Gerhard Schurz
Wahrscheinlichkeit

Grundthemen Philosophie

Herausgegeben von
Dieter Birnbacher
Pirmin Stekeler-Weithofer
Holm Tetens

Gerhard Schurz

Wahrscheinlichkeit

DE GRUYTER

ISBN 978-3-11-042550-5
e-ISBN (PDF) 978-3-11-042036-4
e-ISBN (EPUB) 978-3-11-042056-2

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Einbandabbildung: Martin Zech
Satz: fidus Publikations-Service GmbH, Nördlingen
Druck und Bindung: Hubert & Co. GmbH & Co. KG, Göttingen
♾ Gedruckt auf säurefreiem Papier
Printed in Germany

www.degruyter.com

Persönliches Vorwort

In den über 30 Jahren, in denen ich mich der Wissenschaftstheorie und wissenschaftlichen Philosophie gewidmet habe, war die Wahrscheinlichkeitstheorie (neben anderen formalen Methoden) mein ständiger Begleiter. Dies erklärt sich einerseits dadurch, dass ich (schon aufgrund meiner Doppelausbildung in Naturwissenschaft und Philosophie) immer ein den Fachwissenschaften nahestehender Wissenschaftsphilosoph war, und andererseits durch die Natur der Wissenschaften und ihrer Gegenstände selber. Erstens sind die meisten in den Wissenschaften erforschten Gesetzmäßigkeiten (eventuell mit Ausnahme der klassischen Physik) nicht strikt deterministischer, sondern statistischer Natur. Zweitens ist wissenschaftliches Wissen, außer in formalwissenschaftlichen Gebieten, grundsätzlich unsicher. Aus dem ersten Grund benötigt man die statistische und aus dem zweiten die epistemische Wahrscheinlichkeitstheorie. Eine theoretische Philosophie, die nicht im philosophischen Elfenbeinturm verbleiben möchte, sondern für die Realwissenschaften bedeutsame und anwendbare Ergebnisse sucht, kann nicht im Rahmen der deduktiven Logik verbleiben, sondern muss die Wahrscheinlichkeitstheorie mit einbeziehen. Diese Feststellung heißt keineswegs, dass deshalb die deduktive Logik unwichtig sei oder auf sie verzichtet werden könne. Nein, auch im Kern der Wahrscheinlichkeitstheorie steckt (als notwendiger, aber nicht hinreichender Bestandteil) die deduktive Logik, und wer die letztere ignoriert, kann auch die erstere nicht verstehen.

So erklärt sich, warum die Wahrscheinlichkeitstheorie mein ständiger Begleiter war: weil ich sie für Anwendungen benötigte. Dieter Birnbacher habe ich es zu verdanken, dass aus dieser Situation der Anstoß wurde, meine angesammelten Erfahrungen und Einsichten in diesem Gebiet in Form eines Büchleins zusammenzuschreiben. Das Besondere an diesem Büchlein ist – so meine ich zumindest – dass ich die Wahrscheinlichkeitstheorie nicht, wie üblich, aus einer bestimmten Ecke betrachte und entwickle, sondern die unterschiedlichen Perspektiven, Terminologien und Wissenschaftslager zusammenzubringen versuche.

Wie ich es sehe, befindet sich das Gebiet der Wahrscheinlichkeitstheorie seit Jahrzehnten in einem bedauerlichen Zustand der Lagertrennung. Während in den empirischen Wissenschaften fast ausschließlich von statistischer Wahrscheinlichkeit die Rede ist, verstehen die in der Wissenschaftstheorie einflussreichen Bayesianer Wahrscheinlichkeit grundsätzlich im subjektiven Sinn von rationalen Glaubensgraden, wogegen die dritte Gruppe der mathematischen Wahrscheinlichkeitstheoretiker diesen Interpretationskonflikt systematisch ignoriert. Als anwendungsorientierter Wissenschaftsphilosoph bin ich nicht selten vor die unangenehme Situation gestellt, dass ich, um anwendbare Ergebnisse zu haben, den statistischen Wahrscheinlichkeitsbegriff benutze, während mir Kolleg/inn/en entgegenhalten, es wäre doch „gezeigt worden“, dass der „frequentistische Wahrscheinlichkeitsbegriff“ nicht funktioniere und man besser entweder den bayesianischen Begriff des Glaubensgrades oder den metaphysischen Begriff der objektiven Einzelfallpropensität verwenden solle. Doch die angewandten Wissenschaften verwenden ohne Umstände die Methoden der Statistik – die, wie zu sehen sein wird, bei rechtem Verständnis des Begriffes des Häufigkeitsgrenzwertes sehr gut funktionieren –, während der subjektive Wahrscheinlichkeitsbegriff wie auch der metaphysische Begriff der Einzelfallpropensität in den Wissenschaften eine vergleichsweise untergeordnete Rolle spielen. Diese verzwickte Lage war ein weiterer Anstoß für mich, dieses Buch zu schreiben.

Schlussendlich bin ich zur Auffassung gelangt, dass man beide Wahrscheinlichkeitsbegriffe benötigt, weshalb ich in diesem Buch eine *dualistische* Position entwickle, der es vor allem darum geht, die Brückenprinzipien zwischen beiden Wahrscheinlichkeitsbegriffen herauszuarbeiten. Dies beginnt damit, dass ich die Wahrscheinlichkeitsfunktion über der Algebra der Eigenschaften respektive Propositionen einer interpretierten Sprache entwickle, weil nur so der logische Unterschied zwischen statistischer und epistemischer Wahrscheinlichkeit (erstere besitzt offene Formeln und letztere Sätze als Argumente) explizit gemacht werden kann, was Voraussetzung für die Herausarbeitung der sie verbindenden Brückenprinzipien ist. In Anlehnung an einen bekannten Passus von Kant lässt sich die dualistische Position so formulieren: Subjektive ohne statistische Wahrscheinlichkeitstheorie ist blind (also irrational), statistische ohne subjektive Wahrscheinlichkeitstheorie ist leer (also ohne empirischen Bezug). Die dualistische Position bedeutet jedoch nicht, dass nun alles, was in beiden Positionen behauptet wurde, übernommen werden kann – dies würde schnell zu Widersprüchen führen. Vielmehr müssen in beiden Positionen zugleich gewisse Anteile als „kaum haltbar“ fallen gelassen werden. Beispielsweise sehe ich die radikal-subjektivistische Position im Bayesianismus als ebenso wenig haltbar an wie die These, dass rationale Subjekte zu jeder Sachfrage (auch dann, wenn sie darüber keine Erfahrungen gesammelt haben) rationale apriorische Glaubensgrade in Form fairer Wettquotienten besitzen müssen. Umgekehrt meine ich, dass die traditionelle Auffassung, der statistische Begriff des Häufigkeitsgrenzwerts sei ein zumindest „approximativ“ empirisch gehaltvoller Begriff, ebenfalls korrekturbedürftig ist, da der empirische Gehalt dieses Begriffs nicht deduktiver, sondern induktiver Natur ist und man zu seiner Formulierung den epistemischen Wahrscheinlichkeitsbegriff benötigt. Viele weitere Einsichten dieser Art ergeben sich im dualistischen Ansatz, über den ich jetzt aber nichts weiter verraten will, sondern stattdessen der Leserin und dem Leser viel Vergnügen beim Lesen dieses Büchleins wünsche.

Düsseldorf, im Mai 2014 Gerhard Schurz

Technische Anmerkungen

Der *logisch-mathematische Anhang* enthält wesentliche Ergänzungen, teils einführender und teils fortgeschrittener Art. Anhang 10.1 präsentiert elementare logische Voraussetzungen zum Nachschlagen für den Neuling. Anhang 10.2 erläutert die Details des logischen Aufbaus von Wahrscheinlichkeitsfunktionen über Algebren und logischen Sprachen für den fortgeschrittenen Leser. Anhang 10.3 präsentiert schließlich die Beweise sämtlicher in diesem Buch angeführten Theoreme.

Die Nummerierung von *Abbildungen* erfolgt nach folgendem Muster „Abb. KapitelNr-Nr". Beispiel: „Abb. 2.1-2" ist die 2. Abbildung von Kap. 2.1. Analog werden *Definitionen*, *Merksätze* und *Hervorhebungen* nummeriert. Beispiel: „(Def. 4.2-2)" ist die 2. Definition von Kapitel 4.2. „(3.1-4)" ist die 4. Hervorhebung von Kap. 3.1. Einfache Anführungszeichnen stehen für stilistische und doppelte für wörtliche Zitationen.

Inhaltsverzeichnis

1 Einführung

Der intuitive Begriff der Wahrscheinlichkeit involviert zugleich etwas Objektives („wahr-“) und etwas Subjektives („-scheinlich“). Die frühen Begründer der Wahrscheinlichkeitstheorie hatten diese Doppeldeutigkeit nur unzureichend bemerkt. Erst im 20. Jahrhundert wurde die unterschiedliche Natur der beiden Wahrscheinlichkeitsbegriffe herausgearbeitet.

Die Theorie der Wahrscheinlichkeit entstand im 16. und 17. Jahrhundert, überwiegend im Kontext von *Glücksspielen*.[1] Wichtige Dokumente aus dieser Zeit sind unter anderem ein Manuskript von Galilei aus dem frühen 17. Jahrhundert, ein Briefwechsel zwischen Pascal und Fermat aus dem Jahre 1654, sowie eine 1657 veröffentlichte Schrift von Huygens, alle drei über die Frage, wie die Ergebniswahrscheinlichkeiten bei Glücksspielen mit mehreren Würfeln zu berechnen sind. 1713 erschien die berühmte Schrift von Bernoulli, in der er die Binomialverteilung und das schwache Gesetz der großen Zahlen bewies, und 1763 wurde das Theorem von Bayes veröffentlicht (Bayes und Price 1763). Erst ein halbes Jahrhundert später, 1814, veröffentlichte Laplace seine Essays über Wahrscheinlichkeiten, mit denen man die Geschichte der Wahrscheinlichkeitstheorie oftmals beginnen lässt. Laplaces Wahrscheinlichkeitsansatz war aus heutiger Sicht ein epistemischer, da Laplace allen epistemisch (d. h. erkenntnismäßig) gleichermaßen möglichen Ergebnissen eines Experimentes oder physikalischen Prozesses dieselbe Wahrscheinlichkeit zuschrieb. Allerdings unterschied Laplace dieses subjektive „Gleichverteilungsprinzip“ nicht von der objektiven Gleichwahrscheinlichkeit der Wurfresultate eines regulären Würfels (Laplace 1814, 6f). Erst von Mises (1928, 69) machte den Unterschied deutlich, als er die Frage stellte, wie man wahrscheinlichkeitstheoretisch mit einem irregulären, z. B. einseitig magnetisierten Würfel umgehen sollte, dessen Ergebnisse zwar ebenfalls epistemisch unbekannt, aber statistisch gesehen eben nicht gleichwahrscheinlich sind. Innerhalb von Laplaces Theorie konnte zwischen den beiden Fällen nicht differenziert werden, da nicht zwischen subjektiver und objektiver Wahrscheinlichkeit unterschieden wurde.[2]

Die gegenwärtige Wahrscheinlichkeitstheorie ist durch eine anhaltende Lagertrennung gekennzeichnet. Während in den empirischen Wissenschaften fast ausschließlich von objektiver statistischer Wahrscheinlichkeit die Rede ist, verstehen die in der Wissenschafts- und Erkenntnistheorie einflussreichen Bayesianer Wahrscheinlichkeit im subjektiven Sinn von rationalen Glaubensgraden, wohingegen die mathematischen Wahrscheinlichkeitstheoretiker diesen Interpretationskonflikt systematisch ignorieren. Neben diesen drei Hauptgruppen gibt es einige philosophisch bedeutende Untergruppen. Im objektiven Lager beispielsweise die Propensitätstheoretiker, welche Wahrscheinlichkeiten als objektive physikalische Tendenzen auffassen, im subjektiven Lager die Vertreter der logischen Wahrscheinlichkeitstheorie oder des objektiven Bayesianismus, wobei hier das „objektiv“ nicht im Sinne von „subjekt-extern“, sondern von „intersubjektiv“ aufzufassen ist. Last but not least sei die Gruppe der dualistischen Wahrscheinlichkeitstheorien erwähnt, zu der ich auch meinen Ansatz rechne.

Obwohl es natürlich erscheint, die objektiven Wahrscheinlichkeitsauffassungen den „subjektiven“ Auffassungen gegenüberzustellen, schließe ich mich Gillies (2000, 2) an und bezeichne die Auffassungen von Wahrscheinlichkeit als rationalem Glaubensgrad als die Familie der *epistemischen* Wahrscheinlichkeitstheorien. Denn diese Familie enthält neben den Subjektivisten auch die „logischen“ Wahrscheinlichkeitstheoretiker und „objektiven“ Bayesianer, welche die Bezeichnung ihrer Wahrscheinlichkeiten als „subjektiv“ zurückweisen, weil darin die Bedeutung von „subjektiv variabel“ mitschwingt, die zwar auf einige (z. B. personalistische), aber nicht auf alle epistemische Wahrscheinlichkeitstheorien zutrifft.

Zu den Hauptbegründern der statistischen Wahrscheinlichkeitstheorie zählen u. a. von Mises (1964), Reichenbach (1935, 1949), und Fisher (1956) (als Einführungsliteratur in Statistik sei Hays/Winkler 1970 und Bortz 1985 empfohlen). Die Hauptbegründer der subjektiven Theorie sind u. a. Ramsey (1926) und de Finetti (1970) (einführende Literatur findet sich z. B. in Earman 1992 und Howson/Urbach 1996). Keynes (1921) und Carnap (1950, 1971 und 1980) begründeten die logische Wahrscheinlichkeitstheorie, die den Grundaxiomen der Wahrscheinlichkeitstheorie weitere Axiome oder Prinzipien hinzufügt, welche die Glaubensgrade für alle rationalen Subjekte „apriorisch“ fixieren sollen. Es besteht allerdings weitgehend Einigkeit, dass diese Zusatzaxiome weit über den Bereich des logisch-analytisch Gültigen hinaus gehen, weshalb ich das ‚logische‘ dieser Wahrscheinlichkeitstheorien in Anführungszeichen setze (zum Begriff der logischen bzw. analytischen Wahrheit/Gültigkeit s. Anhang 10.1.3). Repräsentative Einführungen in die mathematische Wahrscheinlichkeitstheorie sind Bauer (1996) oder Billingsley (1995); Überblicke über verschiedene Wahrscheinlichkeitsansätze geben z. B. Fine (1973), Stegmüller (1973a,b), Kutschera (1972, Kap. 2), Howson/Urbach (1996) und insbesondere Gillies (2000).

2 Objektive und epistemische Wahrscheinlichkeit

Die objektive Wahrscheinlichkeit drückt eine subjektunabhängige Eigenschaft der Realität aus. Die subjektive Wahrscheinlichkeit drückt dagegen den Glaubensgrad eines (aktualen oder hypothetischen) rationalen Subjekts aus. Falls es sich dabei um intersubjektive Glaubensgrade handelt, sprechen wir allgemeiner von „epistemischer“ Wahrscheinlichkeit (Näheres dazu weiter unten). Def. 2-1 präsentiert die grundlegende Definition statistischer und subjektiver Wahrscheinlichkeiten, wobei wir uns der symbolischen Schreibweise bedienen. Dabei steht „Fx“ für „x ist ein F“ und „Fa“ für „a ist ein F“; „F“ ist ein Prädikat, das ein Merkmal oder einen Ereignistyp F bezeichnet; „x“ ist eine Individuenvariable und „a“ eine Individuenkonstante, die ein variables resp. bestimmtes Individuum bezeichnen. Eine Kurzeinführung in logische Notationen und Begriffsarten findet sich in Anhang 10.1.

(Def. 2-1) Die *statistische (objektive)* Wahrscheinlichkeit eines Merkmals oder wiederholbaren Ereignistyps, z. B. Fx, ist die relative Häufigkeit seines Eintretens bzw. der Grenzwert seiner relativen Häufigkeit auf lange Sicht. *Formal* schreiben wir dafür ein „kleines“ p(–): p(Fx) steht somit für die Häufigkeit bzw. den Häufigkeitsgrenzwert, mit der beliebige Individuen x eines gegebenen Bereichs die Eigenschaft F besitzen. – Beispiel: Die Häufigkeit von Sonnentagen in Düsseldorf.

Die *epistemische (subjektive)* Wahrscheinlichkeit eines bestimmten Ereignisses bzw. Sachverhaltes, z. B. Fa, ist der rationale Glaubensgrad, in dem ein gegebenes Subjekt oder alle Subjekte eines bestimmten Rationalitätstyps an das Eintreten des Ereignisses glauben. *Formal* schreiben wir dafür ein „großes“ P(–): P(Fa) steht somit für den Glaubensgrad dafür, dass das bestimmte Individuum a die Eigenschaft F besitzt. – Beispiel: Der Grad unseres Glaubens, dass der morgige Tag in Düsseldorf ein Sonnentag sein wird.

Die *relative Häufigkeit* h(Fx) eines Ereignistyps Fx in einem *endlichen* Individuenbereich („domain“) D ist die Anzahl aller F’s in D geteilt durch die Anzahl aller Individuen in D. Für endliche Individuenbereiche identifizieren wir die statistische Wahrscheinlichkeit mit der relativen Häufigkeit, also p(Fx) = h(Fx). Falls D dagegen unendlich ist, ist die relative Häufigkeit undefiniert. In diesem Fall bezieht man sich auf eine zufällige Anordnung der Individuen in D in Form einer (unendlichen) *Zufallsfolge* $(d_1,d_2,\ldots)$. Die relative Häufigkeit $h_n(Fx)$ von F’s unter den ersten n Folgegliedern ist definiert als $h_n(F) =_{def} a_n(F)/n$, mit „$a_n(F)$“ als der Anzahl von F’s in den ersten n Folgegliedern. Man bestimmt nun die statistische Wahrscheinlichkeit p(Fx) als den Grenzwert der relativen Häufigkeiten $h_n(Fx)$ von

F's in n-gliedrigen Anfangsabschnitten der Zufallsfolge, für n gegen unendlich: $p(Fx) = \lim_{n\to\infty} h_n(Fx)$. Der Grenzwertbegriff ist wie folgt definiert:

(Def. 2-2) Der *Grenzwert der relativen Häufigkeit* des Ereignistyps Fx in einer gegebenen Zufallsfolge $(d_1, d_2,...)$ beträgt r, oder kurz $\lim_{n\to\infty} h_n(Fx) = r$, g. d. w. es für jedes noch so kleines $\epsilon>0$ eine Stellenzahl n gibt, sodass für alle $m \geq n$ die relative Häufigkeit $h_m(Fx)$ vom Grenzwert r um weniger als ϵ abweicht.

Zur Veranschaulichung ist in Abb. 2-1 die Konvergenz der relativen Häufigkeiten eines Ereignisses mit Häufigkeitsgrenzwert p(Fx) = 0,6 in zwei Zufallsfolgen dargestellt.

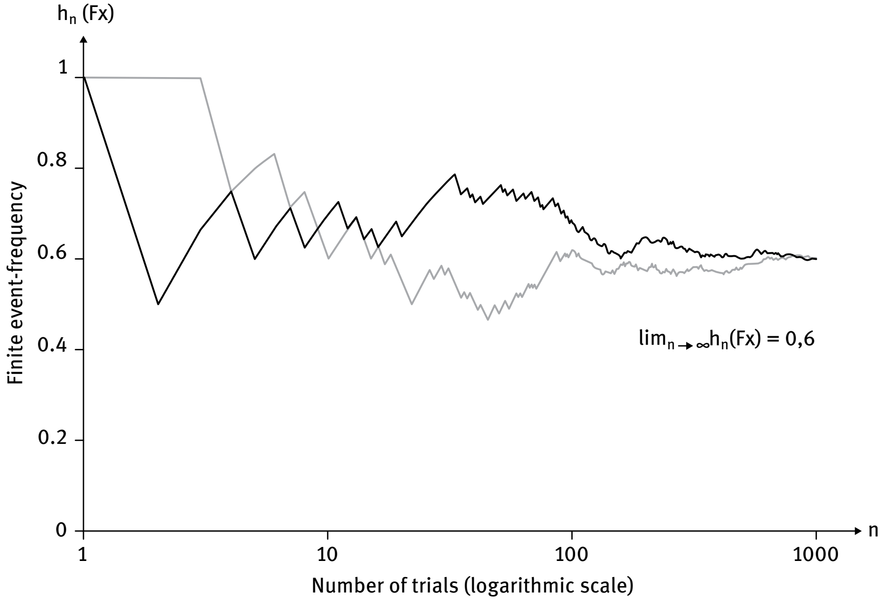


Abb. 2-1: *Konvergenz der relativen Häufigkeiten* eines Ereignisses mit Häufigkeitsgrenzwert p(Fx) = 0.6 in zwei Zufallsfolgen (programmiert in Visual Basic).

Häufigkeitsgrenzwerte sind *theoretische Idealisierungen*, die über die faktisch beobachtbaren Häufigkeiten hinausgehen. Zufallsfolgen werden durch wiederholte Durchführungen von sogenannten Zufallsexperimenten realisiert (Näheres in Kap. 5). Die Ergebnisse dieser Realisierungen bilden die Glieder der Zufallsfolge. Bei der Rede von Häufigkeitsgrenzwerten handelt es sich daher um die Behauptung, dass das zugrundeliegende Zufallsexperiment eine gewisse *Disposition* besitzt, das fragliche Ergebnis Fx mit einer auf lange Sicht konvergierenden

Häufigkeit zu produzieren. Man nennt diese Disposition auch eine „generische Propensität“ (Kap. 5.2). Die Rede von Häufigkeitsgrenzwerten geht ontologisch weit über aktuale Häufigkeiten hinaus, weshalb ich es bevorzuge, von „statistischer“ – anstatt wie häufig im englischsprachigen Bereich von „frequentistischer“ – Wahrscheinlichkeit zu sprechen. Auch das zufällige Ziehen eines Individuums aus einem Individuenbereich D stellt ein Zufallsexperiment dar. Dabei ergibt sich folgender Zusammenhang zu endlichen Häufigkeiten: die Häufigkeit des Merkmals F in einem endlichen Bereich D stimmt genau dann mit der statistischen Wahrscheinlichkeit, zufällig ein F-Individuum aus D zu ziehen, überein, wenn jedes Individuum in D dieselbe statistische Chance besitzt gezogen zu werden – dies ist die grundlegende Forderung an Zufallsziehungen. Formal gesprochen muss für alle $d \in D$ gelten: $p(x=d) = 1/|D|$ (mit „$x=d$“ für „das Ziehungsresultat x war d“ und „$|D|$“ für die Kardinalität von D).

Wann immer sich statistische Wahrscheinlichkeiten auf die Ergebnisse von beliebig oft wiederholbaren Experimenten beziehen, macht es wenig Sinn, diese mit endlichen Häufigkeiten zu identifizieren, da die Ergebnisse von nur endlich vielen Versuchsdurchführungen immer von *Zufälligkeiten* mitbestimmt sind, die die zugrundeliegenden Wahrscheinlichkeiten verzerren. Aus diesem Grund erscheint die Bezugnahme auf Häufigkeitstendenzen im Grenzwert die einzig akzeptable Lösung. Nur wenn sich statistische Wahrscheinlichkeiten explizit auf Verteilungen in einer endlichen Population beziehen, die von keinem bekannten Zufallsexperiment bzw. Kausalprozess erzeugt wurden, ist deren Identifikation mit endlichen Häufigkeiten sinnvoll.

Damit der so definierte statistische Wahrscheinlichkeitsbegriff eine objektive Eigenschaft der Realität ausdrückt, muss der Begriff der Zufallsfolge in zufriedenstellender Weise charakterisiert werden. Dies ist das Hauptproblem der statistischen Theorie, mit dem wir uns in Kap. 5 beschäftigen. Das Hauptproblem der epistemischen Wahrscheinlichkeitstheorie liegt dagegen darin, dass unterschiedliche Subjekte, auch *wenn* sie dieselben Erfahrungen machen, denselben Propositionen unterschiedliche Glaubensgrade zuordnen können. Das Problem der epistemischen Theorie besteht also darin, subjektive Glaubensgrade zu rationalisieren und zu *objektivieren*, womit wir uns in Kap. 6 und 7 beschäftigen.

Zur Interpretation von Eins- und Nullwahrscheinlichkeiten: Im epistemischen Fall bedeutet die Aussage $P(A) = 1$ einfach, dass sich das gegebene Subjekt hinsichtlich der Aussage bzw. Proposition A *sicher* ist, d. h. in keiner Weise an der Wahrheit von A zweifelt – wie immer es um den faktischen Wahrheitswert von A bestellt sein mag. Im statistischen Fall bedarf die Bedeutung von $p(Fx) = 1$ dagegen einer näheren Erläuterung. *Nur* im Falle eines *endlichen* Individuenbereichs ist $p(Fx) = 1$ gleichbedeutend mit dem strikten bzw. ausnahmslosen Allsatz $\forall xFx$ (Alle Individuen sind F), bzw. $p(Fx) = 0$ gleichbedeutend mit $\forall x\neg Fx$. Im Fall eines unendlichen Individuenbereichs ist p(Fx) = 1 dagegen *schwächer* als ∀xFx (und p(Fx) = 0 schwächer als ∀x¬Fx). Denn gegeben eine unendliche Zufallsfolge (d_1, d_2,...) und ein Ereignistyp Fx, dann impliziert p(Fx) = 0 nicht, dass es in dieser Folge *kein* Individuum d_i gibt, welches das Merkmal F hat, sondern lediglich, dass die Häufigkeiten h_n(Fx) *gegen null konvergieren.* Sei die Zufallsfolge beispielsweise die Ordnung der natürlichen Zahlen ℕ, also 1, 2,..., und bezeichne Fx das Prädikat „x ist eine ganzzahlige Potenz von 2". Es gibt unter den natürlichen Zahlen *unendlich* viele ganzzahlige 2er-Potenzen, nämlich alle Zahlen der Form 2^k für ein k∈ℕ. Dennoch gilt $\lim_{k\to\infty}$ p(Fx) = $\lim_{k\to\infty}(k/2^k) = 0$, denn unter den ersten 2^k natürlichen Zahlen befinden sich k Zweierpotenzen, und das Verhältnis von k zu 2^k geht gegen null, wenn k unendlich groß wird. Die statistische Wahrscheinlichkeit dafür, dass eine natürliche Zahl eine (bzw. keine) Zweierpotenz ist, beträgt damit null (bzw. eins). Die statistische Hypothese p(Fx) = 1 lässt somit beliebig und sogar unendlich viele Ausnahmen zu, sofern deren Häufigkeit nur gegen Null konvergiert; sie ist also wesentlich schwächer als die Allaussage ∀xFx.

Statistische Wahrscheinlichkeiten beziehen sich immer auf einen *wiederholbaren* Ereignis*typ* bzw. Sachverhaltstyp, ausgedrückt durch ein Prädikat bzw. eine *offene Formel*, z. B. Fx.[3] Die subjektive Wahrscheinlichkeit bezieht sich dagegen auf ein *bestimmtes* Ereignis oder einen bestimmten Sachverhalt, ausgedrückt in einem *Satz* bzw. einer geschlossenen Formel, z. B. Fa. Denn nur Sätze mit bestimmter Bedeutung, aber nicht offene Formeln mit unbestimmter Bedeutung, können Gegenstand des Glaubens sein und Glaubensgrade besitzen. Ein *Beispiel:* Wenn gesagt wird, die Wahrscheinlichkeit dafür, dass es *morgen* in Düsseldorf regnet, betrage 3/4, so kann dies prima facie keine Häufigkeitsaussage, sondern nur eine epistemische Wahrscheinlichkeitsaussage sein. Denn den morgigen Tag gibt es nur *einmal* – entweder es regnet morgen oder es regnet morgen nicht. Prima facie kann mit einer Einzelfallwahrscheinlichkeit P(Fa) also nur eine epistemische Wahrscheinlichkeitsaussage – eine Aussage über z. B. meinen Glaubensgrad an Fa – gemeint sein. Die korrespondierende statistische Wahrscheinlichkeit p(Fx) kann dagegen nur einem wiederholbaren Ereignistyp Fx zugesprochen werden, etwa dass es an einem beliebigen Tag x in Düsseldorf regnet. Meine subjektive Wahrscheinlichkeit P(Fa_i) kann für verschiedene Individuen a_i beliebig variieren; der Häufigkeitsgrenzwert p(Fx) ist dagegen durch die Klasse aller Fs und durch den Individuenbereich D bzw. das zugrundeliegende Zufallsexperiment festgelegt und von keiner individuellen Instanziierung Fa_i abhängig. Syntaktisch bedeutet dies, dass der statistische Wahrscheinlichkeitsfunktor p(A) *sämtliche* freien Variablen in der Formel A *bindet* (ähnlich wie das ein Quantor tut). Der Ausdruck „p(Fx)" enthält also keine freie Variable; er ist gleichzusetzen mit p({x:Fx}) (mit „{x:Fx}" für „die Menge aller x die Fs sind") und bezeichnet den Häufigkeitsgrenzwert von Fx in D-Zufallsfolgen (vgl. Bacchus 1990, Kap. 3, der hierfür „$p_x(Fx)$" schreibt). Über die Variablen in statistischen Wahrscheinlichkeiten zu quantifizieren, also etwa $\forall x(p(Fx) = 0{,}5)$, wäre daher eine *syntaktische Konfusion*, wogegen die Allquantifikation für subjektive Wahrscheinlichkeiten Sinn macht: $\forall x(P(Fx) = 0{,}5)$ besagt, dass für jedes Individuum d in D der Glaubensgrad der Proposition „d ist ein F" 0,5 beträgt.

Trotz dieser grundlegenden Unterschiede gibt es zwischen den beiden Wahrscheinlichkeitsbegriffen Zusammenhänge. Das bekannteste Prinzip, um statistische Wahrscheinlichkeiten auf subjektive Einzelfallwahrscheinlichkeiten zu übertragen, ist das folgende auf Reichenbach (1949, § 72) zurückgehende Prinzip:

(Def. 2-3) *Prinzip der engsten Referenzklasse:* Die subjektive Wahrscheinlichkeit P(Fa) eines Einzelereignisses wird bestimmt als die (geschätzte) *bedingte* statistische Wahrscheinlichkeit p(Fx|Rx) des entsprechenden Ereignistyps Fx in der *engsten* (nomologischen) Bezugsklasse bzw. Referenzklasse R, von der das zugrundeliegende Subjekt weiß bzw. mit Sicherheit annimmt, dass a in ihr liegt (also Ra gilt).[4]

Das Prinzip der engsten Referenzklasse findet im Alltag und in den Wissenschaften durchgängige Verwendung. Wollen wir z. B. die subjektive Wahrscheinlichkeit dafür bestimmen, dass eine bestimmte Person eine bestimmte Berufslaufbahn einschlägt (Fa), so stützen wir uns auf die uns bekannten Eigenschaften dieser Person als engste Referenzklasse (Ra) und auf die statistische Wahrscheinlichkeit, dass eine Person x mit den Eigenschaften Rx diese Berufslaufbahn einschlägt (p(Fx|Rx)). In der obigen Wetterprognose „die Wahrscheinlichkeit dafür, dass es *morgen* regnet, beträgt 3/4" ist die engste Referenzklasse die vom Meteorologen berücksichtigte *vorausgehende* Wetterentwicklung. Diese Wetterprognose hat gemäß Reichenbachs Prinzip die folgende Deutung: die statistische Wahrscheinlichkeit dafür, dass es an einem Tag regnet, dem eine typologisch gleiche Wetterentwicklung vorausgeht wie dem heutigen Tag, beträgt 3/4. Dies meinen Meteorologen, wenn sie probabilistische Wetterprognosen anstellen.

Reichenbachs Prinzip steht in enger Beziehung zu folgendem Schlussprinzip:

(2-1) *Induktiver Spezialisierungsschluss* (vgl. Carnap 1950, 207f):
Prämisse 1: r % aller Fs sind Gs
Prämisse 2: Dies ist ein F
══════════════ [mit r % Glaubenswahrscheinlichkeit]
Konklusion: Dies ist ein G

Der induktive Spezialisierungsschluss wird auch „direct inference“ (Levi 1977) genannt. Der Doppelstrich „═“ deutet an, dass dieser Schluss unsicher ist, also nur mit einer gewissen Wahrscheinlichkeit (die rechts von Doppelstrich angemerkt ist) von wahren Prämissen zu einer wahren Konklusion führt. Wie bei allen unsicheren Schlüssen gilt das Prinzip der *Gesamtevidenz:* die singuläre Prämisse muss die gesamte für die Konklusion relevante Evidenz enthalten (s. Schurz 2006, 56, Ms. 2.6-4). Unter dieser Zusatzbedingung ist Reichenbachs Prinzip der engsten Referenzklasse eine Anwendung des induktiven Spezialisierungsschlusses, der seinerseits seine tiefere Begründung in dem in Kap. 7.1 besprochenen statistischen Koordinationsprinzip besitzt (Näheres zu induktiven Schlussarten in Kap. 4.1.4).

Mithilfe von Reichenbachs Prinzip der engsten Referenzklasse kann nur die subjektive Wahrscheinlichkeit von *Singulärsätzen* (also Sätzen, die Individuenkonstanten enthalten) durch statistische Wahrscheinlichkeiten bestimmt werden, nicht jedoch die subjektive Wahrscheinlichkeit von *generellen* Hypothesen wie z. B. „Alle Raben sind schwarz“ (formal $\forall x(Fx \rightarrow Gx)$) oder „50 % aller Münzwürfe landen auf Zahl“ (formal $p(Fx) = 0{,}5$). Die subjektive Wahrscheinlichkeit genereller Hypothesen hängt von ihrer subjektiven Ausgangswahrscheinlichkeit ab, die *nicht* auf statistische Wahrscheinlichkeiten zurückführbar ist (s. Kap. 7.3).

Die genaue Ausbuchstabierung des Prinzips der engsten Referenzklasse bringt eine Reihe von Problemen und Verfeinerungen mit sich, die in Kap. 7 erläutert werden. Dort wird auch die Forderung in Def. 2-3 (Klammer), die Bezugsklasse müsse „nomologisch sein“, genauer erläutert.

3 Mathematische Grundlagen der Wahrscheinlichkeit

3.1 Gesetze der Wahrscheinlichkeit

Der statistische und der epistemische Wahrscheinlichkeitsbegriff gehorchen denselben mathematischen Grundgesetzen, die erstmals von Kolmogorov (1933) axiomatisiert wurden. Zur Formulierung dieser Gesetze benutzen wir die üblichen logischen und mengentheoretischen Symbole, insbesondere: ¬ (Negation), ∧ (Konjunktion), ∨ (Disjunktion), → (materiale Implikation), ↔ (Äquivalenz), ∀ (Allquantor), ∃ (Existenzquantor), ∈ (Elementbeziehung), ∪ (Mengenvereinigung), ∩ (Mengendurchschnitt), – (Mengendifferenz), ∅ (leere Menge), ⊆ (unechte oder echte Teilmenge), ⊂ (echte Teilmenge). $f{:}A{\to}B$ (f ist Funktion von Menge A nach Menge B). „$=_{def}$" steht für „ist per definitionem identisch". Näheres zu Symbolik und Grundlagen der Logik und Mengenlehre s. Anhang 10.1. Wir werden den Gehalt von formalisierten Aussagen zwecks Einübung für den Leser oft in Worten wiederholen.

Kolmogorov präsentierte die Wahrscheinlichkeitsaxiome in der mathematisch üblichen *mengenalgebraischen* Darstellung. Hierbei stehen A, B, ... für *Teilmengen* eines sogenannten *Möglichkeitsraumes* Ω (die man auch „Ereignisse" nennt). Dabei ist Ω die Menge aller möglichen Ergebnisse eines *Zufallsexperimentes*. Ein Beispiel ist das Werfen eines Würfels (Ω = {1,2,3,4,5,6}) oder das Ziehen eines Individuums aus dem gegebenen Individuenbereich D, der auch Grundgesamtheit genannt wird (Ω = D). Teilmengen von Ω-Elementen werden als *Disjunktionen* von möglichen Ergebnissen des Experimentes aufgefasst; so entspricht die Menge {1,3,5} beim Würfelwurf der Aussage „es wurde eine 1, 3 oder 5, d. h. eine ungerade Zahl gewürfelt", usw. In der mengenalgebraischen Darstellung muss die Negation ¬A als das Komplement $A^c =_{def} \Omega - A$, die Disjunktion A∨B als die Vereinigung A∪B, und die Konjunktion A∧B als der Durchschnitt A∩B gelesen werden; insbesondere ist $A \vee \neg A = \Omega$ und $A \wedge \neg A = \emptyset$. Es wird angenommen, dass die Algebra von Teilmengen von Ω unter diesen Operationen geschlossen ist (Details in Kap. 3.3). Der mengenalgebraische Aufbau kann sowohl statistisch wie subjektiv interpretiert werden: Im statistischen Fall steht p(A) für den Häufigkeitsgrenzwert des Ereignisses A in einer Zufallsfolge von Experimentrealisierungen, und im subjektiven Fall steht P(A) für die Glaubenswahrscheinlichkeit an das Ereignis A in einer *einzelnen* Experimentrealisierung.

Für die Wissenschaftstheorie ist die *sprachliche* Darstellung der Wahrscheinlichkeitstheorie zu bevorzugen, weil sie den Unterschied zwischen Einzelereignissen und Ereignistypen *explizit* macht. Hierbei stehen A, B,... für offene

Formeln, wenn die Wahrscheinlichkeit im statistischen Sinn aufgefasst wird, und für Sätze, wenn sie im subjektiven Sinn aufgefasst wird.

Die Variablen A, B,... können im Folgenden sowohl als Teilmengen von Ω (mathematisch), als offene Formeln (statistisch) oder als Sätze (epistemisch) gelesen werden; die Wahrscheinlichkeitsgesetze sind immer dieselben. Dass A und B *disjunkt* sind, bedeutet mengenalgebraisch, dass $A \cap B$ leer ist; in der statistischen Lesart, dass die Extension von $A \wedge B$ faktisch (also im dem als ‚faktisch gegeben' betrachteten Modell) leer ist; und in der subjektiven Lesart, dass $A \wedge B$ in allen als epistemisch möglich erachteten Modellen der Sprache unerfüllbar ist. Die Menge aller möglichen Modelle (oder ‚Welten'), kurz „Mod", die der subjektiven Wahrscheinlichkeitsfunktion zugrunde liegt, muss nicht unbedingt mit der Menge aller logisch möglichen Modelle zusammenfallen, sondern kann eine Untermenge davon sein. Wir schreiben im Folgenden $\Box A$ (für „A ist notwendig"), wenn A von allen Modellen in Mod erfüllt wird, und analog „$\Diamond A$" für „A ist möglich", wenn A von mindestens einem Modell in Mod erfüllt wird. „$\Box$" und „$\Diamond$" sind die zwei grundlegenden modallogischen Satzoperatoren; es gilt $\Diamond A$ g. d. w. $\neg\Box\neg A$, oder *in Worten:* A ist möglich g. d. w. die Negation von A nicht möglich ist. Im statistischen Fall behauptet „$\Box A$", dass A exhaustiv ist und „$\Diamond A$", dass A's Extension nicht leer ist.

Die Allquantifikation „für alle A" bezeichnet in der mengenalgebraischen Lesart alle in den Mengenalgebra enthaltenen Teilmengen von Ω (Kap. 3.3), in der statistischen Lesart alle offenen Formeln und in der subjektiv-epistemischen alle Sätze der zugrundeliegenden Sprache.

(Def. 3-1) *Grundaxiome der Wahrscheinlichkeit*
Für alle A, B,... , wobei statt „p" auch „P" stehen kann:
(A1) $p(A) \geq 0$ (Nicht-Negativität)
In Worten: Wahrscheinlichkeiten sind immer größer-gleich null.
(A2) $p(A \vee \neg A) = 1$ (Normierung auf 1)
In Worten: die Wahrscheinlichkeit des gesamten Möglichkeitsraumes ist 1.
(A3) Wenn A, B *disjunkt* sind: $p(A \vee B) = p(A) + p(B)$ (endliche Additivität)
In Worten: für disjunkte Ereignis(typen) addieren sich die Wahrscheinlichkeiten.

Eine Funktion, die Axiome (A1-3) erfüllt, heißt eine (Kolmogorovsche) Wahrscheinlichkeitsfunktion.

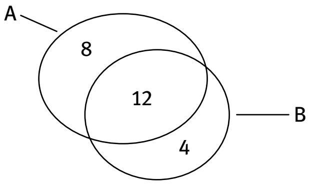
Aus den Grundaxiomen der Wahrscheinlichkeit ergeben sich eine Reihe von Theoremen, von denen in Satz 3.1 die wichtigsten genannt sind. Eine Formel A in n freien Variablen heißt *exhaustiv* im statistischen Fall g. d. w. A im gegeben Modell von *allen* möglichen n-Tupel von Individuen in D erfüllt wird. Im subjektiven Fall heißt ein Satz A exhaustiv g. d. w. alle epistemisch möglichen Modelle der Sprache A wahr machen. In der mengenalgebraischen Lesart schließlich ist A exhaustiv, wenn A mit dem gesamten Möglichkeitsraum Ω zusammenfällt. Eine Folge von n paarweise disjunkten A_i ($1 \leq i \leq n$) heißt eine *Partition* oder Zerlegung von Ω g. d. w. die Disjunktion $A_1 \vee \ldots \vee A_n$ *exhaustiv* ist. Sämtliche im Folgenden angeführten Theoreme gelten – sofern nichts anderes gesagt wird – für *alle* A, B,..., und statt „p" kann wieder „P" stehen: wir ersparen uns im Folgenden diese Zusätze.

(Satz 3-1) *Theoreme unbedingter Wahrscheinlichkeit* (Beweis Anhang 10.3.1)
(T1) $p(\neg A) = 1 - p(A)$ (Komplementärwahrscheinlichkeit)
In Worten: Die Wahrscheinlichkeit der Negation eines Ereignisses ist 1 minus jener des Ereignisses.
(T2) $p(A) \leq 1$ (obere Schranke)
In Worten: Die Wahrscheinlichkeit jedes Ereignisses ist kleiner-gleich 1.
(T3) $p(A \wedge \neg A) = 0$ (Kontradiktion)
In Worten: Ein Widerspruch besitzt die Wahrscheinlichkeit Null.
(T4) Für jede Partition $A_1, \ldots, A_n$: $\sum_{1 \leq i \leq n} p(A_i) = 1$ und $p(B) = \sum_{1 \leq i \leq n} p(B \wedge A_i)$
In Worten: Die Summe der Wahrscheinlichkeiten der Ereignisse einer Partition (A_i: $1 \leq i \leq n$) von Ω addiert sich zu 1, und die Ereignisse ($A_i \wedge B$: $1 \leq i \leq n$) bilden eine Partition von B, deren Wahrscheinlichkeiten sich zu p(B) aufaddieren.
(T5) $p(A_1 \vee A_2) = p(A_1) + p(A_2) - p(A_1 \wedge A_2)$ (allgem. Additionsgesetz)
(T6) Wenn $A_1 \rightarrow A_2 =_{def} \neg A_1 \vee A_2$ exhaustiv ist, dann gilt $p(A_1) \leq p(A_2)$ (Monotonie)
In Worten: Wenn A_1 mit Notwendigkeit A_2 impliziert, dann ist die Wahrscheinlichkeit von A_1 kleiner-gleich der von A_2.
(T7) Ist $A_1 \leftrightarrow A_2$ exhaustiv, dann gilt $p(A_1) = p(A_2)$ (Äquivalenz)

Die Wahrscheinlichkeit von A unter der *Annahme*, dass B vorliegt, nennt man die *bedingte Wahrscheinlichkeit* von A gegeben B. Man schreibt dafür p(A|B) bzw. P(A|B) und definiert diesen Ausdruck gewöhnlich wie folgt:

(Def. 3-2) *Bedingte Wahrscheinlichkeit:* $p(A|B) =_{def} \frac{p(A \wedge B)}{p(B)}$, *sofern* $p(B) > 0$.
(Analog für „P" anstelle von „p".)

In p(A|B) heißt B das *bedingende* Ereignis oder Antecedens und B das *bedingte* Ereignis oder Konsequens. Im endlich-statistischen Fall koinzidiert p(A|B) mit der relativen Häufigkeit von A in der Menge B, die ihrerseits eine Teilmenge des Individuenbereichs D ist – siehe Abb. 3-1. Im unendlich-statistischen Fall koinzidiert p(A|B) mit dem Häufigkeitsgrenzwert von A's in einer Zufallsfolge von B-Individuen.



$p(Bx|Ax) = 12/20 = 3/5$
$p(Ax|Bx) = 12/16 = 3/4$
$|D| = 24$, $p(Ax) = 20/24 = 5/6$
$p(Bx) = 16/24 = 2/3$

Abb. 3-1: *Bedingte statistische Wahrscheinlichkeiten*

Im subjektiv-epistemischen Fall steht $P(A|B)$ für den *hypothetischen* Glaubensgrad an A, unter der Annahme, dass B sicher ist, also mit $P(B) = 1$ geglaubt wird. Wird B tatsächlich mit Sicherheit geglaubt, gilt also $P(B) = 1$, dann folgt $P(A) = P(A|B)$. Zur Vermeidung von Missverständnissen sei darauf hingewiesen, dass $P(A) = 1$, also subjektive Sicherheit bzgl. A, weder logisch noch analytisch impliziert, dass A wahr ist: subjektive Sicherheit ist *fallibel* und die Glaubensfunktion P ist somit unabhängig von der Wahrheitswertfunktion I (s. Anhang 10.1.3 zur logischen Semantik). Man kann diesbezügliche Zusatzannahmen natürlich machen: z. B. nimmt man gelegentlich an, dass das betreffende Subjekt zumindest bzgl. empirischer Evidenzen (Erfahrungen) E infallibel ist, dass hier also $p(E) = 1$ auch $I(E) = 1$ impliziert. Aber solche Annahmen gehen über die Wahrscheinlichkeitstheorie hinaus.

Die gewöhnliche Definition von $p(A|B)$ hat den Nachteil, dass $p(A|B)$ für ein 0-wahrscheinliches Ereignis B nicht definiert ist. Nicht nur unmögliche Ereignisse haben Wahrscheinlichkeit null. Auch kontingente Ereignisse können Wahrscheinlichkeit null besitzen, wie z. B. das erwähnte Beispiel der Wahrscheinlichkeit, aus unendlich vielen natürlichen Zahlen zufällig eine Zahl zu ziehen, die die 2er Potenz einer natürlichen Zahl ist. Daher wurden Methoden entwickelt, die bedingte Wahrscheinlichkeit, statt sie durch die unbedingte zu definieren, direkt zu axiomatisieren, um sie so auf kontingente nullwahrscheinliche Antecedensereignisse ausdehnen zu können. Carnap (1971, 38f) schlug folgendes Axiomensystem vor:

(Def. 3-3) *Direkte Axiomatisierung bedingter Wahrscheinlichkeit:*
Annahme: Die Antecedensereignisse (die in den Axiomen auftreten) sind jeweils nicht leer bzw. möglich:

(B1) $p(A|B) \geq 0$	(Nicht-Negativität)
(B2) $p(A \vee B|B) = 1$	(Folgerung)
(B3) Für disjunkte A und B: $p(A \vee B|C) = p(A|C)+p(B|C)$	(endliche Additivität)
(B4) $p(A \wedge B|C) = p(B|C) \cdot p(A|B \wedge C)$	(allgemeines Multiplikationsprinzip)