:

$\mathcal{P}(A)$ $\quad$ $\mathcal{P}$(Die Bahn kommt in einer Minute)

$\mathcal{P}(A \mid B)$, $\quad$ auch: $\mathcal{P}_B(A)$ [80]

$\mathcal{P}$(Die Bahn kommt in einer Minute | Ich zünde mir jetzt eine an).

[79] Wir verwenden also den Ausdruck ‘kategorische Aussage’ in einem weiten Sinn; in der traditionellen Logik werden als kategorische Aussagen vielfach einzig Tatsachenfeststellungen i.G.z. Aussagen über Möglichkeiten oder Notwendigkeiten bezeichnet.

[80] Die zweite Variante ist eigentlich besser lesbar, aber in der Philosophie selten; wir verwenden daher die erste.

Definition der bedingten Wahrscheinlichkeit:

$\mathcal{P}(A \mid B) = \mathcal{P}(A \wedge B) \,/\, \mathcal{P}(B) \qquad$ für $\mathcal{P}(B) \neq 0$.

Motivation: Die Definition der bedingten Wahrscheinlichkeit ist im Grunde sehr natürlich. Die klassische oder laplacesche Auffassung versteht die kategorische Wahrscheinlichkeit als das Verhältnis der Zahl der 'günstigen' Ausfälle eines Versuchs zur Zahl der 'möglichen' Ausfälle. Gedacht wird dabei an Versuche, in denen die einzelnen Ausfälle gleichwahrscheinlich sind (sog. Laplace-Versuche), wie etwa im Fall eines fairen Würfels:

$$\mathcal{P}(\text{gerade Zahl}) = \frac{\text{Anzahl (gerade Zahlen)}}{\text{Anzahl (Zahlen insgesamt)}} = \frac{\text{Anzahl (2, 4, 6)}}{\text{Anzahl (1, 2, 3, 4, 5, 6)}} = \frac{3}{6}.$$

Bei der bedingten Wahrscheinlichkeit liegt der gleiche (der selbe) Ansatz zugrunde; nur werden allein die Ausfälle betrachtet, die der Bedingung entsprechen (hier fett gesetzt):

$$\mathcal{P}(\text{ger. Zahl} \mid \geq 4) = \frac{\text{Anzahl (gerade Zahlen } \geq 4)}{\text{Anzahl (Zahlen insgesamt } \geq 4)} = \frac{\text{Anzahl (2, }\mathbf{4}\text{, }\mathbf{6}\text{)}}{\text{Anzahl (1, 2, 3, }\mathbf{4, 5, 6}\text{)}} = \frac{2}{3}.$$

7.2. Anwendung: Ein Versuch

Wir betrachten 2 Säcke mit Kaffeebohnen, A und B. Im Sack A befinden sich 8 Bohnen, davon 2 von der Sorte Robusta (R); in B befinden sich 400 Bohnen, davon 300 Robusta. Die übrigen sind jeweils Arabica.

Frage: Wenn wir blind (zufällig) einen der beiden Säcke wählen und dann eine Bohne daraus ziehen, und die ist eine Robusta – mit welcher Wahrscheinlichkeit stammt die dann aus Sack A?

Vorüberlegungen:

Wir wissen: $\mathcal{P}(R \mid A) = \frac{1}{4}$, $\quad \mathcal{P}(R \mid B) = \frac{3}{4}$, $\quad \mathcal{P}(A) = \mathcal{P}(B) = \frac{1}{2}$. Wir suchen: $\mathcal{P}(A \mid R)$.

(i) Die **Definition der bedingten Wahrscheinlichkeit** können wir für zweierlei gebrauchen: Erstens erlaubt sie uns, bedingte Wahrscheinlichkeiten ($\mathcal{P}(X \mid Y)$) und Wahrscheinlichkeiten von Konjunktionen ($\mathcal{P}(X \wedge Y)$) ineinander umzuformen:

$\mathcal{P}(X \mid Y) \longleftrightarrow \mathcal{P}(X \wedge Y)$.

Da für Konjunktionen das **Kommutativgesetz** ($X \wedge Y = Y \wedge X$) gilt, erlaubt dies weiterhin, die bedingten Wahrscheinlichkeiten $\mathcal{P}(X \mid Y)$ und $\mathcal{P}(Y \mid X)$ ineinander umzuformen:

$\mathcal{P}(X \mid Y) \longleftrightarrow \mathcal{P}(X \wedge Y) \longleftrightarrow \mathcal{P}(Y \wedge X) \longleftrightarrow \mathcal{P}(Y \mid X)$.

Die Definition der bedingten Wahrscheinlichkeit ist oft nur zusammen mit dem Kommutativgesetz brauchbar; sehen Sie also Definition und Kommutativgesetz als Einheit.

(ii) Wenn wir die Definition der bedingten Wahrscheinlichkeit anwenden, treten aber auch kategorische Wahrscheinlichkeitsaussagen auf; $\mathcal{P}(A)$ und $\mathcal{P}(B)$ sind bekannt, nicht aber $\mathcal{P}(R)$. Hier brachen wir einen Kniff: Eine Robusta-Bohne R können wir entweder aus Sack A oder aus Sack B erhalten – nicht anders und nicht beides zugleich (mit anderen Worten: die beiden Fälle sind hier erschöpfend und ausschließend). Daher ist

$$\mathcal{P}(R) = \mathcal{P}(R \wedge A) + \mathcal{P}(R \wedge B).$$

Die Idee hinter diesem Kniff ist oft nützlich; wir werden ihr später in anderer Form als 'totale Wahrscheinlichkeit' wiederbegegnen.

Antwort:

$$\begin{aligned}
\mathcal{P}(A \mid R) &= \mathcal{P}(A \wedge R) \;/\; \mathcal{P}(R) && \text{nach Definition}\\
&= \mathcal{P}(R \wedge A) \;/\; \mathcal{P}(R) && \text{Kommutativgesetz}\\
&= \mathcal{P}(R \mid A) \cdot \mathcal{P}(A) \;/\; \mathcal{P}(R) && \text{Zähler: nach Definition}\\
&= \mathcal{P}(R \mid A) \cdot \mathcal{P}(A) \Big/ \big(\mathcal{P}(R \wedge A) + \mathcal{P}(R \wedge B)\big) && \textbf{(Aufgabe)}\\
&= \mathcal{P}(R \mid A) \cdot \mathcal{P}(A) \Big/ \big(\mathcal{P}(R \mid A) \cdot \mathcal{P}(A) + \mathcal{P}(R \mid B) \cdot \mathcal{P}(B)\big) && \text{n. Definition.}
\end{aligned}$$

Jetzt kennen wir alle Ausdrücke, und könnten die Werte einsetzen:

$$= \frac{2}{8} \cdot \frac{1}{2} \Big/ \left(\frac{2}{8} \cdot \frac{1}{2} + \frac{300}{400} \cdot \frac{1}{2}\right) = \frac{1}{8} \Big/ \left(\frac{1}{8} + \frac{3}{8}\right) = \frac{1}{8} \Big/ \frac{1}{2} = \frac{1}{4}.$$

Die Wahrscheinlichkeit, daß eine Robusta-Bohne aus Sack A stammt, ist 1/4.

7.3. Motivation

Der Grund, warum wir Wahrscheinlichkeiten untersuchen, ist der, daß wir auf der Suche nach einer Logik gehaltserweiternder Schlüsse sind, und die Wahrscheinlichkeitstheorie ist ein vielversprechender Ansatz.[81]

(i) Die eine Variante ist dabei die (klassische) **induktive Statistik**. Charles Sanders Peirce[82], einer der drei großen amerikanischen Pragmatisten (neben William James und John Dewey), entwickelte die Grundlagen dieser Theorie 1868 in *A Theory of Probable Inference*. Im 20. Jh. wurde sie von Ronald Fisher und Jerzy Neyman (u. a) erneut entdeckt, und ein sehr großer (und wachsender) Anteil dessen, was wir zu wissen glauben[83], fußt auf diesem Ansatz. Dessen Grundidee ist einfach: Wir möchten wissen, ob das Ergebnis eines Versuchs (E) einfach Zufall ist, und fragen: Wenn es

[81] Ein neuer Versuch ist die Ranking-Theorie, die mit schwächeren Annahmen auskommt.
[82] Aussprache: 'Pörs'.
[83] Warum die Formulierung "… zu wissen *glauben*"? Googlen Sie bitte "Replikationskrise".

Zufall ist, wie (un)wahrscheinlich ist es? Man interessiert sich also für $\mathcal{P}$(Ergebnis | Zufall), und möchte schließen:

> Wenn es ein Zufall ist, dann ist es ein sehr großer Zufall, *oder:*
> Entweder es ist kein Zufall, oder etwas sehr Unwahrscheinliches ist passiert.

(ii) Der andere wichtige Ansatz die die sog. **Bayes'sche Theorie**, benannt nach dem Satz von Bayes (nach Thomas Bayes, dessen Aufsatz *An Essay Towards Solving a Problem in the Doctrine of Chances* 1763 posthum veröffentlicht wurde). Den Satz von Bayes ist einfach ein Zusammenhang zwischen vier Wahrscheinlichkeiten:

$\mathcal{P}(\text{Theorie} \mid \text{Erfahrung}) \longleftrightarrow \mathcal{P}(\text{Erfahrung} \mid \text{Theorie})$, $\mathcal{P}(\text{Theorie})$ und $\mathcal{P}(\text{Erfahrung})$.

Die linke Seite ist interessant: Wie wahrscheinlich ist eine Theorie, gegeben man macht bestimmte Erfahrungen? Absolut kann man es nicht sagen, aber man kann sagen, ob und wann bestimmte Erfahrungen eine Theorie 'bestätigen', diese 'wahrscheinlicher machen' – wenn etwa gilt:

$\mathcal{P}(\text{Theorie} \mid \text{(neue) Erfahrung}) > \mathcal{P}(\text{Theorie})$.

Beide Ansätze beruhen auf bedingten Wahrscheinlichkeiten.

7.4. Zweimaliges Ziehen

Eine wichtiger Kniff bei der Anwendung der Wahrscheinlichkeitstheorie beruht darauf, daß man durch **wiederholte Versuche** seine Sicherheit oft sehr schnell steigern kann. Macht man nur einen oder zwei Versuche, ist das Ergebnis oft ganz ansehnlich, aber mit wiederholten Versuchen wird das Instrument wirklich mächtig, da sich die Wahrscheinlichkeiten unabhängiger Ereignisse multiplizieren.

Frage: Wie groß wäre die Wahrscheinlichkeit, daß die Bohnen aus dem Sack A sind, wenn Sie zweimal ziehen (der Einfachheit halber: mit Zurücklegen) und 2 Robusta-Bohnen hintereinander erwischen?

Antwort: Sei R_1: die erste Bohne ist eine Robusta; R_2: die zweite Bohne ist eine Robusta.

$$
\begin{aligned}
\mathcal{P}(A \mid R_1 \wedge R_2) &= \mathcal{P}(A \wedge R_1 \wedge R_2) \Big/ \mathcal{P}(R_1 \wedge R_2) && \text{nach Definition} \\
&= \mathcal{P}(R_2 \wedge R_1.A) \Big/ \underbrace{\mathcal{P}(R_2 \wedge R_1)}_{\text{Nenner}} && \text{Kommutativgesetz} \\
&= \mathcal{P}(R_2 \wedge R_1.A) \Big/ \text{Nenner} && \text{Zähler:} \\
&= \mathcal{P}(R_2 \mid R_1 \wedge A)\ \mathcal{P}(R_1 \wedge A) \Big/ \text{Nenner} && \text{nach Definition} \\
&= \mathcal{P}(R_2|A)\ \mathcal{P}(R_1 \wedge A) \Big/ \text{Nenner} && \textbf{(Aufgabe)}
\end{aligned}
$$

$$
\begin{aligned}
&= \underbrace{\mathcal{P}(R_2|A)\ \mathcal{P}(R_1|A)\ \mathcal{P}(A)}_{\text{Zähler}} \Big/ \text{Nenner} && \text{nach Definition}\\
&= \text{Zähler} \Big/ \mathcal{P}(R_1 \wedge R_2) && \text{Nenner:}\\
&= \text{Zähler} \Big/ \mathcal{P}(R_2 \wedge R_1.A) + \mathcal{P}(R_2 \wedge R_1 \wedge B) && [\textbf{Aufgabe}]\\
&= \text{Zähler} \Big/ \mathcal{P}(R_2 \mid R_1 \wedge A)\ \mathcal{P}(R_1 \wedge A)\ +\ \mathcal{P}(R_2 \mid R_1 \wedge B)\ \mathcal{P}(R_1 \wedge B) && \text{s.o.}\\
&= \text{Zähler} \Big/ \mathcal{P}(R_2|A)\ \mathcal{P}(R_1|A)\ \mathcal{P}(A)\ +\ \mathcal{P}(R_2|B)\ \mathcal{P}(R_1|B)\ \mathcal{P}(B) && \text{s.o}\\
&= \frac{\mathcal{P}(R_2|A)\ \mathcal{P}(R_1|A)\ \mathcal{P}(A)}{\mathcal{P}(R_2|A)\ \mathcal{P}(R_1|A)\ \mathcal{P}(A)\ +\ \mathcal{P}(R_2|B)\ \mathcal{P}(R_1|B)\ \mathcal{P}(B)}
\end{aligned}
$$

Alle Ausdrücke sind bekannt:

$$= \frac{1}{4}\ \frac{1}{4}\ \frac{1}{2} \Big/ \left(\frac{1}{4}\ \frac{1}{4}\ \frac{1}{2} + \frac{3}{4}\ \frac{3}{4}\ \frac{1}{2}\right) = \frac{1}{32} \Big/ \left(\frac{1}{32} + \frac{9}{32}\right) = \frac{1}{32}\Big/\frac{10}{32} = \frac{1}{10}.$$

Zieht man 2 Robusta-Bohnen hintereinander, beträgt die Wahrscheinlichkeit, daß diese aus Sack A sind, $\frac{1}{10}$.

Mehrfaches Testen ist oft ein mächtiges Instrument: Bei einem Versuch ist $\mathcal{P}(A \mid R_1) = 0,25$, bei zwei Versuchen schon $\mathcal{P}(A \mid R_1.R_2) = 0,1$.

7.5. Häufigkeiten und Wahrscheinlichkeiten

Es gibt aber einen anderen Weg zur Lösung von Aufgaben, der auf jeden Fall geeignet ist, die Ergebnisse zu überprüfen. Menschen (fast alle) sind schlecht im Denken mit Wahrscheinlichkeiten; mit **Häufigkeiten** tun sie sich oft leichter. Die Angabe $\mathcal{P}(A) = 0,5$ bedeutet ja nur, daß bei 10 Versuchen erwartbar 5 mal A herauskommt, bei 1000 folglich 500 mal.

Betrachten wir noch einmal den ersten Bohnenfall (und ignorieren alle Formeln). Wenn man 1000 mal eine Kaffeebohne zöge, würde man 500 mal Sack A erwischen und 500 mal Sack B.

Im Sack A sind 2 von 8 Bohnen von der Sorte Robusta (25%, $\mathcal{P}(A) = 0,25$). Dies bedeutet, von den 500 Versuchen erwischt man vermutlich in einem Viertel der Fälle eine Robusta-Bohne: Das macht insgesamt absolut 125 Robusta-Bohnen aus Sack A.

Im Sack B sind 300 von 400 Bohnen von der Sorte Robusta (75%, $\mathcal{P}(A) = 0,75$). Von den 500 Versuchen dort erwischt man vermutlich also in drei Viertel der Fälle eine Robusta-Bohne: Das macht insgesamt absolut 375 Robusta-Bohnen aus Sack B.

$$\textbf{1000}\text{ Bohnen}\left\{\begin{array}{l} 50\,\%\text{ Sack A: }\textbf{500}\text{ Bohnen}\left\{\begin{array}{l} 25\,\%\text{ Robusta: }\textbf{125 Robusta} \\ \\ 75\,\%\text{ Arabica: 375 Arabica} \end{array}\right. \\ \\ 50\,\%\text{ Sack B: }\textbf{500}\text{ Bohnen}\left\{\begin{array}{l} 75\,\%\text{ Robusta: }\textbf{375 Robusta} \\ \\ 25\,\%\text{ Arabica: 115 Arabica} \end{array}\right. \end{array}\right.$$

Insgesamt hat man $125 + 375 = 500$ Robusta-Bohnen erwischt; von denen sind 125 oder ein Viertel aus Sack A, mit anderen Worten: $\mathcal{P}(A \mid R) = 0,25$.[84]

[84] Wenn Sie jetzt noch große oder krumme Zahlen vermeiden möchten: Betrachten Sie einfach die einzelnen bedingten Wahrscheinlichkeiten einmal in der Bruchschreibweise, also z. B. statt 25% $\frac{1}{4}$, und multiplizieren Sie die Nenner der einzelnen sich ausschließenden Wege: Bei Sack A haben sie $\frac{1}{2}$ und $\frac{1}{4}$ (Wahl von Sack A und Ziehen von Bohne); hier reichen $2 \cdot 4$ Bohnen. Bei Sack B haben sie $\frac{1}{2}$ und $\frac{3}{4}$, also ebenfalls 8. Sie könnten also im Häufigkeitsbild über 8 Bohnen nachdenken: 4 in Sack A, davon eine Robusta; vier in Sack B, davon 3 Robusta. Von 4 Robusta-Bohnen also eine (25%) aus Sack A.

7. Die Axiome von Kolmogorov

8.1. Axiome von Kolmogorov

Annahmen: Seien A und B Aussagen. Dann sind auch die Negation $\neg A$, die Konjunktion $A \wedge B$ und die Disjunktion $A \vee B$ Aussagen. Und:[85]

Sind A und B logisch äquivalent, dann ist $\mathcal{P}(A) = \mathcal{P}(B)$. **(AL)**

Die Axiomatisierung der Wahrscheinlichkeitstheorie geht auf Andrei Kolmogorovs *Grundbegriffe der Wahrscheinlichkeitsrechnung* von 1933 zurück.[86] Es gibt drei Axiome:

Axiom 1. Die Wahrscheinlichkeit einer jeden Aussage ist eine reelle Zahl zwischen 0 und 1:

$0 \leq \mathcal{P}(A) \leq 1.$ **(A1)**

Axoim 2. Die gewisse Aussage (Tautologie) Ω hat die Wahrscheinlichkeit 1:

$\mathcal{P}(\text{Tautologie}) = \mathcal{P}(\Omega) = 1.$ **(A2)**

Axiom 3. Die Wahrscheinlichkeit einer Disjunktion abzählbar vieler sich ausschließender (inkompatibler, unvereinbarer) Aussagen ist gleich der Summe der Wahrscheinlichkeiten der einzelnen Aussagen (**Additivität**). Aussagen heißen inkompatibel, wenn sie paarweise inkompatibel sind.

Sind A und B wechselseitig ausschließend, dann gilt: $\mathcal{P}(A \vee B) = \mathcal{P}(A) + \mathcal{P}(B)$. **(A3)**

Bewundern Sie einen Augenblick die Schönheit des Ganzen. Diese drei Axiome enthalten die Wahrscheinlichkeitstheorie. Dabei ist das erste eigentlich willkürlich (man

[85] Dies ließe sich eigentlich aus den Axiomen beweisen; Michael D. Resnik, *Choices: An Introduction to Decision Theory*, Minneapolis & London 1987, S. 50.

[86] Andrei Nikolajewitsch Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin: Julius Springer, 1933.

könnte auch Prozent-Angaben verwenden); das zweite sagt nur, wo bei 0 und 1 "oben" ist und wo "unten"; nur das dritte Axiom, die Additivität, ist ein spezifischere Aussage (ist aber auch intuitiv plausibel). Dennoch folgen aus diesen drei Axiomen alle komplexeren Aussagen der Wahrscheinlichkeitstheorie, z. B. der Satz von Bayes.

8.2. Folgerungen und Satz von Bayes

Vorbemerkung: Den Beweisen zu den folgenden Aussagen liegen zwei Arten von 'Beweisideen' zugrunde: Entweder, man wendet die Axiome A1, A2 und A3 an, oder, man beginnt mit einer klugen aussagenlogischen Idee ("AL"), wie etwa: "A gdw. $\Omega \wedge A$".

(i) Nützlicher Zusammenhang:

$$\mathcal{P}(A) = \mathcal{P}(A \wedge B) + \mathcal{P}(A \wedge \neg B). \tag{R4}$$

Beweis:

$$\begin{aligned} \mathcal{P}(A) &= \mathcal{P}(A \wedge B \vee A \wedge \neg B) && \text{AL} \\ &= \mathcal{P}(A \wedge B) + \mathcal{P}(A \wedge \neg B) \; \blacksquare && \text{A3} \end{aligned}$$

(ii) Komplementäre Aussagen haben komplementäre Wahrscheinlichkeiten:

$$\mathcal{P}(\neg A) = 1 - \mathcal{P}(A). \tag{R4}$$

Beweis:

$$\begin{aligned} \mathcal{P}(\neg A) &= \mathcal{P}(\Omega \wedge \neg A) && \text{AL} \\ &= \mathcal{P}(\Omega) - \mathcal{P}(\Omega \wedge A) && \text{R4} \\ &= 1 - \mathcal{P}(\Omega \wedge A) && \text{A2} \\ &= 1 - \mathcal{P}(A) \; \blacksquare && \text{AL} \end{aligned}$$

(iii) Disjunktionen überlappender Aussagen:

$$\mathcal{P}(A \vee B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \wedge B) \tag{R6}$$

Beweis:

$$\begin{aligned} \mathcal{P}(A \vee B) &= \mathcal{P}(A \wedge \neg B \vee \neg A \wedge B \vee A \wedge B) && \text{AL} \\ &= \mathcal{P}(A \wedge \neg B) \; + \; \mathcal{P}(\neg A \wedge B) \; + \; \mathcal{P}(A \wedge B) && \text{A3} \\ &= \overbrace{\mathcal{P}(A) - \mathcal{P}(A \wedge B)}^{\mathcal{P}(A \wedge \neg B)} \; + \; \overbrace{\mathcal{P}(B) - \mathcal{P}(A \wedge B)}^{\mathcal{P}(\neg A \wedge B)} \; + \; \mathcal{P}(A \wedge B) && \text{R3} \\ &= \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \wedge B) \; \blacksquare \end{aligned}$$

(iv) Definition der bedingten Wahrscheinlichkeit:

$$\mathcal{P}(A \mid B) = \mathcal{P}(A \wedge B) \;/\; \mathcal{P}(B) \qquad \mid \mathcal{P}(B) \neq 0 \tag{R7}$$

Anmerkung: Die bedingte Wahrscheinlichkeit genügt ebenfalls den Kolmogorow-Axiomen: Seien $\mathcal{P}(B) \neq 0$ und A_1 und A_2 einander ausschließende Aussagen, so gilt:

$$0 \leq P(A \mid B) \leq 1$$

$$\mathcal{P}(\Omega \mid B) = 1$$

$$\mathcal{P}(A_1 \vee A_2 \mid B) = \mathcal{P}(A_1|B) + \mathcal{P}(A_2|B).$$

Beweis der letzten Aussage:

$$\begin{aligned}
\mathcal{P}(A_1 \vee A_2 \mid B) &= \mathcal{P}\big((A_1 \vee A_2) \wedge B\big) \;/\; \mathcal{P}(B) && \text{R7}\\
&= \mathcal{P}\big((A_1 \wedge B) \vee (A_2 \wedge B)\big) \;/\; \mathcal{P}(B) && \text{AL}\\
&= \big(\mathcal{P}(A_1 \wedge B) + \mathcal{P}(A_2 \wedge B)\big) \;/\; \mathcal{P}(B) && \text{A3}\\
&= \mathcal{P}(A_1 \wedge B) \;/\; \mathcal{P}(B) \;+\; \mathcal{P}(A_2 \wedge B)/\mathcal{P}(B) &&\\
&= \mathcal{P}(A_1 \mid B) \;+\; \mathcal{P}(A_2 \mid B) \;\blacksquare && \text{R7}
\end{aligned}$$

Motivation: Geht man von Zufallsapparaturen aus und den erwartbaren relativen Häufigkeiten von deren Ausgängen (wie wir bislang), sind kategorische Wahrscheinlichkeitsaussagen gut motiviert. Aber man könnte auch bedingte Wahrscheinlichkeitsaussagen als basaler oder zentraler betrachten, etwa dann, wenn man bei Wahrscheinlichkeitsaussagen weniger Zufallsapparaturen denkt als vielmehr an sog. subjektive Wahrscheinlichkeiten (Aussagen darüber, für wie glaubhaft jemand eine Aussage hält) – dann erscheinen kategorische Wahrscheinlichkeitsaussagen oft willkürlich, bedingte hingegen nicht (s.u., Vorlesung 10).

(v) Konjunktion/Verbundwahrscheinlichkeit:

$$\mathcal{P}(A \wedge B) = \mathcal{P}(A|B) \cdot \mathcal{P}(B) \qquad \mid \mathcal{P}(B) \neq 0 \tag{R8}$$

Beweis: R7.

(vi) Gesetz der totalen Wahrscheinlichkeit: Die 'totale' (nicht-bedingte) Wahrscheinlichkeit $\mathcal{P}(A)$ ergibt aus der bedingten $\mathcal{P}(A|B)$ und der 'bedingenden' $\mathcal{P}(B)$ als:

$$\mathcal{P}(A) = \mathcal{P}(A|B) \cdot \mathcal{P}(B) + \mathcal{P}(A|\neg B) \cdot \mathcal{P}(\neg B) \qquad \mid \mathcal{P}(B) \neq 0, \mathcal{P}(\neg B) \neq 0 \tag{R9}$$

Beweis s. **Aufgabe**. – Dieser Zusammenhang ist durchaus plausibel:

$$\mathcal{P}(\text{Reichtum}) = \mathcal{P}(\text{Reichtum} \mid \text{Lotto}) \cdot \mathcal{P}(\text{Lotto}) +$$

$$+\, \mathcal{P}(\text{Reichtum} \mid \neg\, \text{Lotto}) \cdot \mathcal{P}(\neg\, \text{Lotto}).$$

Die Wahrscheinlichkeit, reich zu werden, entspricht der Wahrscheinlichkeit, durch eine Lotterie reich zu werden, insoweit man Lotto spielt, plus der Wahrscheinlichkeit, auf anderem Wege reicht zu werden, insofern man das Lottospielen läßt.

(vii) Logische Folgerung: Folgt A aus B, gilt:

$$\mathcal{P}(B) \leq \mathcal{P}(A) \qquad \mid A \text{ folgt logisch aus } B \qquad \textbf{(R10)}$$

Beweis:

$$\begin{aligned} \mathcal{P}(B) &= \mathcal{P}(B \wedge A) && \text{AL} \\ &= \mathcal{P}(A) - \underbrace{\mathcal{P}(A \wedge \neg B)}_{\leq 0} && \text{R3} \\ &\leq \mathcal{P}(A) \;\blacksquare && \text{A1} \end{aligned}$$

(viii) Statistische Unabhängigkeit: Die Aussagen A und B sind statistisch unabhängig gdw.

$$\mathcal{P}(A \mid B) = \mathcal{P}(A). \qquad \mid \mathcal{P}(A) \neq 0, \mathcal{P}(B) \neq 0 \qquad \textbf{(R11)}$$

Für statistisch unabhängige Aussagen bzw. Ereignisse gilt:

Wenn $\mathcal{P}(A \mid B) = \mathcal{P}(A)$, dann $\mathcal{P}(B \mid A) = \mathcal{P}(B)$.

Beweis:

$$\begin{aligned} \mathcal{P}(B \mid A) &= \mathcal{P}(B \wedge A) \,/\, \mathcal{P}(A) && \text{R7} \\ &= \mathcal{P}(B \wedge A) \,/\, \mathcal{P}(A \mid B) && \text{R11} \\ &= \mathcal{P}(B \wedge A) \,/\, \big(\mathcal{P}(A \wedge B) / \mathcal{P}(B)\big) && \text{R7} \\ &= \big(\mathcal{P}(B \wedge A) / \mathcal{P}(A \wedge B)\big) \cdot \mathcal{P}(B) && \\ &= \big(\mathcal{P}(A \wedge B) / \mathcal{P}(A \wedge B)\big) \cdot \mathcal{P}(B) && \text{AL} \\ &= \mathcal{P}(B) \;\blacksquare && \text{AL} \end{aligned}$$

(ix) Der Satz von Bayes (Bayes's Rule): Der sog. Satz von Bayes erlaubt es, viele Wahrscheinlichkeitsprobleme direkt zu lösen, ohne immer neu von den elementareren Sätzen auszugehen. Im Rahmen der sog. Bayes'schen Erkenntnistheorie hat er eine zentrale Stellung (s.u., Vorlesung 10).

Kurze Form: Seien H und E Aussagen und $\mathcal{P}(E) \neq 0$. Dann gilt:

$$\mathcal{P}(H \mid E) = \mathcal{P}(H)\,\mathcal{P}(E \mid H) \;/\; \mathcal{P}(E) \qquad \textbf{(Satz von Bayes)}$$

Beweis:

$$\mathcal{P}(H \mid E) = \mathcal{P}(H \wedge E) \;/\; \mathcal{P}(E) \qquad \text{R7}$$

$$= \mathcal{P}(E \wedge H) \;/\; \mathcal{P}(E) \qquad \text{AL}$$

$$= \mathcal{P}(H)\,\mathcal{P}(E|H) \;/\; \mathcal{P}(E) \;\blacksquare \qquad \text{R7}$$

Längere Form:

$$\mathcal{P}(H \mid E) = \mathcal{P}(H)\,\mathcal{P}(E|H) \;\Big/\; \underbrace{\big(\,\mathcal{P}(E|H)\,\mathcal{P}(H) + \mathcal{P}(E|\neg H)\,\mathcal{P}(\neg H)\,\big)}_{\mathcal{P}(E)\ \text{(totale Wahrscheinlichkeit: R9)}} \qquad \textbf{(Satz von Bayes)}$$

Oder noch allgemeiner für n wechselseitig ausschließende gemeinsam ausschöpfende Aussagen (bzw. Ereignisse) $H_i,\ i = 1, ..., n$, mit $\mathcal{P}(H_i) \neq 0$:

$$\mathcal{P}(H_j \mid E) = \mathcal{P}(H_j) \cdot \mathcal{P}(E \mid H_j) \quad / \quad \textstyle\sum_{i=1}^{n} \mathcal{P}(E \mid H_i) \cdot \mathcal{P}(H_i) \qquad \textbf{(Satz von Bayes)}$$

8.3. Der Basisratenfehlschluß

Die Psychologen Amos Tversky und Daniel Kahneman führten in den 80er Jahren folgenden Versuch durch: Sie stellen Probanden folgende Frage:

> A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
>
> (a) 85% of the cabs in the city are Green and 15% are Blue.
>
> (b) a witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.
>
> What is the probability the the cab involved in the accident was Blue rather than Green? [87]

Die meisten vermuten, daß es wahrscheinlich ein blaues Taxi gewesen sei. Es ist aber (etwas) wahrscheinlicher, daß ein Taxi der Firma Green den Unfall gebaut hat:

$\mathcal{P}(\text{Blue}) = 0,15.$

$\mathcal{P}(\neg\,\text{Blue}) = \mathcal{P}(\text{Green}) = 0,85.$

$\mathcal{P}(\text{“Es war blau!”} \mid \text{Blue}) = 0,8.$

$\mathcal{P}(\text{“Es war blau!”} \mid \neg\,\text{Blue}) = \mathcal{P}(\text{“Es war blau!”} \mid \text{Green}) = 0,2.$

$\mathcal{P}(\text{Blue} \mid \text{“Es war blau!”}) =$

[87] Vgl. Daniel Kahnemann et al. (edd.), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge 1982, S. 156–157.

$$= \frac{\mathcal{P}(\text{Blue}) \cdot \mathcal{P}(\text{“Es war blau!”} \mid \text{Blue})}{\mathcal{P}(\text{“Es war blau!”} \mid \text{Blue})\,\mathcal{P}(\text{Blue}) + \mathcal{P}(\text{“Es war blau!”} \mid \neg\text{Blue})\,\mathcal{P}(\neg\text{Blue})}$$
$$= \frac{0,15 \cdot 0.8}{0,15 \cdot 0.8 + 0,85 \cdot 0,2}$$
$$= \frac{\frac{3}{20} \cdot \frac{4}{5}}{\frac{3}{20} \cdot \frac{4}{5} + \frac{17}{20} \cdot \frac{1}{5}} = \frac{\frac{12}{100}}{\frac{12}{100} + \frac{17}{100}} = \frac{12}{29}.$$

Der Name “Basisraten-Fehlschluß” (base-rate fallacy) stammt daher, daß Tversky und Kahneman behaupteten, Leute ignorierten die sog. Basisrate, d. h. die Informationen über die Häufigkeiten der beiden Taxi-Firmen: $\mathcal{P}(\text{Blue})$ und $\mathcal{P}(\text{Green})$.

8.4. Der Harvard Medical School Test

Eine weitere berühmte Anwendung des Satzes von Bayes liegt in der Interpretation der Ergebnisse von diagnostischen Tests in der Medizin. Unter der **Sensitivität** (sensitivity) eines diagnostischen Test versteht man dessen Fähigkeit, eine Krankheit anzuzeigen (also ein ‘positives’ Testresultat zu geben, wenn jemand eine bestimmte Krankheit hat). Eine hohe Sensitivität ist allein nicht genug für einen aussagekräftigen Test: Ein Test, der immer ‘positive’ Testergebnisse liefert, würde alle Kranken identifizieren, aber auch alle Gesunden. Daher verwendet man noch ein zweites Testcharakteristikum, die **Spezifizität** (specificity) des Tests; sie bezeichnet die Fähigkeit, Gesunde zu identifizieren, also ein ‘negatives’ [88] Testergebnis zu liefern, wenn ein Patient die jeweilige Krankheit nicht hat.

Aber auch ein positives Testergebnis eines Tests mit hoher Sensitivität und hoher Spezifizität ist nicht eindeutig zu interpretieren: In den 70er Jahren stellten W. Casscells, A. Schoenberger und T. B. Graboys 60 Angestellten und Studenten der Harvard Medical School folgende Frage:

> If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person’s symptoms or signs? [89]

Eine “false positive rate” von 5% bedeutet, daß der Test nur in 5% der Fälle bei Gesunden ein positives Ergebnis liefert, also eine hohe Spezifizität hat. “Prevalence” (Prävalenz) bezeichnet die Verbreitung einer Krankheit in einer Bevölkerung.

Fast die Hälfte der Teilnehmer gaben die Antwort 95%; nur 11 Teilnehmer gaben die ‘korrekte’ (vergleiche aber die Anmerkung unten) Antwort: Aufgrund der geringen Verbreitung der Krankheit ist ein positives Testergebnis höchstwahrscheinlich (2%)

[88] Die Ausdrücke ‘positiv’ und ‘negativ’ stehen hier in Anführungszeichen, da sie zu Verwechslungen einladen: Ein ‘positives’ Testergebnis ist für den Patienten natürlich nicht erfreulich.

[89] W. Casscells, A. Schoenberger & T.B. Graboys: Interpretation by Physicians of Clinical Laboratory Results. *New England Journal of Medicine* 299 (18), 1978, S. 999–1001; hier S. 999.

ein falsch-positives:

$$\textbf{1001 Leute}\begin{cases} \textbf{1 krank}\begin{cases}\textbf{1 positives Testergebnis (`richtig positive')}\\ \text{0 negative Testergebnisse}\end{cases}\\ \text{1000 gesund}\begin{cases}\textbf{50 positive Testergebnisse (`falsch positive')}\\ \text{950 negative Testergebnisse}\end{cases}\end{cases}$$

Auch hier könnte man sagen, das Problem bestehe darin, die 'Basisrate' (hier: die Prävalenz) zu vernachlässigen. (Anmerkung: Die Prävalenz sollte freilich nur dann mit der Basisrate gleichgesetzt werden, wenn die Patienten 'zufällig' aus der Bevölkerung gewählt sind – dies sind Patienten selten (man geht ja nicht ohne Anlaß zum Arzt). Daß wir nichts näheres über Symptome oder Krankheitsanzeichen wissen, stellt jedenfalls nicht sicher, daß die Auswahl 'zufällig' ist.)

8.5. Der Konjunktionsfehlschluß (das "Linda problem")

Die Psychologen Amos Tversky und Daniel Kahneman führten in den 80er Jahren auch folgenden Versuch durch (das sog. "Linda problem"):

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discriminiation and social justice, and also participated in anti-nuclear demonstrations.
>
> Linda is a teacher in elementary school.
>
> Linda works in a bookstore and takes Yoga classes.
>
> Linda is active in the feminist movement. (F)
>
> Linda is a psychiatric social worker.
>
> Linda is a member of the League of Women Voters.
>
> Linda is a bank teller. (T)
>
> Linda is an insurance salesperson.
>
> Linda is a bank teller and is active in the feminist movement. (T&F)
>
> A group of 88 undergraduates at [University of British Columbia] ranked the eight statements … by "the degree to which [Linda] resembles the typical member of that class." … More surprising and less acceptable is the finding that the great majority of subjects also rank the conjunctions (… T&F) as more *probable* than their less representative constituents (… T)." [90]

[90] Amos Tversky & Daniel Kahneman: Extensional Versus Intuitive Reasoning: The Conjunction

Frage: Warum finden Tversky und Kahneman dies überraschend und nicht hinnehmbar?

Fallacy in Probability Judgment. *Psychological Review* 90 (4), 1983, S. 293–315; hier S. 297. Vergleichen Sie die Bemerkungen bei Hacking, S. 66.

9. Erwartungswerte und Entscheidungstheorie

9.1. (Frühe) Anwendungen der Wahrscheinlichkeitstheorie

Im 18. und 19. Jahrhundert findet sich kaum eine (mathematische) Darstellung oder Abhandlung der Wahrscheinlichkeitstheorie, in der diese nicht auch durch ihren Nutzen bei der Beurteilung von Fällen motiviert würde, die in der Praxis wichtig sind. Der Mathematiker Pierre-Simon de Laplace etwa betrachtet in seinem *Philosophischen Versuch über Wahrscheinlichkeiten* (1814)[91] auch Erwartungswerte (s.u. [92]) sowie als Anwendungen u. a. Glücksspiele, die Glaubhaftigkeit sog. **testimonialen Wissens** (d.i. des Wissens, das nicht durch Erfahrung oder Nachdenken erworben ist, sondern auf Aussagen anderer Menschen beruht) und Wahlen sowie die Wahl der geeigneten Größe von Gerichtshöfen.

Ein vielfach diskutiertes Beispiel betrifft die Frage der Glaubhaftigkeit von Berichten über außergewöhnliche Ereignisse oder Tatsachen, und besonders die von Berichten über Wunder; viele beziehen sich dabei ausdrücklich auf David Humes Behandlung dieser Frage in *Of Miracles*[93]. Hume hatte behauptet (sinngemäß), Wunder seien (definitionsgemäß) sehr unwahrscheinlich, und Wunderberichte seien unglaubhaft, außer, daß die Berichte falsch sind, wäre das größere Wunder (noch unwahrscheinlicher).

Charles Babbage gibt folgende Analyse (*Nineth Bridgewater Treatise*, 1837, Ch. 10,

[91] Pierre-Simon Laplace, *Essai Philosophique sur les Probabilités*, Paris 1814; eigentlich eine Einführung zur *Théorie Analytique des Probabilités* von 1812. Im folgenden zitiert nach der Ausgabe: Pierre-Simon Laplace, *Philosophical Essay on Probabilities:* Translated from the fifth French edition of 1825 (ed. Andrew Dale). New York 1995.

[92] Ibid., S. 11–14: On expectation.

[93] David Hume, Of Miracles, in id., *An Enquiry Concerning Human Understanding* (ed. Tom Beauchamp). Oxford 1999 [1748].

bes. S. 128–132):

> Die Wahrscheinlichkeit, daß jemand von den Toten aufersteht, betrage 200.000.000.000 zu 1.[94]
>
> Es gebe Zeugen, A und B, einer Auferstehung, die nur in 1 von 100 Fällen lügen.
>
> Die Zeugen seien unabhängig voneinander.

Dann ergebe sich:

99	×	99	=	9801	cases in which A and B agree in truth,
1	×	99	=	99	cases in which B is true and A false,
99	×	1	=	99	cases in which a is true and B false,
1	×	1	=	1	case in which A and B agree in a falsehood.
				10,000	cases.

Wenn beide Zeugen also übereinstimmen, dann nur in 1 von 10 000 Fällen, wenn beide lügen: "It follows, then, that the improbability of the falsehood of the concurring testimony of only *six* such independent witnesses, is already *five times* as great as the improbability against the miracle of a dead man's being restored to life, deduced from Hume's method of estimating its probability solely from experience." [95]

Die Betrachtung von Babbage illustriert einige interessante Punkte: (i) Die Mächtigkeit von mehreren **statistisch unabhängigen** Ereignissen oder Aussagen (vgl. Sie die Abschnitte 6.4 und 7.4). (ii) Man muß sich schon fragen, woher eigentlich die Wahrscheinlichkeitsaussagen in der **Prämissen** kommen (etwa die Glaubhaftigkeit der Zeugen). (iii) Und man muß sich immer fragen, ob das **Wahrscheinlichkeitsmodell** eigentlich stimmt (findet man wirklich 6 unabhängige Zeugen eines Ereignisses? Man könnte ja auch andersherum argumentieren und sagen, daß sich 6 Zeugen in so einer Frage einig sind, spricht dafür, daß sie eben nicht unabhängig sind).[96]

9.2. Entscheidungen und Entscheidungstheorie

Wir wenden uns **Handlungen** zu. Diese werden in der **Entscheidungstheorie** (**decision theory**) [97] untersucht, die man nach der Ausgangsfrage in zwei Bereiche unterteilt: Die **deskriptive Entscheidungstheorie** untersucht, wie Menschen tatsächlich handeln. Die **normative Entscheidungstheorie** untersucht hingegen, wie

[94] Nämlich $1/20\,(100)^5$ (ibid., S. 131).
[95] Ibid., S. 131.
[96] Vgl. zu den Modellen C. Menke, Der Nutzen probabilistischer Modelle testimonialen Wissens, in: C. Spoerhase et al. (edd.), *Unsicheres Wissen. Skeptizismus und Wahrscheinlichkeit 1550–1850.* Berlin/New York 2009, S. 301–316.
[97] Vgl. Hacking, ch. 8–10. Eine gute Darstellung und Diskussion der Entscheidungstheorie mit Schwerpunkt auf philosophischen Fragestellungen gibt auch Michael D. Resnik, Choices: An Introduction to Decision Theory, Minneapolis/London 1987.

Menschen rationalerweise handeln sollten – gemeint ist damit: zweckrational, d. h. so, daß man am ehesten bekommt, was man will. (Diesen Ansatz muß man nicht für zwingend halten: Ein Mensch mit moralischem Empfinden könnte gut Handlungen bevorzugen, die – scheinbar – zu seinem Nachteil sind. Aber man könnte auch sagen, ein solcher Mensch sei einfach nicht-eigennützig, und wolle etwas anderes als ein Egoist, und um dies zu erreichen, solle er oder sie sich dann eben doch (zweck-)rational verhalten.)

Deskriptive und normative Entscheidungstheorie sind nicht unabhängig voneinander: Nur weil Menschen nicht immer zweckrational handeln, gibt es ja überhaupt eine normative Entscheidungstheorie (sonst wäre sie überflüssig); und Wissen darüber, wie Menschen tatsächlich handeln, kann wichtig sein für die Frage sein, wie man selbst handeln sollte. (Beispiele?)

Der **Ansatz der Entscheidungstheorie** ist: Handlungen haben (sichere oder wahrscheinliche oder aber unberechenbare) Folgen, und diese Folgen sind gut oder schlecht: sie haben einen Wert oder Nutzen:

Handlung $\longrightarrow$ Folgen $\longrightarrow$ Wert/Nutzen.

Sprachgebrauch: In der Entscheidungstheorie spricht man i. d. R. von Entscheidungen von **Akteuren** (**agents**).

Entscheidungen. Eine kurze Übersicht zur Einordnung: Man unterscheidet in der Entscheidungstheorie zwischen **Entscheidungen unter Sicherheit (decision under certainty)** – wenn die Folgen des Handelns bekannt (vorhersagbar) sind, **Entscheidungen unter Risiko (decisions under risk)** – wenn die Wahrscheinlichkeiten der Folgen der jeweiligen Handlungen bekannt sind – und schließlich **Entscheidungen unter Ungewißheit (decision under ignorance/uncertainty)** – wenn die Wahrscheinlichkeiten unbekannt sind oder nicht sinnvoll angebbar sind. Im letzten Fall spricht man oft ‘Knightscher Ungewißheit’ (Knightian uncertainty), nach dem Ökonomen Frank Knight, der 1921 in *Risk, Uncertainty and Profit* zwischen Uncertainty und Risk begrifflich unterschieden hatte.[98] Für uns sind im Kontext der Wahrscheinlichkeitstheorie besonders Entscheidungen unter Risiko interessant.

[98] Frank Knight war ein amerikanischer Ökonomen und Mitbegründer der Chicago School. In *Risk, Uncertainty and Profit* heißt es:

“But Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated. [...] It will appear that a *measurable* uncertainty, or ‘risk’ proper, as we shall use the term, is so far different from an *unmeasurable* one that it is not in effect an uncertainty at all.” Frank Knight, *Risk, Uncertainty and Profit.* Boston/New York 1921, S. 19–20; Hervorhebung im Original.

$$\textbf{Entscheidungen unter} \begin{cases} \textbf{Sicherheit} \text{ (Wahrscheinlichkeiten} = 1\text{)} \\ \textbf{Risiko} \text{ (Wahrscheinlichkeiten bekannt)} \\ \textbf{Ungewißheit} \text{ (Wahrscheinlichkeiten unbekannt)} \end{cases}$$

Abb. 9.1. Frank Knights Klassifikation von Entscheidungen (1921).

9.3. Die *Logik von Port Royal*

Ein klassisches Lehrbuch der neuzeitlichen Logik ist die sog. **Logique de Port Royal**[99] (1662) von Antoine Arnauld und Pierre Nicole; der eigentliche Titel ist *La logique ou l'art de penser*.[100] Der Ansatz hier ist (wie man heute sagen würde) ‘psychologisch’ – behandelt wird “die Kunst, seine Vernunft gut zu leiten”. In der Logik von Port Royal findet sich die klassische Untergliederung logischer Abhandlungen: Sie behandelt 1. Begriffe, 2. Urteile, 3. Schlüsse/Argumente und 4. Methodenlehre. Das letzte (16.) Kapitel des 4. Teils behandelt “Urteile, die wir hinsichtlich zukünftiger Ereignisse treffen sollten”. Darin stellen Arnauld und Nicole fest, daß Menschen sich mit Blick auf Ereignisse, an denen sie beteiligt sind und die sie mit Hoffnungen oder Ängsten verknüpfen, vielfach falsch beurteilten: Sie betrachteten allein den möglichen Nutzen oder Schaden, nicht aber die Wahrscheinlichkeit, mit der diese einträten: sie griffen zu allen Vorsichtsmaßnahmen, um Leben und Vermögen zu sichern, und spielten Lotto mit Blick auf den möglichen hohen Gewinn.

Ein **faires Spiel** bestimmen Arnauld und Nicole als eines, in welchem sich Einsatz und Gewinnerwartung die Waage hielten, in dem Sinn, daß man eigentlich (langfristig) nicht erwarten kann, etwas zu gewinnen oder zu verlieren: Setzen 10 Personen jeweils eine Krone ein, und einer der Spieler gewinnt alles, kann zwar jeder 9 Kronen gewinnen (Nutzen), aber es sei auch neun mal wahrscheinlicher, zu verlieren als zu gewinnen. – Diesen Ansatz der Logik von Port Royal verfolgt und erweitert die Entscheidungstheorie.

9.4. Nutzen und Erwartungswerte

Kennt man die Wahrscheinlichkeiten (oder glaubt sie zu kennen), mit denen bestimmte Folgen (Konsequenzen) einer Handlung eintreten, und hat eine Meinung dazu, wie gut oder schlecht man diese findet – wie groß deren **Wert** oder **Nutzen** (**utility**) ist –, dann kann man den **Erwartungswert** (**expected value**) folgendermaßen bestimmen:

$$\text{Erwartungswert (Handlung)} = \mathcal{P}(\text{Folgen} \mid \text{Handlung}) \cdot \text{Wert (Folgen)},$$

[99] Nach Port Royal des Champs, einem Frauenkloster des Zisterzienserordens bei Versailles, dessen Äbtissin, Anqélique Arnauld, die Schwester Antoine Arnaulds war.

[100] Antoine Arnauld, Pierre Nicole, *Logic or the Art of Thinking* (ed. Jill Buroker). Cambridge 1996. Die Übersetzungen folgen (frei) dieser Ausgabe.

oder, mit den Abkürzungen $E(H)$ für den Erwartungswert einer Handlung und $U(F)$ für den Nutzen einer Folge der Handlung:

$$E(H) = \mathcal{P}(F \mid H) \cdot U(F).$$

Hat eine Handlung mehrere verschiedene Folgen (ohne 'Überlappungen'), summiert sich deren erwarteter Wert:

$$E(H) = \sum_i \mathcal{P}(F_i \mid H) \cdot U(F_i).$$

Diese Definition ist sehr geradeheraus und und sollte auch mathematisch keine Schwierigkeiten aufwerfen. Damit läßt sich die Bestimmung eines fairen Spiels der Logik von Port Royal formal fassen: Ein faires Spiel ist eines, in dem der Erwartungswert des Gewinns dem Einsatz entspricht.

9.5. Das St. Petersburg-Spiel

Das St. Petersburg-Spiel ist das folgende: Eine faire Münze wird geworfen, solange, bis sie auf KOPF landet. Geschieht dies beim ersten Wurf, erhalten Sie 2 EUR; geschieht dies beim zweiten Wurf, erhalten Sie 4 EUR; geschieht dies beim dritten Wurf, erhalten Sie 2^3 EUR, und so weiter. Fällt KOPF beim n-ten Wurf, erhalten Sie 2^n EUR.

Was wäre der faire Preis, um bei diesem Spiel mitzuspielen?

$$E(\text{St. Petersb.-Spiel}) = \sum_{i=1}^{\infty} \underbrace{\left(\frac{1}{2^1}\right) 2^1\,\text{EUR}}_{1\,\text{EUR}} + \underbrace{\left(\frac{1}{2^2}\right) 2^2\,\text{EUR}}_{1\,\text{EUR}} + \ldots + \underbrace{\left(\frac{1}{2^i}\right) 2^i\,\text{EUR}}_{1\,\text{EUR}} = \infty\,\text{EUR}$$

Kurz: Jeder Preis sollte Ihnen recht sein! (Sollte er vermutlich nicht.) Irgendetwas stimmt bei diesem sog. **St. Petersburg-Paradox** nicht. (Als **Paradox** oder Paradoxon bezeichnet man eine Menge jeweils für sich plausibler, aber gemeinsam unverträglicher Aussagen.) Aber eine allgemein akzeptierte 'Lösung' (was ein fairer Preis wäre, bzw. worin genau überhaupt das Problem besteht) gibt es leider nicht:

(i) Vielleicht ist für solche Fälle einfach kein Erwartungswert definierbar. (Aber warum nicht?)

(ii) Oder es findet sich in der wirklichen Welt keine Bank, die genug Geld hätte, um das St. Petersburg-Spiel zu finanzieren? (Aber 20 Runden könnte eine Bank finanzieren; dies wären EUR 1.048.576. Wäre dies ein fairer Einsatz?)

(iii) Oder das Problem besteht darin, daß der sog. **Grenznutzen** abnimmt: Irgendwann, wenn man viel Geld hat, hat noch mehr Geld wenig zusätzlichen Wert. Vielleicht wird der Wert sogar negativ (man muß das Geld verwalten und sichern). (Aber beides gilt kaum bei den ersten 20 Runden.) [101]

[101] Vgl. zu diesem Lösungsversuchen Hacking, S. 91–94; Resnik, S. 107–109.

9.6. Die Bayes'sche Regel

Die Definition der erwartbaren Nutzens läßt sich nun verwenden, um zu helfen zu entscheiden, welche von mehreren Handlungsoptionen man "rationalerweise" wählen sollte: **Wähle die Handlung, die den größten Nutzen verspricht – die mit dem höchsten Erwartungswert**. Diese Entscheidungsregel wird (verwirrenderweise) als **Bayes'sche Regel** (Bayes's rule) bezeichnet.

Sie haben EUR 100 für ein Jahr anzulegen, und genau zwei Anlagemöglichkeiten.

Handlung A: Eine Anlage in Kommunalobligationen verspricht eine jährliche Rendite von 2 %.

Handlung B: Eine Anlage in den *Cosmic Energy Dynamics 2000*-Fonds verspricht eine Rendite von 2,5 % nach einem Jahr, wenn wenn es keine Änderung der Gesetzgebung gibt; sonst kann man nur mit Gewinnen von 1 % rechnen (verspricht Ihnen Ihr Finanzberater). Eine Änderung der Gesetzgebung scheint Ihnen aber nicht sehr wahrscheinlich (20 %).

$$E(A) = \underbrace{\mathcal{P}(\text{Rendite} \mid \text{Obligationen})}_{100\,\%} \cdot \underbrace{(2\,\%\text{-Rendite})}_{102\text{ Euro}} = 102 \text{ Euro.}$$

$$E(B) = \underbrace{\mathcal{P}(\text{Rendite} \mid \text{Fond})}_{80\,\%} \cdot \underbrace{(2{,}5\,\%\text{-Rendite})}_{102{,}5\text{ Euro}} + \underbrace{\mathcal{P}(\text{keine Rendite} \mid \text{Fond})}_{20\,\%} \cdot \underbrace{(\text{keine Rendite})}_{100\text{ Euro}}$$
$$= 0,8 \cdot 102{,}5 \text{ Euro} + 0,2 \cdot 100 \text{ Euro} = 82 \text{ Euro} + 20 \text{ Euro} = 102 \text{ Euro.}$$

Es ist also gleich, was Sie im diesem Fall tun (jedenfalls was den Erwartungswert betrifft). Viele Menschen würden freilich, vor eine solche Wahl gestellt, der Anlage in Kommunalobligationen (als festverzinslichen Wertpapieren) den Vorzug vor einem Aktieninvestment geben – sie sind risikoscheu oder (vornehmer) **risikoavers** (s.u.).

9.6. Das Allais-Paradoxon

Das sog. Allais-Paradoxon wurde 1953 von dem Ökomomen Maurice Allais formuliert.[102] Es besteht aus zwei Entscheidungen; gezeigt werden soll, daß die beiden in den jeweiligen Situationen scheinbar besten Entscheidungen miteinander entscheidungstheoretisch unverträglich sind (d. h.: die erste Entscheidung ist falsch, oder die zweite, oder die Entscheidungstheorie ist falsch – oder mehreres davon).

Entscheidung I: Welches der beiden Spiele, P oder P', würden Sie bevorzugen?

P: Sie erhalten 500.000 francs mit 11 %-iger Wahrscheinlichkeit, mit 89 %iger Wahrscheinlichkeit aber nichts. – Oder:

P': Sie erhalten mit 10 %-iger Wahrscheinlichkeit 2.500.000 francs, sonst (90 %) aber

[102] Maurice Allais, Le Comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica* 21 (4), 1953, S. 503–546. Allais gewann 1988 den Alfred-Nobel-Gedächtnispreis für Wirtschaftswissenschaften der Schwedischen Reichsbank für "his pioneering contributions to the theory of markets and efficient utilization of resources".

nichts.

Die meisten wählen P' – sie bevorzugen einen möglichen höheren Gewinn. Die Wahl ergibt vor dem Hintergrund der Bayes'schen Regel Sinn:

$E(P) = 0{,}11 \cdot 500.000\,\text{fr.} = 55.000\,\text{fr.}$

$E(P') = 0{,}10 \cdot 2.500.000\,\text{fr.} = 250.000\,\text{fr.}$

Entscheidung II: Welches der beiden Spiele, Q oder Q', würden Sie bevorzugen?

Q: Sie erhalten 500.000 francs sicher (100 %). – Oder:

Q': Sie erhalten mit 10 %-iger Wahrscheinlichkeit 2.500.000 francs, oder mit 89 %-iger Wahrscheinlichkeit 500.000 francs, sonst (1 %) aber nichts.

Hier wählen die meisten Q – sie bevorzugen einen sicheren Gewinn. Aber:

$E(Q) = 500.000\,\text{fr.}$

$E(Q') = 0{,}1 \cdot 2.500.000\,\text{fr.} + 0{,}89 \cdot 500.000\,\text{fr.} = 695.000\,\text{fr.}$

Dies widerspricht aber der Bayes'schen Regel.

Man kann das Paradoxon noch klarer herausarbeiten. Wir betrachten eine Lotterie mit 100 Losen (wir wechseln wieder zu Häufigkeiten):

	Spiel P oder P'		Spiel Q oder Q'	
Lose 1–89	0	0	5	5
Los 90	**5**	**0**	**5**	**0**
Lose 91–100	**5**	**25**	**5**	**25**

Abb. 9.2. Savages Behandlung des Allais-Paradoxons (vgl. Savage 1972, S. 103).

Da die Zeile 1 (Abb. 9.2) in beiden Entscheidungen jeweils die selben Einträge hat (5 bzw. 0) hat, sollte sie für die Wahl in keiner der beiden Situationen eine Rolle spielen. Da die Zeilen 2 und 3 aber in beiden Entscheidungssituationen identisch sind, sollte dann die Entscheidung in beiden Fällen dieselbe sein! (Man könnte also sagen: Es liegt nicht einfach ein Verstoß gegen die Bayes'sche Regel vor, sondern eine tiefere Inkonsistenz der Bewertung.)

Was kann der Grund für die Bevorzugung von Q gegenüber Q' sein? Zwei klassische Antworten:

(i) Menschen sind **risikoscheu**: "Lieber den Spatz in der Hand als die Taube auf dem Dach."[103] Dafür spricht einiges: Eine Interpretation des höheren erwartbaren

[103] Englische Entsprechung: "A bird in the hand is worth two in the bush." Die Zahl ist willkürlich; vgl. "Capta auis est melior quam mille in gramine ruris."

Werts eines Aktieninvestments gegenüber etwa festverzinslichen Anleihen (oder Sparbüchern) interpretiert diesen als Risiko-Prämie – als Ausgleich dafür, daß Aktien im Wert stärker schwanken ('volatiler' sind), s.h. ein größeres Risiko mit sich bringen. – Ein Grund könnte sein, daß Menschen ein Verlust meist mehr ärgert, als sie ein gleichgroßer Gewinn freut (wie man auch empirisch festgestellt hat).

(ii) Oder das Allais-Paradoxon zeigt (wieder einmal), daß Menschen nicht so gut darin sind, mit Wahrscheinlichkeiten umzugehen: Es wäre klüger ('rationaler'), würde man bei der zweiten Entscheidung die Wahl Q' treffen.[104]

9.7. Grenzen der Entscheidungstheorie

Stellen wir uns zunächst auf den Standpunkt der **normativen Entscheidungstheorie**. Menschen machen Dummheiten, und die Entscheidungstheorie soll helfen, Entscheidungssituationen zu analysieren und bessere Entscheidungen zu treffen. Dann ist es ein guter Ansatz, den Erwartungswert einer Handlung zu errechnen (auch wenn der nicht immer hilft, wie beim St. Petersburg-Spiel), und die Handlung mit dem größten Nutzen zu wählen (Bayes'sche Regel). (Die Bayes'sche Regel macht man oft zum Kriterium (zweck-) rationalen Verhaltens: Nur das Verhalten sei rational, das dieser Regel gehorcht.)

Es gibt aber Fälle, in denen der Erwartungswert zweier Handlungsalternativen übereinstimmt, und doch nicht gleich ist, welche Handlung man bevorzugen sollte (Obligationen vs. Fonds) – jedenfalls dann, wenn man den Wert bzw. Nutzen der Handlungsfolgen allein am Geld festmacht, und die Volatilität (Streuung der Gewinnerwartung) nicht berücksichtigt. (Dies ist konsequent, denn der Erwartungswert berechnet ja gezielt gerade den *erwartbaren* Gewinn.)

Man könnte nun sagen: Wenn jemand sich mit dem Risiko eines Verlustes (bei gleichem Erwartungswert) nicht anfreunden kann, zeigt dies einfach, daß Sicherheit für diese Person einen Eigenwert hat – ein Wert an sich ist – und entsprechend ist es (zweck-) rational, in solchen Fällen die sichere Anlage zu bevorzugen (oder sogar eine sicherere mit geringerem Erwartungswert). Aber dann läuft man Gefahr, daß das Bayes-Kriterium seine Aussagekraft verliert, weil jede Entscheidung einfach die subjektiven Wertpräferenzen abbildet (niemand handelt ja wissentlich zum eigenen Schaden?).[105]

[104] Dies war das Ergebnis der Überlegungen des Statistikers Leonard Savage (Leonard Savage, *The Foundations of Statistics*. New York 1954; 2nd ed. 1972, S. 101–103; vgl. Hacking, S. 110–11).

[105] Dies muß aber auch wieder nicht allgemein gelten; vgl. L. Savages Anmerkung zum Allais-Paradoxon: Insofern Präferenzen subjektiv seien, könnten sie natürlich nicht täuschen oder falsch sein; "but in a different, more subtle way they can be [in error]. [...] A man buying a car for $2,134.56

Vom Standpunkt der **deskriptiven Entscheidungstheorie** stellen sich ähnliche Probleme. Wir suchen ein Modell menschlichen Handelns, und erklären die Wahl einer Handlung mit Bezug auf zwei Größen: die **Überzeugungen** des Akteurs und dessen **Werten** (oder Präferenzen: relativen Bewertungen). Aus beiden berechnet man den Erwartungswert, und das Modell sagt voraus, daß Menschen die Handlung wählen, die den Erwartungswert maximiert (Bayes'sche Regel).

Nimmt man allein monetären Eigennutz als Wert oder Ziel an ("We're Only in It for the Money" [106]) – macht das Modell klare Voraussagen, aber viele Handlungen erscheinen irrational (nicht zweck-rational): Menschen sind risikoavers, altruistisch, ärgern sich mehr über Verluste, als sie sich über gleichhohe Gewinne freuen, bewerten Gewinne nicht immer gleich (abnehmender Grenznutzen).

All diese Faktoren könnte man wiederum einbeziehen, indem man die Nutzenfunktion ändert. Aber dann läuft man wiederum Gefahr, daß das Bayes-Kriterium seine Aussagekraft verliert, weil jede Entscheidung einfach die subjektiven Wertpräferenzen abbildet.

is tempted to order it with a radio installed, which will bring the totel price to $ 2,228.41, feeling that the difference is triffling. But, when he reflects that, if he already had the car, he certainly would not spend $ 93.85 for a radio for it, he realizes that he has made an error." (Leonard Savage, *The Foundations of Statistics.* New York 1954; 2nd ed. 1972, S. 103)

[106] The Mothers of Invention/Frank Zappa, 1968.

10. Signifikanztesttheorie

10.1. Wiederholung und Ausblick

Wiederholung: In Vorlesung 3 hatten wir die Frage, wie man sich gehaltserweiternden Argumenten nähern kann, aufgespalten in zwei Ansätze: Einige gehaltserweiternde Argumente lassen sich mit mit dem Begriff der Wahrscheinlichkeit fassen, nach dem Vorbild von Qualitätskontrollen durch Stichprobenentnahmen; diese haben wir **Wahrscheinlichkeitsschlüsse** genannt.

Bei Wahrscheinlichkeitsschlüssen haben wir wiederum zwei Arten unterschieden: Die eine Art waren **Schlüsse von der Grundgesamtheit (und einer Annahme über die Art der Stichprobenentnahme) auf die Stichprobe** (statistische Syllogismen). Die zweite Art waren **Schlüsse von einer Stichprobe auf eine Grundgesamtheit** (statistische Generalisierungen; ein Sonderfall von enumerativen Induktionen). (Schlüsse von Stichproben auf Stichproben haben wir wegdefiniert: Wir haben sie analysiert als Verbindung eines Schlusses von einer Stichprobe auf eine Grundgesamtheit, und eines Schlusses von dieser Grundgesamtheit auf eine weitere Stichprobe.)

Die Vorlesungen 4–7 behandelten die Grundlagen der **Wahrscheinlichkeitstheorie**. Wahrscheinlichkeitstheorie ist ein Teil der Mathematik, und alle Schlüsse waren letztlich **statistische Deduktionen**, also gültige Schlüsse (nicht gehaltserweiternd, meist im Gegenteil!).

Ausblick: Das Ziel im Folgenden ist es, die beiden Teile zusammenzuführen, und insbesondere betrachten, inwieweit ein tieferes Verständnis der Wahrscheinlichkeitstheorie hilfreich ist, um Schlüsse von einer Stichprobe auf eine Grundgesamtheit besser zu fassen.

Es gibt mehrere Ansätze, wie dies gelingen könnte – diese Ansätze lassen sich aber nicht leicht systematisieren. Sie alle haben Vor- und Nachteile (und einige mehr Nach- als Vorteile). Die wichtigsten Ansätze sind:

– Schätztheorie

– **Signifikanztests** (ein Oberbegriff für sehr verschiedene Verfahren!)

– Fiduzialwahrscheinlichkeiten

– Neyman-Pearsonsche Entscheidungsverfahren

– Bayessche Verfahren, z. B. Bayes Factor

– **Bayessche Erkenntnistheorie**.

Wir behandeln nur zwei, Signifikanztests und Bayessche Erkenntnistheorie: **Signifikanztests**, weil sie die in vielen Wissenschaften (noch) wichtigste Form von statistischen Verfahren sind, zugleich aber in der wissenschaftlichen Praxis zu vielen Problemen führen (es ist nicht sehr übertrieben zu behaupten, daß kaum einer sie versteht); **Bayessche Erkenntnistheorie**, weil sie den gegenwärtig in der Philosophie am meisten verbreiteten Ansatz in der formalen Erkenntnistheorie darstellt.

Signifikanztests sind eine Form von **frequentistischen Verfahren**; gemeint sind damit Verfahren, die die Wahrscheinlichkeiten als langfristige Häufigkeiten deuten: Daß bei einem fairen Würfel die Wahrscheinlichkeit, x Augen zu würfeln, $\mathcal{P}(x) = \frac{1}{6}$ beträgt, sagt einfach nur aus, daß man, wenn man sehr häufig würfelt, in einem Sechstel der Ausfälle x Augen erhält.

Bayessche Verfahren hingegen interpretieren Wahrscheinlichkeiten als **personelle Wahrscheinlichkeiten**: Daß bei einem Würfel, den man für fair hält, die Wahrscheinlichkeit, die Augenzahl x zu würfeln, $\mathcal{P}(x) = \frac{1}{6}$ beträgt, sagt einfach nur aus, daß man jeden Ausfall für gleichwahrscheinlich hält (und also nicht viel Geld darauf wetten würde, daß eine ‘1’ fällt, schon eher aber, daß eine ‘1’, ‘2’, ‘3’ oder ‘4’ fällt.

10.2. Stabilität von Erwartungswerten

In der Entscheidungstheorie bezeichnete der Ausdruck ‘Erwartungswert’ ganz wörtlich den erwarteten Wert oder Nutzen; im folgenden sprechen wir in einem weiteren Sinn auch dann von Erwartungswerten, wenn kein Nutzen mit dem Wert verbunden ist: Erwartungswerte sind Durchschnittswerte von relativen Häufigkeiten, die man aufgrund von Wahrscheinlichkeiten erwarten darf.

Beispiel. Der Erwartungswert der durchschnittlichen Augenzahl eines fairen Würfels ist

$$E(\text{Augen}) = \sum_{i=1}^{6} p_i \, x_i = \tfrac{1}{6}\, 1 + \tfrac{1}{6}\, 2 + \ldots = \tfrac{1}{6}\, (1+2+3+4+5+6) = 3,5.$$

Erwartungswerte verbinden Wahrscheinlichkeiten und Häufigkeiten. Sie haben die folgende Eigenschaft: Bei vielen Wiederholungen eines Experiments, z. B. eines Bernoulli-Experiments, wird die relative Häufigkeit (etwa, mit einer fairen Würfel eine 4 zu werfen), immer **stabiler** der Wahrscheinlichkeit nahekommen, bei einem einzelnen Wurf eine 4 zu werfen, nämlich $\frac{1}{6}$. Präziser:

> Je häufiger ein Experiment wiederholt wird, desto größer ist die Wahrscheinlichkeit, daß die relative Häufigkeit der Ausfälle nahe an dessen Wahrscheinlichkeit liegt. **[Formulierung korrekt?]**

Wir nehmen dies hier als Tatsache hin, als **empirisches Gesetz der großen Zahlen** [107] (vgl. das Beispiel Abb. 10.1). Daß Zusammenhänge zwischen relativen Häufigkeiten und Wahrscheinlichkeiten bestehen, zeigen in der Mathematik die sogenannten Gesetze der großen Zahlen.[108]

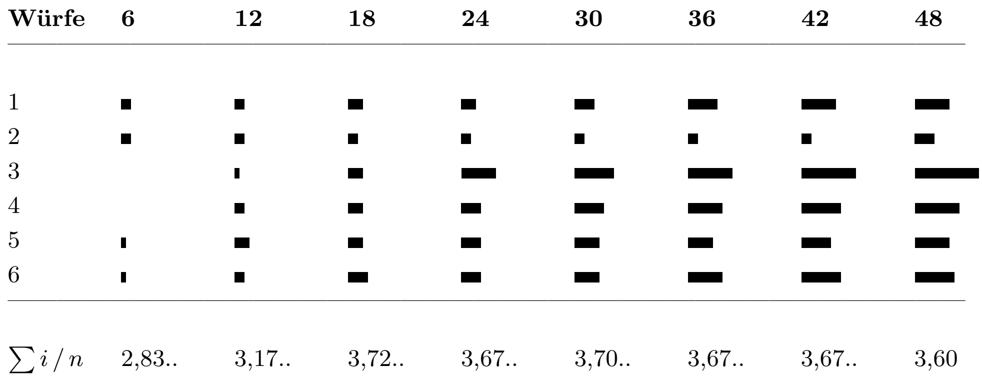

Abb. 10.1. Zufalls-Experiment: Wiederholtes Würfeln, Bielefeld, 20. Juni 2015. In den Spalten die Verteilung nach Augen (i) nach jeweils 6 Würfen (n). Unten die Entwicklung von $\sum i\,/\,n$ (Erwartungswert $E(i) = 3,5$). Man erkennt, daß sich $\sum i\,/\,n$ um $E(i)$ stabilisiert.

10.3. Signifikanztests: ein Beispiel

Unter 'Signifikanztest' (significance test) wird im folgenden eine Art des Testens statistischer Hypothesen verstanden, deren Theorie maßgeblich auf die Arbeiten Sir Ronald Fishers zurückgeht, eines britischen Physikers und Biologen, der 1919 als Statistiker an die Rothamsted Experimental Station kam, um Daten zahlreicher Feldexperimente zu analysieren.[109]

[107] Vgl. die Aussage bei Kolmogoroff 1933, S. 4 [204].
[108] Vgl. zu den Gesetzen der großen Zahlen Hacking, ch. 16.
[109] Fisher verfaßte mehrere Werke zur statistischen Datenanalyse und zum experimentellen Design, die die gegenwärtige wissenschaftliche Praxis maßgeblich mitgeprägt haben (u.a. *Statistical Methods for Research Workers* (1925) und *The Design of Experiments* (1935)). Die Art des Schließens ist

In einem Signifikanztest wird die statistische Hypothese daran gemessen, wie weit die Ergebnisse eines Experiments vom statistisch Erwartbaren abweichen, so daß man schließen kann:

> *Entweder die statistische Hypothese ist falsch, oder das Experiment hatte ein sehr unwahrscheinliches (seltenes) Ergebnis.*

Ist die Abweichung (zu) groß, wird das Testergebnis **statistisch signifikant** genannt (oft verkürzt – und mehrdeutig – signifikant). Die Abweichung von experimentellem Befund und Erwartung wird dabei verwendet, um die Hypothese zu beurteilen.[110]

In der wissenschaftlichen Praxis finden sich oft mehrere statistiyche Betrachtungsweisen vermengt, die sich in Zielsetzung und Interpretation unterscheiden und besser auseinandergehalten würden: Fisher'sche Signifikanztests, Neyman-Pearson-Tests und Bayes'sche Betrachtungen. Wir betrachten sie nacheinander, werden aber auch einen Blick auf typische Schwierigkeiten werfen, die aus der Vermengung entstehen.[111].

Bei der Signifikanztesttheorie nutzt man die erwartete relative Häufigkeit der einzelnen Ausfälle eines Experiments, um zu bestimmen, ob das Testergebnis statistisch signifikant ist (zur Bedeutung s.u.). Ein **Beispiel**:

Hypothese: Wir möchten eine bestimmte Hypothese prüfen (die sog. **Null-Hypothese** (null hypothesis) – eine Begriffsbildung Ronald Fishers); z. B., ob eine bestimmte Münze fair ist. Die Null-Hypothese ist eine **statistische Hypothese**, d. h. eine Aussage über Wahrscheinlichkeiten bzw. (erwartbare) Häufigkeiten. Die Münze ist entweder fair oder nicht (nicht: vermutlich fair, oder wahrscheinlich fair, oder irgendetwas dergleichen. Fair oder nicht fair).

freilich älter – der erste Signifikanztest wird John Arbuthnot zugeschrieben (John Arbuthnott [sic], *An Argument for Divine Providence, taken from the constant Regularity observ'd in the births of both Sexes*, in: Philosophical Transactions 27, 1710, 186–190).

[110] Die Logik erzwingt dies nicht; man könnte (und hat) die Abweichung auch als Kriterium verwendet, um sog. outliers (stark abweichende empirische Befunde) auszusondern. – Zur Geschichte der Signifikanztests vgl. Gigerenzer et al. 1989, ch. 3.

[111] Zur Entwicklung der 'hybriden' Praxis bes. in der Psychologie und den dadurch verursachten Schwierigkeiten s. Gigerenzer 1993. Zu Signifikanztests vgl. Howson/Urbach 2006, ch. 5 b (S. 133–143). Eine Form der 'hybriden' Betrachtungsweise ist das *p value null-hypothesis significance testing* (NHST); vgl. zu den Problemen in der Praxis Wagenmakers 2007. Royall (1997, ch. 3) unterscheidet *p value procedures* und *rejection trials*, und hält beide für problematisch. (p-Wert-Prozeduren entsprechen eher dem frühen Fisher (1925), rejection trials (Signifikanztests im engeren Sinn) gehen auf Fisher 1935 zurück

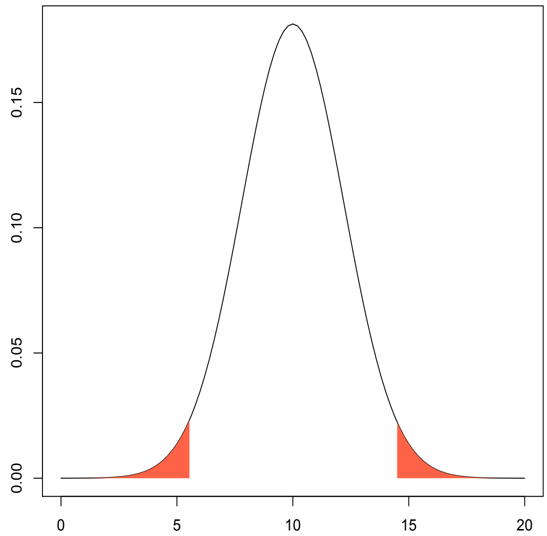

Abb. 10.3. 20 Würfe mit einer Münze; Wahrscheinlichkeit bzw. erwartete relative Häufigkeit wiederholter Versuche mit dem Ergebnis KOPF unter der Annahme, die Münze sei fair (geglättet). Auf der Abszisse ist die Gesamtzahl der KOPF-Ausfälle aufgetragen, auf der Ordinate die sog. Wahrscheinlichkeitsdichte. Am wahrscheinlichsten sind 10 Ausfälle mit KOPF. Rot gekennzeichnet sind die Bereiche, die unter der Zufallshypothese seltene Ausfälle sind: sehr seltenes oder sehr häufiges Vorkommen von KOPF (die 'statistisch signifikanten' Ausfälle bei einem 5 %-Signifikanzniveau).

Test: Wir entscheiden uns dafür, dies zu prüfen, indem wir 20 mal die Münze werfen, und die Anzahl der KOPF-Ausfälle zählen. Bei zahlreichen Wiederholungen dieses Experiments würden wir, wenn die Münze fair ist, bestimmte Ergebnisse mit bestimmter relativer Häufigkeit erwarten (Abb. 10.3).

Interpretation: Nehmen wir an, unser Experiment ergibt ein Resultat ganz an einem der beiden Ränder der Verteilung – etwa nur 3 mal (oder noch seltener) oder 17 mal (oder noch häufiger) KOPF. (Bei unserer Fragestellung ("Ist die Münze fair?"), die sich in der Null-Hypothese widerspiegelt, wäre auch ein Ergebnis, in dem 17 mal oder noch häufiger KOPF fällt, bemerkenswert – starke Abweichungen vom Erwartungswert nach beiden Seiten sprechen ja für eine gewisse Unwucht bei der Münze.) Wie läßt sich dies interpretieren?

Die Wahrscheinlichkeit beträgt (ausschließend: addieren; Rechnung nur interessehalber):

$$\mathcal{P}(\text{Ränder}) = \underbrace{2}_{\text{beide Ränder}} \cdot \underbrace{\left(\binom{20}{0} + \binom{20}{1} + \binom{20}{2} + \binom{20}{3}\right)}_{\text{linker Rand (= rechter Rand)}} \cdot \underbrace{\left(\frac{1}{2}\right)^{20}}_{\mathcal{P}} \approx 0,0025768.$$

Damit kann man folgenden Schluß ziehen:

> **Wenn** die Münze fair ist, **dann** ist ein Ergebnis, bei dem nur 3 mal oder noch seltener KOPF fällt, sehr unwahrscheinlich.

Man sagt, das Ergebnis ist **überzufällig**.

10.4. Die einzelnen Schritte eines Signifikanztests

Der Aufbau eines Signifikanztests folgt folgendem Aufbau/Schritten:

1. Formulieren der **Null-Hypothese H_0** (= was immer passiert, ist Zufall) und des **statistischen Modells**. Hier: Die Münze ist fair. (Andere Kontexte: Ein Medikament ist ohne Wirkung. Ein Düngemittel ist nutzlos.)

Anmerkungen: Die Null-Hypothese ist die Hypothese, **daß, was immer passiert, ein Zufall ist**. Sie muß spezifisch sein. Die Hypothese, daß die Münze nicht fair ist, ließe sich nicht prüfen – sie ist zu unspezifisch und bestimmt keine erwartete Verteilung der möglichen Ausfälle. Man könnte feiner unterscheiden zwischen der substantielle Null-Hypothese ("Die Münze ist fair") und der eigentlichen statistischen Null-Hypothese.[112]

2. Design des Tests. Hier: Wir werden 20 mal die Münze werfen. (Andere Kontexte: 10 Parzellen zufällig auswählen und düngen, weitere 10 nicht – als Vergleichsgruppe – nicht düngen.)

Anmerkungen: Man könnte andere Formen wählen, je nach Kontext. Hier verwenden wir einen Test mit festen Stichprobenumfang (20 Würfe); das scheint natürlich, ist aber nicht zwingend. In der Medizin und der Qualitätskontrolle etwa verwendet man oft sog. sequentielle Tests: Man testet nacheinander, bis eine festgelegte Bedingung erfüllt ist. (Beispiel: Aus einer Kiste Orangen testet man so lange eine Orange nach der nächsten, bis man 5 gefunden hat, die nicht verfault sind. Die Überlegung dahinter: Mehr als 5 gute Orangen möchte man nicht aufschneiden (und wegwerfen), während es egal wäre, wie viele verfaulte man testet. Man kann zeigen, daß der erwartbare Stichprobenumfang (die erwartete Zahl der getesteten Orangen) kleiner ist als bei einem Test mit festen Stichprobenumfang; man spart also Kosten.)

3. Die sog. **Test-Statistik** wählen, d. h. die Art der Anordnung der möglichen Ausfälle. Hier: Wir interessieren uns für die Zahl der KOPF-Ergebnisse unter den 20 Würfen,

[112] Vgl. Gigerenzer et al. 1989, S. 97; Gigerenzer 1993, S. 318 (mit Verweisen).

unabhängig von der Reihenfolge, und ordnen sie der Größe nach von 0 bis 20.

Anmerkungen: Auch dies könnte man anders machen. Wir vermuten, daß die Münze, sollte sie nicht fair sein, 'Schlagseite' hat; dies würde sich in sehr wenigen oder sehr vielen KOPF-Ergebnissen widerspiegeln. Wäre die Münze nicht fair, weil die Würfe nicht unabhängig sind, würde man dies so nicht unbedingt herausfinden – hier bräuchte man eine Vermutung, wie die Münze gezinkt ist, um eine Test-Statistik zu wählen.

4. Festlegen des **Signifikanzniveaus**. Hier: (a) Wir finden beide Ränder der Verteilung (fast keine KOPF und fast nur KOPF-Ergebnisse) gleichermaßen problematisch (sog. zweiseitiger Test), denn beides spricht für ein Ungleichgewicht der Münze. (b) Wir legen beispielsweise fest, daß wir ein Ergebnis dann für 'zu' zufällig ('überzufällig') halten, wenn es nur in 5 von 100 Fällen zu erwarten ist. Technisch gesprochen: Wir verwenden ein Signifikanz-Niveau von $\alpha = 5\,\%$. (Dies verteilt sich hier auf jeweils 2,5 % auf jedem der beiden Ränder der Verteilung.)

Anmerkungen: Die Wahl des Signifikanzniveaus ist eigentlich willkürlich. Man verwendet oft $\alpha = 0{,}01$ oder $\alpha = 0{,}05$. Dies ist eine Tradition, mehr nicht; sie kommt daher, daß diese Werte leicht zu berechnen waren und (bevor es Computer gab) als Tabelle vorlagen. Eine Rolle spielt auch, daß diese Werte forschungsökonomisch brauchbar sind: Sie führen hinreichend oft, aber nicht zu oft, zu signifikanten Resultaten.

Wichtig: Alle diese Testparameter – Null-Hypothese, das Test-Design, die Test-Statistik und das Signifikanzniveau – müssen vor bzw. unabhängig vom Test gewählt werden (dies nennt man die **Prädesignation der Testparameter**).

5. Wir führen das Experiment durch und bestimmen den ***p*-Wert** (p wegen probability): Wir berechnen, wie wahrscheinlich (gegeben unsere erwarteten Häufigkeitsverteilung der möglichen Experiment-Ausfälle) der Ausfall unseres Experiments oder ein noch unwahrscheinlicherer Ausfall ist. (Man könnte sagen: Der p-Wert ist das hypothetische Signifikanzniveau, bei dem der Ausfall des Versuchs gerade noch signifikant gewesen wäre.)

Da der "beobachtete" p-Wert (in Dezimalschreibweise) unhandlich ist, gibt man meist einen (auf)gerundeten Wert als p-Wert an ("berichteter/angegebener p-Wert").

> **Beispiel:** Die genauere Wahrscheinlichkeit in unserem Experiment, 3 mal oder seltener *oder* 17 mal oder häufiger KOPF zu werfen, ist
>
> $\mathcal{P}(\text{Rand}) \approx 0,0025768 < \mathbf{0{,}003}$ **(*p*-Wert)**.

Liegt der p-Wert unterhalb des gewählten Signifikanzniveaus, bezeichnet man das Ergebnis als **(statistisch) signifikant** (auf einem 5 %-Niveau). 'Signifikant' hat hier eine technische Bedeutung – es bedeutet nicht 'bedeutsam' oder 'wichtig' (auch kleinste, ganz irrelevante Effekte können ‘signifikant’ sein).

Das **Signifikanzniveau** ist eine **Eigenschaft des Tests** (s.u.), der ***p*-Wert** hingegen eine **Eigenschaft des Test-Ausfalls** (der Null-Hypothese und der Daten). Wie gesagt, gibt der p-Wert an, wie hoch das Signifikanzniveau hätte gewählt werden können, so daß der tatsächliche Ausfall gerade noch signifikant gewesen wäre.

10.5. Die Interpretation des Ergebnisses: statistische Signifikanz

Was ist die Aussage von ‘Das Testergebnis ist signifikant auf einem Niveau von $x\,\%$’? Die Aussage ist: Gegeben das Zufallsmodell und die Teststatistik (und eine saubere Experiment-Durchführung), gilt:

> **Wenn** die Null-Hypothese richtig ist, **dann** ist das Ergebnis ein großer Zufall.

Oder ($A \supset B$ ist äquivalent zu $\neg A \vee B$):

> (**Entweder**) ist die Null-Hypothese **nicht** richtig, **oder** das Ergebnis ein großer Zufall.

Was das Ergebnis einem **ganz sicher nicht** sagt, ist, ob die Annahme, die Münze sei fair, angesichts der Ergebnisses des Experiments wahrscheinlich ist oder nicht. (Das Ergebnis jedes Lotto-Spiels ist ein sehr unwahrscheinliches Ergebnis – aber spricht dies dafür, daß die Lotterie nicht fair ist?)

Soweit reicht die Logik, und soweit ist alles eindeutig.

(i) Signifikanz: Mißinterpretationen in Medizin und Psychologie

Der Begriff der statistischen Signifikanz hat seit jeder zu Mißinterpretationen eingeladen. Der Mediziner Steven Goodman faßte dies 2008 folgendermaßen zusammen:

> [A]rticles have appeared in the biomedical literature for at least 70 years warning researchers of the interpretive P-value minefield, yet these lessons appear to be either unread, ignored, not believed, or forgotten as each new wave of researchers is introduced to the brave new technical lexicon of medical research.[113]

Er gibt folgende Übersicht über **typische Mißinterpretationen**:[114]

---

1. If $P = .05$, the null hypothesis has only a 5% chance of being true.
2. A nonsignificant difference (eg, $P \geq .05$) means there is no difference between groups.
3. A statistically significant finding is clinically important.

[113] Steven Goodman, A Dirty Dozen: Twelve P-Value Misconceptions, in: Seminars in Hematology 45, 2008, S. 135–140; hier S. 135.
[114] Ibid., Tab. 1, S. 136.

4. Studies with P values on opposite sides of .05 are conflicting.

5. Studies with the same P value provide the same evidence against the null hypothesis.

6. $P = .05$ means that we have observed data that would occur only 5% of the time under the null hypothesis.

7. $P = .05$ and $P \leq .05$ mean the same thing.

8. P values are properly written as inequalities (eg, "$P \leq .05$" when $P = .015$)

9. $P = .05$ means that if you reject the null hypothesis, the probability of a type I error [115] is only 5%.

10. With a $P = .05$ threshold for significance, the chance of a type I error [116] will be 5%.

11. You should use a one-sided P value when you don't care about a result in one direction, or a difference in that direction is impossible.

12. A scientific conclusion or treatment policy should be based on whether or not the P value is significant.

Abb. 10.4. Mißinterpretationen von *p*-Werten nach Goodman (2008). **Aufgabe:** Können Sie sagen, worin hier die Irrtümer bestehen? (Nicht in allen Fällen läßt es sich aufgrund der Vorlesung sagen; Sie müssen ggf. auch etwas weiterlesen.)

In der Praxis finden sich meist schlichte Mißinterpretationen; aber auch in der **Philosophie** gibt es Auffassungen, die jedenfalls nicht unproblematisch sind. Eine häufige ist die folgende:

Statistische Hypothesen lassen sich nicht einfach durch Experimente widerlegen ('falsifizieren'): Die Hypothese 'Alle Schwäne sind weiß' (nicht-statistisch) ist durch die Beobachtung eines einzigen roten Schwans widerlegt; die Hypothese 'Dieser Würfel ist fair' (statistisch) hingegen ist auch durch ein Experiment, in welchem 666 mal 6 hintereinander fällt, nicht als falsch erwiesen. Signifikanztests sollen diese logische Lücke füllen, indem in ihnen definiert wird, wann eine Hypothese praktisch als falsifiziert gilt (nämlich dann, wenn das Ergebnis des Tests 'signifikant' ist) – dann wird diese Hypothese **verworfen** (**rejected**).[117] (Der Begriff des 'Verwerfens' einer Hypothese ist, ebenso wie der der 'Akzeptanz', in der Statistik nicht wirklich definiert, obwohl beide wie Fachtermini verwendet werden.)

Diese Interpretation kann man wählen, aber sie führt zu einige Schwierigkeiten [118] und ist keineswegs die einzig methodologisch mögliche (obgleich auch der (späte) Fisher

[115] Ein Fehler erster Art (type I error) bezeichnet den Fehler, die Null-Hypothese irrtümlicherweise zu 'verwerfen'.
[116] Vgl. die vorangehende Fußnote.
[117] Vgl. etwa Howson/Urbach, S. 131–132.
[118] Vgl. ibid., S. 133 ff.

in diese Richtung dachte).

(ii) Signifikanz: Andere Interpretationen

In der wissenschaftlichen Praxis bedeutet ein signifikantes Ergebnis zumeist schlicht, daß man es **veröffentlichen** kann. Dies ist durchaus sinnvoll motiviert; eine wichtige Bedeutung von Signifikanz, die Fisher betonte, liegt nahe an der üblichen Bedeutung des Ausdrucks: Ein Ergebnis ist signifikant, wenn es Aufmerksamkeit verdient.[119]

Problematisch ist hier allerdings, daß nicht-signifikante Ergebnisse meist *nicht* veröffentlicht werden (sog. Publication bias). Dies ist problematisch, da man auch dann, wenn ein Effekt tatsächlich Zufall ist, selten aber doch ein signifikantes Resultat erwarten kann (bei einem Signifikanz-Niveau von 5 % in einem vom 20 Fällen). Man darf daher weder annehmen, jedes veröffentlichte signifikante Ergebnis sei erwiesen, noch, die nicht-veröffentlichten (nicht-signifikanten) Ergebnisse seien irrelevant.

Eine **plausible Interpretation** ist folgende: Ein **signifikantes Ergebnis** ist eines, das weitere **Beachtung verdient**, und besonders eines, daß es **wert ist, wiederholt und weiter erforscht zu werden**. Dies war die ursprüngliche Funktion bei Fisher, der in den 20er Jahren des letzten Jahrhunderts in der Rothamsted Experimental Station (einem agrarwissenschaftlichen Forschungsinstitut) arbeitete und entscheiden mußte, welche der dort durchgeführten aber nicht ausgewerteten Experimente weiterverfolgt werden sollten.

10.6. Ronald Fishers Tea-testing-Experiment

In *The Design of Experiment* (1935) diskutiert Fisher sich das folgende sehr berühmte Beispiel (das hier zugleich als Beispiel für einen ‘einseitigen Test’ dient):

> A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup.[120]

Das Experiment: Wir nehmen 8 Tassen Tee mit Milch, von denen bei 4en zuerst der Tee, bei vieren zuerst die Milch in die Tasse gegeben wurde. Die Probandin erfährt dies, und daß sie die Tassen in zufälliger (random) Reihenfolge (entschieden etwa durch Würfeln). Sie hat die Aufgabe, unter den 8 Tassen die 4 Tassen herauszufinden, in denen zuerst der Tee eingefüllt wurde.

Das Raisonnement: 4 aus 8 Tassen lassen sich auf 70 verschiedene Möglichkeiten wählen: Man kann $8 \times 7 \times 6 \times 5 = 1680$ Anordnungen (geordnete Mengen) von

[119] Vgl. Gooding 2008, S. 135, zu Fishers ursprünglicher Deutung.

[120] Ronald Fisher, The Design of Experiments, Edinburgh/London 1935, S. 13–29. Der Sinn der Übung erschließt sich vor dem Hintergrund des britischen Klassensystems: die upper classe bevorzugte, zuerst den Tee einzuschenken, die lower classes hielten es umgekehrt; vgl. Ian Hacking, Origins of Randomization in Experimental Design, in: Isis 79/3, 1988, S. 427– 451, bes. S. 449–451.

Tassen wählen, z. B. die Anordnung $(5, 1, 7, 8)$; dabei sind jeweils $4 \times 3 \times 2 \times 1 = 24$ Anordnungen Varianten einer Kombination (ungeordneten Menge), z. B. $\{1, 5, 7, 8\}$; Da es auf die Anordnung nicht ankommt, bleiben $1680 / 24 = 70$ Kombinationen.[121]

Jemand, der nicht zwischen beiden Arten der Mischung unterscheiden kann, wird erwartbar nur in 1 von 70 Fällen mit der Aufteilung richtig liegen (bzw. wird mit einer relativen Häufigkeit richtig liegen, die sich bei zahlreichen Wiederholungen dem Verhältnis 1/70 annähert). (Würde man nur jeweils 3 Tassen verwenden, läge die Wahrscheinlichkeit, zufällig richtig zu liegen, schon bei 1/20 (prüfen!)). Fisher führt aus:

> It is usual and convenient for experimenters to take 5 per cent. as a standard level of significance, in the sense that they are prepared to ignore all results which fail to reach this standard, and, by this means, to eliminate from further discussion the greater part of the fluctuations which chance causes have introduced into their experimental results.[122]

Aber diese Entscheidung, ein 5 %-Signifikanzniveau zu wählen, sei, so Fisher, willkürlich.

Es gibt 16 Arten, zufällig bei 3 bzw. bei 1 Tasse richtig zu liegen, und 36, gerade bei 2 Tassen richtig zu liegen. 16/70 wäre schon allein nicht signifikant auf dem 5%-Level; um so weniger, wenn man auch die noch besseren Ergebnisse berücksichtigt (also $17/70 \approx 0,24$).

Die möglichen Ausfälle werden also in zwei Gruppen geordnet: solche, die eine (im statistischen Sinn) signifikante Abweichung von der Null-Hypothese aufweisen, und solche, für die dies nicht gilt. Die Null-Hypothese könne, so Fisher, durch das Experiment nicht als wahr erwiesen (‘proved’), sondern nur ‘disproved’ werden. Aber dies bedeute nicht, daß damit gezeigt wäre, daß die Probandin die Arten der Tee-Zubereitung erkennen kann – denn diese Hypothese ist vage und nicht exakt (S. 19) – m.a.W., wir wissen nicht, mit welcher Wahrscheinlichkeit etwa 3 richtig bestimmte Tassen zu erwarten sind, wenn es nicht reines Raten ist. Die Null-Hypothese müsse exakt sein, um eine Berechnung der Wahrscheinlichkeitsverteilung zu erlauben.

[121] Dies entspricht dem sog. Binomialkoeffizieten $\binom{8}{4} = \frac{8!}{4!\,(8-4)!} = \frac{8}{4}\frac{7}{3}\frac{6}{2}\frac{5}{1} = 70$.
[122] Fisher 1935, S. 15–16.

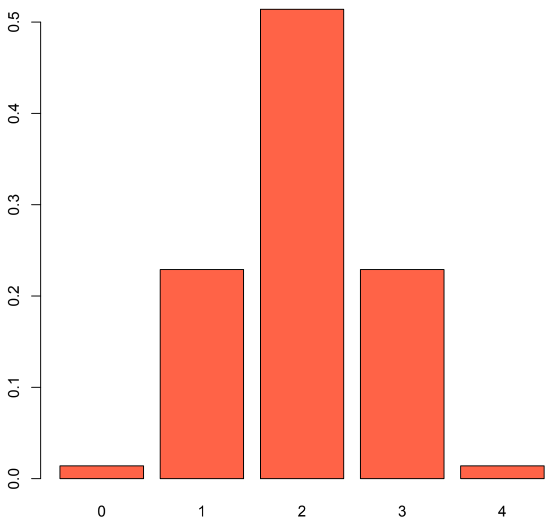

Abb. 10.5. Wahrscheinlichkeiten der jeweils richtig geratenen Tassen unter der Null-Hypothese (d. h., wenn tatsächlich *geraten* wird). Zu erwarten ist, daß 2 von 4 Tassen richtig bestimmt sind; nur die Fälle, daß alle falsch sind (0) und daß alle richtig sind (4), sind deutlich unwahrscheinlich. Beachtet wird im Experiment aber nur der Fall, daß deutlich mehr Tassen richtig bestimmt sind, also nur der rechte Rand (4); aber auch, daß alles falsch ist (0), ist unter der Null-Hypothese kaum zu erwarten.

Anzahl*	**Kombinationen**	**Wahrscheinlichkeit**	***p*-Wert**
4	1	$\frac{1}{70} \approx 0{,}01$	$0{,}01 < \mathbf{0{,}02}$
3	16	$\frac{16}{70} \approx 0{,}23$	$0{,}01{+}0{,}23 = 0{,}24 < \mathbf{0{,}25}$
2	36	$\frac{36}{70} \approx 0{,}51$	
1	16	$\frac{16}{70} \approx 0{,}23$	
0	1	$\frac{1}{70} \approx 0{,}01$	

Abb. 10.6. Wahrscheinlichkeiten, erwartete relative Häufigkeiten und *p*-Werte von Fishers Tea-testing-Beispiel. * ‘Anzahl’ bezeichnet die Anzahl richtig geratener Tassen.

11. Bayes'sche Erkenntnistheorie

11.1. Explikationen von 'Wahrscheinlichkeit'

Die Aussage 'Der Würfel ist fair' bedeutet, daß jede Seite gleichwahrscheinlich geworfen wird und daß die Würfe unabhängig sind. Eine 4 oder 5 fällt mit der Wahrscheinlichkeit 2/6: Wahrscheinlichkeit ist das Verhältnis der günstigen zu den möglichen Ausfällen. Dies ist die sog. klassische oder **Laplace'sche Definition** von Wahrscheinlichkeitsaussagen.

Sie bietet sich für Würfelexperimente an: Die Wahrscheinlichkeit jedes einzelnen sog. Elementarereignisses – die sog. Apriori-Wahrscheinlichkeit – ist bekannt (muß bekannt sein), es gibt endlich viele mögliche Ausfälle eines Experiments (ein Würfel hat endlich viele Seiten), und alle sind gleichwahrscheinlich. – Ist ein Würfel aber nicht fair, ist diese Explikation nicht anwendbar.

Die **statistische Wahrscheinlichkeitsauffassung** (auch relative Häufigkeitsinterpretation oder verkürzt Häufigkeitsinterpretation) geht auf die Mathematiker John Venn [123] und Richard von Mises [124] zurück. Nach von Mises' sog. Limes-Definition von Wahrscheinlichkeit ist die Wahrscheinlichkeit eines Ereignisses (Ausfalls) A der Grenzwert von dessen relativer Häufigkeit bei unendlich vielen Wiederholungen. Wenn der Ausfall A bei n Versuchen n_A mal auftritt, ist

$$\mathcal{P}(A) = \lim_{n\to\infty} \frac{n_A}{n}.$$

Vorzüge: (1) Diese Explikation läßt sich auch auf gezinkte Würfel anwenden. (2) Man muß die Apriori-Wahrscheinlichkeiten nicht kennen: Man kann mit einem gezinkten Würfel sehr oft würfeln und die Apriori-Wahrscheinlichkeiten *schätzen* ("schätzen" ist

[123] John Venn, *The Logic of Chance*. London: MacMillan, 1876.
[124] Richard von Mises, *Wahrscheinlichkeit Statistik und Wahrheit*. Wien: Springer, ^{3}1951.

hier ein Fachbegriff: es gibt eine Schätztheorie), wenn die Verteilung sich stabilisiert. – Einige Probleme seien genannt: (3) Der Sinn einer Wahrscheinlichkeitsaussage über ein einzelnes Experiment ist unklar. (4) Es gibt keine Würfel, mit denen man unendlich oft werfen könnte. Und selbst wenn: (5) Eine Aussage über langfristige Häufigkeiten ist oft nicht hilfreich, wie John M. Keynes bemerkte: "*In the long run* we are all dead."[125] (6) Gar keinen Sinn scheint es zu ergeben, über eine "wahrscheinlich richtige Erklärung" eines Phänomens zu sprechen.

Daneben gibt es die **personelle** oder **subjektive Interpretation** von Wahrscheinlichkeitsaussagen (**personal**, **subjective** oder **Bayesian probability**), die Wahrscheinlichkeitsaussagen ("Die Niederschlagswahrscheinlichkeit beträgt 90 %") als Aussagen über einen bestimmten **Grad der Überzeugung** (**degree of belief**) begreift, den jemand hinsichtlich einer Aussage hat. Wesentlich befördert haben diese Auffassung Bruno de Finetti und Frank Ramsey Mitte des vergangenen Jahrhunderts.[126]

Man spricht oft von Theorien der Wahrscheinlichkeit, so als sei nur eine haltbar oder sinnvoll. Wir sprechen hier einfach von verschiedenen Auffassungen von Wahrscheinlichkeitsaussagen. Die statistische Auffassung von Wahrscheinlichkeitsaussagen fügt sich gut zu **objektiven** oder **physikalischen Wahrscheinlichkeiten** (**chance**): Wahrscheinlichkeitsaussagen sind demnach (stets oder vornehmlich) Aussagen über die Welt, die von Wahrscheinlichkeiten durchdrungen ist (Glücksspiele, Atomzerfälle). Die personelle Auffassung von Wahrscheinlichkeitsaussagen paßt gut zu **subjektiven Wahrscheinlichkeiten** (**credence**): Demnach liegt der Sinn von Wahrscheinlichkeitsaussagen (stets oder vornehmlich) darin, daß wir nicht genug über die Welt wissen: Die Welt ist, wie sie ist, aber wir sind uns unsicher.

Aber natürlich kann man die Frage nach der Bedeutung von Wahrscheinlichkeitsaussagen von der nach der 'Natur' der Wahrscheinlichkeit unterscheiden, und beide anders verbinden: So könnte ich (personell) an objektive Wahrscheinlichkeiten glauben (etwa recht fest überzeugt sein, daß ein Würfel fair ist, oder daß die Quantenmechanik stimmt).[127]

[125] John Maynard Keynes, *Tract on Monetary Reform.* London: MacMillan & Company 1923, S. 80, über die Zusammenhänge zwischen Geldmenge in Inflation: "Now 'in the long run' this is probably true [...] But this *long run* is a misleading guide to current affairs. *In the long run* we are all dead." (Hervorheungen im Original)

[126] Eine Übersicht über die Entwicklung geben Henry E. Kyburg, Jr., Howard E. Smokler, Introduction, in: id. edd. *Studies in Subjective Probability.* New York: Robert E. Krieger, [2] 1980, S. 3–22. – Die Bezeichnung 'personelle Wahrscheinlichkeit' geht auf den Statistiker Leonard Savage zurück.

[127] Ob es überhaupt Wahrscheinlichkeiten in der Welt gibt, ist eigentlich eine gesonderte Frage. Vielleicht gibt es keine: Ein Randomisier-Programm auf einem Computer jedenfalls erzeugt keine zufälligen, sondern einfach schwer vorhersagbare Ausgänge. Auch ein Würfel nicht (wir können vielleicht nicht vorhersagen, wie er fällt, aber vielleicht ließe sich das mit einem sorgfältigen Experiment-Aufbau beheben). Die besten Kandidaten für echte Wahrscheinlichkeiten sind wohl die der Quantenmechanik, etwa die Wahrscheinlichkeit, daß ein radioaktives Atom zerfällt.

11.2. Personelle Wahrscheinlichkeiten

Daß wir uns hinsichtlich der Wahrheit von Aussagen oft unsicher sind, liegt auf der Hand. Dies spiegelt sich auch in der Sprache wieder:

> Ich bin mir ziemlich sicher, daß … – Du solltest davon ausgehen, daß … – Ich würde nicht ausschließen, daß … – Darauf würde ich nicht wetten!

Aber welchen Sinn haben numerische Angaben? Welchen Sinn kann es haben zu sagen, wir seien von A zu 80 % überzeugt? Im Alltag wie in der Wissenschaft finden sich solche Angaben über exakte Überzeugungsgrade kaum.[128]

Ein Ansatz verknüpft die personelle Wahrscheinlichkeit mit Wettquoten. **Wettquote** (**betting rate**) bezeichnet den relativen Gewinn bei Erfolg der Wette. Wenn die Wettquote etwa 1 : 2,5 beträgt, erhält man das 2,5-fache des Einsatzes als Ausschüttung:

> Einsatz × Wettquote = Ausschüttung.

(Gemeint ist damit die sog. Bruttoquote – von der Ausschüttung muß noch der Einsatz abgerechnet werden, um den Gewinn zu erhalten. Bisweilen wird anstelle der Bruttoquote auch die sog. Nettoquote, die **Odds**, angegeben: dies ist das Verhältnis von Gewinn (Ausschüttung – Einsatz) zum Einsatz. Einer Wettquote von 1:4 (Gewinn: 3) entsprechen also Odds von 1:3.)

Damit kann man folgendermaßen argumentieren: *Wenn* ich bereit bin, eine Wette zu akzeptieren mit einer Wettquote von 1:2 *auf* die Aussage, daß es morgen regnet (A), *dann* bedeutet dies, daß meine personelle Wahrscheinlichkeit $\mathcal{P}(A) \geq 1/2$ ist: Ich halte es mindestens für so wahrscheinlich, daß es regnet, wie die kontradiktorische Aussage $\neg A$, daß es nicht regnet. *Wenn* ich nun zugleich bereit wäre, eine Wette mit einer Wettquote von 1:3 *gegen* die Aussage A zu akzeptieren, *dann* bedeutet dies, daß zugleich gilt: $\mathcal{P}(\neg A) \geq 1/3$, also $\mathcal{P}(A) \leq 2/3$. Die personelle Wahrscheinlichkeit befände sich also im Bereich $1/2 \leq \mathcal{P}(A) \leq 2/3$.

Damit kann man definieren: Die **personelle Wahrscheinlichkeit (der Überzeugungsgrad)** $\mathcal{P}(\ \ )$ eines Akteurs hinsichtlich einer Aussage A entspricht gerade der Wettquote, bei der der Akteur gleichermaßen geneigt wäre, *für* bzw. *gegen* A zu wetten.

Die würde bedeuten:

[128] Auch in der Wissenschaft finden sich solche Angaben nicht, wie auch Verfechter personeller Auffassungen von Wahrscheinlichkeitsaussagen zugestehen: "This account [nachdem Wissenschaftler die Wahrscheinlichkeit von Hypothesen ermitteln] bears no resemblance to our ordinary conception of science. Books on electrodynamics, for example, simply list Maxwell's equations as laws; they do not add a degree of confirmation." Richard C. Jeffrey, Valuation and Acceptance of Scientific Hypotheses. *Philosophy of Science* 23 (3), 1956, S. 237-246, hier: S. 246.

$\mathcal{P}(A) = 0.5$ bedeutet: Nichts spricht gegen eine Wette mit Wettquotienten 1:2 (oder besser);

$\mathcal{P}(A) = 0.8$ bedeutet: Nichts spricht gegen eine Wette mit Wettquotienten 8:10 (oder besser).

Man würde im letzten Fall erwarten, bei einem Einsatz von EUR 80 und einer Ausschüttung von EUR 100 in 10 Wetten 8 mal EUR 20 zu gewinnen und 2 mal EUR 80 zu verlieren: no gain, no loss.

Damit hätten wir eine Interpretation, was es bedeuten kann, daß jemand einer Aussage eine bestimmte Wahrscheinlichkeit zuschreibt. Man könnte sie sogar prinzipiell [129] durch durch eine Reihe von Wettangeboten ermitteln.

11.3. Kohärente Überzeugungen

Aber das Nachdenken über Wettquotienten leistet noch etwas mehr: Es kann dazu dienen zu zeigen, daß personelle Wahrscheinlichkeiten auch die Kolmogorov-Axiome erfüllen *sollten*: Dann und nur dann, wenn ein System von personellen Wahrscheinlichkeiten dies tut, sind (wie man sagt) die personellen Wahrscheinlichkeiten **kohärent**, und dies sollten sie sein: denn sonst gibt es kombinierte Wetten, die man alle akzeptieren würde, bei denen man aber insgesamt sicher verliert – sog. **Dutch books** (‘Niederländische Wetten’; woher die Bezeichnung stammt, ist unklar).

1. Es gilt

$0 \leq \mathcal{P}(A) \leq 1$

nach der Definition von Wettquotienten. (Einsatz und Ausschüttung des Quotienten sind positiv).

2. Weiterhin ist

$\mathcal{P}(\Omega) = 1$;

eine geringere Wettneigung auf A würde bedeuten, daß man Wetten gegen A akzeptieren würde (was unklug ist).

3. Das letzte Axiom ist die Additivität bei einander ausschließenden Aussagen A und B:

$\mathcal{P}(A \vee B) = \mathcal{P}(A) + \mathcal{P}(B)$;

Wir geben hier nur ein Beispiel dafür, daß es Dutch books gibt, wenn die Überzeugungsgrade gegen das Additivitätsaxiom verstoßen:

[129] Praktisch setzt dies freilich einiges voraus: Daß der- oder diejenige hinreichend liquide wäre, an Gewinn überhaupt interessiert, nicht während der Reihe von Wettangeboten unkonzentriert wird usw.

M. glaubt (20 %), daß Tyrion in Staffel 6 sterben wird (T); weiterhin (20 %), daß Daenerys in Staffel 6 sterben wird (D); schließlich (50 %), daß Tyrion oder Daenerys in Staffel 6 sterben wird (O). (Daß beide sterben, wäre etwas viel.) C. bietet M. drei Wetten an, gegen die M. bei einer Ausschüttung von jeweils 100 pennies nichts haben kann: 80 pennies gegen T; 80 pennies gegen D; 50 pennies auf O. Dann verliert M. sicher Geld, wenn er die Wetten zugleich akzeptiert (Abb. 10.1):

Staffel 6	80 gegen T	80 gegen D	50 auf O	**Gesamtgewinn**
$\neg T \wedge D$	+20	−80	+50	**−10**
$T \wedge \neg D$	−80	+20	+50	**−10**
$\neg T \wedge \neg D$	+20	+20	−50	**−10**

Abb. 10.1. Gewinne und Verluste der drei Wetten und Gesamtverlust bei einer Verletzung des Additivitätsaxioms (alles in pennies).

Ohne Beweis: **Ein System von Überzeugungsgraden ist genau dann kohärent, wenn es die Kolmogorov-Axiome erfüllt.** Bei der personellen Wahrscheinlichkeitsauffassung werden keine Anforderungen an einzelne Wahrscheinlichkeiten gestellt (Sie können glauben, was sie wollen …), wohl aber an die Überzeugungen hinsichtlich zusammenhängender Aussagen (… aber nicht alles auf einmal).

11.4. Bayes'sche Erkenntnistheorie

Der Ökonom Paul A. Samuelson berichtet, John Maynard Keynes habe auf den 'Vorwurf', er ändere am laufenden Band seine Ansichten, geantwortet: **"When my information change, I change my mind. What do you do, Sir?"** [130] Keynes war kein Verfechter der Bayes'schen Erkenntnistheorie.[131] Diese trägt ihren Namen (wie bereits erwähnt) nach dem Reverent (Geistlichen) Thomas Bayes, dem Verfasser des postum veröffentlichten *Essay Towards Solving a Problem in the Doctrine of Chan-*

[130] Marc Blaug, *John Maynard Keynes: Life, Ideas, Legacy.* Basingstoke, London: Macmillian, 1990.

[131] Im Gegenteil: Keynes ist einer der Hauptvertreter einer weiteren Art von Interpretationen der Wahrscheinlichkeit, die wir hier aussparen: *logischer (bedingter) Wahrscheinlichkeiten.* Die Idee ist hier, daß es so etwas gebe wie eine *objektive* bedingte Wahrscheinlichkeit einer Aussage A angesichts (gegeben) anderer Aussagen B: eine Wahrscheinlichkeit, mit der bestimmte Aussagen (B) eine andere (A) *stützen* (John M. Keynes, *A Treatise on Probability.* London: MacMillan and Co., 1921). Die Schwierigkeiten dieses Ansatzes hat Frank Ramsey – ein zu früh gestorbenes Genie – gut auf den Punkt gebracht:
"[A] more fundamental criticism of Mr Keynes' views [. . .] is the obvious one that there really do not seem to be any such things as the probability relations he describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. I do not perceive them, and if I am to be persuaded that they exist it must be by argument [. . .]"
(Frank Ramsey, Truth and Probability, 1926, in: id., *The Foundations of Mathematics and Other Logical Essays* (ed. R. B. Braithwaite). New York: Humanities Press, 1950, S. 161.)

ces.[132] Bayes'sche Ansätze, die von dieser Arbeit Gebrauch machen, kamen erst in den 50er Jahren des 20. Jahrhunderts wieder in Mode (Keynes starb 1946). Aber das Zitat faßt den Kerngedanken des Bayes'schen Ansatzes in der Erkenntnistheorie.

Die **Bayes'sche Theorie** (**Bayesian theory**) fußt wesentlich dem Satz von Bayes. Zur Erinnerung: Seien H und E Aussagen und $\mathcal{P}(E) \neq 0$. Dann gilt:

Kurze Form:

$$\mathcal{P}(H \mid E) = \mathcal{P}(H)\,\mathcal{P}(E \mid H) \,\big/\, \mathcal{P}(E) \qquad \textbf{(Satz von Bayes)}$$

Längere Form:

$$\mathcal{P}(H \mid E) = \mathcal{P}(H)\,\mathcal{P}(E|H) \,\big/\, \underbrace{\big(\,\mathcal{P}(E|H)\,\mathcal{P}(H) + \mathcal{P}(E|\neg H)\,\mathcal{P}(\neg H)\,\big)}_{\mathcal{P}(E)\ \text{(totale Wahrscheinlichkeit: R9)}} \qquad \textbf{(Satz von Bayes)}$$

Die Bayes'sche Theorie macht nun zwei Annahmen:

(i) Alle in dem Satz von Bayes auftretenden kategorischen oder bedingten Wahrscheinlichkeiten werden als **personelle Wahrscheinlichkeiten** aufgefaßt: Sie bringen allein zum Ausdruck, für wie glaubwürdig die jeweiligen Aussagen gehalten werden. (Es gibt hier keine richtigen oder falschen Einschätzungen dieser Wahrscheinlichkeiten; die einzige Bedingung ist, daß das Gesamtsystem kohärent im oben genannten Sinn ist.)

(ii) Der Satz von Bayes wird als Regel interpretiert, wie man seine **Überzeugungen ändert** (**belief updating/belief revision**), wenn man etwas Neues lernt (sog. Bayes'sche Konditionalisierung (Bayesian conditionalization)). Wenn Überzeugungsänderungen dem Satz von Bayes genügen, ist die Forderung nach Kohärenz erfüllt.

Begrifflichkeit: Die einzelnen kategorischen und bedingten Wahrscheinlichkeiten tragen im Kontext der Bayes'schen Theorie besondere Bezeichnungen:

$\mathcal{P}(H)$ ist die **A-priori-Wahrscheinlichkeit** (**prior probability**, im Plural kurz **priors**). Sie bezeichnet die personelle Wahrscheinlichkeit von H *ohne Wissen um den Befund* E.

$\mathcal{P}(E \mid H)$ ist die **Likelihood** (**likelihood**). Warum dieser Ausdruck Likelihood genannt wird, erschließt sich nicht leicht; 'likelihood' ist eigentlich im Englischen weitgehend synonym zu 'probability' (Wahrscheinlichkeit). Hier fungiert der Ausdruck als Fachbegriff. Die Begriffsbildung stammt von Ronald Fisher, der zwischen 1912 und 1922 die sog. Maximum-Likelihood-Methode entwickelte.[133] Gemeint ist *nicht* die

[132] Thomas Bayes, An Essay Towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* 53, 1764, S. 370–418. (Nachdruck: G. A. Barnard, *Biometrika* 45, 1958. Deutsch in: *Ostwald's Klassiker der Exakten Wissenschaften* 169, 1908.)

[133] Bei Fisher finden sich drei verschiedene Begründungen dieser Methode; vgl. John Aldrich, R. A. Fisher and the Making of Maximum Likelihood 1912–1922. *Statistical Science* 12 (3), 1997, S. 162–176; zur Abgrenzung von probability und likelihood s. S. 169–170. Einen Vorläufer hat Fisher in Carl

Wahrscheinlichkeit der Befunde E angesichts der Hypothese H (bzw. schon, aber dies ist nicht die Pointe), sondern umgekehrt spricht man von der Likelihood der Hypothese H angesichts der Befunde E: Die Maximum-Likelihood-Methode fragt danach, welche Hypothese die Befunde am wahrscheinlichsten macht. Im Bayes'schen Kontext ist dieser Bezug eigentlich unpassend; der Begriff ist aber etabliert und praktisch. Wir verwenden den englischen Ausdruck.[134]

$\mathcal{P}(E)$ ist einfach die (personelle) Wahrscheinlichkeit des Befunds (Experiments, evidence) E.

$\mathcal{P}(H \mid E)$ schließlich nennt man **A-posteriori-Wahrscheinlichkeit** (**posterior probability**). Sie bezeichnet die personelle Wahrscheinlichkeit von H *mit Wissen um den Befund* E.

Die A-posteriori-Wahrscheinlichkeit einer Hypothese H ist soll also proportional zum Produkt von deren A-priori-Wahrscheinlichkeit und der Likelihood sein:

$$\underbrace{\mathcal{P}(H \mid E)}_{\text{A-posteriori-W.}} = \underbrace{\mathcal{P}(H)}_{\text{A-priori-W.}} \cdot \underbrace{\mathcal{P}(E \mid H)}_{\text{Likelihood}} \Big/ \mathcal{P}(E) \qquad \textbf{(Satz von Bayes)}$$

Der Likelihood kommt dabei die zentrale Rolle zu: Während verschiedene Leute L verschiedenen Hypothesen H_i verschiedene personelle A-priori-Wahrscheinlichkeiten $\mathcal{P}_L(H_i)$ zuordnen können (und aus Sicht des Bayes'sche Ansatzes hieran nichts auszusetzen ist), werden ihre A-posteriori-Wahrscheinlichkeiten sich mit immer neuen Erfahrungsbefunden E_x immer weiter annähnern – man spricht etwas flapsig vom '**Ausschwemmen der A-priori-Wahrscheinlichkeiten**' (**washing out of the priors**). Die Hypothese mit der größten Likelihood angesichts der Befunde wird mit immer neuen Befunden immer wahrscheinlicher werden. (Vorausgesetzt ist dabei, daß alle die Likelihoods gleich beurteilen.)

11.5. Beispiel: HRP-2 Schnelltest

Der HRP-2 Schnelltest zur Bestimmung von *Plasomdium falciparum* dient zur Diagnose von *Malaria tropica* und zeigt eine hohe Sensitivität (88%) und Spezifität (95%) (Minimalwerte). Die Inzidenz (Häufigkeit von Neuerkrankungen) von Malaria in Deutschland ist sehr gering.

Sei M die Hypothese, ein Patient habe Malaria, und + ein positives Schnelltestresultat.

Friedrich Gauss.

[134] Im Deutschen ist die technische Entsprechung 'Mutmaßlichkeit' (da kann man gleich beim Englischen Ausdruck bleiben); bisweilen auch 'Erwartbarkeit' (dies ist noch unglücklicher, insofern 'Erwartungen' mit Erwartungswerten verbunden sind, also mit frequentistischen Wahrscheinlichkeitsauffassungen, um die es hier gerade nicht geht).

$\mathcal{P}(M|+) = \mathcal{P}(M)\,\mathcal{P}(+|M) \,\big/\, \big(\,\mathcal{P}(+|M)\,\mathcal{P}(M) + \mathcal{P}(+|\neg M)\,\mathcal{P}(\neg M)\,\big)$

Eine Ärztin A, die diesen Patienten mit HRP-2 testet, könnte nun diesen Patienten nach einer Tropenreise und Vorstellung mit Fieberbeschwerden testen, und eine Malaria-Erkrankung ernsthaft in Betracht ziehen ($\mathcal{P}_A(M) = 0,08$):

$$\mathcal{P}_A(M|+) = \underbrace{\mathcal{P}_A(M)}_{0{,}08}\ \underbrace{\mathcal{P}(+|M)}_{0{,}88} \,\big/\, \big(\ \underbrace{\mathcal{P}(+|M)}_{0{,}88}\ \underbrace{\mathcal{P}_A(M)}_{0{,}08} + \underbrace{\mathcal{P}(+|\neg M)}_{0{,}05}\ \underbrace{\mathcal{P}_A(\neg M)}_{0{,}92}\ \big) \approx 0,6.$$

(Dabei ist $\mathcal{P}_A(\neg M) = 1 - \mathcal{P}_A(M) = 0,92$ und $\mathcal{P}(+|\neg M) = 1 - \text{Spezifizität} = 0,05$.)

Ein Arzt B, der die Anamnese nicht kennt, könnte einen routinemäßigen Test nach einer Tropenreise vermuten, ohne wirklich eine Malaria-Erkrankung anzunehmen ($\mathcal{P}(M) = 0,005$), und das selbe Resultat anders deuten:

$$\mathcal{P}_B(M|+) = \underbrace{\mathcal{P}_B(M)}_{0{,}005}\ \underbrace{\mathcal{P}(+|M)}_{0{,}88} \,\big/\, \big(\ \underbrace{\mathcal{P}(+|M)}_{0{,}88}\ \underbrace{\mathcal{P}_B(M)}_{0{,}005} + \underbrace{\mathcal{P}(+|\neg M)}_{0{,}05}\ \underbrace{\mathcal{P}_B(\neg M)}_{0{,}995}\ \big) \approx 0,08.$$

In beiden Fällen erhöht der positive Test die Wahrscheinlichkeit:

$\mathcal{P}(M|+) > \mathcal{P}(M)$.

Je häufiger der Test wiederholt würde, um so mehr würden sich die Auffassungen beider $\mathcal{P}M = 1$ annähern (wenn die Tests weiter positiv ausfallen): Die A-priori-Wahrscheinlichkeiten werden nach und nach immer unbedeutender.

Dies gilt sowohl bei deterministischen Hypothesen ($\mathcal{P}(E|H) = 1$) wie auch bei statistischen ($\mathcal{P}(E|H) < 1$). Gibt es mehrere Hypothesen, die miteinander verglichen werden, so konvergieren die A-posteriori-Wahrscheinlichkeiten dann, wenn sich die Likelihoods der Hypothesen angesichts der Befunde unterscheiden.

11.6. Objektivität

Der klassische (frequentistische) und der Bayes'sche Ansatz sind die beiden prominentesten Arten der Interpretation von Wahrscheinlichkeitsschlüssen – beide Ansätze treten in vielen Variationen auf.[135]

Verfechter der klassischen und der Bayes'schen Statistik werfen sich gegenseitig vor, daß sich beklagenswerte subjektive Aspekte beim jeweils anderen Ansatz fänden. Verfechter Bayes'scher Ansätze verweisen v.a. auf die Bedeutung möglicher, aber nie realisierter Ausfälle, die Abhängigkeit des Ergebnisses eines Signifikanztests vom (intendierten) Testdesign und der Art der Stichprobenentnahme, der Teststatistik und dem willkürlichen Signifikanzniveau.[136]

[135] Daneben gibt es weitere Ansätze, etwa Likelihood-Verfahren.
[136] Vgl. Howson/Urbach 2006, S. 131–182.

Verfechter frequentistischer Ansätze verweisen auf die Unbestimmtheit der A-priori-Wahrscheinlichkeiten sowie auf die Unmöglichkeit, die Wahrscheinlichkeit der Erfahrungsbefunde $\mathcal{P}(E)$ objektiv zu bestimmen: dafür ist man auch die totale Wahrscheinlichkeit angewiesen, die wiederum auf unbestimmte Ausdrücke führt ($\mathcal{P}(E|\neg H)$; dies entspricht einer (unendlichen?) Summe von Likelihoods alternativer (wechselseitig unverträglicher) Hypothesen – die sog. catchall hypothesis).

Die Bewertung dieser Fragen ist nach wie vor ungeklärt und Gegenstand gegenwärtiger Diskussionen in der philosophischen Forschung.

Literatur

Hacking = Ian Hacking, *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press, [9] 2009.

Hamblin = Charles Leonard Hamblin, *Fallacies*. London: Methuen & Co., 1970.

Howson/Urbach = Colin Howson, Peter Urbach, *Scientific Reasoning: The Bayesian Approach*. La Salla, Il.: Open Court, 1989. Chicago, La Salle 2006. [3] 2006.

Kapp = Ernst Kapp, *Greek Foundations of Traditional Logic*. New York: Columbia University Press, 1942.

Kneale = William Kneale, Martha Kneale, *The Development of Logic*. Oxford: Clarendon Press, 1962.

Salmon = Wesley Salmon, *Logic*. Englewood Cliffs, N.J.: Prentice-Hall, [2] 1973.

Symbole

Logische und semantische Folgerungsbeziehung, metasprachliches Vokabular

$\therefore$; ——— ; *folglich*	logische/semantische Folgerungsbeziehung (zumal zwischen Aussagen)

Aussagenlogik

A, B, C usw.	Platzhalter für Aussagen bzw. Ereignisse
$\neg$	Negation (auch – oder $\sim$); “nicht”
$\wedge$	Konjunktion (auch $\cdot$ oder &); “und”
$\vee$	Adjunktion (Disjunktion); “einschließendes oder”
$\supset$	Subjunktion (materiales Konditional; auch $\rightarrow$); “wenn … dann”
$\equiv$	Bisubjunktion (materiales Bikonditional; auch $\leftrightarrow$); “genau dann … wenn”

Wahrscheinlichkeitstheorie

$\mathcal{P}(A)$	Wahrscheinlichkeit von A
Ω	Tautologie bzw. sicheres Ereignis
$\mathcal{P}(A \mid B)$	bedingte Wahrscheinlichkeit von A gegeben B
$E(H)$	Erwartungswert von H
$U(H)$	Nutzen(funktion) von H

Inferentielle Statistik

p (auch: P)	p-Wert oder P-Wert (Wahrscheinlichkeitsintegral)
H_0	Null-Hypothese
α	Fehlerhäufigkeit für Fehler 1. Art

Geläufige Abkürzungen

gdw.	genau dann, wenn / dann und nur dann, wenn
iff	if and only if
wff	well-formed formula (wohlgeformter Ausdruck; in AL: Satz)