Vorwort

Stochastik ist die Mathematik des Zufalls. Sie ist von größter Bedeutung für die Berufspraxis der Mathematiker. An vielen Schulen hat sie ihren festen Platz gefunden.

Die beiden Hauptgebiete der Stochastik sind Wahrscheinlichkeitstheorie und Statistik. In der Wahrscheinlichkeitstheorie untersucht man zufällige Prozesse mit festen als bekannt angenommenen steuernden Wahrscheinlichkeiten. Dies ist theoretisch und praktisch von eigenständigem Interesse. Darüber hinaus liefert die Wahrscheinlichkeitstheorie Grundlagen für die Statistik, in der aus beobachteten Daten Schlüsse über unbekannte Wahrscheinlichkeiten und über zweckmäßiges Verhalten gezogen werden sollen.

Stochastische Fragen treten in den unterschiedlichsten Problemkreisen auf. Hier einige Beispiele:

- Was sind gute Strategien bei Glücksspielen und anderen Entscheidungsprozessen unter Unsicherheit?
- Welche Wahrscheinlichkeitsaussagen lassen sich über das Wachstum von Populationen und über die Vererbung von Eigenschaften machen?
- Wie übermittelt man ökonomisch Nachrichten?
- Wie vergleicht man mit vorgegebener Sicherheit die Qualität von Heilmitteln oder Produktionsverfahren?
- Was lässt sich über die Genauigkeit von Messungen aussagen?

Dies sind Fragen, die sich nicht ohne Zusatzüberlegungen nur durch den Beweis mathematischer Sätze beantworten lassen. Ein wesentlicher Teil der Schwierigkeit besteht bereits darin, die passenden mathematischen Begriffe zu entwickeln, die es erlauben, diese „realen“ Fragen angemessen mathematisch auszudrücken. Die für Berufspraxis und Schule gleichermaßen wichtige Umsetzung von realen Problemen in eine adäquate theoretische Form kann man wohl nirgends besser üben als in der Stochastik. Die Übungsaufgaben, die oft von der „eingekleideten“ Art sind, sind dabei äußerst wichtig. Der Leser sollte so viele wie möglich lösen.

Ich habe versucht, ein wenig von der Faszination zu vermitteln, die Stochastik ausüben kann. Dies war mir wichtiger als eine möglichst vollständige Abhandlung der praktisch gebräuchlichen Verfahren. Ist das Interesse geweckt, kann ja der Leser weitere Literatur heranziehen. Immerhin wird aber ein gewisser Fundus der Methodenlehre vermittelt, und ich denke, dass der Leser, der hier die Grundideen verstanden hat, sich schnell in systematischere Darstellungen und Handbücher hineinfinden wird.

Das Buch wendet sich an Studenten der Mathematik, der Physik und der Informatik vom dritten Semester an. Es setzt nur Grundkenntnisse aus der Analysis und der linearen Algebra voraus. Nur in einigen späteren Abschnitten würde man eigentlich ein wenig Maßtheorie brauchen. Die Aussagen lassen sich aber auch ohne solche weiter gehenden Vorkenntnisse verstehen, wenn man bereit ist, auf einzelne Beweise (vor allem von Existenzsätzen) zu verzichten. Diese sind in vertiefenden Vorlesungen leicht nachzuholen.

Das Buch enthält mehr Stoff als man bei angemessenem Tempo in einer vierstündigen Vorlesung vermitteln kann. Dies gibt Wahlmöglichkeiten. Die relativ zahlreichen mit einem Stern versehenen Abschnitte, Sätze und Beispiele und die Anhänge können am leichtesten weggelassen werden. Jedenfalls werden sie später nicht unbedingt benötigt.

Allerdings sind darunter viele Rosinen, so dass vieles dafür spricht, lieber einen Teil des Kuchens nicht zu essen.

Vieles aus den ersten Paragraphen ist Schulstoff. Weil Anfänger mit der mathematischen Modellierung realer Experimente oft Schwierigkeiten haben, scheint mir eine ausführliche Darstellung nicht nur für die zukünftigen Lehrer sinnvoll.

Man kann im Prinzip den gesamten Statistikteil auf eine spätere Lehrveranstaltung verschieben, aber Mathematikstudenten mit anderen Studienschwerpunkten und Physikstudenten fehlt oft die Zeit, eine solche zu besuchen.

Wie bei Lehrbüchern üblich habe ich die Quellen in der Regel nicht genannt. Es gibt aber historische Hinweise und Hinweise zum Weiterlesen.

Ich möchte den vielen Mitarbeitern und Freunden herzlich danken, die bei der Entstehung dieses Buches geholfen haben. Petra Küster hat schon bei der Ausarbeitung des Skripts mitgewirkt, das als Grundlage diente. Aus Vorlesungsnotizen von Götz Kersting habe ich manche Anregung geschöpft. Erich Berger, Wolfgang Stadje, Götz Kersting, Uwe Rösler, Hans-Jürgen Döring, Ulrich Wacker, Catherine Pallenberg, Norbert Neumann, Herold Dehling und Heinrich Hering haben Teile des Manuskripts gelesen, und ihre Vorschläge haben zu wesentlichen Verbesserungen geführt. Michael Krawczak hat das schöne Titelbild beigetragen. Das Manuskript haben Frau Schrörs, Frau Zimmer, Frau Graupner, Frau Gieseking und Frau Steffen sehr einwandfrei getippt. Dem Vieweg-Verlag, insb. Frau Schmickler-Hirzebruch, danke ich für die gute Zusammenarbeit. Meiner Frau danke ich für ihr Verständnis dafür, dass ich oft selbst sonntags so schwer vom Schreibtisch wegzukriegen war.

Die fünfte Auflage enthält u.a. einen zusätzlichen Paragraphen über Laufzeitanalysen für rekursive Algorithmen. Ich danke Herrn Uwe Rösler und Herrn Ludger Rüschendorf für wertvolle Hinweise zu diesem aktuellen Thema. Der Abschnitt über nichtparametrische Tests wurde deutlich erweitert.

Herrn Erich Berger danke ich für die sorgfältige Herstellung der neuen Druckvorlage mit LATEX und für unzählige Verbesserungsvorschläge, die ich gerne aufgegriffen habe.

Ich widme dieses Buch meinem Lehrer Konrad Jacobs, der mein Interesse an Stochastik geweckt hat und dem ich viel verdanke.

Göttingen, im Oktober 1999. *Ulrich Krengel*

Zur achten Auflage

Die achte Auflage enthält erweiterte historische Anmerkungen.

Göttingen, im Juli 2005. *Ulrich Krengel*

Kapitel I Diskrete Wahrscheinlichkeitsräume

Zunächst sollen die wichtigsten Grundbegriffe für solche Zufallsexperimente entwickelt werden, die nur endlich viele oder abzählbar viele mögliche Versuchsausgänge haben. Für sie benötigt man weniger theoretischen Hintergrund.

§ 1 Modelle für Zufallsexperimente, Abzählmethoden

Ziel der Wahrscheinlichkeitstheorie ist die Analyse der Gesetzmäßigkeiten, die bei der Beschreibung so genannter „Zufallsexperimente" eine Rolle spielen. Darunter verstehen wir Experimente, deren Ausgänge nicht durch logische oder andere Gründe durch die Versuchsbedingungen determiniert sind. Wenigstens gedanklich sollten die Experimente unter den gleichen Bedingungen wiederholbar sein, und zwar so, dass der Versuchsausgang bei unabhängig angestellten Wiederholungen nicht notwendig stets der gleiche ist, sondern nur statistischen Regelmäßigkeiten folgt.

Beispiele

- Würfelexperimente
- Blindes zufälliges Ziehen von Kugeln oder Losen aus einer Urne
- Kartenspiele

Die gleichen Gesetzmäßigkeiten treten auch bei „Experimenten" ganz anderen Typs auf, etwa

- Geburten (Junge oder Mädchen)
- Ermittlung der Anzahl der Ausschussexemplare in der Tagesproduktion einer Maschine
- Unfallstatistiken
- Registrierung von Lebensdauern
- Anzahl der Ausschläge eines Geiger-Zählers zur Messung von radioaktiver Strahlung.

Dagegen ist die Frage, wie „wahrscheinlich" es ist, dass auf dem Mars einmal Leben existiert hat, von ganz anderer Art. Entweder ist die richtige Antwort ja oder nein, unabhängig davon, ob wir sie kennen oder nicht.

Um nun Gesetzmäßigkeiten in Zufallsexperimenten mathematisch untersuchen zu können, muss man zunächst *mathematische Modelle* dafür bilden, innerhalb derer sie sich rein mathematisch beschreiben und „definieren" lassen. Diese Modellbildung ist ein Vorgang, der auch bei anderen mathematischen Disziplinen auftaucht, etwa in der Geometrie. Zum Beispiel ist eine Kugel für viele Zwecke ein passendes Modell für die Erde. Will man etwa eine Kirchturmhöhe aus einer Längenmessung und aus Winkelmessungen bestimmen, so ist ein Dreieck ein Modell.

Der Übergang von der Wirklichkeit zum Modell ist nie rein logisch begründbar. Er setzt in starkem Maß *Erfahrung* über die Natur des Experiments voraus. Das ist keine Besonderheit der Modelle für Zufallsexperimente. So genügt das ebene Modell der Erdoberfläche vollauf, wenn man eine Landkarte des Landkreises Göttingen herstellen will. Für feine geophysikalische Betrachtungen ist selbst das Modell der Erdkugel zu grob und man betrachtet im feineren Modell Abplattungen. Wir sehen daran auch gleich, dass die Wahl des Modells von der Zielsetzung mitbestimmt wird.

Wir halten also fest, dass es keinen *prinzipiellen* Unterschied zwischen den Rollen, die Modelle in Geometrie und Stochastik spielen, gibt. *Praktisch* scheint es aber Unterschiede zu geben, weil das passende Modell für Zufallsexperimente oft weniger offensichtlich ist. Man täuscht sich leicht darüber, ob ein Versuch wirklich hinreichend viele Symmetrien enthält, um die Annahme zu rechtfertigen, alle Versuchsausgänge seien gleichwahrscheinlich. Häufig lässt sich die Angemessenheit eines Modells nur empirisch prüfen, und das ist stets mit Unsicherheiten behaftet, die mit dem zufälligen Ausgang der Experimente zusammenhängen. Häufiger als in anderen Gebieten wird man schon aus Gründen der mathematischen Praktikabilität bewusst ein relativ grobes Modell in Kauf nehmen müssen. Gelegentlich wird man sogar ohne genügend Erfahrung über die Natur des Experiments ad hoc Modelle entwerfen, um gewisse Phänomene überhaupt erst einmal einer Rechnung zugänglich zu machen, und um zu vergleichen, welche qualitativen Auswirkungen verschiedene Modellannahmen haben.

Zeitweise haben Mathematiker sogar geglaubt, es läge im Wesen der zufälligen Erscheinungen, dass sie sich eben nicht mathematisieren ließen. Jedenfalls hat es — im Gegensatz z.B. zur Geometrie — bis ins zwanzigste Jahrhundert hinein gedauert, bis man eine gesicherte axiomatische Grundlegung gegeben hat. Andererseits macht gerade dieses Phänomen, dass man über Zufallsereignisse mathematisch rigorose Resultate beweisen kann, einen Reiz des Gebietes aus.

Ist man sich der Schwierigkeit der Modellbildung in der Stochastik bewusst, so wird man die Anwendbarkeit der abgeleiteten Resultate auf reale Probleme auch nachträglich noch überprüfen müssen. Eine Art Rückkopplung kann hilfreich sein. Wenn die abgeleiteten Resultate nicht gut mit der Erfahrung übereinstimmen, wird das Modell revidiert werden müssen.

Die Frage der Modellbildung wird für uns ein wiederkehrendes Leitthema sein, das mit der Entwicklung der mathematischen Theorie in Wechselwirkung steht.

1.1 Endliche Wahrscheinlichkeitsräume

Beginnen wir mit einer kleinen Aufgabe: Es sei die Wahrscheinlichkeit dafür zu bestimmen, dass die Summe der bei zwei Würfen eines Würfels erhaltenen Augenzahlen mindestens 10 ist.

Wir können das Ergebnis des aus zwei Würfen bestehenden Zufallsexperiments durch das Paar (i, k) der beobachteten Augenzahlen beschreiben. Z.B. bedeutet $(4, 1)$, dass der erste Wurf eine 4 und der zweite eine 1 ist. Die Menge $\Omega = \{(i,k) : 1 \leq i,k \leq 6\}$ aller möglichen Ergebnisse hat 36 Elemente. Aus Symmetriegründen ist es nahe liegend, sie als gleichwahrscheinlich anzusehen. Jedes $(i,k) \in \Omega$ hat also die Wahrscheinlichkeit $1/36$. Die Menge der Ergebnisse, für die die Summe $i+k$ der Augenzahlen mindestens 10 ist, ist

$$A = \{(6,6), (6,5), (5,6), (6,4), (5,5), (4,6)\}.$$

Da A sechs Elemente hat, ist die gesuchte Wahrscheinlichkeit also $6/36 = 1/6$.

Wir wollen auch Experimente betrachten, für die nicht alle möglichen Ergebnisse gleichwahrscheinlich sind. Werfen wir eine Reißzwecke auf den Fußboden, so landet sie entweder mit der Spitze nach oben oder mit der Spitze schräg nach unten. Bezeichnen wir das erste dieser möglichen Ergebnisse mit o und das zweite mit u, so ist $\Omega = \{o, u\}$ eine Beschreibung der Ergebnismenge. Die Annahme, o und u seien gleichwahrscheinlich, wäre sehr unplausibel. Man könnte versuchen, die Wahrscheinlichkeit p von o durch Ausführung vieler Würfe zu schätzen.

Allgemein ordnen wir einem Zufallsexperiment mit endlich vielen möglichen Ausgängen eine endliche nichtleere Menge Ω zu, deren Elemente ω die Versuchsausgänge bezeichnen. Sie heißen **Ergebnisse** oder oft auch **Stichproben, Realisierungen** oder **Elementarereignisse.** Ω heißt **Ergebnismenge** oder **Stichprobenraum,** gelegentlich auch Grundraum oder Ergebnisraum.

Die Teilmengen von Ω sind die **Ereignisse,** die in unserem Modell in Betracht gezogen werden. Genauer: Wir identifizieren $A \subset \Omega$ mit dem Ereignis, dass ein $\omega \in A$ der beobachtete Versuchsausgang ist. Diese Konvention gestattet es, mengentheoretische Notationen einzusetzen. So bedeutet $A \cap B$ das Ereignis, dass sich A *und* B ereignen, denn $\omega \in A \cap B$ besagt, dass ω in A und in B liegt. Offenbar ist $A \cup B$ das Ereignis, dass sich A *oder* B ereignet. (Das Wort „oder" ist immer im nicht ausschließenden Sinn zu verstehen.) Das Komplement A^c von A in Ω bezeichnet das Ereignis, dass A nicht geschieht. Ereignisse A und B heißen **unvereinbar,** wenn die Mengen A und B disjunkt sind, d.h. wenn ihr Durchschnitt $A \cap B$ die leere Menge $\emptyset$ ist. $\emptyset$ heißt auch das **unmögliche Ereignis.** Ω heißt das **sichere Ereignis.**

Nun müssen den Ereignissen noch Wahrscheinlichkeiten zugeordnet werden. Die Menge der Ereignisse ist mengentheoretisch die Potenzmenge $\mathcal{P}(\Omega)$, d.h. die Menge aller Teilmengen von Ω. Eine Abbildung P von $\mathcal{P}(\Omega)$ in $[0,1]$ heißt **Wahrscheinlichkeitsverteilung** oder **Wahrscheinlichkeitsmaß,** wenn sie die folgenden Eigenschaften hat:

$$P(\Omega) = 1; \tag{1.1}$$

$$P(A) \geq 0 \quad \text{für alle } A; \tag{1.2}$$

$$P(A \cup B) = P(A) + P(B) \quad \text{für alle disjunkten } A, B. \tag{1.3}$$

Ist (1.1) erfüllt, so nennt man P **normiert.** Die letzte der obigen drei Eigenschaften heißt **Additivität** von P. $P(A)$ heißt **Wahrscheinlichkeit** von A. Das Paar (Ω, P) heißt der dem Experiment zugeordnete **Wahrscheinlichkeitsraum.** Er dient uns als Modell für das reale Experiment. Die Wahl von Ω ist oft nahe liegend. Die „richtige" Wahl von P ist nur in wenigen Fällen klar und wird uns noch viel beschäftigen. Die oben genannten „Axiome" (1.1)–(1.3) stecken nur einen Rahmen ab.

Indem wir Wahrscheinlichkeiten nur im Modell definiert haben, sind wir der Frage, was Wahrscheinlichkeiten realer Ereignisse sind, aus dem Weg gegangen. Man versteht wohl allgemein die Wahrscheinlichkeit eines realen Ereignisses als Maß für die Sicherheit oder Häufigkeit, mit der es bei wiederholter Ausführung des Experiments auftreten wird.

Sei etwa A beim Würfeln mit einem Würfel das Ereignis, dass die geworfene Augenzahl höchstens 4 ist. Wirft man den Würfel n-mal, so unterscheiden wir zwischen der **absoluten Häufigkeit** $k_n(A)$, also der Zahl der Würfe unter diesen n Würfen, bei denen A auftritt, und der **relativen Häufigkeit** $h_n(A) = k_n(A)/n$. Empirisch beobachtet man, dass sich die $h_n(A)$ für sehr große n einem Grenzwert $P(A)$ annähern. (Konvergenz lässt sich natürlich empirisch nie prüfen, da wir dazu eine unendliche Folge von Würfen durchführen müssten.) Nehmen wir an, dass für alle Ereignisse A Konvergenz von $h_n(A)$ gegen $P(A)$ vorliegt, so überlegt man sich leicht, dass P die Eigenschaften (1.1)–(1.3) haben muss.

Man hat daher Versuche unternommen, Wahrscheinlichkeiten mit Hilfe von relativen Häufigkeiten zu definieren. Aber dies führt zu großen Schwierigkeiten. Relative Häufigkeiten hängen von der konkret beobachteten Folge von Würfen ab und sind deshalb zufällig. Begrifflich sind damit relative Häufigkeiten sehr verschieden von Wahrscheinlichkeiten. Wir werden später, in § 3, umgekehrt einen Zusammenhang von Wahrscheinlichkeiten und relativen Häufigkeiten herstellen, indem wir ausgehend von dem axiomatischen Modell Konvergenz von relativen Häufigkeiten in geeignetem Sinn beweisen. Dafür ist es aber jetzt noch zu früh.

Nun zurück zu unserem mathematischen Modell. Wir ersparen uns den ziemlich trivialen Beweis der folgenden

Eigenschaften von P: *Für $A, B, A_i \in \mathcal{P}(\Omega)$ gilt*

$$P(A^c) = 1 - P(A), \quad \textit{speziell} \quad P(\emptyset) = 0\,; \tag{1.4}$$

$$A \subset B \;\Rightarrow\; P(A) \leq P(B) \quad (\Rightarrow \textit{ bezeichnet die logische Implikation}); \tag{1.5}$$

$$P(A \setminus B) = P(A) - P(A \cap B) \quad (\textit{mit } A \setminus B = A \cap B^c); \tag{1.6}$$

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i)\,, \quad \textit{falls } A_1, \ldots, A_n \textit{ paarweise disjunkt sind;} \tag{1.7}$$

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} P(A_i) \quad \textit{für beliebige } A_1, \ldots, A_n\,; \tag{1.8}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,. \tag{1.9}$$

Aus (1.7) folgt

$$P(A) = \sum_{\omega \in A} P(\{\omega\}). \tag{1.10}$$

Die Wahrscheinlichkeit eines Ereignisses A ist die Summe der Wahrscheinlichkeiten der Ergebnisse, bei denen A eintritt. P ist also durch die Werte aller $P(\{\omega\})$ mit $\omega \in \Omega$ bestimmt. Wir schreiben auch $P(\omega)$ statt $P(\{\omega\})$. Die Abbildung $\omega \mapsto P(\omega)$ heißt **Wahrscheinlichkeitsfunktion.** Die Wahrscheinlichkeitsverteilung kann also durch Angabe der zugehörigen Wahrscheinlichkeitsfunktion beschrieben werden. Offenbar gilt für eine Wahrscheinlichkeitsfunktion stets

$$P(\omega) \geq 0 \text{ für alle } \omega \in \Omega, \text{ und } \sum_{\omega \in \Omega} P(\omega) = 1\,. \tag{1.11}$$