



Laurie Davies erläutert in seinem Artikel einige Fragestellungen der Wahrscheinlichkeitstheorie und beleuchtet sie als Teil der Mathematik, der auch auf komplexe Systeme und Zustände anwendbar ist.

“There are lies, damned lies, and statistics”

Wahrscheinlichkeit und Statistik

Von Laurie Davies

Wahrscheinlichkeit und Statistik

Die mathematische Stochastik umfasst die Wahrscheinlichkeitstheorie und die Statistik und ist Bestandteil der angewandten Mathematik. Trotz dieser Verankerung in der Mathematik umweht die Wahrscheinlichkeitstheorie und noch mehr die Statistik ein Hauch von

Verrufenheit. Bezüglich der Statistik kennt fast jeder die Redewendung “There are lies, damned lies and statistics“. In diesem Zusammenhang erwähne ich das Buch von meinem Dortmunder Kollegen Walter Krämer¹. Die Wahrscheinlichkeitstheorie ist vielleicht nicht ganz so verrufen, aber es gibt deutliche Vorbehalte. In der WAZ vom

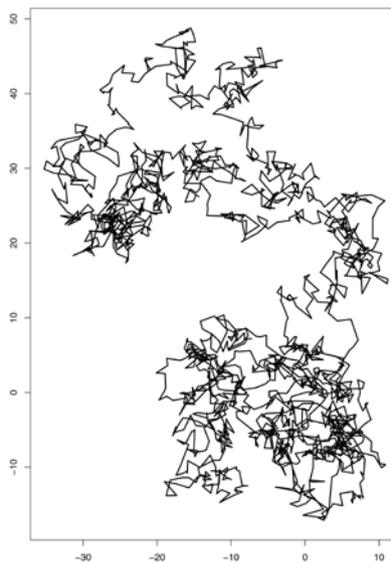
16. Juni 2008 wurde in einer Fußnote, die in Zeitungsartikeln etwas selten sind, Folgendes angemerkt: „Die Wahrscheinlichkeitsrechnung ist strittig, ihr Paradox: Sie wendet mathematische Gesetze auf Zufall an. Gern wird sie so charakterisiert: Wer aus Angst vor Bombenanschlägen ungern fliegt, sollte selber eine Bombe mitnehmen. Denn die

Wahrscheinlichkeit, dass unabhängig voneinander zwei Bomben an Bord sind, ist noch viel niedriger als die, dass eine Bombe an Bord ist.“ Dieses Beispiel ist besonders krass, aber auch das so genannte Ziegenproblem zeigt, dass der Umgang mit Wahrscheinlichkeiten vielen Menschen nicht leicht fällt. Die Argumente und Auseinandersetzungen beziehen sich in den meisten Fällen auf Anwendungen der Wahrscheinlichkeitstheorie: Die Wahrscheinlichkeitstheorie als Bestandteil der Mathematik bleibt von all dem unberührt.

Wahrscheinlichkeitstheorie als Teil der Mathematik

Die Anfänge der Wahrscheinlichkeitstheorie reichen ins 17. Jahrhundert zurück und waren durch Glücksspiele motiviert. Es ging um die Berechnung von Gewinnchancen und die Verteilung der Gelder bei abgebrochenen Spielen. Solche Probleme kann man im Prinzip immer lösen, weil alles, was dazu notwendig ist, die Fähigkeit des Zählens ist. Ein einfaches Beispiel soll dies verdeutlichen. Es wird mit zwei Würfeln gewürfelt und gesucht ist die Wahrscheinlichkeit dafür, dass die Summe der Augenzahlen sechs beträgt. Bei zwei Würfeln gibt es insgesamt 36 Möglichkeiten und von denen ergeben gerade 5 eine Summe von 6. Damit beträgt die Wahrscheinlichkeit $5 / 36$. An dieser Stelle muss man sehr klar zwischen dem mathematischen Modell und dem tatsächlich durchgeführten Versuch unterscheiden. In dem mathematischen Modell werden alle 36 Möglichkeiten als gleich wahrscheinlich angenommen. Wenn man nun innerhalb des Modells andere Wahrscheinlichkeiten berechnen will, so geschieht dies mit derselben Genauigkeit und Präzision wie in allen anderen Bereichen der Mathematik. Man kann sich dabei verrechnen, und viele bekannte Mathematiker haben dies auch prompt getan, sogar Gottfried Wilhelm Leibniz und Isaac Newton. Modelle, in denen die

Elementarereignisse als gleichwahrscheinlich vorausgesetzt werden, werden nach dem französischen Mathematiker Marquis de Pierre Simon Laplace Laplace-Experimente genannt. Laplace war gleichzeitig einer der ersten, der gesehen hat, dass solche Modelle für die Praxis zu speziell sind. In seinem Buch „Philosophischer Versuch über die Wahrscheinlichkeit“ zeigte er anhand von registrierten Geburten, dass die Wahrscheinlichkeiten für einen Knaben oder ein Mädchen



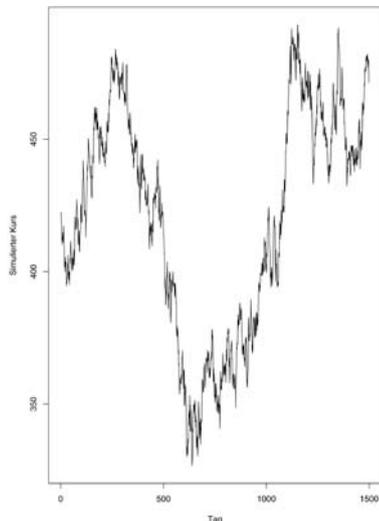
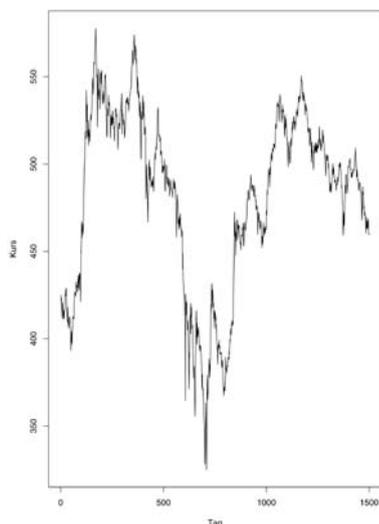
(1) Eine simulierte Brown'sche Bewegung.

nicht gleich sind. Mehr oder weniger gleichzeitig wurde der Anwendungsbereich der Wahrscheinlichkeitstheorie von Zählen auf Messvorgänge erweitert.

Die Erweiterung der Wahrscheinlichkeitstheorie in diese Richtung wurde angestoßen durch Probleme in der Astronomie und der Vermessung. In diesem Zusammenhang ist die so genannte Gauß-Verteilung, in Frankreich die Laplace-Verteilung genannt, entstanden. Auf die nächste Erweiterung der Wahrscheinlichkeitstheorie musste man bis Anfang des 20. Jahrhunderts warten; sie kam aber dann aus zwei ganz unterschiedlichen Richtungen. Im Jahr 1827 beobachtete der schottische Botaniker

Robert Brown, wie Pollenteilchen auf der Oberfläche einer Flüssigkeit sich ständig und ohne erkennbare Regelmäßigkeit bewegten. Eine Simulation ist in Abbildung (1) zu sehen. Die Erklärung kam erst 1905. In diesem Jahr veröffentlichte Einstein drei Arbeiten: „Über einen die Erzeugung und Verwandlung des Lichts betreffenden heuristischen Gesichtspunkt“, „Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen“ sowie „Zur Elektrodynamik bewegter Körper“. Es war die zweite dieser Arbeiten, die ein mathematisches Modell für die von Brown beobachtete Bewegung, die so genannte Brown'sche Bewegung, bereitstellte. Als er die Arbeit schrieb, wusste aber Einstein nichts von der Brown'schen Bewegung. Seine Motivation war die reelle Existenz von Molekülen, die unter anderen von dem berühmten österreichischen Physiker Ernst Mach verneint wurde, nachzuweisen. Für Mach waren Moleküle rein theoretische Konzepte ohne physikalische Existenz. Erst zum Schluss der Arbeit von Einstein findet man eine Bemerkung, dass seine Theorie eine Erklärung für die Brown'sche Bewegung sein könnte. Im Jahr 1900, fünf Jahre vor Erscheinen der Arbeit von Einstein, entwickelte der französische Mathematiker Louis Bachelier in seiner eingereichten Dissertationsarbeit „Théorie de la Spéculation“ ein Modell für die Bewegung von Aktienkursen, das im Wesentlichen identisch mit dem Modell von Einstein war. Das obere Bild von Abbildung (2) zeigt den Kurs des Deutschen Aktienindex DAX an 1500 aufeinander folgenden Tagen. Das untere Bild zeigt eine Simulation einer exponentiellen Brown'schen Bewegung.

Die exponentielle Brown'sche Bewegung ist zu einfach, um die tatsächlich beobachteten Kurse erfolgreich modellieren zu können, aber wesentliche Elemente der Preisbewegung werden schon korrekt



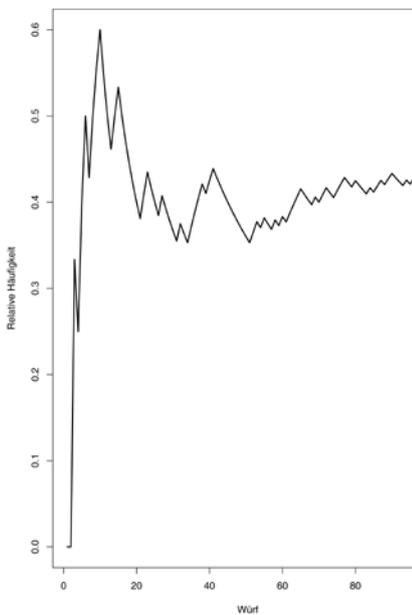
(2) Oben: DAX-Index an 1500 aufeinander folgenden Tagen. Unten: Eine simulierte exponentielle Brown'sche Bewegung.

wiedergegeben. Mit den Arbeiten von Bachelier und Einstein wurden zum ersten Mal zufällige Funktionen Gegenstand einer mathematischen Theorie. Zu diesem Zeitpunkt war aber die Mathematik selbst mit der Problematik überfordert und sowohl die Arbeit von Bachelier wie auch die von Einstein genügten nicht den Anforderungen an Genauigkeit, die die Mathematik verlangt. Dies wurde zum Verhängnis für Bachelier. Obwohl seine Arbeit unter Leitung von Henri Poincaré angenommen wurde, wurde ihr danach

wegen fehlender mathematischer Genauigkeit die verdiente Anerkennung verweigert. Die notwendige mathematische Integrationstheorie, die Bachelier fehlte, um seine Ideen rigoros darzustellen, wurde just zu dieser Zeit und auch in Paris von Henri Lebesgue und anderen französischen Mathematikern entwickelt, eine traurige Ironie. Mit Hilfe der neuen Integrationstheorie gelang es um 1920 dem amerikanischen Mathematiker Wiener zusammen mit anderen, das erste mathematisch einwandfreie Modell für die Brown'sche Bewegung zu entwickeln. Die Brown'sche Bewegung wird öfters zur Ehre Wieners „der Wiener-Prozess“ genannt. Der endgültige Durchbruch der Wahrscheinlichkeitstheorie als eigenständiger Teil der Mathematik geschah 1933 mit der Veröffentlichung des Buches (auf Deutsch) „Grundbegriffe der Wahrscheinlichkeitsrechnung“ des russischen Mathematikers Andrey Kolmogorov. Abgesehen von dem axiomatischen Aufbau führte Kolmogorov die erste mathematisch korrekte Definition des bedingten Erwartungswertes basierend auf dem Satz von Radon-Nikodym ein. Seitdem hat die Wahrscheinlichkeitstheorie eine rasante Entwicklung genommen und ist nun nicht mehr aus der modernen Mathematik wegzudenken. Die Wahrscheinlichkeitstheorie findet Anwendung innerhalb der Mathematik, sogar in der Zahlentheorie, wie auch in vielen anderen Disziplinen wie Physik, Chemie, Biologie, Medizin, Informatik und Soziologie. Obwohl die Wahrscheinlichkeitstheorie als Bestandteil der Mathematik unstrittig ist, gibt es in der Tat Schwierigkeiten, wenn man versucht, sie auf reelle Situationen anzuwenden. Vielleicht liegt das Problem darin, dass im Gegensatz zu Begriffen wie Länge, Gewicht, Zeit und Spannung, Wahrscheinlichkeit zuerst undefiniert und nicht direkt messbar ist. Die zwei verbreitetsten Interpretationen von Wahrscheinlichkeit sind Glaubensgrad und relative Häufigkeit.

Wahrscheinlichkeit als Glaubensgrad

Die Definition von Wahrscheinlichkeit als Glaubensgrad ist personenbezogen (subjektiv) und erfolgt über die Festlegung von Wettquoten. Wir betrachten ein bestimmtes Ereignis, zum Beispiel dass Deutschland 2010 Fußballweltmeister wird. Durch Fragen ermitteln wir die ungünstigste Wettquote, zu der diese Person bereit wäre, eine Wette über den Erfolg der deutschen Mannschaft einzugehen. Genauso können wir die ungünstigste Wettquote ermitteln, zu der diese Person bereit wäre, eine Wette über den Misserfolg der deutschen Mannschaft einzugehen. Nehmen wir an, dass die Person die Quoten 1:4 für Erfolg und 1:1 für Misserfolg nennt. Wenn sie nun 10 Euro auf Erfolg und 20 Euro auf Misserfolg setzt, gewinnt sie, was auch immer passiert: Die Gewinne der einen Wette sind höher als die Verluste der anderen. Dies ist möglich, weil $1/5 + 1/2$ kleiner ist als 1. Wir können nun beide Quoten leicht verschlechtern und immer noch erreichen, dass ein sicherer Gewinn möglich ist. Wenn die ursprünglich angegebenen Quoten wirklich die ungünstigsten wären, würde dies bedeuten, dass unsere Person Quoten ablehnen würde, die einen sicheren Gewinn ermöglichen würden. Hieraus kann man schließen, dass die Summe der Quoten mindestens 1 ergeben muss. Umgekehrt, ist die Summe größer als 1, dann wäre unsere Person bereit, Wetten einzugehen, bei denen ein Verlust sicher wäre. Ein sicherer Verlust oder ein sicherer Gewinn ist nur dann nicht möglich, wenn die Summe der Quoten 1 ergibt. Auf Englisch heißt dieses Argument „Dutch book“-Argument. Man kann einiges dagegen einwenden, aber man kann auf diese Art und Weise über Wahrscheinlichkeiten für nicht wiederholbare Ereignisse oder sogar für Aussagen wie „Ich halte es für wahrscheinlich, dass jede gerade Zahl die Summe zweier Primzahlen ist“ sprechen.



(3) Relative Häufigkeit von „Zahlwürfen“.

Wahrscheinlichkeit als relative Häufigkeit

Die zweite gängige Interpretation von Wahrscheinlichkeit ist als Grenzwert der relativen Häufigkeiten. Abbildung (3) zeigt die Entwicklung der relativen Häufigkeit von „Zahlwürfen“ bei einer Versuchsreihe von 100 Würfeln. Die Abbildung (3) vermittelt den Eindruck, dass die relative Häufigkeit gegen einen Grenzwert konvergiert. Dieser wird dann als Definition der Wahrscheinlichkeit für einen Kopfwurf genommen. Es gibt aber erhebliche Probleme mit dieser Definition. Zuerst kann man anhand einer endlichen Anzahl von Versuchen nie nachweisen, dass ein Grenzwert existiert. Die Hauptschwierigkeit besteht aber darin, zu spezifizieren, wie die Münze geworfen wird. Es ist zum Beispiel nicht erlaubt, die Münze mit „Zahl“ nach oben einfach auf den Boden zu legen. Traditionell wird die Münze mit Hilfe des Daumens hoch in die Luft und mit schnellen Drehungen geworfen. Die Idee ist, dass die Ergebnisse dadurch zufällig werden. Bei der Ziehung der Lottozahlen wird ein erheblicher Aufwand getrieben, um sicherzu-

stellen, dass die Ergebnisse zufällig sind. Dadurch hängt aber nun der Begriff „Wahrscheinlichkeit“ von dem Begriff „Zufall“ ab. Wir müssen nicht nur die Bedingungen so gestalten, dass die Würfe unabhängig sind, sondern auch so, dass sie vergleichbar sind: ändert man die Bedingungen, so ändert sich die Wahrscheinlichkeit. Der Begriff „Wahrscheinlichkeit“ bezieht sich somit nicht auf die Münze alleine, sondern auf die ganze Versuchsanordnung. In der mathematischen Wahrscheinlichkeitstheorie wird eine solche Versuchsanordnung als „unabhängig und identisch verteilt“ bezeichnet. Die beiden Forderungen, Zufall und identische Versuchsbedingungen widersprechen sich, wenigstens im Rahmen der deterministischen Newton’schen Mechanik: Sind die Versuchsbedingungen identisch, dann sind die Ergebnisse identisch. Auf den ersten Blick bietet die Quantenmechanik einen möglichen Ausweg, weil sie in der gängigen Interpretation ein zufälliges Element enthält, nämlich das Kollabieren des Wellenpakets. Es gibt aber deterministische Versionen der Quantenmechanik, die Bohm’sche Mechanik, und somit gibt es keine eindeutige Antwort auf die Quelle des Zufalls. Trotzdem lässt sich der einfache Münzwurf durch die Wahrscheinlichkeitstheorie sehr gut modellieren.

Eine mögliche Erklärung hierfür liefert die Chaostheorie.

Zufall und Chaos

Ein chaotisches System ist ein deterministisches System, bei dem das zeitliche Verhalten des Systems sehr empfindlich von den Anfangsbedingungen abhängt. Dies ist der so genannte Schmetterling-Effekt: Die Anfangsbedingungen werden durch einen Schmetterling gestört und dies führt später zu einem Gewitter. Nun lässt sich für einige chaotische Systeme mathematisch beweisen, dass sie sich wie zufällige Systeme verhalten. Ein Beispiel

für ein chaotisches System ist die Ziehung der Lottozahlen. Das Verfahren ist so angelegt, möglichst chaotisch zu sein. Damit kann man trotz fast identischer Anfangsbedingungen gewährleisten, dass die Ergebnisse verschieden sind. Ob die tatsächlichen Ergebnisse konsistent mit der mathematischen Theorie des Zufalls sind, ist letztendlich eine empirische Frage. Die Zufallszahlen, die ein Computer liefert, sind deterministisch und heißen deswegen Pseudozufallszahlen. Bei Simulationen ist es von großem Vorteil, wenn man genau dieselben Simulationen immer wieder durchführen kann. Dies erreicht man, indem man dieselbe Anfangszahl (seed) für den Pseudozufallserzeuger wählt. Setzt man in meinem Computerprogramm diese Anfangszahl auf 11051944, so bekommt man als erste Zufallszahl 0.3728457. Setzt man die Anfangszahl auf 11051945, so bekommt man als erste Zufallszahl 0.1412734. Damit ähnelt das System zur Erzeugung von Pseudozufallszahlen einem chaotischen System.

Zufall und Komplexität

Einen anderen Zugang zum Zufall bietet die Komplexitätstheorie, die auf Kolmogorov und andere zurückgeht. Man stellt sich eine Folge von Nullen und Einsen vor und muss jetzt ein Computerprogramm schreiben, das diese Folge wiedergibt. Gibt es Regularitäten in der Folge, zum Beispiel 0101010101..., so kann man diese ausnutzen, um ein kurzes Programm zu schreiben. Ist aber die Folge ohne jegliche Regelmäßigkeit, also sozusagen zufällig, dann wird das Programm genauso lang sein, wie die Folge selbst. Dies führt zur Definition der Komplexität einer Folge als Länge des kürzesten Computerprogramms, das die Folge wiedergibt. Es lässt sich zeigen, dass sich die Längen dieser Programme für zwei verschiedene Programmiersprachen durch höchstens eine Konstante unterscheiden. Damit

ergeben alle Programmiersprachen asymptotisch gesehen denselben Wert. Eine zufällige Folge kann man nun definieren als eine, für die das kürzeste Programm genauso lang ist wie die Folge. In der Tat kann man zeigen, dass alle solche Folgen alle statistischen Tests auf Zufälligkeit bestehen. Leider ist die Theorie nicht anwendbar. Zum einen, weil es keine Möglichkeit gibt, das kürzeste Programm zu bestimmen. Zum anderen, weil die Theorie nur asymptotisch ist. Für endliche Folgen können Programme in verschiedenen Sprachen zu verschiedenen Ergebnissen führen.

In Anbetracht der Tatsache, dass komplexe Folgen alle statistischen Tests auf Zufälligkeit bestehen, ist es plausibel, dass man die Wahrscheinlichkeitstheorie auf solche Folgen erfolgreich anwenden kann. Etwas überraschend aber ist die Tatsache, dass man die Wahrscheinlichkeitstheorie auch auf einige einfache Folgen anwenden kann.

1100100100001111110110101010001000100 – Die Folge mutet zuerst als kompliziert an. Sie ist aber einfach und es reicht ein kurzes Programm, die Folge zu reproduzieren. Es handelt sich nämlich um die Binärdarstellung von π . Diese Folge ist im Sinne von Kolmogorov einfach. Trotzdem hat sie bis jetzt alle Tests auf Zufälligkeit bestanden und die Wahrscheinlichkeitstheorie kann man mit Erfolg auf sie anwenden.

Statistik

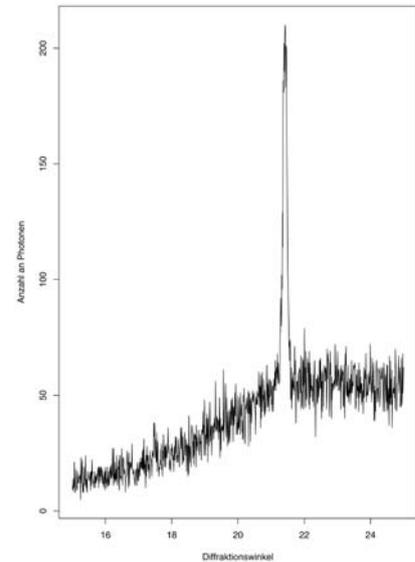
Frequentistische und Bayesianische Statistik

Entsprechend den beiden Interpretationen von Wahrscheinlichkeit gibt es zwei Schulen in der Statistik, die Bayesianische und die frequentistische. Beide benutzen die mathematische Wahrscheinlichkeitstheorie, um die Daten zu modellieren. Als Beispiel betrachten wir Würfe mit einer gefälschten Münze. In beiden Schulen werden

die Ergebnisse einer Wurfreihe als unabhängig und identisch verteilte Bernoulli-Zufallsvariable modelliert, wobei die Wahrscheinlichkeit für Zahl p beträgt. In der Bayesianischen Statistik hat man vor der Durchführung des Experiments Glaubensgrade über den Wert von p . Nun wird das Experiment durchgeführt und anhand der Ergebnisse X_1, \dots, X_n wird der Glaubensgrad durch Anwendung des Satzes von Bayes neu bestimmt. Im frequentistischen Ansatz wird p als ein unbekannter Parameter betrachtet, den man anhand der Daten so gut wie möglich schätzen will. Die Gütekriterien hierfür sind frequentistisch, zum Beispiel der erwartete quadratische Fehler, wobei der Erwartungswert implizit vorausgesetzt, dass der Wurfvorgang immer wieder wiederholt wird. Im Bayesianischen Ansatz wird ein einziger Datensatz analysiert, genauso wie der Begriff von Wahrscheinlichkeit als Ausdruck des Glaubensgrades auf einzelne und nicht wiederholbare Ereignisse angewandt werden kann. In der Praxis kommen Datensätze mit ähnlichen Strukturen immer wieder vor, so dass auch der Bayesianer sich Gedanken über die frequentistischen Eigenschaften seines Verfahrens machen muss. Im Folgenden werden wir uns auf den frequentistischen Ansatz beschränken.

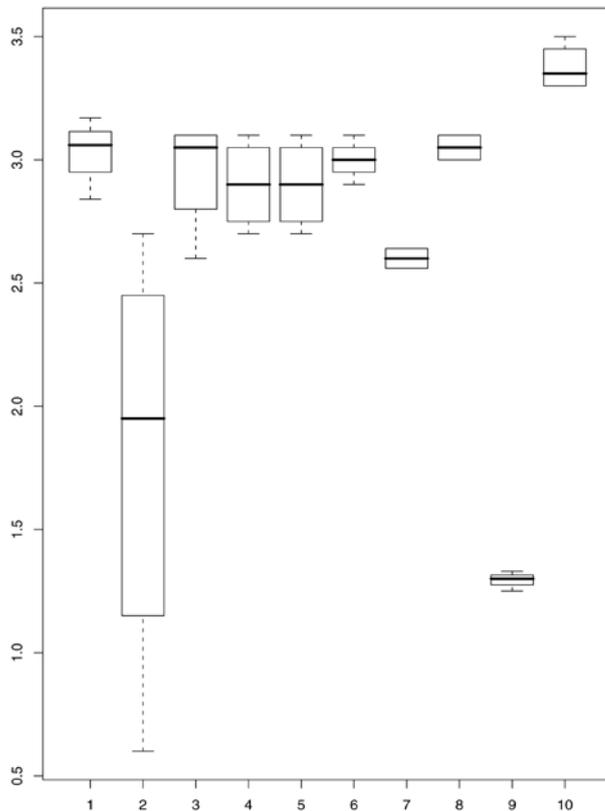
Mathematische Statistik

Die mathematische Statistik operiert innerhalb einer parametrisierten Familie von stochastischen Modellen, die nichts anderes sind als Wahrscheinlichkeitsmaße auf dem Stichprobenraum. Es wird angenommen, dass die Daten einer Verteilung mit einem bestimmten, aber unbekanntem Parameterwert gehorchen. Anhand der Daten will man nun diesen unbekanntem Parameterwert optimal schätzen. Dafür werden Optimalitätskriterien aufgestellt, zum Beispiel minimale Varianz, und dann wird zumindest



(4) Anzahl an Photonen von Röntgenstrahlen in Abhängigkeit vom Diffraktionswinkel.

versucht, das Optimalitätsproblem zu lösen. Als Beispiel nehmen wir wieder den einfachen Münzwurf mit einer gefälschten Münze. Die Wurfreihe wird durch die unabhängig verteilte Bernoulli-Zufallsvariable mit Erfolgswahrscheinlichkeit p modelliert, wobei p nun der Parameter ist. Unter allen Schätzern minimiert das arithmetische Mittel der Daten den zu erwartenden quadratischen Fehler. Damit ist das arithmetische Mittel in dieser Situation optimal. Im Allgemeinen kann das Optimalitätsproblem für eine gegebene Stichprobengröße nicht gelöst werden. In solchen Fällen betreibt man Asymptotik und leitet Verfahren her, die asymptotisch optimal sind. Will man bei dem einfachen Münzwurf $1/p$ statt p schätzen, so gibt es kein Verfahren, das für alle p -Werte optimal ist. Man kann aber zeigen, dass $1/X_n$ asymptotisch optimal ist, wobei X_n das arithmetische Mittel der Daten bezeichnet. Für bestimmte Klassen von parametrisierten Modellen kann man zeigen, dass der so genannte Maximum-Likelihood-Schätzer immer asymptotisch optimal ist. Dies erklärt seine verbreitete Anwendung in der Statistik. Es ist immer gut behaupten zu können,



(5) Ergebnisse eines Ringversuchs im Bereich von Trinkwasserqualität.

dass man den optimalen Schätzer benutzt. Die Optimalität von Maximum-Likelihood gilt allerdings immer nur unter bestimmten Voraussetzungen. Eine davon ist, dass die Dimension des Parameter-raumes klein ist, was nicht immer der Fall ist. Abbildung (4) zeigt einen Ausschnitt von so genannten Dünnschichtdaten. Sie geben die Anzahl an Photonen von Röntgenstrahlen in Abhängigkeit vom Diffraktionswinkel an. Sie wurden von Dieter Mergel, Fachbereich Physik, zur Verfügung gestellt. Das Standardmodell für solche Datensätze ist von der Gestalt $X(t) = f(t) + Z(t)$, das heißt die Daten setzen sich zusammen aus einer Funktion f und Rauschen Z . Hier ist f der Parameter und der dazugehörige Parameterraum ist nun unendlich dimensional.

In einer solchen Situation versagt der Maximum-Likelihood-

Schätzer, weil er die Daten einfach wiedergibt. Auch der übliche Optimalitätsbegriff vom minimalen erwarteten quadratischen Fehler greift nicht mehr: Er ist einfach zu stark und wird durch optimale Konvergenzraten ersetzt. Hochdimensionale Parameterräume und hochdimensionale Datensätze spielen eine immer größere Rolle in der Statistik. Die entsprechende mathematische Theorie wird immer noch entwickelt.

Robuste Statistik

Um 1920 gab es einen Streit zwischen dem englischen Statistiker Fisher und dem englischen Astronomen Eddington über die beste Methode, die Variabilität eines Datensatzes zu bestimmen. Fisher plädierte für die Überlegenheit der Standardabweichung, weil er nachweisen konnte, dass sie für

normalverteilte Daten optimal ist. Eddington argumentierte aus seiner Erfahrung mit astronomischen Daten, dass die mittlere absolute Abweichung vom Mittelwert besser sei, weil sie durch besonders große Beobachtungen weniger beeinflusst ist. Beide hatten recht, wie John W. Tukey 1960 und Huber 1981 zeigten. Wenn die Daten tatsächlich normalverteilt sind, hat Fisher recht. Es genügt aber eine kleine Abweichung von weniger als einem Prozent von der Normalverteilung und schon hat Eddington recht. Das Beispiel zeigt, dass optimale Verfahren sehr stark von den gemachten Annahmen abhängen können, und dass eine kleine Perturbation genügt, um aus einem optimalen Verfahren ein sehr schlechtes Verfahren zu machen. Die robuste Statistik beschäftigt sich mit dieser Problematik. Gesucht sind Verfahren, die sich über ganze Nachbarschaften von Modellen gutartig verhalten. Ein Anwendungsgebiet ist die Behandlung von so genannten Ausreißern.

Abbildung (5) zeigt einen Boxplot eines Ringversuches im Bereich von Trinkwasserqualität. Es waren zehn Labore daran beteiligt und zwei Ausreißer sind klar zu erkennen. Die statistische Auswertung eines solchen Ringversuches ist genau vorgeschrieben und die Herausforderung besteht darin, den Einfluss von Ausreißern auf die Auswertung zu reduzieren. Der statistische Teil von DIN 38402-45 (Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung) basiert auf so genannten "redescending" M-Schätzern, die den Einfluss von Ausreißern allmählich auf null reduzieren.

Wenn die Daten eindimensional sind, ist die Entdeckung von Ausreißern mit grafischen Mitteln einfach. Wenn aber die Daten etwas komplizierter sind, kann es etwas schwieriger sein. Die Daten in der Tabelle der Abbildung (6) geben die Anzahl an keimenden Kopfsalatpflanzen nach Anwendung von

zwei verschiedenen Düngemitteln in jeweils drei Konzentrationen an. Ein Ausreißer ist ohne weiteres nicht zu erkennen. Wendet man eine L_2 -Analyse (kleinste Summe der Quadrate) auf den Datensatz

449	413	326
409	358	291
341	278	312

(6) Keimende Pflanzen in Abhängigkeit von den Düngemittelkonzentrationen.

an, so bekommt man die Residuen in Tabelle (7). Ein Ausreißer ist nicht sichtbar. Wendet man dagegen eine L_1 -Analyse (kleinste Summe der absoluten Residuen) an, so bekommt man die Residuen in Tabelle (8). Ein Ausreißer in der Zelle (3,3) ist nun sichtbar. Die mathematische Theorie hierzu findet man bei Terbeck². Eine genauere Analyse legt die Vermutung nahe, dass der Eintrag in Zelle (3,3) 212 sein sollte. Dieses Beispiel zeigt eine ausgeprägte Schwäche einer L_2 -Analyse, nämlich, dass sie in der Regel nicht geeignet ist, Ausreißer zu entdecken. In dem Beispiel von Tabelle (6) war das größte L_2 -Residuum in der Tat der Ausreißer, aber dies ist nicht immer der Fall.

Die Ausreißer verfälschen eine L_2 -Analyse so stark, dass sie sich dahinter verstecken können. Dieser Effekt heißt "masking effect".

Die L_1 -Analyse hat den Ausreißer entdeckt, aber dies liegt an der speziellen Gestalt der Daten. Im Allgemeinen kann auch eine L_1 -Analyse versagen, insbesondere in der linearen Regression. Ein einfaches Maß für die Robustheit einer Statistik ist der so genannte „finite sample breakdown point“³, hier kurz Bruchpunkt genannt. Man stellt sich eine Stichprobe vom Umfang n vor und man darf k der Werte beliebig verändern. Dann ist der Bruchpunkt der Statistik

k^*/n , wobei k^* die kleinste Anzahl von Stichprobenwerten bezeichnet, die man verändern muss, bis die Statistik beliebig große Werte annehmen kann. Man überlege sich kurz, dass das arithmetische Mittel einen Bruchpunkt von $1/n$ hat, der Median dagegen den optimalen Bruchpunkt von $1/2$. Das Wort „optimal“ bezieht sich hier auf eine Klasse von Schätzern, die verträglich mit einer gegebenen Gruppenstruktur auf dem Stichprobenraum sind⁴. In der linearen Regression hat der Kleinste-Quadrate-Schätzer (L_2 -Schätzer) einen Bruchpunkt von $1/n$ wie auch der L_1 -Schätzer, eine Tatsache, die auf den ersten Blick etwas überraschend ist. Der optimale Bruchpunkt in der linearen Regression beträgt ungefähr $1/2$ und

6.33	20.3	-26.7
9.67	8.67	-18.3
-16.0	-29.0	45

(7) Residuen nach einer L_2 -Analyse.

0	15	-5
0	0	0
0	-12	89

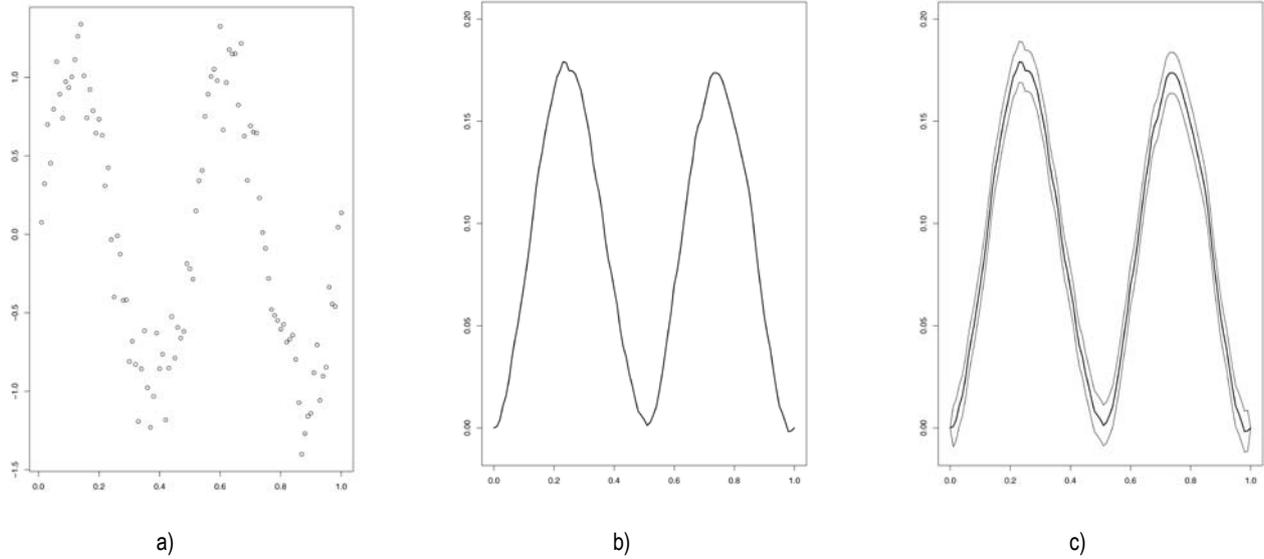
(8) Residuen nach einer L_1 -Analyse.

wird durch den so genannten LMS-Schätzer (Least Median of Squares), der auf Tukey und Hampel zurückgeht, erreicht. Die LMS-Methode bestimmt die Regressionskoeffizienten so, dass der Median der quadratischen Residuen minimiert wird. Schätzer mit optimalen Bruchpunkten heißen "high breakdown estimators". In den letzten 20 Jahren sind viele solcher Schätzer entwickelt worden und sie sind inzwischen unverzichtbare Werkzeuge in der Datenanalyse. Ihre Hauptschwäche besteht darin, dass sie nur für kleine Stichproben genau berechnet werden können.

Für große Stichproben gibt es nur – wenn auch meistens sehr effektive – heuristische Algorithmen. Dies ist ein inhärentes Problem: Für bestimmte "exact fit"-Probleme ist es bekannt, dass sie algorithmisch schwierig sind, und, weil sie durch optimale robuste Schätzer gelöst werden, sind diese auch algorithmisch schwierig. Inzwischen sind "high breakdown"-Schätzer Bestandteil der meisten Statistikprogramme.

Nicht-parametrische Regression

Die nicht-parametrische Regression beschäftigt sich mit der Schätzung der Funktionen f in dem Modell $X(t) = f(t) + Z(t)$. Die Bezeichnung „nicht-parametrisch“ ist aus mathematischer Sicht falsch, weil f selbst der Parameter ist, sie ist aber die Standardbezeichnung. Das Modell eignet sich für die Analyse von Datensätzen wie in Abbildung (4), und es lässt sich auf zwei-, drei- und vierdimensionale Bilder (Raum und Zeit) verallgemeinern. Abbildung (4) zeigt nur die ersten 1000 Punkte eines Datensatzes der Länge 7001. Das Problem besteht darin, die Anzahl, die Positionen und Höhen der lokalen Maxima zu bestimmen⁵. Eine besonders effektive Methode dies zu erreichen, ist die „Straffe-Saite-Methode“⁶, die nun kurz erklärt wird. Das erste Bild von Abbildung (9) zeigt eine verrauschte Sinus-Kurve. Zuerst werden die Daten aufintegriert und es ergibt sich das zweite Bild von Abbildung (9). Wir bilden nun einen Schlauch, indem wir die aufintegrierten Daten um 5 nach unten und 5 nach oben verschieben, mit Ausnahme der beiden Endpunkte, die erhalten bleiben. Wir bekommen nun das dritte Bild von Abbildung (9). Jetzt muss man sich eine Saite vorstellen, die in (0, 0) anfängt und in dem letzten Punkt der aufintegrierten Daten aufhört. Die Saite wird straff gezogen, muss aber innerhalb des Schlauches bleiben. Das Ergebnis ist das vierte Bild von



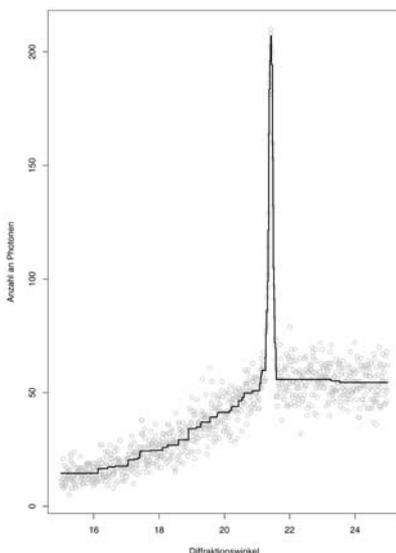
(9 a-e) Darstellung der „Straffe-Saite-Methode“.

Abbildung (9). Schließlich wird die Saite differenziert, die so entstandene Ableitung ist ein Schätzer für die Funktion f . Das Endergebnis ist Bild e von Abbildung (9). Die Beziehung zwischen der straffen Saite und der Anzahl der lokalen Extremwerte ist die Folgende. Von allen Funktionen, die durch den Schlauch verlaufen, hat die Ableitung der straffen Saite die minimale Anzahl an lokalen Extremwerten. Das Endergebnis hängt von der Breite des

Schlauches ab. Es ist aber möglich, die Breite aus den Daten automatisch zu bestimmen. Bei Datensätzen, die erhebliche lokale Variationen in der Glattheit aufweisen, ist der automatische Schlauch nicht überall gleich breit. An Stellen, an denen die Funktion sich stark verändert, ist er eng, an anderen Stellen ist er breit. Abbildung (10) zeigt das Ergebnis der automatischen Methode für die Daten von Abbildung (4).

nature. A deterministic nature does not per se imply the inapplicability of probability theory. It can be shown mathematically that complex sequences of numbers and certain deterministic chaotic systems behave like random systems and can be well described using probability theory.

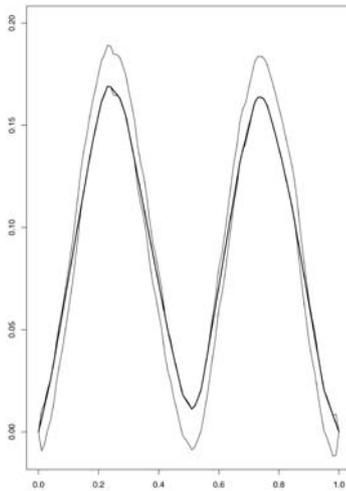
Statistics is one form of applied probability theory and is consequently more influenced by problems of interpretability than is probability theory itself. In its pure form Bayesian statistics is concerned with degrees of belief and how these should be altered in the light of experimental evidence. Frequentist statistics is more concerned with providing procedures which are optimal for repeated applications under similar circumstances. In practice both pure forms are tempered by practical necessities. The article describes two special areas of statistics, robust statistics and non-parametric regression. Robust statistics is concerned with the detection and accommodation of outlying or aberrant observations as they occur for example in interlaboratory tests and the analysis of variance. Non-parametric regression is concerned with estimating a functional



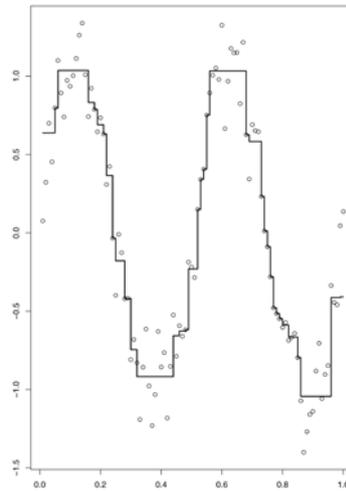
(10) Das Ergebnis der „Straffe-Saite-Methode“ für die Daten von Abbildung (4).

Summary

The article discusses some of the problems concerning the interpretation of the concept of probability. It is emphasized that the mathematical theory of probability, being a part of mathematics, is not affected by such considerations, but that there are problems with subjective and frequentist interpretations when it comes to applications. In particular the frequentist requirement that the experimental conditions remain the same for all repetitions is in conflict with deterministic theories of



d)



e)

relationship between two variables when this is contaminated by noise. The example discussed comes from the area of thin film physics where the intensity of diffracted x-rays depends on the angle of diffraction.

(with discussion). *Annals of Statistics*, 29(1):1-65.
 – Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In Bickel, P. J., Doksum, K. A., and J. L. Hodges, J., editors, *A Festschrift for Erich L. Lehmann*, 157-184, Belmont, California. Wadsworth.]
 – Krämer, W. (2008). *So lügt man mit Statistik*. Piper, München, 10th edition.
 – Terbeck, W. und Davies, P. L. (1998). Interactions and outliers in the two-way analysis of variance. *Annals of Statistics*, 26:1279-1305.

Anmerkungen

- 1) Krämer, 2008
- 2) Terbeck, 1998
- 3) Donoho and Huber, 1983
- 4) Davies and Gather, 2005, Davies and Gather, 2006
- 5) Davies et al., 2008
- 6) Davies and Kovac, 2001

Literatur

- Davies, P. L. and Gather, U. (2005). Breakdown and groups (with discussion). *Annals of Statistics*, 33:977-1035.
 – Davies, P. L. and Gather, U. (2006). Addendum to the discussion of ‘Breakdown and Groups’. *Annals of Statistics*, 34:1577-1579.
 – Davies, P. L., Gather, U., Meise, M., Mergel, D., and Mildenerger, T. (2008). Residual based localization and quantification of peaks in x-ray diffractograms. *Annals of Applied Statistics*. to appear.
 – Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution

Der Autor

Laurie Davies, Jahrgang 1944, studierte von 1962 bis 1966 Mathematik und Wirtschaftswissenschaften an der Universität Cambridge. Von 1966 bis 1971 besuchte er die London School of Economics and Political Science, wo er 1971 in Mathematik promovierte. Von 1971 bis 1973 war er Wissenschaftlicher Mitarbeiter an der Universität Konstanz, wo er sich 1974 habilitierte. Von 1973 bis 1974 war er Lecturer am Department of Statistics der Universität Sheffield. Er folgte 1975 einem Ruf an die Universität Münster und seit 1977 ist er Universitätsprofessor an der Universität in Essen. Seine Forschungsinteressen umfassen die Grundlagen der Statistik, die robuste Statistik und die nichtparametrische Statistik.

DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

ub | universitäts
bibliothek

Dieser Text wird über DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

DOI: 10.17185/duepublico/73817

URN: urn:nbn:de:hbz:464-20210204-165558-3

Alle Rechte vorbehalten.