Stochastik 1 für Studierende der Informatik

Hendrik Baumann
hendrik.baumann@uni-hamburg.de

Sommersemester 2016

Kapitel 1

Einleitung

Die *Stochastik* beschäftigt sich mit der Beschreibung und Analyse zufälliger Ereignisse und Prozesse und ihrer Einbettung in mathematische Modelle. Etymologisch kommt der Begriff Stochastik aus dem Griechischen und bedeutet etwa „die Kunst des geschickten Vermutens“.

1.1 Historische Anfänge

Im Jahr 1654 stellte der Chevalier de Méré stellt Blaise Pascal (1623-1662) Fragen bezüglich Glücksspielen. In diesen ging es um Gewinnchancen, faire Einsätze oder auch Ruinwahrscheinlichkeiten. Pascal begann daraufhin einen Briefwechsel mit Pierre de Fermat (1601-1665), der Antworten fand. In diesem Briefwechsel kam der Begriff „Wahrscheinlichkeit“ noch nicht vor.

Ein überliefertes Beispiel ist der Abbruch eines mehrrundigen Glücksspiels: Zwei Spieler (A und B) werfen neunmal eine faire Münze. Fällt in einer Runde Zahl, so gewinnt A diese, fällt Bild, so gewinnt B die Runde; Gesamtsieger ist, wer zuerst fünf Runden für sich entschieden hat. Der Anfangseinsatz geht vollständig an den Gesamtsieger. Durch höhere Gewalt muss das Spiel nach sieben Runden beim Stand von 4:3 für A abgebrochen werden. Wie soll der Einsatz aufgeteilt werden?

Für damalige Spieler war es naheliegend, den Einsatz

- im Verhältnis der gewonnenen Runden, also 4:3,
- oder umgekehrt zum Verhältnis der noch zum Gesamtsieg benötigten gewonnenen Runden, also 2:1,

aufzuteilen. Pascal und Fermat kamen auf die Idee, die Gewinnchancen zu ermitteln, und den Einsatz im Verhätnis der Gewinnchancen aufzuteilen. Spieler B kann nur gewinnen, wenn in den letzten beiden Runden jeweils Bild fällt. Die Erfahrung zeigte: Wird eine Münze sehr oft geworfen, so fällt in etwa der Hälfte der Fälle Bild. Wird sehr oft eine Münze zweimal hintereinander geworfen, so fällt in etwa einem Viertel der Fälle zweimal Bild. Vom Spielstand 4:3 aus startend wird bei sehr häufiger Wiederholung der letzten zwei Würfe Spieler B noch in etwa einem Viertel aller Fälle gewinnen, Spieler A in den restlichen Fällen, also in etwa drei Viertel aller Spiele mit Zwischenstand 4:3 nach sieben Runden. Die Aufteilung sollte also im Verhältnis 3:1 erfolgen.

Pascal und Fermat gingen davon aus, dass die Glücksspielsituation

- zwar unterschiedliche Ausgänge haben kann,
- aber im Prinzip beliebig oft wiederholbar ist.

Durch diese Eigenschaften werden *Zufallsexperimente* charakterisiert. Die Ausgänge eines Zufallsexperiment werden *Ereignisse* genannt (dieser Begriff wird später sauber definiert).

Wird ein Zufallsexperiment n mal ausgeführt, lässt sich für ein fixiertes Ereignis die *relative Häufigkeit* bestimmen, d.h. es wird gezählt, wie oft das betrachtete Ereignis eingetreten ist, und diese Anzahl wird durch die Anzahl n der Experimente geteilt. Die Erfahrung zeigte Pascal und Fermat, dass sich die relative Häufigkeit für steigendes n stabilisiert, d.h. dass sie für $n \to \infty$ konvergiert.

Dieser Grenzwert wird seit Jacob Bernoulli (1654-1705) und Abraham de Moivre (1667-1754) als *Wahrscheinlichkeit* bezeichnet. Im Laufe der Zeit (bis 1931) gab es durch verschiedene Mathematiker (Fermat, Pascal, Huygens, Bernoulli, de Moivre, Bayes, Euler, Lagrange, Laplace, Legrende, Gauß, Poisson, Tschebyscheff, Lyapunov, Markov) wesentliche Fortschritte in der Wahrscheinlichkeitsrechnung. Trotzdem war sie innerhalb der Mathematik immer noch sehr umstritten, da sie sich scheinbar nicht mit der Unterscheidung zwischen wahren und falschen Aussagen beschäftigte.

In die heutige Form wurde die Wahrscheinlichkeitstheorie in den Jahren 1931-1933 durch Andrej Nikolajewitsch Kolmogorov (1903-1987) gebracht. Seine fundamentale Leistung war, Wahrscheinlichkeiten axiomatisch zu definieren, d.h. sie muss gewissen Axiomen folgen. Die ursprüngliche Motivation ging dabei nicht verloren, da die Axiome für Wahrscheinlichkeiten den Eigenschaften relativer Häufigkeiten nachempfunden sind. Durch diese Axiomatisierung wurde die Stochastik zu einer echten mathematischen Disziplin, und seitdem entwickelt sie sich rasant.

1.2 Zufall

Zusätzlich zu unklaren Begrifflichkeiten birgt die Stochastik für viele Menschen ein weiteres Problem: Für den menschlichen Beobachter wirken zum Beispiel

- das Ergebnis des Wurfes mit einer fairen Münze,
- die Wartezeit an einer Ampel,
- der Zeitpunkt des Zerfalls eines Atomkerns und
- der Börsenkurs einer Aktie

dem „Zufall“ unterworfen. Es stellt sich aber die wesentliche Frage: Was ist eigentlich Zufall?

Der Ausgang des Wurfs einer Münze hängt kausal, insbesonder auch deterministisch von verschiedenen Einflussfaktoren wie Wurfkraft, Effet, Raumklima, ... ab, und die physikalischen Gesetze für diese Kausalität sind prinzipiell auch bekannt. Ähnlich hängt die Wartezeit an einer Ampel kausal (deterministisch) und sogar vergleichsweise einfach von dem Verkehr und der Steuerung der Ampel ab (Länge der Grün- und Rotphasen, Kontaktschaltungen, ...) ab. Trotzdem wird in beiden Fällen von Glück und Pech, also Synonymen für Zufall gesprochen.

Im Fall des Münzwurfs, oder vergleichbar auch eines Würfelwurfs liegt ein Grund in der hohen Komplexität der physikalischen Zusammenhänge. Darüber hinaus sind die Einflussfaktoren für einen Spieler nicht hinreichend genau messbar. Im Beispiel der Wartezeit an einer Ampel sind die Zusammenhänge weniger komplex, aber zur Bestimmung der Wartezeit fehlen den aktiven Verkehrsteilnehmern die nötigen Informationen (andere Verkehrsteilnehmer und die Ampelsteuerung betreffend).

Insgesamt lässt sich feststellen, dass der Begriff Zufall umgangssprachlich (und auch in der mathematischen Modellierung, wie sie in dieser Vorlesung diskutiert wird) fast ausschließlich verwendet wird, wenn ein Ereignis zwar kausale Ursachen hat,

- aber diese Kausalität hochkomplex ist und Informationen über die Ursachen nicht oder nicht genügend genau messbar sind,
- die Ursachen nicht beobachtbar sind oder
- die genaue Form der Kausalität bisher nicht erfassbar ist.

In den beiden anderen Beispielen lässt sich darüber streiten, ob tatsächlich echter Zufall vorliegt. Beim Zerfall von Atomkernen gibt es gibt zumindest keine lokale Ursachen für die konkreten Zeitpunkte. Ob in der Quantenphysik echter (objektiver) Zufall existiert, führt auf eine physik-philosophische Diskussion. Der Börsenkurs ist stark von der Entscheidung (Kaufen und Verkaufen) anderer Marktteilnehmer abhängig. Inwieweit der freie Wille anderer Menschen einen echten Zufall darstell oder zumindest beinhalt, führt auf philosophische, theologische und psychologische Diskussionen.

1.3 Stochastische Modellierung

In der mathematischen Modellierung werden sehr allgemein eine oder mehrere Zielgrößen als Funktionen von Einflussfaktoren charakterisiert. Im deterministischen Fall liefern gleiche Werte der Einflussfaktoren gleiche Werte der Zielgrößen.

Im stochastischen Fall sind hingegen bei gleichen Werten der Einflussfaktoren verschiedene Werte der Zielgrößen möglich. Meist ist dies so zu interpretieren, dass die Zielgröße von weiteren, nicht beobachtbaren oder nicht messbaren Ursachen abhängt, diese werden als „Zufallseinfluss“ modelliert.

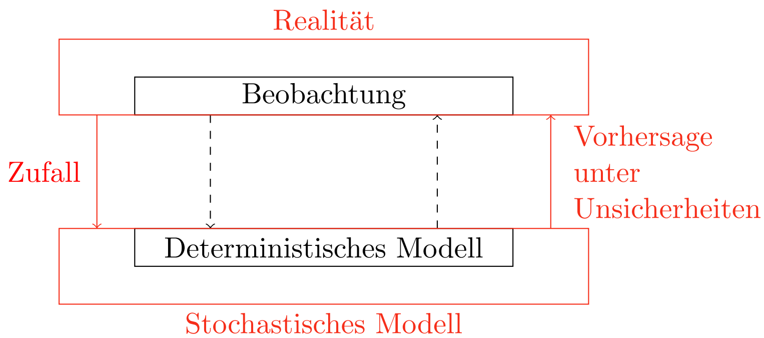
Bei der Modellierung dieser Zufallseinflüsse ist vor allem eine Quantifizierung des Zufalls nötig, dies wird einen wesentlichen Platz in dieser Veranstaltung einnehmen, und erfolgt anhand von *Verteilungstypen* und Parameter.

Um dies zu veranschaulichen werden zwei Experimente betrachtet, in denen jeweils die Zahlen von 0 bis 5 die möglichen Ausgänge sind.

- Experiment 1: Beschrifte fairen Würfel mit Zahlen 0 bis 5, betrachte Ergebnis eines einfachen Wurfs.

- Experiment 2: Werfe faire Münze fünfmal, zähle, wie häufig Zahl erscheint.

Im ersten Experiment erscheint die Annahme sinnvoll, dass jede ede Zahl mit der gleichen Wahrscheinlichkeit auftritt, also $\frac{1}{6}$. Im zweiten Beispiel hingegen sollten die Anzahlen unterschiedlich wahrscheinlich sein, beispielsweise ist die Wahrscheinlichkeit für fünfmal Zahl $\left(\frac{1}{2}\right)^5 = \frac{1}{32}$.



Eine wesentliche Aufgabe der Stochastik ist die *Quantifizierung von Unsicherheiten*

- bei der Modellierung der Zufallseinflüsse und
- bei der Vorhersage.

Durch Vergleich (testweise) vorhergesagter Daten und realer Daten sollten die stochastischen Modelle dann auch validiert werden können. Im Rahmen dieser Veranstaltung wird Basiswissen vermittelt, es werden Rechenregeln für Wahrscheinlichkeiten und wichtige Verteilungstypen hergeleitet, und es werden einige Grenzwertsätze (häufige Wiederholung eines Experiments) diskutiert. Weitere Bereiche der Stochastik sind

- Stochastische Prozesse: Zeitliche geordnete Prozesse mit Zufallseinflüssen (Markovketten, Brownsche Bewegung, Warteschlangentheorie,...).
- Stochastische Simulation: Nachbildung komplexer stochastischer Situationen auf dem Rechner.
- Statistik: Parameterschätzung, Modellvalidierung, Auswertung stochastischer Simulationen,...

Da nahezu jedes realistische Problem mit Zufallseinflüssen behaftet ist, sind die Einsatzgebiete der Stochastik entsprechend vielfältig. Unter anderem finden sich stochastische Modellierungen in folgenden Bereichen:

- Glücksspiele: Faire Einsätze, Gewinnchancen, ...
- Finanzmärkte: Optionspreisbewertungen,...
- Physik, Chemie: Kinetik, Teilchenbewegungen, molekulare Reaktionen, Versuchsplanung,...

- Biologie, Medizin, Pharmazie: Populationsentwicklungen, Ausbreitung von Epidemien, Medikamententests, Therapiebewertungen,...
- Informatik: Analyse von Algorithmen (zufällige Daten), randomisierte Algorithmen, Monte-Carlo-Methoden, Leistungsbewertung von Servern und Rechnernetzen,...
- Industrie und Dienstleistung: Verkehrsinfrastruktur, Logistik, Lagerhaltung, Produktionsprozesse, Callcenter,...
- ...

Beispiel 1.3.1. Bei der Gestaltung von Bahnfahrplänen ist eine Berücksichtigung der Zufallseinflüsse sinnvoll (oder sogar notwendig). Zunächst entstehen zufällig Verspätungen, diese plfanzen sich aufgrund des dicht befahrenen Schienennetzes und zu ermöglichenden Anschlussmöglichkeiten fort und verstärken sich teilweise. Wenn die Quellverspätungen und die Verspätungsfortpflanzung quantifiziert sind, kann ein robuster Fahrplan entwickelt werden, d.h. ein Fahrplan, bei dem sich häufig auftretende Verspätungen nicht stark fortpflanzen.

Beispiel 1.3.2. In einem Service-Callcenter rufen Kunden zu zufälligen Zeitpunkten an, die Gesprächsdauern hängen von der konkreten Anfrage ab, und sind ebenfalls zufällig. Sind keine Service-Mitarbeiter verfügbar, so muss ein Kunde später noch einmal anrufen, und wird (spätestens nach mehreren vergeblichen Versuchen) unzufrieden. Umgekehrt sind die Betreiber solcher Callcenter an einer möglichst hohen Auslastung ihrer Mitarbeiter interessiert. Auch hier müssen zunächst die Zufallseinflüsse quantifiziert werden, außerdem sollte ein sinnvolles LoS (Level of Service) definiert werden. Beispielsweise wird eine (hohe) Wahrscheinlichkeit vorgegeben, mit der ein anrufender Kunde innerhalb einer Minute mit einem Mitarbeiter verbunden wird. Die Analyse des Modells soll schließlich dazu führen, dass eine (minimale) Anzahl von Mitarbeitern gefunden wird, mit der dieses Ziel erreicht werden kann.

Beispiel 1.3.3. Auch Produktionsprozesse unterliegen vielfältigen stochastischen Einflüssen (Qualität der Grundwerkstoffe, Fehler durch Mensch und Maschinen, ...). Dadurch müssen die Zeiten, die Werkstücke in einzelnen Produktionsabschnitten verbringen, als zufällig angesehen werden (beispielsweise kann eine Nachbearbeitung erforderlich werden oder eine Maschine braucht wegen eines Materialfehlers für ihre Aufgabe einfach länger). In der Konsequenz bauen sich Warteschlangen zwischen den Produktionsstationen auf. Dies hat zur Folge, dass Warteplatz in der Produktion eingeplant werden muss, die Werkstücke binden Kapital (sie sind „totes Kapital“), und es gibt zeitliche Unregelmäßigkeiten der Produktion. Auch hier müssen zunächst die Zufallseinflüsse und ihre Fortpflanzung quantifiziert werden. Anschließend sollte ein Produktionsprozess gestaltet werden, bei dem möglichst wenig Warteplatz benötigt wird, und die Durchlaufzeit eines Produktes von Produktionsbeginn bis Fertigstellung möglichst gering ist.

So wichtig der Einsatz von Stochastik in vielen praktischen Problemstellungen ist, so groß sind auch die Probleme bei der Vermittlung stochastischer Resulatate. Beispielsweise möchten Anwender häufig konkrete Vorhersagen für die Durchführung eines einzelnen Experimentes. Dies ist bei zufälligen Einflüssen natürlich nicht möglich; die Aussagen der Wahrscheinlichkeitstheorie haben für Anwendungen nur bei häufiger Wiederholung oder einer langen Laufzeit Sinn.

Hinzu kommt noch, dass einige Aussagen der Stochastik zu kontraintuitiven Ergebnissen führen. Hierzu wird es im Rahmen dieser Veranstaltung Beispiele geben.

Kapitel 2

Diskrete Wahrscheinlichkeitsräume – Grundbegriffe

2.1 Ergebnisse und Ereignisse

Im Beispiel vom Chevalier de Méré aus der Einleitung wurde das Münzwurfspiel abgebrochen, wenn noch zwei Runden ausstehen. Aufgrund des Spielstandes von 4:3 für Spieler A gewinnt Spieler B genau dann, wenn bei den letzten beiden Würfen Bild erscheint. Hier gibt es zwei mögliche stochastische Modellierungen:

- Das Experiment wird mit zwei Ausgängen modelliert, nämlich „A gewinnt" oder „B gewinnt", die Wahrscheinlichkeiten sind dann $\frac{3}{4}$ bzw. $\frac{1}{4}$.
- Das Experiment wird mit vier Ausgängen modelliert, nämlich (Zahl,Zahl), (Zahl,Bild), (Bild,Zahl), (Bild,Bild). Hier sind die Wahrscheinlichkeiten jeweils $\frac{1}{4}$.

Wird die zweite Modellierung gewählt, so sind die ersten drei Ausgänge günstig für A, der letzte für B. Welche Modellierung sollte gewählt werden?

Ein weiteres Beispiel für diese Fragestellung ergibt sich, wenn beim Wurf mit zwei fairen Würfeln die Augensumme von Interesse ist. Hier gibt es sogar drei plausible Modellierungen:

- Das Experiment wird mit 11 Ausgängen, nämlich den Zahlen $2, 3, \dots, 12$ modelliert.
- Das Experiment wird mit 21 Ausgängen, den Paaren (i, j) mit $1 \leq i \leq j \leq 6$ (die Würfelergebnisse der Größe nach aufsteigend geordnet), modelliert.
- Das Experiment wird mit 36 Ausgängen, den Paaren $(i, j) \in \{1, \dots, 6\}^2$ (gedanklich: ein roter und ein grüner Würfel, die Augenzahl des roten Würfels wird zuerst aufgeführt), modelliert.

Bei der zweiten und dritten Modellierung beinhaltet das Ereignis „7 gewürfelt" mehrere Ausgänge des Zufallsexperiments. Auch hier stellt sich wieder die Frage, welche Modellierung die richtige ist.

Tatsächlich sind alle Modellierungen möglich, aber in Hinblick, Wahrscheinlichkeiten zuzuordnen, ist in den beiden Beispielen (und auch in vielen weiteren Situationen) die jeweils detaillierteste (d.h. die letzte) Modellierung die beste. Interessante Ereignisse können dann aus mehreren möglichen Ausgängen der detaillierten Modellierung zusammengesetzt werden. Darüber hinaus bieten detaillierte Modelle den Vorteil, dass sie für ähnliche Fragestellungen wiederverwendet werden können.

Definition 2.1.1. Die möglichen Ausgänge eines Zufallsexperiments heißen *Ergebnisse* oder *Elementarereignisse*. Die Menge aller Ergebnisse heißt *Ergebnismenge* oder *Ergebnisraum*.

Besonders gebräuchlich ist die Bezeichnung Ω für Ergebnismengen. Einfache Beispiele sind

- der Wurf mit einem fairen Würfel: $\Omega = \{1, \dots, 6\}$.
- der Wurf mit zwei fairen Würfeln: $\Omega = \{1, \dots, 6\}^2$.
- die Anzahl wartender Kunden an einem Schalter: $\Omega = \mathbb{N}_0$.
- die Wartezeit an einer Ampel: $\Omega = [0, \infty)$.

Die Mächtigkeit der Ergebnismengen in den einfachen Beispielen unterscheidet sich wesentlich. Es wird sich herausstellen, dass Ergebnismengen mit höchstens abzählbar unendlich vielen Elementen einfacher zu behandeln sind.

Definition 2.1.2. In der Stochastik heißt eine Menge mit endlich vielen oder abzählbar unendlich vielen Elementen *diskret*.

Achtung: Der Begriff „diskret“ ist in der Mathematik nicht eindeutig.

Ergebnisse geben also den detaillierten Ausgang eines Experimentes wieder. Wird beim Wurf mit zwei fairen Würfeln $\Omega = \{1, \dots, 6\}^2$ gesetzt, so umfasst die Formulierung „Augensumme 7 gewürfelt“ umfasst die Ergebnisse $(1, 6), (2, 5), \dots, (6, 1)$, d.h. „Augensumme 7 gewürfelt“ ist eine *Menge von Ergebnissen*.

Definition 2.1.3. Ist $\Omega \neq \emptyset$ eine diskrete Ergebnismenge, so heißen die Teilmengen von Ω *Ereignisse*. Die Menge aller Ereignisse, also die Potenzmenge, wird mit 2^Ω bezeichnet.

Rein formal sind Ergebnisse keine Ereignisse: Beim Würfelwurf ist $(1, 6)$ ein Ergebnis, die Menge $\{(1, 6)\}$ ist ein Ereignis, da sie eine einelementige Teilmenge der Ergebnismenge ist.