

Vorlesungsskript - Bayes Statistik

Dominikus Vogl¹

9. Juli 2009

¹University of Cologne

Inhaltsverzeichnis

1	Einleitung	3
2	Wissenschaftstheorie I	4
2.1	Kritischer Erkenntnisfortschritt	6
2.1.1	Karl Popper – der "strenge" Prüfer	6
2.1.2	Logik der Forschung – Falsifikationismus	7
2.1.3	Alles ist möglich – eigentlich! oder der brave Empirist	8
3	Wissenschaftstheorie II	11
3.1	Wissenschaftstheorie von Balzer	11
3.1.1	Wissenschaft als Prozess	11
3.1.2	Bestätigung	11
3.1.3	Strukturen	14
3.1.4	Daten	15
4	Wahrscheinlichkeitstheorie I	18
4.1	Statistische Theorien	18
4.1.1	Wahrscheinlichkeitsfunktion	19
4.2	Wahrscheinlichkeiten nach Rohwer und Pötter	20
4.2.1	Unterschiedliche Generalisierungsprobleme	20
4.2.2	Epistemische und aleatorische Wahrscheinlichkeiten	21
4.2.3	Stochastische Inferenz	21
4.2.4	Deskriptive Generalisierungsprobleme	21
4.2.5	Modale Generalisierungsprobleme	22
4.2.6	Stochastische Ablaufschemas	22
4.2.7	stochastische Modelle	22
4.2.8	Fiktive Populationen	23
4.2.9	Wahrscheinlichkeit und Chance	23
5	Wahrscheinlichkeitstheorie II	25
5.1	Grundmodell der statistischen Methodik	25
5.2	Zufallsvariable	25
5.2.1	Zufallsauswahl	25
5.2.2	Formale Definition einer Zufallsvariable	26
5.3	Bedingte Wahrscheinlichkeiten	28
5.3.1	Bedingte Wahrscheinlichkeiten	28
6	Graphische Darstellung bedingter Wahrscheinlichkeiten	31

7	Satz von Bayes	33
7.1	Quantifizierung von Unsicherheit	33
7.1.1	Zusammenfassung des Satzes von Bayes	33
7.2	Wahrscheinlichkeitsaussagen	34
7.3	Was heißt Bayes Statistik?	36
7.3.1	Wiederholung Inferenzstatistik	36
7.3.2	Bayesianische Ansätze	38
7.3.3	Noch ein genauerer Blick auf bayesianische Verfahren	38
8	Bayes Verfahren	41
8.1	Einleitung	41
8.2	Die aposteriori Verteilung	42
8.2.1	Definition der aposteriori Verteilung	42
8.3	Konjugierte Verteilungen, Bayes Lernen	42
8.3.1	Das Beta-Binominal Modell	43
8.3.2	Gamma-Poisson-Modell	43
8.3.3	Selbstkonjugiertheit der Normalverteilung	44
9	Ergänzung: Bayesianische Statistik	46
10	Monte-Carlo-Methode	48
10.1	Einleitung	48
10.1.1	Geschichte der Monte-Carlo-Methode	48
10.2	Funktion der Monte-Carlo-Methode	49
10.2.1	Pseudopopulation	49
10.2.2	Zufallsgenerator einer Gleichverteilung	49
10.2.3	Die numerische Bestimmung einer Fläche	51
10.3	Markov-Ketten-Monte-Carlo-Methode	51
11	Erata	53

Kapitel 1

Einleitung

Beginnen möchte ich mit einem Satz von Joan Míro, der sagte: "...der weite frei Raum, in dem Bewegung ohne Ende abläuft. . .".

Die Gesellschaft ist solch ein Raum, in dem Bewegung ohne Ende abläuft. Es gibt Unterschiede, die sich zu messen und zu erklären die Wissenschaft anspricht. Hierfür werden Methoden entwickelt, genutzt und geprüft. In der empirischen Sozialwissenschaft beruhen die Methoden auf wahrscheinlichkeitstheoretischen Überlegungen. Der Rolle der Wahrscheinlichkeitstheorie soll in dem folgenden erläutert werden. Ziel ist es den Anwender mit einem enormen wissenschaftsgeschichtlichen Erkenntnisfortschritt vertraut zu machen, um so das Urteil und die Aussagekraft wissenschaftlicher Ergebnisse zu erhöhen. . . . um die Weite des Raumes und die endlose Bewegung tatsächlich zu erleben. . .

Kapitel 2

Wissenschaftstheorie I

In den nächsten 30 Minuten möchte ich kein vollständiges Bild der Wissenschaftstheorie geben. Mir liegt vielmehr daran Wissenschaft als einen im Prinzip offenen Prozess zu beschreiben. Bewusst polarisiere ich ein wenig, um als Dozent nicht vorgefertigte Meinungen zu vermitteln, sondern unterschiedliche Denkweisen zu zeigen. Wissenschaft ist ein Abenteuer und ich hoffe, dass ich diesen Gedanken transportieren kann.

Eine einzelwissenschaftliche, etwa eine physikalische Untersuchung kann ohne weitere Umschweife mit der Bearbeitung ihres Problems beginnen. Sie kann sozusagen, mit der Tür ins Haus fallen; es ist ja ein "Haus" da: ein wissenschaftliches Lehrgebäude, eine allgemein anerkannte Problemsituation. Der Forscher kann es deshalb auch dem Leser überlassen, die Arbeit in den Zusammenhang der Wissenschaft einzuordnen.

In einer anderen Lage befindet sich der Philosoph. Er steht nicht vor einem Lehrgebäude, sondern vor einem Trümmerfeld (in dem es freilich auch Schätze zu entdecken gibt). An eine allgemein anerkannte Problemsituation kann er nicht anknüpfen, denn daß es eine solche nicht gibt, das allein dürfte vielleicht allgemein anerkannt sein; taucht doch sogar in den philosophischen Auseinandersetzungen immer wieder die Frage auf, ob die Philosophie es überhaupt mit echten "Problemen" zu tun habe.

Wer diese Frage bejaht, wer deshalb den Versuch auch nicht für zwecklos hält, den traurigen Zustand, den man philosophische Diskussion nennt, zu überwinden, der kann, wenn er sich zu keiner der streitenden Schulen bekennt, wohl nur den einen Weg gehen: am Anfang anzufangen. (Karl Popper, Wien 1934)[18, XIII, Hervorhebung im Original]

In welchem Spannungsfeld bewegt sich der Forscher? Dieser Frage ging der Philosoph Karl R. Popper (1902 - 1994) Zeit seines Lebens nach. Ersetzt man in dem eben erwähnten Zitat die Disziplin "Philosophie" durch Soziologie, so befinde zumindest ich mich in der mir vertrauten und doch noch unbekanntem Welt der Soziologie und wissenschaftlichen Forschung.

Wie komme ich auf diesen Gedanken?

Auf den ersten Blick hin sind Poppers Worte längst Geschichte. Für mich hat das Zitat allerdings nicht an Aktualität verloren. Der Kern der Aussage liegt weniger in der Zustandsbeschreibung einer Wissenschaft, auch nicht im Streit vorhandener Schulen über ihre wissenschaftlichen Problemsituationen und die Abgrenzung zu anderen Disziplinen. Es geht darum, einen Weg *neu* anzufangen. Popper führt diese Überlegung aber weiter, sein Verständnis des kritischen Rationalismus zielt darauf ab, durch strenge Prüfung, immer wieder am Anfang anzufangen.

Ich möchte noch einen Versuch wagen, wissenschaftliches Handeln zu beschreiben. Der Sozialwissenschaftler Vilfredo Pareto (1848 - 1923) zieht im Paragraph 2 seines *Trattato di sociologia generale* den Schluss:

Beschäftigen wir uns mit der Suche nach den Beziehungen zwischen den sozialen Tatsachen, und man möge diesem Studium den Namen geben, den man will. Wenig wiegt die Methode, durch die man das Wissen um diese Beziehungen erwirbt. Das Ziel alleine ist uns wichtig; wenig und selbst gar nicht die Mittel, die es zu erreichen gestatten. [16, 1]

Das sind radikale Gedanken, die ich etwas präzisieren möchte.

Paretos Gedanken zielen auf das ab, was Ralf Dahrendorf eine Wissenschaft ohne "Namensschildchen" und "Zäune" nennt und meint damit eine Wissenschaft, in der nicht Institutionen die Methodik vorgeben und für sich beanspruchen, sondern vielmehr die "Suche" nach "Beziehungen" die Methodik bestimmt [8, 45]. Paretos Gedanken sind eine Einschränkung und sind keineswegs eine Rechtfertigung für Beliebigkeit, keine Begründung für lockere Handhabung in der wissenschaftlichen Vorgehensweise.

Paretos wie auch Poppers einleitende Sätze weisen darauf hin, dass ein wissenschaftliches Gebäude vorhanden ist. Der Weg in diesem Gebäude aber ist mehr als eine reine Anpassung an Konventionen, geltende Normen und Sanktionen – ein Weg, den Dahrendorf treffend als "Abenteuer" für den Menschen beschreibt – ein Weg, der "die Unterwerfung unter eigene Konventionen und Regeln, damit den Erwerb eines eigenen Gewissens" bedeutet [8, 45].

Die einführenden Zitate weisen darauf hin, dass in der Wissenschaft eine kritische Betrachtung notwendig ist, um so dem Ziel näher zu kommen, mehr über die Beziehungen der Tatsachen der Gesellschaft in Erfahrung zu bringen. Wissenschaft ist auch eine Prüfung für sich selbst, denn es ist eine Reise ins Unbekannte – ein "Abenteuer".

Unser Augenmerk richtet sich auf zwei Gegenstände. Einerseits die Tatsachen, wie dem Satz: 'Heute, hier befinden sich XX Personen zu dieser Uhrzeit an diesem Ort' – oder 'heute scheint die Sonne, es liegt kein Schnee'; andererseits gilt es die Gründe für diese Tatsachen zu ergründen – wie das Wort schon sagt.

Ferner wollen wir uns Fragen, wie sich Aussagen verallgemeinern lassen – welche Kriterien sind hierfür notwendig und wie soll man zu einer allgemeingültigen Aussage kommen?

Dahrendorf beschreibt unser Problem passend: Der Mensch ist unvollkommen und seine Lage durch Ungewissheit geprägt.[9, 224]

Ein Ziel dieses Kurses ist es die vorhandenen Unsicherheiten zu erkennen und dadurch ein Gespür zu bekommen was wir tatsächlich wissen und was wir als Wissen-Schaffende tatsächlich behaupten dürfen.

Im Hinterkopf sollten wir auch die Frage behalten, was eine Disziplin wie die Soziologie an Erkenntnis liefern kann. Allerdings sollte jede und jeder Einzelne diese Frage für sich selbst beantworten, noch nicht jetzt, sondern bei seiner Abschlussarbeit. Wir hier an der Universität können nur gebetsmühlenartig auf das Problem der Unsicherheit hinweisen und Verfahren vermitteln, wie man die Unsicherheit einschätzen kann.

Eine Möglichkeit der Unsicherheitsreduktion ist die Verwendung von wissenschaftlichen Theorien, eine andere Möglichkeit ist die Arbeit mit empirischen Daten.

2.1 Kritischer Erkenntnisfortschritt

Dieses Kapitel befasst sich mit Theorien – genauer mit der Entstehung, Prüfung und Bestätigung einer Theorie. Zwei Überzeugungen sollen hierfür vorgestellt werden. Nach Karl Popper bilden sich Theorien aus einem logischen Denkprozess und zeichnen sich durch Prüfbarkeit aus. Aus der Sicht Paul Feyerabends beginnt eine Theorie bereits mit nichtwiderlegbaren Vermutungen, die beständig geprüft werden. Auf den Gegensatz und das Gemeinsame der Denkansätze soll nun näher eingegangen werden.

Vorweg sei gesagt, dass dies kein vollständiger Überblick über die Wissenschaftstheorie ist, sondern bestimmte Schlagworte, auf die wir im Verlauf der Veranstaltung noch häufiger stoßen werden, in Erinnerung gerufen werden sollen.

2.1.1 Karl Popper – der "strenge" Prüfer

Karl Popper oder der "strenge" Prüfer, man könnte auch sagen: der "ewige" Sucher. Wie man auch immer Karl Popper nennen möchte, das Ziel ist mehr über Poppers Einstellung zur Wissenschaft zu erfahren.

Poppers Überlegungen zur Theorieentwicklung liegen einem Evolutionsverständnis ohne eindeutige Tendenz zugrunde. Die Natur besteht aus Prozessen, die zwischen Determinismus und Zufall (ersten beiden Schlagwörter) ablaufen und auf unterschiedlichen Ebenen der Wirklichkeit stattfinden. Insgesamt unterscheidet Popper drei Ebenen, die er 'Welt 1', 'Welt 2' und 'Welt 3' nennt. Die Welt 1 ist die Welt der physikalischen Gegenstände – die Welt der natürlichen Elemente und der lebendigen Organismen. Daneben gibt es die Welt 2 der psychischen Zustände und Prozesse – die Welt der subjektiven Bewusstseinserebnisse. Darüber hinaus existiert die Welt 3 – die Welt der Erzeugnisse des menschlichen Geistes, deren Inhalte unserem Denken entstammen. Darunter fallen Mythen, Theorien, Werkzeuge, Kunstwerke und wissenschaftliche Entdeckungen. Sie sind von uns selbst geschaffen, obwohl sie nicht immer Ergebnisse planvollen Handelns gewesen sind [19, 63].

Zu Beginn einer wissenschaftlichen Erneuerung stehen Theorien der Welt 3. Dabei ist wissenschaftliches Verstehen der Versuch, eine Problemstellung abstrakt zu erfassen, wobei durch diesen intellektuellen Prozess die Probleme der Theorien erst offensichtlich werden. Das Ergebnis der Überlegungen ist eine neue Theorie. Ob mit der Erneuerung schon eine Veränderung der Sicht auf die Welt geschieht, bleibt vorerst offen. Die Theorie wird in Schriftform gebracht,

vervielfältigt und kommt dann möglicherweise zu einer praktischen Anwendung in Welt 1.

Welt 3-Gegenstände sind abstrakt, doch sie sind gleichzeitig wirklich, da sie die Welt verändern können. Allerdings nur durch das Eingreifen des Menschen und durch die Erfassung des Problems in Welt 2 als psychischen Prozess. Als Werkzeug für die Menschen dient dabei die Sprache, eine Wahrnehmungsform, die nicht gegeben ist, sondern durch einen Lernprozess und aktiver Tätigkeit erworben wird. Durch die Art und Weise Dinge wahrzunehmen und sprachlich darzustellen entwickelt sich – laut Popper – eine eigene Persönlichkeit [19, 73].

Popper beschreibt hier die Folgen des wissenschaftlichen Tuns für den Menschen unter sehr allgemeinen Gesichtspunkten. Die Forderung nach wissenschaftlicher Erkenntnis bleibt nichtsdestotrotz unvermindert bestehen, dabei, so Popper, sollte ein Problem nicht klein geredet und durch den Wunsch nach Einfachheit 'weg'-erklärt werden. Nur durch eine vollständige Erfassung des Problems wird verhindert, ein wissenschaftliches Urteil vorwegzunehmen oder dem Urteil voreilig die Zustimmung zu verweigern [19, 89].

Auf einen Punkt möchte ich noch hinaus: Durch Poppers abstrakte Aufteilung der Welt werden die Unterschiede von wissenschaftlichen Fragen im Vergleich zu alltagsgegenständlichen Fragen ersichtlich. Kurz gesagt, behandelt die Wissenschaft Probleme nicht Fragen. Für Dahrendorf gibt es eine klare Trennung zwischen Problemen und Fragen. Probleme stellen wir uns selbst, Fragen bekommen wir gestellt. Fragen betreffen das Leben und kein Mensch kann sich ihnen entziehen, sie sind aber mit den verfügbaren Mitteln prinzipiell lösbar. Probleme hingegen betreffen die Wissenschaft. Sie sind esoterisch, speziell und nur wenigen bekannt, was die Welt-der-Probleme ein wenig absurd erscheinen lässt [8, 17]. Gerade das Absurde in der Wissenschaft bietet die Möglichkeit, Fragen als Problem zu formulieren und so mehr über das Unbekannte der Welt zu lernen – die "Rätsel des menschlichen Wissens", so Popper [18, XXIV].

Also sollten wir, wenn wir Wissenschaft betreiben im Augenmerk haben, dass sich unser Tun im Rahmen seiner selbst gewählten Fragestellung - einer "gedachten Notwendigkeit" abspielt [8, 22]. Soziologie als Erfahrungswissenschaft¹ ist die Gelegenheit, innerhalb einer selbst gestellten Fragestellung, über Erfahrungen der Gesellschaft zu staunen und Tatsachen zu entdecken, die nach einem naiven "Warum?" verlangen [8, 256].

Mein Fazit der des bisherigen Stoffes lautet: Somit sind Probleme der selbst gestellte Antrieb, mehr über ausgewählte Bereiche der Welt zu lernen, im Gedankenaustausch mit anderen Menschen und um den Preis der Abstraktion.

2.1.2 Logik der Forschung – Falsifikationismus

Im Interesse der Wahrheitssuche und der Befreiung vom Irrtum müssen wir uns dazu erziehen, unsere eigenen Ideen ebenso kritisch betrachten zu können wie die Ideen, gegen die wir kämpfen. [17, 116]

¹Der Wortstamm Empirie stammt vom griechisch *empireia*: Erfahrung, Erfahrungswissen. Informationen werden durch gezielte Beobachtungen im Feld oder Labor gewonnen. Objekte und Sachverhalte werden untersucht. Als nicht-empirisch gelten die Mathematik und die Philosophie – insbesondere die Logik, Rechtswissenschaften, Theologie (Dogmatik) – da nicht-widerlegbare Aussagen (Tautologien) getroffen werden und Aussagen nach rein logisch-formalen Kriterien getroffen werden.

Popper beschreibt in seinem Buch 'Logik der Forschung' (Erstausgabe 1936) die Prinzipien des kritischen Rationalismus. Kurz gesagt charakterisiert Popper wissenschaftliches Forschen dadurch, dass Theorien aufgestellt und "systematisch" überprüft werden. Nur so ist eine Überprüfung der gebildeten Kategorien und Hypothesen möglich.

Wissenschaft läuft nach klaren Schritten ab [18, 6-8]:

- der logische Vergleich der Folgerungen - insbesondere die Suche nach inneren Widersprüchen,
- die Untersuchung der logischen Form einer Theorie hinsichtlich der empirischen Prüfbarkeit ihrer Aussagen,
- der Vergleich mit anderen Theorien und die Bewertung des wissenschaftlichen Fortschritts,
- die Prüfung durch die praktische empirische Anwendung.

Die empirische Forschung bewirkt für sich alleine betrachtet noch keine erfahrungswissenschaftliche Aussage. Sie gelingt erst durch das Aufstellen von empirisch prüfbareren Sätzen und zweitens durch die Möglichkeit der Kontrolle mittels empirischer Ergebnisse. So gesehen ist Falsifikation ein Kriterium, welches ermöglicht, Sätze, Hypothesen und Theorien an der Empirie scheitern zu lassen. Die Falsifikation ist somit ein "Abgrenzungskriterium", weil sich durch sie die Wissenschaft als widerlegbare Form der Beobachtung von anderen unwiderlegbaren Formen der Beobachtungen wie der Kunst oder der Metaphysik unterscheidet [18, 9]. Trotz dieser Vorgaben ist Wissenschaft nicht vorgegeben oder determiniert. Popper begreift Wissenschaft als ein Spiel mit Regeln: "[Und so] können wir die Untersuchung der Regeln des Wissenschaftsspiels, der Forschungsarbeit, auch *Logik der Forschung* nennen." [18, 26, Hervorhebung im Original]

Nach Popper ist menschliches Wissen niemals wahr, sondern immer nur bewährt. Positivistische Verifikation – den Schluss von Einzelsätzen auf allgemeingültige Aussagen – lehnt Popper ab und schlägt stattdessen das Prinzip der deduktiven Logik vor. Denn diese schließt nicht nur von den Prämissen auf die logisch folgende Konklusion, sondern modifiziert sie auch, falls sich die Konklusion als falsch erweisen sollte. Hierfür werden intersubjektiv nachprüfbarere Sätze gebildet, die im Widerspruch zur prüfenden Theorie stehen. Ausgehend von einer theoretischen Überlegung entsteht Schritt für Schritt eine Annäherung an die Wahrheit, die im Grunde genommen ein Lernen aus den überprüften Fehlern darstellt.

Poppers Falsifikationismus wurde vielfach kritisiert, unter anderem von Imre Lakatos und Paul Feyerabend.

2.1.3 Alles ist möglich – eigentlich! oder der brave Empirist

Laut Imre Lakatos (1922 – 1974) handelt es sich bei der Entwicklung von wissenschaftlichen Theorien um einen historischen Prozess, bei dem es nicht nur den logischen Inhalt einer Theorie zu betrachten gilt, sondern auch die Entscheidungen der Menschen für eine Theorie berücksichtigt werden muss [12, 312f.].

Nach Lakatos ist eine Theorie zum Zeitpunkt ihrer Entstehung noch unzureichend begründet. Sie enthält Widersprüche und die zu untersuchende Tatsache sowie deren genaue Beurteilung sind noch unklar. Im Laufe der geschichtlichen Entwicklung verringert die Theorie ihre Mängel und erweist sich als brauchbar. Somit ist es nicht treffend, von einer einzigen Theorie zu sprechen, vielmehr gilt es, sich mit einer Folge von Theorien, einem so genannten "Forschungsprogramm" auseinander zusetzen.

Ein Forschungsprogramm durchläuft unterschiedliche Prozesse, bestehend aus Wachstum, Stagnation und Degeneration, wobei die Entwicklung unbestimmt ist. Sagt eine Theorie Tatsachen erfolgreich voraus, wächst sie solange, bis die Vorhersagekraft abnimmt und befriedigende Erklärungen seltener werden. Der degenerative Prozess setzt ein, wenn die Theorie ad hoc oder durch zufällige Ergänzungen verändert wird. Die Veränderung ist notwendig, da zwischenzeitlich Theorien und Programme existieren, die diese Änderungen bereits haben. Lakatos, verdeutlicht, dass die Entscheidung für ein Forschungsprogramm nicht durch Regeln, sondern höchstens durch Richtlinien bestimmt wird und so, im Unterschied zu Poppers Falsifikationsforderung, degenerierte Programme nicht zwangsläufig verschwinden müssen. Man kann an Forschungsprogramme glauben, selbst wenn diese degeneriert und überholt sind. Statt Regeln handelt es sich hierbei um freie Entschlüsse in konkreten historischen Situationen, was bedeutet, dass eine in Vergessenheit geratene Theorie im Kern vorhanden ist und es von den Menschen und ihren Entscheidungen abhängt, ob die Erklärungen wieder an Bedeutung gewinnen.

Als Beurteilungskriterien für eine der vorhandenen Theorien nennt Lakatos drei Punkte [12, 328]:

- wie Popper fordert auch Lakatos die Orientierung an Basissätzen, die einen realen Sachverhalt beinhalten,
- es gilt ein Rationalitätskriterium,
- aus dem Rationalitätskriterium folgt, dass Regeln universell anwendbar sein müssen.

Feyerabend kritisiert die letzten beiden Kriterien. Sein Wunsch ist die Mannigfaltigkeit der Meinungen und eine Methode, die Vielfalt fördert. Jede Einschränkung ist dabei hinderlich. Die "Irreduzibilität", die tolerante Haltung gegenüber der Verwendung gleicher Begriffe in unterschiedlichen Sinnzusammenhängen, ist zu unterstützen. Die Forderung der "Sinninvarianz", wonach der Sinn eines Begriffs im Zuge des Fortschrittes der Erkenntnis erhalten bleibt und eine Verwendung in einem anderen Zusammenhang nicht möglich ist, ist hingegen abzulehnen. Nach seinem Verständnis führen alternative Bedeutungen letzten Endes entweder zu einer Bestätigung einer vorhandenen Theorie oder aber zu einer absolut neuen Erkenntnis, einer, wie Feyerabend betont, revolutionären Entdeckung [11, 321-325].

Feyerabends Methodik besteht nicht darin bestehende Erklärungen ständig zu verbessern oder durch mindestens gleich gute Theorien zu ersetzen, sondern er regt vielmehr zu eigenständigen Kriterien und Erklärungen an, die im Extremfall die bisherige Theorie von einem Moment auf den Anderen widerlegen.

(...) ein braver Empirist ist ein Denker, der nicht bei Theorien stehen bleibt, die allgemein akzeptiert sind, selbst wenn sie fehlerfrei

und in hohem Grad konfirmiert sind. Wohl wissend, dass die radikalste und wirkungsvollste Kritik einer Theorie jene ist, die mit Hilfe von Alternativen geführt wird, wird er versuchen, solche Alternativen zu finden. [11, 325]

Ein wichtiger Bestandteil dieses Prozesses sind Vermutungen und freies Erfindungsstreben. Damit Vermutungen zu einem späteren Zeitpunkt bewährte wissenschaftliche Theorien ersetzen und ergänzen können, ist die Methodik nicht beliebig, denn die Vermutungen werden kritisch geprüft: "Ein braver Empirist muss also ein kritischer Metaphysiker sein." [11, 325] Für Feyerabend sind die zu untersuchenden Sätze zu Beginn einer wissenschaftlichen Betrachtung vernachlässigbar und noch reine Spekulation, da die empirische Prüfung erst mit der Entwicklung einer Theorie bedeutender wird. Vor allem kritisiert er Erklärungen, die dadurch begründet werden, dass sie aus Mangel an besseren Alternativen gewählt worden sind [11, 323-326].

In diesem Kapitel sollten gegensätzlichen Herangehensweisen gezeigt werden, die sich im Grundverständnis der Wissenschaft sehr ähneln. Für Popper und Feyerabend steht die wissenschaftliche Erkenntnis im Vordergrund. Beide vertreten im Kern eine ähnlich kritische Haltung, verfolgen ein gemeinsames Ziel und platzieren den kritikfähigen Mensch als Wissen-Schaffender in den Vordergrund. Die jeweiligen Wege dort hin erweisen sich als unterschiedlich, wenn nicht gar gegensätzlich. Popper fordert, die Konsequenzen wissenschaftlicher Sätze logisch und empirisch zu testen und bei Bedarf die Prämissen dieser Sätze zu ändern. Feyerabend hingegen spricht sich für eine nicht vorgegebene, anfangs sogar beliebige Methodik aus und erst die empirische Prüfung zeigt, ob die Vermutung nicht nur Metaphysik ist, sondern auch zu wissenschaftlicher Erklärungskraft führt. Popper fordert zur Prüfung einer Theorie die Wahl geeigneter Basissätze, was für Feyerabend eine ungenügende Forderung ist, da Basissätze so gewählt werden, dass sie nicht den Tatsachen entsprechen, sondern im Einklang mit der Theorie stehen.

Unabhängig davon, welcher Weg gewählt wird, ob Poppers Überprüfung vorhandener Theorien oder Feyerabends Nachspüren neuer Felder, die Ergebnisse ähneln sich. Denn der kritische Rationalismus sowie die kritische Metaphysik haben ihre Gemeinsamkeit in der empirischen Überprüfbarkeit. Wenn dieses kritische Element vorhanden ist, ist die Herangehensweise zweitrangig, das Ergebnis auf lange Sicht aber gleich. Fehlt jedoch die Überprüfbarkeit, so handelt es um zwei unversöhnliche Gegensätze: die starre dogmatische Überzeugung auf der einen und die zufällige vage Beliebigkeit auf der anderen Seite. Daher kann die These insofern konkretisiert werden, dass wissenschaftliche Forschung sowohl der Versuch ist bestehende Theorien zu überprüfen, als auch nach neuen Alternativen zu suchen.

Kapitel 3

Wissenschaftstheorie II

Wilfredo Pareto spricht von der Erforschung der sozialen Tatsachen als Ziel soziologischer Forschung. Diesen Gedanken möchte ich heute aufgreifen und ihn in einen interdisziplinären Kontext setzen. Gesellschaftswissenschaften als Teil der Wissenschaft also, mit seinen Eigenheiten und Besonderheiten.

3.1 Wissenschaftstheorie von Balzer

Balzer [1] versucht in seinem Buch, die Natur- und Sozialwissenschaften nicht isoliert zu betrachten, sondern speziell den Problemen in den Sozialwissenschaften gerecht zu werden.

3.1.1 Wissenschaft als Prozess

Balzers Ansatz ist ein strukturalistischer Ansatz. Balzer spricht grundsätzlich von Systemen (Zuständen), die sich in einem Möglichkeitsraum befinden.

”Ein Prozess ist ein ständiger Übergang vom Möglichen zum Wirklichen, eine Folge von Zuständen in einem Raum, dem sogenannten Zustandsraum, der auch alternative, bloß mögliche Zustände enthält.”

Wissenschaft, wie wir sie analysieren, befindet sich in einem ständigen Wandel durch neue Daten und neue Theorien – Balzer spricht auch von Spannungen durch ständige Neuerungen. Wegen der Spannungen kommt es in der Wissenschaft zu Umbrüchen, in denen bisherige Grundmodelle durch neue Grundmodelle ersetzt werden, die die Spannungen auflösen und Kohärenz erzeugen.

3.1.2 Bestätigung

Allgemein

Wissenschaftliche Theorien sollen gut bestätigt sein. Ereignisse sollen prinzipiell wiederholbar sein und die beschriebenen Sachverhalte sollen tatsächlich wiederholt eintreten. Was ist eine Bestätigung?

Was ist, wenn die Systeme sich nicht durch Daten ausschöpfen lassen? Dies ist nicht möglich, wenn die intendierten Systeme in die Zukunft reichen oder

wenn die Datengewinnung aus praktischer Sicht sehr aufwendig ist. Daher soll eine *schwächerer* Form der Bestätigung angenommen werden. Diese ist eine Bestätigung, wenn die Annahme gilt, dass ein System hypothetisch – wie bei der induktiven Bestätigung "im Limes" – durch Daten ausgeschöpft ist. Ist dies der Fall, so kann die 'schwache' Bestätigung als eine 'starke' Bestätigung betrachtet werden und daher von einem induktiven Bestätigungsbegriff ausgegangen werden.

Die Bestätigung beginnt in der Regel mit vielen Beobachtungen und Daten, die wenn nicht voll beobachtet werden können – Ausschöpfung nicht möglich ist – durch Deduktion (z.B. statistische Methoden) ergänzt wird. So erfolgt eine Einbettung in bestehende Theorien oder eine Erklärung alltäglicher Begebenheiten. Reine deduktive Theorien sind zu schwach, sie müssen auch immer ein induktives Element enthalten, um Menschen zu überzeugen.

induktive Methoden

Unter induktiver Methode versteht man die "Ableitung von Hypothesen oder Gesetzmäßigkeiten aus vorgegebenen Daten". Voraussetzung ist, dass hinreichend viele Daten gewonnen wurden und das System somit ausgeschöpft wurde. Induktion bedeutet nun, dass man aus den gewonnenen Daten auf die richtigen Hypothesen schließt. Da bei der induktiven Methode keine Theorie vorhanden ist, können keine Basissätze formuliert werden (Basissätze sind Sätze, "die im Vokabular der Theorie formuliert sind.") und damit auch keine Daten im Sinne einer Theorie erhoben werden.

Aus der gegebenen Datenstruktur werden Axiome¹ abgeleitet, die zu einer Modellklasse passen, so dass die Datenstruktur zu einem Modell dieser Klasse passt.

Doch sollte man zwei grundlegende Fälle unterscheiden: Zum einen gibt es statische Systeme, die keine zeitliche Entwicklung durchmachen und deren axiomatische Erfassung keinen Zeitbezug hat. Zum anderen gibt es dynamische Systeme, die zum Zeitpunkt der Untersuchung noch existieren und in die Zukunft reichen, aber nicht müssen (abgestorbene Sonnensysteme, untergegangene Zivilisationen). Dynamische Systeme, die in die Zukunft reichen können nicht ausschöpfend erfasst werden, denn die Zukunft ist unbekannt.

Bei Induktion handelt es sich formal gesprochen darum, dass ein Modell x aus einer Klasse M , bis auf einen *Passungsgrad* ε ("Epsilon") zur Datenstruktur z passt. Nach und nach wird die Passung besser $\varepsilon \rightarrow \varepsilon'$. Im Idealfall wird es ein Modell geben, dass als Struktur mit der Datenstruktur identisch ist, aber Hypothesen liefert, die die Zusammenhänge in den Daten erklären.

Für das Entdecken von Zusammenhängen gibt es Computerprogramme, die nach festen Regeln – dies deutet auf die Strenge der Methode hin – vorgehen.²

¹Axiomatik: Nach Borovcnik wird ein Begriff wie Wahrscheinlichkeit nicht definiert, sondern der Begriff wird festgelegt und seine Bedeutung abgeklärt. Üblicherweise wird in der Mathematik hierfür eine Theorie verwendet, innerhalb dieser die Begrifflichkeit implizit festgelegt und abgegrenzt wird. Für den Wahrscheinlichkeitsbegriff ist dies durch die Axiomatik von Andrey Kolmogorov (1903-1987) geschehen (vgl. Kolmogorov 1933). Darin werden die Eigenschaften einer Wahrscheinlichkeit erklärt. Was inhaltlich unter dem Begriff Wahrscheinlichkeit zu verstehen ist bleibt dabei offen. So gesehen ist der Kolmogorovsche Axiomatik gleichermaßen für die objektivistische wie die subjektivistische Wahrscheinlichkeit anwendbar, da die inhaltliche Bedeutung von der Axiomatik unberührt bleibt (vgl. Borovcnik 1984: 133).

²Meschinelle Entdeckung wie sie in der KI verwendet durchkempt einen Suchraum und

deduktive Bestätigung

Eine deduktive Bestätigung drückt sich durch bewertende Aussagen aus, die eine Beziehung zwischen einem Ereignis und einer Theorie ausdrückt: "Durch das Vorliegen der Daten ist die Theorie nun besser bestätigt."

Allgemein wird aus einer Theorie bzw. ihren Hypothesen ein atomarer Satz gewonnen. Anschließend wird unabhängig von der Auswahl, das durch den Satz beschriebene Ereignis durch Beobachtung oder Experiment verifiziert. Tritt ein Ereignis ein, so liegt keine Verifikation – keine endgültige Bestätigung einer Theorie – vor, sondern die Theorie hat sich bewährt. Erst viele eintretende Ereignisse machen eine Theorie "brauchbar" [1, 308], wobei wir die potentiellen Daten in einem System bei der Bestätigung berücksichtigen sollten.³ Deduktion liegt auch vor, wenn der gleiche atomare Satz auf zwei unterschiedlichen Wegen gewonnen wird. Einerseits kann die logische Ableitung eines atomaren Satzes aus der Theorie erfolgen, andererseits ist die Verifikation über Beobachtung oder Experiment möglich. In all diesen Fällen stellt die Bestätigung einer benutzten Theorie eine Bestätigung durch Übereinstimmung dar [1, 309].

Auch die statistischen Testverfahren sind als deduktive Bestätigung einzuordnen. Statistische Testverfahren testen keinen atomaren Satz, sondern eine "probabilistische Aussage über die Wahrscheinlichkeit, dass ein Merkmal eine bestimmte Ausprägung, bzw. der Stichprobenkennwert einen bestimmten Wert annimmt" [1, 310]. In diesem Fall liegt *Übereinstimmung bis auf Wahrscheinlichkeit* von zwei unabhängigen Verfahren vor. Einerseits wird theoretisch die Wahrscheinlichkeit für das Auftreten eines Ereignisses abgeleitet, andererseits wird das Auftreten eines Ereignisses durch Datenbeschaffung und Berechnung bestimmt. Ist die Wahrscheinlichkeit groß und das Ereignis tritt ein, so besteht Übereinstimmung – natürlich ist auch der gegensätzliche Fall von Übereinstimmung möglich, wenn eine kleine Wahrscheinlichkeit besteht und sich das Ereignis nicht beobachten lässt.

Bestätigung durch Erfolg

Bestätigung durch Erfolg ist die angemessenste Form der Bestätigung in den "weichen" Disziplinen, in denen aus ethischen Gründen oder ökonomischen (finanziellen) Gründen das gesamte System datenmäßig nicht ausgeschöpft werden kann.

Probabilistische Bestätigung

Die Grundidee ist, den Bestätigungsgrad einer Theorie zu berechnen. Hierfür wird ein Wahrscheinlichkeitsraum zugrundegelegt und für die Sätze der Theorie Wahrscheinlichkeiten bestimmt. $p(h|e)$ ist dann der Grad, indem die Hypothese h durch die Daten e bestätigt wird. Theorien können auch verglichen werden, wenn $p(h|e) < p(h'|e)$, also h' durch die Daten e besser bestätigt wird. Leider funktioniert dieser probabilistische Ansatz nicht: Satzmengen können überabzählbar sein, so dass Wahrscheinlichkeiten nicht gebildet werden können und Wahrscheinlichkeitsräume müssten Allaussagen genügen können.

sucht befriedigende Ergebnisse. Ziel ist es nicht eine vollständige Erklärung zu finden, wie es die Informatik anstrebt. Heuristiken (Regeln) bestimmen welche Bereiche untersucht werden.

³Ein Vergleich von Barrow und seinen Graphiken zum Wesen wissenschaftlicher Forschung kann hier dienen.

3.1.3 Strukturen

empirische Theorien

Definition einer Theorie *Theorien* dienen dazu einen Zustand zu beschreiben. "Theorien stellen das in einem Zustand vorhandene Wissen in objektiver, meist sprachlich niedergelegter Weise dar. Sie sind dann in Form von Hypothesen, Sätzen oder Axiomen relativ leicht identifizierbar und in zusammengehörige Gruppen klassifizierbar. Theorien, die ein gewisses Maß an Kohärenz aufweisen, erlauben die Definition oder Konstruktion von Modellen. *Modelle* können wir uns in erster Näherung denken als begriffliche Konstrukte zur "Abbildung" realer Systeme oder zum Umgang mit solchen." [1, 16]

Es gibt unterschiedliche Theorien: kleine, "lokale" Theorien, die ein Gesetz enthalten bis hin zu den "reifen" Theorien, die viele Gesetze umfassen (Forschungsprogramme, Theorienetze und Theorieevolution). Ausgereifte Theorien findet man vor allem in der Physik (Thermodynamik, Mechanik) aber auch die ökonomische Gleichgewichtstheorie fällt in den Bereich. [1, 48]

(!) Ein lokale Theorie besteht aus vier Komponenten :

- einer Klasse von Modellen: **M**
- einer Menge intendierter Systeme: **I**
- einer Menge von Datenstrukturen: **D**
- einem Approximationsparameter: **U**

Eine Theorie **T** ist somit: $\mathbf{T}=(\mathbf{M},\mathbf{I},\mathbf{D},\mathbf{U})$

- (!) **Modelle** sind durch bestimmte Sätze, Axiome, Hypothesen, Gesetze und Annahmen der Theorie charakterisiert. Ebenso legen Satzmenge Modelle fest, so dass es keine Unterscheidung zwischen Modellen und Gesetzen gibt.
- (!) **Daten** sind – wie Gesetze – Sätze, die hypothetischen Charakter haben. Daten bestätigen Gesetze, bekommen aber auch erst durch die Gesetze Bedeutung. Eine Theorie, deren Daten und Gesetzen zusammenpassen nennen wir "brauchbar". Die Methoden der Datenfindung können dabei sehr komplex sein – siehe Messmethoden – und die Daten werden meist über die Sinne nicht direkt wahrgenommen. Wir befassen uns mit den Daten auf abstraktem Wege.
- (!) **Intendierte Systeme** der Theorie sind diejenigen Phänomene (Systeme) auf deren Untersuchung sich die Wissenschaftsgemeinschaft festlegt. Dies ist notwendig, da sich die Systeme meist nicht durch die beobachteten Daten festlegen lassen. Deswegen werden für die Systeme Modelle aufgestellt, die die Struktur der Systeme und deren Funktionsweisen in geschlossener und übersichtlicher Weise darstellen. Andererseits liefert die genaue Untersuchung eines Systems eine zugehörige Datenstruktur. Das theoretische Bild eines realen Systems besteht aus den Gesetzen der Theorie und den aus dem System stammenden Daten [1, 52].

(!) **Approximationsparameter** bedeutet die Beschreibung der Passung der Daten und der Gesetze. Gesetze passen immer nur approximativ zu ihren Daten, d. h. es werden Approximationen, Passungsgrade oder Signifikanzen festgelegt, die für die Akzeptanz der Theorie gelten.

(!) Zusammengefasst kann für die *Entstehung einer Theorie* gesagt werden, dass sie:

1. Intendierende Systeme festgemacht werden, die neue oder interessante Phänomene enthalten.
2. Für deren Modellbildung werden Modelle konstruiert. Modelle sind geistige Konstrukte – meist sprachlich –, die mit einem geeigneten Repräsentationsformalismus – also der Sprache – Ausdruck finden.
3. Aus den intendierten Systemen werden unter Anleitung der Modelle geeignete Daten gewonnen und man untersucht, wie gut die Modelle die intendierten Systeme abbilden.
4. Abschließend wird beurteilt, wie gut die Modelle mit den gewonnenen Daten zusammenpassen. Hierfür wird die Güte der Anpassung mit einem Approximationsformalismus bestimmt. Dieser wird übernommen oder neu entwickelt.

3.1.4 Daten

Daten drücken Sachverhalte aus. Wie oben beschrieben passen Daten nur immer approximativ zu den Theorien. Daten sind die singulären Gegenstücke zu der allgemeinen Theorien und bilden ein *Spannungsverhältnis* zwischen allgemeiner Systematisierung und fortwährendem Test.

Messmethoden

Messmethoden liefern einen systematischen, weil regelgeleiteten Zugang zu den Daten. Hierzu gehört auch die statistische Analyse. Die Entwicklung von Messmethoden kann den Zustand eines Wissenschaftssystems ändern (beispielsweise die Erfindung des elektrischen Thermometers, ein notwendiges Instrument bei der Entwicklung der Quantenphysik). Der PC ist heute ein Instrument, das nicht nur unterstützt sondern auch vorantreibt – so können aus Daten neue Theorien konstruiert werden (Data-Mining).

Unterscheidung in harte und weiche Daten

Wissenschaftliche Veränderungen hängen mit der Natur der zu untersuchenden Objekte zusammen. Hierbei gibt es stabile und einfache Objekte, wie sie in der Physik untersucht werden. Diese Objekte oder Systeme lassen sich schneller und erfolgreicher beforschen als veränderliche und komplexe Objekte, wie sie in den Sozialwissenschaften auftreten. Stabile Objekte ermöglichen eine unkomplizierte Datenerhebung, wohingegen die Datenerhebung bei komplexen Objekten wesentlich schwieriger ist [?, 24]. Als Basissätze gelten jene Daten einer Theorie, die die Eigenschaften haben: Wiederholbarkeit der Bestimmung und Einigkeit unter den Benutzern. Eine mögliche Skala von harten bis weichen Daten

reicht von: Naturwissenschaften – Physik, Chemie, Biologie, die historischen- und Geisteswissenschaften, Linguistik und Sozialwissenschaften – Psychologie, Soziologie, Ökonomie, Politologie, bis zu den praktischen Disziplinen – Jura und Medizin.

Harte Daten sind konstruierbare Daten, die von Maschinen entstammen, materiellen Gegenständen oder experimentellen Anordnungen. Dann folgen die Daten statischer Natur: sie ändern sich kaum im Zeitverlauf, zu finden in den Naturwissenschaften oder in literarischen Texten (historische Wissenschaften). Vergängliche Daten liegen dann vor, wenn eine gewisse Periodizität festgestellt werden kann – die Wiederholungen sind fast identisch (das Sonnensystem, Pendelschwingungen). Die weichsten Daten – und damit die seltensten – stammen von Systemen, die nicht periodisch sind, nur in einzelnen, oder sehr wenigen Exemplaren auftreten und für die auf theoretischer Seite nur qualitative und vage Vorhersagen möglich sind (beispielsweise das Rätssystem kurz nach der Oktoberrevolution, die altägyptische Form der Staatsökonomie).

In den Naturwissenschaften liegen oft tausende Fälle vor (experimentelle Designs von unterschiedlichen Materialien können 1000-fach wiederholt werden), in den Sozialwissenschaften wenige Hundert (und das ist meist schon viel).

Entscheidend bei Daten, egal welchen Grades, ist ihre Wiederholbarkeit. Ist die Annahme der Wiederholbarkeit verletzt, so gibt es keinen Grund mehr von einem Datenbegriff auszugehen. Weiche und harte Daten lassen sich noch nach der Objektivität beurteilen, also daran, inwieweit Übereinstimmung herrscht, nur ist verständlicherweise bei weichen Daten der Einigungsprozess weniger klar strukturiert.

Datenstrukturen

Datenstrukturen erfassen die in einem intendierten System vorkommenden Informationen, ohne dass alle in dem System vorkommenden Daten erfasst werden müssen. In einer Datenstruktur lassen sich statistische Merkmale, wie Mittelwert und Standardabweichung definieren, die für die Güte der Passung, dem *Passungsgrad* zwischen Daten und Modellen benötigt werden.

Ausgangspunkt ist ein Messreihe von n wiederholbaren Messungen eines bestimmten Merkmals. Für die meisten Messreihen gilt, dass die Werte normalverteilt sind, d.h. die Werte häufen sich um den Mittelwert μ ("mü"). Die meisten Werte einer Messreihe liegen in einem bestimmten Bereich $]\mu - \sigma, \mu + \sigma[$

Die Standardabweichung $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$ wird als die Wurzel der durchschnittlichen quadratischen Abweichung σ ("sigma") vom Mittelwert zum Mittelwert μ bezeichnet. Die Standardabweichung ist also der Abstand der Grenzen von seinem Mittelwert oder auch ein Maß der Streuung der Werte um den Mittelwert.

Approximative Passung

Die Passung von Daten und statistischen Theorien bzw. statistischen Modellen ist ein weitläufiges Feld, in das wir in den folgenden Wochen immer tiefer einsteigen werden. Hier nur allgemeine Überlegungen.

Die Grundfrage lautet, wie die Abweichung der Daten von den theoretischen Funktionen (den wissenschaftlichen Tatsachen) modelliert werden kann. Wir können nie von einer idealen Passung von Daten und Theorien ausgehen, da

die Daten nicht einheitlich sind, sondern immer nur eine ungefähre – *approximativen* – Passung vorliegt. Der Passungsgrad ε wird pragmatisch festgelegt, er ist Konvention oder besitzt, wie Balzer es formuliert eine "*meta-theoretischen Status*".

Als *Passungsgrad* wird definiert, dass ein gegebenes Modell x bei einer gegebenen Datenstruktur z bis auf den Grad ε ("epsilon") zueinander passen [1, 212]. Hierfür benötigen wir einen Approximationsapparat \mathbf{U} . Dieser besagt, dass es eine Schranke für eine befriedigende Passung "bis auf ε " gibt. Wenn ein zulässiger Passungsgrad ε_0 ("epsilon Null") vorliegt und das Modell x nur bis auf $\varepsilon > \varepsilon_0$ zu den Daten z passt, dann ist die approximative Passung von x und z unbefriedigend. Als Schlussfolgerung müsste man sagen, dass das bisherige Modell angesichts der Daten kein befriedigendes Modell ist.

Ausgehend von einer Messreihe stellt sich die Frage, welches der "wahre" Wert – derjenige Wert, der die Theorie sicher bestätigt – ist. Der Mittelwert wird aus statistischen Gründen – auf die wir noch näher eingehen – als der passendste Wert für den "wahren" Wert betrachtet. Zu erinnern gilt, dass der wahre Wert theoretisch beschrieben wird, real aber unbekannt ist. Abhilfe schafft der Mittelwert μ , da die meisten der formal möglichen Werte einer Messreihe im Bereich $]\mu - \sigma, \mu + \sigma[$ liegen. Die Standardabweichung wird oft als Maß dafür genommen, ob ein gemessener Wert noch eine tolerable Abweichung vom "wahren" Wert hat. Eine befriedigende Passung vom gemessenen Wert vom theoretischen Wert ist dann gegeben, wenn der gemessene Wert minus dem theoretischen Wert kleiner der Standardabweichung ist. Liegen keine theoretischen Annahmen für den theoretischen Wert vor, so kann auch der Mittelwert der Messreihe genommen werden.

Kapitel 4

Wahrscheinlichkeitstheorie I

4.1 Statistische Theorien

Dieser Abschnitt ist eine Zusammenfassung von Balzer [1, 107-117].

Statistik behandelt die Häufigkeitsverteilungen einer Population. Es gibt eine Menge P von Objekten, die eine oder mehrere Eigenschaften mit variabler Ausprägung aufweisen. Als Population können wir Menschen betrachten, die die Eigenschaft Alter mit variabler Ausprägung M haben¹; in der Ökonomie sind es Güterklassen, die als Ausprägungen unterschiedliche Preise haben; oder in der Genetik kommen bei Menschen unterschiedliche Augenfarben vor. Jedes Element hat genau eine Ausprägung, d. h. jedem Element der Population P kann genau eine Eigenschaft M zugeordnet werden. So kann man die Situation in einer Funktion darstellen: $h : P \rightarrow M$. h ist die Anzahl von Objekten einer Population, die die verschiedenen Ausprägungen des Merkmals haben. Von Interesse sind dabei nicht die absoluten Werte, sondern die *Zahlenverhältnisse*, in denen die verschiedenen Ausprägungen zueinander stehen.

Betrachten wir Zahlenverhältnisse, so spricht man von *relativen Häufigkeiten*. Die relative Häufigkeit ist eine Eigenschaft von Merkmalsausprägungen, so dass man sagen kann: Eine bestimmte Ausprägung hat in einer Population P die relative Häufigkeit α ("Alpha"). Relative Häufigkeiten sind keine Eigenschaft von einzelnen Objekten, sondern eine Eigenschaft von Mengen von Objekten (alle Blauäugigen, alle Personen über 65 Jahre).

Intuitiv sprechen wir nicht von relativen Häufigkeiten, sondern von Eigenschaften eines Objektes, denn allgemein spricht man: die relative Häufigkeit, dass ein zufälliges Kind blaue Augen habe sei $10/20$. Korrekterweise müsste man sprechen: Die Menge der blauäugigen Kinder hat die Menge $10/20$. Ein tatsächliches Ereignis wird als die Menge all der Individuen mit einer bestimmten Ausprägung bestimmt, verbal wird aber von dem Objekt gesprochen, dass eine bestimmte Ausprägung hat – also von möglichen Ergebnissen – von Wahrscheinlichkeiten.

Balzer meint, dass man heute nicht mehr von abstrakten Merkmalsausprägungen spricht, sondern, dass es in unserem heutigen Wahrscheinlichkeitsverständnis selbstverständlich ist, nicht von Mengen, sondern von zufällig heraus-

¹Untere Grenze = 0 Jahre, obere Grenze = etwas über 100 Jahre

gegriffenen Individuen (Objekten) einer Population zu sprechen. Man sagt: "das Ereignis, dass ein beliebig herausgegriffenes Individuum die Merkmalsausprägung a besitze, oder das Ereignis, dass ein Experiment ein bestimmtes Ergebnis zeige, habe die relative Häufigkeit α . Er nennt dies den "Aufstieg der "probabilistischen" Weltsicht" [1, 108, Hervorhebung im Original].

Für die Sozialwissenschaften, aber auch für andere Wissenschaften, trifft es zu, dass für bestimmte Populationen eine genaue Bestimmung der relativen Häufigkeiten nicht möglich ist, weil sie ökonomisch nicht machbar (Einkommensverteilung aller Deutschen) oder die Population eine offene, potentiell unendliche Menge bildet (Experimente in der Quantenmechanik können beliebig oft wiederholt werden).

Wie kann man mit nun mit relativen Häufigkeiten und Verteilungen umgehen, die ja von großem wissenschaftlichem Interesse sind? Dieser Frage stellt sich die statistische Theorie.

Verwirklicht wurde dies historisch über ein (meta-)theoretisches Konstrukt, dessen Bedeutung durch präzise Annahmen (Axiome) festgelegt wird und im Spezialfall einer relativen Häufigkeit entspricht: dem *Begriff der Wahrscheinlichkeit*. Die grobe Idee hinter dem Konstrukt der Wahrscheinlichkeit ist, dass relative Häufigkeiten durch reelle Zahlen zwischen 0 und 1 ersetzt werden und zum Umgang mit den Zahlen Rechenregeln erstellt werden (*Axiome für Wahrscheinlichkeitsräume*).

Durch die Konstruktion von Wahrscheinlichkeiten kann man das Problem von potentiell unendlichen Grundgesamtheiten (wodurch bei relativen Häufigkeiten der Nenner nicht definiert wäre) umgehen. Die Wahrscheinlichkeit ist ein Größen- und Maßverhältnis, das sich nicht durch ganzzahlige Brüche zweier Ereignisse fassen lässt, denn das eine Ereignis drückt aus, welche bestimmte Eigenschaft ein Objekt hat und das andere Ereignis drückt aus, dass ein Objekt irgendeine Eigenschaft hat. Wahrscheinlichkeiten sind damit ein abstraktes Gegenstück zu relativen Häufigkeiten.

4.1.1 Wahrscheinlichkeitsfunktion

Es lässt sich eine Wahrscheinlichkeitsfunktion p definieren, so dass bestimmten Ereignissen (oder Teilmengen) einer Population Ω ("Omega") Zahlen zwischen 0 und 1 zugeordnet werden. Die Wahrscheinlichkeitsfunktion p ist dann eine Funktion in einem reellen abgeschlossenen Zahlenintervall $[0, 1]$. Die Wahrscheinlichkeitsfunktion bildet einen Wahrscheinlichkeitsraum, der bestimmte mathematische Voraussetzungen – Axiome – erfüllt.²

Eine Zufallsvariable drückt die Wahrscheinlichkeit aus, dass ein bestimmtes Objekt aus der Grundgesamtheit beobachtet wird – jeder Ausprägung wird die Wahrscheinlichkeit ihres Vorkommens zugeordnet. Vereinfacht wird dies durch Annahmen über die Verteilung der Zufallsvariable.

Verteilungen von Zufallsvariablen sind mathematische Formen, die meist über eine Dichte definiert werden. Eine Dichte von Verteilungen ist eine Funktion mit der Eigenschaft, dass sich die Werte der Verteilung über ein Integral

²u.a. a) Ω ist eine nicht-leere Menge, b) die Wahrscheinlichkeit der Teilmengen ist 1; und weitere

der Dichte ausdrücken lassen. Als *statistisches Element* wird die Kombination einer Zufallsvariable mit einem Wahrscheinlichkeitsraum genannt. Eine *Verteilungshypothese* ist dann ein Satz, der für ein statistisches Element die genaue Form der Zufallsvariable über die konkrete Grundgesamtheit festlegt.

Die bekannteste Verteilung ist die Normalverteilung, die durch eine Dichte in der bekannten Glockenform definiert wird.

$$f(x|\mu, \sigma) = P(X = x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (4.1)$$

Neben der Normalverteilung gibt es u. a. die Binominalverteilung, Chi-Quadrat-Verteilung, die F-, die Poisson- und die t-Verteilung. Die Form einer Verteilung ist eine Gleichung, in der Parameter vorkommen. Für die Normalverteilung sind dies die Parameter Mittelwert μ und Varianz σ^2 . Bei numerischer Fixierung der Parameter erhält man eine Funktion, die je nach Parameterwahl variiert, allerdings bleibt die charakteristische Verteilungsform erhalten.

Eine *reine statistische Theorie* ist eine Theorie, deren Modelle aus einem statistischen Element und deren Axiome aus einer Verteilungshypothese für dieses Element bestehen. Meist legt die Verteilungshypothese die Verteilung der Zufallsvariable über eine Dichte fest. So kann mathematisch durch die Berechnung aus einem Integral die Wahrscheinlichkeit bestimmt werden, dass eine Person in einer Population ein bestimmtes Alter hat.

Die bisherigen Begriffe deuten auf ein rein mathematisches Konstrukt hin, doch dem ist nicht so. Deutlich ist, dass die Zufallsvariable empirisch bestimmt wird, aber auch der Wahrscheinlichkeitsraum ist empirischer Natur, denn die Kategorisierung der Variable (die Bestimmung der einzelnen Zustände) beruht auf Beobachtungen und kann nicht mathematisch hergeleitet werden. Reine statistische Theorien sind also empirische Theorien, auch wenn sie nur eine Verteilung der Zufallsvariable festlegen.

Statistische Theorien weisen zwei Charakteristika auf: Entweder sie werden verwendet um Phänomene zu erklären, die aus einer *großen Menge von Objekten* bestehen und deren Erfassung zu teuer ist, so dass man sich auf Mittelwerte und andere Verteilungscharakteristika beschränkt oder sie beschreiben Vorgänge, die *nicht-deterministischer Natur* sind, wie beispielsweise dem Würfelwurf oder Spinexperimenten in der Quantenphysik.

4.2 Wahrscheinlichkeiten nach Rohwer und Pötter

4.2.1 Unterschiedliche Generalisierungsprobleme

- Deskriptive Generalisierungsproblem: Wie lassen sich Aussagen über eine Verteilung eines Merkmals einer Stichprobe auf eine real existierende Grundgesamtheit verallgemeinern? Angenommen wird beispielsweise, dass es für eine bestimmte Grundgesamtheit von Haushalten eine bestimmte Verteilung des Einkommens geben kann und das man diese Verteilung mit Stichproben näherungsweise ermitteln kann.
- Modales Generalisierungsproblem: Was wird in noch nicht beobachteten Situationen in einer fiktiven Grundgesamtheit geschehen? Hierbei handelt

es sich um eine Aussage in der Zukunft.

4.2.2 Epistemische und aleatorische Wahrscheinlichkeiten

Entscheidend für eine Unterscheidung des Wahrscheinlichkeitsbegriffs ist deren Verwendung.

- Epistemische Wahrscheinlichkeiten drücken eine Qualifizierung eines unsicheren Sachverhaltes aus. Eine Quantifizierung ist meist nicht möglich, eine vergleichende Abschätzung allerdings schon: das eine Ereignis ist wahrscheinlicher als das andere Ereignis.
- Aleatorische Wahrscheinlichkeiten quantifizieren Wahrscheinlichkeitsaussagen. Die Quantifizierung wird ermöglicht, da man die Regeln für das Zustandekommen der Wahrscheinlichkeiten kennt. Dieser Ansatz lässt sich zurückführen auf Jacob Bernoulli (1654-1705), der Überlegungen zu Verfahren anstellt, die Sachverhalte erzeugen.

Zufallsvariablen sind das Ergebnis von Zufallsgeneratoren und beziehen sich auf mögliche Ereignisse (Würfeln). Im Gegensatz dazu beziehen sich statistische Variablen auf realisierte Sachverhalte (Einkommensverteilung).

4.2.3 Stochastische Inferenz

Ein stochastisches Inferenzproblem liegt vor, wenn die Konstruktion eines Zufallsgenerators unbekannt ist und aus den beobachteten Daten lediglich Vermutungen über die Wahrscheinlichkeitsverteilung des Zufallsgenerators möglich sind, weil die Regeln unbekannt sind.

Im Gegensatz dazu ist die Wahrscheinlichkeitsverteilung in einer Urne mit bekannter Anzahl schwarzer und weißer Kugeln gegeben.

4.2.4 Deskriptive Generalisierungsprobleme

Ist die Art der Stichprobenauswahl bekannt (beispielsweise Zufallsauswahl), so stellt sich kein Inferenzproblem. Ein Auswahlgenerator ist ein bekannter Zufallsgenerator, der zur Auswahl der Elemente verwendet wird (Zufallsauswahl, Quotenauswahl). Allerdings kann die Annahme des Auswahlgenerators verletzt sein.

Die stochastische Inferenz löst die Frage, wie oft bei dem gleichen Verfahren der Datengewinnung ähnliche Ergebnisse erzeugt werden (95% der Fälle beispielsweise) und man den gleichen Schätzwert erhalten würde. Dies sagt nichts über die Qualität der tatsächlich gegebenen Daten aus. Letztlich werden durch die hypothetische Wiederholung einer realisierten Stichprobe – unter der Annahme eines bekannten Auswahlgenerators – quantifizierbare Wahrscheinlichkeiten gebildet.

Generalisierte Aussagen sind zeitlich und räumlich begrenzt, da sich die beobachteten Daten auf einen Zeitpunkt beziehen und Aussagen über vergangenen und zukünftige Beobachtungen nicht zulassen. Die Tragweite von Aussagen, die mit einem Auswahlgenerator getroffen wurden, reicht nicht in die Zukunft.

Prozesse, die für wissenschaftliches Handeln notwendig sind:

- *datengenerierende Prozesse*: Verfahren der Stichprobenziehung
- *datenerzeugende Prozesse*: zu untersuchender Sachverhalt muss bereits vorhanden sein
- *substantielle Prozesse*: Prozesse, durch die Sachverhalte entstehen und sich ändern, sie sind die Grundvoraussetzung für realisierte Sachverhalte.

4.2.5 Modale Generalisierungsprobleme

Modale Generalisierungsprobleme beziehen sich nicht auf eine Referenzgesamtheit mit bereits realisierten Prozessen, sondern man stellt die Frage wie Prozesse ablaufen und wahrscheinlich ablaufen werden. Man kann hierfür keine Grundgesamtheit definieren. Für die unbekannt Situation wird ein *Ablaufschema* (Modell) konstruiert, das eine Situation definiert und sich daher auf eine mögliche Grundgesamtheit bezieht.

Modelle sind keine Abbildungen der Realität, sondern Ablaufschemas, die mögliche Prozesse beschreiben. Ablaufschemas können nur festlegen, welche Abläufe als möglich betrachtet werden sollen, nicht jedoch, welche Abläufe tatsächlich möglich sind (Würfelwurf als ideale Situation vs. Würfelwurf in der Realität).

Allerdings sollten in der Wissenschaft die Ablaufschemas quantifiziert werden, um so Wahrscheinlichkeitsaussagen über das Eintreten der möglichen Zustände zuzulassen. Hierfür wird stochastische Inferenz angewendet.

4.2.6 Stochastische Ablaufschemas

Entspricht das Ablaufschema einem Zufallsgenerator, so spricht man von einem stochastischen Ablaufschema. Nun können auch statistische Daten, die aus bisherigen Beobachtungen gewonnen wurden, generalisiert werden.

Allerdings handelt es sich um eine Fiktion – was man immer im Hinterkopf behalten sollte. Die Fiktion beruht auf der Überlegung, dass die Werte – wie bei einem Würfelwurf – durch den Zufallsprozess entstanden sind. Abstrakt gesprochen sind die beobachteten Ergebnisse Prozesse, die durch die Aktivierung eines Zufallsgenerators entstanden sind, wodurch ein stochastisches Ablaufschema unterstellt wird und es sich damit um objektivierbare Eigenschaften des Zufallsgenerators handelt. Allerdings – um wieder in die Realität zurückzukehren – hängt das Verhalten eines beobachteten Menschen/ Ereignisses nicht von der Aktivierung eines Zufallsgenerators ab, sondern beruht auf anderen Tatsachen.

4.2.7 stochastische Modelle

Stochastische Modelle ermöglichen die Umwandlung eines modalen Generalisierungsproblems in ein stochastisches Inferenzproblem. Voraussetzung für die Anwendung stochastischer Modelle sind wiederholbare Prozesse, die sich zusätzlich im Zeitverlauf nicht ändern dürfen, sondern sich nach gleichen Regeln wiederholen. Die Vorteile des Verfahrens ist die Modellierung bzw. Formulierung einer Problemstellung, so dass sie zeitlos gilt und intersubjektiv nachprüfbar ist. Weiterhin kann durch die Quantifizierung ein höheres Niveau an Eindeutigkeit erzeugt werden – der Informationsgehalt einer Theorie kann gesteigert werden.

4.2.8 Fiktive Populationen

Erst in den 1930er und 1940er Jahren hat es sich durchgesetzt, dass man von einer Zufallsauswahl ausgehen sollte um wahrscheinlichkeitstheoretische Inferenzprobleme zu lösen. Maßgeblich dazu beigetragen haben die Gedanken von R. A. Fisher (1890-1962) und Jezy Neyman (1894-1981). Weist Fisher noch auf den Unterschied zwischen realen Sachverhalten und durch Experimente erzeugte Sachverhalte hin, so ist dies bei Neyman nicht mehr relevant. In Deutschland fand allerdings die Verwendung von mathematischen Hilfsmitteln und Stichproben bis in die 1960er Jahre wenig Beachtung. Zählen und Häufigkeiten bzw. Mittelwerte feststellen war erstrebenswerter.

4.2.9 Wahrscheinlichkeit und Chance

Rohwer behauptet, dass sich der Wahrscheinlichkeitsbegriff zu einem bloßen rhetorisch verwendetem Deutungsschema entwickelt hat. Vorstellen sollte man sich einen Zufallsgenerator, der Situationen eines bestimmten Typs erzeugt. Laplace gibt einen Vorschlag, dass man sich die Geburt wie das Ziehen aus einer Urne vorstellen sollte. Kennt man das Verteilungsgesetz, also die Proportion, in denen die Urne die beiden Sorten von Kugeln enthält, so kann man die Wahrscheinlichkeit des Geschlechts vorhersagen. Die Gründe für das Zustandekommen des Prozesses (die Geburt) können nicht genannt werden, daher ist es für die wissenschaftliche Arbeit unerlässlich eine Antworten zu finden, wie ein Zufallsgenerator aktiviert wird, wie also der Prozess in Gang gesetzt wird.

Chance

Der Begriff Chance bezieht sich in erster Linie einen handlungspraktischen Sinn, denn man spricht von Handlungsmöglichkeiten und wie sie sich realisieren lassen. Dabei spielt der Kontext eine Rolle und die subjektive Einschätzung von Handlungen. Die Sozialstatistik benötigt aber einen Chancebegriff, der sich nicht auf Handlungschancen von Subjekten, sondern auf Eigenschaften von Situationen, in denen sich soziale Akteure befinden, bezieht. Man spricht dann von Situationen als in der Realität vorhandene Begebenheiten. Die Realisierungsmöglichkeiten können durch Wahrscheinlichkeiten abschätzbar gemacht werden.

Der Wahrscheinlichkeitsbegriff bezieht sich auf einen Zufallsgenerator, einem Verfahren, das nach bestimmten Regeln unabhängig von Ort und Zeit abläuft. Der Chancebegriff bezieht sich auf Situationen und nicht auf ein Verfahren, in der mit gewissen Wahrscheinlichkeiten mögliche Vorgänge geschehen.

klassische Wahrscheinlichkeit

Aleatorische wie auch klassische Wahrscheinlichkeiten sind dadurch gekennzeichnet, dass sie quantifiziert werden können. In der klassischen³ Begriffsbildung geht man davon aus, dass sich die Wahrscheinlichkeiten durch relative Häufigkeiten bilden lassen. Richard von Mises (1883-1953) spielt bei dieser Idee eine bedeutende Rolle.

Von Mises bezieht sich auf die Überlegungen von Pierre-Simon Laplace (1749-1827), der mit seiner Definition von Wahrscheinlichkeit die Grundlage

³auch klassische, frequentistische, objektivistische oder Laplace-Wahrscheinlichkeit genannt

für das heute gebräuchliche Wahrscheinlichkeitsverständnis liefert. Demnach ist die Wahrscheinlichkeit eines Ereignisses die Anzahl der günstigen Fälle dividiert durch die Anzahl der gleich möglichen Fälle. Laplace wiederum bezieht sich auf Jacob Beroulli (1655-1705) der Überlegungen zu idealen Glückspielsituationen macht und das Gesetz der Großen Zahl entwickelt.

$$\text{Wahrscheinlichkeit eines Ereignisses} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der gleich möglichen Fälle}} \quad (4.2)$$

Laplace führt nun das *Prinzip des unzureichenden Grundes* ein um so von gleichen Möglichkeiten zu sprechen. Das Prinzip des unzureichenden Grundes besagt, dass man dann von 'gleich möglichen' Sachverhalten ausgehen kann, wenn man keinen Grund angeben kann, das eines der Sachverhalte 'eher möglich' ist. Diese Definition verwischt die Tatsache, dass man Ursprünglich bei der Definition von Wahrscheinlichkeiten von einem Zufallsgenerator ausgegangen ist und nicht vom Prinzip des unzureichenden Grundes.

Von Mises definiert Wahrscheinlichkeiten als den Grenzwert relativer Häufigkeiten, da er von einer unendlichen Anzahl unabhängiger Versuche desselben Zufallsexperimentes ausgeht. Die relative Häufigkeit ist der daraus resultierende Grenzwert, also das Auftreten eines Merkmals in der Beobachtungsfolge bei unbegrenzter Fortsetzung der Versuche. Von objektivistischen Wahrscheinlichkeiten spricht man, da die Wahrscheinlichkeit auf Beobachtungen eines Objektes zurückzuführen sind. Es wird also unterstellt, dass ein Objekt eine hypothetisch konzipierbare Größe besitzt – eine hypothetische Konstante.

Damit handelt es sich nicht mehr um aleatorische Wahrscheinlichkeiten, die sich als Eigenschaften von Zufallsgeneratoren definieren lassen sondern um *frequentistischen Wahrscheinlichkeiten* oder F-Wahrscheinlichkeiten. F-Wahrscheinlichkeiten sind unbekannt und nur näherungsweise durch empirische prüfbare relative Häufigkeiten feststellbar, sie sind also ein *Grenzwert von relativen Häufigkeiten*, wie ihn von Mises definiert. Die Konstante bezieht sich auf Prozesse, die Situationen hervorbringen durch die dann relative Häufigkeiten bestimmt werden können.

Der Unterschied von F-Wahrscheinlichkeiten und aleatorischen Wahrscheinlichkeiten liegt darin, dass sich aleatorische Wahrscheinlichkeiten auf Eigenschaften eines Verfahren beziehen und nicht wie die F-Wahrscheinlichkeiten auf die erzeugenden Situationen des Verfahrens, also auf die Folge von Ereignissen.

Problematisch bei dem Ansatz für die Sozialstatistik ist, dass man es nicht immer mit einer Folge von Ereignissen zu tun hat, sondern auch mit abgegrenzten Mengen.

Kapitel 5

Wahrscheinlichkeitstheorie II

5.1 Grundmodell der statistischen Methodik

Zur Wiederholung der Grundproblematik der Vorlesung¹. Statistik kann als eine Methode aufgefasst werden, Information einer – abstrakt gesprochen – irgendwie gearteten Wirklichkeit zu gewinnen. Jede angewandte Statistik steht nun in Beziehung zu ihrem Untersuchungsgegenstand. Um sich mehr Klarheit über diese Beziehung zu verschaffen, sollte man sich anfangs die grundlegenden Charakteristika der Methodik vergegenwärtigen.

Brachinger meint, dass die Sozialstatistik methodisch auf relativ einfache Elementarbausteine zurückgreift. Methodisch handelt es sich um wohlabgegrenzte diskrete Menge von empirischen Objekten, (Erhebungseinheiten), die die Grundgesamtheit bilden. Ferner wird angenommen, dass es möglich ist, jedem Objekt genau eine wohldefinierte Eigenschaft (Merkmalsausprägung) zuzuordnen. Nach der Zuordnung allen Elementen der Grundgesamtheit oder einer geeigneten – weil repräsentativen – Stichprobe und nach einer geeigneten Klassifikation der einzelnen Elemente der Grundgesamtheit bzw. der Stichprobe erhält man eine Häufigkeitsverteilung. Mit geeigneten Analysemethoden werden nun die gesuchten Informationen aus den Daten heraus"destiliert".

Balzer hat schon darauf hingewiesen, dass die statistische Analyse nicht als ein rein logisches Verfahren betrachtet werden sollte, empirisch daran ist die Bestimmung der Grundgesamtheit, die Klassifikation der Merkmalsausprägungen und die raum-zeitliche Einordnung der gewonnenen statistischen Ergebnisse.

5.2 Zufallsvariable

5.2.1 Zufallsauswahl

Von einer Zufallsauswahl bzw. repräsentativen Stichproben spricht man, wenn jede Einheit aus der Grundgesamtheit die *gleiche Chance* hat in die Stichprobe zu kommen. Ist diese Chance für einzelne Einheiten nicht gegeben, so handelt es sich um systematische Verzerrungen und man spricht von nicht-repräsentativen Stichproben.

¹Dieser Abschnitt bezieht sich auf [4, 10]

5.2.2 Formale Definition einer Zufallsvariable

Vergleiche im Folgenden [10, 322]. Gegeben ist ein Zufallsexperiment mit der Ergebnismenge Ω . Eine Zufallsvariable X ist eine Abbildung, die jedem $\omega \in \Omega$ (jedem Ergebnis einer Ergebnismenge, gesprochen "klein Omega Element von groß Omega") eine reelle Zahl $X(\omega) = x$ zuordnet. Von Interesse ist aber nicht die Abbildung von X , sondern man möchte etwas über die Verteilung von X wissen und so jeder Merkmalsausprägung der Zufallsvariable x_i eine Wahrscheinlichkeit p_i zuordnen. Die Wahrscheinlichkeitsverteilung gibt an mit welcher Wahrscheinlichkeit eine Zufallsgröße einen definierten Wert annimmt; formal: $P(X = x_i) = p_i$ (gesprochen: die Wahrscheinlichkeit, mit der die Zufallsvariable X den Wert x_i annimmt ist gleich p_i).

Von Interesse ist ja die *Verteilung der Zufallsvariable* X , daher die Wahrscheinlichkeitsverteilung von X . Die Wahrscheinlichkeitsverteilung ist die Zuordnung von Wahrscheinlichkeiten für X : $P(X)$. Möchte man beispielsweise wissen wie viele Wohnungen in Köln weniger oder gleich 1000,- Euro kosten und hat man einen geeigneten Datensatz, so ist sind ω die einzelnen Elemente des Datensatzes Ω , X die Zufallsvariable Wohnungsmiete mit $x = X(\omega)$ und die Ergebnismenge des interessierenden Ereignisses $X \leq x$ ist $X \leq 1000$. Eine Folge von Zufallsvariable (Nettomieten in Köln) X_1, \dots, X_n mit den *Merkmalsausprägung* x_i Preis in Euro erhält man in obigem Beispiel, indem man eine unabhängig und identisch verteilte Zufallsstichprobe (ähnlich wie bei einem Würfelwurf) unter allen Kölner Mietwohnungen zieht.

Verteilungsfunktion

Kennt man die Wahrscheinlichkeiten $P(X \leq x)$ für die Ereignisse der Form $X \leq x$, so lassen sich die Wahrscheinlichkeiten für andere Ereignisse durch eine Funktion berechnen. Eine Funktion $F(x)$, die jedem $x \in \mathfrak{R}$ die Wahrscheinlichkeit $P(X \leq x)$ zuordnet heißt Verteilungsfunktion von X :

$$F(x) = P(X \leq x). \quad (5.1)$$

Unabhängig und identische Wiederholungen

Sei X eine diskrete oder stetige Zufallsvariable mit Erwartungswert μ und Varianz σ^2 und einer bestimmten Verteilungsfunktion F . Der zu X gehörende Zufallsvorgang werde n -mal unabhängig wiederholt. Für jeden Teilversuch erhält man eine bestimmte Zufallsvariable X_i . Die Zufallsvariablen X_1, \dots, X_n sind unabhängig und besitzen alle die gleiche Verteilungsfunktion und damit insbesondere den gleichen Erwartungswert μ und die gleiche Varianz σ^2 . Dann kann man sagen: X_1, \dots, X_n sind unabhängig identisch verteilt (i.i.d.)² wie X .

Diese Annahme ist bei vielen Zufallsstichproben vom Umfang n für ein Merkmal zumindest näherungsweise erfüllt und diese Annahmen sind Voraussetzung für weitere statistischen Gesetze, wie das Gesetz der großen Zahl und den zentralen Grenzwertsatz.

Gesetz der großen Zahlen und Grenzwertsätze

Vergleiche im Folgenden: [10, 307-317]

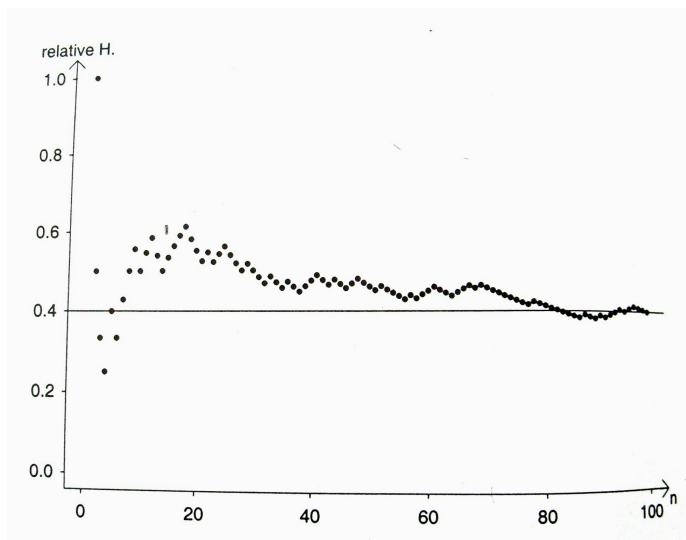
²independent identical distributed

Ausgangsüberlegung ist ein Bernoulli-Versuch (Würfelwurf, Urnenexperiment) mit gleichbleibender Wahrscheinlichkeit $P(A)$ für das Auftreten des Ereignisses A .

Gesetz der großen Zahl

Das Gesetz der großen Zahl besagt, dass bei großem n die relative Häufigkeit für das Auftreten von A nahe bei der Wahrscheinlichkeit $P(A)$ liegt und für $n \rightarrow \infty$ konvergiert die relative Häufigkeit gegen die Wahrscheinlichkeit $P(A)$.

$$f_n(X = 1) \rightarrow P(A) = P(X = 1) \text{ für } n \rightarrow \infty \quad (5.2)$$



Graphik aus [10, 308]

Das Bild zeigt, wie mit wachsender Anzahl der Versuchswiederholungen die relative Häufigkeit f_n einer binominalverteilten Zufallsvariable $X \sim B(1, \pi)$ tendenziell immer näher am wahren Werte $\pi = 0.4$ konvergiert.

Das Gesetz der großen Zahl sagt aus, dass die Wahrscheinlichkeit $P(\mu - c \leq \hat{X}_n \leq \mu + c)$ mit der das arithmetische Mittel in ein beliebig vorgegebenes Intervall $[\mu - c \leq \mu + c]$ fällt, gegen 1 konvergiert, wenn $n \rightarrow \infty$. Für großes n ist damit $P(\mu - c \leq \hat{X}_n \leq \mu + c)$ nahe bei 1. Man sagt, dass das empirische arithmetische Mittel \hat{X}_n nach *Wahrscheinlichkeit gegen μ konvergiert*.

Der Erwartungswert des arithmetischen Mittels ist dann $E(\bar{X}_n) = \mu$ und die Varianz ist $Var(\bar{X}_n) = \frac{\sigma^2}{n}$. Man sieht, dass mit steigendem n die Varianz umgekehrt proportional ist und gegen null geht.

Hauptsatz der Statistik

Vergleiche im Folgenden: [10, 311].

Gezeigt werden soll, dass eine empirische Verteilung umso näher an einer theoretischen Verteilung liegt, je größer die Anzahl der Wiederholungen n ist.

Der Hauptsatz der Statistik besagt, dass für Zufallsstichproben, bei denen X_1, \dots, X_n unabhängig und identisch wie das interessierende Merkmal X verteilt sind, die unbekannte Verteilung $F(x)$ von X durch die empirische Verteilungsfunktion $F_n(X)$ für $n \rightarrow \infty$ gut approximiert wird. Stimmen umge-

kehrt $F_n(X)$ und eine theoretische Verteilung $F(x)$, etwa die Normalverteilung, schlecht überein, so entstammen die Daten vermutlich einer anderen Verteilung.

Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz rechtfertigt die Annahme der zumindest approximativ³ Normalverteilung für Zufallsvariablen, die sich additiv aus vielen zufälligen Einflüssen ähnlicher Größenordnung erklären lassen, ferner rechtfertigt es die Normalverteilungsapproximation für Schätz- und Teststatistiken der induktiven Statistik.

Entscheidend für den zentralen Grenzwertsatz ist, dass keine der unabhängigen Zufallsvariablen X_i die anderen deutlich dominiert. Dann konvergieren X_1, \dots, X_n unabhängig identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 approximativ gegen eine Standardnormalverteilung: $F_n(z) \rightarrow \Phi(z)$ oder $Z_n \stackrel{a}{\sim} N(0, 1)$ [10, 312].

Der zentrale Grenzwertsatz besagt, dass sich die Verteilung der Mittelwerte bei wachsendem Stichprobenumfang an eine Normalverteilung annähern und dabei die Varianz proportional zum Stichprobenumfang schrumpft. Dies kann anhand der Standardabweichung gezeigt werden.

Für die Standardabweichung gilt:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{\hat{\sigma}^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}^2)}{n(n-1)}} \quad (5.3)$$

Sie hängt damit von der Populationsvarianz $\frac{\sum_{i=1}^n (x_i - \bar{x}^2)}{(n-1)}$ und der Stichprobengröße n ab. Bei der Bildung von Konfidenzintervallen wird auch der zentrale Grenzwertsatz verwendet, da ab einer Größe von $n = 30$ 95 % der normalverteilten Zufallsgröße im Intervall von $[\mu - 1.96 \cdot \sigma_{\bar{x}}; \mu + 1.96 \cdot \sigma_{\bar{x}}]$ liegen.

5.3 Bedingte Wahrscheinlichkeiten

5.3.1 Bedingte Wahrscheinlichkeiten

Ziel der Sozialforschung ist es, Wahrscheinlichkeiten nicht nur bezogen auf gesamte Populationen, sondern auch auf Teilgesamtheiten zu erhalten [10, 199-203]. Vor allem werden wir ständig mit Wahrscheinlichkeiten über Teilpopulationen konfrontiert:

- 30% aller Autounfälle werden von Fahrern unter 25 Jahre verursacht. Mit der Population Unfallverursacher und der Bedingung Autofahrer jünger als 25 Jahre.
- die Arbeitslosenquote unter Akademikern beträgt 5%. Mit der Population Akademiker und der Bedingung arbeitslos.

Bewertet wird die Chance für das Eintreten eines Ereignisses A bei Kenntnis weiterer Informationen über ein bereits eingetretenes Ereignis B . So kann man sagen, dass die Unfallwahrscheinlichkeit mit dem Alter variiert und ebenso variiert die Wahrscheinlichkeit arbeitslos zu sein, mit dem Ausbildungsgrad.

³approximativ steht für asymptotisch oder auch durch $\stackrel{a}{\sim}$

Definition: Bedingte Wahrscheinlichkeit

Berechnet wird die Wahrscheinlichkeit für ein Ereignis A gegeben das Ereignis B oder auch die Wahrscheinlichkeit A unter (der Bedingung) B als der Quotient aus der Wahrscheinlichkeit für die Schnittmenge von A und B , also das gemeinsame Ereignis, und der Wahrscheinlichkeit für das bedingende Ereignis.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (5.4)$$

Allerdings meint Borovcnik, dass das Wort 'bedingend' irreführend ist, denn ein Ereignis kann eine Wahrscheinlichkeit auf unterschiedlichste Art beeinflussen: es kann begünstigen, benachteiligen, oder gänzlich irrelevant bzw. unabhängig sein [3, 199]. Empirisch bleibt daher zu bedenken, dass ein bedingendes Ereignis genauestens erkannt und definiert werden muss, dann erst kann die Wahrscheinlichkeitsaussage auf Grundlage der neuen Information neu bewertet werden.

Produktsatz

Der Produktsatz besagt, dass:

$$P(A \cap B) = P(A|B) \cdot P(B) \quad (5.5)$$

und analog:

$$P(A \cap B) = P(B|A) \cdot P(A). \quad (5.6)$$

Unabhängigkeit zweier Ereignisse

Man spricht von stochastischer Unabhängigkeit, wenn für die Wahrscheinlichkeit eines Ereignisses A , das Eintreten des Ereignisses B unbedeutend ist, dann ist:

$$P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)} \quad (5.7)$$

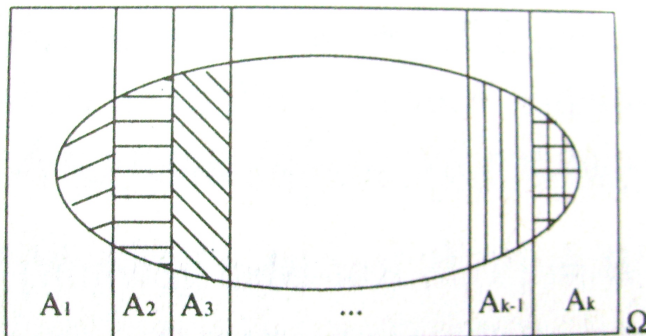
Daraus folgt, dass die Schnittmenge zweier unabhängiger Ereignisse (z. B. Ergebnisse eines Würfelwurfes: Die Wahrscheinlichkeit für das Paar zwei mal Eins (1,1) ist $P(A \cap B) = P(\{(1, 1)\}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(A) \cdot P(B)$), also

$$P(A \cap B) = P(A) \cdot P(B). \quad (5.8)$$

Totale Wahrscheinlichkeit

Betrachtet man einen Ereignisraum Ω als eine Vereinigung von Mengen A_1, \dots, A_k , so spricht man auch von einer disjunkten Zerlegung. Dabei müssen die Mengen paarweise disjunkt sein, sich also gegenseitig ausschließen, d.h. $A_i \cap A_j = \emptyset$. Beispielsweise kann man eine Population in Männer und Frauen einteilen oder aber Alterskohorten im 10-Jahresabstand bilden und beobachten wie häufig ein Ereignis B (Erwerbstätigkeit, Autounfall) innerhalb dieser Gruppen vorkommt. Die Wahrscheinlichkeit für das Ereignis B ist nun die Summe der Einzelwahrscheinlichkeiten pro Teilpopulation. Beispielsweise interessiert uns die Variation der Arbeitslosenquote bei Erwerbstätigen nach Bildungsgrad. Der Ereignisraum

sind alle Erwerbstätigen, die man in k (hier drei) disjunkte Kategorien (Akademiker, handwerkliche Ausbildung, sonstige) einteilt. Die Wahrscheinlichkeit, dass eine beliebige Person einer der gebildeten Kategorie angehört, muss bekannt sein. Somit kann man nun für jede der interessierenden drei Kategorien die Wahrscheinlichkeit der Arbeitslosigkeit bestimmen.



Graphik aus [10, 207]

Mathematisch lässt sich die Graphik wie folgt ausdrücken:

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k) \quad (5.9)$$

oder, da es sich um bedingte Wahrscheinlichkeiten handelt:

$$P(B) = P(B|A_1) \cdot P(A_1) + \dots + P(B|A_k) \cdot P(A_k). \quad (5.10)$$

Somit lässt sich der Satz der totalen Wahrscheinlichkeit definieren als:

$$P(B) = \sum_{i=1}^k P(B|A_i) \cdot P(A_i) \quad (5.11)$$

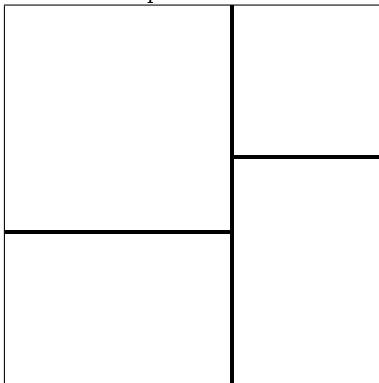
für eine A_1, \dots, A_k disjunkte Zerlegung von Ω .

Kapitel 6

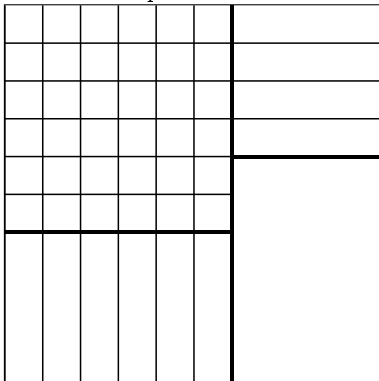
Graphische Darstellung bedingter Wahrscheinlichkeiten

Hierfür finden Sie die Unterlagen in einem separaten PDF-File. Prüfungsrelevant ist die Darstellung der Informationen in einem Einheitsquadrat, dem Satz von Bayes und dem Wahrscheinlichkeitsbaum mit absoluten Werten.

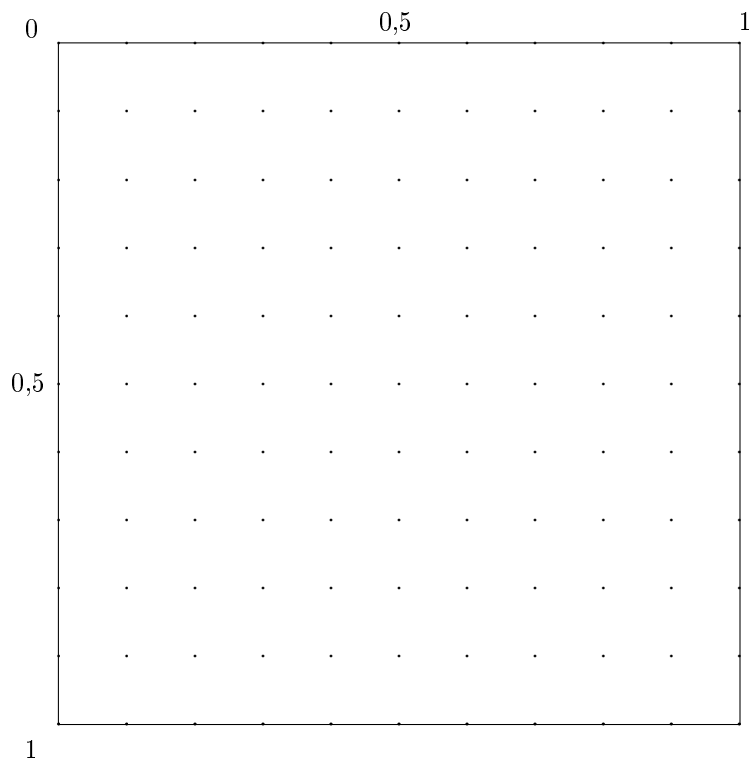
Einheitsquadrat mit Linien



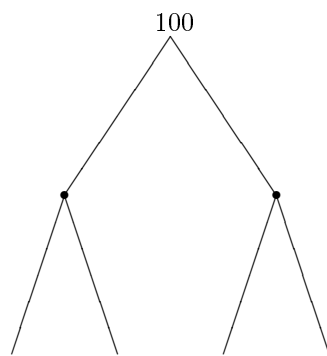
Einheitsquadrat mit unterschiedlich markierten Flächen



Vorlage Einheitsquadrat



Vorlage Wahrscheinlichkeitsbaum mit absoluten Werten:



Kapitel 7

Satz von Bayes

7.1 Quantifizierung von Unsicherheit

Wissenschaftliches Handeln bedeutet unsicheres Wissen in sicheres Wissen zu transformieren. Möglich ist die Transformation, da das Ausmaß der Unsicherheit bekannt ist und quantifiziert werden kann [20, 89]. Durch die Quantifizierung wissen wir, wie oft wir bei dem Befolgen einer bestimmten Regel Fehler begehen und daher haben wir ein Wissen über die Menge der Unsicherheit und können darauf aufbauend Entscheidungsregeln festlegen.

Für die Quantifizierung der Unsicherheit gibt es keine Master-Methode, denn die Ansätze sind umstritten.

Thomas Bayes (? – 1761) hat versucht die Unsicherheit durch eine *a priori* Verteilung zu quantifizieren. Die *a priori* Verteilung drückt den Glauben an verschiedenen Hypothesen vor einer Beobachtung der Daten aus und wird mit $P(A)$ bezeichnet. $P(b|A)$ wird das Wissen über die Wahrscheinlichkeitsverteilung der Daten (b), gegeben einer Hypothese (A) genannt. $P(b)$ wird die Randverteilung der beobachteten Daten genannt. Mit diesen Informationen lässt sich nun die bedingte Verteilung, gegeben der Daten, bestimmen [20, 92]:

$$P(A|b) = \frac{P(A) \cdot P(b|A)}{P(b)} \quad (7.1)$$

$P(A|b)$ wird auch als *aposteriori* Verteilung bezeichnet und drückt das subjektive Wissen nach einer Beobachtung aus. Demnach wird die Wahrscheinlichkeitsbewertung von A durch die Kenntnis von b begünstigt [2, 146]:

$$b \text{ begünstigt } A \iff P(A|b) > P(A) \quad (7.2)$$

Problematisch ist die induktive Komponente dieses Vorgehens, denn die *a priori* Verteilung ist subjektiver Natur.

7.1.1 Zusammenfassung des Satzes von Bayes

Die Bestandteile des Satzes von Bayes sind [2, 148-149]:

- A : Hypothese oder Annahme über mögliche Umweltzustände,
- b : Beobachtungsergebnisse,
- $P(A)$: apriori Wahrscheinlichkeit der Hypothese A ,
- $P(b|A)$: Likelihood der Hypothese A für die Beobachtung b ,
- $P(A|b)$: aposteriori Wahrscheinlichkeit der Hypothese A nach der Beobachtung b , wird auch bayesianische Umkehrwahrscheinlichkeit genannt.
- $P(b)$: Wahrscheinlichkeit der Beobachtung b , auch Randverteilung von b genannt.

Der Satz von Bayes lautet somit:

$$P(A|b) = \frac{P(b|A) \cdot P(A)}{P(b)} \quad (7.3)$$

Damit kann man mögliche Umweltzustände unter bestimmten subjektiven Annahmen wahrscheinlichkeitstheoretisch bestimmen.

7.2 Wahrscheinlichkeitsaussagen

Objektivistischer / frequentistischer Wahrscheinlichkeitsbegriff

Hinter dem Begriff der klassischer Statistik verbirgt sich ein, im 20. Jahrhundert dominanter, inferenzstatistischer Ansatz. Personen wie Fisher, Neyman und Pearson lassen sich damit verbinden sowie die Verfahren der Maximum-Likelihood-Schätzung, der Signifikanztest und Konfidenzintervalle. Bei der Inferenz spielt die "Wahrscheinlichkeitsanalyse" der Stichprobenstatistik¹ eine zentrale Rolle. Es wird die "Wahrscheinlichkeit der beobachteten Stichprobenstatistik bestimmt" und damit beurteilt ob ein Schluss auf die Grundgesamtheit möglich ist. Objektivistische Wahrscheinlichkeiten beruhen auf einem Inferenzkonzept, das unabhängig von apriori-Einschätzungen ist. Die Hauptvertreter der modernen Statistik sind K. Pearson (1847-1936), R. A. Fischer (1890-1962), Jerzy Neyman (1894-1981), Egon S. Pearson (1895-1980) und Abraham Wald (1902-1950).

Uns interessieren derzeit wie Wahrscheinlichkeitsaussagen für das Eintreten eines Ereignisses A gewonnen werden. Vergleiche im Folgenden auch [10, 179ff]. In der Statistik wird für Wahrscheinlichkeit eines Ereignisses A die Notation $P(A)$ verwendet, wobei P für 'probability'² steht. Die Axiomen von Kolmogorov definieren die Eigenschaften und Rechenregeln von Wahrscheinlichkeiten, wobei nicht gesagt wird, was Wahrscheinlichkeiten sind³.

Wie gezeigt werden konnte, lässt sich zwischen relativen Häufigkeiten und Wahrscheinlichkeiten eine Verbindung herstellen, die auf einer Häufigkeitsinterpretation – und nicht individuellen Interpretation – der Wahrscheinlichkeit beruht. Ausgangsidee ist ein Zufallsexperiment, das beliebig oft wiederholt werden kann. "Ermittelt man nun bei n unabhängigen Wiederholungen dieses Experiments jeweils die relative Häufigkeit eines Ereignisses A , so lässt sich die Wahrscheinlichkeit für das Eintreten von A als den Wert ansehen, bei dem sich die relative Häufigkeit von A mit wachsendem n stabilisiert." [10, 180]

¹Die Charakteristika der Stichprobendaten

²aus dem Lateinischen: probare, to prove (testen, prüfen) und ability, also überprüfbar

³vgl.[10, 180, 182]

Diese Auffassung wird auch als objektivistischer Wahrscheinlichkeitsbegriff bezeichnet.

Inferenzstatistische Ansätze

Das Wahrscheinlichkeitsverständnis beruht auf der frequentistischen Annahme, dass die Wahrscheinlichkeit eines Ereignisses von der relativen Häufigkeit bei unendlicher Wiederholung des Wahrscheinlichkeitsprozesses, der zu diesem Ereignis führt, abhängt. Hypothetisch wird also ein Ereignis – wie eine Stichprobe – unendlich oft wiederholt und der jeweilige Mittelwert bestimmt. Die Proportion dieser Ziehungen führt schließlich zu der Wahrscheinlichkeit eines Ereignisses. Voraussetzung sind natürlich wiederholbare Ereignisse wie die Stichprobenziehung aus einer Population – was bei kleinen Datensätzen oder kleinen Grundgesamtheiten nicht immer möglich ist.

Das Konzept lässt sich formalisieren, wobei θ ein hypothetischer Populationsparameter ist und X den beobachteten Daten entspricht. Untersucht wird nun die Wahrscheinlichkeit, dass die Daten bei gegebenem Populationsparameter auftreten. Es handelt sich um eine konditionale Datenwahrscheinlichkeit $P(X|\theta)$. Bei der Maximum-Likelihood-Methode wird nun der Parameter θ bestimmt, der die Datenwahrscheinlichkeit $P(X|\theta)$ maximiert. Für θ wird keine Wahrscheinlichkeit bestimmt, sondern Hypothesen über Werte von θ geprüft. Diese Hypothesen werden angenommen oder verworfen, je nachdem, mit welcher Wahrscheinlichkeit die beobachteten Daten X unter Annahme von θ beobachtet werden können.

subjektivistischer Wahrscheinlichkeitsbegriff

(!) Dem objektivistischen (dem Objekt zugehörigem) Wahrscheinlichkeitsbegriff steht der subjektivistische (dem Subjekt zugehörige) Wahrscheinlichkeitsbegriff gegenüber. Bei dem subjektiven Wahrscheinlichkeitsbegriff ist die Wahrscheinlichkeit für das Eintreten eines Ereignisses A abhängig vom Glauben der jeweiligen Person an das Eintreten und kann pro Person variieren und ist nicht eine Eigenschaft des Objektes. Die Quantifizierung dieser Wahrscheinlichkeit geschieht durch eine *imaginäre Wette* auf das Eintreten des Ereignisses A . Der Wettquotient ist das Verhältnis von Einsatz und Gewinn zu dem eine Person bereit ist auf das Eintreten des Ereignisses A zu wetten.

Die Vertreter der subjektivistischen Wahrscheinlichkeit sind Bruno de Finetti (1906 – 1986) und Leonhard J. Savage (1917 – 1971), die sich bemüht haben, dass die subjektivistischen Wahrscheinlichkeiten mit den Axiomen von Kolmogorov konsistent sind, daher wird mit von einem idealtypischen rationalen Subjekt ausgegangen, das nach definierten Rationalitätskriterien handelt. De Finetti ist der Auffassung, dass eine subjektive Meinung objektiv dargestellt werden kann und sich daher als Forschungsgegenstand anbietet. Dadurch können nun Informationen abgebildet werden, die nicht gleichwahrscheinlich und symmetrischer Natur sind, sondern vom Wissensstand des Beobachters abhängen.

(!) Der subjektiven Auffassung liegt ein *prinzipielles Unwissen über die Umweltzustände* zugrunde, was aber nicht an einer den Ereignissen innewohnenden Zufälligkeit liegt, sondern an der Beschränktheit der Menschen bei der Bestimmung von Gesetzmäßigkeiten. Die Gesetzmäßigkeiten sind nicht probabilisti-

scher Natur, sondern *deterministisch* bestimmt. Zukünftige Ereignisse sind also daher unsicher, weil ein Subjekt nicht genug Informationen über das Eintreten des Ereignisses hat und daher als Wahrscheinlichkeitseinschätzung eine Wette über mögliche Ausgänge abschließen muss. Mit zusätzlichem Wissen, wird die Aussage genauer.

Diskussion des subjektivistischen und objektivistischen Wahrscheinlichkeitsbegriffs

Unter den Statistikern und den Wissenschaftstheoretikern gibt es keine einheitliche Meinung über die beste Methode.

Problematisch an der frequentistischen/objektivistischen Auffassung von Wahrscheinlichkeiten ist *Gleichwahrscheinlichkeit symmetrischer Situationen* und die *Begründung der Wahrscheinlichkeit in der Unendlichkeit*. Ebenso spricht man von einer logischen Wahrscheinlichkeit, da die statistischen Wahrscheinlichkeit als der Grenzwert der relativen Häufigkeiten logisch definiert wird, aus der Theorie begründet wird und auf das Gesetz der großen Zahl zurückzuführen ist. Allerdings ist es aber nicht gesagt, dass es hierfür auch eine tatsächliche Notwendigkeit gibt.

Prominente Kritiker sind Wolfgang Stegmüller [21] [22] und Rudolf Carnap [7]. Diese weisen auf die Notwendigkeit einer subjektive Bewertung aller statistischen Verfahren hin.

Zusammenfassend die beiden großen Schulen der Statistik:

objektiv	subjektiv
objektbezogen Glücksspiel (Würfel)	subjektbezogen wahrscheinlich (Betrachter) Gott würfelt nicht
alleatorisch	epistemisch
Eintretensneigung	Grad des Dafürhaltens
Zufall	Unsicherheit (deterministisches Weltbild)
Häufigkeitsinterpretation	Wettinterpretation
vollkommene Unsicherheit über die Umweltzustände	perfektes probabilistisches Wissen über die Umweltzustände
Auswertung der Stichprobe mit im Durchschnitt erfolgreichen Verfahren	Aufdatieren von subjektivem apriori Wissen; Dazulernen nach dem Satz von Bayes

Dies ist ein grober Überblick über die Unterschiede zweier Schulen in der Statistik. Im Folgenden werden nun die Konsequenzen für die Anwendung der Wahrscheinlichkeitsrechnung erläutert.

7.3 Was heißt Bayes Statistik?

7.3.1 Wiederholung Inferenzstatistik

Die Aufgabe der Inferenzstatistik ist es Schlüsse von einer Stichprobe auf eine Grundgesamtheit zu ziehen. Entscheidend hierbei ist die Wahrscheinlichkeitsanalyse einer Stichprobenstatistik. Die Stichprobenstatistik (z. B. Mittelwert

oder der Standardfehler ist dann eine Maß für die Streuung der Stichprobenstatistik) ist eine *charakteristische Zusammenfassung der Stichprobendaten*. Die Stichprobenstatistik ist variabel und ist von der zu untersuchenden Zufallsvariable abhängig und ist daher selbst eine Zufallsvariable. Ist die Verteilung in der Grundgesamtheit bekannt, so kann die Verteilung der Stichprobenstatistik exakt abgebildet werden, anderenfalls können mit Hilfe des zentralen Grenzwertsatzes *approximative Aussagen* getroffen werden und anschließend der Standardfehler der Schätzung (nicht der Stichprobenstatistik) und Konfidenzintervalle berechnet werden.

Anschließend wird die Hypothese aufgestellt, dass die Stichprobenstatistik (z. B. \bar{X}) dem Populationsparameter μ (Populationsmittelwert)⁴ gleicht. Populationsparameter sind in der Regel unbekannt und werden aus der Stichprobenstatistik geschätzt. Dabei wird die *Wahrscheinlichkeit der beobachteten Stichprobenstatistik* bestimmt und dann der Schluss auf die Grundgesamtheit gezogen.

Als Beispiel nennt Broscheid [6, 45] den T-Test, der die Differenz zwischen Durchschnitten in zwei Bevölkerungen bestimmt. Die Hypothese (die Nullhypothese) lautet, dass die Differenz zwischen den Bevölkerungen Null ist und die in der Stichprobe beobachtete Differenz (die beobachtete Stichprobenstatistik) das Resultat der Zufallsstichprobenziehung.

Anschließend wird angenommen dass die Nullhypothese wahr ist. Dann wird die Wahrscheinlichkeit bestimmt, mit der eine Zufallsstichprobe zu einer Differenz zwischen den Durchschnitten führt, die gleich der beobachteten ist oder größer. Ist die Wahrscheinlichkeit gering, so wird die Nullhypothese verworfen (der Richtwert ist meist 0.05) und die Unterschiede als signifikant – also überzufällig – bezeichnet.

In der klassischen statistischen Inferenz kann man die Datenwahrscheinlichkeit $P(X)$ der beobachteten Daten bzw. der Stichprobenstatistik, als eine konditionale Datenwahrscheinlichkeit darstellen, die durch die Annahme hypothetischer Populationsparameter θ bedingt sind, also $P(X|\theta)$. In der Maximum-Likelihood-Analyse wird der Populationsparameter θ bestimmt, der die konditionale Datenwahrscheinlichkeit $P(X|\theta)$ maximiert. Unterschiedliche Werte von θ werden akzeptiert, aufgrund der oben beschriebenen Wahrscheinlichkeit mit der die beobachteten Daten unter Annahme von θ beobachtet werden können. Über den Populationsparameter θ selbst werden keine Wahrscheinlichkeitsaussagen gemacht! Eigenes Vorwissen über die Wahrscheinlichkeit eines Populationsparameteres fließen in die Analyse nicht mit ein, es werden einzig die beobachteten Daten mit ihren charakteristischen Merkmalen und die Wahrscheinlichkeit dieser in einer Grundgesamtheit beurteilt. Wie schon oben gezeigt wurde entspricht die Wahrscheinlichkeit eines bestimmten Stichprobenmittelwertes der Proportion dieses Mittelwertes in einer hypothetischen, unendlichen Reihe von Stichprobenziehungen und Mittelwertsmessungen. Geeignet ist die frequentistische Analyse dann, wenn es sich um Stichprobenziehungen handelt, aber nicht wenn es sich um einmalige oder seltene Ereignisse handelt.

⁴Stichprobenstatistik und ihre Kennwerte der zentralen Tendenz oder der Streuung werden durch lateinische Buchstaben gekennzeichnet, wohingegen Populationsparameter durch griechische Buchstaben ausgedrückt werden.

7.3.2 Bayesianische Ansätze

Folgendes Kapitel bezieht sich auf den Artikel von Andreas Broscheid [6].

(!) Bayesianische Ansätze sind Alternativen zu bekannten statistischen Inferenzverfahren. Broscheid spricht nicht von der "Methode" sondern von "Ansätzen", denn es gibt mehrere bayesianische Modelle. Allen gemein ist, dass sie mit Hilfe der *bayesianischen Umkehrwahrscheinlichkeit* geschätzt werden.

(!) Die Kenntnis über die Datenwahrscheinlichkeit $P(X|\theta)$ ist das Ziel der klassischen Inferenzstatistik. Das Ziel der bayesianischen Statistik ist die Beurteilung der Populationparameter durch eine Wahrscheinlichkeitsverteilung $P(\theta|X)$. Es geht darum in der Analyse $P(\theta|X)$ zu ermitteln. Broscheid verwendet folgende Darstellung der bayesianischen Umkehrwahrscheinlichkeit:

$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{\int P(X|\theta) \cdot P(\theta) d\theta} \quad (7.4)$$

Die a priori Wahrscheinlichkeit $p(\theta)$ drückt das persönliche Wissen oder Unwissen (Literaturrecherche, Ergebnisse vorhergehender Untersuchungen) über den Forschungsgegenstand aus und muss nicht die relative Häufigkeit sein. Vielmehr handelt es sich um die subjektive Unsicherheit über das Geschehen oder Vorkommen eines Ereignisses, das man in der Form von Wahrscheinlichkeiten ausdrücken kann.

(!) Der bayesianische Ansatz ersetzt die Annahme der Asymptotik durch *explizite Annahmen über die Parameterverteilung*. Der Vorteil dieses Begriffes ist, dass er flexibler eingesetzt werden kann als der klassische Wahrscheinlichkeitsbegriff, andererseits können die Ergebnisse *stark verzerrt* werden, wenn die subjektive Einschätzung so stark ist, dass die beobachteten Daten keinerlei Einfluss mehr auf das Endergebnis der posteriori Wahrscheinlichkeit haben.

Bayesianische Ansätze eignen sich, um Nicht-Zufallsstichproben zu analysieren. Frequentistische Ansätze sind hierfür ungeeignet, da "die Wahrscheinlichkeit eines Ereignisses, als dessen relative Häufigkeit bei unendlicher Wiederholung des Prozesses, der zu dem Ereignis führt", definiert wird [6, 49]. *Nichtzufallsstichproben* sind nicht wiederholbar und daher sind Wahrscheinlichkeitsaussagen über Nichtzufallsstichproben bedeutungslos – wenn man es streng betrachtet. Aus bayesianischer Sicht ist dies möglich, da die *Wahrscheinlichkeitsaussage nicht auf der Wiederholbarkeit der zugrundeliegenden stochastischen Prozesse beruht*, was erleichtert wird, wenn für eine Analyse die Beobachtungen unabhängig und gleichförmig verteilt sind.⁵

7.3.3 Noch ein genauerer Blick auf bayesianische Verfahren

Die Maximum-Likelihood-Methode nimmt die beobachteten Daten und schätzt daraus die unbekanntes Populationparameter, indem bei beobachteten Daten der wahrscheinlichste Wert bestimmt wird. Alle Information zur Schätzung des unbekanntes Parameters ist demnach in der Likelihood-Funktion enthalten. Bayesianischer Verfahren verwenden ebenso die Likelihood-Funktion, so dass Gill [13, 62] meint, dass jedes Likelihood-Modell ein Bayes-Modell ist, welches

⁵Gleichverteilung: es gibt keine Präferenz für eines der Ereignisse; im diskreten Fall tritt jeder mögliche Zustand mit der gleichen Wahrscheinlichkeit ein, im stetigen Fall ist die Dichte konstant.

uninformative priori-Wahrscheinlichkeiten verwendet. Wir können festhalten, dass die posteriori-Wahrscheinlichkeit ein Kompromiss aus den Daten (Likelihood) und dem Vorwissen (priori) ist, dies kann je nach Stärke des Vorwissens und nach Anzahl der Daten variieren.

Anbei nochmals eine Zusammenfassung von Gill über die Hauptunterschiede bayesianischer und frequentistischer Ansätze:

	Wahrscheinlichkeitsinterpretation
objektiv:	Beobachtung von einer unendlichen Serie identischer Wiederholungen. Wahrscheinlichkeit wird mit $P(data \theta)$ bezeichnet wobei θ eine Hypothese meist die Nullhypothese H_0 ist.
subjektiv:	Wahrscheinlichkeit ist der Grad des Vertrauens des Beobachters vor und nach der Beobachtung der Daten, also $p(\theta data)$.
	Was ist fix und was ist variabel?
objektiv:	Daten sind eine variable iid-(unabhängig und identisch verteilte) Zufallsauswahl, wobei die Parameter der Grundgesamtheit als fix (dem Objekt zugehörig) angenommen werden.
subjektiv:	Die Daten sind eine fixe Auswahl aus der Grundgesamtheit, wobei die Parameter variabel und unbekannt sind und durch Verteilungen beschrieben werden können.
	Zusammenfassung der Ergebnisse
objektiv:	Punktschätzung und Standardfehler. Das 95% Konfidenzintervall besagt, dass in 19 von 20 Fällen das Intervall den wahren Wert im Durchschnitt beinhaltet.
subjektiv:	Posteriori-Wahrscheinlichkeiten werden durch den Mittelwert und die Quantile beschrieben. Das Highest-Posteriori-Density-(HDI)Intervall zeigt das Intervall mit der höchsten posteriori-Wahrscheinlichkeit.
	Inferenz-Modell
objektiv:	Deduktion von $P(data H_0)$ mit festgelegtem α . H_0 wird abgelehnt, wenn $P(data H_0) < \alpha$, H_0 wird angenommen, wenn $P(data H_0) \geq \alpha$.
subjektiv:	Induktion von $P(\theta data)$, wobei mit $P(\theta)$ begonnen wird. $100(1 - \alpha)\%$ des höchsten Wahrscheinlichkeitsniveaus im $1 - \alpha$ HPD.
	Qualitätschecks
objektiv:	Berechnung der Fehler 1. und 2. Art. Manchmal wird auch der Power bestimmt (Robustheit der Ergebnisse). In der Regel finden die p -Werte Beachtung.
subjektiv:	Posteriori Predictive Checks. Sensitivitätsanalysen, Bayes-Faktor

Beide Ansätze unterscheiden sich oftmals nicht sehr stark, da der zentrale Grenzwert und die Normalverteilungsannahme von beiden Ansätzen verwendet wird.

Handelt es sich um Aussagen über kleine Populationen finden bayesianische Ansätze eine bessere Grundlage, da sie Begründung in der Unendlichkeit nicht benötigen. Der Hauptvorwurf an die bayesianische Sichtweise ist, dass die Priori-Annahmen einen großen Einfluss auf kleine Stichproben haben. Daher werden in der Statistik *robuste Schätzer* entwickelt, die gegenüber Prioris robuster und nicht so sensibel reagieren.

Kapitel 8

Bayes Verfahren

8.1 Einleitung

Zu Beginn des Kapitels lohnt nochmals der Blick zurück. In dem Krankheitsbeispiel ist A der Zustand 'krank', b ist ein 'positives Testergebnis'. Die Informationen werden mit dem Satz von Bayes ausgewertet.

$$P(A|b) = \frac{P(b|A) \cdot P(A)}{P(b)} \quad (8.1)$$

Wie unterscheiden sich nun $P(b|A)$ und $P(A|b)$ und welche fortführenden Schlüsse können gezogen werden?

$P(b|A)$ beschreibt ein Ergebnis, bei vollem Kenntnisstand über A . Die Aussage: wenn jemand krank ist, hat er ein positives Testergebnis, kann als allgemeine Gesetzmäßigkeit verstanden werden und mit einer Wahrscheinlichkeit (z. B. 80%) angegeben werden.

$P(A|b)$ beschreibt ein Ergebnis über A , nachdem b beobachtet wurde: die Wahrscheinlichkeit, wenn ein positives Ergebnis vorliegt, dass eine Person krank ist. A ist also unbekannt und erst nach der Beobachtung wird die Wahrscheinlichkeit von A beurteilt.

Man kann auch von A als einem unbekanntem Parameter θ ("Theta") sprechen und von b als einer Beobachtung in Form einer Stichprobe.

Ziel ist es mehr über den unbekanntem Parameter zu lernen. Da im bayesianischen Ansatz der Parameter θ als eine unbekannte Zufallsgröße betrachtet wird kann Vorwissen über θ in die Analyse mit einfließen.

(!) Kurz gesagt ist der *bayesianische Ansatz* eine Kombination aus Vorwissen und beobachteten Daten, die zu genauem Wissen über einen Parameter führt:

$f(\theta)$	Priori-Wissen
$f(X \theta)$	Stichprobenmodell
$f(\theta X)$	Posteriori-Wissen: Neues Wissen über Parameter θ

$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{\int P(X|\theta) \cdot P(\theta) d\theta} \quad (8.2)$$

8.2 Die aposteriori Verteilung

Stichproben verändern die Information, daher wird die Stichprobe mit dem Vorwissen gewichtet, so dass man ein aposteriori Wissen erhält.

Da das wahre θ zufällig ist – weil unbekannt – kann man nur eine Wahrscheinlichkeitsverteilung für den Parameter angeben. Die Dichte von θ wird mit π (gesprochen "Pi"), also $\pi(\theta)$ ("Dichte Pi von Theta") angegeben.

Die Funktion $f_\theta(x)$ wird bei Bayes auch als die bedingte Dichte $f(x|\theta)$ ("die Dichte von X wenn Theta gegeben") oder als Likelihood bezeichnet.

8.2.1 Definition der aposteriori Verteilung

Gegeben ist die Dichte $f_\theta(\cdot)$ und ein apriori Verteilung mit der Dichte $f(\cdot)$. Dann heisst für jedes $x \in X$ die aposteriori Verteilung von θ gegen x

$$f(\theta|x) = c(x) \cdot f(x|\theta) \cdot f(\theta) \quad (8.3)$$

Die aposteriori Verteilung von θ gegeben x besitzt eine Normierungskonstante $c(x)$.

(!) Die formale Schreibweise mit Erklärung:

$$\underbrace{f(\theta|x)}_{\text{aposteriori Vtlg.}} = \underbrace{c(x)}_{\text{Normierungskonstante}} \cdot \underbrace{f(x|\theta)}_{\text{Stichprobenmodell}} \cdot \underbrace{f(\theta)}_{\text{Vorwissen}}$$

Bayes Postulat

(!) Das *Bayes Postulat* besagt, dass nach der Beobachtung der Stichprobe die aposteriori Verteilung die volle Information enthält. Sie beschreibt also das Wissen über den Parameter vollständig. Anschließende Analysen haben sich auf die aposteriori zu stützen.

8.3 Konjugierte Verteilungen, Bayes Lernen

Konjugiertheit

Eine Verteilungsfamilie Π ("groß Pi") von apriori Verteilungen heißt zu einer Menge P ("groß Rho") von Stichprobenverteilungen konjugiert, wenn für jede priori $\pi(\cdot) \in \Pi$ und jedes $p(\cdot) \in P$ die zugehörige aposteriori Verteilung wieder ein Element von Π ist. Man sagt dann auch, dass jedes Element von $\pi(\cdot) \in \Pi$ zu P konjugiert ist.

(Priori + Stichprobenverteilung \rightarrow Posteriori).

(!) Als Beispiel ließen sich anführen:

Priori	Stichprobenverteilung	\rightarrow	Posteriori
Beta	Binominal/Bernoulli	\rightarrow	Beta
Gamma	Poisson	\rightarrow	Gamma
Normal	Normal	\rightarrow	Normal

8.3.1 Das Beta-Binomial Modell

(!) *Prüfungsrelevant ist die Herleitung der Posteriori Verteilung aus der Stichprobenverteilung und der Priori-Verteilung.*

Ausgangssituation ist ein Stichprobenmodell mit einer Bernoulli (Binomial) Verteilung: $X \sim \text{Bi}(n, \pi)$; dabei ist n bekannt, π ist unbekannt und wird apriori geschätzt. Als apriori Verteilung für π gilt eine Beta Verteilung: $\pi \sim \text{Be}(\alpha, \beta)$; dabei sind die Parameter $\alpha, \beta > 0$ und bekannt. Gemäß dem Bayes Postulat müsste die aposteriori Verteilung – nach der Berechnung der apriori Verteilung mit der Beobachtung – wiederum eine Beta Verteilung sein.

Ausgangssituation ist ein Stichprobenmodell (Bernoulli Verteilung) zum Parameter π (Die Wahrscheinlichkeit ja = π , die Wahrscheinlichkeit nein = $(1 - \pi)$). Die Wahrscheinlichkeit π ist: $p(x|\pi) = \pi^x(1 - \pi)^{1-x}$ mit $x \in \{0, 1\}$. und der Verteilung:

$$f(x|\pi) = \binom{n}{x} \pi^x(1 - \pi)^{n-x} \text{ mit } 0 < \pi < 1 \quad (8.4)$$

Als apriori Verteilung dient eine Beta Verteilung.

$$f(\pi) = \frac{1}{B(\alpha, \beta)} \pi^{\alpha-1}(1 - \pi)^{\beta-1} \text{ mit } 0 < \pi < 1 \quad (8.5)$$

Falls $\alpha = \beta = 1$ so handelt es sich um eine Gleichverteilung auf $[0, 1]$. Für die vorhandenen Informationen (apriori Verteilung und Stichprobe) wird der Satz von Bayes angewendet, der zu der allgemeinen Form

$$f(\pi|x) = c(x) \cdot f(x|\pi) \cdot f(\pi)$$

führt.

$$\begin{aligned} f(\pi|x) &= \binom{n}{x} \cdot \pi^x(1 - \pi)^{n-x} \cdot \frac{1}{B(\alpha, \beta)} \cdot \pi^{\alpha-1}(1 - \pi)^{\beta-1} \\ &\propto \pi^{x+\alpha-1}(1 - \pi)^{n-x+\beta-1} \\ &\propto \pi^{\alpha'-1}(1 - \pi)^{\beta'-1} \end{aligned} \quad (8.6)$$

Die aposteriori Verteilung ist wiederum eine Beta Verteilung, aber mit den Parametern $\alpha' = \alpha + x$ und $\beta' = \beta + n - x$ also $\pi|x \sim \text{Be}(\alpha + x, \beta + n - x)$

Dabei sind $\binom{n}{x}$ und $\frac{1}{B(\alpha, \beta)}$ konstant und unabhängig von π und können für die weitere Bestimmung des Parameters vernachlässigt werden, werden aber – vereinfacht dargestellt – in der Normierungskonstante $c(x)$ berücksichtigt. Die Posteriori-Verteilung ist daher nicht identisch (=) sondern bis auf eine Normierungskonstante *proportional* (\propto) zur Stichprobenverteilung und Priori-Verteilung.

8.3.2 Gamma-Poisson-Modell

(!) *Prüfungsrelevant ist die Herleitung der Posteriori Verteilung aus der Stichprobenverteilung und der Priori-Verteilung.*

Die Poisson-Verteilung beschreibt das Auftreten seltener Ereignisse. Gegeben ist eine Stichprobe X_1, \dots, X_n eines zum Parameter λ poissonverteilten Merkmales (beispielsweise erzielte Tore in einer Fussballsaison oder Spielerkarriere).

Die Poisson-Stichprobenverteilung lautet:

$$\prod_i f(x|\lambda) = \frac{1}{\prod_{i=1}^n x_i!} \cdot \lambda^{\sum x_i} \cdot \exp^{-n \cdot \lambda} \quad (8.7)$$

Die Gamma Verteilung der Priori lautet:

$$\pi(\lambda) = \frac{b^a}{\Gamma(a)} \cdot \lambda^{a-1} \cdot \exp^{-b \cdot \lambda} \quad (8.8)$$

Gesucht ist wiederum:

$$\pi(\lambda|x) = c(x) \cdot f(x|\lambda) \cdot \pi(\lambda)$$

wobei wiederum vereinfachend alle Komponenten, die unabhängig von λ sind ($\frac{1}{\prod_{i=1}^n x_i!}$ und $\frac{b^a}{\Gamma(a)}$) in einer Normierungskonstante zusammengefasst werden,

$$\begin{aligned} \pi(\lambda|x) &= \frac{1}{\prod_{i=1}^n x_i!} \cdot \lambda^{\sum x_i} \cdot \exp^{-n \cdot \lambda} \cdot \frac{b^a}{\Gamma(a)} \cdot \lambda^{a-1} \cdot \exp^{-b \cdot \lambda} \\ &\propto \lambda^{\sum x_i + a - 1} \exp^{-(b+n) \cdot \lambda} \\ &\propto \lambda^{a' - 1} \cdot \exp^{-b' \cdot \lambda} \end{aligned} \quad (8.9)$$

mit $a' = \sum x_i + a$ und $b' = (n + b)$. Die Posteriori-Verteilung ist damit eine Gamma-Verteilung zzgl. Informationen aus den Daten.

8.3.3 Selbstkonjugiertheit der Normalverteilung

(!) *Prüfungsrelevant ist nicht die Herleitung der Posteriori Verteilung, allerdings die Interpretation der Gewichtungsfaktoren für den unbekannt Parameter.*

Hat man als Stichprobenverteilung eine Normalverteilung, so kann man als Priori-Verteilung eine Normalverteilung verwenden, wodurch man als Posteriori-Verteilung wiederum eine Normalverteilung erhält. Dies nennt man die Selbstkonjugiertheit der Normalverteilung.

Gegeben ist eine i.i.d. Stichprobe einer normalverteilten Zufallsvariable X_1, \dots, X_n mit bekannter Varianz σ^2 und unbekanntem Parameter μ . Als Priori-Verteilung für den unbekannt Parameter μ dient eine Normalverteilung mit den Parametern ν "Nü" und ϕ^2 "Phi-Quadrat", so ist die posteriori-Verteilung $\pi(\mu|X)$ eine Normalverteilung mit den Parametern ν' "Nü-Strich" und ϕ'^2 "Phi-Quadrat-Strich".

Es kann bewiesen werden – was wir nicht tun –, dass:

$$\begin{aligned}
\nu' &= \frac{\bar{x}\phi^2 + \nu\frac{\sigma^2}{n}}{\phi^2 + \frac{\sigma^2}{n}} \\
&= \underbrace{\bar{x} \cdot \frac{\phi^2}{\phi^2 + \frac{\sigma^2}{n}}}_{\text{beobachtete Tatbestände}} + \underbrace{\nu \cdot \frac{\frac{\sigma^2}{n}}{\phi^2 + \frac{\sigma^2}{n}}}_{\text{Vorvermutung}} \quad (8.10)
\end{aligned}$$

Man kann sagen, dass es sich bei dem Term mit dem \bar{x} um die beobachteten Tatbestände handelt und bei dem Term mit dem ν um die Vorvermutung.

(!) Da der unbekannte Parameter ν' (oder auch $\hat{\mu}$) nach der Beobachtung die volle Information besitzt, kann er als der HPD (highest-posteriori-density) Schätzer betrachtet werden. Der Schätzer variiert je nach Zusammenspiel von Vorwissen und Beobachtung hinsichtlich drei Komponenten: der Fallzahl der Beobachtung n , der Varianz der Stichprobe σ^2 und hinsichtlich der Varianz des Vorwissens ϕ^2 .

n n groß \rightarrow \bar{x} (Stichprobe) viel Gewicht
 n klein \rightarrow ν (Vorvermutung) viel Gewicht

σ^2 σ^2 groß \rightarrow ν viel Gewicht
 σ^2 klein \rightarrow \bar{x} viel Gewicht

ϕ^2 ϕ^2 groß \rightarrow \bar{x} viel Gewicht
 ϕ^2 klein \rightarrow ν viel Gewicht

Kapitel 9

Ergänzung: Bayesianische Statistik

(!) Dieses Kapitel ist in folgenden Themen prüfungsrelevant: Berechnung Erwartungswert und Modus; Gewichtung posteriori Erwartungswert; Tabellen mit Unterschiede der Glaubwürdigkeitsregionen (Formeln werden angegeben).

Die Beta-Verteilung hat die Form:

$$B(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (9.1)$$

mit $B(\alpha, \beta)^{-1}$ als Normierungskonstanten der Verteilung.
Der Wertebereich ist $\alpha, \beta > 0$,
der Erwartungswert

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

der Modus ist

$$Mod(X) = \frac{\alpha - 1}{\alpha + \beta - 2}$$

und die Varianz ist

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Für $\alpha = \beta = 1$ erhält man eine stetige Gleichverteilung auf dem Intervall $(0, 1)$.

Für die Posteri-Verteilung einer Beta-Verteilung ist der Erwartungswert:

$$E(\pi|x) = \frac{\alpha + x}{\alpha + \beta + n}$$

Der Modus der Verteilung (der häufigste Wert und damit das Maximum der Verteilung) ist:

$$Modus(\pi|x) = \frac{\alpha + x - 1}{\alpha + \beta + n - 2}$$

Bei einer Beta Gleichverteilung als Prior ($\alpha = \beta = 1$) entspricht der Modus dem *Maximum der Stichprobenverteilung* $\bar{x} = \frac{x}{n}$ – auch Maximum-Likelihood-Schätzwert (ML-Schätzer) genannt.

Dies wird umso ersichtlicher, wenn man bedenkt, dass der posteriori-Erwartungswert ein *gewichtetes Mittel* des priori-Erwartungswertes $\frac{\alpha}{\alpha+\beta}$ und des Stichprobenschätzers (ML-Schätzer) \bar{x} ist. Die Gewichtung ist proportional zu $(\alpha+\beta)$ und zu n , so dass mit steigendem Stichprobenumfang n das Gewicht des ML-Schätzers \bar{x} erhöht wird. Dies kann durch Umformung der posteriori-Verteilung gezeigt werden:

$$\begin{aligned} E(\pi|x) &= \frac{\alpha + x}{\alpha + \beta + n} \\ &= \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \frac{\alpha}{\alpha + \beta} + \frac{n}{\alpha + \beta + n} \cdot \frac{x}{n} \end{aligned}$$

Siehe im Folgenden die Tabelle von [14, 143]. Sie verdeutlicht, wie sich mit steigendem Stichprobenumfang die Kennwerte der posteriori-Verteilung der Stichprobenverteilung annähern.

(!) Die posteriori-Verteilung ist eine Kombination aus einer Priori Betaverteilung $\pi \sim Be(3, 2)$ und einer Stichproben Binominalverteilung $\pi \sim Bi(n, \pi)$.

n	x	Erwartungswert	Modus	Median	2.5% Quantil	97.5% Quantil
10	0	0.08	0.0	0.06	0.00	0.28
10	1	0.17	0.1	0.15	0.02	0.41
10	5	0.50	0.5	0.50	0.23	0.77
100	0	0.01	0.0	0.01	0.00	0.04
100	10	0.11	0.1	0.11	0.06	0.17
100	50	0.50	0.5	0.50	0.40	0.60

(!) Der Erwartungswert ist im Falle $n = 10$ und $x = 0$ oder $x = 1$ größer als bei $n = 100$ und $x = 0$ oder $x = 10$. Dies hängt damit zusammen, dass es sich bei der aposteriori Verteilung um einen *gewichteten Wert* aus Vorwissen und Stichprobenmodell handelt. Mit steigender Stichprobengröße verliert somit das Vorwissen an Gewicht. Der Modus ist sowohl für 10 als auch für 100 Fälle identisch, allerdings verändern sich die Intervallgrenzen, so dass die Schätzung bei mehr Beobachtungen n genauer wird.

Die 2.5% und 97.5% Quantile (beiden linken Spalten) können mit den Konfidenzintervallen (KI) der Stichprobenverteilung verglichen werden (beiden rechten Spalten). Mit steigender Fallzahl n nähern sich die Werte der Posteriori-Verteilung den Werten des Konfidenzintervalls der Stichprobenverteilung an.

n	x	2.5% Quantil	97.5% Quantil	KI Untergrenze	KI Obergrenze
10	0	0.00	0.28	0.00	0.18
10	1	0.02	0.41	0.00	0.37
10	5	0.23	0.77	0.22	0.78
100	0	0.00	0.04	0.00	0.02
100	10	0.06	0.17	0.05	0.17
100	50	0.40	0.60	0.40	0.60

Kapitel 10

Monte-Carlo-Methode

Laut Held [14, 185] ist ein grundlegendes Problem der Bayes Inferenz die "Integration zur Berechnung der Normalisierungskonstante".

Die Monte-Carlo Methode und die Markov-Chain-Monte-Carlo Methode (MCMC) erlauben es dieses Problem zu umgehen, indem sie Zufallszahlen aus der Posteriori-Verteilung simulieren und dadurch Integration vermeiden, gleichzeitig aber Kenngrößen der a-posteriori-Verteilung ermitteln.

10.1 Einleitung

Die Statistik beruht auf Erwartungen über Zufallsvariablen:[5, 82]

- Hypothesentests: Der p -Wert einer Stichprobe ist die Wahrscheinlichkeit, weitere Stichproben zu beobachten, die mindestens so extrem sind wie die beobachtete Stichprobe unter der Bedingung, dass die Nullhypothese wahr ist. Sogesehen ist der p -Wert der Erwartungswert einer Zufallsvariable, die 0 ist, wenn eine Stichprobe weniger extrem ist und 1, wenn die Stichprobe extremer ist.

Die gesuchten Schätzer können analytisch ermittelt werden oder über große Stichproben approximiert (ermittelt) werden.¹ Man kann aber auch durch Computer gestützt Verfahren diese Werte ermitteln. Hierfür tut man so, als hätte man echte Zufallszahlen, tatsächlich handelt es sich aber um genauestens definierte Werte, die fast zufällig erscheinen und daher als zufällig behandelt werden können. Diese Zahlen heißen Pseudo-Zufallszahlen. Vorstellen kann man sich zwei Personen, die durch einen Vorhang getrennt sind. Der Programmierer kennt die den Algorithmus bzw. die Mechanismen – für ihn sind die Zahlen deterministisch und vorhersagbar, der Benutzer hat diese Information nicht, daher sind die Werte aus seiner Sicht zufällig [5, 83].

10.1.1 Geschichte der Monte-Carlo-Methode

Während dem zweiten Weltkrieg wurden für den Bau der Atombombe Zufallsprozesse generiert, um die Neutronendiffusion in spaltbarem Material zu simulieren. Als Zufallszahlen benutzte man damals Listen von Roulettepielen

¹Diese Annahme beruht auf dem Gesetz der großen Zahl

aus dem Casino in Monte-Carlo, das dadurch zum Namensgeber des Verfahrens wurde. Die Monte-Carlo-Methode (MC-Methode) dient dazu Vorhersagen zu generieren. Die MC-Methode hat sich in den letzten Jahrzehnten weiterentwickelt, was die Schätzgenauigkeit erhöhte. Hauptgrund sind rechnergestützte Verfahren (Speicherkapazität), die die Simulation von tausenden Fällen erlauben. Praktisch für uns heute: Zufallszahlen müssen nicht mehr per Hand generiert und in Tabellen gespeichert werden.

10.2 Funktion der Monte-Carlo-Methode

Für ein Modell wird eine fiktive Population mit bestimmten Eigenschaften bestimmt. Diese "Pseudo"-Population dient dann als Grundgesamtheit. Aus der Grundgesamtheit werden nun durch Vorwissen bedingte Zufallsstichproben gezogen. Ausgangspunkt für weitere Analysen sind damit konstruierte Zahlenwerte, die einer Verteilung oder einem Zusammenhang unterliegen.

Aus der Population bzw. Grundgesamtheit werden nun in Schleifen Pseudo-Stichproben – Pseudo-Samples – gezogen, wobei in jedem Durchgang der interessierende Parameter θ geschätzt wird, der in einem Parametervektor θ gespeichert wird. Nach einer ausreichenden Anzahl von Wiederholungen und Schätzung von Schätzern θ nähert man sich der Dichte der Stichprobenkennwertverteilung an. Die Stichprobenkennwertverteilung wird über die relative Häufigkeit geschätzt.

10.2.1 Pseudopopulation

Wie oben angedeutet, wird versucht durch eine fiktive Population, die reale Welt widerzuspiegeln. Deren Eigenschaften sollen die Eigenheiten der Population darstellen: deren soziale Mechanismen, Prozesse und Strukturen. Nur so kann man sich ein Bild über die statistischen Kennwerte machen.

Die Eigenheiten einer Population lassen sich unterscheiden in

- feste deterministische Größen und Konstanten,
- stochastische Größen.

Deterministische Größen sind fest vorgegeben durch die Erzeugung von Folgen mit einer Rekursionsformel wie: $a_{n+1} = a_n + 2$ für $n = 0 \dots 99$, was bei einem Startwert von $a_0 = 0$ zu der Zahlenreihe 0, 2, 4, 6, 8, ... führt. Ebenso können Konstanten $\beta_0 = 1.4$ festgelegt werden.

Stochastische Zufallszahlen werden mittels Generatoren (Würfelwurf) erzeugt. Als Resultat erhält man dann gleichwahrscheinliche Pseudozufallszahlen.

10.2.2 Zufallsgenerator einer Gleichverteilung

(!) der gesamte Abschnitt ist prüfungsrelevant.

Gemische Generatoren funktionieren beispielsweise nach folgender Formel [15, 28]: $x_{n+1} = (a \cdot x_n + b) \bmod M$ ²

² $\bmod M$ steht für Modulus "lat: mit Maß" und ist der Rest aus der Division zweier ganzer Zahlen. Beispiel: $7 \bmod 2 = 1$, gesprochen: "sieben modulo zwei gleich eins", denn $7 : 2 = 3$, Rest 1, denn $(2 \cdot 3 + 1) = 7$ (zwei passt drei mal in die 7 – der Rest ist also eins. So ist auch $7 \bmod 3 = 1$).

Setzt man $a = 7$, $b = 2$ und $M = 10$, so ergibt sich bei einem Startwert von $x_0 = 9$ für den nächsten Wert $x_1 = (7 \cdot 9 + 2) \bmod 10$ der Rest von $\frac{65}{10} = 5$. Dies ist dann der neue Startwert und man erhält $x_2 = (7 \cdot 5 + 2) \bmod 10$ mit dem Rest von $\frac{37}{10} = 7$. Für den nächsten Durchlauf $x_3 = (7 \cdot 7 + 2) \bmod 10$ ergibt sich ein Rest von $\frac{51}{10} = 1$. Führt man die Rechnung weiter ergibt sich folgende Tabelle:

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
9	5	7	1	9	5	7	1

Abschließend werden die Werte noch durch M geteilt, so dass die Werte zwischen 0 und 1 normiert werden und so wie Wahrscheinlichkeiten benützt werden können: $u_n = x_n/M$.

u_0	u_1	u_2	u_3	u_4	u_5	u_6	u_7
9/10=0.9	5/10=0.5	7/10=0.7	1/10=0.1	9/10=0.9	5/10=0.5	7/10=0.7	1/10=0.1

Dieses Beispiel eines Pseudozufallsgenerators gelangt ziemlich bald in eine *Wiederholungsschleife* hinein. Verwendet man sehr große Primzahlen, so gelangt man später in die Wiederholungsschleife und verhindert ferner dass ein Zufallsgenerator konstante Werte erzeugt. Heutige Zufallsgeneratoren wiederholen sich erst nach Hunderten von Millionen Pseudozufallszahlen.

Ein weiteres Beispiel soll den Prozess nochmals verdeutlichen, wobei als M -Wert eine Primzahl verwendet wird, um zu verhindern, dass M durch 171 geteilt werden kann:

$$\begin{aligned} x_n &= 171 \times x_{n-1} \pmod{30\,269} \\ u_n &= x_n/30\,269 \end{aligned}$$

Mit Anfangswert $x_0 = 27\,218$.

Das R-Output ergibt:

```
[1] 0.76385080 0.61848756 0.76137302 0.19478675 0.30853348 0.75922561
[7] 0.82757937 0.51607255 0.24840596 0.47741914 0.63867323 0.21312234
[13] 0.44391952 0.91023820 0.65073177 0.27513297 0.04773861 0.16330239
[19] 0.92470845 0.12514454 0.39971588 0.35141564 0.09207440 0.74472232
[25] 0.34751726 0.42545178 0.75225478 0.63556774 0.68208398 0.63636063
[31] 0.81766824 0.82126929 0.43704780 0.73517460 0.71485678 0.24051009
[37] 0.12722587 0.75562457 0.21180085 0.21794575 0.26872378 0.95176583
[43] 0.75195745 0.58472364 0.98774324 0.90409330 0.59995375 0.59209092
[49] 0.24754700 0.33053619
```

Ein anderer Startwert (14 385) führt zu einer anderen Reihe von Zufallszahlen:

```
[1] 0.26581651 0.45462354 0.74062572 0.64699858 0.63675708 0.88546037
[7] 0.41372361 0.74673759 0.69212726 0.35376127 0.49317784 0.33341042
[13] 0.01318180 0.25408834 0.44910635 0.79718524 0.31867587 0.49357428
[19] 0.40120255 0.60563613 0.56377812 0.40605900 0.43608973 0.57134362
```

[25] 0.69975883 0.65875979 0.64792362 0.79493872 0.93452047 0.80299977
 [31] 0.31296045 0.51623773 0.27665268 0.30760844 0.60104397 0.77851928
 [37] 0.12679639 0.68218309 0.65330867 0.71578182 0.39869173 0.17628597
 [43] 0.14490072 0.77802372 0.04205623 0.19161518 0.76619644 0.01959100
 [49] 0.35006112 0.86045129

Die Zufallszahlen können nun verwendet werden um als X und Y-Werte eines 0-1-Quadrates zu dienen.

10.2.3 Die numerische Bestimmung einer Fläche

(!) *der gesamte Abschnitt ist prüfungsrelevant.*

Mit Hilfe von Zufallszahlen kann die Inhalt einer beliebigen Fläche bestimmt werden. So können auch Zufallszahlen gewonnen werden, die einer gewünschten Verteilung entsprechen.

Möchte man die Fläche eines Integrals bestimmen, so ist dies das Verhältnis von Punkten innerhalb einer Fläche zu den Punkten eines Rechtecks, das nochmals um die Größe des Rechtecks gewichtet wird.

$$\frac{\text{Anzahl der Punkte innerhalb der Fläche}}{\text{Gesamtanzahl der Punkte}} \times (\text{Höhe der Box} \times \text{Länge der Box})$$

Bei der Bestimmung handelt es sich um eine relative Häufigkeit, so dass mit größer werdender Punktzahl, gemäß dem *Gesetz der großen Zahl*, auch die Schätzung genauer wird.

10.3 Markov-Ketten-Monte-Carlo-Methode

In der Bayes Statistik gibt es unterschiedliche Verfahren, zur Bestimmung von Integralen mittels Monte-Carlo-Methoden wie beispielsweise die Verwerfungsmethode (Rejection Sampling) oder das Importance Sampling. Auf die häufig verwendete Form der Markov-Ketten-Monte-Carlo-Methode soll noch eingegangen werden.

Die Markov-Ketten(Chain)-Monte-Carlo-Methode ist für hochdimensionale Posteriori-Verteilungen anwendbar.

(!) Zur numerischen Schätzung mehrdimensionaler bayesianischer Modelle haben sich Monte-Carlo-Methoden durchgesetzt, in denen sukzessive Simulationsschritte *Markov-Ketten* bilden, also schwach miteinander korrelieren.

Der Gibbs-Sampler ist eine Methode. Angenommen, ein bayesianisches Modell hat zum Ziel, die a-posteriori-Wahrscheinlichkeitsverteilung dreier Populationsparameter zu schätzen, die als α , β und γ bezeichnet werden. Der Gibbs-Sampler beginnt dann damit, dass beliebige Startwerte für die drei Parameter gewählt werden. Daraufhin wird jeweils ein Zufallswert von den a-posteriori Verteilungen der einzelnen Parameter gezogen (der hochgestellte Index bezeichnet die erst Runde des Simulationsvorgangs):

$$\alpha^1 \sim P(\alpha|\beta, \gamma, X)$$

$$\begin{aligned}\beta^1 &\sim P(\beta|\alpha, \gamma, X) \\ \gamma^1 &\sim P(\gamma|\alpha, \beta, X)\end{aligned}$$

Die a-posteriori Startwerte sind durch die Startwerte der jeweils anderen Parameter bestimmt. In der zweiten Runde der Simulation wird der Vorgang wiederholt, wobei die Parameterverteilungen jetzt durch die Werte der jeweils anderen Parameter bedingt sind, die in der ersten Runde gezogen wurden:

$$\begin{aligned}\alpha^2 &\sim P(\alpha|\beta^1, \gamma^1, X) \\ \beta^2 &\sim P(\beta|\alpha^1, \gamma^1, X) \\ \gamma^2 &\sim P(\gamma|\alpha^1, \beta^1, X)\end{aligned}$$

Es lässt sich zeigen, dass der Prozess, wenn er nur oft genug wiederholt wird (in der Regel mehrere 10.000 Runden, wobei die ersten 4.000-6.000 Runden verworfen werden, da der Sampler einige Runden braucht, um von den Startwerten auf die Zufallsfunktion zu konvergieren), unter relativ allgemeinen Bedingungen die a-posteriori-Wahrscheinlichkeitsverteilung numerisch beschreibt. Auf dieser Grundlage können dann die a-posteriori-Verteilungen der Parameter u. a. wie Mittelwerte, Varianzen und Quantile beschrieben werden.

Kapitel 11

Erata

Hier finden Sie die Änderungen die in dem Skript seit dem 07.07.2009 vorgenommen wurden.

2009-07-07: Tabelle auf Seite 45: geändert wurde die Varianz der Vorvermutung zu:

ϕ^2 groß $\rightarrow \bar{x}$ viel Gewicht

ϕ^2 klein $\rightarrow \nu$ viel Gewicht

2009-07-07: Gleichung 8.6 auf Seite 43. Hier stand bisher für die Berechnung von b' ein falscher Wert. Das n aus der Bernoulliverteilung hat gefehlt.

Ebenso wurde in diesem Kapitel der Parameter θ durch π ersetzt und a durch α und b durch β .

2009-07-08: Erklärung zur Tabelle mit Erwartungswert und Modus auf Seite 47 (obere der beiden Tabellen) verändert.

2009-07-08: Einige (!)-Abschnitte aus den ersten Kapiteln wurden gelöscht

2009-07-08: Kapitel 6 auf Seite 31 mit Vorlagen zur graphischen Darstellung des Einheitsquadrates und des Wahrscheinlichkeitsbaumes wurde hinzugenommen. Die Verwendung des Einheitsquadrates wird in einem extra PDF-Dokument erklärt.g

Literaturverzeichnis

- [1] Wolfgang Balzer. *Die Wissenschaft und ihre Methode: Grundsätze der Wissenschaftstheorie. Ein Lehrbuch*. Karl Alber, Freiburg/München, 1997.
- [2] Wolfgang Bea. *Stochastisches Denken: Analysen aus kognitionspsychologischer und didaktischer Perspektive*, volume Psychologie des Entscheidungsverhaltens und des Konfliktes. Lang, Frankfurt am Main u.a., 1994.
- [3] Manfred Borovcnik. *Stochastik im Wechselspiel von Intuitionen und Mathematik*, volume Lehrbücher und Monographien zur Didaktik der Mathematik. B.I.-Wissenschaftsverlag, Mannheim, 1992.
- [4] Hans Wolfgang Brachinger. Statistik zwischen lüge und wahrheit: Zum wirklichkeitsbezug wirtschafts- und sozialstatistischer aussagen. *Wirtschafts- und Sozialstatistisches Archiv*, (1):5–26, 2007.
- [5] John W. Braun and Duncan J. Murdoch. *A First Course in Statistical Programming with R*. Cambridge University Press, Cambridge, 2007.
- [6] Andreas Broscheid. Bayesianische ansätze zur analyse kleiner fallzahlen. In Peter Kriwy and Christiane Gross, editors, *Klein aber fein! Quantitative empirische Sozialforschung mit kleinen Fallzahlen*, pages 43–64. VS Verlag, Wiesbaden, 2009.
- [7] Rudolf Carnap. Statistische und induktive wahrscheinlichkeit. In Lorenz Krüger, editor, *Erkenntnisprobleme der Naturwissenschaften: Texte zur Einführung in die Philosophie der Wissenschaft*, pages 193–204. Kiepenheuer und Witsch, Köln - Berlin, 1970.
- [8] Ralf Dahrendorf. *Pfade aus Utopia: Arbeiten zur Theorie und Methode der Soziologie*. Piper, München, 1974.
- [9] Ralf Dahrendorf. *Lebenschancen: Anläufe zur sozialen und politischen Theorie*. Suhrkamp, Frankfurt am Main, erste auflage 1979 edition, 1979.
- [10] Ludwig Fahrmeir, Rita Künstler, Iris Pigeot, and Gerhard Tutz. *Statistik: der Weg zur Datenanalyse*, volume 3. verbesserte Auflage. Springer-Verlag, Berlin, 2001.
- [11] Paul Feyerabend. Wie wird man ein braver empirist? ein aufruf zur toleranz in der erkenntnistheorie. In Lorenz Krüger, editor, *Erkenntnisprobleme der Naturwissenschaften: Texte zur Einführung in die Philosophie der Wissenschaft*, pages 302–335. Kiepenheuer und Witsch, Köln - Berlin, 1970.

- [12] Paul Feyerabend. *Der Wissenschaftstheoretische Realismus und die Autorität der Wissenschaften*, volume Wissenschaftstheorie, Wissenschaft und Philosophie. ViewegundSohn, Braunschweig/ Wiesbaden, 1978.
- [13] Jeff Gill. *Bayesian Methods: A Social and Behavioral Sciences Approach*, volume Second Edition. Chapman und Hall/CRC, Boca Rato, 2008.
- [14] Leonhard Held. *Methoden der statistischen Infernez: Likelihood und Bayes*. Spektrum Akademischer Verlag, Heidelberg, 2008.
- [15] Walter Hengartner and Radu Theodorescu. *Einführung in die Monte-Carlo-Methode*. Hansa, München, 1978.
- [16] Vilfredo Pareto. *Traite de Sociologie Generale*. Lausanne-Paris, 1917.
- [17] Karl R. Popper. Selbstbefreiung durch das wissen. In Leonhard Reinisch, editor, *Der Sinn der Geschichte*, pages 100–116. Beck, München, 1961.
- [18] Karl R. Popper. *Logik der Forschung*. Mohr, Tübingen, 1966.
- [19] Karl R. Popper and John C. Eccles. *Das Ich und sein Gehirn*. Piper, München, 1977.
- [20] C. Rahakrishna Rao. *Was ist Zufall? Statistik und Wahrheit*. Prentice Hall Verlag, München u. a., 1995.
- [21] Wolfgang Stegmüller. Das problem der kausalität. In Lorenz Krüger, editor, *Erkenntnisprobleme der Naturwissenschaften: Texte zur Einführung in die Philosophie der Wissenschaft*, pages 156–173. Kiepenheuer und Witsch, Köln - Berlin, 1970.
- [22] Wolfgang Stegmüller. *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, volume 1. Halbband: Personelle Wahrscheinlichkeit und Rationale Entscheidung, 2. Halbband: Personelle und statistische Wahrscheinlichkeit. Springer, Berlin u.a., 1973.