Einführung in die Stochastik

für Studierende der Informatik im Bachelorstudiengang

TU Braunschweig

SS 2007

Dr. Lothar Schüler

Institut für Mathematische Stochastik
Technische Universität Braunschweig
Pockelsstr. 14 * 38106 Braunschweig
Tel. 0531-391-7569 (7567) * Telefax 391-7564
EMAIL: L.Schueler@TU-BS.DE

Homepages:

http://www.math.tu-bs.de/stochastik/schueler.htm
http://www.math.tu-bs.de/stochastik/vorl/EinSt.htm

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1

Einführung

In der **Stochastik** beschäftigt man sich mit der Mathematik des **Zufalls**, d.h. man versucht ein mathematisches Kalkül zu entwickeln, das es gestattet, "zufällige Phänomene" zumindestens als Massenerscheinungen zahlenmäßig zu beschreiben. Dabei wird die Stochastik üblicherweise grob in die beiden Teilgebiete

- **Wahrscheinlichkeitstheorie**:
  Mathematische Modelle und deren Beziehungen zueinander, und
- **Statistik**:
  Anpassung der Modelle an die Realität,

aufgegliedert.

Hier werden wir uns nur mit den Methoden der Wahrscheinlichkeitstheorie beschäftigen. Dabei werden so weit möglich Begriffsbildungen vermieden, die maßtheoretische Überlegungen erfordern.

1.1 Modellbildung

Methoden der Stochastik finden heute in vielen Bereichen Anwendung, wie etwa in der Informatik (Analyse von Algorithmen, Informationstheorie, Warteschlangen, stochastische Betriebssysteme, . . .), in den Ingenieurwissenschaften und der Betriebswirtschaft (Statistik, Entscheidungstheorie, Qualitätskontrolle, . . .) oder in der Medizin (Statistik, Signal– und Bildverarbeitung).

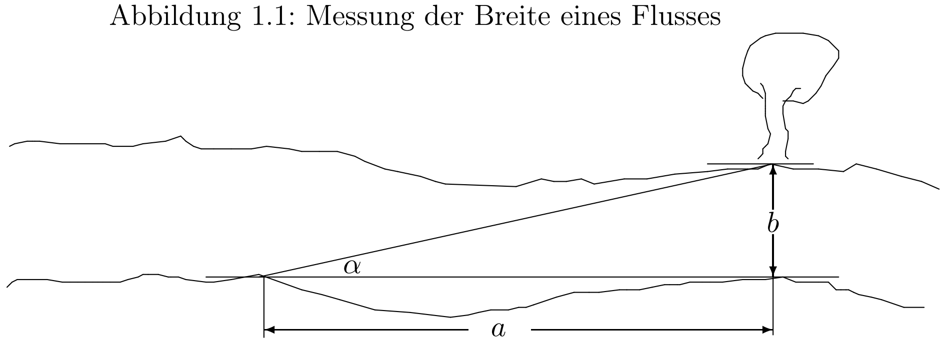
Zugrunde liegt diesen Modellen immer die Wahrscheinlichkeitstheorie. Sie ist ein formales Modell eines bestimmten Ausschnitts der Wirklichkeit. Derartige Modelle werden durch Abstraktion bestimmter Aspekte eines meist fachwissenschaftlich aufbereiteten Sachverhalts der Realität gewonnen. Im mathematischen Modell müssen logische Schlüsse gezogen werden, die ggfls. unter

Ausnutzung zusätzlich aus der Realität gewonnener Daten Aussagen in diesem Modell ermöglichen. Diese Aussagen müssen dann wieder in der Realität interpretiert werden.

Manche klassische mathematischen Modelle (z. B. natürliche Zahlen) sind uns so vertraut, dass man den Unterschied zwischen Modell und Wirklichkeit kaum noch erfährt.

Natürliche Zahlen bilden das "natürliche" Modell, wenn es um Anzahlen geht. Interessieren dagegen Fragen wie "Länge", "Abstand", etc., wird man meist die euklidische Geometrie benutzen, die wohl historisch aus der Landvermessung (Nilüberflutungen in Ägypten) entstanden ist. Die Rolle des mathematischen Modells kann man z.B. bei der Messung der Breite eines Flusses in folgende 4 Schritte unterteilt denken:

Abbildung 1.1: Messung der Breite eines Flusses



1. Abgrenzung und Präzisierung der Fragestellung durch Begriffe und Methoden der jeweiligen "Fachwissenschaft".
   Hier: Mit Methoden der Landvermessung wird festgelegt, was "Breite des Flusses" bedeutet, welche Genauigkeit benötigt wird und vieles andere mehr.
   Mit geeigneten fachwissenschaftlichen Methoden werden Daten ermittelt.
   Hier: Messen des Winkels α und der Entfernung a.

2. Als Mathematisches Modell kommt die analytische Geometrie der Ebene in Frage. Mit den Begriffen "Strecke", "Winkel" usw. kann eine rein mathematische Beschreibung der obigen Situation gebildet werden. In dieser Beschreibung kommen die fachspezifischen Ausdrücke wie "Fluss" oder "Ufer" nicht mehr vor.

3. Die analytische Geometrie liefert nun den mathematischen Schluss $b = a \cdot \tan\alpha$. Dieser Aussage kommt im Rahmen des gewählten mathematischen Modells (Axiome der analytischen Geometrie der Ebene) *absolute* mathematische Gültigkeit zu.

4. Zur Anwendung der mathematischen Theorie ist eine "Rückinterpretation" der mathematischen Aussagen in die Realität notwendig.
   Hier: Die Breite des Flusses beträgt $\tan(\text{Messwert } \alpha) \times \text{Messwert } a$.

Bei der Rückinterpretation muss man beachten, dass nur diejenigen Tatsachen berücksichtigt werden können, die zuvor in das gewählte mathematische Modell einbezogen worden sind. So bleiben hier Änderungen des Wasserstandes oder bei großen Flüssen die Kugelgestalt der Erde unberücksichtigt. Deshalb kann die Gültigkeit eines mathematischen Schlusses stark eingeschränkt sein.

Bei der Wahl des mathematischen Modelles kommt es daher darauf an, die für den Sachverhalt "wichtigen" Aspekte herauszufiltern und auf mathematische Objekte abzubilden.

Generell ist ein Modell zur Erklärung eines empirischen Sachverhalts soweit geeignet, wie die Modellaussagen bei ihrer Rückinterpretation ein sinnvolles Handeln in der Realität ermöglichen. Daher ist es eine wichtige Aufgabe eines verantwortungsbewussten Informatikers, sich mit der Eignung bzw. den Grenzen der Anwendbarkeit formaler Modelle kritisch auseinanderzusetzen.

Welche "Ausschnitte" aus der Realität können mit Hilfe der Wahrscheinlichkeitstheorie (**WT**) modelliert werden?

Es sind Situationen, in denen Aussagen über Ereignisse verlangt werden, über deren Eintritt keine Gewissheit besteht. Dabei kann die "Ungewissheit" auf Faktoren beruhen, die die Situation beeinflussen, über die aber keine (vollständige) Information vorliegt. Dafür sind verschiedene Ursachen denkbar:

- nicht alle Faktoren sind bekannt, u.U. gibt es keine vollständige fachwissenschaftliche Erklärung des beobachteten Phänomens wie bei vielen biologischen und medizinischen Vorgängen;
- Messungen stoßen an prinzipielle Grenzen (physikalische, ethische, gesetzliche, ...);
- Aufwand und Kosten sind zu hoch.

Derartige Situationen wollen wir als "zufallsabhängig" oder kurz als **Zufallsexperiment** bezeichnen.

Dabei kann die Wahrscheinlichkeitstheorie keine neuen Informationen über ungewisse Sachverhalte "erfinden". Sie kann nur bereits vorliegende Informationen *verdichten* und *umformen*. Die Ungewissheit kann dadurch gemildert aber nicht beseitigt werden. Auch wenn wir wissen, dass die Chance beim Würfeln keine Sechs zu würfeln fünfmal so hoch ist, wie die Chance eine Sechs zu würfeln, bleibt die Ungewissheit, welche Augenzahl sich beim nächsten Würfeln ergibt.

Sinnvolle Aussagen dieser Art sind meist dann erzielbar, wenn es sich nicht um einmalige Ereignisse, sondern um "Massenphänomene" handelt, wenn wir also eine Reihe von vergleichbaren Situationen betrachten können.
Als Ausgangspunkt für mathematische Schlüsse wird das beobachtete Auftreten der Ereignisse in der Vergangenheit benutzt. Daraus wird typischerweise auf das Auftreten in der Zukunft geschlossen, zumindest auf das Auftreten in einer langen Reihe von gleichartigen zukünftigen Situationen.

So meint die obige Aussage zur Chance, eine Sechs zu würfeln, dass in einer sehr langen Reihe von Würfen das Ereignis "Augenzahl ≤ 5" etwa fünfmal häufiger eintritt wie das Ereignis "Augenzahl $= 6$". Gewonnen wird diese Aussage aus einer in diesem Fall offensichtlichen Übertragung der Beobachtung, dass alle Augenzahlen etwa gleichhäufig auftreten, auf zukünftige Würfelergebnisse.

Man kann also gewisse "stabile" Beobachtungen (hier: relative Häufigkeiten) machen und daraus Gesetzmäßigkeiten ableiten, die sich in einer großer Zahl von einzelnen, nicht vorhersehbaren Zufallsexperimenten "im Mittel durchsetzen".

Im allgemeinen werden nicht alle Umstände bei der Wiederholung eines Zufallsexperiments gleich sein. Mehrfaches Werfen eines Würfels wird man beispielsweise gut als Wiederholung **eines** Experiments ansehen können; verschiedene Spielergebnisse einer Fußballmannschaft jedoch nur näherungsweise. Trotzdem gewinnt man auch bei Betrachtung solcher nicht–identischer Wiederholung eine gewisse Information, die zur Grundlage von *Handeln unter Unsicherheit* gemacht werden kann; z.B. bei Fußballwetten.

Beobachtete relative Häufigkeiten von Ereignissen werden wir zur empirischen Grundlage der sog. *mathematischen Wahrscheinlichkeit* machen (Häufigkeitsinterpretation).

Ein große Schwierigkeit bei der Aufstellung eines stochastischen Modells (etwa im Unterschied zu einem geometrischen Modell) besteht darin, dass die zu modellierenden Sachverhalte nicht immer "mit bloßem Auge" erkennbar sind, sondern eine genaue (fachwissenschaftliche) Analyse (z.B. der Abhängigkeit verschiedenen Ereignisse voneinander) erfordert. So kann es für eine empirische Situation verschiedene Modelle geben, bei denen jeweils unterschiedliche Aspekte der realen Situation berücksichtigt sind. Größen wie die relative Häufigkeit werden in der Praxis selten bewusst (wie bei Gewinnchancen in einen Spiel) eingesetzt, obwohl ihre unbewusste Anwendung (z. B. beim Autofahren) eine große Rolle spielt. Für das richtige Verständnis der mathematischen Theorie ist daher die Unterscheidung zwischen Realität und Modell besonders wichtig.

Im folgenden werden wir als mathematisches Modell für die Ergebnisse eines Zufallsexperiments und seine relativen Häufigkeiten den sog. **Wahrscheinlichkeitsraum** entwickeln. Sind dabei überabzählbar viele Ergebnisse möglich (z.B. bei der Beobachtung zufallsbedingter Zeitdauern, modelliert als reelle Zahlen), so ist ein größerer mathematischer Aufwand erforderlich (z.B. Integrale statt Summen). Daher beschränken wir uns zunächst auf Modelle für Zufallsexperimente, die nur abzählbar oder sogar endlich viele Ausgänge haben können. Die darin abgeleiteten Begriffe und Aussagen werden wir dann mit meist nur geringen Modifikationen auf den überabzählbaren Fall übertragen.

1.2 Historische Entwicklung

In ihrer Entwicklungsgeschichte hat die Wahrscheinlichkeitstheorie immer weitere Anwendungsgebiete gefunden, deren spezifische Fragestellungen zu einem weiteren Ausbau dieser Theorie geführt haben. Auch durch die Informatik entstehen immer wieder neue Anwendungsgebiete für diese Theorie.

Zum Verständnis der Grundbegriffe der Wahrscheinlichkeitstheorie ist es nützlich, zunächst zum historisch ältesten Anwendungsbereich, der lange auch der einzige war, zurückzukehren, nämlich dem **Glücksspiel**.

Schon im 16. Jahrhundert begann man sich für die "Gesetzmäßigkeiten von zufälligen Ereignissen" zu interessieren, wie das "Liber de ludo alae" des italienischen Mathematikers **Geronimo Cardano** (1501–1576) zeigt, wo Probleme beim Werfen mit 3 Würfeln behandelt werden. Am Typ der Fragestellung erkennt man, dass die Wahrscheinlichkeitstheorie zu dieser Zeit noch im wesentlichen mit der **Kombinatorik** zusammenfällt. Zu erwähnen ist in diesem Zusammenhang noch **Jakob I. Bernoulli** (1654–1704), auf den das *Gesetz der großen Zahlen* zurückgeht (vergl. Kapitel 5). Grundlegend sind auch die Arbeiten des französischen Mathematikers **Abraham de Moivre** (1667–1754), der als erster einen Spezialfall des sog. *zentralen Grenzwertsatzes* (vergl. Kapitel 7) bewies und dessen Buch "the doctrine of chances" (1718) großen Einfluss auf die Entwicklung der Wahrscheinlichkeitstheorie hatte.

Mit den Fortschritten in der Grundlegung der Analysis sowie mit erweiterten Anwendungsmöglichkeiten, besonders in der Wirtschaft und der Physik, damals auch im juristischen Bereich, entwickelte sich die Wahrscheinlichkeitstheorie mehr und mehr zu einer ernst zunehmenden mathematischen Theorie. Besonders beeinflusst wurde diese Theorie im 18. und beginnenden 19. Jahrhundert von **Piere Simone Laplace** (1749–1829), der als erster eine analytische Darstellung der Wahrscheinlichkeitstheorie vorstellte, in der u.a. die folgende Definition des Begriffes **Wahrscheinlichkeit** benutzt wurde:

> Die gesuchte *Wahrscheinlichkeit eines Ereignisses* finden wir durch Zurückführen aller Ereignisse derselben Art auf eine gewisse Anzahl gleichmöglicher Fälle, d.h. solcher, über deren Existenz wir in gleicher Weise unschlüssig sind, und durch Bestimmung der dem Ereignis günstigen Fälle.

Im Vergleich zu unserem heutigen Verständnis von Wahrscheinlichkeit ist diese Definition sehr eingeschränkt und basiert noch stark auf der kombinatorischen Methode zur Bestimmung der Wahrscheinlichkeit für ein Ereignis A durch den Quotienten

$$\mathrm{P}\,(\mathrm{A}) \;=\; \frac{\text{Anzahl der für A günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}\;.$$

Von dieser Formel machen wir auch heute noch Gebrauch.

Laplace benutzte seine Ergebnisse für astronomische Studien. Sicher auch von der Astronomie angeleitet, gab der Mathematiker und Astronom **Carl Friedrich Gauß** (1777–1855) eine wahrscheinlichkeitstheoretische Begründung für die in der Fehlerrechnung viel benutzte *Methode der kleinsten Fehlerquadrate* an.

Starken Einfluss auf die Entwicklung der Wahrscheinlichkeitstheorie im 19. und 20. Jahrhundert hatte die berühmte russische Schule, mit der sich Namen wie **Pafnuti–Lwowitch Chebychev**

(1821–1894), **Andrej Andrejewitch Markov** (1856–1922) und besonders **Andrej Nikolajewitch Kolmogorov** (1903–1987) verbinden. Zu einer geschlossenen mathematischen Theorie (im Gegensatz zu einer Summe von Einzelergebnissen) hat sich die Wahrscheinlichkeitstheorie(WT) erst mit Kolmogorov entwickelt. Seitdem ist die WT unübersehbar gewachsen und in vielen Zweigen weiterentwickelt worden.

Anwendungen hat die WT zunächst beim Glücksspiel und wahrscheinlich schon sehr früh bei dem mit dem Glücksspiel verwandten **Versicherungswesen** (Berechnung der erforderlichen Prämien in Lebensversicherungen, Beantwortung von Fragen nach der Wahrscheinlichkeit für den Ruin einer Versicherungsgesellschaft) gefunden. Weitere Bereiche eröffneten sich durch die Bevölkerungsstatistik und wirtschaftliche Problemstellungen. Bei den wissenschaftlichen Anwendungen waren es vor allem die Astronomie und die Physik, die zur Weiterentwicklung dieser Theorie geführt haben. Heute gibt es kaum noch eine Wissenschaft, die sich nicht stochastischer Methoden bedient.

Trotz dieser Konsolidierung der Wahrscheinlichkeitstheorie in Theorie und Praxis ist die mehr philosophische Diskussion um eine Begründung der Wahrscheinlichkeitstheorie noch nicht abgebrochen: zum Überblick werden deshalb im folgenden Abschnitt verschiedene Begründungen des Begriffs **Wahrscheinlichkeit** diskutiert.

1.3 Der Begriff der Wahrscheinlichkeit

Im normalen wie im wissenschaftlichen Alltag wird häufig das Wort "wahrscheinlich" verwendet. Die Aufgabe der Wahrscheinlichkeitstheorie ist es, diesen Gebrauch so zu präzisieren, dass er für korrektes Argumentieren sowie zur Lösung von Problemen nützlich ist. "Präzisieren" heißt dabei, eine Theorie anzugeben, die mathematisch korrekt ist und gleichzeitig die Bedeutung des Wortes "wahrscheinlich" in solchen Sätzen möglichst genau trifft. Man hat also zu überlegen,was die Bedeutung in den jeweiligen Sätzen ist. Dabei stellt sich heraus, dass in "wahrscheinlich" eine ganze Reihe verschiedener Bedeutungen stecken, die mehr oder weniger gut mathematischer Behandlung zugänglich sind, die teilweise miteinander mathematisch übereinstimmen, wenn auch nicht inhaltlich, die sich aber teilweise völlig widersprechen.

Im folgenden wird eine Liste von Theorien der Wahrscheinlichkeit angegeben, aus der ersichtlich werden soll, dass die Debatte um die Bedeutung dieses Begriffs keineswegs abgeschlossen ist. Eine Beschäftigung mit solchen Auseinandersetzungen ist sicherlich nützlich. Die Relevanz solcher Grundsatzdebatten sollte jedoch nicht überschätzt werden.

<u>Verschiedene Begründungen der Wahrscheinlichkeitstheorie</u>

<u>Statistische Begründung:</u>
Macht man eine große Anzahl gleichartiger Experimente, so stellt sich oft heraus, dass die relative Häufigkeit des Eintretens eines bestimmten Ergebnisses mit wachsender Anzahl von Beobachtungen gegen einen bestimmten Wert zu konvergieren scheinen. Man könnte also diese relative

Häufigkeit als W. auffassen. Dann wäre aber die W. abhängig vom Ausgang einer Versuchsreihe, also die W. für "Kopf" beim Münzwurf einmal 42/100, ein anderes mal 53/100. Man sollte also besser den Grenzwert der Folge der relativen Häufigkeiten bei fortgesetztem Münzwurf als Definition nehmen. Neben der Frage nach der Existenz bleibt das Problem, wie man ihn feststellen will, da man stets nur endlich viele Versuche durchführen kann und jede endlich Zahlenfolge stets zu einer unendlichen Zahlenfolge mit einem beliebig vorgebbaren Grenzwert fortgesetzt werden kann. **Mises** hat versucht, dieses Problem zu lösen, indem er forderte, dass die endliche Folge "zufällig" sein müsse. Das Problem, was nun "zufällig" heißt, konnte auch er nicht in den Griff bekommen. Der erste brauchbare (?) Ansatz dazu stammt von **Schnorr** durch den Begriff der *Berechenbarkeit*. Danach ist *zufällig*, was *nicht berechenbar* ist.

Festzuhalten ist, dass die statistische Begründung auf beliebig oft wiederholbare Ereignisse zugeschnitten ist.

Kombinatorische Wahrscheinlichkeit:
Statistische Ergebnisse und kombinatorische überlegungen stützen sich oft gegenseitig. Man kann Ereignisse in vielen Fällen elementar zerlegen und durch Auszählen aller möglicher elementarer Ereignisse die Wahrscheinlichkeit als Quotient der "günstigen" durch die "möglichen" Fälle definieren. Dabei bleibt offen, was i.a. ein "mögliches" Ereignis ist. Warum verwendet man beim Münzwurf ("Kopf" oder "Zahl") nicht die 4 Ereignisse "Kopf", "Zahl", "Münze bleibt auf dem Rand stehen" und "Münze bleibt in der Luft"? Es ist offensichtlich notwendig, Relevanzkriterien aufgrund naturwissenschaftlicher überlegungen oder auch subjektiver Überzeugungen zu finden.

Logische Wahrscheinlichkeit (nach **Carnap**):
Die logische Wahrscheinlichkeit nimmt das Problem ernst, dass die Auswahl der zu beachtenden Ereignisse sowie die Zuordnung von W. zu Ereignissen oft aufgrund apriorischer überlegungen durchgeführt werden müssen. Diese überlegungen beruhen oft auf Informationen über frühere Ereignisse oder Naturgesetze o.ä.. Man kann nun diese Informationen I als Bedingungen für das Eintreten eines Ereignisses E auffassen und die W. von E als Maß dafür, in welchem Grad E von diesen Bedingungen logisch impliziert wird, d.h. $\mathrm{W(E|I)} = 1$, falls E logisch aus I folgt, bzw. $= 0$, falls die Negation von E logisch aus I folgt. Für die Definition der übrigen Werte sind dann geeignete(?) Festlegungen zu treffen.

Eine wesentliche Beschränkung der logischen W. ist, dass nur Ereignisse behandelt werden können, die sich in einer sehr einfachen formalen Sprache ausdrücken lassen, und dass sie die ebenfalls nicht unproblematische Begründung der Logik voraussetzt.

Die obigen Theorien machen den Begriff der W. an objektiven Naturvorgängen fest. Demgegenüber folgen nun einige "subjektive" Theorien.

Subjektive Theorie nach **de Finetti**:
Der Grundgedanke ist, dass wir nie wissen, ob eine Information wahr ist, sondern höchstens von ihrer Wahrheit überzeugt sein können. Radikaler formuliert: Wir können Wahrscheinlichkeiten nicht behaupten, sondern nur ausdrücken, in welchem Grad wir vom Eintreffen eines Ereignisses überzeugt sind. Falls wir bereit sind, unsere Überzeugung aufgrund von Erfahrung jeweils in "rationaler" Weise zu ändern, kann bei bestimmten Aussagen dieser Grad gegen einen bestimmten

Wert konvergieren, der dann die W. des Ereignisses ist, das durch die Aussage beschrieben wird.

Problematisch ist hier vor allem, wie hier das "rationale Verhalten" festgelegt werden soll. Demgegenüber steht der Vorzug, dass die Theorie selbst erklärt, wie man zu Wahrscheinlichkeitsaussagen kommt, anstatt das irgendwelchen Vorüberlegungen zu überlassen. Das erfordert einen grundsätzlich anderen mathematischen Apparat, man benutzt u.a. Methoden der *Spieltheorie*. Die Ergebnisse stimmen aber in vielen Punkten mit den vorher beschriebenen Theorien überein.

Entscheidungstheoretische Wahrscheinlichkeit:
Die WT wird hier als Theorie mit der Hauptaufgabe, Entscheidungen zu fällen, aufgefasst: Fällt beim Würfeln hintereinander 100-mal die 6, ist dann ein unwahrscheinliches Ereignis eingetreten, oder ist der Würfel falsch? Der gemeinsame Ansatz verschiedener Varianten dieser Methode ist etwa folgender:
Eine Person möge eine Funktion angeben, die den Schaden oder Nutzen angibt, den das Eintreten eines bestimmten Ereignisses für sie bedeutet. Sie kann nun eine Entscheidungsstrategie suchen, die in einem bestimmten Sinn optimal ist, nämlich den möglichen Verlust minimiert. Existiert eine solche Strategie, so ist dadurch eine subjektive Wahrscheinlichkeit dieses Ereignisses bestimmt.

Fuzzy – Wahrscheinlichkeitstheorie:
Sie benutzt als Grundlage die sog. Fuzzylogik, die keine absoluten Mengenzugehörigkeiten kennt (Grundlage der Cantorschen Mengendefinition), sondern auch zulässt, dass Elemente mit bestimmten "Zugehörigkeitsgraden" zu mehreren Mengen gehören. Sie soll hier nicht behandelt werden.

Dies war nur eine Auswahl der wichtigsten Theorien. Ihre Vielzahl und die Energie mit der sich teilweise die Anhänger der Theorien untereinander bekämpfen, könnte den Versuch nahe legen, auf die WT vollständig zu verzichten. Gewisse philosophische und weltanschauliche Positionen (radikaler Determinismus z. B.) legen das auch nahe.

Betrachtet man jedoch die Ergebnisse neuerer physikalischer Grundlagenforschung (Quantenmechanik, chaotische dynamische Systeme usw.), so erscheint diese Position reichlich obsolet. Vereinfachend und überspitzt formuliert legt die moderne Physik nahe, dass bei allen Vorgängen in der Natur der Zufall allgegenwärtig ist. Dass gewisse Vorgänge deterministisch abzulaufen scheinen, liegt an einer makrokosmischen Betrachtungsweise und ist im wesentlichen eine Folge des *Gesetzes der großen Zahlen* (vergl. Kapitel 5).

1.4 Literatur

Die nachfolgend aufgeführten Bücher stellen einen kleinen Ausschnitt aus einem überreichlichen Vorrat an guten Büchern zu diesem Themenkreis dar. Dass ein Buch hier nicht aufgeführt ist heißt nicht, dass es sich nicht lohnt es zu benutzen. Die Auswahl ist subjektiv auf die Bedürfnisse der Vorlesung hin zugeschnitten.

<u>Literatur zur Wahrscheinlichkeitstheorie für Studierende der Informatik</u>

Bauer, H.: Wahrscheinlichkeitstheorie, W. de Gruyter, Berlin, 5. Auflage , 2002.

Behnen, K.: Grundkurs Stochastik, Teubner, Stuttgart, 1995.

Bosch, K.: Elementare Einführung in die Wahrscheinlichkeitsrechnung, Vieweg, studium Basiswissen, 8. Aufl., 2003.

Bosch, K.: Lotto und andere Zufälle, Vieweg-Verlag, 1999.

Greiner, M. & Tinhofer, G.: Stochastik für Studienanfänger der Informatik, Hanser-Verlag, 1996.

Greiner, M.: Stochastik für Studierende der Informatik - Ausgewählte Aufgaben zur Vertiefung und Prüfungsvorbereitung , CS-Press Dr. Christian Sutter, 1997.

Henze, N.: Stochastik für Einsteiger, Vieweg, 5. Aufl., 2004.

Hübner, G.: Stochastik. Eine anwendungsorientierte Einführung für Informatiker, Ingenieure und Mathematiker. Mathematische Grundlagen der Informatik, Vieweg-Verlag, 4. Aufl., 2003.

Jondral, F. und Wiesler, A.: Wahrscheinlichkeitsrechnung und stochastischer Prozesse, Grundlagen für Ingenieure und Naturwissenschaftler, Teubnerverlag, 2. Auflage 2002.

Krengel, U.: Einführung in die Wahrscheinlichkeitstheorie und Statistik, Vieweg Studium, Aufbaukurs Mathematik, 7. Aufl., 2003.

Kapitel 2

Kombinatorische Grundaufgaben

Einen breiten Raum nehmen in der Informatik Wahrscheinlichkeiten ein, die mit kombinatorischen Überlegungen berechnet werden können. Dazu müssen immer wieder Elementeanzahlen von geeignet gewählten **Mengen** und **Listen** ermittelt werden. Daher sind nachfolgend die wichtigsten **kombinatorischen Grundaufgaben** und deren Lösungen zusammengestellt.

Bei **Mengen** spielt die Reihenfolge der Elemente keine Rolle, jedes Element kommt genau einmal vor; so ist $\{1,4\} = \{4,1\} = \{1,4,1\}$. Bei **Listen** ist die Reihenfolge der Elemente signifikant. In **mehrfachen Listen** können im Gegensatz zu **einfachen** Listen Elemente auch mehrfach erscheinen, also $(1,4) \neq (4,1) \neq (1,4,1)$.

Definition 2.1 *Eine* **Grundmenge** Ω *von der Mächtigkeit n (Anzahl der Elemente von* Ω *ist* $|\Omega| = n$*) sei vorgegeben. Eine Auswahl von r Elementen aus* Ω *heißt* **Kombination** *oder* **Probe aus** Ω **vom Umfang** r. *Kann jedes Element von* Ω *maximal einmal in die Probe gewählt werden, spricht man von einer Probe* **ohne Wiederholung**, *sonst von einer Probe mit Wiederholung. Ist die Reihenfolge der Elemente von Bedeutung, spricht man von einer* **geordneten**, *sonst von einer* **ungeordneten** *Probe. In einer ungeordneten Stichprobe können ohne Informationsverlust die Elemente umgeordnet also sortiert werden.*

In Kurzform sagt man, eine Probe heißt

Bezeichnung	*falls sie ... ist*
Menge	*ungeordnet ohne Wiederholungen*
mehrfache Menge	*ungeordnet mit Wiederholungen*
(mehrfache) Liste	*geordnet mit Wiederholungen*
einfache Liste	*geordnet ohne Wiederholungen*

Eine Liste der Länge 2 , 3 , k heißt auch **Paar** , **Tripel** , k**-Tupel**.

Für die Ermittlung von Wahrscheinlichkeiten ist es erforderlich festzustellen, wie viele verschiedene Proben der obigen Art aus einer Grundmenge gezogen werden können.

Satz 2.1 *Aus einer Grundmenge Ω vom Umfang n können die folgenden Anzahlen unterschiedlicher Proben des Umfangs r gezogen werden:*

Typ	*Anzahl*	
mehrfache Liste	$A(n,r) = n^r$	$(r \in I\!N)$
einfache Liste	$B(n,r) = \frac{n!}{(n-r)!}$	$(r \leq n)$
Menge	$C(n,r) = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$	$(r \leq n)$
mehrfache Menge	$D(n,r) = \binom{n+r-1}{r}$	$(r \in I\!N)$

Satz 2.2 *Seien* $A_1, \ldots, A_k$ *beliebige Mengen. Dann gilt*

a) $|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \cdot |A_2| \cdots |A_k|$ *und*

b) $\left|\bigcup\limits_{i=1}^{k} A_i\right| = \sum\limits_{i=1}^{k} |A_i|$, *falls die A_i disjunkt sind.*

Allgemein gilt: $\left|\bigcup\limits_{i=1}^{k} A_i\right| \leq \sum\limits_{i=1}^{k} |A_i|$

Bemerkung: Der Beweis kann durch Induktion über k geführt werden.

Beispiel 2.1 *Wie viele Möglichkeiten gibt es, einen Lottoschein korrekt auszufüllen?*

Auf einem korrekt ausgefüllten Lottoschein sind 6 verschiedene Zahlen angekreuzt. für das erste Kreuz hat man 49, dass das zweite 48, ... und dass das sechste Kreuz $49-5$ Möglichkeiten. Es gibt also $49 \cdot 48 \cdots 44 = \dfrac{49!}{(49-6)!} = B(49,6) = 10.068.347.520$ *verschiedene Arten, einen Lottoschein auszufüllen.*

Vorsicht: Die Anzahl der unterschiedlichen korrekt ausgefüllten Lottoscheine ist kleiner, da oben das Ankreuzen der gleichen 6 Zahlen in unterschiedlicher Reihenfolge jeweils mehrfach gezählt wird. für die Anzahl unterschiedlicher (korrekter) Lottoscheine gilt also

$$\frac{10.068.347.520}{6!} = \binom{49}{6} = 13.983.816 \ .$$

Beispiel 2.2 *Wie viele n–äre Zahlen der Länge r gibt es? Eine n–äre Zahl ist eine Zahl zur Basis n.*

n	*Bezeichnung*
2	*Binärzahlen, Dualzahlen*
8	*Oktalzahlen*
10	*Dezimalzahlen*
16	*Hexadezimalzahlen*

Allgemein sind dies n^r, es gibt also (für $r = 3$) 8 verschiedene Binär–, 512 Oktal–, 1000 Dezimal– und 4096 Hexadezimalzahlen mit drei Ziffern.

Beispiel 2.3 *Wie viele Elemente besitzt die Potenzmenge $\mathcal{P}(A_n)$ einer Menge A_n der Mächtigkeit n?*

Die Anzahl der r-elementigen Teilmengen beträgt offenbar

$$C(n,r) \;=\; \binom{n}{r} \qquad \Longrightarrow \qquad |\mathcal{P}(A_n)| \;=\; \sum_{r=0}^{n} \binom{n}{r}$$

Unter Benutzung des **binomischen Lehrsatzes**

$$(a+b)^n \;=\; \sum_{i=0}^{n} \binom{n}{i}\, a^i\, b^{n-i} \tag{2.1}$$

erhalten wir für die obige Anzahl

$$|\mathcal{P}(A_n)| \;=\; \sum_{i=0}^{n} \binom{n}{i} 1^i 1^{n-i} \;=\; (1+1)^n \;=\; 2^n \quad .$$

Für das Rechnen mit **Binomialkoeffizienten** $\binom{n}{r}$ gelten folgende leicht zu beweisende Rechenregeln:

$$\binom{n}{r} \;=\; \binom{n}{n-r} = \frac{n(n-1)\cdots(n-r+1)}{r!} \tag{2.2}$$

$$\binom{n}{r} \;=\; \binom{n-1}{r} \;+\; \binom{n-1}{r-1} \tag{2.3}$$

$$\binom{n}{r} \;=\; \frac{n}{r}\binom{n-1}{r-1} \tag{2.4}$$

$$\binom{n}{r} \;=\; 0 \quad \text{für } r > n \tag{2.5}$$

Beispiel 2.4 *In einer "kleinen Übungsgruppe" mit 15 Teilnehmern sollen Untergruppen mit zwei oder drei Teilnehmern gebildet werden. Wie viele verschiedene Aufteilungen sind möglich? Dabei soll auch unterschieden werden, in welche Übungsgruppe gleicher Größe ein Student kommt (unterschiedliche Übungsleiter).*

Die Teilnehmer bilden entweder 5 *Untergruppen mit je* 3 *Teilnehmern oder* 6 *Untergruppen mit* 3×2 *und* 3×3 *oder* 7 *Untergruppen mit* 6×2 *und* 1×3 *Teilnehmern. Damit erhalten wir also dass die Gesamtanzahl*

$$\begin{aligned} & E(15;3,3,3,3,3) \;+\; E(15;2,2,2,3,3,3) \;+\; E(15;2,2,2,2,2,2,3) \\ &= \frac{15!}{3!^5} + \frac{15!}{2!^3 \cdot 3!^3} + \frac{15!}{2!^6 \cdot 3!} \;=\; 4.330.326.000 \quad . \end{aligned}$$

Dabei nennt man

$$E\{n; m_1, \ldots, m_k\} \;=\; \frac{n!}{m_1! \cdots m_k!} \;=\; \binom{n}{m_1 \; \cdots \; m_r}$$

$$\text{mit} \qquad n, m_1, \ldots, m_k \in I\!N \quad \text{und} \quad m_1 + \cdots + m_k = n$$

den **Multinomialkoeffizienten** *n über m_1 bis m_k.*

Kapitel 3

Diskrete Wahrscheinlichkeitsräume

In der Wahrscheinlichkeitstheorie sollen Modelle für Erscheinungen der Wirklichkeit entwickelt, beschrieben und ausgewertet werden, in denen der **Zufall** eine Rolle spielt. Dabei wollen wir uns nicht auf die Diskussion über Existenz eines "mystischen" Zufalls einlassen. Immer dann, wenn das Ergebnis eines Experiments (**Versuchsausgang**) nicht eindeutig (deterministisch) aus seinen reproduzierbaren Grundvoraussetzungen feststeht, wollen wir von einem **Zufallsexperiment** sprechen. Ein Zufallsexperiment muss wie jedes andere physikalische Experiment unter "gleichen Bedingungen" wiederholbar sein (und trotzdem unterschiedliche Ausgänge haben können).

Es soll also "zufällig" sein, welcher der möglichen Ausgänge des Experimentes eintritt.

Dazu verwenden wir die folgenden Bezeichnungen:

Ω	**Menge der Versuchsausgänge**, Stichprobenraum, Raum der Elementarereignisse, Merkmalraum, (Ereignisraum)
$\omega \in \Omega$	(möglicher) **Versuchsausgang,** (Elementarereignis?), Realisierung, (Stichprobe)

Wir sprechen von einem **diskreten Zufallsexperiment**, wenn die Anzahl der Versuchsausgänge **endlich** oder **höchstens abzählbar unendlich** ist.

In diesem Abschnitt setzen wir also

$$\begin{aligned} \Omega &= \{\omega_1, \omega_2, \ldots, \omega_n\} \qquad \text{oder} \\ \Omega &= \{\omega_1, \omega_2, \omega_3, \ldots\} \quad \text{mit } |\Omega| = |\mathbb{N}| \end{aligned}$$

voraus.

Beispiel 3.1 (Münzwurf) $\Omega = \{K, Z\} \quad (n = 2)$
K = *"Kopf liegt oben"* , **Z** = *"Zahl liegt oben"*

Beispiel 3.2 (Würfeln mit einem Würfel) $\Omega = \{1, 2, \ldots, 6\}$ $(n = 6)$

$$\omega_i = i = \text{"}\textit{Augenzahl i liegt oben}\text{"} \quad (i = 1, \ldots, 6)$$

Führt man nun diese Zufallsexperimente mehrfach durch, so wird man bei korrekter Durchführung (bzw. "korrektem" Würfel) erwarten, dass alle Versuchsausgänge bei "hinreichend großer" Anzahl von Versuchen "etwa" gleich oft vorkommen. Sei also

$$r_N(\omega_i) := \frac{\text{Anzahl der Ausgänge } \omega_i \text{ unter } N \text{ Versuchen}}{N}$$

die **relative Häufigkeit** des Versuchsausgangs ω_i, dann erwartet man hier für große N

$$r_N(\omega_i) \approx \frac{1}{n} \quad .$$

Allgemein wird man nicht bei jedem Zufallsexperiment erwarten, dass alle Versuchsausgänge in etwa gleich oft realisiert werden.

Beispiel 3.3 (Würfeln mit zwei Würfeln) *Es werden jeweils die Augensummen als Versuchsausgang notiert, also*

$$\Omega = \{2, 3, \ldots, 12\} \quad .$$

Führt man dieses Experiment hinreichend oft durch, pendeln sich die relativen Häufigkeiten zwar auch in der Nähe fester Wert ein. Jedoch sind diese teilweise deutlich unterschiedlich.

Die Folge dieser relativen Häufigkeiten scheinen mit wachsendem N eine immer bessere Näherung für das zu werden, was wir umgangssprachlich die **Wahrscheinlichkeit** des entsprechenden Versuchsausgangs nennen. Leider ist es nicht möglich, hier einen Grenzübergang wie in der Analysis durchzuführen und den daraus resultierenden Grenzwert als mathematische Definition zu verwenden. Denn

- es ist nicht möglich den Grenzübergang experimentell durchzuführen,
- auch ein Gedankenexperiment könnte zu beliebigen Grenzwerten führen, da zum Beispiel nicht auszuschließen ist, wenn auch sehr unwahrscheinlich, dass bei beliebig vielen Experimenten auch mit korrekten Würfeln nie die Sechs gewürfelt wird, ...

So geht man für die mathematische Definition der Wahrscheinlichkeit so vor, dass man aus "logischen" Überlegungen heraus bei einfachen Grundexperimenten den Versuchsausgängen Wahrscheinlichkeiten zuordnet und sich dabei stark an dem anschaulichen "Vorbild" der relativen Häufigkeit orientiert.

Bevor wir zu einer Definition kommen, müssen wir bedenken, dass wir solche Wahrscheinlichkeiten nicht nur für einzelne Versuchsausgänge, sondern auch für weiter gefasste **Ereignisse**, zum Beispiel für Fragestellungen definieren müssen wie

- Mit welcher Wahrscheinlichkeit ist die Augensumme zweier Würfel gerade?
- Mit welcher Wahrscheinlichkeit ist die Augensumme kleiner als 5?
- Mit welcher Wahrscheinlichkeit liegen zwei "Buben" im "Skat"?

Alle diese Fragen suchen nach Wahrscheinlichkeiten für das Auftreten von Versuchsausgängen, die zu einer Teilmenge der Menge aller Versuchsausgänge gehören.

Wir definieren deshalb

Definition 3.1 *Sei Ω die Menge aller Versuchsausgänge eines diskreten Zufallsexperiments. Dann heißt jede Teilmenge A von Ω ein* **Ereignis**

$$A \quad \textit{heißt} \quad \textbf{Ereignis} \qquad \Longleftrightarrow \qquad A \subseteq \Omega$$

Speziell heißen
$\emptyset$ *unmögliches Ereignis*
Ω *sicheres Ereignis*

Die Menge aller Ereignisse eines diskreten Zufallsexperiments ist die Potenzmenge

$$\mathcal{P}(\Omega) = \{A \mid A \subseteq \Omega\} \quad .$$

Beispiel 3.4 (Würfeln mit einem Würfel) *Die Versuchsausgänge ω_i sind die jeweils gewürfelten Augenzahlen $(i = 1, \ldots, 6)$. Dann kann man u.a. die folgenden Ereignisse formulieren:*

$$\begin{array}{rcl}
A_i = \{i\} & : & \textit{"Augenzahl i liegt oben"} \ (i = 1, \ldots, 6) \quad \textit{(Elementarereignis)} \\
A = A_1 \cup A_2 = \{1, 2\} & : & \textit{"Augenzahl} \leq 2\textit{"} \\
B = A_1 \cup A_3 \cup A_5 & : & \textit{"Augenzahl ungerade"} \\
C = \bigcup_{i=1}^{6} A_i = \Omega & : & \textit{"Wurf mit beliebiger Augenzahl" (sicheres Ereignis)} \\
D = B \cap A = A_1 & : & \textit{"Augenzahl ungerade und} \leq 2\textit{"}
\end{array}$$

Für das Operieren mit Ereignissen steht nun der gesamte Mengenkalkül zur Verfügung, wobei wir die folgenden Verabredungen treffen:

... tritt ein	wenn ... eintritt
A	ein Versuchsausgang $\omega_i \in A$
$A \cup B$	A oder B
$A \cap B$	sowohl A als auch B
$\overline{A} := A^c$	A nicht
$A \backslash B$	ein Versuchsausgang aus A, der nicht zu B gehört

Auch den übrigen Mengenoperationen kann der folgende Sinn unterlegt werden:

$A \subset B$	:	Wenn A eintritt, ist B automatisch eingetreten (A impliziert B)
$A \cap B = \emptyset$	:	A und B können nicht gleichzeitig eintreten (disjunkt)
$A \backslash B$	:	$= \{\omega : \omega \in A \ \wedge \ \omega \notin B\} = A \cap \overline{B}$

Zur Vereinfachung führen wir noch folgende Schreibweisen ein:

$$\begin{aligned} A \cap B = \emptyset \quad &\Longrightarrow \quad A \cup B \;:=\; A + B \qquad \text{bzw.} \\ A_1, A_2, A_3, \ldots \text{ paarweise disjunkt} \quad &\Longrightarrow \quad \bigcup_i A_i \;:=\; \sum_i A_i \\ A \cap B \quad &:= \quad A \cdot B \;:=\; AB \end{aligned}$$

Die einelementigen Teilmengen von Ω bezeichnet man als **Elementarereignisse**, so dass wir jedes Ereignis auch als Summe von Elementarereignissen schreiben können.

$$A \;=\; \{\omega_1, \ldots, \omega_r\} \;=\; \sum_{i=1}^{r} \{\omega_i\} \qquad \forall A \subseteq \Omega$$

Satz 3.1 *Es sei* $2 \leq n \leq \infty$ *und* $A_i \subset \Omega$ $(1 \leq i \leq n)$ *seien Ereignisse, dann gilt*

(a) $\displaystyle\bigcup_{i=1}^{n} A_i \;=\; A_1 \;+\; \sum_{i=2}^{n} \left(A_i \setminus \bigcup_{j=1}^{i-1} A_j \right)$ *(Zerlegung in disjunkte Ereignisse)*

(b) Sind B_i $(1 \leq i \leq n)$ *paarweise disjunkt und* $A \subset \sum_{i=1}^{n} B_i$, *so gilt*

$$A \;=\; \sum_{i=1}^{n} (A \cap B_i)$$

(Zerlegung des Ereignisses A *mit Hilfe der "Fallunterscheidung"* B_i, *insbesondere für den Fall* $\Omega = \sum_{i=1}^{n} B_i$.*)*

Wenn wir nun für beliebige Ereignisse $A \in \mathcal{P}(\Omega)$ eine Wahrscheinlichkeit einführen, soll dieser Wahrscheinlichkeitsbegriff möglichst stark am Vorbild der relativen Häufigkeit orientiert sein. Dazu stellen wir einige **Eigenschaften der relativen Häufigkeiten** r_N zusammen:

(3.1) $$r_N(A) \;:=\; \frac{\text{Anzahl der Ausgänge } \omega_i \in A \text{ unter } N \text{ Versuchen}}{N}$$

(3.2) $$r_N(A) \;\geq\; 0 \qquad \forall A \subseteq \Omega \qquad \text{(Positivität)}$$

(3.3) $$r_N(\Omega) \;=\; 1 \qquad \text{(Beschränktheit, Normierung)}$$

(3.4) $$A_i \text{ paarw. disjunkt} \;\Longrightarrow\; r_N\Big(\sum_i A_i\Big) \;=\; \sum_i r_N(A_i) \qquad (\sigma-\text{Additivität})$$

Unter einer "Wahrscheinlichkeit" (besser: Wahrscheinlichkeitsfunktion) wollen wir nun eine Funktion verstehen, die jedem Ereignis A eine Zahl zuordnet, die Wahrscheinlichkeit des Ereignisses. Dabei wollen wir nur solche Funktionen zulassen, die die obigen Eigenschaften der relativen Häufigkeiten besitzen, also

Definition 3.2 (Kolmogorov – Axiome)
Die Funktion $P : \mathcal{P}(\Omega) \longrightarrow I\!R^1$ *heißt* **Wahrscheinlichkeit** *oder* **Wahrscheinlichkeitsmaß** *auf* Ω *(genauer auf* $\mathcal{P}(\Omega)$*) genau dann, wenn*

(3.5) $\quad P(A) \geq 0 \qquad \forall\, A \subseteq \Omega$

(3.6) $\quad P(\Omega) = 1$

(3.7) $\quad A_i \subseteq \Omega$ *paarw. disjunkt* $\Longrightarrow P\Big(\sum\limits_i A_i\Big) = \sum\limits_i P(A_i)$.

Das Tripel $(\Omega, \mathcal{P}(\Omega), P)$ *heißt dann* **diskreter Wahrscheinlichkeitsraum (WR)**.

Bemerkung:

Bei diskreten Wahrscheinlichkeitsräumen ist die Angabe der Menge aller Ereignisse $\mathcal{P}(\Omega)$ redundant. Häufig wird deshalb auch das Paar (Ω, P) als **diskreter Wahrscheinlichkeitsraum** bezeichnet.

Sind nun die Wahrscheinlichkeiten $p_i := P(\{\omega_i\})$ für alle Elementarereignisse $(i = 1, \ldots, n)$ bekannt, so kann wegen der obigen Additivität die Wahrscheinlichkeit für jedes
$A = \{\omega_1, \ldots, \omega_r\} \in \mathcal{P}(\Omega)$ ermittelt werden.

$$P(A) = P(\{\omega_1, \ldots, \omega_r\}) = P\left(\sum_{i=1}^{r} \{\omega_i\}\right) = \sum_{i=1}^{r} P(\{w_i\}) = \sum_{i=1}^{r} p_i \qquad \forall A \subseteq \Omega$$

Damit gilt wegen $P(\Omega) = 1$ auch die sinnvolle Normierung: $\quad \sum\limits_i p_i = 1$.

Aus der obigen Definition lassen sich nun mehr oder weniger leicht die folgenden Eigenschaften der Wahrscheinlichkeiten nach Kolmogorov ableiten.

Satz 3.2 (Rechenregeln für Wahrscheinlichkeiten)
Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum mit $A; B; \ldots \subseteq \Omega$, dann gilt

(3.8) $$P(\emptyset) = 0$$

(3.9) $$P(\overline{A}) = 1 - P(A)$$

(3.10) $$A \subseteq B \implies P(A) \leq P(B)$$

(3.11) $$P(A \backslash B) = P(A) - P(A \cap B)$$

(3.12) $$B \subseteq A \implies P(A \backslash B) := P(A - B) = P(A) - P(B) \quad \textit{(Subtraktivität)}$$

(3.13) $$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \textit{bzw. allgemeiner}$$

(3.14) $$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i) - \sum_{i_2} \sum_{i_1 < i_2} P(A_{i_1} A_{i_2}) + \sum_{i_3} \sum_{i_2 < i_3} \sum_{i_1 < i_2} P(A_{i_1} A_{i_2} A_{i_3}) \mp \cdots + (-1)^{n+1} P\Big(\bigcap_{i=1}^{n} A_i\Big)$$

(3.15) $$\sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i A_j) \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Ungleichung von **Bonferroni**

<u>Beweise</u>:

(3.8) $P(A) = P(A + \emptyset) = P(A) + P(\emptyset) \implies$ Beh.

(3.9) $1 = P(\Omega) = P(A + \overline{A}) = P(A) + P(\overline{A}) \implies$ Beh.

(3.10) $B = A + B\overline{A} \implies P(B) = P(A) + P(\overline{A}B) \geq P(A)$

(3.11) $A = A\Omega = A(B + \overline{B}) = AB + A\overline{B} = AB + A\backslash B$
$\implies P(A) = P(AB) + P(A\backslash B) \implies P(A\backslash B) = P(A) - P(AB)$

(3.12) $B \subseteq A \implies AB = B \implies P(A\backslash B) = P(A) - P(B)$

(3.13) $A \cup B = (A\backslash B) + B \implies P(A \cup B) = P(A\backslash B) + P(B) = P(A) - P(AB) + P(B)$

Die übrigen Formeln werden durch vollständige Induktion über n bewiesen.

Im Nachfolgenden werden wir für verschiedene wichtige Zufallsexperimente die Wahrscheinlichkeitsräume angeben.

3.1 Der Laplace – Raum

Beschreibung des Experiments:

Ein Zufallsexperiment habe N $(N < \infty)$ verschiedene Ausgänge, von denen keiner bevorzugt werde ("Alle Versuchsausgänge sind gleichwahrscheinlich").

Wahrscheinlichkeiten der Elementarereignisse:

Sei p_i die Wahrscheinlichkeit für das Elementarereignis

$$\text{"Der } i\text{–te Versuchsausgang ist eingetreten"} := \{\omega_i\} ,$$

dann gilt wegen $\sum\limits_{i=1}^{N} p_i = 1$ natürlich

$$P\big(\{\omega_i\}\big) := p_i = \frac{1}{N} \qquad \forall\, i = 1, \ldots, N$$

Definition 3.3 *Der Wahrscheinlichkeitsraum* $(\Omega, \mathcal{P}(\Omega), P)$ *heißt* **Laplace – Raum**

$$\begin{aligned} \Longleftrightarrow \qquad & (L_1) \quad |\Omega| = N < \infty \qquad \textit{und} \\ & (L_2) \quad P\big(\{\omega\}\big) = \frac{1}{N} \quad \forall\, \omega \in \Omega \qquad \textit{gilt.} \end{aligned}$$

Wahrscheinlichkeiten für beliebige Ereignisse

Sei $A := \{\omega_{i_1}, \ldots, \omega_{i_k}\}$, dann ergibt sich

$$P(A) = \sum_{j=1}^{k} p_{i_j} = \frac{k}{N} = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der günstigen Ausgänge}}{\text{Anzahl der möglichen Ausgänge}}$$

Die Wahrscheinlichkeit eines beliebigen Ereignisses A in einem **Laplace–Raum** ergibt sich als Quotient der Anzahl der "für A günstigen Ausgänge" durch die Anzahl N der möglichen Ausgänge.

Beispiel 3.5 (Würfeln mit einem Würfel)

$$\omega_i = \textit{"Augenzahl i wird gewürfelt"} \qquad (i = 1, \ldots, 6)$$

$$p_1 = p_2 = \cdots = p_6 = \frac{1}{6}$$

$$P(\textit{"gerade Augenzahl"}) = P(\{2,4,6\}) = \frac{3}{6} = \frac{1}{2}$$

Beispiel 3.6 (Würfeln mit 2 Würfeln)
Zur Beschreibung des Modells als Laplace–Raum muss man als Elementarereignisse solche Versuchsausgänge betrachten, die untereinander sicher "gleichwahrscheinlich" sind.

1. Versuch: $\omega_i \;=\;$ *"Augensumme = i"* , $i = 2, \ldots, 12$
Es ist sehr zweifelhaft, oft die Augensumme 2 genauso oft wie die Augensumme 7 vorkommt. Zur Überprüfung kann man eine größere Anzahl von Würfelversuchen durchführen.

2. Versuch:

$$\Omega \;=\; \big\{(i,j) : i,j \in \{1,2,3,4,5,6\}\big\}$$

also: i = *Augenzahl des 1. Würfels und* j = *Augenzahl des 2. Würfels.*

Jetzt gibt es keinen Grund daran zu zweifeln, dass alle Elementarereignisse $\omega = (i,j)$ *die gleiche Wahrscheinlichkeit besitzen.*

$$|\Omega| \;=\; (\textit{Anzahl der Augenzahlen des 1. W.}) \times (\textit{Anzahl der Augenzahlen des 2. W.}) \;=\; 36$$

Für alle $\omega \in \Omega$ *gilt also* $P(\{\omega\}) = 1/36$.

Ebenso gilt

$$P(\textit{"Augensumme} = m\textit{"}) \;=\; \frac{\textit{Anzahl der Paare aus } \Omega \textit{ mit Augensumme } m}{36}$$

Damit ergibt sich die Tabelle 3.1 für die Wahrscheinlichkeiten für das Auftreten der Augensumme m beim Würfeln mit 2 Würfeln:

Tabelle 3.1: Augensummen beim Würfeln mit zwei Würfeln

m	A_m	$p_m \;=\; P(A_m)$
2	(1,1)	$1/36 \;=\; 0.028 \;=\; 2,8\%$
3	(1,2) , (2,1)	$2/36 \;=\; 0.056 \;=\; 5,6\%$
4	(1,3) , (2,2) , (3,1)	$3/36 \;=\; 0.083 \;=\; 8,3\%$
5	(1,4) , (2,3) , (3,2) , (4,1)	$4/36 \;=\; 0.111 \;=\; 11,1\%$
6	(1,5) , (2,4) , (3,3) , (4,2) , (5,1)	$5/36 \;=\; 0.139 \;=\; 13,9\%$
7	(1,6) , (2,5) , (3,4) , (4,3) , (5,2) , (6,1)	$6/36 \;=\; 0.167 \;=\; 16,7\%$
8	(2,6) , (3,5) , (4,4) , (5,3) , (6,2)	$5/36 \;=\; 0.139 \;=\; 13,9\%$
9	(3,6) , (4,5) , (5,4) , (6,3)	$4/36 \;=\; 0.111 \;=\; 11,1\%$
10	(4,6) , (5,5) , (6,4)	$3/36 \;=\; 0.083 \;=\; 8,3\%$
11	(5,6) , (6,5)	$2/36 \;=\; 0.056 \;=\; 5,6\%$
12	(6,6)	$1/36 \;=\; 0.028 \;=\; 2,8\%$

3.2 Urnenmodelle

Beispiel 3.7 *In einer Urne befinden sich N Kugeln, von denen R rot und S schwarz ($N = R+S$) sind. Nun wird daraus eine Kugel "zufällig" gezogen. Wie groß ist die Wahrscheinlichkeit, dass die Kugel rot ist?*

"Zufällig" soll hier heißen, dass unabhängig von ihrer Farbe jede Kugel die gleiche Wahrscheinlichkeit besitzt, gezogen zu werden. Man sieht sofort, dass es keinen Sinn hat, die Ereignisse "rote/schwarze Kugel wird gezogen" zu Elementarereignissen zu wählen, da dann kein Laplace–Raum vorliegt.

Stattdessen denken wir uns die Kugeln durchnumeriert (von 1 bis N), wobei die ersten R Kugeln rot und die restlichen S Kugeln schwarz sind, und betrachten

$$\{i\} \;=\; \text{"Kugel } i \text{ wird gezogen"} \qquad (i = 1, \ldots, N)$$

als i-tes Elementarereignis mit $\Omega = \{1, 2 \ldots, N\}$.

Nach Voraussetzung gilt hier die Gleichwahrscheinlichkeit, so dass wir leicht

$$P(\text{"rote Kugel"}) \;=\; \frac{R}{N}$$

erhalten. Analog ergibt sich

$$P(\text{"schwarze Kugel"}) \;=\; \frac{S}{N} \;=\; \frac{N-R}{N} \;=\; 1 \;-\; \frac{R}{N}\;.$$

Nun betrachten wir die Ereignisse

$$\{SK\} = \text{"schwarze Kugel"} \qquad \text{und} \qquad \{RK\} = \text{"rote Kugel"}$$

als Elementarereignisse eines anderen Wahrscheinlichkeitsraumes mit $\Omega = \{SK, RK\}$ mit $P(\{RK\}) = \frac{R}{N} \neq \frac{S}{N} = P(\{SK\})$ falls $R \neq S$ gilt. Selbstverständlich gilt auch hier

$$P(\{SK\}) \;+\; P(\{RK\}) \;=\; \frac{S}{N} + \frac{R}{N} \;=\; \frac{S+R}{N} \;=\; 1$$

Dieses einfache Beispiel wollen wir dadurch erweitern, dass wir nun nicht nur einmal, sondern n-mal hintereinander in die Urne greifen, um eine Kugel zu ziehen. Nun lautet die Frage:

> Wie groß ist die Wahrscheinlichkeit $P(A_r)$, bei n-maligem Ziehen aus einer Urne mit R roten und S schwarzen Kugeln genau r rote Kugeln zu ziehen?

Dabei muss man offensichtlich zwei unterschiedliche Situationen unterscheiden:

I. Ziehen ohne Zurücklegen:
Nach jedem Zug wird die Farbe der Kugel notiert und die Kugel beiseite gelegt. Die Zahl der Kugeln in der Urne hat sich also um eine verringert. Das ist offensichtlich gleichwertig damit, dass man alle n Kugeln gleichzeitig zieht. Dabei muss offensichtlich $n \leq N$ gelten

II. Ziehen mit Zurücklegen:
Jede gezogene Kugel wird nach Notieren der Farbe vor dem nächsten Zug wieder in die Urne zurückgelegt. Der nachfolgende Zug findet also wieder unter den gleichen Bedingungen wie der vorhergehende statt. Hier gilt $n \in \mathbb{N}$ (Probe = mehrfache Menge).

(a) Urnenmodell ohne Zurücklegen (**Urnenmodell I**)

$$\Omega_{\mathrm{I}} = \{\omega : \omega = (i_1, \dots, i_n), i_j \in \{1, \dots, N\}, i_1 < i_2 < \cdots < i_n\}$$

(Probe vom Umfang n aus N ohne Wdhlg. ohne Beachtung der Reihenfolge = einfache Menge)

Wieder seien dabei die ersten R Kugeln rot und die übrigen schwarz. Dabei bedeutet die letzte Ungleichungskette, dass wir die Reihenfolge, in der die Kugeln gezogen wurden, nicht beachten wollen (die Nummern also grundsätzlich der Größe nach geordnet aufgeschrieben werden).
Offensichtlich liegt hier ein Laplace–Modell vor mit

$$|\Omega_{\mathrm{I}}| = \binom{N}{n}.$$

Wir müssen nun nur noch feststellen, wie viele Kombinationen der obigen Art existieren, die genau r rote Kugeln enthalten.

$$P(A_r) = \frac{\text{Anzahl der obigen Kombinationen mit genau } r \text{ Komponenten } i_j \leq R}{\binom{N}{n}}$$

Eine solche Kombination ist offensichtlich dadurch gekennzeichnet, dass genau r Kugeln rot und $s = n - r$ Kugeln schwarz sind. Davon gibt es aber offensichtlich

$$\binom{R}{r} \cdot \binom{S}{s} = \binom{R}{r} \cdot \binom{N-R}{n-r}$$

Stück.

Damit erhalten wir also

$$P(A_r) \;=\; \frac{\binom{R}{r}\cdot\binom{N-R}{n-r}}{\binom{N}{n}}\;,$$

falls $n-r \le N-R$, also $r \ge n-N+R = n-S$ und damit $\max\{0, n-S\} \le r \le \min\{R,n\}$ gilt. Anderenfalls ist diese Wahrscheinlichkeit 0.

Befinden sich in einer Urne R rote, W weiße und S schwarze Kugeln mit $R+S+W = N$ so ermittelt man analog für das Ereignis

$A_{r,s}$ = "unter n gezogenen Kugeln befinden sich genau r rote und s schwarze Kugeln"

$$P(A_{r,s}) \;=\; \frac{\binom{R}{r}\binom{S}{s}\binom{N-R-S}{n-r-s}}{\binom{N}{n}}$$

Beispiel 3.8 (Lotto: 6 aus 49)
Wie groß ist die Wahrscheinlichkeit, beim Zahlenlotto genau k Richtige zu haben, also von den aus 49 *Zahlen zufällig ausgewählten 6 Zahlen genau k richtig vorhergesagt zu haben? Dabei soll die ebenfalls gezogene Zusatzzahl nicht berücksichtigt werden.*

Es werden 6 Zahlen ohne Zurücklegen gezogen (Urnenmodell I). Fasst man in Gedanken die getippten 6 *Zahlen als die roten Kugeln auf und werden aus allen* $N = 49$ *Kugeln genau* $n = 6$ *gezogen, so erhält man*

$$P(A_k) \;=\; \frac{\binom{6}{k}\cdot\binom{43}{6-k}}{\binom{49}{6}} \qquad (k = 3,4,5,6)\;.$$

*Berücksichtigt man, dass noch eine siebte Zahl (**Zusatzzahl**) gezogen wird, so gilt für die Wahrscheinlichkeit, dass man neben der Zusatzzahl weitere* $k = 3,4,5$ *Zahlen "richtig" angekreuzt hat, offensichtlich*

$$P(A_{k+1}) \;=\; \frac{\binom{6}{k}\cdot\binom{42}{6-k-1}}{\binom{49}{6}} \qquad (k = 3,4,5)\;.$$

Für den Fall, dass man neben "Sechs Richtigen" auch noch die "richtige" **Superzahl** *hat, gilt ebenso*

$$P(A_{6+S}) \;=\; \frac{1}{10\cdot\binom{49}{6}} \;=\; \frac{1}{10}\cdot P(A_6)\;.$$

(b) <u>Urnenmodell mit Zurücklegen</u> (**Urnenmodell II**)

Die Urne enthält wie im Modell I R rote und S schwarze Kugeln. Jetzt wird lediglich vor dem nächsten Zug die zuvor gezogene Kugel wieder in die Urne zurückgelegt, so dass bei jedem Ziehen die gleiche Situation wie am Anfang vorliegt. So kann also die gleiche Kugel auch mehrfach gezogen werden.

Tabelle 3.2: Wahrscheinlichkeiten für k Richtige beim Lotto

$$\begin{array}{rclclclcl}
P(A_3) & = & \frac{\binom{6}{3}\cdot\binom{43}{3}}{\binom{49}{6}} & = & \frac{20\cdot 12341}{13983816} & = & 1.76504038\% & \approx & 1/57 \\
P(A_{3+1}) & = & \frac{\binom{6}{3}\cdot\binom{42}{2}}{\binom{49}{6}} & = & \frac{20\cdot 287}{4661272} & = & 0.12314\% & \approx & 1/812 \\
P(A_4) & = & \frac{\binom{6}{4}\cdot\binom{43}{2}}{\binom{49}{6}} & = & \frac{15\cdot 903}{13983816} & = & 0.09686197\% & \approx & 1/1032 \\
P(A_5) & = & \frac{\binom{6}{5}\cdot\binom{43}{1}}{\binom{49}{6}} & = & \frac{6\cdot 43}{13983816} & = & 0.00184499\% & \approx & 1/54201 \\
P(A_{5+1}) & = & \frac{\binom{6}{5}\cdot\binom{42}{0}}{\binom{49}{6}} & = & \frac{6}{13983816} & = & 0.00004\% & \approx & 1/2327749 \\
P(A_6) & = & \frac{\binom{6}{6}\cdot\binom{43}{0}}{\binom{49}{6}} & = & \frac{1\cdot 1}{13983816} & = & 0.00000715\% & = & 1/13983816 \\
P(A_{6+S}) & = & \frac{1}{10}\cdot\frac{\binom{6}{6}\cdot\binom{43}{0}}{\binom{49}{6}} & = & \frac{1\cdot 1}{139838160} & = & 0.000000715\% & = & 1/139838160
\end{array}$$

Damit liegt eine andere Situation vor, die andere gleichwahrscheinliche Elementarereignisse zu ihrer Beschreibung erfordert (Probe mit Wiederholung unter Beachtung der Reihenfolge = mehrfache Liste):

$$\Omega_{\mathrm{II}} \;=\; \left\{\omega :\; \omega = (i_1, \ldots, i_n)\;,\; i_j \in \{1, \ldots, N\}\right\}$$

Die so beschriebenen Elementarereignisse scheinen unter den gegebenen Bedingungen gleichwahrscheinlich zu sein. Mit

$$|\Omega_{\mathrm{II}}| \;=\; N^n$$

können wir nun die Wahrscheinlichkeit, genau r rote Kugeln unter den n gezogenen zu finden, ermitteln, wenn wir wissen, wie viele der obigen n–lässt genau r rote Kugeln enthalten ($|A_r|$).

1. Zunächst befinden sich die r roten Kugeln auf r Plätzen unter den n gezogenen Kugeln. Dafür gibt es insgesamt $\binom{n}{r}$ verschiedene Möglichkeiten solche Plätze auszusuchen.
2. Für jede dieser Möglichkeiten gibt es $R^r \cdot (N-R)^{n-r}$ verschiedene Arten r rote und $s = n - r$ schwarze Kugeln aus der Urne auf die entsprechenden Plätze zu verteilen.

Damit erhalten wir also

$$|A_r| \;=\; \binom{n}{r} \cdot R^r \cdot (N-R)^{n-r}$$

und für die gesuchte Wahrscheinlichkeit

$$P(A_r) \;=\; \binom{n}{r} \cdot \frac{R^r \cdot (N-R)^{n-r}}{N^{r+(n-r)}} \;=\; \binom{n}{r}\left(\frac{R}{N}\right)^r \left(1 - \frac{R}{N}\right)^{n-r} \quad .$$

Bemerkung

Das Urnenmodell II wird häufig als Approximation für das schwieriger zu berechnende Urnenmodell I benutzt. Ist die Anzahl der Züge n klein gegen die Anzahl der roten (R) und der schwarzen ($N-R$) Kugeln, so dass unabhängig, ob bei einem Zug eine rote oder schwarze Kugel gefunden wurde, die Gesamtsituation (das Verhältnis zwischen roten und schwarzen Kugeln) nahezu unverändert bleibt, müssen beide Modelle nahezu gleiche Ergebnisse (Wahrscheinlichkeiten) liefern.

Es gilt also im Modell I:

$$n \ll R, N-R \qquad \Longrightarrow \qquad P(A_r) \approx \binom{n}{r}\left(\frac{R}{N}\right)^r\left(1-\frac{R}{N}\right)^{n-r} .$$

Beispiel 3.9 (Volksbefragung) 75% *der* 40 *Millionen wahlberechtigten Bürger eines Staates seien mit ihrer Regierung unzufrieden . Von einem Meinungsforschungsinstitut wird eine Befragung von* 20 *"zufällig ausgewählten" Bürgern durchgeführt, ob sie tatsächlich unzufrieden sind. Unter der Annahme, dass jeder eine "ehrliche" Antwort gibt, soll die Wahrscheinlichkeit ermittelt werden, dass mehr als die Hälfte der Befragten positiv zur Regierung eingestellt sind, also höchstens* 10 *Befragte sich negativ äußern. R ist hier die Anzahl der negativ eingestellten Bürger.*

$$N = 4\cdot 10^7 \quad , \quad R = 3\cdot 10^7 \quad , \quad n = 20 \quad , \quad \frac{R}{N} = \frac{3}{4} = 0{,}75 = 75\% .$$

Gesucht ist also die Wahrscheinlichkeit, dass die Anzahl r der negativen Antworten höchstens 10 *ist.*

$$\begin{aligned} A_r &:= \text{"genau } r \text{ Befragte sind unzufrieden"} \\ B &:= \sum_{r=0}^{10} A_r \qquad \Longrightarrow \qquad P(B) = \sum_{r=0}^{10} P(A_r) \end{aligned}$$

$$\begin{aligned} P(B) &= \sum_{r=0}^{10} \frac{\binom{R}{r}\cdot\binom{N-R}{n-r}}{\binom{N}{n}} \\ &\approx \sum_{r=0}^{10}\binom{n}{r}\left(\frac{R}{N}\right)^r\left(1-\frac{R}{N}\right)^{n-r} = \sum_{r=0}^{10}\binom{20}{r}\left(\frac{3}{4}\right)^r\left(\frac{1}{4}\right)^{n-r} \\ &= \frac{1}{4^{20}}\sum_{r=0}^{10}\binom{20}{r}\cdot 3^r \\ &= \frac{1}{4^{20}}(1\cdot 1 + 20\cdot 3 + 190\cdot 9 + \cdots + 167960\cdot 19683 + 184756\cdot 59049) \\ &\approx 1{,}39\% \end{aligned}$$

3.3 Besetzungsmodelle

Insbesondere in der statistischen Physik interessiert man sich für die Verteilung von Zuständen verschiedener Elementarteilchen. Dabei können Zustände wie Ort, Geschwindigkeit, Energie quantifiziert betrachtet werden, so dass auch hier diskrete Wahrscheinlichkeitsräume mit höchstens abzählbar vielen Elementarereignissen (Zuständen) vorliegen.

Hier verwendet man ebenfalls Urnenmodelle, wobei es darum geht, n Kugeln (Teilchen) auf N Urnen (Zustände, Zellen eines Phasenraums) zu verteilen.

(a) Maxwell – Boltzmann – Statistik

Voraussetzung: Jede der n Kugeln fällt zufällig unabhängig von allen anderen in eine der N Urnen. Dabei ist es gleichgültig, wie viele Kungeln sonst noch in dieser urne liegen. Dazu fassen wir die Urnen und Kugeln als unterscheidbar auf und lassen pro Urne sind beliebig viele Kugeln zu. Wir notieren also, welche der n Kugeln in welcher der N Urnen liegt, und benutzen das Urnenmodell II. Ein zufälliges Verteilen der Kugeln auf die Urnen bedeutet hier also, dass jede Urne die gleiche Wahrscheinlichkeit von $1/N$ besitzt, von einer speziellen Kugel belegt zu sein. Versteht man unter dem Ereignis "$i_k = j$", dass die k-te Kugel in Urne j liegt, gilt:

$$\begin{aligned}
\Omega &= \{(i_1,\ldots,i_n) \,:\, i_1,\ldots,i_n \in \{1,\ldots,N\}\} \;, \\
|\Omega| &= N^n \;. \\
A_k^{(m)} &= \text{"In Urne } m \text{ befinden sich genau } k \text{ Kugeln"} \\
|A_k^{(m)}| &= (\text{Anz. d. Mögl., } k \text{ Kugeln aus } n \text{ auszuwählen}) \times \\
&\qquad\qquad (\text{Anz. d. Mögl., } n-k \text{ Kugeln auf } N-1 \text{ Zellen zu verteilen}) \\
&= \binom{n}{k}(N-1)^{n-k} \\
\Longrightarrow \qquad & P\left(A_k^{(m)}\right) = \binom{n}{k}\left(\frac{1}{N}\right)^k\left(1-\frac{1}{N}\right)^{n-k}
\end{aligned}$$

(b) Bose – Einstein – Statistik

Voraussetzung: Die Kugeln sind nicht unterscheidbar und pro Urne sind beliebig viele Kugeln zugelassen. Hier ist nur wichtig, wie viele (und nicht welche) Kugeln in einer Urne liegen. Eine zufällige Verteilung der Kugeln soll hier bedeuten, dass alle möglichen Verteilungen der Kugeln auf die Urnen die gleiche Wahrscheinlichkeit besitzen.

Eine mögliche Verteilung ω der Kugeln auf die Urnen wird dadurch dargestellt, dass n Kugeln und $N-1$ senkrechte Trennwände(Striche) linear nebeneinander in einer beliebigen Reihenfolge aufgezeichnet werden.

Abbildung 3.1: Zustand in der Bose–Einstein–Statistik

$$\omega = \bigcirc \;|\;|\; \bigcirc \;\; \bigcirc \,|\, \bigcirc \;|\;|\;|\; \bigcirc \;\; \bigcirc \;\; \bigcirc \;\; \bigcirc \;|$$

$n = 8$ Kugeln sind auf $N = 8$ Urnen verteilt. Dabei sind jeweils zwischen zwei aufeinanderfolgenden Strichen die Inhalte einer Urne angegeben. So enthält die erste Urne 1 Kugel (Anzahl der Kugeln vor dem ersten Strich), die zweite ist leer, die dritte enthält 2 und die vierte 1 Kugel. Die 5. und 6. Zelle sind leer, die 7. enthält 4 Kugeln und die 8. ist wiederum leer (Kugeln nach dem 7. Strich).

Auf diese Weise ergibt sich die Elementeanzahl von Ω als die Anzahl der Möglichkeiten, n Kugeln auf insgesamt $N + n - 1$ Plätze für Kugeln und Trennwände zu legen. Also gilt

$$|\Omega| = \binom{N-1+n}{n} = D(N,n) \; .$$

$$\begin{aligned}
A_k^{(m)} &= \text{"In Urne } m \text{ befinden sich genau } k \text{ Kugeln"} \\
|A_k^{(m)}| &= (\text{Anz. d. Mögl., } n-k \text{ Kugeln auf } N-1 \text{ Urnen zu verteilen}) \\
&= \binom{(N-1) \;-\; 1 \;+\; (n-k)}{n \;-\; k} = \binom{N+n-k-2}{n-k} \\
\Longrightarrow \qquad & P\left(A_k^{(m)}\right) = \frac{\binom{N+n-k-2}{n-k}}{\binom{N+n-1}{n}}
\end{aligned}$$

(c) Fermi – Dirac – Statistik

Voraussetzung: Wie bei der Bose–Einstein–Statistik, aber es ist nur eine Kugel pro Zelle erlaubt.

Ein mögliche Verteilung besteht aus einer Auswahl von n Urnen aus den insgesamt N für die n Kugeln. Es gilt also

$$|\Omega| = \binom{N}{n} \; .$$

$A_1^{(m)}$ ist also das Ereignis, dass die m-te Urne eine Kugel enthält.

$$\begin{aligned}
|A_1^{(m)}| &= (\text{Anz. d. Mögl., } n-1 \text{ Kugeln auf } N-1 \text{ Urnen zu verteilen}) \\
&= \binom{N-1}{n-1} \\
\Longrightarrow \qquad & P\left(A_1^{(m)}\right) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{(N-1)!\; n!\; (N-n)!}{(n-1)!\; (N-n)!\; N!} = \frac{n}{N}
\end{aligned}$$

Kapitel 4

Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Eine wichtige Aufgabe der Wahrscheinlichkeitstheorie besteht darin, Verfahren zu entwickeln, die es gestatten, Wahrscheinlichkeiten komplizierter Ereignisse auf Wahrscheinlichkeiten einfacher Ereignisse zurückzuführen. Häufig betrachtet man dabei mehrere Ereignisse in einem Wahrscheinlichkeitsraum und möchte wissen, ob die Kenntnis des Eintreten eines Ereignisses, Informationen über die Wahrscheinlichkeit des Auftretens eines anderen Ereignisses zur Folge hat.

Beispiel 4.1 *Die* $N = 1050$ *Wähler in einem Dorf bei Braunschweig werden befragt, welche Partei sie bei der letzten Bundestagswahl gewählt haben. Die Ergebnisse sind nach männlichen und weiblichen Personen aufgeschlüsselt.*

Anzahl	*CDU*	*SPD*	*FDP*	*Grüne*	*Sonstige*	*Gesamt*
männlich	*205*	*204*	*24*	*47*	*20*	*500*
weiblich	*273*	*169*	*28*	*55*	*25*	*550*
Gesamt	*478*	*373*	*52*	*102*	*45*	*1050*

Wenn wir nun "zufällig" einen der Wähler auswählen (jeder hat unabhängig von Geschlecht und gewählter Partei die gleiche Chance ausgewählt zu werden), betrachten wir die folgenden Ereignisse:

$$
\begin{array}{lcll}
\Omega & = & \text{"Wähler im Dorf"} & (|\Omega| \; = \; N \; = \; 1050) \\
M & = & \text{"männlich"} & (|M| \; = \; m \; = \; 500) \\
W & = & \text{"weiblich"} & (|W| \; = \; N - m) \; = \; 550) \\
C & = & \text{"CDU – Wähler"} & S \; = \; \text{"SPD – Wähler"} \\
F & = & \text{"FDP – Wähler"} & G \; = \; \text{"Grün – Wähler"} \\
R & = & \text{"Wähler einer sonstigen Partei"} &
\end{array}
$$

Unschwer erkennt man, dass die Zahl der männlichen CDU-Wähler

$$|C \cap M| \;=\; 205 \;,$$

beträgt und die der weiblichen FDP-Wähler

$$|F \cap W| \;=\; 28 \;.$$

Zunächst ermitteln wir die Wahrscheinlichkeiten, dass ein zufällig ausgewählter Wähler männlich bzw. weiblich ist:

$$\begin{aligned} P(M) &= \frac{|M|}{|\Omega|} = \frac{m}{N} = \frac{500}{1050} = 0,48 = 48\% \;, \\ P(W) &= \frac{|W|}{|\Omega|} = \frac{N-m}{N} = \frac{550}{1050} = 0,52 = 52\% \;. \end{aligned}$$

Ebenso durch Division durch die Gesamtanzahl N aller Wähler erhalten wir die Wahrscheinlichkeit einen weiblichen FDP-Wähler aus allen Wahlberechtigten auszuwählen zu

$$P(F \cap W) \;=\; \frac{|F \cap W|}{|N|} \;=\; \frac{28}{1050} \;=\; 0,03$$

Insgesamt ergeben sich in obiger Tabelle die entsprechenden **Wahrscheinlichkeiten für die Durchschnittsereignisse** *(Wahlverhalten differenziert nach Geschlecht) ebenso zu*

Wahrscheinl.	*CDU*	*SPD*	*FDP*	*Grüne*	*Sonstige*	*Gesamt*
männlich	*0,20*	*0,19*	*0,02*	*0,04*	*0,02*	*0,48*
weiblich	*0,26*	*0,16*	*0,03*	*0,05*	*0,02*	*0,52*
Gesamt	*0,46*	*0,36*	*0,05*	*0,09*	*0,04*	*1,00*

also beispielsweise:

$$P(S \cap M) \;=\; \frac{|S \cap M|}{|N|} \;=\; \frac{204}{1050} \;=\; 0,19$$

Zur Ermittlung der Wahrscheinlichkeit auf einen Mann, bzw. eine Frau zu stoßen bei Auswahl nur unter den Anhängern der FDP, müssen wir die Grundmenge Ω anpassen

$$\Omega' \;=\; F \quad und \quad M' \;=\; M \cap F$$

und erhalten die **bedingte Wahrscheinlichkeit von** M **unter der Bedingung** F *bzw. (***bei gegebenem** F*) durch*

$$P(M') \;=\; \frac{|M'|}{|\Omega'|} \;=\; \frac{|M \cap F|}{|F|} \;=\; \frac{\frac{|M \cap F|}{|N|}}{\frac{|F|}{|N|}} \;=\; \frac{P(M \cap F)}{P(F)} \;=:\; P(M \mid F) \;,$$

Analog erhalten wir die **bedingten Wahrscheinlichkeiten** *für die Auswahl eines Mannes bzw. einer Frau* **bei gegebenem Wahlverhalten** *durch Division jeweils durch die Gesamtanzahl der entsprechenden Wähler:*

bed. Wahrscheinl.	*CDU*	*SPD*	*FDP*	*Grüne*	*Sonstige*
männlich	*0,43*	*0,55*	*0,46*	*0,46*	*0,44*
weiblich	*0,57*	*0,45*	*0,54*	*0,54*	*0,56*
Gesamt	*1,00*	*1,00*	*1,00*	*1,00*	*1,00*

Wissen wir umgekehrt, dass der Betroffene männlich ist, ergibt sich die Wahrscheinlichkeit, dass er ein SPD-Wähler ist durch $\Omega' = M$ *und* $S' = S \cap M$ *und wir erhalten* **bedingte Wahrscheinlichkeit von** S **unter der Bedingung** M *bzw. (***bei gegebenem** M*) durch*

$$P(S') = \frac{|S'|}{|\Omega'|} = \frac{|S \cap M|}{|M|} = P(S \mid M) \quad (\textit{ bedingte W. von } S \textit{ unter der Bedingung } M) .$$

*Analog erhalten wir die bedingte Wahrscheinlichkeit, einen Wähler einer bestimmten Partei zu finden, wenn wir nur unter den Männern bzw. Frauen auswählen (***bedingte Wahrscheinlichkeit für das Wahlverhalten bei gegebenem Geschlecht***):*

Wahrscheinl.	*CDU*	*SPD*	*FDP*	*Grüne*	*Sonstige*	*Gesamt*
männlich	*0,41*	*0,41*	*0,05*	*0,09*	*0,04*	*1,00*
weiblich	*0,50*	*0,31*	*0,05*	*0,10*	*0,04*	*1,00*

Das Wissen, ob der/die Befragte ein Mann oder eine Frau ist, stellt eine Vorabinformation dar, die zu veränderten Wahrscheinlichkeiten für die Ereignisse unterschiedlicher "Wahlverhalten" führt.

Definition 4.1 (Bedingte Wahrscheinlichkeit)
Sei $(\Omega, \mathcal{P}(\Omega), P)$ *ein diskreter Wahrscheinlichkeitsraum mit den Ereignissen* $A, B \subseteq \Omega$ *und* $P(B) > 0$*, dann ist*

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

die **bedingte Wahrscheinlichkeit von** A **unter der Bedingung** B *bzw. die bed. W. von* A *bei gegebenem* B*.*

Satz 4.1 (Rechenregeln für bedingte Wahrscheinlichkeiten)

(4.1) $$B \subseteq A \implies P(A|B) = 1$$

(4.2) $$A \cap B = \emptyset \implies P(A|B) = 0$$

(4.3) $$A_1, A_2, \ldots \text{ paarw. disjunkt} \implies P\Big(\sum_i A_i \,\Big|\, B\Big) = \sum_i P(A_i|B)$$

(4.4) $$P(\overline{A}|B) = 1 - P(A|B)$$

(4.5) $$P\Big(\bigcap_{i=1}^n A_i\Big) \neq 0 \implies$$

$$P\Big(\bigcap_{i=1}^n A_i\Big) = P(A_n|A_{n-1}\cdots A_1)\cdot P(A_{n-1}|A_{n-2}\cdots A_1)\cdots P(A_2|A_1)\cdot P(A_1)$$

Beweis:

(4.1) $$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(4.2) $$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$$

(4.3)

$$P\Big(\sum_i A_i \,\Big|\, B\Big) = \frac{P\Big(\big(\sum_i A_i\big)\cap B\Big)}{P(B)} = \frac{P\Big(\sum_i A_i B\Big)}{P(B)} = \frac{\sum_i P(A_i B)}{P(B)}$$
$$= \sum_i \frac{P(A_i B)}{P(B)} = \sum_i P(A_i|B)$$

(4.4) $$1 = \frac{P(\Omega B)}{P(B)} = P(\Omega|B) = P(A+\overline{A}\mid B) = P(A|B) + P(\overline{A}|B)$$

(4.5)

$$\begin{aligned} P\Big(\bigcap_{i=1}^n A_i\Big) &= P(A_n|(A_{n-1}\cap\cdots\cap A_1))\cdot P(A_{n-1}\cap\cdots\cap A_1) \\ &= P(A_n|A_{n-1}\cdots A_1)\cdot P(A_{n-1}|A_{n-2}\cdots A_1)\cdot P(A_{n-2}\cap(A_{n-3}\cap\cdots\cap A_1)) \\ &= P(A_n|A_{n-1}\cdots A_1)\cdot P(A_{n-1}|A_{n-2}\cdots A_1)\cdots P(A_3|A_2A_1)\cdot P(A_2|A_1)\cdot P(A_1) \end{aligned}$$

■

Beispiel 4.2 (Signalübertragung) *Über einen "Kanal" werden die "Signale" 0 bzw. 1 übertragen. Aufgrund von Störungen können sich das gesendete und das empfangene Signal unterscheiden. Als Versuchsausgänge betrachten wir die Paare*

$$(i,j) \quad \textit{mit} \quad i,j \in \{0,1\} \quad \textit{und} \quad i = \textit{gesendetes } (j = \textit{empfangenes}) \textit{ Signal}\,.$$

Dann gilt also $\Omega = \{(0,0)\ ,\ (0,1)\ ,\ (1,0)\ ,\ (1,1)\}$ *(i.a. kein Laplaceraum) .*

Nun betrachten wir die folgenden Ereignisse:

$$\begin{array}{rcl} S_i & := & \text{"}i \text{ wird gesendet"} = \{(i,0),(i,1)\} \quad \text{für} \quad i=0,1 \quad , \\ E_i & := & \text{"}i \text{ wird empfangen"} = \{(0,i),(1,i)\} \quad \text{für} \quad i=0,1 \quad \text{und} \\ F & := & \text{"Fehler tritt auf"} = \{(1,0),(0,1)\} \ . \end{array}$$

In der Praxis kennt man sowohl die Wahrscheinlichkeiten $P(S_0)$ *und* $P(S_1)$ *mit denen die Signale gesendet werden, als auch die Fehlerwahrscheinlichkeiten für beide Signale*

$$f_i := P(F \mid S_i) \qquad i=0,1 \ ,$$

mit

$$f_0 = P\big(\{(0,1)\}\,|S_0\big) \neq P(\{(0,1)\}) \qquad \text{bzw.} \qquad f_1 = P\big(\{(1,0)\}\,|S_1\big) \neq P(\{(1,0)\}\ .$$

Die Angabe dieser beiden Fehlerwahrscheinlichkeiten f_0 *und* f_1 *ist in der Regel sinnvoll, da Übertragungsfehler häufig frequenzabhängig, also von der Art der übertragenen "Nachricht" abhängig sind.*

Natürlich interessiert man sich auch für den "totalen Fehler" des Übertragungskanals

$$\begin{array}{rcl} P(F) & = & P(F \cap \Omega) = P(F \cap (S_0 + S_1)) \\ & = & P(F \cap S_0) + P(F \cap S_1) \\ & = & P(F|S_0) \cdot P(S_0) + P(F|S_1) \cdot P(S_1) \\ & = & f_0 \cdot P(S_0) + f_1 \cdot P(S_1) \end{array}$$

Diese Berechnung der "totalen Wahrscheinlichkeit" eines Ereignisses Tupel sich auch verallgemeinern:

Satz 4.2 (Formel von der totalen Wahrscheinlichkeit)
Sei $(\Omega, \mathcal{P}(\Omega), P)$ *ein diskreter WR mit*

$$B, A_1, A_2, \ldots \subseteq \Omega \qquad \text{und} \qquad A_1, A_2, \ldots \ \text{paarweise disjunkt}\ ,$$

dann gilt

$$B \subseteq \sum_i A_i \qquad \Longrightarrow \qquad P(B) = \sum_i P(B|A_i) \cdot P(A_i)$$

Falls einzelne A_i *die Wahrscheinlichkeit 0 besitzen, also die bedingten Wahrscheinlichkeiten* $P(B|A_i)$ *nicht definiert sind, bleibt die Formel gültig, wenn die entsprechenden Summanden weggelassen werden.*

<u>Beweis</u>:

$$
\begin{aligned}
P(B) &= P\Big(B \cap \sum_i A_i\Big) = P\Big(\sum_i B \cap A_i\Big) \\
&= \sum_i P(B \cap A_i) = \sum_i P(B|A_i) \cdot P(A_i)
\end{aligned}
$$

$\blacksquare$

Beispiel 4.3 *Ein Informatikstudent ist begeisterter Angler. Dazu sucht er abwechselnd drei verschiedene Seen auf (Ereignisse S_1, S_2 und S_3). Seine Freundin weiß aus langjähriger Beobachtung, dass er diese Seen mit den folgenden Wahrscheinlichkeiten aufsucht:*

$$P(S_1) = \frac{1}{2} \quad \textit{(Mitanglerin ?)}\,, \quad P(S_2) = P(S_3) = \frac{1}{4}\,.$$

Die Wahrscheinlichkeit, innerhalb einer Stunde einen Fisch zu angeln (Ereignis F), ist an den Seen unterschiedlich, nämlich

$$P(F|S_1) = \frac{2}{3} \quad , \quad P(F|S_2) = \frac{3}{4} \quad \textit{und} \quad P(F|S_3) = \frac{4}{5}\,.$$

An einem Sonntag geht er angeln, ohne sein Ziel mitzuteilen. Mit welcher Wahrscheinlichkeit angelt er innerhalb einer Stunde einen Fisch?

$$
\begin{aligned}
P(F) &= P(F|S_1) \cdot P(S_1) + P(F|S_2) \cdot P(S_2) + P(F|S_3) \cdot P(S_3) \\
&= \frac{2}{3} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{4} + \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{3} + \frac{3}{16} + \frac{1}{5} \\
&= \frac{173}{240} = 72,1\%
\end{aligned}
$$

Nach einer Stunde ruft der Student bei seiner Freundin an, dass er noch keinen Fisch geangelt hat, ohne mitzuteilen an welchem See er sich befindet. Wie groß ist die Wahrscheinlichkeit, dass er am See 1 angelt ($P(S_1|\overline{F})$) ?

Satz 4.3 (Formel von Bayes)
Unter den Voraussetzungen des Satzes 4.2 mit $P(B) > 0$ gilt

$$P(A_k) > 0 \qquad \Longrightarrow \qquad P(A_k|B) = \frac{P(B|A_k) \cdot P(A_k)}{\sum_i P(B|A_i) \cdot P(A_i)} \quad \forall k$$

<u>Beweis</u>:

$$P(A_k|B) = \frac{P(BA_k)}{P(B)} = \frac{P(B|A_k) \cdot P(A_k)}{\sum_i P(B|A_i) \cdot P(A_i)}$$

$\blacksquare$

Beispiel 4.4 (Angler) *Für die gesuchte Wahrscheinlichkeit aus Beispiel 4.3 dafür, dass der Student nach seinem Anruf sich am See 1 (Mitanglerin) aufhält, gilt demnach:*

$$P(S_1|\overline{F}) \;=\; \frac{P(\overline{F}|S_1)\cdot P(S_1)}{P(\overline{F})} \;=\; \frac{\frac{1}{3}\cdot\frac{1}{2}}{1\;-\;\frac{173}{240}} \;=\; \frac{40}{67} \;=\; 59,7\% \;>\; P(S_1) \;=\; 50\%$$

Häufig kommt es vor, dass die Wahrscheinlichkeit für das Eintreten eines Ereignisses A nicht davon abhängt, ob eine Hypothese (Ereignis: H) zutrifft oder nicht gilt. In der Sprache der bedingten Wahrscheinlichkeiten hieße das

$$P(A|H) \;=\; P(A)\;.$$

Umgangssprachlich sagt man, dass das Ereignis A von der Hypothese H unabhängig ist.

Sei F das Ereignis, dass ein zufällig ausgewählter Student einen Fisch innerhalb einer Stunde fängt, und B das Ereignis, dass ein zufällig ausgewählter Student die CDU wählt, so scheint es klar zu sein, dass F von B unabhängig ist. Bestehen Zweifel an dieser Aussage, müsste man durch wiederholtes Beobachten versuchen, die Abhängigkeit bzw. Unabhängigkeit zu belegen.

Besonders interessant ist dies bei Ereignissen wie z.B.
A = "zufällig ausgewählter Mensch stirbt an Lungenkrebs" und
B = "zufällig ausgewählter Mensch ist Raucher" .

Mathematisch wählt man für die Definition der Unabhängigkeit nicht den obigen intuitiven Zugang mit Hilfe der bedingten Wahrscheinlichkeiten, sondern fordert:

Definition 4.2 *Seien $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter WR und $A, B \subseteq \Omega$ beliebige Ereignisse, dann heißen A und B (stochastisch)* **unabhängig** *genau dann, wenn gilt:*

$$P(A\cap B) \;=\; P(A)\cdot P(B)\;.$$

Der folgende Satz zeigt, dass diese Definition i.a. dem intuitiven Zugang entspricht.

Satz 4.4 *Unter den Voraussetzungen der Definition 4.2 gilt für alle Ereignisse B mit $P(B) > 0$*

$$A, B \textit{ unabhängig} \qquad \Longleftrightarrow \qquad P(A|B) \;=\; P(A)$$

<u>Beweis</u>:

$\Longrightarrow$:
$$P(A|B) \;=\; \frac{P(A\cap B)}{P(B)} \;=\; \frac{P(A)\cdot P(B)}{P(B)} \;=\; P(A)$$

bzw.

$\Longrightarrow$:
$$P(A\cap B) \;=\; P(A|B)\cdot P(B) \;=\; P(A)\cdot P(B)$$

Im allgemeinen wird man ohne weitere Prüfung feststellen, dass Ereignisse, die in keinem kausalen Zusammenhang stehen, auch stochastisch unabhängig sind. So sind zwei aufeinanderfolgende Würfe beim Würfeln oder Ziehungen beim Zahlenlotto sicher auch stochastisch unabhängig. Es ist also nicht zu erwarten, dass eine Strategie beim Ausfüllen des Lottoscheins, die von den vorhergehenden Lottozahlen Gebrauch macht ("keine Zahl erneut tippen", "gesamte Folge wiederholen") irgendeinen geführt auf die Gewinnchancen hat (auch keinen negativen!).

Ebenso oft wird man auch zwischen stochastisch abhängigen Ereignissen einen kausalen Zusammenhang vermuten, der aber i.a. nicht unmittelbar vorhanden sein muss. Auch stochastische Unabhängigkeit ist keine Garantie für das Fehlen eines kausalen Zusammenhangs. Man sollte also vorsichtig mit der Interpretation solcher Ergebnisse sein, sondern nachgewiesene stochastische Abhängigkeiten nur dazu benutzen, gezielt nach kausalen Zusammenhängen zu suchen.

Beispiel 4.5 *So haben beispielsweise statistische Untersuchungen ergeben, dass parallel zur Abnahme der Geburten in Niedersachsen auch die Zahl der Weissstörche abgenommen hat. Aus der stochastischen Abhängigkeit auf einen kausalen Zusammenhang zu schließen, erscheint jedoch etwas abenteuerlich.*

Satz 4.5

(1) A, B *unabhängig* $\Longrightarrow$ $A, \overline{B}$, $\overline{A}, B$ *und* $\overline{A}, \overline{B}$ *unabhängig*

(2) $AB = \emptyset$ *und* $P(A), P(B) > 0$ $\Longrightarrow A, B$ *nicht unabhängig (= abhängig)*

Beweis:

(1) : A, B unabhängig $\Longrightarrow$

$$\begin{aligned} P(A\overline{B}) &= P(A \backslash AB) = P(A) - P(AB) = P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot (1 - P(B)) = P(A) \cdot P(\overline{B}) \end{aligned}$$

$$P(\overline{A}B) = P(B\overline{A}) = P(B) \cdot P(\overline{A}) = P(\overline{A}) \cdot P(B)$$

$$\begin{aligned} P(\overline{A}\,\overline{B}) &= P(\overline{A} \backslash (\overline{A}B)) = P(\overline{A}) - P(\overline{A}B) \\ &= P(\overline{A}) - P(\overline{A}) \cdot P(B) = (1 - P(A)) \cdot (1 - P(B)) = P(\overline{A}) \cdot P(\overline{B}) \end{aligned}$$

(2) : Der Beweis wird indirekt geführt.

$$A, B \text{ unabhängig} \quad \Longrightarrow$$

$$0 < P(A) \cdot P(B) = P(AB) = P(\emptyset) = 0 \qquad \text{Widerspruch!}$$

$\blacksquare$

Beispiel 4.6 (Geschlechtsverteilung in einer Familie) *Wir betrachten nur Familien mit drei Kindern. Es werde das Geschlecht jedes Kindes (m = "männlich" und w = "weiblich") dem Alter nach geordnet notiert.*

$$\Omega \;=\; \{mmm, mmw, mwm, \dots\} \qquad mit \qquad |\Omega| \;=\; 2^3 = 8$$

Wir betrachten die folgenden Ereignisse :

H : *"Eine zuf. ausgewählte Familie hat mindest. einen Jungen und mindest. ein Mädchen"*
A : *"Eine zuf. ausgewählte Familie hat höchstens ein Mädchen"*

Hier kann man einen kausalen Zusammenhang weder unmittelbar feststellen noch verneinen.

Nun gilt aber:

$$\begin{aligned}
H &= \Omega\backslash\{mmm, www\} \quad mit \quad |H| \;=\; 8-2 \;=\; 6 \\
A &= \{mmm, mmw, mwm, wmm\} \quad mit \quad |A| \;=\; 4 \\
P(H) &= \frac{2\cdot 3}{8} \;=\; \frac{3}{4} \quad , \quad P(A) \;=\; \frac{4}{8} \;=\frac{1}{2} \\
A\cap H &= \{mmw, mwm, wmm\} \quad mit \quad |A\cap H| \;=\; 3 \\
P(A\cap H) \;=\; \frac{3}{8} &= \frac{3}{4}\cdot\frac{1}{2} \;=\; P(H)\cdot P(A)
\end{aligned}$$

Die Ereignisse H und A sind also unabhängig.

Aber*: Man rechnet leicht nach, dass die entsprechenden Ereignisse in Familien mit 2 oder 4 Kindern* <u>*nicht*</u> *unabhängig sind.*

Beispiel 4.7 (Ziegenproblem) *In einer amerikanischen Fernsehshow kann der Kandidat ein Auto gewinnen, wenn er errät hinter welcher von drei Türen das Auto steht. Hinter den anderen Türen steht je eine Ziege. Das Raten läuft in zwei Stufen ab. Zunächst tippt der Kandidat auf eine Tür. Dann öffnet der Moderator, der weiß, wo das Auto steht, eine nichtgetippte Tür, hinter der eine Ziege steht. Jetzt bekommt der Kandidat noch einmal die Möglichkeit seine erste Entscheidung für eine Tür zu revidieren. Lohnt sich für ihn eine Änderung, sollte er besser bei seiner ersten Entscheidung bleiben oder ist es gleichgültig was er tut?*

Der Begriff der Unabhängigkeit von 2 Ereignissen Tupel sich auf verschiedene Arten auf die Unabhängigkeit von n Ereignissen übertragen.

Definition 4.3 (paarweise Unabhängigkeit)
Die Ereignisse $A_1, A_2, \dots, A_n$ heißen **paarweise unabhängig***, wenn für alle Paare mit A_i, A_j $(i \neq j)$*

$$P(A_i \cap A_j) \;=\; P(A_i)\cdot P(A_j)$$

gilt.

Definition 4.4 (vollständige Unabhängigkeit)
Die Ereignisse $A_1, A_2, \ldots, A_n$ *heißen* **vollständig unabhängig**, *wenn für jedes* $k \geq 2$ *und für jede Teilauswahl* $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ *(*$i_1, \ldots, i_k$ *paarweise verschieden) der Länge* k *aus diesen Ereignissen*

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) \;=\; P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

gilt.

Bemerkung:

Für die Definition 4.3 müssen $\binom{n}{2}$ Gleichungen und für Definition 4.4 insgesamt

$$\binom{n}{2} + \binom{n}{3} + \cdots : + \binom{n}{n} \;=\; 2^n - n - 1$$

Gleichungen überprüft werden (einschließlich der Gleichungen für Definition 4.3). Damit gilt auch

$$A_1, \ldots, A_n \quad \text{vollst. unabh.} \qquad \Longrightarrow \qquad A_1, \ldots, A_n \quad \text{paarw. unabh.}$$

Die Umkehrung gilt i.a. nicht.

Satz 4.6 *Sind die Ereignisse* $A_1, A_2, \ldots A_n \subseteq \Omega$ *vollst. unabhängig und gilt* $P(A_i) > 0$ $\forall\ i = 1, \ldots, n$, *dann gilt für beliebige Auswahlen* $A_{i_1}, \ldots, A_{i_k}$ *der Länge* $k = 1, \ldots, n-1$

$$P(A_j \mid A_{i_1} \cdots A_{i_k}) \;=\; P(A_j) \qquad \forall\ j \neq i_l\ (l = 1, \ldots, k)\ .$$

Beweis:

Wegen der vollständigen Unabhängigkeit der A_i gilt wegen $P(A_i) > 0\ (i = 1, \ldots, n)$ auch für alle Teilauswahlen der A_i

$$P(A_{i_1} \cdots A_{i_k}) \;=\; P(A_{i_1}) \cdots P(A_{i_k}) \;>\; 0\ .$$

So existiert auch die folgende bedingte Wahrscheinlichkeit und es gilt

$$\begin{aligned} P(A_j \mid A_{i_1} \cdots A_{i_k}) \;&=\; \frac{P(A_{i_1} \cdots A_{i_k} \cdot A_j)}{P(A_{i_1} \cdots A_{i_k})} \;=\; \frac{P(A_{i_1}) \cdots P(A_{i_k}) \cdot P(A_j)}{P(A_{i_1}) \cdots P(A_{i_k})} \\ &=\; P(A_j)\ . \end{aligned}$$

$\blacksquare$

Beispiel 4.8 *Ein stochastisches Experiment bestehe aus einem regulären Münzwurf (die Ereignisse "Kopf" =* K *und "Zahl" =* Z *treten jeweils mit der W.* $1/2$ *ein) und dem Wurf mit einem Würfel ("1" =* 1 *, ... , "6" =* 6 *jeweils mit W.* $1/6$*). Zwischen beiden Wurfergebnissen besteht offensichtlich kein kausaler Zusammenhang, so dass für die Kombinationsergebnisse (Unabhängigkeit) gilt:*

1. **Zufallsexperiment Münzwurf***:*

$$WR: \qquad (\Omega_M, \mathcal{P}(\Omega_M), P_M)\ ,\ \Omega_M \ = \ \{K, Z\}\ ,\ P_M(\{K\}) \ = \ P_M(\{Z\}) \ = \ \frac{1}{2}$$

2. **Zufallsexperiment Würfeln***:*

$$WR: \qquad (\Omega_W, \mathcal{P}(\Omega_W), P_W)\ ,\ \Omega_W \ = \ \{1, \ldots, 6\}\ ,\ P_W(\{i\}) \ = \ \frac{1}{6} \quad \forall i$$

Für den Gesamt – Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ *gilt dann*

$$\Omega \ = \ \Omega_W \times \Omega_M \ = \ \{1, \ldots, 6\} \times \{K, Z\} \ = \ \Big\{(i, m)\ :\ i \in \{1, \ldots, 6\}\ ,\ m \in \{K, Z\}\Big\}$$

mit

$$P(\{i\} \cap \{m\}) \ = \ P_W(\{i\}) \cdot P_M(\{m\}) \ = \ \frac{1}{6} \cdot \frac{1}{2} \ = \ \frac{1}{12} \qquad (i = 1, \ldots, 6\ ,\ m = K, Z)$$

Damit ist die Wahrscheinlichkeit P auf dem obigen **Produktraum** *wegen der Unabhängigkeit die obige* **Produktwahrscheinlichkeit** $P = P_W \otimes P_M$.

Nun wollen wir diese Bezeichnungen auf Kombinationen von n unabhängigen Zufallsexperimenten ausdehnen.

Definition 4.5 (Produktraum)
$(\Omega_1, \mathcal{P}(\Omega_1), P_1)\ ,\ \ldots\ ,\ (\Omega_n, \mathcal{P}(\Omega_n), P_n)$ *seien n diskrete Wahrscheinlichkeitsräume.*

$(\Omega, \mathcal{P}(\Omega), P)$ *heißt der zugehörige* **Produktraum***, wenn gilt*

$$\Omega \ = \ \Omega_1 \times \cdots \times \Omega_n \tag{4.6}$$

$$P(\{(\omega_1, \ldots, \omega_n)\}) \ = \ P_1(\{\omega_1\}) \cdots P_n(\{\omega_n\}) \qquad \forall \omega = (\omega_1, \ldots, \omega_n) \in \Omega \tag{4.7}$$

$P = P_1 \otimes \cdots \otimes P_n$ *heißt die* **Produktwahrscheinlichkeit** *von $P_1, \ldots, P_n$*

Bemerkung (Rechnen in Produkträumen):

Sei $(\Omega, \mathcal{P}(\Omega), P)$ der obige diskrete Produktraum, dann gilt für beliebige $A_i \in \Omega_i$:

$$P(A_1 \times \cdots \times A_n) \ = \ P_1(A_1) \cdots P_n(A_n)$$

Beispiel 4.9 (Bernoulli–Experiment, Binomialverteilung)
Im Urnenmodell II des Abschnitts 3.2 haben wir durch Zurücklegen der jeweils gezogenen Kugel erreicht, dass bei jedem erneuten Ziehen die gleiche Situation herrscht, aufeinanderfolgende Ziehungen sich also nicht gegenseitig beeinflussen können. Die Ergebnisse aller Ziehungen sind also vollständig unabhängig.

In einer Verallgemeinerung dazu betrachten wir zunächst die i-te Ziehung mit den möglichen Ausgängen "Erfolg"(E) oder "Misserfolg"($\overline{E}$).

$$\Omega_i \;=\; \{E, \overline{E}\} \qquad (i = 1, \ldots, n)$$

und erhalten für den gesamten Raum

$$\Omega \;=\; \Omega_1 \times \cdots \times \Omega_n \;=\; \{E, \overline{E}\}^n \;.$$

Ebenso erhalten wir für die Wahrscheinlichkeiten

$$P_i(\{E\}) \;=\; p \qquad \textit{bzw.} \qquad P_i(\{\overline{E}\}) \;=\; q \;=\; 1 - p \qquad (i = 1, \ldots, n)$$

Wegen der Unabhängigkeit der Ziehungen gilt also

$$P \;=\; P_1 \otimes \cdots \otimes P_n \;.$$

Nun interessieren wir uns für das Ereignis

$$A_k \;=\; \text{"in } n \text{ Ziehungen genau } k\text{-mal Erfolg"} \;.$$

Ein $\omega = (\omega_1, \ldots, \omega_n)$ *mit genau k-mal "Erfolg" besitzt die Produktwahrscheinlichkeit*

$$P(\{\omega\}) \;=\; p^k (1-p)^{n-k} \;.$$

Insgesamt gibt es aber $\binom{n}{k}$ *verschiedene solcher Ereignisse mit dieser W. (Anzahl der Möglichkeiten, k Erfolgsereignisse auf die n Ziehungen zu verteilen), so dass insgesamt*

$$P(A_k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, \ldots, n)$$

gilt.

Definition 4.6 (Bernoulli–Experiment) *Ein Zufallsexperiment mit zwei möglichen Ausgängen E (Erfolg) oder $\overline{E}$ (Ingenieure) mit den Wahrscheinlichkeiten*

$$P(E) \;=\; p \qquad \textit{und} \qquad P(\overline{E}) \;=\; 1 - p$$

heißt **Bernoulli-Experiment**. *Eine n-malige unabhängige Ausführung eines Bernoulli–Experimentes heißt* **n-faches Bernoulli–Experiment**.

Satz 4.7 *Für die Wahrscheinlichkeit, bei einem n-fachen Bernoulli–Experiment genau k–mal Erfolg zu haben, gilt*

$$P(A_k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, \ldots, n) \;.$$

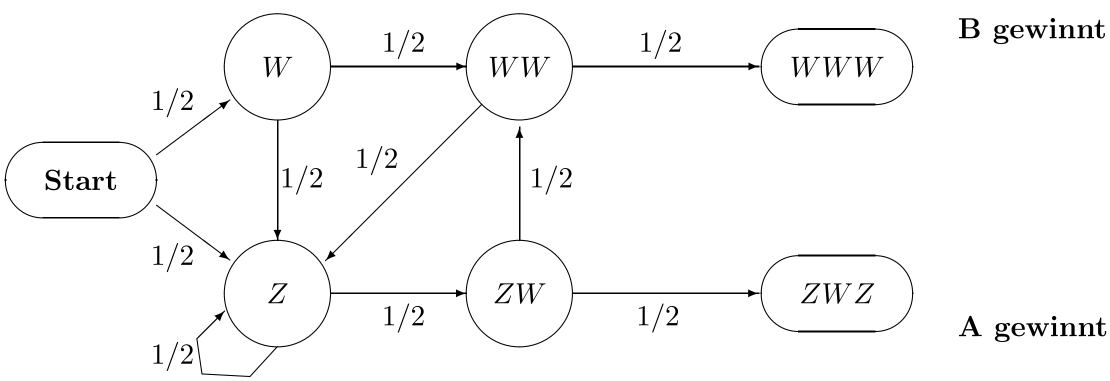
Zum Abschluss wollen wir noch das folgende Spiel betrachten:

A und B werfen eine symmetrische Münze (Wahrscheinlichkeit für Wappen (W) und Zahl (Z) jeweils 1/2), bis eine der folgenden Dreiersequenzen auftritt. A gewinnt, falls zuerst ZWZ und B, falls zuerst WWW auftaucht. Da B diese Vorlesung bereits zum zweiten Mal hört und weiß, dass beide Sequenzen gleich wahrscheinlich sind, hält er dieses Spiel für fair. Wie groß ist aber nun die Erfolgswahrscheinlichkeit für beide Spieler tatsächlich?

Wollen wir den Wahrscheinlichkeitsraum nach dem bisherigen Muster aufschreiben, so werden wir feststellen, dass das zwar möglich, aber außerordentlich schwierig ist. So wollen wir die möglichen Ausgänge und ihre Wahrscheinlichkeiten anders notieren. So ist es z.B. nicht erforderlich, die gesamte geworfenen Sequenz zu notieren, sondern nur die für den Fortgang erforderlichen letzten Sequenzen. So reicht bei der Sequenz $WWZZWWZW$ das Festhalten der beiden letzten Ergebnisse ZW, da nur sie für die Entscheidung über den Gewinn bedeutsam sind.

Das Diagramm enthält die relevanten Wurfergebnisse (Zustände) und die bedingten Wahrscheinlichkeiten (jeweils 1/2), mit denen der jeweilige Folgezustand erreicht wird.

Abbildung 4.1: übergangsdiagramm für ein Münzwurfspiel



Nun wollen wir die Wahrscheinlichkeit q für den Sieg des Spielers A ermitteln. Dabei verwenden wir die folgenden Bezeichnungen:

$$q(S) \;:=\; P(\text{A gewinnt} \mid \text{Spiel im Zustand } S) \qquad \forall\, S = \text{"relevanter Zustand"}$$

mit

$$S \;\in\; \{W, Z, WW, ZW, WWW, ZWZ\}$$

Aus der Formel für die totale Wahrscheinlichkeit gewinnen wir durch

$$\begin{aligned} q &= P(\text{A gewinnt}) \\ &= P(\text{A gewinnt} \mid \text{Spiel im Zustand } Z) \times P(\text{Spiel im Zustand } Z) \\ &\quad + P(\text{A gewinnt} \mid \text{Spiel im Zustand } W) \times P(\text{Spiel im Zustand } W) \\ &= \frac{1}{2} \cdot q(Z) \;+\; \frac{1}{2} \cdot q(W) \end{aligned}$$

die folgenden Gleichungen:

$$q = \frac{1}{2} \cdot q(W) + \frac{1}{2} \cdot q(Z) \,, \tag{4.8}$$

$$q(W) = \frac{1}{2} \cdot q(WW) + \frac{1}{2} \cdot q(Z) \,, \tag{4.9}$$

$$q(Z) = \frac{1}{2} \cdot q(Z) + \frac{1}{2} \cdot q(ZW) \,, \tag{4.10}$$

$$q(WW) = \frac{1}{2} \cdot q(Z) \quad \text{und} \tag{4.11}$$

$$q(ZW) = \frac{1}{2} + \frac{1}{2} \cdot q(WW) \,. \tag{4.12}$$

Die Gleichungen 4.9 – 4.12 stellen ein eindeutig lösbares lineares Gleichungssystem für die obigen bedingten Wahrscheinlichkeiten dar. Nach kurzer Rechnung ergibt sich:

$$q(Z) = q(ZW) = \frac{2}{3} \quad , \quad q(WW) = \frac{1}{3} \quad \text{und} \quad q(W) = \frac{1}{2} \,.$$

Setzt man diese Lösungen in die Gleichung 4.8 ein, so erhält man

$$q = \frac{7}{12} > \frac{1}{2} \,.$$

Spieler A hat größere Chancen, das Spiel zu gewinnen. Es ist also nicht fair.

Kapitel 5

Zufallsvariable

5.1 Definition und Beispiele

Einem diskreten Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{P}(\Omega), P\big)$ liegt stets ein konkretes Zufallsexperiment mit den möglichen Ausgängen $\omega \in \Omega$ zugrunde, die zunächst verbal beschrieben werden. In der Regel werden wir Versuchsausgänge bzw. daraus zusammengesetzte Ereignisse numerisch bewerten, d.h. i.a. durch Angabe einer reellen Zahl festlegen.

Beispiel 5.1 (n–faches Bernoulli–Experiment) $\Omega = \{E, \overline{E}\}^n$ *mit* $P(\{E\}) = p$ *und den interessierenden Ereignissen*

$$A_k \;=\; \text{"Genau } k\text{–mal Erfolg"} \qquad \text{mit} \qquad P(A_k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \;.$$

Hier interessiert man sich nicht mehr für die Ausgangsereignisse, sondern nur noch für die Größe k*. Wir ordnen jedem Versuchsausgang* ω *die Anzahl der "Erfolge"* X *zu. Die tatsächliche Größe von* X *hängt also von* ω *ab:*

$$X(\omega) \;=\; \text{Anzahl der Erfolge in } \omega$$

Damit stellt X *eine Funktion*

$$X \;:\; \Omega \;\longrightarrow\; I\!R$$

dar.

Diese Funktion X *(numerische Bewertung der Versuchsausgänge* ω*, bzw. der interessierenden Ereignisse* A_k*) benutzen wir nun dazu, die Ereignisse* A_k *kürzer auszudrücken:*

$$A_k \;=\; \{\omega \;:\; X(\omega) = k\} \quad \text{bzw.}$$

$$P(A_k) \;=\; P\Big(\{\omega \;:\; X(\omega) = k\}\Big) \;=:\; P(X \;=\; k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \;.$$

Man sagt dann auch, "**Die Zufallsgröße X ist binomialverteilt**", *und schreibt*

$$X \sim Bi(n,p) \; .$$

Beispiel 5.2 (Hypergeometrische Verteilung) *Der zugrunde liegende Wahrscheinlichkeitsraum wird durch das Urnenmodell I (Ziehen ohne Zurücklegen) beschrieben. Auch hier interessieren wir uns i.a. nur für die Anzahl r der gezogenen "roten Kugeln" und bezeichnen sie mit X .*

$$\begin{aligned} A_r &= \text{"Anzahl der roten Kugeln beim } n\text{–fachen Ziehen ohne Zurücklegen } = r\text{"} \\ &= \{\omega \, : \, X(\omega) = r\} \quad \text{und damit} \\ P(A_r) &= P\Big(\{\omega \, : \, X(\omega) = r\}\Big) \; = \; P(X \; = \; r) \; = \; \frac{\binom{R}{r} \cdot \binom{N-R}{n-r}}{\binom{N}{n}} \end{aligned}$$

Man sagt: **Die Zufallsgröße X ist hypergeometrisch verteilt**

und schreibt $\quad X \sim Hyp(n; N, R) \quad .$

Definition 5.1 (Diskrete Zufallsvariable) *Sei $\big(\Omega, \mathcal{P}(\Omega), P\big)$ ein diskreter Wahrscheinlichkeitsraum, dann heißt eine Abbildung*

$$X \, : \, \Omega \longrightarrow \mathbb{R}$$

eine **diskrete (reellwertige) Zufallsvariable (ZV)** *auf dem obigen WR. In der Literatur werden auch die Bezeichnungen* **Zufallsgröße** *oder seltener (weil irreführend)* **zufällige Funktion** *benutzt.*

Mit Hilfe des Begriffs der Zufallsvariablen sind wir in der Lage, eine Reihe von häufig interessierenden Ereignissen eleganter mathematisch auszudrücken.

$[X = x] \; := \; \{\omega \in \Omega \, : \, X(\omega) = x\}$ "X nimmt den Wert x an"
$[X \in A] \; := \; \{\omega \in \Omega \, : \, X(\omega) \in A\}$ "X nimmt einen Wert aus A an"
$[X \leq x] \; := \; \{\omega \in \Omega \, : \, X(\omega) \leq x\}$ "X nimmt einen Wert kleiner oder gleich x an"

Definition 5.2 *Sei X eine diskrete ZV auf dem WR $(\Omega, \mathcal{P}(\Omega), P)$, dann heißen*
$\Omega_X \; = \; \{x \in \mathbb{R} \, : \, X(\omega) = x$ *für mindestens ein* $\omega \in \Omega\}$ *der* **Bildraum** *von Ω unter X und*
$f_X \, : \, \Omega_X \; \mathbb{R}^1$ *mit* $f_X(x) \; := \; P(X = x)$ *die* **(diskrete) Verteilungsdichte** *bzw. die* **Zähldichte** *von X.*
Umgekehrt wird durch f_X auf Ω_X eine Wahrscheinlichkeit

$$P_X(B) \; := \; \sum_{x \in B} f_X(x) \qquad \forall \; B \subseteq \Omega_X$$

festgelegt. $\big\{\big(x, f_X(x)\big) \, : \, x \in \Omega_X\big\}$ *heißt auch die* **(Wahrscheinlichkeits–)Verteilung von** X.

Bemerkungen:

1. Man sagt auch, dass die Zufallsvariable $X : \Omega \longrightarrow \Omega_X$ den Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{P}(\Omega), P\big)$ auf den Wahrscheinlichkeitsraum $\big(\Omega_X, \mathcal{P}(\Omega_X), P_X\big)$ abbildet und schreibt:
$$X \;:\; \big(\Omega, \mathcal{P}(\Omega), P\big) \quad\longrightarrow\quad \big(\Omega_X, \mathcal{P}(\Omega_X), P_X\big) \quad \textbf{(von } X \textbf{ induzierter WR)}\ .$$

2. Für die Wahrscheinlichkeiten P_X gilt für alle $B \subseteq \Omega_X$:
$$\begin{aligned} P_X(B) &= \sum_{x:x\in B} f_X(x) \\ &= \sum_{x:x\in B} P(X=x) \;=\; \sum_{x:x\in B} P(\{\omega \;:\; X(\omega)=x\}) \\ &= P(\{\omega \;:\; X(\omega)\in B\}) \;=\; P\big(X^{-1}(B)\big) \end{aligned}$$
$X^{-1}(B)$ ist das **Urbild** von B bzgl. der Zufallsvariablen (Abbildung) X, also die Menge der Versuchsausgänge aus Ω, die zum Eintreten des Ereignisses $B \subseteq \Omega_X$ führen.

3. Statt "X ist ZV auf dem diskreten WR $(\Omega, \mathcal{P}(\Omega), P)$" schreiben wir i.a. kürzer "X ist diskrete ZV auf Ω".

4. Seien X, Y, Z diskrete reellwertige ZV auf Ω und A, B, C Teilmengen von $\mathbb{R}$, so schreiben wir:
$$P\Big(\{\omega : X(\omega)\in A\} \cap \{\omega : Y(\omega)\in B\} \cap \{\omega : Z(\omega)\in C\}\Big) \;=:\; P\big(X\in A, Y\in B, Z\in C\big)$$

5. Wie üblich ist $X+Y$ definiert als die Abbildung
$$X+Y \;:\; \omega \;\longmapsto\; X(\omega)+Y(\omega)\ .$$
Analog sind $\dfrac{X}{Y}$, falls $Y(\omega)\neq 0$ für alle $\omega\in\Omega$ und
$$X\cdot Y\ ,\ X^k\ ,\ g(X)\ ,\ h(X,Y)\ ,\ldots$$
mit $h \;:\; \omega \;\longmapsto\; h\big(X(\omega), Y(\omega)\big) \in \mathbb{R}$ definiert und ebenfalls diskrete Zufallsvariable.

Beispiel 5.3 (Geometrische Verteilung) *Ein Bernoulli–Experiment mit der Erfolgswahrscheinlichkeit p für das Ereignis E wird so oft wiederholt, bis zum ersten Mal der Erfolg E eintritt. Wir interessieren uns also für die Versuchsausgänge aus*
$$\Omega \;=\; \Big\{E, \overline{E}\,E, \overline{E}\,\overline{E}\,E, \overline{E}\,\overline{E}\,\overline{E}\,E, \ldots\Big\}$$
Da wir uns hier nur für die Anzahl der Misserfolg vor dem ersten Erfolg interessieren, betrachten wir die Zufallsvariable
$$X(\omega) \;=\; \text{"Anzahl der Misserfolg in } \omega\text{"}\ .$$

X kann also die Werte $k = 0, 1, 2, \ldots$ annehmen.

Legt man fest, dass die Durchführung eines Bernoulli–Experiments eine Zeiteinheit beträgt, kann man die obige ZV auch als **Wartezeit** *bis zum ersten Eintreten eines Erfolges bezeichnen.*

Man sieht leicht

$$f_X(k) \;=\; P(X = k) \;=\; P\Big(\underbrace{\overline{E}\cdots\overline{E}}_{k\,-\,mal}\,E\Big) \;=\; (1-p)^k p \qquad \forall\, k = 0, 1, 2, \ldots$$

<u>Bemerkung</u>

- Eine diskrete Zufallsvariable X mit der Dichte
$$f_X(k) \;=\; (1-p)^k p \qquad \forall k = 0, 1, 2, \ldots \qquad (0 < p < 1)$$
heißt **geometrisch verteilt** mit dem Parameter p und schreibt
$$X \;\sim\; \mathrm{Geo}(p)\ .$$
- Die bezeichnung ist in der literatur nicht einheitlich. Oft wird unter einer geometrisch verteilten Zufallsvariablen X^* die Zufallsvariable verstanden, die die Anzahl aller Versuche (also nicht nur der fehlversuche) bis zum ersten Erfolg zählt. Wegen
$$X^* \;=\; X \;+\; 1$$
lassen sich die nachfolgenden Ergebnisse auch leicht auf die Verteilung von X^* übertragen.
$$f_{X^*}(k) \;=\; P(X^* = k) \;=\; (1-p)^{k-1} p \qquad \forall\, k = 1, 2, \ldots$$

Geometrisch verteilte Zufallsvariable besitzen dabei die folgende interessante Eigenschaft:

Satz 5.1 *Sei X eine geometrisch verteilte Zufallsvariable mit dem Parameter p. Für jedes $l \in I\!N$ ist dann die bedingte Wahrscheinlichkeit*

$$P(X = k + l \mid X \geq k) \;=\; P(X = l)\ ,$$

also unabhängig von k. Man sagt auch, dass die ZV X kein Gedächtnis besitzt.

<u>Beweis</u>:

$$\begin{aligned}
P(X = k + l \mid X \geq k) \;&=\; \frac{P\big(X = k + l\ ,\ X \geq k\big)}{P(X \geq k)} \;=\; \frac{P(X = k + l)}{P(X \geq k)} \\
&=\; \frac{(1-p)^{k+l} \cdot p}{p \cdot \Big((1-p)^k + (1-p)^{k+1} + \cdots\Big)} \\
&=\; \frac{(1-p)^{k+l}}{(1-p)^k\,(1 + (1-p) + (1-p)^2 + \cdots)} \\
&=\; \frac{(1-p)^l}{\frac{1}{1-(1-p)}} \;=\; (1-p)^l \cdot p \;=\; P(X = l)
\end{aligned}$$

Hinweis:

Für die **geometrische Reihe** gilt:

$$\sum_{i=0}^{\infty} q^i \;=\; \frac{1}{1-q} \qquad (|q| \;<\; 1)$$

Dieser Satz besitzt die folgende Interpretation:

Die Tatsache, dass in einer Folge von Bernoulliexperimenten bis zu einem bestimmten Zeitpunkt (k) kein Erfolg eingetreten ist ($X \geq k$), verändert die Wartezeit zum nächsten Erfolg nicht.

Die meisten Menschen erwarten dagegen, dass nach einer Reihe von Misserfolgen die Wahrscheinlichkeit für einen Erfolg beim nächsten Mal größer ist als nach einer Folge von Erfolgen. Das ist bei Bernoulli–Experimenten offensichtlich falsch!

5.2 Erwartungswerte

Bemerkungen:

1. Bernoulli–Experiment mit dem Parameter p:
   Wieviele Erfolge erwartet man, wenn das Experiment n–mal unabhängig durchgeführt wird?

   $$\begin{array}{llllll} p \;=\; \frac{1}{2} & \Longrightarrow & \cdots \Longrightarrow & \frac{n}{2} - \text{mal} & ! & \text{(Münzwurf)} \\ p \;=\; \frac{1}{6} & \Longrightarrow & \cdots \Longrightarrow & \frac{n}{6} - \text{mal} & ! & \text{(Würfeln)} \end{array}$$

   Also bei n Versuchen erwarten wir allgemein, dass etwa $n \cdot p$ Erfolge eintreten.

2. Geometrische Verteilung:
   Ein Spiel bestehe aus einer Folge von unabhängigen Bernoulli–Experimenten mit dem Parameter p. Sobald das erste Mal ein Erfolg eintritt, werden G DM ausgezahlt. Für jeden Fehlwurf muss 1 DM bezahlt werden. Wie groß muss G sein, damit das Spiel ”fair” ist bzw. sich für den Anbieter ”lohnt”?
   oder
   Wie viele Fehlversuche erwarte ich bis zum Eintreten des ersten Erfolges?
   In Gedanken führen wir das Spiel N–mal durch und ermitteln die relative Häufigkeit $r_N(i)$ $(i = 0, 1, \ldots)$ der Spiele mit i Fehlversuchen vor dem ersten Erfolg. Da die Anzahl der Spiele mit genau i Fehlversuchen $N \cdot r_N(i)$ beträgt, müssen also insgesamt

   $$\sum_{i=0}^{\infty} N \cdot r_N(i) \cdot i \quad \text{DM}$$

bezahlt werden. Im Mittel müssen also pro Spiel DM

$$\sum_{i=0}^{\infty} r_N(i) \cdot i$$

bezahlt werden. Nun ersetzen wir idealisiert die relativen Häufigkeiten durch die entsprechenden Wahrscheinlichkeiten für die Wartezeit i. So erhalten wir für die erwarteten Zahlungen für ein Spiel:

$$\sum_{i=0}^{\infty} i \cdot P(X = i) \ .$$

Dabei ist die Zufallsvariable X (Anzahl der Misserfolg) geometrisch verteilt, so dass wir für diese "erwartete Anzahl" weiter

$$\begin{aligned} \sum_{i=0}^{\infty} i \cdot (1-p)^i \cdot p &= \cdots = p \cdot \frac{1-p}{p^2} \quad \text{(vergl. Beispiel 5.4)} \\ &= \frac{1-p}{p} = \frac{1}{p} - 1 \end{aligned}$$

erhalten. G muss also gleich diesem Betrag gewählt werden, damit das Spiel "fair" ist. Für ein faires Spiel gilt also

$$\begin{aligned} p &= 1 &\Longrightarrow\quad G &= 0 \\ p &= \frac{1}{2} &\Longrightarrow\quad G &= 1 \quad \text{und} \\ p &= \frac{1}{6} &\Longrightarrow\quad G &= 5 \ . \end{aligned}$$

Definition 5.3 (Erwartungswert) *Ist X diskrete Zufallsvariable auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{P}(\Omega), P\big)$ und gilt*

$$\sum_{x \in \Omega_X} |x| \cdot P(X = x) \ < \ \infty \ ,$$

so heißt

$$E\{X\} \ := \ \sum_{x \in \Omega_X} x \cdot P(X = x)$$

der **Erwartungswert** *der Zufallsvariablen X.*

Bemerkung:

Die obige Forderung nach der absoluten Konvergenz der Reihe

$$\sum_{x \in \Omega_X} x \cdot P(X = x)$$

ist notwendig, um zu sichern, dass für unendliche Stichprobenräume Ω_X ein Umordnen der Summanden (andere Nummerierung der Versuchsausgänge) nicht zu einem anderen Erwartungswert führt.

Beispiel 5.4 (Ermittlung verschiedener Erwartungswerte)

1. $X \sim Bi(n,p)$

$$E\{X\} = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i (1-p)^{n-i} = \cdots = np \qquad \textit{(vergl. Beispiel 5.13)}$$

2. $X \sim Geo(p) \quad (0 < p < 1)$

$$\begin{aligned} E\{X\} &= \sum_{i=1}^{\infty} i \cdot p(1-p)^i = p(1-p)\sum_{i=1}^{\infty} i(1-p)^{i-1} = p(1-p)\sum_{i=1}^{\infty} \frac{d}{dx} x^i \Big|_{x=1-p} \\ &= p(1-p)\frac{d}{dx}\sum_{i=0}^{\infty} x^i \Big|_{x=1-p} =^1 p(1-p)\,\frac{d}{dx}\frac{1}{1-x}\Big|_{x=1-p} \\ &= p(1-p)\,\frac{1}{(1-x)^2}\Big|_{x=1-p} = \frac{p(1-p)}{p^2} = \frac{1-p}{p} = \frac{1}{p} - 1 \end{aligned}$$

Satz 5.2 (Existenz des Erwartungswertes) *Falls $E\{X\}$ existiert, gilt*

$$E\{X\} = \sum_{\omega \in \Omega} P(\{\omega\}) \cdot X(\omega) \ ,$$

und gilt umgekehrt

$$\sum_{\omega \in \Omega} P(\{\omega\}) \cdot \big|X(\omega)\big| < \infty \ ,$$

so existiert $E\{X\}$, und es gilt die obige Formel.

Bemerkung:

Wegen

$$x \cdot P(X = x) = x \cdot \sum_{\omega \,:\, X(\omega)=x} X(\omega)$$

tauchen in der obigen Formel dieselben Summanden wie in der Definition auf. Lediglich die Reihenfolge ist verändert, was im Falle der Existenz (absolute Konvergenz) keine Rolle spielt.

[1]Potenzreihe mit einem Konvergenzradius $r = 1 > 1 - p$

Satz 5.3 (Linearität des Erwartungswertes)

(a) Ist $c \in I\!R$ und X die konstante Abbildung nach c, d.h. es gilt $X(\omega) = c \quad \forall\, \omega \in \Omega$, also $P(X = c) = 1$ und $P(X \neq c) = 0$, dann existiert der Erwartungswert von X und es gilt

$$E\{X\} = c\ .$$

(b) Seien X und Y zwei diskrete Zufallsvariable auf dem gleichen Wahrscheinlichkeitsraum, deren Erwartungswerte existieren, und a und b beliebige reelle Zahlen, dann existiert der folgende Erwartungswert, und es gilt

$$E\{a \cdot X + b \cdot Y\} = a \cdot E\{X\} + b \cdot E\{Y\}\ .$$

<u>Beweis</u>:

(a)

$$\sum_{x \in \Omega_X} |x| \cdot P(X = x) = \sum_{x \neq c} |x| \cdot \underbrace{P(X = x)}_{=\,0} + |c| \cdot \underbrace{P(X = c)}_{=\,1} = |c| < \infty$$

$$\mathrm{E}\{X\} := \mathrm{E}\{c\} = \sum_{x \neq c} x \cdot P(X = x) + c \cdot P(X = c) = 0 + c \cdot 1 = c$$

(b) Die Existenz des Erwartungswertes ist nach Satz 5.2 wegen der Existenz der Einzelerwartungswerte und

$$\sum_{\omega \in \Omega} \underbrace{|aX(\omega) + bY(\omega)|}_{\leq\ |a||X|+|b||Y|} \cdot P(\{\omega\}) \leq |a| \underbrace{\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\})}_{<\,\infty} + |b| \underbrace{\sum_{\omega \in \Omega} |Y(\omega)| \cdot P(\{\omega\})}_{<\,\infty}$$

gesichert.

So erhalten wir also

$$\begin{aligned}
\mathrm{E}\{aX + bY\} &= \sum_{\omega \in \Omega} \big(a\, X(\omega) + b\, Y(\omega)\big) \cdot P(\{\omega\}) \\
&= a \sum_{\omega \in \Omega_X} X(\omega) \cdot P(\{\omega\}) + b \sum_{\omega \in \Omega_Y} Y(\omega) \cdot P(\{\omega\}) \\
&= a \cdot \mathrm{E}\{X\} + b \cdot \mathrm{E}\{Y\}
\end{aligned}$$

$\blacksquare$

<u>Bemerkung</u>:

Durch Induktion zeigt man sofort, dass für beliebige Zufallsvariable $X_1, \ldots, X_n$ und beliebige $a_1, \ldots, a_n \in I\!R$

$$\mathrm{E}\left\{\sum_{i=1}^{n} a_i X_i\right\} = \sum_{i=1}^{n} a_i\, \mathrm{E}\{X_i\} \quad \text{und ebenso} \quad \mathrm{E}\{aX + b\} = a \cdot \mathrm{E}\{X\} + b$$

gilt, falls alle Einzelerwartungswerte $\mathrm{E}\{X_i\}$ existieren.

Beispiel 5.5 (Sortieren von Listen) *Wir betrachten zufällig angeordnete Listen aus n Elementen, für die eine eindeutige "richtige" Anordnung existiert. Als Maß für die* **Ordnung einer Liste** *zählen wir die Elemente der Liste, die bereits auf ihrem "richtigen" Platz stehen.*

$$\Omega \; := \; \{\omega \, : \, \omega \text{ Permutation der Zahlen } 1, \ldots, n\} \; = \; \mathcal{P}^n \quad \text{mit} \quad |\Omega| = n!$$

Unter einer "zufällig" angeordneten Liste wollen wir ein Laplace–Modell verstehen, d.h. alle möglichen Listen sind gleichwahrscheinlich, also

$$P(\omega) \; = \; \frac{1}{n!} \qquad \forall \, \omega \in \Omega \; .$$

X sei die Zufallsvariable, die zählt, wie viele Elemente der Liste auf ihrem richtigen Platz stehen.

Es ist nun sehr schwierig, die Dichte von X, d.h. die Wahrscheinlichkeiten $P(X = j)$ für $j = 0, \ldots, n$ explizit auszurechnen. Dagegen können wir relativ einfach den Erwartungswert von X mit der obigen Formel ermitteln. Es gilt nämlich

$$\begin{aligned}
X \; &= \; X_1 \; + \; \cdots \; + \; X_n \\
\text{mit} \qquad & X_i(\omega) \; := \; \begin{cases} 1 & \text{falls } i\text{-tes Element von } \omega \text{ auf richtigen Platz} \\ 0 & \text{sonst} \end{cases} \\
\text{und} \qquad & E\{X_i\} \; = \; 1 \cdot P(X_i \; = \; 1) \; = \; \frac{(n-1)!}{n!} \; = \; \frac{1}{n} \qquad \forall \; i = 1, \ldots, n \\
E\{X\} \; &= \; E\{X_1\} \; + \cdots \; + \; E\{X_n\} \; = \; n \cdot \frac{1}{n} \; = \; 1
\end{aligned}$$

Beispiel 5.6 (Hypergeometrische Verteilung) *X sei hypergeometrisch verteilt (vergleiche Beispiel 5.2):*

$$X \; \sim \; Hyp(n; N, R) \quad \Longleftrightarrow \quad P(X = r) \; = \; \frac{\binom{R}{r} \cdot \binom{N-R}{n-r}}{\binom{N}{n}} \; .$$

Auch hier ist es nicht einfach, den Erwartungswert (erwartete Anzahl der roten Kugeln bei n-maligem Ziehen ohne Zurücklegen) auf direkte Weise zu ermitteln.

$$X_i \; := \; \begin{cases} 1 & \text{falls beim } i\text{-ten Zug eine rote Kugel gezogen wird} \\ 0 & \text{sonst} \end{cases}$$

Damit erhalten wir $\quad X \; = \; X_1 \; + \; \cdots \; + \; X_n \quad$ *und*

$$\begin{aligned}
E\{X\} \; &= \; E\{X_1\} \; + \; \cdots \; + \; E\{X_n\} \\
\text{mit} \qquad & E\{X_i\} \; = \; P(X_i = 1) \; = \; P(X_1 = 1) \; = \; \frac{R}{N} \quad \forall \; i = 1, \ldots, n \; ,
\end{aligned}$$

(Wenn nur ein einzelner Zug unabhängig von den anderen betrachtet wird, ist es gleichgültig, welcher Zug es ist, da die Reihenfolge der gefundenen roten und schwarzen K. gleichgültig ist.

$$E\{X\} \; = \; \frac{nR}{N}$$

Wer misstrauisch ist, führe die folgende Rechnung aus:

$$E\{X\} \;=\; \sum_{r=1}^{n} r \cdot \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}} \;=\; \cdots \;=\; \frac{nR}{N}\;.$$

Definition 5.4 (Indikatorvariable) *Sei* $\big(\Omega, \mathcal{P}(\Omega), P\big)$ *ein diskreter WR und* $A \subseteq \Omega$ *ein beliebiges Ereignis, dann heißt die Zufallsvariable* I_A *mit*

$$I_A(\omega) \;:=\; \begin{cases} 1 & \textit{falls } \omega \in A \\ 0 & \textit{falls } \omega \notin A \end{cases} \qquad \textbf{Indikatorvariable} \textit{ über } A\;.$$

Bemerkung:

$$\mathrm{E}\{I_A\} \;=\; 1 \cdot P(I_A = 1) \;+\; 0 \cdot P(I_A = 0) \;=\; P(\omega \in A) \;=\; P(A)$$

Beispiel 5.7

$$X_i \;=\; \begin{cases} 1 & \textit{falls ein Erfolg im } i\textit{-ten Versuch eintritt,} \\ 0 & \textit{falls kein Erfolg im } i\textit{-ten Versuch eintritt,} \end{cases} \qquad i = 1, \ldots, n$$

$$X \;\sim\; Bi(n,p) \quad , \quad X \;=\; \sum_{i=1}^{n} X_i \qquad \Longrightarrow \qquad E\{X\} \;=\; \sum_{i=1}^{n} E\{X_i\} \;=\; np$$

Definition 5.5 (Bedingter Erwartungswert) *Sei* X *eine Zufallsvariable mit existierendem Erwartungswert auf dem diskreten WR* $\big(\Omega, \mathcal{P}(\Omega), P\big)$ *und* $A \subseteq \Omega$ *ein beliebiges Ereignis mit* $P(A) > 0$ *, dann heißt*

$$E\{X \mid A\} \;:=\; \sum_{\omega\in\Omega} X(\omega) \cdot P\big(\{\omega\} \mid A\big) \;=\; \sum_{x\in\Omega_X} x \cdot P(X = x|A)$$

der **bedingte Erwartungswert** *von* X *bei gegebenem* A *(bzw. unter der Bedingung* A*). Falls* $P(A) = 0$ *gilt, kann man* $E\{X|A\}$ *beliebig wählen.*

Bemerkung:

$$P(A) > 0 \text{ und } \omega \notin A \qquad \Longrightarrow \qquad P(\{\omega\}|A) \;=\; 0$$

Damit erhalten wir

$$\begin{aligned} \mathrm{E}\{X|A\} \;&=\; \sum_{\omega\in\Omega} X(\omega) \cdot P(\{\omega\}|A) \;=\; \sum_{\omega\in A} X(\omega) \cdot P(\{\omega\}|A) \\ &=\; \sum_{\omega\in A} X(\omega) \cdot \frac{P(\{\omega\} \cap A)}{P(A)} \;=\; \sum_{\omega\in A} X(\omega) \cdot \frac{P(\{\omega\})}{P(A)} \end{aligned}$$

Satz 5.4 (Unbedingter Erwartungswert) *Sei X eine Zufallsvariable mit existierendem Erwartungswert auf dem diskreten WR $(\Omega, \mathcal{P}(\Omega), P)$ und sei $A_1, A_2, \ldots \subseteq \Omega$ eine Folge paarweise disjunkter Ereignisse mit $\sum_i A_i = \Omega$, dann gilt analog der Formel für die totale Wahrscheinlichkeit:*

$$E\{X\} = \sum_i E\{X \mid A_i\} \cdot P(A_i) \ .$$

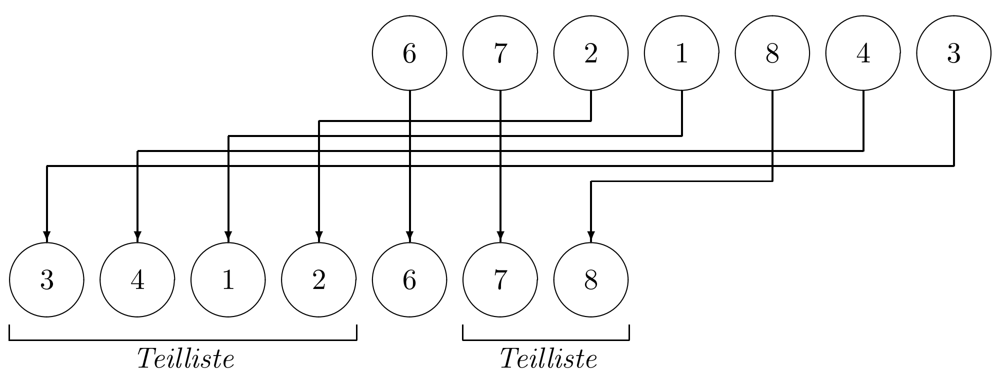
<u>Beweis</u>:

$$\begin{aligned} \mathrm{E}\{X\} &= \sum_{\omega\in\Omega} X(\omega) \cdot P(\{\omega\}) = \sum_i \sum_{\omega\in A_i} X(\omega) \cdot P(\{\omega\}) \\ &= \sum_i \sum_{\omega\in A_i} X(\omega) \cdot \frac{P(\{\omega\})}{P(A_i)} \cdot P(A_i) \\ &= \sum_i \mathrm{E}\{X \mid A_i\} \cdot P(A_i) \end{aligned}$$

Beispiel 5.8 (Quicksort) *Ein schneller Algorithmus zum Sortieren einer Liste von n Zahlen.*

<u>Gegeben</u>: Liste aus n unterschiedlichen natürlichen Zahlen

<u>Vorgehen</u>:
Im ersten Schritt wird das erste Objekt der Liste mit allen anderen verglichen ($n-1$ Vergleiche). Alle kleineren Objekte werden davor, die übrigen dahinter einsortiert, wobei die interne Rangfolge dieser Objekte untereinander teilweise umgekehrt wird.



Jetzt werden Teillisten der Elemente gebildet, die nun vor bzw. hinter dem ersten Element stehen, und der obige Algorithmus auf beide Teillisten angewendet. Das Verfahren wird rekursiv fortgesetzt, bis jede Teilliste eine Länge von einem Element erreicht hat.

Die benötigte Rechenzeit für diesen Sortieralgorithmus hängt im wesentlichen von der Anzahl der notwendigen Vergleiche ab. Wir wollen daher wissen, wie viele Vergleiche "im Mittel" benötigt werden, um eine zufällig angeordnete Liste der Länge n zu ordnen.

Dazu konstruieren wir zunächst einen geeigneten WR (vergl. Beispiel 5.5).

$$\Omega \;:=\; \{\omega \,:\, \omega \in \mathcal{P}^n\} \qquad \textit{mit} \qquad |\Omega| \;=\; n!$$

Betrachten wir nun alle möglichen Listen als gleichwahrscheinlich (Laplaceraum), erhalten wir

$$P(\{\omega\}) \;=\; \frac{1}{n!} \qquad \forall\, \omega \in \Omega\ .$$

Die Zufallsvariable X_n zähle nun die Anzahl der Vergleiche, die dieser Algorithmus durchführen muss, um eine Liste ω der Länge n zu sortieren.

Wir interessieren uns also für

$$E\{X_n\} \;=\; \frac{1}{n!}\sum_{\omega\in\Omega} X_n(\omega) \;=\; ?$$

Dazu betrachten wir die folgenden Ereignisse

$$A_x \;:=\; \{\omega \,:\, \textit{Beim Anfangsschritt rückt das 1. Element auf Platz } x\} \qquad \forall\ 1 \le x \le n\ .$$

Dabei gilt natürlich

$$P(A_x) \;=\; \frac{(n-1)!}{n!} \;=\; \frac{1}{n} \qquad .$$

Anschließend sind dann zwei Listen der Länge $(x-1)$ bzw. $(n-x)$ zu sortieren.

Wir lösen das Problem nun rekursiv. Zunächst gilt für die erwartete (mittlere) Anzahl $M(n)$ der notwendigen Vergleiche für eine Liste der Länge n:

$$\begin{aligned}
M(n) \;:=\; E\{X_n\} &= \sum_{x=1}^{n} E\{X_n \mid A_x\}\cdot P(A_x)\\
&= \frac{1}{n}\sum_{x=1}^{n} E\{X_n \mid A_x\} \;=\; \frac{1}{n}\sum_{x=1}^{n}\Big[(n-1) \,+\, M(x-1) \,+\, M(n-x)\Big]\\
&= (n-1) \,+\, \frac{1}{n}\Big[\sum_{x=0}^{n-1} M(x) \,+\, \sum_{x=1}^{n} M(n-x)\Big] \;=\; (n-1) \,+\, \frac{2}{n}\sum_{x=0}^{n-1} M(x) \qquad (5.1)
\end{aligned}$$

Dabei gilt $\quad M(0) \;=\; M(1) \;=\; 0 \quad$ *und weiter*

$$\begin{aligned}
M(2) &= 1 \,+\, \frac{2}{2}\Big[M(0) \,+\, M(1)\Big] \;=\; 1 \quad \textit{und}\\
M(3) &= 2 \,+\, \frac{2}{3}\Big[M(0) \,+\, M(1) \,+\, M(2)\Big] \;=\; \frac{8}{3}\ , \quad \textit{u.s.w.}
\end{aligned}$$

Allgemein erhalten wir aus Gleichung 5.1

$$nM(n) \;=\; n(n-1) \,+\, 2\sum_{x=0}^{n-1} M(x)\ .$$

Ersetzt man hier n *durch* $n-1$*, erhält man*

$$(n-1)M(n-1) \;=\; (n-1)(n-2) \;+\; 2\sum_{x=0}^{n-2} M(x)$$

Subtrahiert man die zweite von der ersten Gleichung, so ergibt sich

$$\begin{array}{rcl} nM(n) \;-\; (n-1)M(n-1) & = & 2(n-1) \;+\; 2M(n-1)\ , \\ \Longrightarrow \qquad\qquad nM(n) & = & 2(n-1) \;+\; (n+1)M(n-1)\ . \end{array}$$

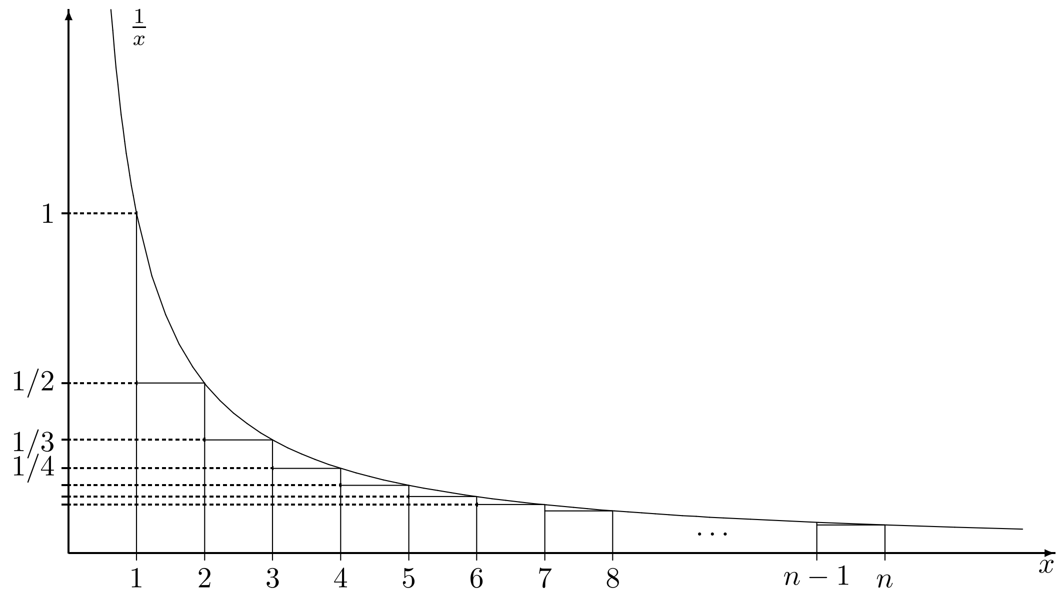
Division dieser Gleichung durch $n(n+1)$ *ergibt*

$$\begin{array}{lrcl} & \dfrac{M(n)}{n+1} & = & \dfrac{2(n-1)}{n(n+1)} \;+\; \dfrac{M(n-1)}{n} \\ \textit{bzw.} & \dfrac{M(n)}{n+1} \;-\; \dfrac{M(n-1)}{n} & = & 2\left[\dfrac{2}{n+1} \;-\; \dfrac{1}{n}\right]\ . \end{array}$$

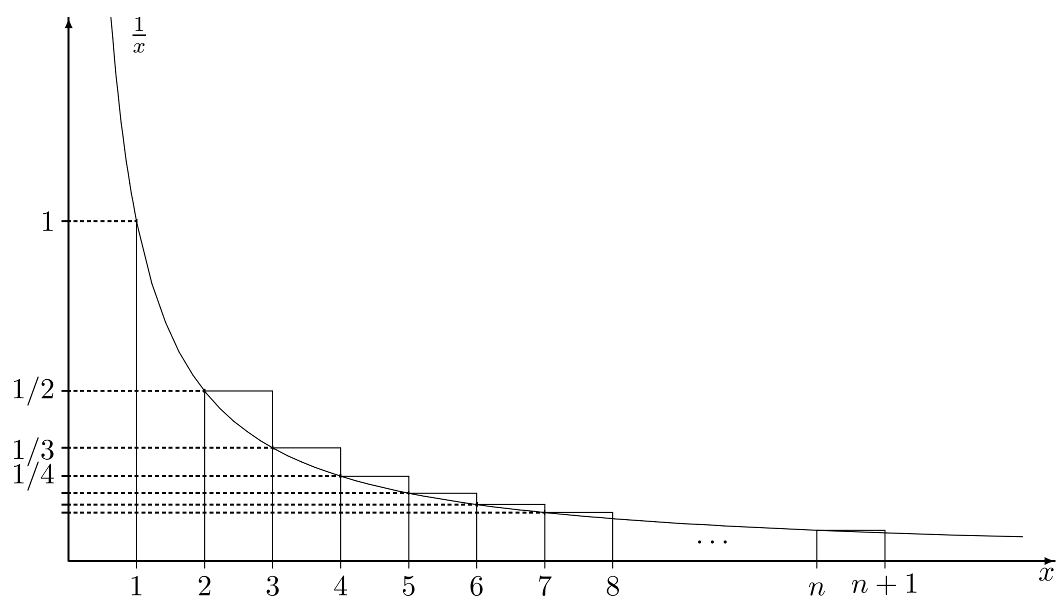
Ebenso gilt aber

$$\begin{aligned} \frac{M(n)}{n+1} \;&=\; \left(\frac{M(n)}{n+1} \;-\; \frac{M(n-1)}{n}\right) \;+\; \left(\frac{M(n-1)}{n} \;-\; \frac{M(n-2)}{n-1}\right) \\ &\quad + \cdots + \underbrace{\left(\frac{M(1)}{2} \;-\; \frac{M(0)}{1}\right)}_{=0} \;+\; \underbrace{\frac{M(0)}{1}}_{=\,0} \\ &=\; \sum_{i=2}^{n}\left(\frac{M(i)}{i+1} \;-\; \frac{M(i-1)}{i}\right) \;=\; 2\sum_{i=2}^{n}\left(\frac{2}{i+1} \;-\; \frac{1}{i}\right) \\ &=\; \sum_{i=2}^{n}\frac{4}{i+1} \;-\; \sum_{i=1}^{n-1}\frac{2}{i+1} \;=\; \frac{4}{n+1} \;+2\sum_{i=1}^{n-1}\frac{1}{i+1} \;-\; 2\ . \end{aligned} \tag{5.2}$$

Die folgenden Bilder machen die Gültigkeit zweier benötigter Ungleichungen deutlich:



$$\sum_{i=1}^{n-1} \frac{1}{i+1} \;=\; \frac{1}{2} \;+\; \frac{1}{3} \;+\; \cdots \;+\; \frac{1}{n} \;\leq\; \int_1^n \frac{1}{x} dx \;=\; \log n$$



$$\log(n+1) \;-\; \log 2 \;=\; \int_2^{n+1} \frac{1}{x} dx \;\leq \sum_{i=1}^{n-1} \frac{1}{i+1}$$

Aus Gleichung 5.2 erhalten wir schließlich

$$\frac{4}{n+1} - 2 + 2\log(n+1) - 2\log 2 \leq \frac{M(n)}{n+1} \leq \frac{4}{n+1} - 2 + 2\log n$$

und damit wegen $\lim\limits_{n\to\infty} \dfrac{\log\{n+1\}}{\log n} = 1$ *asymptotisch*

$$\lim_{n\to\infty} \frac{M(n)}{n\log n} = 2 \quad , \quad \textit{also} \quad M(n) = O\big(n\log n\big) \ .$$

Tabelle 5.1: Mittlere Anzahl der Vergleiche in QUICKSORT

n	$M(n)$	Anz. paarw. Vergleiche
2	1.00	1
3	2.67	3
4	4.83	6
5	7.40	10
10	24.44	45
100	656.24	4.950
10000	155789.27	499.950.000

In Tabelle 5.1 werden die obigen mittleren Anzahlen der notwendigen Vergleiche, mit der Anzahl $\binom{n}{2}$ der notwendigen paarweisen Vergleiche beim klassischen Vorgehen für ausgewählte n verglichen.

Häufig will man nicht den Erwartungswert einer Zufallsvariablen selbst, sondern den Erwartungswert einer Funktion

$$f \ : \ \Omega_X \ \longrightarrow \ I\!R$$

dieser Zufallsvariablen ermitteln.

Satz 5.5 *Ist X eine diskrete Zufallsvariable auf $(\Omega, \mathcal{P}(\Omega), P)$ und $f : \Omega_X \longrightarrow I\!R$, dann gilt*

$$E\{f(X)\} = \sum_{x\in\Omega_X} f(x)P(X=x) \ ,$$

falls die obige Reihe absolut konvergent ist.

Beweis:

$Y := f(X)$ ist ebenfalls eine Zufallsvariable auf dem obigen WR. daher gilt

$$\begin{aligned}
\mathrm{E}\{Y\} &= \sum_{y\in\Omega_Y} y\cdot P(Y=y) \;=\; \sum_{y\in\Omega_Y} y\cdot P(f(X)=y) \\
&= \sum_{y\in\Omega_Y} y \sum_{x\in\Omega_X:\, f(x)=y} P(X=x) \;=\; \sum_{y\in\Omega_Y}\;\sum_{x\in\Omega_X:\, f(x)=y} f(x)\cdot P(X=x) \\
&= \sum_{x\in\Omega_X} f(x)\cdot P(X=x)
\end{aligned}$$

$\blacksquare$

Beispiel 5.9

$$E\{|X|\} \;=\; \sum_{x\in\Omega_X} |x|\cdot P(X=x)$$

5.3 Höhere Momente und erzeugende Funktionen

Definition 5.6 *Sei X eine diskrete Zufallsvariable auf $(\Omega, \mathcal{P}(\Omega), P)$, dann heißen im Falle der Existenz*

$$\begin{array}{rcll}
M_k & := & E\left\{X^k\right\} \quad (k\in I\!N) & k\text{-tes } \textbf{Moment} \text{ von } X, \\
\mu_k & := & E\left\{(X - E\{X\})^k\right\} \quad (k\in I\!N) & k\text{-tes } \textbf{zentriertes Moment} \text{ von } X \text{ und} \\
M_{(k)} & := & E\{X(X-1)\cdots(X-k+1)\} & \textbf{faktorielles Moment} \text{ der Ordnung } k.
\end{array}$$

Speziell heißen:

$$\begin{array}{lll}
M_1 \;=\; E\{X\} \;=\; M_{(1)} & & \text{Erwartungswert von } X, \\[2ex]
\mu_2 \;:=\; Var\{X\} & & \textbf{Varianz} \text{ von } X, \\[2ex]
S\{X\} \;:=\; \sqrt{Var\{X\}} & & \textbf{Streuung} \text{ von } X, \\[2ex]
V\{X\} \;:=\; \dfrac{S\{X\}}{E\{X\}} & & \textbf{Variationskoeffizient}, \\[2ex]
\gamma_1 \;:=\; \dfrac{\mu_3}{(Var\{X\})^{3/2}} & & \textbf{Charliersche Schiefe}, \\[2ex]
g_2 \;:=\; \dfrac{\mu_4}{(Var\{X\})^2} & & \textbf{Kurtosis} \text{ und} \\[2ex]
\gamma_2 \;:=\; \dfrac{\mu_4}{(Var\{X\})^2} - 3 & & \textbf{Exzeß}.
\end{array}$$

Bemerkungen:

1. $$M_{(k)} \;=\; \sum_{x \in \Omega_X} x(x-1)\cdots(x-k+1)\cdot P(X=x)$$

2. Der Erwartungswert einer Zufallsvariablen beschreibt ihre Lage (Lageparameter). Die Varianz beschreibt, wie weit ihre Realisierungen "im quadratischen Mittel" von diesem Lageparameter abweichen. Es wäre hier auch sinnvoll
$$\mathrm{E}\{|X \;-\; \mathrm{E}\{X\}\,|\}$$
zu betrachten. Diese Größe wäre aber für praktische Berechnungen schwieriger zu handhaben. Dagegen hätte die Größe
$$\mu_1 \;=\; \mathrm{E}\{X \;-\mathrm{E}\{X\}\}$$
keinen Sinn, da sie für alle Zufallsvariablen mit existierendem Erwartungswert stets den Wert 0 hat.

3. Der Variationskoeffizient berücksichtigt, dass die absoluten Werte der Streuung einer Zufallsvariablen in Abhängigkeit von der Größe des Erwartungswertes zu bewerten sind.

4. Die Charliersche Schiefe ist 0, falls die Zufallsvariable symmetrisch zum Erwartungswert verteilt ist. Sie stellt also ein Maß für die Asymmetrie (Schiefe) der Verteilung dar.

5. Die Bedeutung der Kurtosis wird erst für stetige Zufallsvariable klar.

Satz 5.6 (Rechenregeln für Momente) *Im Falle der Existenz gilt:*

(5.3) $$M_0 \;=\; E\{1\} \;=\; 1$$

(5.4) $$E\{a\cdot g(X) \;+\; b\cdot h(X)\} \;=\; a\cdot E\{g(X)\} \;+\; b\cdot E\{h(X)\} \qquad \textbf{Linearität}$$

(5.5) $$Var\{X\} \;=\; E\left\{X^2\right\} \;-\; \underbrace{(E\{X\})^2}_{=:\; E^2\{X\}}$$

(5.6) $$Var\{a\cdot X \;+\; b\} \;=\; a^2\cdot Var\{X\}$$

Beweis:

5.3 $X \equiv 1\,,$ d.h. $P(X=1)=1$ und $P(X\neq 1)=0 \quad\Longrightarrow\quad \mathrm{E}\{X\} = 1\cdot P(X=1)+0 \;=\; 1$

5.4
$$\begin{aligned}\mathrm{E}\{a\cdot g(X)+b\cdot h(X)\} &= \sum_{x\in\Omega_X}(a\cdot g(x)+b\cdot h(x))\cdot P(X=x)\\ &= a\cdot\sum_{x\in\Omega_X} g(x)\cdot P(X=x) \;+\; b\cdot\sum_{x\in\Omega_X} h(x)\cdot P(X=x)\\ &= a\cdot\mathrm{E}\{g(X)\} \;+\; b\cdot\mathrm{E}\{h(X)\}\end{aligned}$$

5.5

$$
\begin{aligned}
\operatorname{Var}\{X\} &= \mathrm{E}\left\{(X-\mathrm{E}\{X\})^2\right\} = \mathrm{E}\left\{X^2 - 2X\cdot\mathrm{E}\{X\} + \mathrm{E}^2\{X\}\right\} \\
&= \mathrm{E}\left\{X^2\right\} - 2\mathrm{E}\{X\}\cdot\mathrm{E}\{X\} + \mathrm{E}^2\{X\} = \mathrm{E}\left\{X^2\right\} - \mathrm{E}^2\{X\}
\end{aligned}
$$

5.6

$$
\begin{aligned}
\operatorname{Var}\{aX+b\} &= \mathrm{E}\left\{(aX+b)^2\right\} - \mathrm{E}^2\{aX+b\} \\
&= \mathrm{E}\left\{a^2X^2+2abX+b^2\right\} - (a\,\mathrm{E}\{X\}+b)^2 \\
&= a^2\,\mathrm{E}\left\{X^2\right\} + 2ab\,\mathrm{E}\{X\} + b^2 - a^2\,\mathrm{E}^2\{X\} - 2ab\,\mathrm{E}\{X\} - b^2 \\
&= a^2\cdot\left(\mathrm{E}\left\{X^2\right\} - \mathrm{E}^2\{X\}\right) = a^2\cdot\operatorname{Var}\{X\}
\end{aligned}
$$

Beispiel 5.10 (Binomialverteilung) $X \sim Bi(n,p)$

$$
\begin{aligned}
Var\{X\} &= E\left\{X^2\right\} - E^2\{X\} = E\left\{X^2\right\} - (np)^2 \\
E\left\{X^2\right\} &= \sum_{k=0}^{n} k^2\binom{n}{k}p^k(1-p)^{n-k} \\
k^2\binom{n}{k} &= \frac{k^2\cdot n!}{k!\,(n-k)!} = k\cdot n\cdot\frac{(n-1)!}{(k-1)!\,(n-1-(k-1))!} = k\cdot n\cdot\binom{n-1}{k-1} \\
E\left\{X^2\right\} &= np\sum_{k=1}^{n} k\binom{n-1}{k-1}p^{k-1}(1-p)^{n-1-(k-1)} \\
&= np\sum_{k=0}^{n-1}(k+1)\binom{n-1}{k}p^k(1-p)^{n-1-k} \\
&\quad\ mit\ \ Y \sim Bi(n-1,p) \\
&= np\underbrace{\sum_{k=0}^{n-1}k\cdot\binom{n-1}{k}p^k(1-p)^{n-1-k}}_{=\ E\{Y\}\ =\ (n-1)\cdot p} + np\underbrace{\sum_{k=0}^{n-1}\binom{n-1}{k}p^k\ (1-p)^{n-1-k}}_{=\ \sum_{i=0}^{n-1}P(Y=k)\ =1} \\
&= n(n-1)p^2 + np = n^2p^2 - np^2 + np \\
Var\{X\} &= np - np^2 = np(1-p)
\end{aligned}
$$

Satz 5.7 *Im Falle der Existenz der entsprechenden Momente der ZV X gilt*

$$
\mu_k = \sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i}M_i\,M_1^{k-i}\ .
$$

Beweis:

$$\mu_k \;=\; \mathrm{E}\left\{(X \;-\; \mathrm{E}\{X\})^k\right\} \;=\; \mathrm{E}\left\{\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}X^i\big[\mathrm{E}\{X\}\big]^{k-i}\right\} \;=\; \sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i}M_iM_1^{k-i}$$

$\blacksquare$

Ein wichtiges Hilfsmittel zur Berechnung von Wahrscheinlichkeiten und Momenten ist die sog. **erzeugende Funktion**.

Definition 5.7 *X sei diskrete ZV mit dem Wertebereich $\Omega_X \subseteq \{0,1,2,\ldots\}$ und $P(X=i)=p_i$ $(i \in I\!N_0)$. Dann heißt die reelle Funktion g_X mit*

$$g_X(t) \;:=\; \sum_{i\in\Omega_X} t^i \cdot P(X=i) \;=\; \sum_{i\in\Omega_X} p_i \cdot t^i \;=\; E\left\{t^X\right\}$$

die **erzeugende Funktion** *der Zufallsvariablen X. Ihr Definitionsbereich ist der Definitionsbereich der obigen Potenzreihe.*

Bemerkung:

Wegen

$$\left|\sum_i p_i \cdot t^i\right| \;\leq\; \sum_i p_i \cdot |t|^i \;\overset{|t|\leq 1}{\leq}\; \sum_i p_i \;=\; 1 \;<\; \infty\;,$$

ist die Potenzreihe glm. konvergent für $|t| \leq 1$. Ihr Konvergenzradius ist also ≥ 1.

Aus der Analysis wissen wir, dass g_X damit für $|t| < 1$ stetig und beliebig oft differenzierbar, und an der Stelle $t = +1$ von links und an der Stelle $t = -1$ von rechts differenzierbar ist. Dabei darf gliedweise differenziert werden, da eine Potenzreihe im Inneren ihres Konvergenzgebietes glm. stetig ist, bzw. am Rande des Gebietes jeweils glm. halbseitig stetig ist.

Betrachten wir die folgenden Ableitungen

$$\begin{aligned}
\frac{d^k}{dt^k}\,g_X(t) \;&=\; \frac{d^k}{dt^k}\sum_{i=0}^{\infty} t^i\;P(X=i) \quad \text{mit} \quad \frac{d^k}{dt^k}t^i \;=\; \begin{cases} 0 & \text{falls } i<k\;, \\ i(i-1)\cdots(i-k+1)t^{i-k} & \text{falls } i \geq k \end{cases} \\
&=\; \sum_{i=k}^{\infty} i(i-1)\cdots(i-k+1)\cdot t^{i-k}P(X=i) \;=\; k!\cdot P(X=k) \;+\; t\cdot(\cdots)\;,
\end{aligned}$$

dann gilt für $t=0$

$$\frac{d^k}{dt^k}\,g_X(0) \;=\; k!\cdot P(X=k) \qquad \forall\; k=0,1,2,\ldots$$

Aus der erzeugenden Funktion g_X lassen sich also die Einzelwahrscheinlichkeiten durch Differentiation ermitteln.

$$P(X=k) \;=\; \frac{1}{k!}\,g_X^{(k)}(0) \qquad \forall\; k=0,1,2,\ldots$$

Damit ist der Name erzeugende (bzw. auch wahrscheinlichkeitserzeugende) Funktion gerechtfertigt. Aus der erzeugenden Funktion kann die Verteilung der Zufallsvariablen X gewonnen werden.

Beispiel 5.11 (geometrische Verteilung) $X \sim Geo(p)$ *mit* $0 < p < 1$.

$$g_X(t) = \sum_{k=0}^{\infty} t^k (1-p)^k p = p \cdot \frac{1}{1 - t(1-p)}$$

für alle

$$|t(1-p)| < 1 , \quad d.h. \quad |t| < \underbrace{\frac{1}{1-p}}_{\geq 1} .$$

Beispiel 5.12 (Binomialverteilung) $X \sim Bi(n,p)$.

Aus dem binomischen Lehrsatz erhalten wir

$$g_X(t) = \sum_{i=0}^{n} t^i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^{n} \binom{n}{i} (tp)^i (1-p)^{n-i} = (tp + (1-p))^n .$$

Satz 5.8

$$P(X = k) = \frac{1}{k!} g_X^{(k)}(0) \qquad \forall\ k = 0, 1, 2, \ldots \tag{5.7}$$

$$\lim_{t \uparrow 1} g_X^{(k)}(t) = g_X^{(k)}(1-) = M_{(k)} \tag{5.8}$$

Dabei sind beide Seiten der zweiten Gleichung genau dann endlich, falls M_k *endlich ist.*

Beweis:

$$g_X^{(k)}(t) = \sum_{i=k}^{\infty} i(i-1)\cdots(i-k+1) t^{i-k} P(X = i)$$

Wegen

$$|t| \leq 1 \qquad \Longrightarrow \qquad |i(i-1)\cdots(i-k+1)t^{i-k}| \leq i^k$$

und

$$\Longrightarrow \qquad g_X^{(k)}(t) \leq \sum_{i=0}^{\infty} i^k P(X = i) \leq \mathrm{E}\left\{X^k\right\} < \infty ,$$

ist die obige Reihe glm. abs. konvergent in $t \in [-1, 1]$.

Damit dürfen beide Grenzübergänge vertauscht werden.

$$\begin{aligned} g_X^{(k)}(1-) &= \sum_{i=k}^{\infty} i(i-1)\cdots(i-k+1) \left(\lim_{t \uparrow 1} t^{i-k} \right) P(X = i) \\ &= \sum_{i=k}^{\infty} i(i-1)\cdots(i-k+1) P(X = i) = M_{(k)} \end{aligned}$$

Bemerkung:

Aus den faktoriellen Momenten der Ordnungen 1 bis k lassen sich auch die zentrierten und nicht zentrierten Momenten dieser Ordnungen ermitteln.

$$\begin{aligned}\mathrm{Var}\{X\} &= \mathrm{E}\{X^2\} - \mathrm{E}^2\{X\} = \mathrm{E}\{X^2\} - \mathrm{E}\{X\} + \mathrm{E}\{X\} - \mathrm{E}^2\{X\} \\ &= \mathrm{E}\{X(X-1)\} + \mathrm{E}\{X\}(1-\mathrm{E}\{X\}) = M_{(2)} + M_{(1)}(1-M_{(1)})\end{aligned}$$

Beispiel 5.13 (Binomialverteilung) $X \sim Bi(n,p)$

$$X := \sum_{i=1}^{n} X_i \quad mit \quad X_1,\dots,X_n \sim Bi(1,p) \qquad \textit{(vergl. Beispiel 5.16)}$$

$$\begin{aligned}Var\{X\} &= Var\left\{\sum_{i=1}^{n} X_i\right\} = ? \\ g_X(t) &= \big(tp+(1-p)\big)^n \\ g'_X(t) &= np\big(tp+(1-p)\big)^{n-1} \\ E\{X\} &= M_{(1)} = g'_X(1-) = np \\ g''_X(t) &= n(n-1)p^2\big(tp+(1-p)\big)^{n-2} \\ M_{(2)} &= g''_X(1-) = n(n-1)p^2 \\ E\{X\} &= np \\ Var\{X\} &= M_{(2)} + M_{(1)}\big(1-M_{(1)}\big) = n(n-1)p^2 + np(1-np) \\ &= n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p)\end{aligned}$$

Beispiel 5.14 (Geometrische Verteilung) $g_X(t) = \frac{p}{1-t(1-p)} \quad \forall\ |t| < \frac{1}{1-p}$

$$\begin{aligned}g'_X(t) &= \frac{p(1-p)}{\big(1-t(1-p)\big)^2} \\ g'_X(1-) &= \frac{1-p}{p} = \frac{1}{p} - 1 = E\{X\} \\ g''_X(t) &= \frac{2p(1-p)^2}{\big(1-t(1-p)\big)^3} \\ g''_X(1-) &= 2\left(\frac{1-p}{p}\right)^2 \\ Var\{X\} &= 2\left(\frac{1-p}{p}\right)^2 + \frac{1-p}{p}\left(1 - \frac{1-p}{p}\right) \\ &= \frac{1-p}{p}\left(2\,\frac{1-p}{p} + 1 - \frac{1-p}{p}\right) \\ &= \frac{1-p}{p^2}\,(2 - 2p + p - 1 + p) = \frac{1-p}{p^2}\end{aligned}$$

5.4 Unabhängigkeit von Zufallsvariablen

Wir wollen nun den Begriff der Unabhängigkeit von Ereignissen auf Zufallsvariable $X_1, \ldots, X_n$ übertragen. Dazu ist es notwendig, dass diese Zufallsvariablen auf dem gleichen (**gemeinsamen**) WR $(\Omega, \mathcal{P}(\Omega), P)$ definiert sind. Sind die X_i Zufallsvariable auf verschiedenen Ω_i $(i = 1, \ldots, n)$ definiert, entstammen also unterschiedlichen Zufallsexperimenten, wählt man

$$\Omega \;:=\; \Omega_1 \times \cdots \times \Omega_n \;.$$

Sind darüberhinaus die Zufallsexperimente unabhängig, kann man für P die Produktwahrscheinlichkeit, also insgesamt den Produktwahrscheinlichkeitsraum gem. Definition 4.5 wählen. In jedem Fall identifiziert man für alle $\omega = (\omega_1, \ldots \omega_n) \in \Omega$ mit $\omega_i \in \Omega_i$ $(i = 1, \ldots, n)$

$$X_i(\omega) \;:=\; X_i(\omega_i) \qquad \forall\; i = 1, \ldots, n \;.$$

Dann gilt

$$\begin{aligned} & P(\underbrace{X_1 = x_1}_{:=B_1 \subseteq \Omega_1}, \ldots, \underbrace{X_n = x_n}_{:=B_n \subseteq \Omega_n}) = P\big(B_1 \times \cdots \times B_n\big) \\ = \;\; & P_1(B_1) \cdot P_2(B_2) \cdots P_n(B_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n) \end{aligned}$$

besitzt.

Definition 5.8 (Vollständige Unabhängigkeit von Zufallsvariablen) *Die Zufallsvariablen* $X_1, \ldots, X_n$ *auf* $(\Omega, \mathcal{P}(\Omega), P)$ *heißen genau dann (vollständig)* **unabhängig**, *wenn*

$$P(X_{i_1} \in C_1, \ldots, X_{i_k} \in C_k) \;=\; P(X_{i_1} \in C_1) \cdots P(X_{i_k} \in C_k)$$

für alle $C_j \subseteq \Omega_{X_{i_j}}$ $(j = 1, \ldots, n)$ *und beliebige Teilauswahlen* $i_1, \ldots, i_k$ *aus* $\{1, \ldots, n\}$ *und alle* $2 \leq k \leq n$ *gilt.*

Bemerkungen:

1. Wie bei der Unabhängigkeit von Ereignissen gibt es auch den Begriff der paarweisen Unabhängigkeit von Zufallsvariablen, der geringere Anforderungen stellt. Wir werden ihn aber im Verlauf dieser Vorlesung nicht benötigen.

2. Man zeigt leicht, dass es bei diskreten Wahrscheinlichkeitsräumen zur Überprüfung der Unabhängigkeit gem. Definition 5.8 genügt,

$$P\,(X_1 = x_1 \,,\; \ldots \;,\; X_n = x_n) \;=\; P(X_1 = x_1) \cdots P(X_n = x_n)$$

   für alle $x_i \in \Omega_i$ $(i = 1, \ldots, n)$ zu zeigen.

3. Die obigen ZV $X_1, \ldots, X_n$, deren gemeinsamer WR der Produktraum ist, sind demnach unabhängig.

Beispiel 5.15 (Bernoulli–Experiment) *Die Ausgänge eines Bernoulli–Experimentes mit* $\Omega = \{E, \overline{E}\}$, $P(\{E\}) = p$ *und* $P(\{\overline{E}\}) = 1 - p$ *werden durch die Zufallsvariable*

$$X(\omega) \;=\; \begin{cases} 1 & \text{falls } \omega = E \\ 0 & \text{falls } \omega \neq E \end{cases}$$

beschrieben. Für das n–fache Bernoulli–Experiment (n–fache unabhängige Ausführung) sind die Zufallsvariablen $X_1, \ldots X_n$ *für die Einzelexperimente gem. dieser Verteilung unabhängige Zufallsvariable auf dem Produktwahrscheinlichkeitsraum*

$$(\Omega^n, \mathcal{P}(\Omega^n), P^n) \qquad \text{mit} \qquad \Omega^n = \mathop{\times}_{i=1}^{n} \Omega \quad , \quad P^n = \bigotimes_{i=1}^{n} P \; .$$

Satz 5.9 (Rechenregeln für unabhängige Zufallsvariable) *Seien* $X_1, \ldots, X_n$ *n unabhängige Zufallsvariable mit endlichen Erwartungswerten, dann gilt*

1. *Die ZV* $X_1 \cdot X_2 \cdots X_n$ *besitzt einen endlichen Erwartungswert, nämlich*

$$E\{X_1 \cdot X_2 \cdots X_n\} \;=\; E\{X_1\} \cdot E\{X_2\} \cdots E\{X_n\}$$

2. *Existieren zusätzlich die Varianzen der* X_i $(i = 1, \ldots, n)$ *, gilt*

$$Var\left\{\sum_{i=1}^{n} a_i \cdot X_i\right\} \;=\; \sum_{i=1}^{n} a_i^2 \cdot Var\{X_i\}$$

3. *Existieren die erzeugenden Funktionen* $g_{X_1}, \ldots, g_{X_n}$ $\forall\, |t| \leq t_0$ *dann gilt:*

$$g_{\sum_{i=1}^{n} X_i}(t) \;=\; \prod_{i=1}^{n} g_{X_i}(t) \quad \forall\; |t| \leq t_0$$

<u>Beweis</u>:

Es genügt den Beweis jeweils für $n = 2$ zu führen.

1. Nach Voraussetzung gilt $\mathrm{E}\{|X_1|\}, \mathrm{E}\{|X_2|\} \;<\; \infty$ und mit $Z := X_1 \cdot X_2$

$$\begin{aligned}
\mathrm{E}\{|Z|\} \;&=\; \sum_{z \in \Omega_Z} |z| \cdot P(Z = z) \\
&=\; \sum_{z \in \Omega_Z} |z| \sum_{x_1 \in \Omega_{X_1}} \; \sum_{x_2 \in \Omega_{X_2} :\, x_1 \cdot x_2 = z} P(X_1 = x_1 \,,\, X_2 = x_2) \\
&=\; \sum_{x_1 \in \Omega_{X_1}} \sum_{x_2 \in \Omega_{X_2}} \underbrace{|x_1 \cdot x_2|}_{|x_1| \cdot |x_2|} \cdot P(X_1 = x_1) \cdot P(X_2 = x_2) \\
&=\; \mathrm{E}\{|X_1|\} \cdot \mathrm{E}\{|X_2|\} \;<\; \infty
\end{aligned}$$

Damit ist die Existenz nachgewiesen. Führt man die gleiche Rechnung ohne Betragsstriche aus, erhält man die obige Formel.

2. Auf den Nachweis der Existenz soll hier verzichtet werden.

$$\begin{aligned}
\mathrm{Var}\{a_1X_1+a_2X_2\} &= \mathrm{E}\left\{(a_1X_1+a_2X_2)^2\right\} \;-\; \mathrm{E}^2\{a_1X_1+a_2X_2\} \\
&= \mathrm{E}\left\{a_1^2X_1^2 \;+\; 2a_1a_2X_1X_2 \;+\; a_2^2X_2^2\right\} \;-\; \big(a_1\mathrm{E}\{X_1\} \;+\; a_2\mathrm{E}\{X_2\}\big)^2 \\
&= a_1^2\mathrm{E}\left\{X_1^2\right\} \;+\; 2a_1a_2\mathrm{E}\{X_1\cdot X_2\} \;+\; a_2^2\mathrm{E}\left\{X_2^2\right\} \\
&\quad -a_1^2\mathrm{E}^2\{X_1\} \;-\; 2a_1a_2\mathrm{E}\{X_1\}\mathrm{E}\{X_2\} \;-\; a_2^2\mathrm{E}^2\{X_2\} \\
&= a_1^2\mathrm{Var}\{X_1\}+a_2^2\mathrm{Var}\{X_2\}+2a_1a_2\left(\mathrm{E}\{X_1X_2\}-\mathrm{E}\{X_1\}\mathrm{E}\{X_2\}\right) \\
&= a_1^2\mathrm{Var}\{X_1\} \;+\; a_2^2\mathrm{Var}\{X_2\}
\end{aligned}$$

3. Mit $X_1 \ldots, X_n$ sind auch $t^{X_1}, \ldots, t^{X_n}$ vollständig unabhängig. Damit erhalten wir für alle $|t| \leq t_0$:

$$g_{\sum X_i(t)} \;=\; \mathrm{E}\left\{t^{\sum X_i}\right\} \;=\; \mathrm{E}\left\{\prod_{i=1}^n t^{X_i}\right\} \;=\; \prod_{i=1}^n \mathrm{E}\left\{t^{X_i}\right\} \;=\; \prod_{i=1}^n g_{X_i}(t)\ .$$

Beispiel 5.16 (Binomialverteilung) *Ist die ZV $X \sim Bi(n,p)$, so lässt sie sich als Summe von n unabhängigen Bernoullivariablen X_i mit*

$$P(X_i=1)=p \quad bzw. \quad P(X_i=0)=1-p\ (i=1,\ldots,n)$$

darstellen. Es gilt also

$$X \;=\; \sum_{i=1}^n X_i\ .$$

Damit erhalten wir

$$\begin{aligned}
E\{X\} &= E\left\{\sum_{i=1}^n X_i\right\} \;=\; \sum_{i=1}^n E\{X_i\} \;=\; n\cdot\big(1\cdot p \;+\; 0\cdot(1-p)\big) \;=\; np \\
Var\{X\} &= Var\left\{\sum_{i=1}^n X_i\right\} \;=\; \sum_{i=1}^n Var\{X_i\} \;=\; n\cdot\left(E\left\{X_1^2\right\} \;-\; E^2\{X_1\}\right) \\
&= n\big(1^2\cdot p\;-p^2\big) \;=\; np(1-p)
\end{aligned}$$

Definition 5.9 *Sind X_1 und X_2 Zufallsvariable deren Erwartungswerten existieren, so heißen*

$$X_1, X_2\ \textbf{unkorreliert} \quad \Longleftrightarrow \quad E\{X_1\cdot X_2\} \;=\; E\{X_1\}\cdot E\{X_2\}\ .$$

Allgemein heißt

$$cov\{X_1,X_2\} \;:=\; E\{(X_1-E\{X_1\})\cdot(X_2-E\{X_2\})\} \;=\; E\{X_1\cdot X_2\} \;-\; E\{X_1\}\cdot E\{X_2\}$$

die **Kovarianz** *von X_1 und X_2.*

Bemerkung:

1. Aus der Unabhängigkeit zweier Zufallsvariabler folgt stets deren Unkorreliertheit. dass die Umkehrung i.a. nicht gilt, wird in den Übungen gezeigt.

2. Man sieht leicht, dass für den Beweis des Satzes 5.9 auch die paarweise Unkorreliertheit der Zufallsvariablen genügt.

3. Aus der vorletzten Zeile des Beweises von Satz 5.9 (2) erhalten wir

$$\mathrm{Var}\{a_1 X_1 + a_2 X_2\} = a_1^2 \mathrm{Var}\{X_1\} + a_2^2 \mathrm{Var}\{X_2\} + 2a_1 a_2 \mathrm{cov}\{X_1, X_2\}$$

und damit allgemein

$$\begin{aligned} \mathrm{Var}\left\{\sum_{i=1}^{n} \alpha_i X_i\right\} &= \sum_{i=1}^{n} \alpha_i^2 \mathrm{Var}\{X_i\} + \sum_{i \neq j} \alpha_i \alpha_j \mathrm{cov}\{X_i, X_j\} \\ &= \sum_{i=1}^{n} \alpha_i^2 \mathrm{Var}\{X_i\} + 2 \sum_{i<j} \alpha_i \alpha_j \mathrm{cov}\{X_i, X_j\} \end{aligned}$$

5.5 Gesetz der großen Zahlen

Der folgende Satz belegt den Zusammenhang zwischen der Varianz einer Zufallsvariablen und der Wahrscheinlichkeit für ihr Abweichen von ihrem Erwartungswert.

Satz 5.10 (Chebychev – Ungleichung) *Ist X eine ZV auf dem diskreten WR $\big(\Omega, \mathcal{P}(\Omega), P\big)$ mit endlichem Erwartungswert $E\{X\}$ und endlicher Varianz $Var\{X\}$, so gilt*

$$P\big(|X - E\{X\}| \geq t\big) \leq \frac{Var\{X\}}{t^2} \qquad \forall\, t > 0\,.$$

<u>Beweis</u>:

$$\begin{aligned} & P\big(|X - \mathrm{E}\{X\}| \geq t\big) \\ = & \sum_{x \in \Omega_X :\, |x - EX| \geq t} P(X = x) \leq \sum_{x \in \Omega_X :\, |x - EX| \geq t} \underbrace{\frac{|x - \mathrm{E}\{X\}|^2}{t^2}}_{\geq 1} \cdot P(X = x) \\ \leq & \frac{1}{t^2} \sum_{x \in \Omega_X} \big(x - \mathrm{E}\{X\}\big)^2 \cdot P(X = x) = \frac{\mathrm{Var}\{X\}}{t^2} \end{aligned}$$

$\blacksquare$

Das folgende Beispiel belegt, dass diese Ungleichung scharf ist, es also eine Zufallsvariable gibt, für die in der Chebychev–Ungleichung sogar das Gleichheitszeichen gilt. Die Ungleichung kann deshalb nicht mehr allgemein verbessert werden.

Beispiel 5.17 *Die Zufallsvariable X nehme die Werte 0 , $+k$ und $-k$ mit den folgenden Wahrscheinlichkeiten an:*

$$P(X = -k) = P(X = +k) = \frac{1}{2k^2} \qquad \textit{und} \qquad P(X = 0) = 1 - \frac{1}{k^2}\,.$$

Man rechnet leicht nach, dass

$$\begin{aligned} E\{X\} &= -k \cdot \frac{1}{2k^2} + k \cdot \frac{1}{2k^2} + 0 \cdot \left(1 - \frac{1}{k^2}\right) = 0 \\ & \textit{und} \\ Var\{X\} &= E\Big\{(X - E\{X\})^2\Big\} = E\big\{X^2\big\} = \frac{k^2}{2k^2} + \frac{k^2}{2k^2} + 0 = 1 \end{aligned}$$

gilt. Somit erhalten wir

$$P\big(|X - E\{X\}|) \geq k\big) = P(X = +k) + P(X = -k) = \frac{1}{k^2} = \frac{Var\{X\}}{k^2}\,.$$

Trotzdem ist diese Ungleichung in vielen praktischen Fällen (insbes. für $t < \mathrm{Var}\{X\}$) wenig nützlich, da sie häufig nur die triviale Abschätzung

$$P\big(|X - \mathrm{E}\{X\}| \geq t\big) \leq 1$$

liefert.

<u>Bemerkung</u>:

Oft findet man die Chebychev – Ungleichung auch in der folgenden Form, in der t als Vielfaches der Streuung $S(X) = \sqrt{\mathrm{Var}\{X\}}$ ausgedrückt wird.

$$t := k \cdot S(X) \qquad \Longrightarrow \qquad P\big(|X - \mathrm{E}\{X\}| \geq k \cdot S(X)\big) \leq \frac{\mathrm{Var}\{X\}}{k^2 \cdot \mathrm{Var}\{X\}} = \frac{1}{k^2}$$

Wir werden nun diese Ungleichung benutzen, um den folgenden Satz zu beweisen.

Satz 5.11 ((Schwaches) Gesetz der großen Zahlen) *Für jedes $n \in \mathbb{N}$ seien $X_1, \ldots, X_n$ paarweise unkorrellierte Zufallsvariablen auf Ω, die alle den gleichen endlichen Erwartungswert besitzen und deren Varianzen gleichmäßig beschränkt sind. Es gelte also*

$$E\{X_1\} = \cdots = E\{X_n\} = \mu < \infty \quad \textit{und} \quad Var\{X_i\} \leq \sigma^2 < \infty \; (i = 1, \ldots, n)$$

Bezeichnet man mit $S_n := X_1 + \cdots + X_n$ die Folge der n-ten Partialsummen, so gilt für die Folge der Mittelwerte $\overline{X}_n := \frac{1}{n} S_n$

$$\lim_{n \to \infty} P\big(\,|\overline{X}_n - \mu)| \geq \varepsilon\big) = 0 \qquad \forall\, \varepsilon > 0\,.$$

<u>Beweis</u>:

Wegen

$$\mathrm{E}\{S_n\} = \mathrm{E}\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} \underbrace{\mathrm{E}\{X_i\}}_{\mu} = n \cdot \mu$$

gilt

$$\begin{aligned} P\big(\,|\overline{X}_n - \mu)| \geq \varepsilon\big) &= P\big(\,|S_n - n \cdot \mu)| \geq n \cdot \varepsilon\big) \\ &= P\big(\,|S_n - \mathrm{E}\{S_n\})| \geq n \cdot \varepsilon\big) \\ &\leq \frac{\mathrm{Var}\{S_n\}}{n^2 \varepsilon^2} = \frac{\mathrm{Var}\{X_1\} + \cdots + \mathrm{Var}\{X_n\}}{n^2 \varepsilon^2} \\ &\leq \frac{n \sigma^2}{n^2 \varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2} \longrightarrow 0 \end{aligned}$$

$\blacksquare$

Bemerkungen:

1. Die Voraussetzung der paarweisen Unkorreliertheit ist erfüllt, falls die X_i unabhängig sind.
2. Die glm. Beschränktheit der Varianzen liegt vor, wenn alle Varianzen darüberhinaus endlich und untereinander gleich sind.

Interpretation:

Führen wir beliebig viele Zufallsexperimente (unabhängig) durch und ermitteln jedes Mal den Wert einer Zufallsvariablen X, so gilt für den Mittelwert dieser Variablen, dass die Wahrscheinlichkeiten für beliebig kleine Abweichungen vom Erwartungswert bei wachsender Anzahl der Experimente gegen 0 strebt. Man sagt auch, dass der Mittelwert $\overline{X}_n$ **stochastisch gegen den Erwartungswert konvergiert** und schreiben

$$\overline{X}_n \xrightarrow{P} \mathrm{E}\{X\} \ .$$

Beispiel 5.18 (Binomialverteilung) *Ist die ZV X binomialverteilt, so gilt*

$$X \ \sim \ Bi(n,p) \qquad \Longrightarrow \qquad X \ = \ \sum_{i=1}^{n} X_i$$

mit

$$X_i \ \textit{unabhängig} \ , \ E\{X_i\} \ = \ p \ , \ \ Var\{X_i\} \ = \ p(1-p) \qquad \forall \ i = 1, \ldots, n \ .$$

Damit gilt gem. des schwachen Gesetzes der großen Zahlen

$$P\big(\,\big|\overline{X}_n \,-\, p)\big| \ \geq \ \varepsilon\big) \quad \longrightarrow \quad 0 \quad , \quad \textit{also} \ \ \overline{X}_n \xrightarrow{P} p \ .$$

$\overline{X}_n$ ist aber gerade die **relative Häufigkeit der Erfolge** *der Bernoulli–Experimente. So zeigt der Satz, dass diese relative Häufigkeit mit wachsendem n mit immer kleinerer Wahrscheinlichkeit von der Wahrscheinlichkeit p für einen solchen Erfolg abweicht.*

Auch für den Sortieralgorithmus **Quicksort** lässt sich ein solches *Gesetz der großen Zahlen* beweisen.

Beispiel 5.19 (Quicksort) *Wir wissen, dass für die erwartete Anzahl $E\{X_n\} \ := \ M(n)$ von erforderlichen Vergleichen zum Ordnen einer Liste der Länge n*

$$\lim_{n\to\infty} \frac{M(n)}{n \log n} \ = \ 2$$

gilt. Durch eine ähnliche Rechnung zeigt man $Var\{X_n\} \ \leq \ c \cdot n^2$ mit einer von n unabhängigen Konstanten $c > 0$.

Betrachten wir nun die Zufallsvariablen $Z_n(\omega) := \frac{X_n(\omega)}{n \log n}$, *erhalten wir*

$$P\big(\,|\overline{Z}_n - E\{Z_n\})| \geq \varepsilon\big) \leq \frac{Var\{Z_n\}}{\varepsilon^2} = \frac{Var\{X_n\}}{n^2(\log n)^2\varepsilon^2} \leq \frac{c}{\varepsilon^2(\log n)^2} \; .$$

Wegen

$$E\{Z_n\} = \frac{E\{X_n\}}{n \log n} = \frac{M_n}{n \log n} \longrightarrow 2 \; ,$$

betrachten wir direkt

$$P\big(\underbrace{|Z_n(\omega) - 2| \geq \varepsilon}_{:= A_\varepsilon}\big) \; .$$

Für alle $\omega \in A_\varepsilon$ *gilt nun*

$$|Z_n(\omega) - E\{Z_n\}| \geq \frac{\varepsilon}{2} \quad \textit{oder} \quad |E\{Z_n\} - 2| \geq \frac{\varepsilon}{2} \; .$$

Anderenfalls könnte man wegen

$$\begin{aligned} |Z_n(\omega) - 2| &= |Z_n(\omega) - E\{Z_n\} + E\{Z_n\} - 2| \leq |Z_n(\omega) - E\{Z_n\}\,| + |E\{Z_n\} - 2| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

zeigen.

Wird n_0 *nun so groß gewählt, dass für alle* $n \geq n_0$ $|E\{Z_n\} - 2| < \frac{\varepsilon}{2}$ *gilt, gilt für* $n \geq n_0$ *und* $\omega \in A_\varepsilon$ *natürlich*

$$|Z_n(\omega) - E\{Z_n\}| \geq \frac{\varepsilon}{2}$$

und damit für alle $n \geq n_0$

$$\begin{aligned} P\big(|Z_n(\omega) - 2| \geq \varepsilon\big) &\leq P\left(|Z_n - E\{Z_n\}| \geq \frac{\varepsilon}{2}\right) \\ &\leq \frac{c}{(\varepsilon \cdot \log n)^2/4} \longrightarrow 0 \; . \end{aligned}$$

Damit gilt für die Anzahl X_n *der notwendigen Vergleiche von QUICKSORT für das Sortieren einer Liste der Länge* n

$$\lim_{n\to\infty} P\left(\left|\frac{X_n}{n \log n} - 2\right| \geq \varepsilon\right) = 0 \qquad \textit{bzw.} \qquad \frac{X_n}{n \log n} \xrightarrow{P} 2 \; .$$

Kapitel 6

Approximationen der Binomialverteilung

Im vorhergehenden Kapitel haben wir die binomialverteilte Zufallsvariable

$$X \;\sim\; Bi(n,p) \qquad \text{mit} \qquad P(X\;=\;k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \quad \forall\; k = 0,1,\ldots,n$$

eingeführt.

6.1 Die Poisson – Verteilung

Jetzt wollen wir mit der **Poisson – Verteilung** eine weitere wichtige Verteilung auf der Menge der nichtnegativen ganzen Zahlen

$$\mathbb{N}_0 \;=\; \{0,1,2\ldots\}$$

kennen lernen.

Dazu betrachten wir für eine feste reelle Zahl $\lambda > 0$ die Wahrscheinlichkeitsverteilung

$$\{(k,p_k)\;:\;k\in\mathbb{N}_0\} \qquad \text{mit} \qquad p_k \;=\; p_k(\lambda) \;:=\; \frac{\lambda^k}{k!}\,e^{-\lambda} \quad \forall\; k\in\mathbb{N}_0\;.$$

Zunächst überzeugen wir uns mit Hilfe der Taylorreihe für e^x

$$e^x \;=\; \sum_{k=0}^{\infty}\frac{x^k}{k!} \qquad \forall\; x\in\mathbb{R}\;,$$

dass hierdurch tatsächlich eine Wahrscheinlichkeitsverteilung gegeben ist.

Tatsächlich gilt

$$\sum_{k=0}^{\infty} p_k(\lambda) \;=\; e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \;=\; e^{-\lambda}\cdot e^{\lambda} \;=\; 1\;.$$

Definition 6.1 *Eine Zufallsvariable* X *mit dem Wertebereich* $\Omega_X = \mathbb{N}_0$ *und der Verteilung*

$$P(X = k) \;=\; p_k(\lambda) \;=\; \frac{\lambda^k}{k!}\, e^{-\lambda} \qquad \forall\; k \in \mathbb{N}_0$$

heißt **poissonverteilt** *mit dem Parameter* λ. *Wir schreiben dann*

$$X \;\sim\; Po(\lambda)\;.$$

Satz 6.1 (Poisson–Verteilung) *Für* $X \sim Po(\lambda)$ *gilt*

(6.1) $$E\{X\} \;=\; \lambda \qquad \textit{und}$$

(6.2) $$Var\{X\} \;=\; \lambda$$

<u>Beweis</u>:

$$\begin{aligned}
\mathrm{E}\{X\} &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \;=\; \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda \cdot e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \;=\; \lambda \cdot e^{-\lambda} \cdot e^{\lambda} \;=\; \lambda \\
\mathrm{E}\left\{X^2\right\} &= \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} \;=\; e^{-\lambda} \sum_{k=1}^{\infty} \big(k(k-1) + k\big) \cdot \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=2}^{\infty} k(k-1) \cdot \frac{\lambda^k}{k!} \;+\; \lambda \;=\; \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \;+\; \lambda \\
&= \lambda^2 \;+\; \lambda \\[1em]
\mathrm{Var}\{X\} &= \lambda^2 \;+ \lambda \;-\; \lambda^2 \;=\; \lambda
\end{aligned}$$

$\blacksquare$

Diese Poissonverteilung ergibt sich nun auch aus einem Grenzübergang aus der Binomialverteilung.

Dazu gehen wir von einer Folge von Zufallsvariablen

$$X_n \;\sim\; Bi(n, p) \qquad (n \in \mathbb{N})$$

aus und lassen zu, dass p von der Anzahl der Bernoulli–Experimente abhängt, also $p = p_n$ gilt.

$$X_n \;\sim\; Bi(n, p_n)$$

Satz 6.2 (Poisson'scher Grenzübergang)
Gilt $X_n \sim Bi(n, p_n)$ *mit* $E\{X_n\} = np_n \longrightarrow \lambda > 0$ *für* $n \longrightarrow \infty$ *und damit auch* $p_n \longrightarrow 0$, *so gilt weiter*

$$\lim_{n\to\infty} P(X_n = k) \;=\; \frac{\lambda^k}{k!}\, e^{-\lambda} \;=\; p_k(\lambda)$$

Bemerkungen:

1. Setzt man den Erwartungswert der binomialverteilten Zufallsvariablen X_n
$$\mathrm{E}\{X_n\} \;=\; np_n \;:=\; \lambda_n \;,$$
so lautet die Aussage des Satzes
$$\lambda_n \;\longrightarrow\; \lambda \qquad \Longrightarrow \qquad P(X_n = k) \;\longrightarrow\; p_k(\lambda) \;=\; \frac{(\lambda)^k}{k!}e^{-\lambda} \;.$$
Man sagt auch, dass die Folge der Zufallsvariablen X_n **verteilungskonvergent** ist, bzw. "in Verteilung" gegen eine poissonverteilte Zufallsvariable X konvergent ist, und schreibt dafür
$$X_n \;\overset{\mathcal{D}}{\curvearrowright}\; X \;\sim\; \mathrm{Po}(\lambda) \qquad \text{oder} \qquad X_n \;\overset{\mathcal{D}}{\curvearrowright}\; \mathrm{Po}(\lambda) \;.$$

2. Der Satz besagt: Für große n $(n \gg k \;,\; np)$ gilt
$$X \;\sim\; Bi(n,p) \qquad \Longrightarrow \qquad P(X = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \;\approx\; p_k(np) \;=\; \frac{(np)^k}{k!}\, e^{-np} \;.$$

3. Ist $X_n \sim Bi(n, p_n)$ so gilt in der Situation der Bemerkung 1 für große n
$$\begin{aligned} \mathrm{E}\{X_n\} \;&=\; np_n \;:=\; \lambda_n \;\approx\; \lambda \quad \text{und} \\ \mathrm{Var}\{X_n\} \;&=\; np_n(1-p_n) \;=\; \lambda_n \left(1 - \frac{\lambda_n}{n}\right) \;\approx\; \lambda\left(1 - \frac{\lambda}{n}\right) \;\approx\; \lambda \end{aligned}$$

Beweis (Satz 6.2):

Mit $\lambda_n = np_n$ und $\lambda_n \to \lambda$ erhalten wir für jedes feste k
$$\begin{aligned} P(X_n = k) \;&=\; \binom{n}{k} p_n^k (1-p_n)^{n-k} \;=\; \binom{n}{k} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\ &=\; \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \frac{\lambda_n^k}{n^k} \cdot \frac{\left(1 - \frac{\lambda_n}{n}\right)^n}{\left(1 - \frac{\lambda_n}{n}\right)^k} \\ &=\; \underbrace{\frac{\lambda_n^k}{k!}}_{\longrightarrow \frac{\lambda^k}{k!}} \cdot \underbrace{\frac{1}{\left(1 - \frac{\lambda_n}{n}\right)^k}}_{\longrightarrow 1} \cdot \underbrace{\left(1 - \frac{1}{n}\right)}_{\longrightarrow 1} \cdot \underbrace{\left(1 - \frac{2}{n}\right)}_{\longrightarrow 1} \cdots \underbrace{\left(1 - \frac{k-1}{n}\right)}_{\longrightarrow 1} \cdot \left(1 - \frac{\lambda_n}{n}\right)^n \\ &\longrightarrow\; \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \left(1 - \frac{\lambda_n}{n}\right)^n \;, \text{ falls dieser Grenzwert existiert.} \end{aligned}$$

Um den Grenzwert zu ermitteln, betrachten wir weiter
$$\log\left(1 - \frac{\lambda_n}{n}\right)^n \;=\; n \cdot \log\left(1 - \frac{\lambda_n}{n}\right) \;=\; -\lambda_n \cdot \frac{\log\left(1 - \frac{\lambda_n}{n}\right)}{-\frac{\lambda_n}{n}} \;\longrightarrow\; -\lambda \;,$$

denn es gilt für $x_n \to 0$ nach der Regel von l'Hospital

$$\frac{\log(1 - x_n)}{-x_n} = -\frac{d}{dx}\log(1-x)\Big|_{x=0} = \frac{1}{1-x}\Big|_{x=0} = 1 .$$

Also gilt

$$\log\left(1 - \frac{\lambda_n}{n}\right)^n \longrightarrow -\lambda \qquad \Longrightarrow \qquad \left(1 - \frac{\lambda_n}{n}\right)^n \longrightarrow e^{-\lambda}$$

und damit die Behauptung.

Satz 6.3 *Für die erzeugende Funktion einer poissonverteilten Zufallsvariablen*

$$X \sim Po(\lambda)$$

gilt:

$$g_X(t) = e^{-\lambda(1-t)} .$$

Beweis:

$$g_X(t) = \sum_{k=0}^{\infty} t^k \cdot \underbrace{P(X = k)}_{p_k} = e^{-\lambda} \cdot \underbrace{\sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!}}_{e^{\lambda t}}$$

$$= e^{-\lambda + \lambda t} = e^{-\lambda(1-t)}$$

Nun führen wir den Poissonschen Grenzübergang für die erzeugende Funktion durch. Dazu betrachten wir die Folge der Zufallsvariablen

$$X_n \sim Bi(n, p_n)$$

für $n \to \infty$, $p_n \to 0$ mit $np_n \to \lambda$. Dann erhalten wir für die zugehörige Folge der erzeugenden Funktionen nach Beispiel 5.12

$$\begin{aligned} g_{X_n}(t) &= \big(tp_n + 1 - p_n\big)^n = \big(p_n(t-1) + 1\big)^n \\ &= \left(1 + \frac{np_n(t-1)}{n}\right)^n \qquad \text{mit} \quad np_n(t-1) \to \lambda(t-1) \\ &\longrightarrow e^{\lambda(t-1)} = e^{-\lambda(1-t)} \\ &= g_X(t) \qquad \text{und} \qquad X \sim Po(\lambda) \end{aligned}$$

Allgemein kann man in der Analysis den folgenden Satz beweisen.

Satz 6.4 *Seien* $X_1, X_2, \ldots$ *diskrete Zufallsvariable mit den Wahrscheinlichkeitsverteilungen*

$$P(X_n = k) \;=\; p_{n,k} \qquad \forall\ n \in \mathbb{N}$$

und den erzeugenden Funktionen $g_1, g_2, \ldots$ *Konvergiert nun die Folge der erzeugenden Funktionen gegen eine Funktion* $g(t)$, *gilt also*

$$\lim_{n\to\infty} g_n(t) \;=\; g(t) \qquad \forall\ t\ ,$$

und ist die Grenzfunktion g *erzeugende Funktion einer Zufallsvariablen* X *mit der Wahrscheinlichkeitsverteilung*

$$P(X = k) \;=\; \pi_k \qquad \forall\ k \in \mathbb{N}\ ,$$

dann gilt auch

$$\lim_{n\to\infty} p_{n,k} \;=\; \pi_k \qquad \forall\ k \in \mathbb{N}\ .$$

Wir wollen nun untersuchen, in welchen Situationen die so eingeführte *Poisson–Verteilung* als Modell sinnvoll eingesetzt werden kann.

Zunächst einmal betrachten wir ein n-faches Bernoulli–Experiment. Ist n groß und p entsprechend klein, können wir also $np = \lambda$ setzen, sprechen wir auch von der n-fachen unabhängigen Wiederholung eines Zufallsexperients mit "seltenem" Erfolg. Man sagt deshalb auch, dass die Poissonverteilung die **Verteilung der seltenen Ereignisse** ist.

Wenn zufallsabhängige Vorkommnisse sich so über einen Zeitraum verteilen, dass ihre Anzahl pro Zeiteinheit im Durchschnitt über längere Zeit hinweg als relativ konstant angesehen werden kann und diese Anzahlen in beliebigen Zeitabschnitten unabhängig voneinander sind, benutzt man das Poisson–Modell als Träger eines **zeitlich homogenen Chaos**.

Beispiel 6.1 (Zeitlich homogenes Chaos) *An einer Zentraleinheit eines Rechners werden die eingehenden Jobs registriert, wobei sich eine durchschnittliche Häufigkeit von* λ *Jobs pro Zeiteinheit ergibt. Dann ist die Wahrscheinlichkeit, in einem kleinen Zeitintervall der Länge* δ *einen Job zu registrieren, durch* $\lambda\delta + o(\delta)$ *gegeben. Dabei ist* $o(\delta)$ *das* Landausche "klein o" [1]. $o(\delta)$ *strebt also schneller gegen* 0 *als* δ.

Nun teile man ein Zeitintervall $[0, t]$ *der Länge* t *in* N *gleiche Teile, so dass die Wahrscheinlichkeit für einen Job in einem beliebigen Teilintervall gleich*

$$\frac{\lambda t}{N} \;+\; o\left(\frac{t}{N}\right)$$

ist, wobei wir das obige δ *gleich* t/N *gesetzt haben. Natürlich wird* $\lambda t/N$ *beliebig klein, wenn wir* t *festhalten und* N *hinreichend groß wählen.*

Zunächst sei vorausgesetzt, dass die Wahrscheinlichkeit für mehr als einen Job in einer Zeiteinheit für jedes kleine Teilintervall vernachlässigt werden kann, so dass in jedem Teilintervall 0

[1] $f(\delta) = o\big(g(\delta)\big) \iff \lim_{\delta\to 0} \frac{f(\delta)}{g(\delta)} \;=\; 0$

oder 1 *Job eintreffen. Schließlich nehmen wir noch an, dass diese Anzahlen stochastisch unabhängig sind, was durch die Erfahrung gerechtfertigt wird. Wir können diese Anzahlen also als unabhängige Bernoulli–Variable behandeln, die die Werte* 0 *bzw.* 1 *mit den Wahrscheinlichkeiten* $1 - \frac{\lambda t}{N}$ *bzw.* $\frac{\lambda t}{N}$ *annehmen.*

Ist X *also die Zufallsvariable, die zählt, wie viele Jobs im Intervall* $[0, t]$ *eintreffen, gilt*

$$X \sim Bi\left(N, \frac{\lambda t}{N}\right) .$$

Da N *frei wählbar ist und beliebig groß gewählt werden kann, ist es auch hier möglich, die Poisson–Verteilung als Modell zu verwenden:*

$$X \sim Po(\lambda t) .$$

Dieses Modell wird benutzt, um Ankunftsströme (Poissonstrom) bei Warteschlangen (z.B. in Computernetzen) zu modellieren.

6.2 Die Normalverteilung

Die Approximation der Binomialverteilung durch die Poisson–Verteilung ist noch recht neu. Erst die Betrachtung der Huftritttoten in den preußischen Armeen hat die Poisson–Verteilung in die Wahrscheinlichkeitstheorie eingeführt. Andererseits experimentierte man schon sehr frühzeitig mit Münzwurf–Experimenten, zählte Würfelergebnisse aus und maß unzählige physikalische und biologische Größen, die man durch binomialverteilte Zufallsgrößen modellieren konnte.

Da man früher noch keine Rechner zur Verfügung hate, war es kaum möglich, die Wahrscheinlichkeiten für eine Zufallsvariable $X \sim Bi(n, p)$

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

für große n, festes p und beliebiges k exakt zu ermitteln. Man benötigt **normalerweise** eine möglichst gute , leicht zu berechnende Approximation.

Beispiel 6.2 *Man nehme an, dass eine vollkommene Münze* 100*–mal geworfen wird. Mit welcher Wahrscheinlichkeit zeigt sie mindestens* 45*–, aber höchstens* 55*–mal Kopf. Die Lösung*

$$\sum_{k=45}^{55} \binom{100}{k} \frac{1}{2^{100}} = \frac{1}{2^{100}} \cdot \sum_{k=45}^{55} \binom{100}{k}$$

ist nicht zufrieden stellend, solange wir keine Vorstellung von der Größenordnung dieser Wahrscheinlichkeit haben. Liegt sie nun in der Nähe von $\frac{3}{4}$, $\frac{1}{2}$ *oder* $\frac{1}{10}$ *?*

Mit Hilfe des Satzes 5.9(3) lassen sich leicht die folgenden Aussagen verifizieren:

Satz 6.5 (Addition unabhängiger Binomial– und Poissonverteilungen)

1. *Sind* $X_1 \sim Po(\lambda_1)$, $X_2 \sim Po(\lambda_2)$ *und beide unabhängig, so gilt:*
$$X_1 \,+\, X_2 \;\sim\; Po(\lambda_1 \,+\, \lambda_2)\ .$$
2. *Sind* $X_1 \sim Bi(n_1, p)$, $X_2 \sim Bi(n_2, p)$ *und beide unabhängig, so gilt:*
$$X_1 \,+\, X_2 \;\sim\; Bi(n_1 + n_2, p)\ .$$

Beweis:

1. $$g_{X_1+X_2}(t) \;=\; e^{-\lambda_1(1-t)} \cdot e^{-\lambda_2(1-t)} \;=\; e^{-(\lambda_1+\lambda_2)(1-t)}$$
2. $$g_{X_1+X_2}(t) \;=\; (tp \,+\, 1-p)^{n_1} \cdot (tp \,+\, 1-p)^{n_2} \;=\; (tp \,+\, 1-p)^{n_1+n_2}$$

Definition 6.2 *Eine Zufallsvariable* X *heißt* **Standardzufallsvariable**, *falls ihre beiden ersten Momente existieren und*
$$E\{X\} \;=\; 0 \qquad \textit{und} \qquad Var\{X\} \;=\; 1$$
gilt. Besitzt eine beliebige Zufallsvariable X *endliches erstes und zweites Moment mit* $Var\{X\} > 0$ *, so heißt*
$$X_0 \;:=\; \frac{X \,-\, E\{X\}}{\sqrt{Var\{X\}}}$$
die zugehörige **standardisierte Zufallsvariable**.

Bemerkung:

Die standardisierte Zufallsvariable X_0 ist eine Standardzufallsvariable, denn es gilt
$$\mathrm{E}\{X_0\} \;=\; \frac{1}{\sqrt{\mathrm{Var}\{X\}}} \cdot \underbrace{\mathrm{E}\{X \,-\, \mathrm{E}\{X\}\}}_{0} \;=\; 0$$
$$\mathrm{Var}\{X_0\} \;=\; \frac{1}{\mathrm{Var}\{X\}} \cdot \underbrace{\mathrm{Var}\{X \,-\, \mathrm{E}\{X\}\}}_{\mathrm{Var}\{X\}} \;=\; 1$$

Nun wenden wir uns wieder dem Problem der Approximation der Wahrscheinlichkeiten einer binomialverteilten Zufallsvariablen zu. Dabei erinnern wir uns daran, dass wir sie stets als Summe von unabhängigen Bernoullivariablen darstellen können.

Sei nun $X_1, X_2, \ldots$ eine Folge unabhängiger Bernoullivariabler mit der Erfolgswahrscheinlichkeit p, so stellt die Folge der Partialsummen

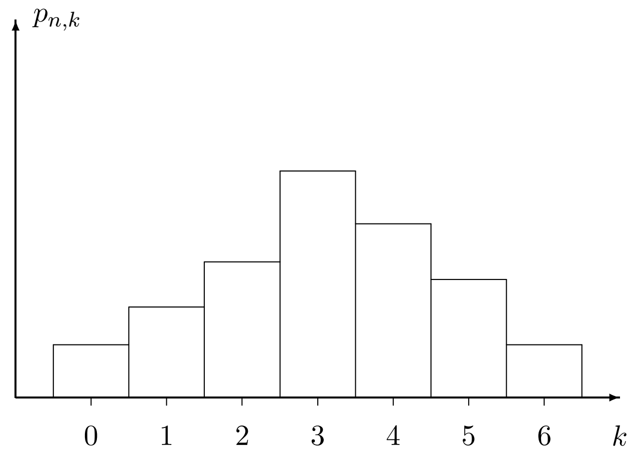
$$S_n \;:=\; \sum_{i=1}^{n} X_i \qquad \forall\, n \in I\!N$$

eine Folge von binomialverteilten Zufallsvariablen

$$S_n \;\sim\; Bi(n,p) \qquad \forall\, n \in I\!N$$

dar. Ihre Wahrscheinlichkeitsverteilungen wollen wir uns in dem folgenden Diagramm (Bild 6.2) anschaulich machen.

Abbildung 6.1: Wahrscheinlichkeitsverteilung einer Binomialverteilung



Dabei stellen die

$$P_n(S_n = k) \;:=\; P(S_n = k) \;:=\; p_{n,k} \;=\; \binom{n}{k} p^k (1-p)^{n-k} \qquad \forall\, k, n$$

für festes n die abgebildeten Wahrscheinlichkeitsverteilungen dar.

In Anlehnung an eine in der Statistik übliche Bezeichnungsweise nennen wir eine solche grafische Darstellung **Histogramm**. Die Flächen der obigen Rechtecke stellen wegen ihrer Breite 1 die Wahrscheinlichkeiten für die Werte dar, über denen sie errichtet sind. Die Gesamtfläche aller Rechtecke ist dabei gleich 1. Auch rechnet man leicht nach, dass der "Gipfel" des Histogramms in unmittelbarer Nähe des Erwartungswertes $\mathrm{E}\{S_n\} = np$ liegt.

Lassen wir nun n wachsen, während p fest bleibt, erkennt man unschwer, dass

$$P(S_n = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k} \;\longrightarrow\; 0 \qquad \forall\, k$$

gilt. Da aber mit $n \to \infty$ auch $np \to \infty$ folgt, wandert der "Gipfel" mit wachsendem n immer weiter nach rechts, während das gesamte "Gebirge" immer flacher wird ("zerfließt").

Um das Abwandern zu verhindern, betrachten wir nun eine neue Zufallsvariable, nämlich

$$S_n \;-\; \mathrm{E}\{S_n\} \;=\; S_n \;-\; np\,.$$

Hier wandert der Gipfel zwar nicht mehr nach rechts, an dem "Zerfließen" ändert sich jedoch nichts. Deshalb wollen wir die Zufallsvariable jetzt vollständig standardisieren.

Die entsprechende Standardvariable von S_n lautet

$$Z_n \;:=\; \frac{S_n \;-\; \mathrm{E}\{S_n\}}{\sqrt{\mathrm{Var}\{S_n\}}} \;=\; \frac{S_n \;-\; np}{\sqrt{np(1-p)}}\,.$$

Wegen $\mathrm{Var}\{Z_n\} = 1$ erhält man aus der Ungleichung von *Chebychev*

$$P(|Z_n| \;\geq\; x) \;\leq\; \frac{1}{x^2} \qquad \forall\, x > 0\,,\; n \in I\!N\,.$$

Für $x = 3$ gilt damit

$$P(|Z_n| \;\geq\; 3) \;\leq\; \frac{1}{3^2} = \frac{1}{9}$$

und die Zufallsvariable Z_n nimmt für jedes n mit einer Wahrscheinlichkeit von mindestens $8/9$ Werte im Intervall $[-3, +3]$ an. Ein Zerfließen mit wachsendem n ist also ausgeschlossen.

Jetzt wollen wir analog zur Abbildung 6.2 ein Histogramm der Wahrscheinlichkeiten

$$p_{n,k} \;=\; P(S_n \;=\; k) \;=\; P\Bigg(Z_n \;=\; \underbrace{\frac{k-np}{\sqrt{np(1-p)}}}_{:=\, x_{n,k}}\Bigg) \;=\; P\big(Z_n \;=\; x_{n,k}\big) \qquad \forall\; k, n$$

erstellen. Da die "Trägerpunkte" $x_{n,k} = \frac{k-np}{\sqrt{np(1-p)}}$ den Abstand $b(n) = \frac{1}{\sqrt{np(1-p)}}$ voneinander haben, ist es erforderlich, alle Wahrscheinlichkeiten mit $\sqrt{np(1-p)}$ zu multiplizieren, damit für die Gesamtfläche F aller Rechtecke

$$F \;=\; \sum_k \underbrace{\sqrt{np(1-p)} \;\cdot\; p_{n,k}}_{:=\, h_n(x_{n,k})} \cdot b(n) \;=\; \sum_k p_{n,k} \;=\; 1$$

gilt. Wir betrachten also die Funktion

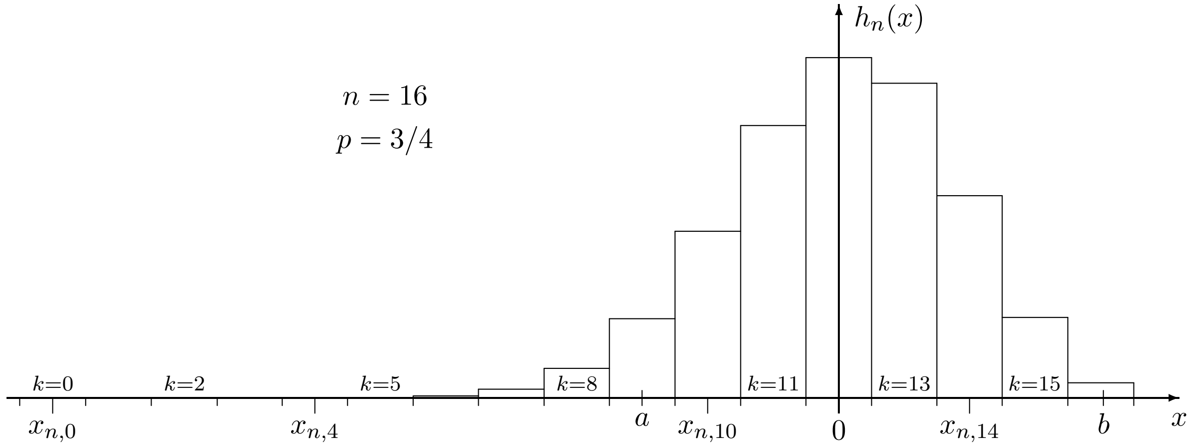
$$h_n(x) \;=\; \sqrt{np(1-p)} \;\cdot\; p_{n,k} \qquad \forall\; x_{n,k} - \frac{b(n)}{2} \;<\; x \;\leq\; x_{n,k} + \frac{b(n)}{2}\,.$$

Für $p = \frac{3}{4}$ und $n = 16$ ergibt sich das Bild 6.3 auf der folgenden Seite.

Lässt man nun n wachsen, stellt man eine zunehmende Stabilisierung bei gleichzeitiger Glättung von $h_n(x)$ fest.

$$\begin{aligned} P\left(a \;\leq\; \frac{S_n \;-\; np}{\sqrt{np(1-p)}} \;\leq\; b\right) \;&=\; \sum_{a \leq k \leq b} P\left(\frac{S_n \;-\; np}{\sqrt{np(1-p)}} \;=\; x_{n,k}\right) \\ &=\; \sum_{a \leq k \leq b} h_n(x_{n,k}) \cdot b(n) \;\approx\; \int_a^b \varphi(x)\, dx \end{aligned}$$

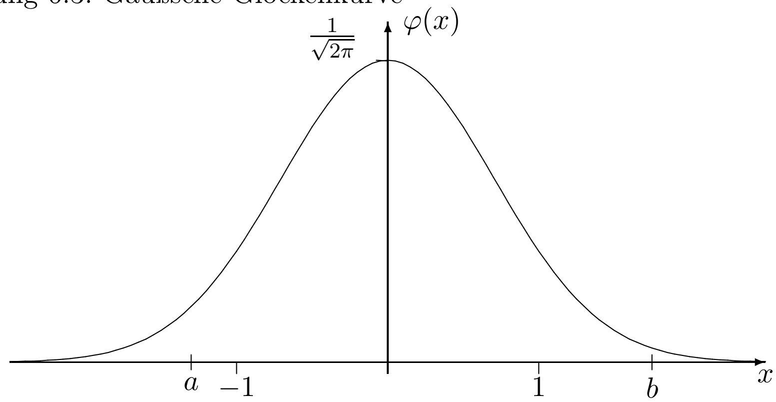
Abbildung 6.2: Wahrscheinlichkeitsverteilung einer standardisierten Binomialverteilung



Tatsächlich lässt sich zeigen, dass die Folge der h_n gegen die Grenzfunktion φ (**Gaußsche Glockenkurve**)

$$\lim_{n\to\infty} h_n(x) \;:=\; \varphi(x) \;=\; \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \qquad \forall\ x \in \mathbb{R}$$

Abbildung 6.3: Gaußsche Glockenkurve



konvergiert und der folgende Satz gilt:

Satz 6.6 (De Moivre – Laplace) *Ist* $(S_n)_{n\in\mathbb{N}}$ *eine Folge von binomialverteilten Zufallsvariablen*

$$S_n \;\sim\; Bi(n,p) \qquad \forall\ n \in \mathbb{N}\ ,$$

dann gilt

$$\lim_{n\to\infty} P_n\left(a \;\leq\; \frac{S_n \;-\; E\{S_n\}}{\sqrt{Var\{S_n\}}} \;\leq\; b\right) \;=\; \int_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx \qquad \forall\ -\infty \leq a \leq b \leq +\infty\ .$$

Diese Konvergenz ist sogar gleichmäßig für alle $[a,b] \subset \mathbb{R}$.

Der Beweis kann mit der Stirlingschen Formel

$$n! \;=\; \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\, e^{\omega(n)} \qquad \text{mit} \qquad \frac{1}{12\left(n \,+\, \frac{1}{2}\right)} \;<\; \omega(n) \;<\; \frac{1}{12n}$$

(vergl. *Krickeberg / Ziezold* [2]) geführt werden. Wir werden im folgenden Kapitel noch in allgemeinerer Form darauf zurückkommen.

Bemerkung:

(Für große n gilt damit für $S_n \sim Bi(n,p)$: '

$$\begin{aligned} P(x_u \le S_n \le x_o) &= P\left(\frac{x_u - np}{\sqrt{np(1-p)}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{x_o - np}{\sqrt{np(1-p)}}\right) \\ &\approx \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{x_u - np}{\sqrt{np(1-p)}}}^{\frac{x_o - np}{\sqrt{np(1-p)}}} e^{-\frac{1}{2}x^2} dx \end{aligned}$$

Beispiel 6.3 *Mit welcher Wahrscheinlichkeit fällt bei* 100*-maligem Münzwurf mindestens* 45*– aber höchstens* 55*–mal Kopf.*

$$P(44.5 \le S_n \le 55.5) \;\approx\; \frac{1}{\sqrt{2\pi}} \int\limits_{-1.1}^{1.1} e^{-\frac{1}{2}x^2} dx$$

Dabei ist die Festsetzung der Grenzen 44.5 *und* 55.5 *relativ willkürlich. Darüberhinaus bleibt das Problem der Berechnung des Integrals. Auf beide Probleme werden wir im folgenden Kapitel eingehen.*

[2]K. Krickeberg, H. Ziezold: "Stochastische Methoden", Springer–Verlag, Berlin, 1977 .

Kapitel 7

Stetige Zufallsvariable

Im letzten Kapitel sind wir auf Wahrscheinlichkeiten gestoßen, die sich durch Integrale approximativ berechnen lassen. Für $S_n \sim Bi(n, p)$ gilt

$$\lim_{n\to\infty} P_n\left(a \leq \frac{S_n - \mathrm{E}\{S_n\}}{\sqrt{\mathrm{Var}\{S_n\}}} \leq b\right) = \int_a^b \underbrace{\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2}}_{\varphi(x)} dx \qquad \forall -\infty \leq a \leq b \leq +\infty .$$

Kann man sich nun vorstellen, dass es eine Zufallsvariable X gibt, für die unmittelbar

$$P\big(a \leq X \leq b\big) = \int_a^b \varphi(x) dx \qquad \forall -\infty \leq a \leq b \leq +\infty$$

gilt?

Zunächst müsste eine solche Zufallsvariable reelle Zahlen als Werte annehmen, wobei jedoch für alle $a \in \mathbb{R}$

$$P(X = a) = P(a \leq X \leq a) = \int_a^a \varphi(x)\, dx = 0 \qquad \forall\, a \in \mathbb{R}$$

gelten sollte.

Eine diskrete Zufallsvariable kann diese Eigenschaft nicht haben. Wir müssen also den Begriff des Wahrscheinlichkeitsraumes erweitern.

Beispiel 7.1 *Es soll die Höhe "Normalnull" (NN) festgelegt werden. Dazu wird an einem festen Pegel am Meeresufer zu festen Zeitpunkten die Höhe des Wasserstandes X abgelesen. Der "mittlere" Wasserstand soll als NN bezeichnet werden. Klar ist, dass der Wasserstand ständig in nicht vollständig vorhersehbarer Weise schwankt (Wellenbewegungen, Tidenhub). Daher können wir diese Ablesungen als Versuchsausgänge eines zufälligen Experimentes auffassen. Wegen der* *möglichst genauen Ablesung und der Vielzahl solcher Versuchsausgänge wählen wir als Menge der Versuchsausgänge*

$$\Omega \;=\; I\!R \;.$$

Auf diesem Ω müssen wir nun adäquat eine Zufallsvariable X zu definieren, die es uns u.a. gestattet, ihren Erwartungswert als NN zu bezeichnen.

7.1 Allgemeine Wahrscheinlichkeitsräume

Leider stoßen wir dabei auf das folgende Problem.

In der **Maßtheorie** wird gezeigt, dass es für die meisten folgenden Modelle nicht möglich ist, auf der Potenzmenge einer überabzählbaren Menge (hier : $I\!R$) Wahrscheinlichkeiten gem. den Axiomen von Kolmogorov festzulegen. Das geht nur, wenn wir uns auf eine echte Teilmenge $\mathcal{S}$ dieser Potenzmenge beschränken [1].

Damit wir mit diesen Ereignissen $A \in \mathcal{S}$ weiterhin rechnen können wie wir das mit Ereignissen aus der Potenzmenge gewöhnt sind, müssen wir fordern, dass diese Ereignismenge $\mathcal{S}$ zumindest gegenüber den Operationen $\cap$, $\cup$, und der Negation abgeschlossen ist.

Definition 7.1 *Ein nichtleeres Mengensystem $\mathcal{S} \subseteq \mathcal{P}(\Omega)$ heißt genau dann σ –* **Algebra (über Ω)***, wenn*

$$A \in \mathcal{S} \implies \overline{A} \in \mathcal{S} \tag{7.1}$$

$$A_1, A_2, A_3, \ldots \in \mathcal{S} \implies \bigcup_i A_i \in \mathcal{S} \tag{7.2}$$

Bemerkung:

1. Man zeigt leicht, dass in einer σ – Algebra $\mathcal{S}$ auch

   $$A_1, A_2, A_3, \ldots \in \mathcal{S} \implies \bigcap_i A_i \in \mathcal{S}$$

   gilt.

2. Da eine σ – Algebra $\mathcal{S} \neq \emptyset$ ist, existiert eine Teilmenge

   $$A \subset \Omega \quad \text{mit} \quad A \in \mathcal{S}$$

   und damit auch

   $$\overline{A} \in \mathcal{S} \quad \implies \quad A \cap \overline{A} = \emptyset \in \mathcal{S} \quad \text{und} \quad A \cup \overline{A} = \Omega \in \mathcal{S} \;.$$

   Damit gilt für jede σ – Algebra $\mathcal{S}$ über Ω: $\qquad \emptyset, \Omega \in \mathcal{S}$.

[1] H. Bauer: "Wahrscheinlichkeitstheorie", W.de Gruyter, 4. Auflage, Berlin, 1991

3. Die kleinste σ – Algebra über Ω enthält lediglich Ω und $\emptyset$ $(\mathcal{S} = \{\emptyset, \Omega\})$. Diese σ–Algebra ist i.a. nicht reichhaltig genug für das Modell.

Da wir uns hier meist mit Versuchsausgängen $\omega \in I\!R$ beschäftigen, interessiert besonders der folgende Satz, der in der Maßtheorie bewiesen wird:

Satz 7.1 *Es existiert eindeutig die kleinste σ–Algebra $\mathcal{B}^1$ über $I\!R^1$, die alle Intervalle $(a, b] \subset I\!R$ enthält. Sie heißt die* **Borelsche σ–Algebra**.

Bemerkungen:

Auf der Borelschen σ–Algebra ist es möglich, für alle später folgenden Modelle eine Wahrscheinlichkeit P festzulegen, die den Axiomen von Kolmogorov genügt.
Die Borelsche σ–Algebra enthält alle einelementigen Teilmengen $\{x\}$ für $x \in I\!R$ und alle halboffenen, offenen und abgeschlossenen Intervalle aus $I\!R$.

Definition 7.2 $\big(\Omega, \mathcal{S}, P\big)$ *heißt* **Wahrscheinlichkeitsraum**, *falls $\mathcal{S}$ eine σ–Algebra über Ω ist und $P : \mathcal{S} \to I\!R$ den Axiomen von Kolmogorov (3.5 – 3.7) genügt. Die Elemente von $\mathcal{S}$ heißen Ereignisse.*

(7.3) $\qquad P(A) \ \geq\ 0 \qquad \forall\, A \,\in\, \mathcal{S}$

(7.4) $\qquad P(\Omega) \ =\ 1$

(7.5) $\qquad A_i \in \mathcal{S}$ *paarw. disjunkt für abzählbar viele i* $\Longrightarrow\ P\Big(\sum\limits_i A_i\Big) \ =\ \sum\limits_i P(A_i)\ .$

Aus der Gültigkeit dieser Gleichungen folgert man leicht, analog zu den Formeln (3.8) bis (3.15), die folgenden **Rechenregeln** für den Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ mit den Ereignissen $A, B, \ldots \in \mathcal{S}$:

(7.6) $\qquad P(\emptyset) \ =\ 0$

(7.7) $\qquad P(\overline{A}) \ =\ 1 \,-\, P(A)$

(7.8) $\qquad A \subseteq B \quad\Longrightarrow\quad P(A) \,\leq\, P(B)$

(7.9) $\qquad P(A \backslash B) \ =\ P(A) \,-\, P(A \cap B)$

(7.10) $\qquad B \subseteq A \,\Longrightarrow\, P(A \backslash B) \ =\ P(A) \,-\, P(B)$

(7.11)
$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \ =\ \sum_{i=1}^{n} P(A_i) \,-\, \sum_{i_2} \sum_{i_1 < i_2} P(A_{i_1} A_{i_2}) + \sum_{i_3} \sum_{i_2 < i_3} \sum_{i_1 < i_2} P(A_{i_1} A_{i_2} A_{i_3}) \,\mp \cdots \,+\, (-1)^{n+1} P\Big(\bigcap_{i=1}^{n} A_i\Big)$$

(7.12)
$$\sum_{i=1}^{n} P(A_i) \,-\, \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i A_j) \ \leq\ P\left(\bigcup_{i=1}^{n} A_i\right) \ \leq\ \sum_{i=1}^{n} P(A_i)$$

Außerdem lassen sich aus den Axiomen die folgenden später benötigten Aussagen für monoton absteigende bzw. aufsteigende Mengenfolgen ableiten:

$$\text{(7.13)} \quad \{A_i\} \downarrow \bigcap_i A_i : \; A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \quad \Longrightarrow \quad \lim_{i\to\infty} P(A_i) = P\left(\lim_{i\to\infty} A_i\right) = P\left(\bigcap_{i=1}^{\infty} A_i\right)$$

$$\text{(7.14)} \quad \{A_i\} \uparrow \bigcup_i A_i : \; A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \quad \Longrightarrow \quad \lim_{i\to\infty} P(A_i) = P\left(\lim_{i\to\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

Ebenso überträgt man die Begriffe **Unabhängigkeit**, **bedingte Wahrscheinlichkeit** und die dafür ermittelten Rechenregeln auch auf allgemeine Wahrscheinlichkeitsräume.

Wir betrachten jetzt eine Zufallsvariable X

$$X \;:\; (\Omega, \mathcal{S}, P) \;\longrightarrow\; (I\!R, \mathcal{B}, P_X)\;,$$

deren Wahrscheinlichkeitsverteilung mit Hilfe der Gaußschen Glockenkurve als

$$P_X\big((a,b]\big) \;=\; \int\limits_a^b \varphi(x)\,dx \qquad \forall\; -\infty \;\leq\; a \;\leq\; b \;<\; \infty$$

festgelegt ist.

Da für alle $A \in \mathcal{B}$

(a) $\quad P_X(A) \;=\; \int\limits_A \varphi(x)\,dx \;\geq\; 0,$

(b) $\quad P_X(I\!R) \;=\; \int\limits_{-\infty}^{\infty} \varphi(x)\,dx \;=\; 1$ und

(c) $\quad P_X\left(\sum\limits_i A_i\right) \;=\; \int\limits_{\sum A_i} \varphi(x)\,dx \;=\; \sum\limits_i \int\limits_{A_i} \varphi(x)\,dx$

gilt, stellt P_X eine Wahrscheinlichkeit auf der σ–Algebra $\mathcal{B}$ dar.

Zur Festlegung der Wahrscheinlichkeit P auf der gesamten σ–Algebra genügt es bei **diskreten** Zufallsvariablen P auf Ω festzulegen, d.h.

$$P(\{\omega\}) \quad \forall \omega \in \Omega$$

vorzugeben.

Da im oben Beispiel wegen $P(X = a) = P_X(\{a\}) = \int_a^a \varphi(x)\, dx = 0 \quad \forall\ a \in I\!R$ jedes Elementarereignis $\{a\}$ des Bildraums $I\!R$ die Wahrscheinlichkeit 0 besitzt, ist klar, dass in einem solchen Wahrscheinlichkeitsraum ein geeignetes P nicht durch Angabe der $P_X(\{a\}) \quad \forall\ a \in I\!R$ festgelegt werden kann.

Wir betrachten daher

Satz 7.2 *Ist* $(I\!R^1, \mathcal{B}^1, P)$ *ein Wahrscheinlichkeitsraum, dann besitzt die Funktion*

$$F\ :\ I\!R^1 \longrightarrow I\!R^1 \qquad \textit{mit} \qquad F(x) \ = \ P\big((-\infty, x]\big) \qquad \forall\ x \in I\!R^1$$

die folgenden Eigenschaften:

(7.15) $\quad F(\cdot)$ *monoton nichtfallend ,*

(7.16) $\quad \lim\limits_{x\to\infty} F(x) = 1$ *und* $\lim\limits_{x\to-\infty} F(x) = 0$ *und*

(7.17) $\quad F(\cdot)$ *rechtsseitig stetig .*

F *heißt die* **Verteilungsfunktion** *des Wahrscheinlichkeitsmaßes* P.

Für das Beispiel der Wahrscheinlichkeit auf der Basis der Gaußschen Glockenkurve gilt

$$F(x) \ = \ P\big((-\infty, x]\big) \ = \ \int\limits_{-\infty}^{x} \varphi(x)\, dx \ := \ \Phi(x)\ .$$

<u>Beweis</u>(Satz 7.2):

(a) **Monotonie**:

$$\begin{aligned} x_1 \ \le\ x_2 \ &\Longrightarrow\ (-\infty, x_1] \ \subseteq\ (-\infty, x_2] \\ &\Longrightarrow\ F(x_1) \ = \ P\big((-\infty, x_1]\big) \ \le\ P\big((-\infty, x_2]\big) \ = \ F(x_2) \end{aligned}$$

(b) **Grenzwerte**:

$$\begin{aligned} \lim_{x\to\infty} F(x) \ &= \ \lim_{x\to\infty} P\big((-\infty, x]\big) \ = \ P(I\!R^1) \ = \ 1 \\ \lim_{x\to-\infty} F(x) \ &= \ \lim_{x\to-\infty} P\big((-\infty, x]\big) \ = \ P(\emptyset) \ = \ 0 \end{aligned}$$

(c) **Stetigkeit von rechts**:

$$\lim_{x\downarrow x_0} F(x) \ = \ \lim_{x\downarrow x_0} P\big((-\infty, x]\big) \ \overset{(7.13)}{=} \ P\left(\bigcap_{x>x_0} (-\infty, x]\right) \ = \ P\big((-\infty, x_0]\big) \ = \ F(x_0) \quad \forall\ x_0 \in I\!R^1$$

$\blacksquare$

Bemerkung:

I. a. ist die Verteilungsfunktion F nicht linksseitig stetig.

$$\begin{aligned}\lim_{x\uparrow x_0} F(x) &= \lim_{x\uparrow x_0} P\big((-\infty,x]\big) \overset{7.14}{=} P\left(\bigcup_{x<x_0}(-\infty,x]\right) = P\big((-\infty,x_0)\big)\\ &= P\big((-\infty,x_0]\setminus\{x_0\}\big) = P\big((-\infty,x_0]\big) - P(\{x_0\}) \neq F(x_0)\\ &\quad\text{falls}\quad P(\{x_0\}) > 0\end{aligned}$$

Ist P mit Hilfe der Gaußschen Glockenkurve definiert, ist F auch linksseitig stetig, da dann $P(\{x\}) = 0$ für alle $x \in \mathbb{R}$ gilt (s. vorige Seite).

Satz 7.3 *Ist $F : \mathbb{R}^1 \longrightarrow \mathbb{R}^1$ eine Funktion mit den Eigenschaften 7.15 – 7.17, so ist*

$$\big(\mathbb{R}^1,\mathcal{B}^1,P\big) \qquad \textit{mit} \qquad P\big((a,b]\big) = F(b) - F(a) \qquad \forall\ -\infty < a < b < +\infty$$

ein Wahrscheinlichkeitsraum. $P = P_F$ heißt dann das von F erzeugte Wahrscheinlichkeitsmaß.

Auf den Beweis soll hier verzichtet werden.

Ebenso wie bei diskreten Wahrscheinlichkeitsräumen können wir auch Zufallsvariable

$$X : \Omega \longrightarrow \mathbb{R}^1$$

auf einem allgemeinen Wahrscheinlichkeitsraum $\big(\Omega,\mathcal{S},P\big)$ benutzen. Wollen wir dabei jedoch für beliebige $A \in \mathcal{B}^1$ Ereignisse durch

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$$

festlegen, müssen wir hier zusätzlich beachten, dass zwar stets

$$X^{-1}(A) \subseteq \Omega \qquad \text{und damit} \qquad X^{-1}(A) \in \mathcal{P}(\Omega)$$

gilt. Da i.a. $\mathcal{S} \subset \mathcal{P}(\Omega)$ (echte Teilmenge) gilt, ist aber nicht gesichert, dass auch $X^{-1}(A) \in \mathcal{S}$ gilt und damit ein Ereignis ist, dessen Wahrscheinlichkeit P wir ermitteln können.

Falls $X : \Omega \longrightarrow \mathbb{R}^1$ jedoch die Eigenschaft besitzt, dass für jede Teilmenge

$$A \subseteq \mathbb{R}^1 \qquad \text{mit} \qquad A \in \mathcal{B}^1$$

auch

$$X^{-1}(A) \in \mathcal{S}$$

gilt, nennt man X $\mathcal{S}$ – $\mathcal{B}^1$ – **messbar** und bezeichnet sie als **Zufallsvariable**.

Betrachten wir speziell einen Wahrscheinlichkeitsraum $\big(\mathbb{R}^1,\mathcal{B}^1,P\big)$, so heißt $X : \mathbb{R} \to \mathbb{R}$ $\mathcal{B}^1$ – $\mathcal{B}^1$ – messbar oder kurz messbar sein.

Für einen diskreten Wahrscheinlichkeitsraum ist diese Messbarbarkeit trivialerweise wegen $\mathcal{S} = \mathcal{P}(\Omega)$ gegeben. Hinzugefügt sei, dass alle stetigen Funktionen $f : \mathbb{R} \to \mathbb{R}$ messbar sind. Im folgenden werden wir stets ohne weitere Überprüfung die Messbarkeit der angesprochenen Zufallsvariablen als gegeben hinnehmen.

7.2 Verteilungsfunktionen und Dichten

Definition 7.3 *Sei X eine reellwertige Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$. Dann heißt die Funktion*

$$F_X : I\!R \longrightarrow I\!R \qquad \textit{mit} \qquad F_X(x) := P(X \leq x) \quad \forall\, x \in I\!R$$

die **Verteilungsfunktion** *der Zufallsvariablen X.*

Mit Hilfe der Verteilungsfunktion einer Zufallsvariablen X können wir sofort die Wahrscheinlichkeit angeben, mit der die Zufallsvariable Werte im halboffenen Intervall $(a, b]$ annimmt:

$$P(X \in (a, b]) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a) \; .$$

Beispiel 7.2 (Binomialverteilung) $X \sim Bi(n, p)$

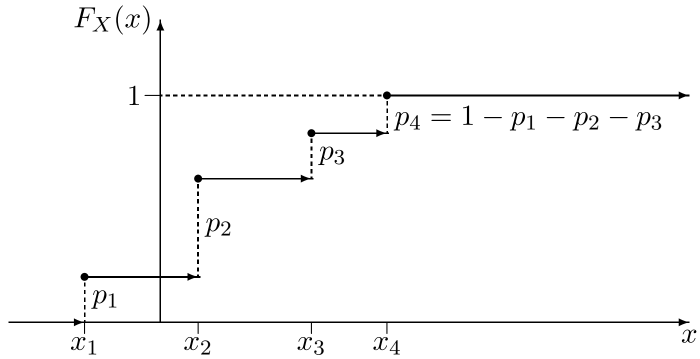
$$F_X(x) = P(X \leq x) = \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \left\{ \begin{array}{ccc} 0 & & -\infty < x < 0 \\ (1-p)^n & & 0 \leq x < 1 \\ (1-p)^n + np(1-p)^{n-1} & & 1 \leq x < 2 \\ \vdots & \textit{für} & \vdots \\ \sum\limits_{k=1}^{i} \binom{n}{k} p^k (1-p)^{n-k} & & i \leq x < i+1 \\ \vdots & & \vdots \\ 1 & & x \geq n \end{array} \right.$$

<u>Bemerkung</u>:

Die Verteilungsfunktion einer **diskreten Zufallsvariablen** X, also einer Zufallsvariablen mit diskretem Wertebereich und mit $P(X = x_k) = p_k$, ist eine **reine Treppenfunktion** mit den Sprungstellen x_k und den Sprunghöhen p_k (vergl. Abb. 7.1).

Abbildung 7.1: Verteilungsfunktion einer diskreten Zufallsvariablen



F_X springt bei x_j
um $p_j = \mathrm{P}\,(X = x_j)$
$(j = 1, \ldots, n)$

Definition 7.4 *Besitzt die Verteilungsfunktion F_X einer Zufallsvariablen X die Darstellung*

$$F_X(x) \;=\; \int\limits_{-\infty}^{x} f_X(t)\,dt \qquad \forall\; x \in \mathbb{R}^1$$

mit einer stückweise stetigen Funktion $f_X \;:\; \mathbb{R}^1 \;\longrightarrow\; [0, \infty)$ *mit*

$$\int\limits_{-\infty}^{\infty} f_X(t)\,dt \;=\; 1\;,$$

so heißt die Zufallsvariable X **stetig** *mit der* **(Wahrscheinlichkeits)-Dichte** *f_X.*

Bemerkung

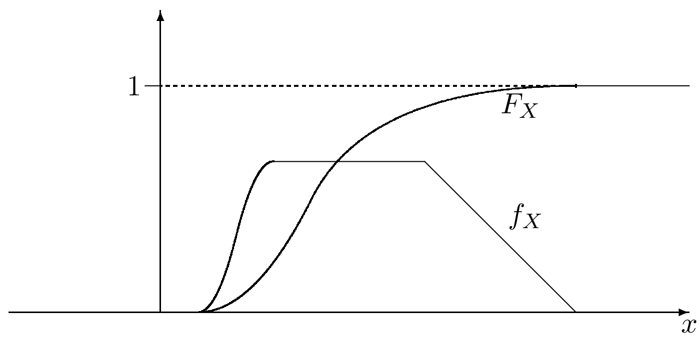
Die Verteilungsfunktion einer stetigen Zufallsvariablen besitzt keine Sprungstellen und ist stetig.

Satz 7.4 *Gilt für die stückweise stetige Funktion* $f \;:\; \mathbb{R}^1 \;\longrightarrow\; \mathbb{R}^1$

$$f(x) \geq 0 \quad \forall\; x \in \mathbb{R}^1 \quad \mathit{mit} \quad \int\limits_{-\infty}^{\infty} f(x)\,dx \;=\; 1\;,$$

so ist f Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen X.

Abbildung 7.2: Verteilungsfunktion einer stetigen Zufallsvariablen



F_X ist stetig,

$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)\, dt$

Beweis:

Zu zeigen ist lediglich, dass die Funktion

$$F(x) \;=\; \int\limits_{\infty}^{x} f(t)\, dt$$

die leicht nachprüfbaren Eigenschaften 7.15 – 7.17 besitzt.

$\blacksquare$

Bemerkung:

Ist X eine reellwertige Zufallsvariable auf $\big(\Omega, \mathcal{S}, P\big)$ und $B \in \mathcal{B}$ ein beliebiges Ereignis, so gilt

$$P(X \in B) \;:=\; \int\limits_B dF_X(x) \;=\; \begin{cases} \int_B f_X(x)\, dx & \text{falls } X \text{ stetig} \\ \sum_{x_i \in B} P(X = x_i) & \text{falls } X \text{ diskret} \end{cases}$$

Dabei nennt man $\int_B dF_X(x)$ **Lebesgue–Stieltjes Integral**.

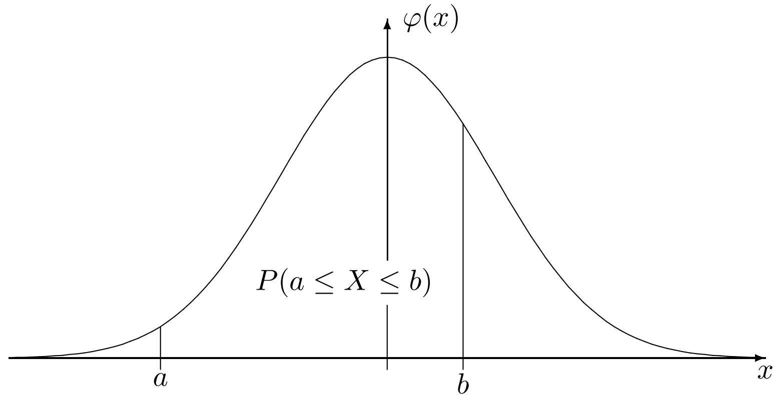
Beispiel 7.3 $\quad f_X(x) \;=\; \varphi(x) \;=\; \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$

Auf den Beweis, dass $\int_{-\infty}^{\infty} \varphi(t)\, dt \;=\; 1$ ist, soll hier verzichtet werden. Dafür wollen wir die Wahrscheinlichkeit, dass die Zufallsvariable X Werte in dem Intervall $[a, b]$ bzw. (a, b) annimmt, ermitteln.

$$P(a \le X \le b) \;=\; P(a < X < b) \;=\; P(a < X \le b) \;=\; F_X(b) \;-\; F_X(a) \;=\; \int\limits_a^b \varphi(t)\, dt$$

Leider können wir dieses Integral nicht explizit auswerten, so dass wir noch immer keinen Zahlwert für die Wahrscheinlichkeit im Beispiel 6.3 angeben können.

Abbildung 7.3: Wahrscheinlichkeit für ein Intervall



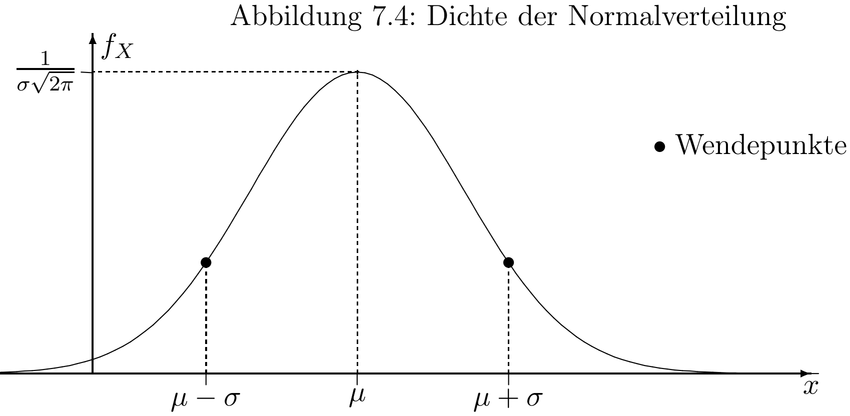
Definition 7.5 *Die stetige Zufallsvariable X besitzt eine* **Normalverteilung** *mit den Parametern μ $(-\infty < \mu < \infty)$ und σ^2 $(0 < \sigma^2 < \infty)$, falls X die Wahrscheinlichkeitsdichte*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad (x \in \mathbb{R})$$

besitzt. Man schreibt dann kurz $X \sim \mathcal{N}(\mu, \sigma^2)$. Für $\mu = 0$ und $\sigma^2 = 1$ heißt die Zufallsvariable X standardnormalverteilt.

$$X \sim \mathcal{N}(0, 1^2) \qquad \textbf{Standardnormalverteilung}$$

Abbildung 7.4: Dichte der Normalverteilung



Bezeichnungen:

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\, x^2\right\}$$ heißt Dichte der Standardnormalverteilung (Gaußsche Glockenkurve)

$$\Phi(x) := \int_{-\infty}^{x} \varphi(t)\, dt$$ heißt Verteilungsfunktion der Standardnormalverteilung

Tabelle 7.1: Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8349	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Ablesebeispiele:

$$\Phi(1.97) = 0.9756$$

$$\Phi(-0.27) = 1 - 0.6064 = 0.3936$$

Auf der vorhergehenden Seite ist $\Phi(x)$ für $x \geq 0$ tabelliert. Die übrigen Werte ergeben sich wegen der Symmetrie der Dichtefunktion φ aus

$$\begin{aligned}
\Phi(-x) \;&=\; \int_{-\infty}^{-x} \varphi(x)\,dx \;=\; 1 \;-\; \int_{-x}^{\infty} \varphi(x)\,dx \\
& \qquad\qquad y := -x \quad , \quad dy = -dx \quad , \quad \varphi(x) = \varphi(-y) = \varphi(y) \\
&=\; 1 \;-\; \int_{-\infty}^{x} \varphi(x)\,dx \;=\; 1 \;-\; \Phi(x) \qquad (x \;\geq\; 0)\;.
\end{aligned}$$

$\blacksquare$

$$\begin{aligned}
& Y \;\sim\; \mathcal{N}(\mu, \sigma^2) \;\;\Longrightarrow \\
& F_Y(x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ -\frac{1}{2} \left(\frac{t-\mu}{\sigma} \right)^2 \right\} dt \\
& \qquad\qquad \left(y \;:=\; \frac{t-\mu}{\sigma} \quad , \quad dy \;=\; \frac{1}{\sigma}\,dt \right) \\
& \qquad\;\;=\; \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}y^2}\,dy \;=\; \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(y)\,dy \;=\; \Phi\left(\frac{x-\mu}{\sigma} \right)
\end{aligned}$$

Es gilt also weiter

$$\begin{aligned}
F_{(Y-\mu)/\sigma}(x) \;&=\; P\left(\frac{Y-\mu}{\sigma} \leq x \right) \;=\; P(Y \leq \sigma x + \mu) \;=\; F_Y(\sigma x + \mu) \\
&=\; \Phi\left(\frac{\sigma x + \mu - \mu}{\sigma} \right) \;=\; \Phi(x)\;.
\end{aligned}$$

<u>Bemerkungen</u>: $\quad X \;\sim\; \mathcal{N}(\mu, \sigma^2) \quad\Longrightarrow\quad Y \;:=\; \frac{X-\mu}{\sigma} \;\sim\; \mathcal{N}(0,1)$

$$Y \;\sim\; \mathcal{N}(0, 1^2) \quad\Longrightarrow\quad X \;:=\; \sigma Y \;+\; \mu \;\sim\; \mathcal{N}(\mu, \sigma^2)$$

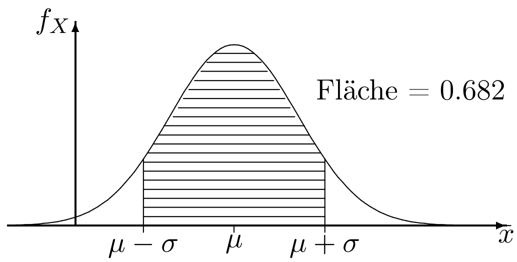
Für eine ZV $X \sim \mathcal{N}(\mu, \sigma^2)$ erhalten wir allgemein

$$\begin{aligned}
\mathrm{P}\left(a \leq X \leq b\right) \;&=\; \mathrm{P}\left(a < X < b\right) \\
&=\; \int_{a}^{b} f_X(t)\,dt \;=\; F_X(b) \;-\; F_X(a) \;=\; \Phi\left(\frac{b-\mu}{\sigma} \right) \;-\; \Phi\left(\frac{a-\mu}{\sigma} \right)
\end{aligned}$$

und für $k > 0$:

$$\begin{aligned}
\mathrm{P}\,\underbrace{\left(\mu \;-\; k\cdot\sigma \;\leq\; X \;\leq\; \mu \;+\; k\cdot\sigma\right)}_{k\cdot\sigma\text{–Bereich}} \;&=\; \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma} \right) \;-\; \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma} \right) \\
&=\; \Phi(k) \;-\; \underbrace{\Phi(-k)}_{1-\Phi(k)} \;=\; 2\cdot\Phi(k) \;-\; 1
\end{aligned}$$

$$
\begin{array}{lll}
k = 1 & : & \mathrm{P}\big(\mu - \sigma \le X \le \mu + \sigma\big) = 2 \cdot 0.8413 - 1 = 68.26\% \\
k = 2 & : & \mathrm{P}\big(\mu - 2\cdot\sigma \le X \le \mu + 2\cdot\sigma\big) = 2 \cdot 0.9772 - 1 = 95.44\% \\
k = 3 & : & \mathrm{P}\big(\mu - 3\cdot\sigma \le X \le \mu + 3\cdot\sigma\big) = 2 \cdot 0.9987 - 1 = 99.74\%
\end{array}
$$



Definition 7.6 *Die stetige Zufallsvariable X besitzt eine* **Gleichverteilung** *auf dem Intervall [0,1] , falls X die Wahrscheinlichkeitsdichte*

$$
f_X(x) = \begin{cases} 1 & , \text{ falls } 0 \le x \le 1 \\ 0 & , \text{ sonst} \end{cases}
$$

besitzt. Man schreibt dann kurz

$$
X \sim U[0,1] \quad \text{oder} \quad X \sim G[0,1]
$$

Bemerkungen:

1. Die Verteilungsfunktion der Gleichverteilung auf [0,1] lautet:

$$
F_X(x) = \int_0^x f_X(x)\, dx = \begin{cases} 0 & , \text{ falls } x \le 0 \\ x & , \text{ falls } 0 < x \le 1 \\ 1 & , \text{ falls } x > 1 \end{cases}
$$

2. Für alle $0 \le a < b \le 1$ gilt :

$$
P(X \in (a,b]) = P_X\big((a,b]\big) = F_X(b) - F_X(a) = b - a = \text{Länge}\big((a,b]\big) .
$$

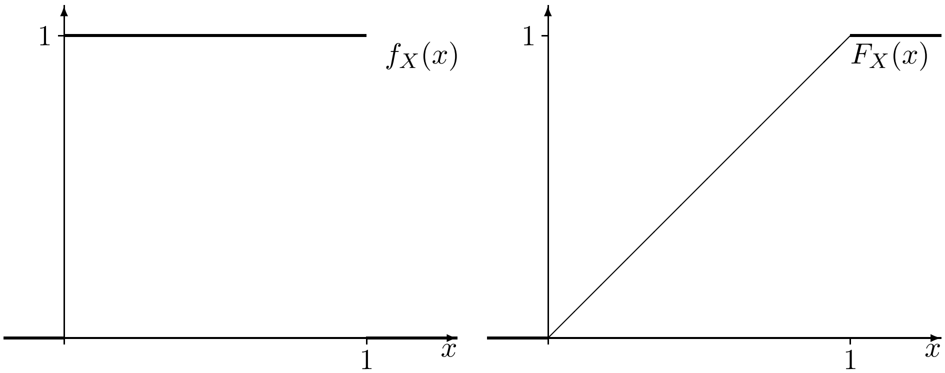
Die Wahrscheinlichkeit, dass die gleichverteilte Zufallsvariable X Werte im Intervall $(a,b]$ annimmt, ist also nur von der Länge, nicht aber der expliziten Lage des Intervalls abhängig.

3. Die obige Aussage bleibt gültig auch wenn das Intervall offen oder abgeschlossen ist.

$$
\begin{aligned}
P_X\big([a,b]\big) &= P_X\big((a,b]\big) + \underbrace{P_X(a)}_{0} = P_X\big((a,b]\big) \\
P_X\big((a,b)\big) &= P_X\big((a,b]\big) - \underbrace{P_X(b)}_{0} = P_X\big((a,b]\big)
\end{aligned}
$$

Diese Rechnung kann man für alle stetigen Zufallsvariablen X durchführen.

Abbildung 7.5: Dichte und Verteilungsfunktion der Gleichverteilung



4. Gilt $X \sim \mathrm{U}(0,1)$, dann nimmt die Zufallsvariable

$$Y := (b-a)X + a$$

für alle $a, b \in I\!R$ mit $a < b$ nur Werte auf dem Intervall $[a, b]$ an. Für die Verteilungsfunktion der Zufallsvariablen Y erhalten wir:

$$\begin{aligned} F_Y(x) &= P(Y \leq x) = \mathrm{P}\,((b-a)X + a \leq x) = \mathrm{P}\left(X \leq \frac{x-a}{b-a}\right) \\ &= F_X\left(\frac{x-a}{b-a}\right) = \int\limits_0^{\frac{x-a}{b-a}} dt \\ &\qquad \text{mit der Substitution } y := (b-a)t + a \text{ erhalten wir} \\ &= \int\limits_a^x \frac{dt}{b-a} \end{aligned}$$

Für die Dichte von Y erhalten wir nun durch Differenzieren

$$f_Y(x) = \begin{cases} \frac{1}{b-a}\,, & \text{falls } a \leq x \leq b\,, \\ 0 & \text{sonst} \end{cases}$$

und sagen Y heißt **gleichverteilt** auf dem Intervall $(a, b]$ und wir schreiben

$$Y \sim \mathrm{U}(a, b]\ .$$

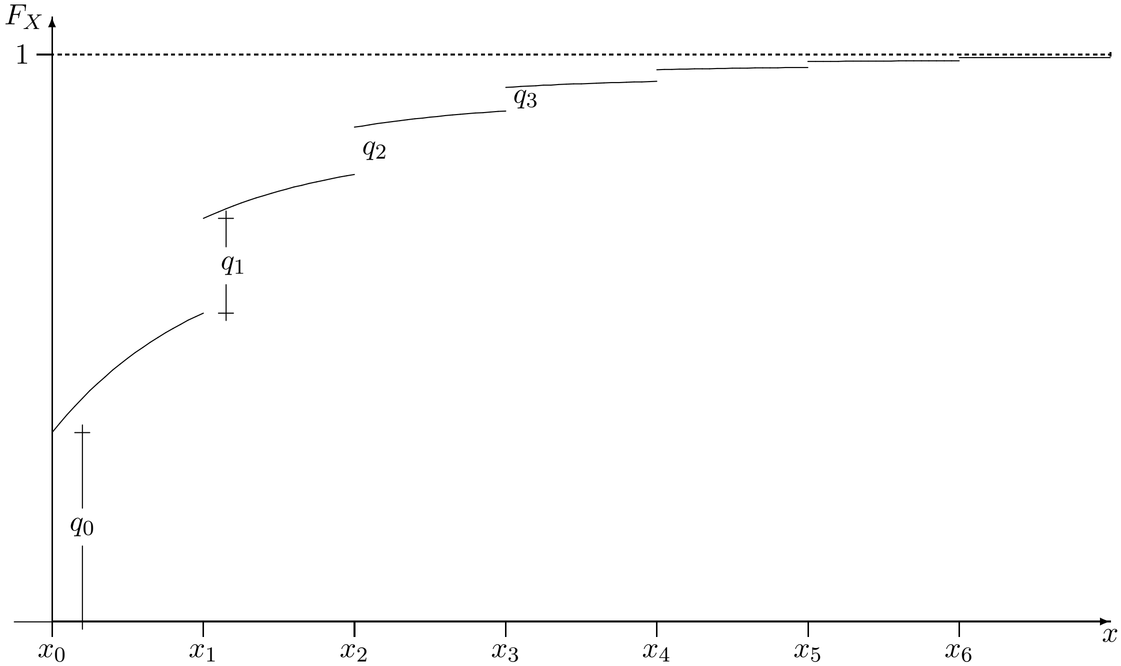
5. Gilt umgekehrt $Y \sim \mathrm{U}(a, b]$, erhalten wir durch die Transformation $X := \frac{Y-a}{b-a}$ eine Zufallsvariable X, die Werte im Intervall $(0, 1]$ annimmt. Sie besitzt die Verteilungsfunktion

$$\begin{aligned} F_X(x) &= F_{(Y-a)/(b-a)}(x) = P\left(\frac{Y-a}{b-a} \leq x\right) = P(Y \leq (b-a)x + a) \\ &= F_Y\big((b-a)x + a\big) = \int\limits_a^{(b-a)x+a} \frac{1}{b-a}\,dt = \frac{1}{b-a}\big[(b-a)x + a - a\big] = x\ . \end{aligned}$$

Es gilt also $\quad X = \frac{Y-a}{b-a} \sim \mathrm{U}(0, 1]\ .$

Beispiel 7.4 (Gemischte Zufallsvariable) *Die folgende Grafik zeigt die Verteilungsfunktion der Lebensdauer X eines elektronischen Bauteils als Funktion seiner Einsatzdauer. Zu den Zeiten x_k mit $k = 0, 1, \ldots$ wird das Teil jeweils eingeschaltet, was mit einer gewissen Wahrscheinlichkeit q_k zum sofortigen Ausfall führt.*

Abbildung 7.6: Verteilungsfunktion von Lebensdauern



Diese Verteilungsfunktion gehört weder zu einer stetigen noch zu einer diskreten Zufallsvariablen X. Zunächst besitzt die Verteilungsfunktion F_X Sprungstellen $x_0, x_1, \ldots$ mit den Sprunghöhen q_k mit

$$q_k \;=\; P(X = x_k) \;=\; F_X(x_k) \;-\; \underbrace{\lim_{\varepsilon \downarrow 0} F_X(x_k - \varepsilon)}_{:= F_X(x_k - 0)} \quad \textit{für} \quad k = 0, 1, 2, \ldots \;.$$

Für die Summe dieser Sprunghöhen gilt natürlich

$$p \;:=\; \sum_k q_k \;<\; 1 \;.$$

Verändert man diese Verteilungsfunktion, in dem an allen Sprungstellen die Sprunghöhe zu 0 gesetzt wird, so erhält man eine stetige, monoton nicht fallende Funktion mit der Eigenschaft, dass ihr Limes für x gegen ∞ gleich $1 - p$ ist. Dividieren wir diese Funktion durch $1 - p$ erhalten wir also die Verteilungsfunktion einer stetigen Zufallsvariablen X_s, den **stetigen Anteil** *der Zufallsvariablen X. Analog erhalten wir die Verteilungsfunktion einer diskreten Zufallsvariablen X_d, den* **diskreten Anteil** *von X, wenn wir jede Sprunghöhe q_k durch p dividieren und entsprechend*

$$P(X_d = x_k) \;:=\; p_k \;=\; \frac{1}{p} \cdot q_k \quad \forall\, k$$

setzen.

Insgesamt haben wir damit die obige Verteilungsfunktion als gewichtete Summe der Verteilungsfunktionen der diskreten Zufallsvariablen X_d und der stetigen Zufallsvariablen X_s folgendermaßen dargestellt:

$$F_X(x) \;=\; p \cdot F_{X_d}(x) \;+\; (1-p) \cdot F_{X_s}(x) \;.$$

Die Zufallsvariable X besitzt also Sprungstellen x_k mit den Sprunghöhen $q_k = p \cdot p_k$. Die Summe dieser Sprunghöhen ergibt dabei den Mischungsparameter p für den diskreten Anteil an der Gesamtverteilung.

Für eine solche **gemischte Zufallsvariable** *X auf einem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$, erhalten wir für beliebige Ereignisse $B \in \mathcal{S}$*

$$P(X \in B) \;=\; \int\limits_B dF_X(x) \;=\; p \int\limits_B dF_{X_d}(x) \;+\; (1-p) \int\limits_B dF_{X_s}(x)$$

$$= \; p \sum_{x_k \in B} \underbrace{P(X_d = x_k)}_{p_k} \;+\; (1-p) \int\limits_B f_{X_s}(x)\, dx \tag{7.18}$$

$$= \; \sum_{x_k \in B} \underbrace{P(X = x_k)}_{q_k} \;+\; \int\limits_B f(x)\, dx \tag{7.19}$$

mit $f(x) \;=\; \frac{d}{dx} F_X(x) \;=\; (1-p) \frac{d}{dx} F_{X_s}(x)$ *an allen Stetigkeitsstellen x von F_X. Dabei beachte man, dass f keine Wahrscheinlichkeitsdichte ist, da*

$$\int\limits_{-\infty}^{\infty} f(x)\, dx \;=\; 1 \;-\; p \;<\; 1$$

ist.

7.3 Momente

Auch bei stetigen Zufallsvariablen interessieren wir uns für ihren "mittleren Wert", den Erwartungswert. Analog zu den diskreten Zufallsvariablen erhält man allgemein:

Definition 7.7 *Ist X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$, so heißt*

$$E\{g(X)\} \;=\; \int\limits_{-\infty}^{\infty} g(x)\, dF_X(x) \;:=\; \begin{cases} \sum\limits_k g(x_k) \cdot P(X = x_k) & , \quad X \text{ diskret mit Werten } x_k, \\ \int\limits_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx & , \quad X \text{ stetig mit Dichte } f_X, \end{cases}$$

der **Erwartungswert** *von $g(X)$, falls die Reihe bzw. das Integral absolut konvergent sind.*

Bemerkung:

1. Damit können alle Momente wie Varianz, ... auch von stetigen Zufallsvariablen gebildet werden.
2. Alle abgeleiteten Rechenregeln für Momente gelten auch für stetige Zufallsvariable.
3. Für gemischte Zufallsvariable wie in Beispiel 7.4 mit der Verteilungsfunktion

$$F_X(x) \;=\; p \cdot F_{X_d}(x) \;+\; (1-p) \cdot F_{X_s}(x)$$

gilt dann

$$\begin{aligned}
\mathrm{E}\{g(X)\} &= \int_{-\infty}^{\infty} g(x)\, dF_X(x) \\
&= p \cdot \int_{-\infty}^{\infty} g(x)\, dF_{X_d}(x) \;+\; (1-p) \cdot \int_{-\infty}^{\infty} g(x)\, dF_{X_s}(x) \\
&= p \sum_k g(x_k) \cdot P(X_d = x_k) \;+\; (1-p) \int_{-\infty}^{\infty} g(x) \cdot f_{X_s}(x)\, dx && (7.20) \\
&= p\, \mathrm{E}\{g(X_d)\} \;+\; (1-p)\, \mathrm{E}\{g(X_s)\} \\
&= p \sum_k g(x_k) \cdot \underbrace{P(X_d = x_k)}_{q_k\,/\,p} + (1-p) \int_{-\infty}^{\infty} g(x) \cdot \underbrace{f_{X_s}(x)}_{f(x)/(1-p)}\, dx && (7.21) \\
&= \sum_k g(x_k) \cdot q_k \;+\; \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx
\end{aligned}$$

mit $f(x) \;=\; \frac{d}{dx} F_X(x) \;=\; (1-p) \frac{d}{dx} F_{X_s}(x)$ an allen Stetigkeitsstellen x von F_X.

Beispiel 7.5 (Normalverteilung) $X_0 \sim \mathcal{N}(0, 1^2)$

$$E\{X_0\} \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{x \cdot e^{-\frac{1}{2}x^2}}_{\textit{ungerade Funktion}}\, dx \;=\; 0\,, \quad \textit{falls} \quad \int_0^{\infty} x \cdot e^{-\frac{1}{2}x^2}\, dx \;<\; \infty$$

Wegen

$$\frac{d}{dx}\left(-e^{-\frac{1}{2}x^2}\right) \;=\; x \cdot e^{-\frac{1}{2}x^2}$$

gilt aber

$$\int_0^{\infty} x \cdot e^{-\frac{1}{2}x^2}\, dx \;=\; -e^{-\frac{1}{2}x^2}\Big|_0^{\infty} \;=\; 1 \;<\; \infty\,.$$

Etwas mühsamer rechnet man $Var\{X_0\} = 1$ *aus, so dass* X_0 *eine Standardzufallsvariable ist.*

Allgemein erhält man für $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned} E\{X\} &= \mu \quad \textit{und} \\ Var\{X\} &= \sigma^2 \end{aligned}$$

Für die zugehörige Standardzufallsvariable $X_0 = \frac{X-\mu}{\sigma}$ *erhält man*

$$\begin{aligned} F_{X_0}(y) &= P(X_0 \leq y) = P\left(\frac{X-\mu}{\sigma} \leq y\right) = P(X \leq \sigma y + \mu) \\ &= F_X(\sigma y + \mu) = \Phi\left(\frac{\sigma y + \mu - \mu}{\sigma}\right) = \Phi(y)\ , \end{aligned}$$

also

$$X_0 \sim \mathcal{N}(0, 1^2)\ .$$

Beispiel 7.6 (Gleichverteilung) $X \sim U[a,b] \iff f_X(x) = \begin{cases} \frac{1}{b-a} & \textit{falls } a \leq x \leq b \\ 0 & \textit{sonst} \end{cases}$

$$\begin{aligned} E\{X\} &= \int_a^b \frac{1}{b-a} x\, dx = \frac{1}{b-a} \left.\frac{x^2}{2}\right|_a^b \\ &= \frac{1}{b-a} \cdot \frac{b^2-a^2}{2} = \frac{a+b}{2} \end{aligned}$$

Speziell :

$$X \sim U[0,1] \implies E\{X\} = \frac{1}{2}$$

$$\begin{aligned} Var\{X\} &= E\left\{X^2\right\} - E^2\{X\} \quad \textit{mit} \\ E\left\{X^2\right\} &= \frac{1}{b-a} \int_a^b x^2\, dx = \frac{1}{b-a} \frac{b^3-a^3}{3} = \frac{b^2+ab+a^2}{3} \quad , \quad \textit{also} \\ Var\{X\} &= \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{1}{12}\left(a^2 - 2ab + b^2\right) = \frac{(b-a)^2}{12} \end{aligned}$$

Speziell :

$$X \sim U[0,1] \implies Var\{X\} = \frac{1}{12}$$

Für die Standardzufallsvariable $X_0 := \frac{X - \frac{a+b}{2}}{\frac{b-a}{\sqrt{12}}}$ *ergibt sich nach kurzer Rechnung*

$$X_0 \sim U[-\sqrt{3}\,, +\sqrt{3}] \qquad \textit{mit} \qquad Var\{X_0\} = \frac{(2\sqrt{3})^2}{12} = 1\ .$$

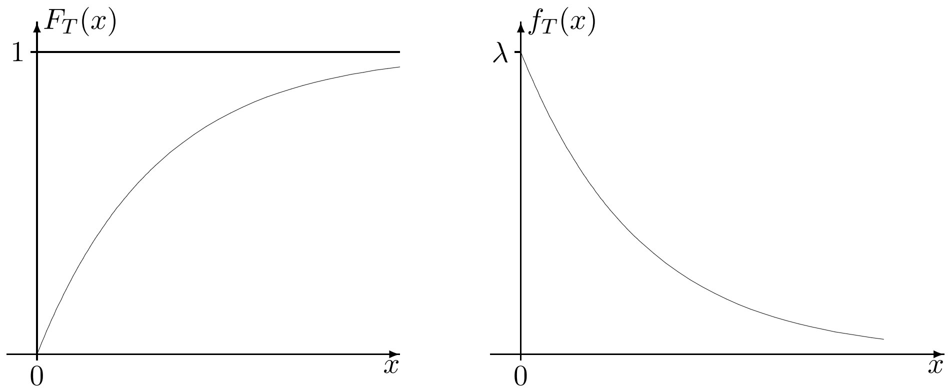
Beispiel 7.7 (Exponentialverteilung) *Die Zufallsvariable T gebe den zeitlichen Abstand der Ausfälle eines technischen Systems an. Dann gilt häufig mit einem geeigneten $\lambda > 0$ für die Verteilungsfunktion*

$$F_T(x) \;=\; P(T \leq x) \;=\; \begin{cases} 1 \;-\; e^{-\lambda x} & \textit{falls} \quad x \geq 0 \;, \\ \quad 0 & \textit{sonst} \;. \end{cases}$$

Für ihre Dichte gilt dann entsprechend

$$f_T(x) \;=\; \begin{cases} \lambda\, e^{-\lambda x} & \textit{falls} \quad x \geq 0 \;, \\ \quad 0 & \textit{sonst} \;. \end{cases}$$

Abbildung 7.7: Verteilungsfunktion und Dichte der Exponentialverteilung



Man nennt dann die Zufallsvariable T **exponentialverteilt** *mit dem Parameter λ und schreibt*

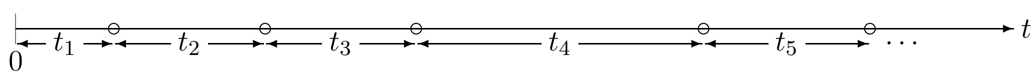
$$T \;\sim\; Exp(\lambda) \;.$$

Zählt man nun mit der Zufallsvariablen N die Anzahl der Ausfälle dieses Systems in einem Intervall der Länge t, so kann man zeigen, dass

$$N \;\sim\; Po(\lambda t)$$

gilt. Die Ausfallzeitpunkte des Systems stellen also ein homogenes Chaos *mit der Dichte λ (Häufigkeit der Ausfälle pro Zeiteinheit) dar.*

Abbildung 7.8: Ausfallzeitpunkte eines technischen Systems



Die Zufallsvariablen T_j geben dabei den Abstand der Ausfallzeitpunkte an ($j \in \mathbb{N}$).

Der Erwartungswert der Zufallsvariablen T (mittlere Dauer zwischen zwei Ausfällen) müsste dann der Kehrwert von λ, der mittleren Anzahl der Ausfälle pro Zeiteinheit sein.

Tatsächlich rechnet man mit partieller Integration nach der Formel

$$\int u\,dv \;=\; uv - \int v\,du$$

$$\begin{aligned} E\{T\} &= \int_0^\infty x\, f_T(x)\,dx \;=\; \lambda \int_0^\infty \underbrace{x}_{u} \cdot \underbrace{e^{-\lambda x}\,dx}_{dv} \\ &= \lambda \left[+\frac{x}{\lambda} e^{-\lambda x}\Big|_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x}\,dx \right] \;=\; -\underbrace{x\, e^{-\lambda x}\Big|_0^\infty}_{0} - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty \;=\; \frac{1}{\lambda} \end{aligned}$$

Ebenso ermittelt man die Varianz zu

$$Var\{T\} \;=\; \frac{1}{\lambda^2}\,.$$

Bei den diskreten Zufallsvariablen haben wir gesehen, dass die erzeugenden Funktionen ein gutes Instrument darstellen, um Momente zu berechnen, aber auch um Konvergenzaussagen machen zu können. Ein vergleichbares Instrument stellen für allgemeine Zufallsvariable die sog. **charakteristischen Funktionen** dar.

Definition 7.8 *Ist X eine Zufallsvariable, dann heißt die Funktion $\varphi_X : \mathbb{R} \longrightarrow \mathbb{C}$ mit*

$$\varphi_X(t) \;=\; E\{e^{itX}\} \;=\; \int_{-\infty}^{\infty} e^{itx}\,dF_X(x) \;=\; \begin{cases} \int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx & \text{falls } X \text{ stetig} \\ \sum_k e^{itx_k} P(X = x_k) & \text{falls } X \text{ diskret} \end{cases}$$

charakteristische Funktion *der Zufallsvariablen X.*

Bemerkungen:

1. Dabei verstehen wir unter dem Integral über eine komplexwertige Funktion

   $$g : \mathbb{R} \longrightarrow \mathbb{C} \qquad \text{mit} \qquad g(x) \;=\; \operatorname{Re} g(x) + i \cdot \operatorname{Im} g(x)$$

   die natürliche Aufspaltung in die entsprechenden Integrale über reellwertige Funktionen

   $$\int_{-\infty}^{\infty} g(x)\,dx \;=\; \int_{-\infty}^{\infty} \operatorname{Re} g(x)\,dx + i \cdot \int_{-\infty}^{\infty} \operatorname{Im} g(x)\,dx\,.$$

2. Man kann zeigen, dass die charakteristische Funktion φ_X eindeutig die Verteilungsfunktion F_X festlegt. Es gibt eine explizite Umkehrformel, auf die hier nicht eingegangen werden soll.

3. Wegen

$$e^{itx} = \cos tx + i \sin tx \qquad \Longrightarrow \qquad \left|e^{itx}\right| = \sqrt{\cos^2 tx + \sin^2 tx} = 1$$

gilt

$$|\varphi_X(t)| = \left| \int_{-\infty}^{\infty} e^{itx}\, dF_X(x) \right| \leq \int_{-\infty}^{\infty} \left|e^{itx}\right| dF_X(x) = \int_{-\infty}^{\infty} dF_X(x) = 1 \text{ (glm. beschr. in } t) \, .$$

Damit existiert $\varphi_X(t)$ für alle $t \in I\!R$.

<u>Eigenschaften und Rechenregeln</u>:

(7.22) $$|\varphi_X(t)| \leq 1 \qquad \forall\, t \in I\!R \quad \text{(nur Werte im Einheitskreis)}$$

(7.23) $$\varphi_X(0) = \mathrm{E}\left\{e^0\right\} = 1$$

(7.24)
$$\begin{aligned}
\varphi_X(-t) &= \mathrm{E}\left\{e^{-itX}\right\} = \int_{-\infty}^{\infty} (\cos tx - i \sin tx)\, dF_X(x) \\
&= \int_{-\infty}^{\infty} \cos tx\, dF_X(x) - i \int_{-\infty}^{\infty} \sin tx\, dF_X(x) \\
&= \overline{\varphi_X(t)} \qquad \text{(konjugiert komplex)} \\
\varphi_X^{(k)}(t) &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{itx}\, dF_X(x) = \int_{-\infty}^{\infty} \frac{d^k}{dt^k} e^{itx}\, dF_X(x) \\
&= \int_{-\infty}^{\infty} (ix)^k e^{itx}\, dF_X(x) \\
\Longrightarrow \quad \varphi_X^{(k)}(0) &= i^k \int_{-\infty}^{\infty} x^k\, dF_X(x) = i^k \cdot M_k \qquad \text{(falls } M_k \text{ ex.)}
\end{aligned}$$

(7.25) $$\Longrightarrow \quad M_k = i^{-k} \varphi_X^{(k)}(0)$$

(7.26) $$\varphi_{aX+b}(t) = \mathrm{E}\left\{e^{it(aX+b)}\right\} = e^{itb}\mathrm{E}\left\{e^{iatX}\right\} = e^{itb}\varphi_X(at)$$

Beispiel 7.8 $X \sim Po(\lambda)$

$$\varphi_X(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\left(\lambda e^{it}\right)^k}{k!}}_{e^{\lambda e^{it}}} = e^{\lambda(e^{it}-1)}$$

Beispiel 7.9 $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\varphi_X(t) = \int_{-\infty}^{\infty} \frac{e^{itx}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$$

$$z := \frac{x-\mu}{\sigma}, \quad x = \sigma z + \mu, \quad dx = \sigma dz$$

$$= \int_{-\infty}^{\infty} e^{it(\mu+\sigma z)} e^{-z^2/2} \frac{dz}{\sqrt{2\pi}} = \frac{e^{it\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}+it\sigma z} dz$$

$$= \frac{e^{it\mu - \frac{1}{2}t^2\sigma^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[z^2 - 2it\sigma z - t^2\sigma^2\right]\right\} dz$$

$$= \exp\left\{it\mu - \frac{1}{2}t^2\sigma^2\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-it\sigma)^2} dz = \exp\left\{it\mu - \frac{1}{2}t^2\sigma^2\right\},$$

weil das obige Integral $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-it\sigma)^2} dz = 1$ *ist, wie man mit Methoden der Funktionentheorie leicht nachweisen kann.*

Für die Standardnormalverteilung ($\mu = 0$ *und* $\sigma^2 = 1$) *gilt speziell:*

$$\varphi_{X_0}(t) = e^{-t^2/2}$$

Satz 7.5 *Seien* X *und* Y *unabhängige Zufallsvariable, dann gilt*

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) \qquad \forall\, t \in \mathbb{R}$$

Beweis:

$$\varphi_{X+Y}(t) = \mathrm{E}\left\{e^{it(X+Y)}\right\} = \mathrm{E}\left\{e^{itX} \cdot e^{itY}\right\} = \mathrm{E}\left\{e^{itX}\right\} \cdot \mathrm{E}\left\{e^{itY}\right\}$$
$$= \varphi_X(t) \cdot \varphi_Y(t)$$

■

Bemerkung: $X_1, \ldots, X_n$ unabh. $\Longrightarrow$ $\varphi_{\sum X_i}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t)$

7.4 Der Zentrale Grenzwertsatz

Wir betrachten nun eine Verallgemeinerung des Satzes von *De Moivre – Laplace* für allgemeine Zufallsvariable den wir ohne Beweis angeben werden.

Definition 7.9 (Verteilungskonvergenz) *Die Folge der Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ heißt genau dann* **verteilungskonvergent** *gegen die Zufallsvariable X, wenn*

$$\lim_{n\to\infty} F_{X_n}(x) \;=\; F_X(x)$$

an allen Stetigkeitsstellen x von F_X gilt. Wir schreiben dann

$$X_n \;\xrightarrow{\mathcal{D}}\; X \;.$$

In Verallgemeinerung des Satzes von Moivre–Laplace (Satz 6.6) betrachten wir nun eine Folge beliebiger Zufallsvariabler $\{X_n\}_{n\in\mathbb{N}}$ auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$, deren Erwartungswerte $\mathrm{E}\{X_n\}$ und Varianzen $\mathrm{Var}\{X_n\}$ für alle $n \in \mathbb{N}$ existieren. Auch hier schreiben wir

$$S_n \;=\; \sum_{i=1}^{n} X_i \qquad , \qquad Z_n \;=\; \frac{S_n \;-\; \mathrm{E}\{S_n\}}{\sqrt{\mathrm{Var}\{S_n\}}} \;.$$

Wir sagen nun, dass für die Folge $\{X_n\}$ ein **zentraler Grenzwertsatz** gilt, wenn die Folge der Wahrscheinlichkeit Z_n verteilungskonvergent gegen eine standardnormalverteilte Zufallsvariable X ist, also

$$Z_n \;\xrightarrow{\mathcal{D}}\; X \qquad \text{mit} \qquad X \;\sim\; \mathcal{N}(0, 1^2) \;.$$

gilt. Man sagt auch, dass die Folge $\{S_n\}$ **asymptotisch normalverteilt** ist.

Damit gilt näherungsweise

$$\sum_{i=1}^{n} X_i \;\sim\; \mathcal{N}\left(\sum_{i=1}^{n} \mathrm{E}\{X_i\} \;,\; \mathrm{Var}\left\{ \sum_{i=1}^{n} X_i \right\} \right) \;.$$

Diesen Sachverhalt, wollen wir mit

$$\sum_{i=1}^{n} X_i \;\approx\; \mathcal{N}\left(\sum_{i=1}^{n} \mathrm{E}\{X_i\} \;,\; \mathrm{Var}\left\{ \sum_{i=1}^{n} X_i \right\} \right)$$

bzw.

$$S_n \;=\; \sum_{i=1}^{n} X_i \;\approx\; \mathcal{N}\left(\mu \;,\; \sigma^2\right)$$

mit $\mu = \sum_{i=1}^{n} X_i$ und $\sigma^2 = \sum_{i=1}^{n} \mathrm{Var}\{X_i\}$ kennzeichnen.

Satz 7.6 *$\{X_n\}_{n\in\mathbb{N}}$ sei eine Folge identisch nach F_X verteilter, vollständig unabhängiger Zufallsvariabler mit existierendem Erwartungswert $E\{X\}$ und existierender Varianz $Var\{X\} > 0$, dann ist die Folge der Z_n asymptotisch normalverteilt, und es gilt*

$$Z_n \;:=\; \frac{S_n \,-\, nE\{X\}}{\sqrt{n\,Var\{X\}}} \quad \xrightarrow{\mathcal{D}} \quad X_0 \qquad \textit{mit} \qquad X_0 \;\sim\; \mathcal{N}(0,1^2)\;.$$

Der Beweis erfolgt mit Hilfe der charakteristischer Funktionen und soll hier nicht geführt werden.

Bemerkung:

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;\approx\; \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ falls $X_1,\ldots,X_n$ *iid* nach F_X mit $\mathrm{E}\{X\} = \mu$ und $\mathrm{Var}\{X\} = \sigma^2$.

Beispiel 7.10 *$X_1, X_2, \ldots$ seien unabhängige Zufallsvariable, die die gleiche Verteilung wie $X \;\sim\; U[a,b]$ besitzen.*

Wegen $E\{X\} \;=\; \frac{a+b}{2}$ und $Var\{X\} \;=\; \frac{1}{12}(b-a)^2$ gilt

$$Z_N \;:=\; \frac{\sum_{i=1}^n X_i \;-\; \frac{n}{2}(a+b)}{(b-a)\sqrt{n/12}} \quad \xrightarrow{\mathcal{D}} \quad X_0\;.$$

Damit ist für große n

$$F_{Z_n}(x) \;\approx\; \Phi(x) \qquad \textit{bzw.}$$

$\sum_{i=1}^{n} X_i$ *näherungsweise wie X verteilt mit $X \;\sim\; \mathcal{N}\left(\frac{n}{2}(a+b)\,,\, \frac{n}{12}(b-a)^2\right)$, also*

$$F_{\sum X_i}(x) \;\approx\; \Phi\left(\frac{x \;-\; \frac{n}{2}(a+b)}{(b-a)\sqrt{n/12}}\right)\;.$$

Bemerkung:

Speziell für $n = 12$ gilt: $X_1,\ldots,X_{12} \;\sim\; \mathrm{U}(0,1]$ (vollst. unabh.) $\Longrightarrow$ $\sum_{i=1}^{12} X_i \;-\; 6 \;\approx\; \mathcal{N}(0,1^2)$.

Beispiel 7.11 (Satz von De Moivre–Laplace(1733)) *Die Zufallsvariablen $X_1, X_2, \ldots$ seien unabhängige Bernoullivariable mit $P(X_i = 1) = p$ und $P(X_i = 0) = 1-p$.*

Damit gilt wegen $E\{X_i\} = p$ und $Var\{X_i\} = p(1-p)$

$$\frac{\sum_i X_i \;-\; np}{\sqrt{np(1-p)}} \qquad \xrightarrow{\mathcal{D}} \qquad X_0\;,$$

wobei $X_0 \sim \mathcal{N}(0,1^2)$ ist.

Da $\sum_{i=1}^{n} X_i \ \sim \ Bi(n,p)$ *gilt, können wir wegen dieser Verteilungskonvergenz die Verteilungsfunktion einer binomialverteilten Zufallsvariablen*

$$Y \ \sim \ Bi(n,p)$$

für große n folgendermaßen approximieren:

$$F_Y(k) \ = \ P(Y \le k) \ = \ P(Y < k+1) \ \approx \ \Phi\left(\frac{k \ + \ 0.5 \ - \ np}{\sqrt{np(1-p)}}\right) \ .$$

Berücksichtigt man die sog. **Faustformel von** *Pfanzagl, nach der der Stichprobenumfang hinreichend groß für diese Approximation ist, falls*

$$np \ > \ 5 \quad und \quad n(1-p) \ > \ 5$$

sind, dann gilt folgende Näherung:

$$P(k \le Y \le l) \ \approx \ \Phi\left(\frac{l - np + 0.5}{\sqrt{np(1-p}}\right) \ - \ \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1-p}}\right) \ .$$

Allgemeiner gilt der folgende Satz, dessen Beweis man z.B. bei *Feller* [2] oder *Renyi*[3] nachlesen kann.

Satz 7.7 *Seien die Zufallsvariablen* $X_1, X_2, \cdots$ *vollständig unabhängig mit existierenden Erwartungswerten* $E\{X_i\}$ *und Varianzen* $Var\{X_i\} > 0$ *für alle* $i \in I\!N$ *, dann gilt für die Folge der Zufallsvariablen*

$$Z_n \ := \ \frac{\sum_{i=1}^{n} \big(X_i \ - \ E\{X_i\}\big)}{c_n} \qquad mit \qquad c_n \ := \ \sqrt{\sum_{i=1}^{n} Var\{X_i\}}$$

unter der **Lindeberg − Bedingung**

$$\lim_{n\to\infty} \frac{1}{c_n^2} \sum_{i=1}^{n} \int\limits_{x:\ |x-E\{X_i\}|>\varepsilon\cdot c_n} (x \ - \ E\{X_i\})^2 \, dF_{X_i}(x) \ = \ 0 \qquad \forall\, \varepsilon > 0 \ ,$$

dass

$$Z_n \ \xrightarrow{\mathcal{D}} \ X_0 \qquad mit \qquad X_0 \ \sim \ \mathcal{N}(0,1^2)$$

gilt.

Bemerkung:

Die Lindebergbedingung ist erfüllt, falls die Varianzen der Zufallsvariablen X_i **gleichgradig beschränkt** sind, d.h.

$$\mathrm{Var}\,\{X_i\} \ \le \ c^2 \ < \ \infty \qquad \forall\, i \in I\!N$$

gilt.

[2] W. Feller: "An Introduction to Probability Theory and its Applications", Vol.I, Wiley & Sons, New York, 1968.

[3] A. Renyi: "Wahrscheinlichkeitstheorie", VEB Wissenschaften, Berlin, 1966.

Kapitel 8

Multivariate Zufallsvariable

Beispiel 8.1 *Um für Simulationen die erforderlichen Informationen über Engpässe innerhalb eines Computers zu ermitteln, werden eine Reihe von Daten erhoben. Zu verschiedenen zufälligen Zeitpunkten, werden Anforderungen an bestimmte Betriebsmittel, wie Arbeitsspeicher (S), Platten ($P_1, \ldots, P_r$), Drucker ($D_1, \ldots, D_m$), ... ermittelt. Alles sind Realisierungen von Zufallsvariablen, die jedoch nicht unabhängig sind. Es hat also wenig Sinn, sie einzeln zu betrachten. Gerade die Abhängigkeiten zwischen Ihnen muss der Gegenstand der Untersuchung sein. Deshalb fasst man alle Variablen zu einem Zufallsvektor (= Vektor von Zufallsvariablen) zusammen:*

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = (X_1, \ldots, X_n)^T = \left(S, P_1, \ldots, P_r, D_1, \ldots, D_m, \ldots\right)^T$$

Bezeichnet ω einen der zufällig ausgewählten Zeitpunkte, so beschreibt $\boldsymbol{X}(\omega)$ den Zustand des Computers zu diesem Zeitpunkt. $\omega \in \Omega$ ist also wie bisher ein Versuchsausgang, dem aber mit $\boldsymbol{X}$ ein Vektor reeller Zahlen zugeordnet wird. Dabei stellt

$$\boldsymbol{X} \; : \longrightarrow \; I\!R^n$$

eine Funktion von n Veränderlichen dar.

Beispiel 8.2 *Es werden Fahrzeuge beobachtet, die eine Kreuzung passieren. Die Auswahl der beobachteten Fahrzeuge stellt eine zufällige Stichprobe aus der Grundmenge aller Fahrzeuge dar, die in der betreffenden Stadt fahren. Bezeichnet ω ein spezielles Fahrzeug, so wird das Zufallsexperiment durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ beschrieben. Nun interessiert uns von jedem Fahrzeug eine Reihe unterschiedlicher Daten, wie* Anzahl der Insassen (X_1), Geschwindigkeit (X_2), Hubraumgröße (X_3) *und der* Typ des Fahrzeugs (X_4). *Es ist also nicht möglich, einen Versuchsausgang durch die Angabe des Wertes einer Zufallsvariabler X zu beschreiben. Dazu sind hier 4 Zufallsvariable $X_1, \ldots, X_4$ bzw. ein Vektor von Zufallsvariablen $Z = (X_1, \ldots, X_4)^T$ erforderlich.*

Definition 8.1 (Zufallsvektor) *Auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{S}, P)$ *seien die reellwertigen Zufallsvariablen*

$$X_1, \ldots, X_n \;:\; \Omega \;\longrightarrow\; I\!R^1$$

definiert. Die (vektorwertige) Funktion

$$\boldsymbol{Z} \;:\; \Omega \;\longrightarrow\; I\!R^n \qquad mit \qquad \boldsymbol{Z}(\omega) \;=\; \big(X_1(\omega), \ldots, X_n(\omega)\big)^T \quad \forall\, \omega \in \Omega$$

heißt (reeller) **Zufallsvektor** *bzw.* $\boldsymbol{n}$ **– dimensionale Zufallsvariable**.

Dabei interessieren wir uns für die Ereignisse $\{\omega : \boldsymbol{Z}(\omega) \in C\}$ *für alle* $C \in \mathcal{B}^n$*. Dabei ist* $\mathcal{B}^n$ *die* n **– dimensionale Borelsche** σ**-Algebra** *über dem* $I\!R^n$*. Darunter versteht man die eindeutig bestimmte kleinste* σ*-Algebra über dem* $I\!R^n$*, die alle* n*-dimensionalen Intervalle (Quader) enthält.*

Setzen wir nun

$$P_{\boldsymbol{Z}}(C) \;:=\; P\big(\{\omega \;:\; \boldsymbol{Z}(\omega) \in C\}\big) \qquad \forall\; C \in \mathcal{B}^n \;,$$

so stellt $(I\!R^n, \mathcal{B}^n, P_{\boldsymbol{Z}})$ *den von* $\boldsymbol{Z}$ *induzierten Wahrscheinlichkeitsraum dar. Für* $n = 1$ *entspricht dies völlig der Definition des vorangegangenen Kapitels.*

$P_{\boldsymbol{Z}}$ *heißt die* **Verteilung** *des Zufallsvektors* $\boldsymbol{Z}$ *bzw. die* **gemeinsame Verteilung** *der Zufallsvariablen* $X_1, \ldots, X_n$*.*

Bemerkung:

In der Definition wurde der Hinweis unterlassen, dass $\boldsymbol{Z}$ nur dann Zufallsvariable sein kann, wenn die Funktion

$$\boldsymbol{Z} \;:\; \big(\Omega, \mathcal{S}, P\big) \;\longrightarrow\; \big(I\!R^n, \mathcal{B}^n, P_{\boldsymbol{Z}}\big)$$

$\mathcal{S}$ – $\mathcal{B}^n$ – messbar ist, d.h. für alle Urbilder

$$\boldsymbol{Z}^{-1}(C) \in \mathcal{S} \qquad \forall\; C \in \mathcal{B}^n$$

gilt. Auch hier wollen wir auf diese Eigenschaft, die u.a. alle stetigen Funktionen besitzen, nicht weiter eingehen.

8.1 Multivariate Verteilungen

Definition 8.2 *Ist* $\boldsymbol{Z}$ *ein n-dimensionaler Zufallsvektor auf dem Wahrscheinlichkeitsraum* $\big(\Omega, \mathcal{S}, P\big)$*, so heißt die Funktion*

$$F_{\boldsymbol{Z}} \;:\; I\!R^n \;\longrightarrow\; I\!R$$

mit

$$F_{\boldsymbol{Z}}(x_1, \ldots, x_n) \;=\; F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \;=\; P\big(X_1 \le x_1 \wedge X_2 \le x_2 \wedge \ldots \wedge X_n \le x_n\big)$$

die **Verteilungsfunktion** *des Zufallsvektors* $\boldsymbol{Z}$*, bzw. die* **gemeinsame Verteilungsfunktion** *der Zufallsvariablen* $X_1, \ldots, X_n$.

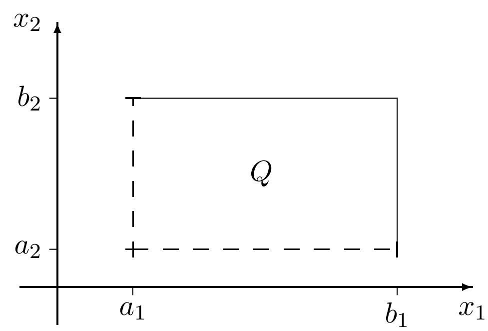
Die Wahrscheinlichkeit, dass der zweidimensionale Zufallsvektor $\boldsymbol{Z} := (X_1, X_2)^T$ *einen Wert im Quader*

$$Q := \left\{(x_1, x_2)^T \ : \ a_1 < x_1 \leq b_1 \ \textit{und} \ a_2 < x_2 \leq b_2\right\} \in \mathcal{B}^2$$

annimmt, lässt sich leicht zu

$$P(\boldsymbol{Z} \in Q) = F_{\boldsymbol{Z}}(b_1, b_2) - F_{\boldsymbol{Z}}(a_1, b_2) - F_{\boldsymbol{Z}}(b_1, a_2) + F_{\boldsymbol{Z}}(a_1, a_2)$$

ermitteln.



Existiert eine nichtnegative, integrable Funktion

$$f_{\boldsymbol{Z}} \ : \ \mathbb{R}^n \longrightarrow \mathbb{R} \qquad \textit{mit} \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{Z}}(t_1, \ldots, t_n) \ dt_n \cdots dt_1 = 1$$

und

$$F_{\boldsymbol{Z}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\boldsymbol{Z}}(t_1, \ldots, t_n) \ dt_n \cdots dt_1 \qquad \forall \ x_1, \ldots, x_n \in \mathbb{R} \ ,$$

so heißt $f_{\boldsymbol{Z}}$ *die (Wahrscheinlichkeits -)***Dichte** *des Zufallsvektors* $\boldsymbol{Z}$ *(gem. Dichte der Zufallsvariablen* $X_1, \ldots, X_n$*). Der Zufallsvektor* $\boldsymbol{Z}$ *heißt dann* **stetig***, und es gilt für alle* $C \in \mathcal{B}^n$

$$\begin{aligned} P(\boldsymbol{Z} \in C) &= \int \cdots \int_C f_{\boldsymbol{Z}}(t_1, \ldots, t_n) \ dt_n \cdots dt_1 \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} I_C(t_1, \ldots, t_n) \cdot f_{\boldsymbol{Z}}(t_1, \ldots, t_n) \ dt_n \cdots dt_1 \end{aligned}$$

mit der Indikatorvariablen $\quad I_C(t_1, \ldots, t_n) := \begin{cases} 1 & \textit{falls } (t_1, \ldots, t_n)^T \in C \\ 0 & \textit{sonst} \end{cases} \quad .$

Bemerkung:

1. Durch Vorgabe einer nichtnegativen integrablen Funktion $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ mit

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \ldots, t_n)\, dt_n \cdots dt_1 \;=\; 1$$

ist die Verteilung eines stetigen Zufallsvektors $\boldsymbol{Z}$ durch $f_{\boldsymbol{Z}} :\equiv f$ eindeutig festgelegt.

2. Besitzt die Verteilungsfunktion $F_{\boldsymbol{X}}$ eines Zufallsvektors $\boldsymbol{X}$ stetige partielle Ableitungen

$$\frac{\partial}{\partial x_i} F_{\boldsymbol{X}}(x_1, \ldots, x_n) \qquad \forall\, i = 1, \ldots, n \;,$$

so ist die Reihenfolge der Ableitungen und Integrationen gleichgültig und $\boldsymbol{X}$ ist ein stetiger Zufallsvektor mit der Dichte

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n) \;=\; \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\boldsymbol{X}}(x_1, \ldots, x_n) \qquad \forall\, \boldsymbol{x} = (x_1, \ldots, x_n)^T \;\in\; \mathbb{R}^n \;.$$

Beispiel 8.3 (Multivariate Normalverteilung) *Der n–dimensionale Zufallsvektor $\boldsymbol{X}$ heißt* **multivariat normalverteilt** *mit dem Erwartungsvektor $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$ und der symmetrischen, streng positiv definiten*[1] *Kovarianzmatrix $\Sigma = \big((\sigma_{ij})\big)_{i,j=1,\ldots,n}$, falls für ihre Dichte gilt:*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\} \qquad \forall\, \boldsymbol{x} \in \mathbb{R}^n \;.$$

Wir schreiben dann

$$\boldsymbol{X} \;\sim\; \mathcal{N}(\boldsymbol{\mu}, \Sigma) \;.$$

Speziell für $k = 1$ gilt

$$\Sigma \;=\; \big((\sigma^2)\big) \quad , \quad \boldsymbol{\mu} \;=\; (\mu) \quad , \quad \Sigma^{-1} \;=\; \frac{1}{\sigma^2} \quad , \quad |\Sigma| \;=\; \sigma^2 \;,$$

und damit

$$\boldsymbol{X} \;\sim\; \mathcal{N}(\mu, \sigma^2) \;.$$

Für $k = 2$ mit $\boldsymbol{X} \;=\; (X, Y)^T$ und $\boldsymbol{\mu} \;=\; (\mu_x, \mu_y)^T$,

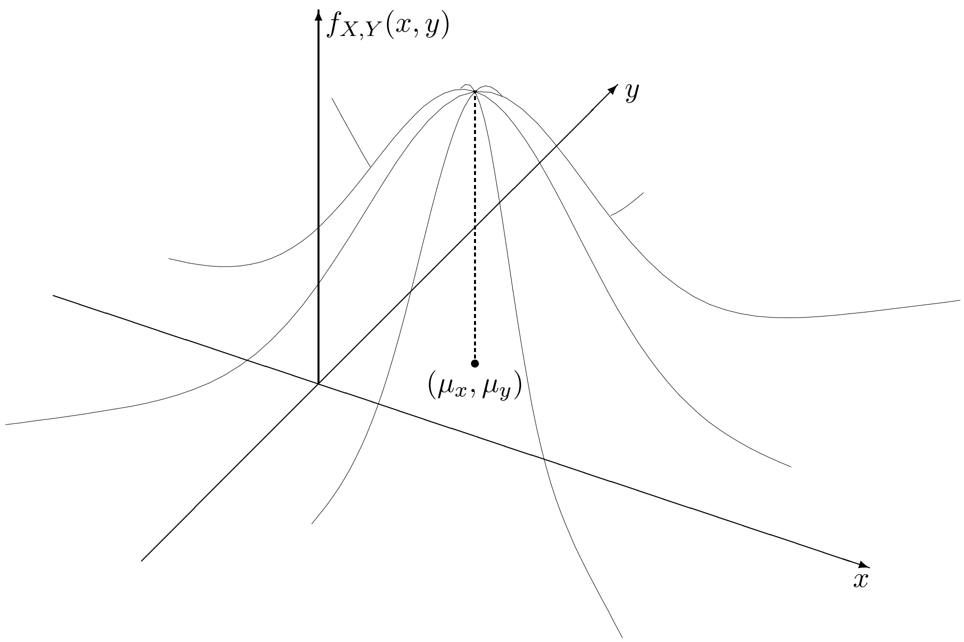
$$\Sigma \;=\; \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \qquad \textit{mit} \qquad |\sigma_{xy}| \;<\; \sigma_x^2, \sigma_y^2 \;,\; \textit{also} \quad \sigma_{xy}^2 \;< \sigma_x^2 \cdot \sigma_y^2$$

[1]Eine Matrix Σ heißt **streng positiv definit**, falls für alle $x \in \mathbb{R}^n$ mit $x \neq 0$ $\quad x^T \Sigma x > 0$ gilt. Das hat u.a. $\sigma_{ii} := \sigma_i^2 > 0$, $\sigma_i^2 > |\sigma_{ij}|\ \forall\, j \neq i$ und $\det \Sigma := |\Sigma| > 0$ zur Folge.

erhält man nach kurzer Rechnung mit dem Korrelationskoeffizienten $\rho := \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\sigma_y^2}}$ $\big(|\rho|<1\big)$ *die Dichte der* **bivariaten Normalverteilung**

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\cdot\frac{x-\mu_x}{\sigma_x}\cdot\frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

Abbildung 8.1: Dichte der bivariaten Normalverteilung



Bemerkung:

Die Verteilung des Zufallsvektors $\boldsymbol{X} = (X_1,\ldots,X_n)^T$ wird durch die gemeinsame Verteilungsfunktion $F_{X_1,\ldots,X_n}$ vollständig festgelegt. In welchem Zusammenhang stehen nun die einzelnen Verteilungsfunktionen F_{X_i} $(i=1,\ldots,n)$ mit der gemeinsamen Verteilungsfunktion $F_{\boldsymbol{X}}$?

Es gilt:

$$\begin{aligned}
F_{X_i}(x_i) &= P\Big(X_i \le x_i\Big) = P(X_i \le x_i\ ,\ X_j \in (-\infty,+\infty)\ \forall\ j \ne i\Big) \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{x_i}\cdots\int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(t_1,\ldots,t_i,\ldots,t_n)\,dt_n\cdots dt_i\cdots dt_1 \\
&= \int_{-\infty}^{x_i}\underbrace{\Big(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(t_1,\ldots,t_i,\ldots,t_n)\,dt_n\cdots dt_{i+1}dt_{i-1}\cdots dt_1\Big)}_{:=\,f_{X_i}(t_i)}\,dt_i
\end{aligned}$$

Ist $F_{\boldsymbol{X}}$ die Verteilungsfunktion des stetigen Zufallsvektors $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, so ergeben sich die **Marginaldichten** der ebenfalls stetigen Einzelvariablen X_i zu

$$f_{X_i}(t_i) \; = \; \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(t_1, \ldots, t_i, \ldots, t_n)\, dt_n \cdots dt_{i+1} dt_{i-1} \cdots dt_1 \qquad (i = 1, \ldots, n) \; .$$

Für die obige Multinormalverteilung zeigt man mit einiger Rechnung auf diese Weise, dass

$$(X_1, \ldots, X_n) \; \sim \; \mathcal{N}\big(\boldsymbol{\mu}, \Sigma\big) \qquad \Longrightarrow \qquad X_i \; \sim \; \mathcal{N}\big(\mu_i, \sigma_{ii}\big) \quad (i = 1, \ldots, n)$$

gilt.

Beachte:

Die gemeinsame Verteilung der Zufallsvariablen $X_1, \ldots . X_n$ legt alle Marginalverteilungen fest. Umgekehrt legen die Marginalverteilungen noch nicht die gemeinsame Verteilung vollständig fest. Bei der Normalverteilung legen die Marginalverteilungen beispielsweise nur die Hauptdiagonale von Σ, nicht aber die übrigen Matrixelemente fest. Wenn die Einzelvariablen X_i normalverteilt sind, muss die gemeinsame Verteilung aller X_i nicht einmal eine multivariate Normalverteilung darstellen.

Beispiel 8.4 (Bivariate diskrete Zufallsvektoren) *Das nachfolgende Beispiel behandelt zwar nur bivariate Variable, doch können sämtliche Bezeichnungen ohne Schwierigkeiten, wenn auch mit erheblichem Schreibaufwand, auf n–dimensionale diskrete Variablen übertragen werden.*

$$\boldsymbol{X} \; = \; \begin{pmatrix} X \\ Y \end{pmatrix} \qquad \textit{mit} \qquad P(X = x_i, Y = y_j) \; = \; p_{ij} \quad (i = 1, \ldots, m \; , \; j = 1, \ldots, l)$$

Tabelle 8.1: Kontingenztafel eines bivariaten Zufallsvektors

Y / X	y_1	y_2	$\cdots$	y_l	$P(X = x_i)$
x_1	p_{11}	p_{12}	$\cdots$	p_{1l}	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2l}	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{ml}	$p_{m\cdot}$
$P(Y = y_j)$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot l}$	1

Dabei sieht man leicht:

$$
\begin{aligned}
p_i &:= P(X = x_i) = P\Big(\{X = x_i\} \cap \sum_{j=1}^{l} \{Y = y_j\}\Big) \\
&= \sum_{j=1}^{l} P(X = x_i, Y = y_j) = \sum_{j=1}^{l} p_{ij} =: p_{i\cdot} \\
q_j &:= P(Y = y_j) = \cdots = \sum_{i=1}^{m} p_{ij} =: p_{\cdot j} \\
1 &= \sum_{i=1}^{m}\sum_{j=1}^{l} p_{ij} = \sum_{i=1}^{m} p_{i\cdot} = \sum_{j=1}^{l} p_{\cdot j}
\end{aligned}
$$

Auch hier bestimmen die Randverteilungen (x_i, p_i) *bzw.* (y_j, q_j) *die gemeinsame Verteilung* $\big((x_i, y_j)\,,\, p_{ij}\big)$ *nicht vollständig.*

In Definition 5.8 haben wir die **vollständige Unabhängigkeit** von Zufallsvariablen definiert und in der anschließenden Bemerkung gezeigt, dass für diskrete Zufallsvariable die Unabhängigkeit gegeben ist, wenn die gemeinsame Verteilung der Zufallsvariablen als Produktverteilung der Einzelvariablen festgelegt ist, also

$$P\left(X_1 = x_1\,,\, \ldots\,,\, X_n = x_n\right) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

für alle $x_i \in \Omega_i \ (i = 1, \ldots, n)$ gilt. Entsprechend zeigen wir nun für stetige Zufallsvariable:

Bemerkung:

Die stetigen Zufallsvariablen $X_1, \cdots, X_n$ sind vollständig unabhängig, wenn für ihre gemeinsame Dichte

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \qquad \forall\ x_1, \ldots, x_n \in \mathbb{R}$$

gilt.

Allgemein kann man für beliebige Zufallsvariable zeigen:

Satz 8.1 *Seien* $X_1, \ldots, X_n$ *Zufallsvariable auf dem Wahrscheinlichkeitsraum* $\big(\Omega, \mathcal{S}, P\big)$. *Dann sind diese Zufallsvariablen genau dann vollständig unabhängig, wenn*

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \qquad \forall\ x_1, \ldots, x_n \ \in\ \mathbb{R}$$

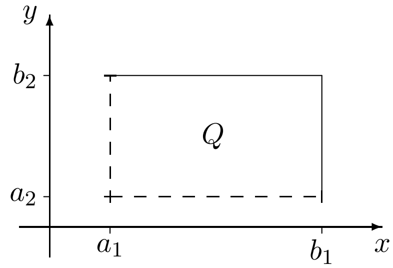
gilt.

Zusammenfassung:

1. $\boldsymbol{X} \;=\; (X_1, \ldots, X_k)^T \;:\; \Omega \;\longrightarrow\; I\!R^k$ **k-dim. Zufallsvektor**

$$F_{X_1,\ldots,X_k}(x_1, \ldots, x_k) \;=\; F_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; P(\boldsymbol{X} \le \boldsymbol{x}) \;=\; P(X_1 \le x_1, \ldots, X_k \le x_k)$$

2. Speziell für $k = 2$ erhält man $\Big(\boldsymbol{X} \;=\; (X, Y)^T \;,\; \boldsymbol{x} \;=\; (x, y)^T\Big)$



$$P(Q) \;=\; P(\boldsymbol{X} \in Q) \;=\; F_{\boldsymbol{X}}(b_1, b_2) \;-\; F_{\boldsymbol{X}}(a_1, b_2) \;-\; F_{\boldsymbol{X}}(b_1, a_2) \;+\; F_{\boldsymbol{X}}(a_1, a_2)$$

3.

$$\boldsymbol{X} \text{ stetig} \;\Longleftrightarrow\; F_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \int_{-\infty}^{x} \int_{-\infty}^{y} f_{\boldsymbol{X}}(u, v)\, dv\, du \;=\; F_{XY}(x, y)$$

$$\Longrightarrow\; P(\boldsymbol{X} \in Q) \;=\; \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{\boldsymbol{X}}(x, y)\, dy\, dx$$

$$\boldsymbol{X} \text{ diskret} \;\Longrightarrow\; P(\boldsymbol{X} \in Q) \;=\; \sum_{a_1 < x_i \le b_1} \; \sum_{a_2 < y_j \le b_2} \underbrace{P(X = x_i, Y = y_j)}_{p_{ij}}$$

4. $X_1, \;\ldots\;, X_k$ unabhängig

$$\Longleftrightarrow\; F_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; F_{X_1}(x_1) \cdots F_{X_k}(x_k) \forall\, \boldsymbol{x} \in I\!R^k$$

$$\Longleftrightarrow\; f_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \prod_{i=1}^{n} f_{X_i}(x_i) \qquad \forall\, \boldsymbol{x} \in I\!R^k \qquad \boldsymbol{X} \text{ stetig}$$

$$\Longleftrightarrow\; P(X = x_i, Y = y_j) \;=\; P(X = x_i) \cdot P(Y = y_j) \qquad \forall\, i, j \qquad X, Y \text{ diskret}$$

8.2 Bedingte Verteilungen

Im Kapitel 4 haben wir den Begriff der bedingten Wahrscheinlichkeit für Ereignisse eines Wahrscheinlichkeitsraumes $\big(\Omega, \mathcal{S}, P\big)$ eingeführt.

$$P(B > 0) \qquad \Longrightarrow \qquad P(A|B) \;=\; \frac{P(AB)}{P(B)} \qquad \forall\; A, B \;\in\; \mathcal{S}$$

Wir wollen diesen Begriff nun auf Zufallsvariable übertragen.

Sei (X,Y) ein **bivariater diskreter Zufallsvektor** mit $P(X=x_i, Y=y_j) = p_{ij}$.

$$P(X=x_i) \mid Y=y_j) = \frac{P(X=x_i \wedge Y=y_j)}{P(Y=y_j)} = \frac{p_{ij}}{p_{\cdot j}} =: p_{i|j}$$

falls $p_{\cdot j} = P(Y=y_j) > 0$ ist.

Wegen $\sum_i p_{i|j} = \frac{1}{p_{\cdot j}} \sum_i p_{ij} = 1$ stellen die Werte $p_{i|j}$ für festes j die Wahrscheinlichkeitsverteilung einer bedingten Zufallsvariablen $X|Y=y_j$ dar.

Definition 8.3 *Sei (X,Y) ein bivariater diskreter Zufallsvektor auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$ und $y_j \in Y(\Omega)$ ein Wert von Y mit $P(Y=y_j) > 0$, dann ist*

$$\big(x_i\ ,\ p_{i|j}\big)_{x_i \in X(\Omega)} \qquad \textit{mit} \qquad p_{i|j} = P(X=x_i \mid Y=y_j)$$

die **bedingte Wahrscheinlichkeitsverteilung** *der Zufallsvariablen X unter der Bedingung $Y=y_j$.*

Bemerkung:

$$\mathrm{E}\{X \mid Y=y_j\} = \sum_i x_i \cdot P(X=x_i \mid Y=y_j) = \sum_i x_i \cdot p_{i|j}$$

$$\mathrm{E}\{X\} = \sum_i x_i \cdot P(X=x_i) = \sum_i x_i \cdot p_{i\cdot}$$

$$\begin{aligned}
P(X=x_i) &= \sum_j P\left(X=x_i|Y=y_j\right) \cdot P(Y=y_j) \\
\mathrm{E}\{X\} &= \sum_i x_i \sum_j P\left(X=x_i|Y=y_j\right) \cdot P(Y=y_j) \\
&= \sum_j \left(\sum_i x_i P\left(X=x_i|Y=y_j\right) \right) P(Y=y_j) \\
&= \sum_j \mathrm{E}\{X|Y=y_j\} P(Y=y_j) =: E_Y\{\mathrm{E}\{X|Y\}\}
\end{aligned}$$

Die letzte Gleichung bezeichnet man auch als **Satz vom totalen Erwartungswert**.

Für stetige Zufallsvektoren (X,Y) hat die gleiche Begriffsbildung wegen $P(Y=y) = 0$ für alle $y \in I\!R$ keinen Sinn.

Sind jedoch die gemeinsame Dichte $f_{X,Y}(x,y)$ für alle $x \in I\!R$ und ihre Marginaldichte $f_Y(y)$ in y stetig mit $f_Y(y) > 0$, so ist

$$P\big(Y \in (y-\varepsilon, y+\varepsilon)\big) > 0 \qquad \forall\, \varepsilon > 0\ ,$$

und die folgende Definition sinnvoll.

Definition 8.4 *Sei (X, Y) ein stetiger bivariater Zufallsvektor auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$, so heißt im Falle der Existenz*

$$\lim_{\varepsilon \downarrow 0} P\big(X \leq x \mid Y \in (y-\varepsilon\ ,\ y+\varepsilon)\big) \ = \ \lim_{\varepsilon \downarrow 0} \frac{P\big(X \leq x\ ,\ y-\varepsilon < Y < y+\varepsilon\big)}{P\big(y-\varepsilon < Y < y+\varepsilon\big)} \ =: \ F_{X|Y=y}(x)$$

die **bedingte Verteilungsfunktion von *X*** *unter der Bedingung $Y = y$. Man schreibt auch*

$$F_{X|Y=y}(x) \ = \ F_{X|Y}(x|y) \ .$$

Gilt darüberhinaus

$$F_{X|Y=y}(x) \ = \ \int_{-\infty}^{x} f(t)\, dt \qquad \forall\ x \in I\!R\ ,$$

mit einer nichtnegativen integrablen Funktion

$$f\ :\ I\!R \longrightarrow I\!R \qquad \text{mit} \qquad f(t) \ := \ f_{X|Y=y}(t)\ ,$$

dann heißt $f_{X|Y=y}$ die **bedingte Dichte** *der Zufallsvariablen X unter der Bedingung $Y = y$.*

Satz 8.2 *Sei (X, Y) ein stetiger bivariater Zufallsvektor auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$ und sei die Marginaldichte f_Y stetig in y mit $f_Y(y) > 0$, dann existiert die bedingte Dichte $f_{X|Y=y}$, und es gilt*

$$f_{X|Y=y}(x) \ = \ \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad \forall\ x \in I\!R\ .$$

Beweis:

$$\begin{aligned}
F_{X|Y=y}(x) \ &= \ \lim_{\varepsilon \downarrow 0} \frac{P\big(X \leq x\ ,\ y-\varepsilon < Y < y+\varepsilon\big)}{P\big(y-\varepsilon < Y < y+\varepsilon\big)} \\
&= \ \lim_{\varepsilon \downarrow 0} \frac{\frac{1}{2\varepsilon}\left[F_{X,Y}(x\ ,\ y+\varepsilon) \ - \ F_{X,Y}(x\ ,\ y-\varepsilon)\right]}{\frac{1}{2\varepsilon}\left[F_Y(y+\varepsilon) \ - \ F_Y(y-\varepsilon)\right]} \\
&= \ \frac{\frac{\partial}{\partial y}\ F_{X,Y}(x,y)}{f_Y(y)} \ = \ \frac{1}{f_Y(y)} \cdot \frac{\partial}{\partial y} \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\, du\, dv \\
&= \ \frac{1}{f_Y(y)} \int_{-\infty}^{x} f_{X,Y}(u,y)\, du \ \overset{*}{=} \ \int_{-\infty}^{x} \underbrace{\frac{f_{X,Y}(u,y)}{f_Y(y)}}_{:=\ f_{X|Y=y}(u)}\, du
\end{aligned}$$

Bemerkung:

Damit lässt sich der bedingte Erwartungswert von X unter der Bedingung $Y = y$ folgendermaßen berechnen:

$$\mathrm{E}\{X \mid Y = y\} \ = \ \int_{-\infty}^{\infty} x \cdot f_{X|Y=y}(x)\, dx\ .$$

Bemerkung:

Analog der Formel für die totale Wahrscheinlichkeit erhalten wir aus $(*)$

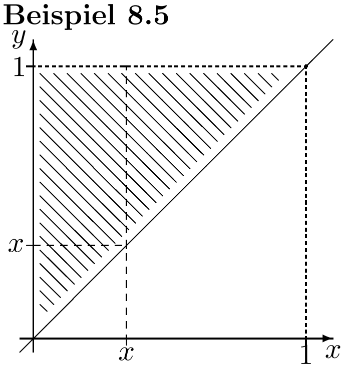
$$F_X(x) = P(X \le x) = P(X \le x\,,\, Y \in \mathbb{R}) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(u,y)\,du\,dy}_{\overset{*}{=}\ f_Y(y)\cdot F_{X|Y=y}(x)}$$

$$= \int_{-\infty}^{\infty} F_{X|Y=y}(x)\cdot f_Y(y)\,dy \ := \ \mathrm{E}_Y\left\{F_{X|Y}(x)\right\}$$

Analog gilt

$$F_Y(y) = \int_{-\infty}^{\infty} F_{Y|X=x}(y)\cdot f_X(x)\,dx \ := \ \mathrm{E}_X\left\{F_{Y|X}(y)\right\}$$

Beispiel 8.5



$$f_{X,Y}(x,y) = \begin{cases} 2 & \textit{für} \quad 0 \le x \le y < 1\ , \\ 0 & \textit{sonst}\ . \end{cases}$$

Dichte der Gleichverteiling auf dem linken oberen Einheitsdreieck

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_{x}^{1} 2\,dy = 2(1-x) \qquad (\,0 \le x < 1\,) \qquad \boldsymbol{X} \not\sim U[0,1)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{0}^{y} 2\,dx = 2y \qquad (\,0 \le y < 1\,) \qquad \boldsymbol{Y} \not\sim U[0,1)$$

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{y} \ne f_X(x) \quad (\,0 \le x \le y < 1\,) \quad \Longrightarrow \quad X|Y=y \ \sim\ U[0,y]$$

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{1-x} \ne f_Y(y) \quad (\,0 < x < y < 1\,) \quad \Longrightarrow \quad Y|X=x \ \sim\ U[x,1)$$

$$f_{X,Y}(x,y) \ne f_X(x)\cdot f_Y(y) = \frac{1}{y(1-x)} \quad (0 \le x, y < 1) \qquad \Longrightarrow \quad X,Y \ \textit{nicht unabhängig.}$$

Bemerkung:

Ist $C \in \mathcal{B}^k$ mit $\int\limits_C \cdots \int dx_k \cdots dx_1 \; := \; F(C) \; < \; \infty$, so heißt der Zufallsvektor $\boldsymbol{X}$ auf C **gleichverteilt**, falls

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \begin{cases} \frac{1}{F(C)} & \text{für } \boldsymbol{x} \in C \;, \\ 0 & \text{sonst} \end{cases}$$

gilt.

Beispiel: C = Kreis um (m_x, m_y) mit dem Radius r.

8.3 Funktionen und Momente von Zufallsvektoren

Wir wollen nun Funktionen $g_j : I\!R^n \longrightarrow I\!R \; (j = 1, \ldots, k)$ mehrerer Zufallsvariabler $X_1, \ldots, X_n$ bzw. eines Zufallsvektors $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ auf $\big(\Omega, \mathcal{S}, P\big)$betrachten. Im Falle der Messbarkeit der Funktionen g_j ist der Vektor $\big(g_1(\boldsymbol{X}), \ldots, g_k(\boldsymbol{X})\big)^T \; := \; \boldsymbol{g}(\boldsymbol{X})$ ebenfalls ein Zufallsvektor auf dem obigen Wahrscheinlichkeitsraum.

In einer Reihe von Spezialfällen haben wir bereits Momente dieses Zufallsvektors berechnet. So gilt für $k = 1$ und $g_1(X_1, \ldots, X_n) \; := \; X_1 \; + \; \cdots \; + \; X_n$

$$\mathrm{E}\left\{g_1(\boldsymbol{X})\right\} \;=\; \sum_{i=1}^{n} \mathrm{E}\left\{X_i\right\} \;.$$

Nun wollen wir aber darüberhinaus die gesamte Verteilung des neuen Zufallsvektors bestimmen.

Beispiel 8.6 $(X_1, \ldots, X_n)^T$ *sei ein Zufallsvektor auf* $\big(\Omega, \mathcal{S}, P\big)$. *Dann betrachten wir die Zufallsvariablen* M_n *und* N_n *mit*

$$\begin{aligned} M_n(\omega) \;&:=\; \max\{X_1(\omega), \ldots, X_n(\omega)\} \qquad \forall \; \omega \in \Omega \quad \textit{und} \\ N_n(\omega) \;&:=\; \min\{X_1(\omega), \ldots, X_n(\omega)\} \qquad \forall \; \omega \in \Omega \;. \end{aligned}$$

Besitzen nun $X_1, \ldots, X_n$ *die gleiche Verteilungsfunktion* F_X *und sind vollständig unabhängig,* *so gilt*

$$\begin{aligned}
F_{M_n}(z) &= P(M_n \leq z) = P\Big(\bigcap_{i=1}^{n}\{X_i \leq z\}\Big) = \prod_{i=1}^{n} P(X_i \leq z) \\
&= \prod_{i=1}^{n} F_{X_i}(z) = F_X^n(z) \qquad \forall\, z \in I\!R \qquad \textit{und} \\
F_{N_n}(z) &= P(N_n \leq z) = 1 - P(N_n > z) = 1 - P\Big(\bigcap_{i=1}^{n}\{X_i > z\}\Big) \\
&= 1 - \prod_{i=1}^{n} P(X_i > z) = 1 - \prod_{i=1}^{n}\big(1 - F_{X_i}(z)\big) \\
&= 1 - \big(1 - F_X(z)\big)^n \qquad \forall\, z \in I\!R\ .
\end{aligned}$$

Ist X darüberhinaus stetig ($X_1, \ldots, X_n$ stetige Zufallsvariable), gilt weiter

$$\begin{aligned}
f_{M_n}(z) &= n \cdot F_X^{n-1}(z) \cdot f_X(z) \qquad \forall\, z \in I\!R \qquad \textit{und} \\
f_{N_n}(z) &= n\big(1 - F_X(z)\big)^{n-1} f_X(z) \qquad \forall\, z \in I\!R\ .
\end{aligned}$$

Allgemein gilt für stetige Zufallsvektoren der folgende **Transformationssatz für Dichten**:

Satz 8.3 *Sei $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ ein stetiger Zufallsvektor auf dem Wahrscheinlichkeitsraum $\big(\Omega, \mathcal{S}, P\big)$ mit der Dichte $f_{\boldsymbol{X}} = f_{X_1,\ldots,X_n}$ und $\boldsymbol{X} : \Omega \longrightarrow R \subseteq I\!R^n$, dann gilt für jede stetige, umkehrbare Transformation*

$$\boldsymbol{g} = (g_1, \ldots, g_n)^T \ :\ R \longrightarrow S \subseteq I\!R^n$$

mit

$$\boldsymbol{Y} := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} := \boldsymbol{g}(\boldsymbol{X}) = \begin{pmatrix} g_1(X_1, \ldots, X_n) \\ \vdots \\ g_n(X_1, \ldots, X_n) \end{pmatrix}$$

und der Umkehrtransformation

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \boldsymbol{g}^{-1}(\boldsymbol{Y}) =: \boldsymbol{h}(\boldsymbol{Y}) = \begin{pmatrix} h_1(Y_1, \ldots, Y_n) \\ \vdots \\ h_n(Y_1, \ldots, Y_n) \end{pmatrix}$$

mit der Funktionaldeterminante

$$J = \det \begin{pmatrix} \frac{\partial}{\partial y_1} h_1(y_1, \ldots, y_n) & \cdots & \frac{\partial}{\partial y_n} h_1(y_1, \ldots, y_n) \\ \vdots & & \vdots \\ \frac{\partial}{\partial y_1} h_n(y_1, \ldots, y_n) & \cdots & \frac{\partial}{\partial y_n} h_n(y_1, \ldots, y_n) \end{pmatrix} \neq 0 \qquad \forall\, \boldsymbol{y} \in S$$

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |J| \cdot f_{\boldsymbol{X}}\big(\boldsymbol{g}^{-1}(\boldsymbol{y})\big)$$

bzw.

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = |J| \cdot f_{X_1,\ldots,X_n}\big(h_1(y_1, \ldots, y_n), \ldots, h_n(y_1, \ldots, y_n)\big)\ .$$

Der Beweis benutzt unmittelbar den entsprechenden Transformationssatz für Integrale aus der Analysis.

Beispiel 8.7 (k-fache Faltung) $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ *sei ein stetiger Zufallsvektor. Wir interessieren uns für die Verteilung der Summe seiner Komponenten*

$$Z \; := \; X_1 \; + \; \cdots \; + \; X_n \; .$$

Zunächst ergänzen wir diese Funktion zu einer Transformation im Sinne des Satzes 8.3.

$$\left.\begin{array}{lclllllll} Y_1 & := & X_1 & + & X_2 & + & \cdots & + & X_n \\ Y_2 & := & & & X_2 & & & & \\ \vdots & & & & & & & & \\ Y_n & := & & & & & & & X_n \end{array}\right\} \Longrightarrow \left\{\begin{array}{lclllllll} X_1 & := & Y_1 & - & Y_2 & - & \cdots & - & Y_n \\ X_2 & := & & & Y_2 & & & & \\ \vdots & & & & & & & & \\ X_n & := & & & & & & & Y_n \end{array}\right. .$$

Damit gilt für die Funktionaldeterminante

$$J \; = \; \begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix} \; = \; 1 \; \neq \; 0 \qquad \forall \, \boldsymbol{y} \, \in \, I\!R^n \; .$$

Mit Satz 8.3 erhalten wir weiter

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) \; = \; f_{X_1,\ldots,X_n}\Big(y_1 - \sum_{i=2}^{n} y_i \; , \; y_2, \ldots, y_n\Big) \cdot 1 \qquad \textit{und}$$

$$f_Z(z) \; = \; f_{Y_1}(z) \; = \; \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f_{X_1,\ldots,X_n}\Big(z - \sum_{i=2}^{n} y_i \; , \; y_2, \ldots, y_n\Big) \, dy_2 \cdots dy_n \; .$$

Speziell für unabhängige Zufallsvariable $X_1, \ldots, X_n$ *mit der gleichen Wahrscheinlichkeitsdichte* f_X *gilt*

$$\begin{aligned} f_{\sum X_i}(z) \; &= \; \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f_X\Big(z - \sum_{i=2}^{n} y_i\Big) \prod_{i=2}^{n} f_X(y_i) \, dy_2 \cdots dy_n \\ &= \; \int\limits_{-\infty}^{\infty} f_X(y_n) \int\limits_{-\infty}^{\infty} f_X(y_{n-1}) \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f_X\Big(z \; - \; \sum_{i=2}^{n} y_i\Big) \cdot f_X(y_2) \, dy_2 \cdots dy_n \end{aligned}$$

Die letzte Integration bezeichnet man als **Faltungsintegral** *oder auch als* **$\boldsymbol{k}$–fache Faltung von $\boldsymbol{f_X}$**.

Für $k = 2$ *ergibt sich*

$$f_{X+Y}(z) \; = \; \int\limits_{-\infty}^{\infty} f_X(y) f_X(z-y) \, dy \; .$$

Sind $X_1, \ldots, X_n \sim \mathcal{N}(0, 1^2)$ *und vollständig unabhängig, dann gilt*

$$\Longrightarrow \qquad f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\exp\left\{-\frac{1}{2}x_i^2\right\}}{\sqrt{2\pi}} = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\}$$

$$\Longrightarrow \qquad f_{\sum X_i}(z) = (2\pi)^{-n/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\left(z - \sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2\right]\right\} dy_n \cdots dy_2$$

$$= \cdots = \frac{1}{\sqrt{2\pi n}} \exp\left\{-\frac{1}{2}\left(\frac{z}{\sqrt{n}}\right)^2\right\}$$

$$\Longrightarrow \qquad \sum_{i=1}^{n} X_i \sim \mathcal{N}(0, n)$$

Allgemein gilt:

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad \forall\, i = 1, \ldots, n \quad (\textit{iid}^2) \qquad \Longrightarrow \qquad \sum_{i=1}^{n} X_i \sim \mathcal{N}\Big(\sum_{i=1}^{n} \mu_i\,,\ \sum_{i=1}^{n} \sigma_i^2\Big)$$

Beispiel 8.8 *Seien* $Z := X \cdot Y$ *und* X, Y *stetige Zufallsvariable mit der gemeinsamen Dichte* $f_{X,Y}$.

$$\begin{matrix} U = X \cdot Y \\ V = Y \end{matrix} \quad \Longrightarrow \quad \begin{matrix} X = \frac{U}{V} \\ Y = V \end{matrix} \quad \Longrightarrow \quad J = \begin{vmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{vmatrix} = \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v} \neq 0$$

$$\Longrightarrow \qquad f_{U,V}(u, v) = f_{X,Y}\left(\frac{u}{v}, v\right) \cdot \frac{1}{|v|}$$

$$\Longrightarrow \qquad f_U(u) = f_{X \cdot Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}\left(\frac{u}{v}, v\right) \frac{1}{|v|}\, dv$$

Beispiel 8.9 $Z := \frac{aX}{Y}$ *mit* $Y(\omega) \neq 0 \quad \forall\, \omega \in \Omega\,,\ a \neq 0$

$$\begin{matrix} U = \frac{aX}{Y} \\ V = Y \end{matrix} \quad \Longrightarrow \quad \begin{matrix} X = \frac{1}{a} \cdot UV \\ Y = V \end{matrix} \quad \Longrightarrow \quad J = \begin{vmatrix} \frac{v}{a} & \frac{u}{a} \\ 0 & 1 \end{vmatrix} = \frac{v}{a} \neq 0$$

[2] <u>i</u>dentically <u>i</u>ndependend <u>d</u>istributed: alle X_i besitzen die gleiche Verteilung und sind vollständig unabhängig

$$\Longrightarrow \qquad f_{U,V}(u,v) \;=\; f_{X,Y}\left(\frac{uv}{a}, v\right) \cdot \frac{|v|}{|a|}$$

$$\Longrightarrow \qquad f_U(u) \;=\; f_{\frac{aX}{Y}}(u) \;=\; \int\limits_{-\infty}^{\infty} f_{X,Y}\left(\frac{uv}{a}, v\right) \frac{|v|}{|a|}\, dv$$

Definition 8.5 *Ist* $\boldsymbol{X}$ *ein Zufallsvektor auf dem Wahrscheinlichkeitsraum* $\big(\Omega, \mathcal{S}, P\big)$ *und* $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ *eine* $\mathcal{B}^n$ – $\mathcal{B}$ – *messbare Funktion, dann heißt*

$$E\{g(\boldsymbol{X})\} \;=\; \int\limits_{\mathbb{R}^n} g(\boldsymbol{x})\, dF_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_n)\, dF_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

der **Erwartungswert** *von* $g(\boldsymbol{X})$.

Im Falle der Existenz heißen:

$$E\{\boldsymbol{X}\} \;=\; \begin{pmatrix} E\{X_1\} \\ \vdots \\ E\{X_n\} \end{pmatrix}$$ *der* **Erwartungsvektor** *von* $\boldsymbol{X}$,

$E\left\{X_1^{k_1} \cdots X_n^{k_n}\right\}$ **gemeinsames Moment** *der Ordnung* $k = \sum\limits_{i=1}^{n} k_i$,

$E\big\{(X_1 - E\{X_1\})^{k_1} \cdots (X_n - E\{X_n\})^{k_n}\big\}$ **gemeinsames zentriertes Moment**.

Die gemeinsamen 2. zentrierten Momente heißen:

$$E\{(X_i - E\{X_i\}) \cdot (X_j - E\{X_j\})\} \;=\; \begin{cases} Var\{X_i\} & \text{für } i = j \quad \textbf{Varianz} \text{ von } X_i \\ cov\,\{X_i, X_j\} & \text{für } i \neq j \quad \textbf{Kovarianz} \text{ von } X_i \text{ und } X_j \end{cases}$$

Die Matrix der 2. zentrierten Momente
$Cov\{\boldsymbol{X}\} \;=\; \Big(\Big(E\big\{\big(X_i - E\{X_i\}\big) \cdot \big(X_j - E\{X_j\}\big)\big\}\Big)\Big)_{i,j=1,\ldots,n}$ *heißt* **Kovarianzmatrix** *von* $\boldsymbol{X}$.

Bemerkungen:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \text{ diskret} \quad \Longrightarrow \quad \mathrm{E}\{g(X,Y)\} \;=\; \sum_i \sum_j g(x_i, y_j) \cdot P(X = x_i, Y = y_j)$$

$$\begin{pmatrix} X \\ Y \end{pmatrix} \text{ stetig} \quad \Longrightarrow \quad \mathrm{E}\{g(X,Y)\} \;=\; \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(x,y) \cdot f_{X,Y}(x,y)\, dx\, dy$$

$$\mathrm{cov}\,\{X, X\} \quad = \quad \mathrm{Var}\,\{X\}$$

$$\boldsymbol{X} \;\sim\; \mathcal{N}\big(\boldsymbol{\mu}, \Sigma\big) \qquad \Longrightarrow \quad \mathrm{E}\,\{\boldsymbol{X}\} \;=\; \cdots \;=\; \boldsymbol{\mu} \quad \text{und} \quad \mathrm{Cov}\{\boldsymbol{X}\} \;=\; \cdots \;=\; \Sigma$$

Satz 8.4 (Ungleichung von Cauchy–Schwarz) *Sind X und Y Zufallsvariable mit existierenden Varianzen und $Var\{X\}, Var\{Y\} < \infty$, dann existiert $cov\{X,Y\}$, und es gilt*

(a) $E^2\{X \cdot Y\} \;\leq\; E\{X^2\} \cdot E\{Y^2\}$

(b) Ist $E\{X^2\} > 0$, *dann gilt in (a) das Gleichheitszeichen genau dann, wenn ein* $\alpha \in I\!R$ *existiert mit* $P(\alpha X + Y = 0) = 1$. *Man sagt dann auch: Mit Wahrscheinlichkeit 1 besteht ein linearer Zusammenhang zwischen X und Y.*

Beweis:

(a) $\mathrm{E}\{X^2\} = 0$

$$\Longrightarrow \quad P(X=0)=1 \quad \left(\text{sonst: } \exists\, \varepsilon > 0 \text{ mit } P(X>\varepsilon) = p > 0 \quad \Longrightarrow \quad \mathrm{E}\{X^2\} \geq p \cdot \varepsilon^2 > 0\right)$$

$$\Longrightarrow \quad \mathrm{E}\{XY\} = 0 \quad \Longrightarrow \quad \mathrm{E}^2\{XY\} = 0 \leq \mathrm{E}\{X^2\} \cdot \mathrm{E}\{Y^2\} = 0$$

Allgem. gilt $\quad 0 \;\leq\; \mathrm{E}\{(\alpha X + Y)^2\} \;=\; \alpha^2 \mathrm{E}\{X^2\} + 2\alpha \mathrm{E}\{XY\} + \mathrm{E}\{Y^2\} \qquad \forall \alpha \in I\!R$

$\mathrm{E}\{X^2\} > 0$: $\quad$ wähle $\alpha := -\dfrac{\mathrm{E}\{XY\}}{\mathrm{E}\{X^2\}}$

$$\Longrightarrow \quad 0 \;\leq\; \frac{\mathrm{E}^2\{XY\}}{\mathrm{E}\{X^2\}} - 2\frac{\mathrm{E}^2\{XY\}}{\mathrm{E}\{X^2\}} + \mathrm{E}\{Y^2\} \quad \Longrightarrow \quad \frac{\mathrm{E}^2\{XY\}}{\mathrm{E}\{X^2\}} \;\leq\; \mathrm{E}\{Y^2\}$$

(b) $\quad$ "=" in (a) $\quad \Longleftrightarrow \quad \mathrm{E}\{(\alpha X + Y)^2\} = 0 \quad \Longleftrightarrow \quad P(\alpha X + Y = 0) = 1$

Definition 8.6 *Sind X, Y Zufallsvariable auf $(\Omega, \mathcal{S}, P)$ mit $0 < Var\{X\}, Var\{Y\} < \infty$, so heißt*

$$\rho(X,Y) := \frac{cov\{X,Y\}}{\sqrt{Var\{X\}\, Var\{Y\}}}$$ **Korrelationskoeffizient** *von X und Y.*

X und Y heißen **unkorreliert** $\quad \Longleftrightarrow \quad \rho(X,Y) = 0 \quad (\Longleftrightarrow \quad cov\{X,Y\} = 0)$

Satz 8.5 *Sind X und Y Zufallsvariable auf $(\Omega, \mathcal{S}, P)$ mit $0 < Var\{X\}, Var\{Y\} < \infty$, dann gilt*

1. *X und Y unabhängig $\Longrightarrow$ X und Y sind unkorreliert.*
2. $|\rho(X,Y)| \;\leq\; 1$

3. $\quad |\rho(X,Y)| = 1 \implies \exists\, \alpha \in I\!R \;\; \textit{mit} \;\; P(\alpha X + Y = 0) = 1$

4. $\quad (X,Y)^T$ *bivariat normalverteilt* $\implies \Big(X,Y$ *unabhängig* $\iff X,Y$ *unkorreliert*$\Big)$

Bemerkungen:

1. Zum Beweis der ersten Aussage des Satzes benötigt man die folgenden Rechenregeln:

$$\begin{aligned} \operatorname{cov}\{X,Y\} &= \mathrm{E}\{(X - \mathrm{E}\{X\}) \cdot (Y - \mathrm{E}\{Y\})\} \\ &= \mathrm{E}\{X \cdot Y - Y \cdot \mathrm{E}\{X\} - X \cdot \mathrm{E}\{Y\} + \mathrm{E}\{X\} \cdot \mathrm{E}\{Y\}\} \\ &= \mathrm{E}\{X \cdot Y\} - \mathrm{E}\{X\} \cdot \mathrm{E}\{Y\} \end{aligned} \tag{8.1}$$

Sind X und Y unabhängig, gilt (Beweis hier nur zur stetige Zufallsvariable, für diskrete Variable verläuft er analog)

$$\begin{aligned} \mathrm{E}\{X \cdot Y\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot \underbrace{f_{X,Y}(x,y)}_{=f_X(x)\cdot f_Y(y)}\, dx\, dy \\ &= \int_{-\infty}^{\infty} x f_X(x)\, dx \cdot \int_{-\infty}^{\infty} y f_Y(y)\, dy = \mathrm{E}\{X\} \cdot \mathrm{E}\{Y\} \end{aligned} \tag{8.2}$$

Damit gilt Aussage (1).
Die Umkehrung dieser Aussage gilt i.a. nicht.

2. $$\begin{aligned} \operatorname{cov}\{X+Z,Y\} &= \mathrm{E}\{(X+Z)Y\} - \mathrm{E}\{X+Z\}\ \mathrm{E}\{Y\} \\ &= \mathrm{E}\{XY + ZY\} - (EX + \mathrm{E}\{Z\}) \cdot \mathrm{E}\{Y\} \\ &= \mathrm{E}\{XY\} - \mathrm{E}\{X\}\ \mathrm{E}\{Y\} + \mathrm{E}\{ZY\} - \mathrm{E}\{Z\}\ \mathrm{E}\{Y\} \\ &= \operatorname{cov}\{X,Y\} + \operatorname{cov}\{Z,Y\} \end{aligned}$$

Damit gilt allgemein

$$\operatorname{cov}\left\{\sum_i X_i, \sum_j Y_j\right\} = \sum_i \sum_j \operatorname{cov}\{X_i, Y_j\} \tag{8.3}$$

Definition 8.7 *Ein n–dimensionaler Zufallsvektor* $\boldsymbol{X}$ *auf einem Wahrscheinlichkeitsraum* $\big(\Omega, \mathcal{S}, P\big)$ *heißt genau dann* **unkorreliert**, *wenn seine Kovarianzmatrix eine Diagonalmatrix ist, also*

$$Cov\{\boldsymbol{X}\} = diag(Var\{X_1\}, \ldots, Var\{X_n\})$$

gilt.

Index

Anhang A

Übungsaufgaben

A.1 Kombinatorik

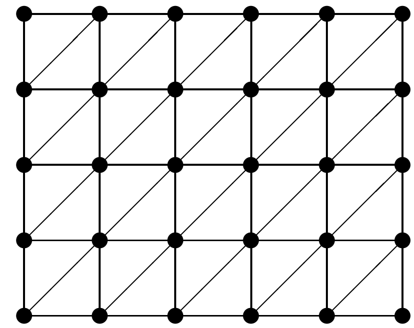
1. Wie viele verschiedene ungeordnete Proben vom Umfang r kann man mit Wiederholungen aus einer Grundmenge Ω ($|\Omega| = n$) ziehen?

2. Wie viele verschiedene Sitzordnungen gibt es für 5 Männer und 5 Frauen an einem runden Tisch, wenn nur zwischen den Geschlechtern und nicht zwischen individuellen Personen unterschieden werden soll? Dabei sollen Ordnungen, die durch Drehen des Tisches ineinander übergehen nicht unterschieden werden. Was ändert sich, wenn Männer und Frauen in "gemischter Reihe" sitzen sollen? Was ergibt sich, wenn die Personen nicht als Individuen gesehen werden sondern nur das Geschlecht interessiert?

3. Auf wie viele Arten kann man 6 Bücher nebeneinander in ein Regal stellen, wenn 3 von ihnen gleich (nicht unterscheidbar) sind?

4. Wie groß ist die Anzahl der verschiedenen Würfe mit 5 nicht unterscheidbaren Würfeln? Was ändert sich, wenn die Würfel unterschiedliche Farben haben, also unterscheidbar sind?

5. Im unten stehenden Raster darf man nur nach rechts, oben oder rechts oben laufen. Wie viele verschiedene Wege gibt es, von A nach B?



6. Wie viele Tischordnungen gibt es an einem runden Tisch mit 12 Personen, wenn 2 bestimmte Personen mindestens 3 Plätze auseinander sitzen sollen. (Wie in Aufgabe 2 , ohne Berücksichtigung der Geschlechter).

A.2 Diskrete Wahrscheinlichkeitsräume

1. In einer Kirschtorte befinden sich drei Kerne. Die Torte werde in 4 gleiche Teile zerschnitten. Man berechne die Wahrscheinlichkeiten der folgenden Ereignisse:

   (a) alle Kerne befinden sich in einem bestimmten Stück,

   (b) alle Kerne konzentrieren sich in einem Stück,

   (c) die Kerne verteilen sich in verschiedenen Stücken.

2. Zwei Freunde vereinbaren das folgende Knobelspiel: Wer beim Münzwurf zuerst fünfmal gewonnen hat, erhält 100 DM. Beim Spielstand von 4 : 3 fällt die Münze in einen Gully. Man einigt sich darauf, das Spiel nicht fortzusetzen, sondern das Geld entsprechend den Gewinnchancen zu verteilen. Der Zurückliegende will die 100 DM daraufhin im Verhältnis 4 : 3 aufteilen. Der andere ist damit nicht einverstanden. Wie sollte die Aufteilung tatsächlich erfolgen?

3. Eine Urne enthalte n Lose mit genau einem Gewinnlos. Zwei Personen ziehen nacheinander jeweils ein Los. Die zweite Person behauptet, dass für sie die Chance, den Gewinn zu erzielen, geringer sei. Stimmt das?

4. Bei einer Tombola gibt es 100 Lose, von denen 95 Nieten sind. Sie kaufen 10 Lose. Wie groß ist die Wahrscheinlichkeit, mindestens einmal etwas zu gewinnen, d.h. mindestens ein Gewinnlos zu ziehen? Mit welcher Wahrscheinlichkeit ziehen Sie genau 7 Nieten?

5. Es werde mit 3 Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit, dass genau (mindestens) 2 Würfel eine gerade Zahl zeigen? Wie groß ist die Wahrscheinlichkeit, dass die Augensumme 12 beträgt?

6. Wie oft muss man mit zwei Würfeln mindestens würfeln, damit die Wahrscheinlichkeit, mindestens einmal eine höhere Augensumme als 10 zu würfeln, mindestens 80% beträgt?

7. Wie groß ist die Wahrscheinlichkeit p dafür, dass von n Schülern einer Klasse mindestens zwei am selben Tag Geburtstag haben? Wie groß muss n mindestens sein, damit p größer als 0.5 wird?

8. Über die Informatik-StudentInnen im 5. Fachsemester liegen folgende Informationen vor:

   - 85% aller StudentInnen haben den Stochastikschein.
   - Von allen StudentInnen mit bestandenem Vordiplom haben 35% nur den Stochastikschein aber nicht den Numerikschein und 45% sowohl den Stochastik- als auch den Numerikschein.
   - Erfahrungsgemäß bestehen 30% der StudentInnen mit dem Stochastikschein nicht das Vordiplom.

   Sind die Ereignisse

   $$V \;=\; \text{"StudentIn hat Vordiplom"} \quad \text{und} \quad S \;=\; \text{"StudentIn hat Stochastikschein"}$$

   unabhängig?

9. Gegeben sei ein Parallelrechner mit $n > 1$ unterscheidbaren Prozessoren und $k \in I\!N$ voneinander unterscheidbaren Jobs.

   (a) Wie viele Möglichkeiten gibt es, diese Jobs auf die einzelnen Prozessoren zu verteilen, wenn jedem Prozessor

      i. höchstens ein Job zugeteilt werden darf (für $k \leq n$)?

      ii. auch mehrere Jobs zugeteilt werden dürfen?

   (b) Man bestimme für den Fall 9(a)ii die Wahrscheinlichkeit, dass mindestens ein Prozessor keinen Job zugeteilt bekommt.

   (c) Nun seien $n = 4$ und $k = 10$. Man bestimme die Wahrscheinlichkeit, dass der erste Prozessor 3, der zweite 4, der dritte 2 und der vierte 1 Job(s) zugeteilt bekommt.

10. Zu einer Feier bringen n Personen jeweils ein Geschenk mit. Die Geschenke werden zufällig unter den n Personen verteilt.

    (a) Man berechne die Wahrscheinlichkeit p_n dafür, dass keine Person das eigene Geschenk erhält.

    (b) Man berechne $\lim_{n\to\infty} p_n$!

    (Hinweis: Man verwende die Formel (3.14) der Vorlesung mit A_i=\{i-te Person erhält ihr eigenes Geschenk\}.)

11. Drei ideale Münzen werden unabhängig voneinander geworfen. Dabei werden folgende Ereignisse betrachtet:

    A: "Es tritt höchstens einmal Wappen auf",

    B: "Es tritt mindestens jede Seite einmal auf".

    Sind die Ereignisse A und B unabhängig?

12. Drei Gücksspielautomaten gleichen Typs sind mit unterschiedlichen Zufallsmechanismen ausgestattet worden, was aber von außen nicht erkennbar ist. Ein professioneller Spieler erzielt bei Automat A mit der Wahrscheinlichkeit $1/6$, bei Automat B mit der Wahrscheinlichkeit $1/3$ und bei Automat C immer einen Gewinn pro Spiel. Er wählt nun einen der Automaten zufällig aus, und spielt daran 1 (2, 5, n)–mal. Bei jedem Spiel gewinnt er.
Wie groß ist die Wahrscheinlichkeit, dass er am Automaten A (B, C) spielt?

13. In einer Kist mit 120 Eiern sind davon 25% faul. 6 Eier werden zufällig herausgegriffen (ohne Zurücklegen). Wie groß ist die Wahrscheinlichkeit, dabei höchstens ein faules Ei zu haben? Berechnen Sie diese Wahrscheinlichkeit
   (a) exakt unter Benutzung des entsprechenden Urnenmodells, bzw.
   (b) approximativ und erklären die geringen Unterschiede.

14. In fünf Parallelklassen gibt es folgende Schülerverteilungen:

Klasse	1	2	3	4	5
Mädchen	15	10	15	14	12
Jungen	15	16	13	11	16

   Eine Klasse wird zufällig ausgelost und daraus wieder eine Person.
   (a) Mit welcher Wahrscheinlichkeit wird ein Mädchen ausgelost.
   (b) Man weiß, dass ein Junge ausgelost wurde. Bestimmen Sie die Wahrscheinlichkeit p_i dafür, dass er aus der i-ten Klasse stammt für $i = 1, 2, 3, 4, 5$.

15. Drei ideale Münzen werden unabhängig voneinander geworfen. Dabei werden folgende Ereignisse betrachtet:
   A: ”Es tritt höchstens einmal Wappen auf”,
   B: ”Es tritt mindestens jede Seite einmal auf”.
   Sind die Ereignisse A und B unabhängig?

16. Auf dem Jahrmarkt wird ein Spiel angeboten, das aus n Bernoulliexperimenten mit der Erfolgswahrscheinlichkeit $P(E) = p$ besteht. Die Mitspieler dürfen wetten, wie oft der Erfolg E dabei eintritt. Auf welche Anzahl $k = 0, 1, \ldots, n$ sollte man setzen, um mit möglichst hoher Wahrscheinlichkeit zu gewinnen?

17. Ein Spieler bietet Ihnen das folgende Spiel an. Er hat drei Würfel, auf denen jede der angegebenen Augenzahlen zweimal aufgemalt sind:
   Würfel 1 : $1, 5, 9$
   Würfel 2 : $3, 4, 8$
   Würfel 3 : $2, 6, 7$
   Sie dürfen sich einen Würfel aussuchen. Der Spieler nimmt einen der anderen Würfel. Jeder würfelt nun einmal. Wer die höchste Augenzahl hat, gewinnt den Einsatz von DM 1. Wer besitzt die größere Gewinnwahrscheinlichkeit, wenn ihr Gegenspieler sich für seine Gewinnmaximierung optimal verhält?

18. Aus einem Teich werden k Fische gefangen, markiert und wieder ausgesetzt, um die Anzahl N aller Fische im Teich zu schätzen. Nach einer Woche werden wieder k Fische auf einmal gefangen. Unter diesen befinden sich r markierte.

    (a) Mit welcher Wahrscheinlichkeit $P(N, k, r)$ tritt dieses Ereignis ein, wenn jeder Fisch, gleichgültig, ob er markiert ist oder nicht, die gleiche Chance hat gefangen zu werden?

    (b) Für welches N ist $P(N, 50, 4)$ maximal?

19. Die Urne U_1 enthalte 5 weiße und 3 schwarze Kugeln, die Urne U_2 4 weiße und 5 schwarze und die Urne U_3 3 weiße und 7 schwarze Kugeln.

    (a) Aus jeder Urne werde zufällig eine Kugel entnommen. Wie groß ist die Wahrscheinlichkeit, dass genau zwei dieser Kugeln schwarz sind?

    (b) Eine Urne wird zufällig ausgewählt. Aus dieser Urne werden 4 Kugeln ohne Zurücklegen gezogen. Bestimmen Sie die Wahrscheinlichkeit, dass sich darunter mindestens eine Kugel jeder Farbe befindet!

    (c) Eine Kugel wird zufällig aus Urne U_1 gezogen und in Urne U_2 gelegt. Anschließend wird ebenso eine Kugel aus U_2 in U_3 gelegt. Jetzt wird eine Kugel zufällig aus U_3 gezogen. Wie groß ist die Wahrscheinlichkeit, dass diese Kugel weiß ist?

20. Es wird mit einem roten und einem schwarzen Würfel je einmal gewürfelt. A sei das Ereignis "Der rote Würfel zeigt eine gerade Augenzahl", B das Ereignis "Der schwarze Würfel zeigt eine ungerade Augenzahl" und C das Ereignis "Die Augensumme ist ungerade".
    Zeigen Sie, dass A, B und C zwar paarweise, nicht aber vollständig unabhängig sind!

21. Eine bestimmte Krankheit werde von den Erregertypen E_1, E_2 und E_3 mit den Wahrscheinlichkeiten $p_1 = 0.6$, $p_2 = 0.3$ und $p_3 = 0.1$ verursacht. Ein gleichzeitiges Auftreten verschiedener Erregertypen kann praktisch ausgeschlossen werden. Das Einnahme des Medikaments M führt mit den Wahrscheinlichkeiten 0.8 (bei E_1), 0.6 (bei E_2) bzw. 0.3 (bei E_3) zu einem Heilerfolg. Mit Wahrscheinlichkeit 0.2 tritt als Folge der Behandlung unabhängig vom Erregertyp und Heilerfolg eine Nebenwirkung auf.

    Berechnen Sie die Wahrscheinlichkeiten dafür, dass

    (a) bei der Behandlung mit M kein Heilerfolg erreicht wird,

    (b) der Erregertyp E_3 vorliegt, wenn ein Heilungserfolg durch M erreicht wird,

    (c) der Erregertyp E_1 vorliegt, wenn eine Nebenwirkung auftritt und der Erregertyp E_2 aufgrund der Krankheitsgeschichte ausgeschlossen werden kann.

22. Alle Deutschen sollen in einer Reihenuntersuchung auf eine Krankheit untersucht werden, die bei 2000 Personen etwa einmal vorkommt. Der dabei verwendete Test diagnostiziert diese Krankheit bei einem tatsächlich Erkrankten mit einer Wahrscheinlichkeit von 95%. Bei einem Gesunden liefert der Test in etwa 10 von 100 Fällen eine Fehldiagnose.
    Mit welcher Wahrscheinlichkeit liegt bei einer Person, für die der Test angezeigt hat, diese Krankheit tatsächlich vor?
    Wie groß ist die Wahrscheinlichkeit, dass bei der Reihenuntersuchung eine Erkrankung "übersehen" wird?
    Interpretieren Sie die Ergebnisse!

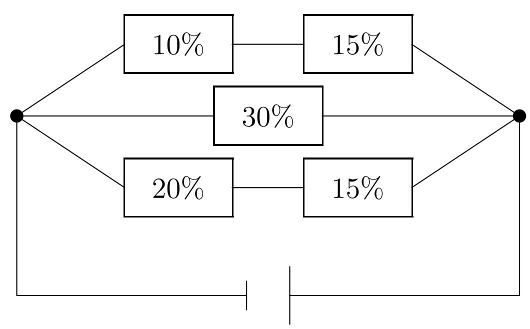
23. Ein Automobilhersteller verwendet Zündkerzen dreier verschiedener Hersteller. Dabei liefert Hersteller A 25%, Hersteller B 35% und Hersteller C 40% der verwendeten Kerzen. Erfahrungsgemäß sind von Hersteller A 5% der gelieferten Kerzen defekt, von Hersteller B 4% und von C 2%.

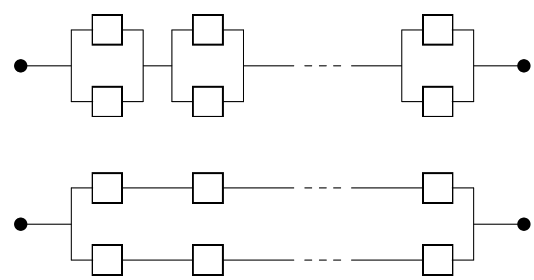
    (a) Beim Einbau einer Zündkerze wird festgestellt, dass sie defekt ist. Wie groß ist die Wahrscheinlichkeit, dass sie von Hersteller A, B bzw. C stammt?

    (b) Alle gelieferten Zündkerzen werden einer automatischen Qualitätskontrolle unterworfen, die 2% der einwandfreien und 98% der defekten Kerzen als schadhaft aussortiert. Wie groß ist die Wahrscheinlichkeit, dass eine von dieser Kontrolle freigegebene Kerze tatsächlich einwandfrei ist?

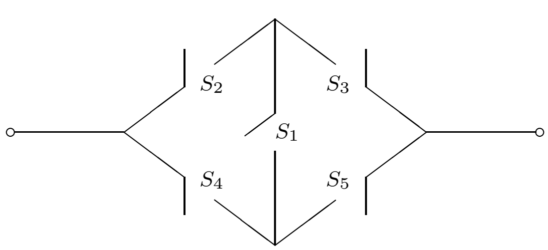
24. (a) In der linken Zeichnung ist eine Schaltung mit 5 Elementen dargestellt; die Prozentzahlen geben an, mit welcher Wahrscheinlichkeit das Element in einem gegebenen Zeitraum versagt und den Stromkreis an dieser Stelle unterbricht. Dabei sei das Versagen jedes Elementes unabhängig vom Versagen die übrigen Elemente. Mit welcher Wahrscheinlichkeit wird im betrachteten Zeitraum der Hauptstromkreis unterbrochen?

    (b) Wichtige elektrische Systeme werden dupliziert, um die Ausfallwahrscheinlichkeit niedrig zu halten. Ermitteln Sie die Ausfallwahrscheinlichkeiten der beiden Systeme in der rechten Figur, wenn jedes der $2n$ Elemente die gleiche Versagenswahrscheinlichkeit p unabhängig von den übrigen Elementen besitzt. Welches System ist sicherer?





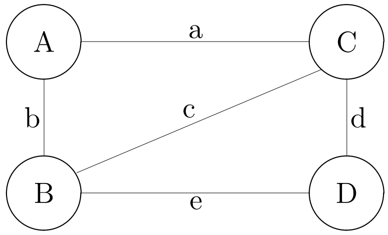
25. Eine elektrische Leitung sei durch das abgebildete Schaltwerk unterbrochen.



Im diesem Schaltwerk sei jeder der Schalter $S_1, \ldots, S_5$ mit einer Wahrscheinlichkeit von $p \in (0,1)$ geöffnet und die Stellungen der Schalter seien voneinander unabhängig. Unter der Bedingung, dass das Schaltwerk den Strom durchlässt, berechne man die Wahrscheinlichkeit, dass der Schalter S_1 geöffnet ist.

Ermitteln Sie zum Schluss diese Wahrscheinlichkeit explizit für $p = 1/2$.

26. Zwischen den Modulen A, B, C und D bestehen die Kabelverbindungen a,. . . ,e:



Jede dieser Verbindungen sei unabhängig von den übrigen mit einer Wahrscheinlichkeit von $0 < q < 1$ gestört. Mit welcher Wahrscheinlichkeit lässt sich eine Nachricht von C nach D störungsfrei übermitteln, falls $q = 0.2$ ist?

27. Ein Student muss einen bestimmten Schein erwerben. Die erforderliche Prüfung kann er ohne jede Vorbereitung mit der Wahrscheinlichkeit p bestehen. Wie viele Prüfungsteilnahmen muss er einplanen, damit er mit einer Wahrscheinlichkeit von mindestens $0,8$ diese Prüfung mindestens einmal besteht? (Geben Sie auch den Zahlwert für $p = 0,2$ an!). Dabei wird angenommen, dass die Ergebnisse aller Prüfungsteilnahmen unabhängig sind und p immer gleich bleibt.

28. Ein Mann spielt Roulette und hat für höchstens vier Spiele Zeit. Er beginnt mit einem Chip und hört mit Spielen auf, falls er keinen Chip mehr hat. In jedem Einzelspiel gewinnt er mit Wahrscheinlichkeit $1/3$ zwei Chips oder verliert mit Wahrscheinlichkeit $2/3$ einen Chip (Setzen von einem Chip auf "oberes", "mittleres" oder "unteres Drittel" ohne Berücksichtigung der "Null"). Die Zufallsvariable X sei die Anzahl der Chips nach Ende der Spielserie (4 oder weniger Spiele). Bestimmen Sie alle möglichen Spielverläufe (Diagramm wie in Abb. 4.1 der Vorlesung) und die Wahrscheinlichkeitsverteilung von X. Mit welcher Wahrscheinlichkeit verliert der Spieler alles? Welche Gewinnerwartung besitzt er? Was ändert sich, wenn er statt dessen in allen Runden auf die "einfachen Chancen" setzt, also pro Runde jeweils mit Wahrscheinlichkeit $1/2$ einen Chip gewinnt oder verliert?

29. Ein Punkt befindet sich zum Zeitpunkt 0 im Ursprung der Zahlengeraden. Innerhalb eines festen Zeitintervalls bewegt er sich unabhängig von seiner derzeitigen Position mit Wahrscheinlichkeit von je $1/2$ um eine Einheit nach rechts oder links. Die Zufallsvariable X_n bezeichne die Position des Punktes nach n Zeitintervallen. Bestimmen Sie die Wahrscheinlichkeitsverteilung von X_5 und $|X_6|$. Bestimmen Sie den erwarteten Abstand des Punktes vom Nullpunkt nach 6 Schritten ($\mathrm{E}\{|X_6|\}$).

30. Einem Gefäß, das n Kugeln unbekannter Farbe enthält, werde auf gut Glück eine Kugel entnommen. Sie sei weiß und werde in das Gefäß zurückgelegt. Wie groß ist die Wahrscheinlichkeit dafür, dass das Gefäß lauter weiße Kugeln enthält, wenn vor Ziehen der ersten Kugel alle Annahmen über die Anzahl der weißen Kugeln gleichwahrscheinlich waren? Betrachten Sie zur Kontrolle den Sonderfall $n = 1$.

31. Die 32 Karten eines Kartenspiels werden folgendermaßen bewertet:
    4 Asse mit je 11 Punkten,
    je 4 Könige, Damen, Buben und Zehnen mit je 10 Punkten und
    je 4 Neunen, Achten und Siebenen mit 0 Punkten.

    Zwei Karten werden zufällig ausgewählt und verdeckt in den "Skat" gelegt (Ziehen ohne

Zurücklegen!). Die Zufallsvariable X sei die Summe der Bewertungszahlen beider Karten im "Skat".

(a) Geben Sie den Wertevorrat von X an!

(b) Bestimmen Sie die Wahrscheinlichkeitsverteilung von X !

(c) Alle Mitspieler können den "Skat ersteigern". Der Höchstbietende zahlt sein "Gebot" als Einsatz und erhält nach Aufdecken des Skats die Summe der Bewertungspunkte in DM ausgezahlt.

   i. Bis zu welchem Betrag sollten Sie mitbieten, wenn die erwartete Auszahlung höher als der Einsatz sein soll ?

   ii. Bis zu welchem Betrag sollten Sie mitbieten, wenn die Wahrscheinlichkeit, dass der Gewinn höher als der Einsatz ist, größer als 1/2 sein soll?

32. Ihnen wird das folgende Spiel angeboten. Sie müssen aus einem verdeckt liegenden Skatspiel (32 Karten, darunter 12 "Bilder" und 4 "Asse") eine Karte ziehen. Ziehen Sie ein "As", erhalten Sie 10 DM. Ziehen Sie ein "Bild", müssen Sie mit einem "fairen" Würfel (Augenzahlen: 1 – 6) würfeln und erhalten die gewürfelte Augenzahl als Gewinn in DM. In allen anderen Fällen erhalten Sie nichts. Welchen Gewinn haben Sie im Mittel zu erwarten? Lohnt sich das Spiel für Sie, wenn der Einsatz 3 DM beträgt?

33. Die Zufallsvariable X sei binomialverteilt mit $\mathrm{E}\{X\} = 6$ und $\mathrm{Var}\{X\} = 4$. Man bestimme die Parameter dieser Binomialverteilung.

34. Man betrachte eine Folge unabhängiger Bernoulli-Experimente mit der Erfolgswahrscheinlichkeit p. Die Zufallsvariable X_r gebe die Anzahl der Misserfolg bis zum r-ten Erfolg an. Hierbei ist $r = 1, 2 \ldots$ ein fester Parameter.

(a) Man zeige, dass für die erzeugende Funktion $g_{X_r}(t)$ gilt:

$$g_{X_r}(t) \;=\; \left(\frac{p}{1-t(1-p)}\right)^r \quad |t| \;<\; \frac{1}{1-p}\;.$$

(Hinweis: X_r ist die Summe der Wartezeiten zwischen den einzelnen Erfolgen.)

(b) Man berechne den Erwartungswert und die Varianz von X_r.

(c) Geben Sie die Wahrscheinlichkeitsverteilung von X_r an.

35. Sie würfeln mit einem Tetraeder, dessen 4 Seiten mit 1, 2, 3 bzw. 4 beschriftet sind. Alle Seiten besitzen die gleiche Wahrscheinlichkeit, unten zu liegen (Augenzahl = 1, 2, 3, 4). Ihr Konkurrent würfelt mit einem üblichen korrekten Würfel (Augenzahlen 1, …, 6). Hat er eine höhere Augenzahl als Sie, zahlen Sie Ihm die Differenz der Augenzahlen in DM aus. Anderenfalls dürfen Sie erneut mit dem Tetraeder würfeln und bekommen von ihm die Differenz zwischen der Summe Ihrer beiden Augenzahlen und seiner zuvor gewürfelten Augenzahl in DM ausgezahlt.
Geben Sie den Wertevorrat und den Erwartungswert der Zufallsvariablen X (= Ihr Gewinn in DM) an. Mit welcher Wahrscheinlichkeit ist Ihr Gewinn positiv? Begründen Sie, ob sich das Spiel für Sie "lohnt" (Ist es "fair"?).

36. X sei eine nichtnegative, ganzzahlige Zufallsvariable. Zeigen Sie

$$\mathrm{E}\{X\} \; = \; \sum_{k=1}^{\infty} P(X \geq k) \; .$$

Verwenden Sie die Formel, um den Erwartungswert einer geometrischen Verteilung zu ermitteln.

37. Die Zufallsvariable X sei gleichverteilt auf $\{1, \dots n\}$, d.h. es gilt $P(X = k) = 1/n$ für $k = 1, \dots, n$.

    (a) Berechnen Sie die erzeugende Funktion $g_X(t)$.

    (b) Ermitteln Sie den Erwartungswert und die Varianz von X mit Hilfe der erzeugenden Funktion.

38. In einer Zentraleinheit eines Rechners werden zu einem bestimmten Zeitpunkt 7 Jobs mit einer Bearbeitungsdauer von je 160 Sekunden, 5 mit je 220 und 8 mit je 280 Sekunden erwartet. Die Reihenfolge des Eintreffen der Jobs ist zufällig. Die Bearbeitungsdauer für den ersten Job kann also als eine Zufallsvariable X aufgefasst werden.

    (a) Geben Sie die Wahrscheinlichkeitsverteilung der Zufallsvariablen X an.

    (b) Bestimmen Sie den Erwartungswert der Zufallsvariablen X.

    (c) Wie groß ist die Varianz der Zufallsvariablen $Y := 2 \cdot X \; - \; 400$.

39. Eine diskrete Zufallsvariable X heißt **logarithmisch** verteilt mit dem Parameter $\vartheta \in (0, 1)$, falls gilt:

$$X \sim \mathrm{LOG}(\vartheta) \quad \Longleftrightarrow \quad P(X = k) \; = \; \begin{cases} \alpha \, \frac{\vartheta^k}{k} & \text{für } k = 1, 2, \dots \\ 0 & \text{sonst} \end{cases}$$

Bestimmen Sie α und $\mathrm{E}\{X\}$.

40. Es seien X und Y zwei unabhängige logarithmisch verteilte Zufallsvariable (vergl. Aufgabe 39).

    (a) Man berechne $P(X = Y)$ und $P(X > Y)$.

    (b) Man bestimme den Erwartungswert von $Z := \min\{X, Y\}$. (Hinweis: Es ist günstig, von den Wahrscheinlichkeiten $P(Z \geq k)$ auszugehen.)

41. Zwei Kinder K_1 und K_2 werfen abwechselnd Ringe auf ein Ziel. Das Kind, das zuerst trifft, gewinnt. Die Trefferwahrscheinlichkeiten der Kinder seien unabhängig voneinander p_1 bzw. p_2 für jeden Wurf $(0 < p_1, p_2 < 1)$. Mit welcher Wahrscheinlichkeit gewinnt K_1,

    (a) falls es als erstes wirft,

    (b) falls mit dem Wurf einer idealen Münze entschieden wird, wer als erstes wirft.

42. Man betrachte ein n-faches Bernoulli–Experiment mit Erfolgswahrscheinlichkeit $p = 2/3$. Wie groß muss n mindestens sein, damit die relative Häufigkeit der Erfolge mit einer Wahrscheinlichkeit von mindestens 0.95 betragsmäßig um nicht mehr als 0.01 von p abweicht? (Hinweis: Man verwende die Chebychev–Ungleichung.)

43. An einer Straße werden Kraftfahrzeuge gezählt. Dabei wird festgestellt, dass für beide Richtungen die Anzahl der pro Minute passierende Fahrzeuge poissonverteilt mit den Parametern $\lambda = 1$ (nach links) und $\lambda = 2$ (nach rechts) ist. Beide Richtungen können als stochastisch unabhängig angesehen werden.

    (a) Wie groß ist die Wahrscheinlichkeit, dass in einer Minute in jeder Richtung höchstens ein Fahrzeug vorbeifährt?

    (b) Mit welcher Wahrscheinlichkeit werden insgesamt, d.h. in beiden Richtungen zusammen, mindestens 2 Fahrzeuge in einer Minute beobachtet?

44. Die Zufallsvariable X sei poissonverteilt mit dem Parameter $\lambda = 4$. Bestimmen Sie

    (a) $P(X < 4)$ und

    (b) $P(X = 1 \mid X > 0)$.

    (c) Ermitteln Sie $P(|X - \mathrm{E}\{X\}| > 4)$ auf 4 Dezimalen genau und

    (d) vergleichen Sie das Ergebnis mit der zugehörigen Chebychev-Abschätzung.

45. Die Zufallsvariable N habe eine Poissonverteilung mit Parameter $\lambda > 0$. Für jedes $n = 0, 1, \ldots$ seien $X_1, X_2, \ldots, X_n$ Bi(1,p)-verteilte Bernoullivariable, so dass $N, X_1, \ldots, X_n$ vollständig unabhängig sind. Durch den Ansatz

$$Y := \begin{cases} 0 & \text{falls} \quad N = 0\,, \\ \sum\limits_{i=1}^{n} X_i & \text{falls} \quad N = n \end{cases}$$

wird eine Zufallsvariable Y definiert. Man berechne die erzeugende Funktion von Y und identifiziere die zugehörige Verteilung.

46. Die Anzahl der an einem Tag in einer Telefonzentrale vermittelten Telefongespräche sei eine mit dem Parameter $\lambda > 0$ poissonverteilte Zufallsvariable. Man nehme an, dass jedes Gespräch unabhängig von den anderen mit Wahrscheinlichkeit p eine Auslandsverbindung erfordert. Bestimmen Sie die erwartete Anzahl dieser Auslandgespräche.
(Hinweis: Man benutze Aufgabe 45.)

47. Bei einer Lieferung von Kondensatoren sei deren Kapazität (in pF) $\mathcal{N}(100, 0.25^2)$ – verteilt. Wie viel Prozent Ausschuss sind zu erwarten, wenn die Kapazität der Kondensatoren mindestens 99.7pF betragen soll?
Wie hoch ist der Ausschussanteil, wenn die Kapazität höchstens 100.7pF betragen darf?
Wie muss man Toleranzgrenzen $100 + c$ und $100 - c$ wählen, damit 95% dieser Kondensatoren eine Kapazität zwischen diesen Grenzen besitzen?

48. Von $N = 1050$ Schülern eines Gymnasiums stammen $R = 370$ aus einer geschiedenen Ehe. Für die statistische Auswertung der Leistungen des aktuellen Abiturjahrgangs mit $n = 130$ Schülern kann aus Datenschutzgründen die Anzahl der darunter befindlichen Schüler aus geschiedenen Ehen nicht explizit nachgeprüft werden. Sie stellt also eine Zufallsvariable X dar.
Berechnen Sie mit Hilfe einer Normalapproximation der hypergeometrischen Verteilung die

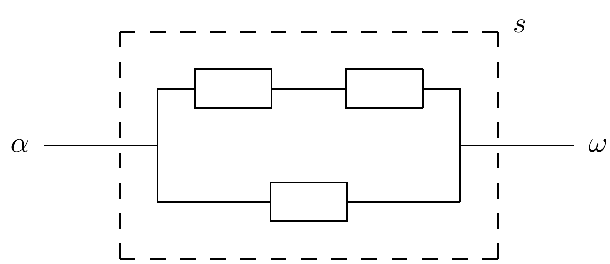
Wahrscheinlichkeit, dass im Abiturjahrgang mindestens 40 Schüler aus geschiedenen Ehen stammen. Hinweis: $$\operatorname{Var}\{X\} = n \cdot \frac{N-n}{N-1} \cdot \frac{R}{N}\left(1 - \frac{R}{N}\right) .$$

49. Eine Maschine produziere Teile, die mit Wahrscheinlichkeit $p = 0.015$ defekt sind. In einer Kiste befinden sich 100 solcher Teile.
    (a) Berechnen Sie die Wahrscheinlichkeit dafür, dass höchstens zwei Teile defekt sind.
    (b) Vergleichen Sie das Ergebnis unter (a) mit einer geeigneten Approximation durch die Poissonverteilung.
    (c) Wie viele derartige Teile müssten in der Kiste sein, damit die Wahrscheinlichkeit dafür, dass mindestens 100 intakte Teile in der Kiste sind, größer als 95% wird? (Hinweis: Benutzen Sie den Poissonschen Grenzwertsatz.)

50. Durch Berechnung der jeweiligen Wahrscheinlichkeiten nehmen Sie zu der folgenden Behauptung Stellung:
    Beim dreimaligen Würfeln sind die Ereignisse "Augensumme = 11" und "Augensumme = 12" gleichwahrscheinlich, da beide Summen auf 6 verschiedene Arten dargestellt werden können.
    $11 = 6+4+1 = 6+3+2 = 5+5+1 = 5+4+2 = 5+3+3 = 4+4+3$
    $12 = 6+5+1 = 6+4+2 = 6+3+3 = 5+5+2 = 5+4+3 = 4+4+4$

51. Aus einer Urne mit 3 roten und 4 schwarzen Kugeln und aus einer Urne mit 2 roten, 2 weißen und 3 schwarzen Kugeln wird zufällig je eine Kugel gezogen. Wie groß ist die Wahrscheinlichkeit, dass beide gezogenen Kugeln die gleiche Farbe haben?

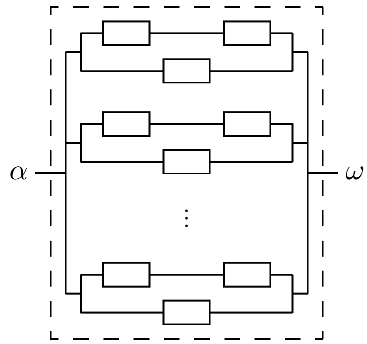
52. Es sei das folgende aus drei Bauteilen bestehende System s gegeben, das genau dann ausfällt, wenn keine intakte Verbindung zwischen α und ω besteht:



    Auf die drei angegebenen Positionen müssen in beliebiger Folge drei Bauteile b_1, b_2, b_3 eingesetzt werden, die mit den Wahrscheinlichkeiten $\alpha_1, \alpha_2, \alpha_3$ unabhängig voneinander ausfallen können.
    (a) Verteilen Sie die Bauteile so, dass das unter geeigneten Unabhängigkeitsannahmen mit möglichst hoher Wahrscheinlichkeit nicht ausfällt.
    (b) Aufgrund eines Produktionsfehlers seien nur noch minderwertige Bauteile verfügbar mit einer Ausfallwahrscheinlichkeit von jeweils 90%, allerdings zu günstigeren Preisen und in einer großen Stückzahl. Vergleichen Sie die beiden folgenden Varianten zur

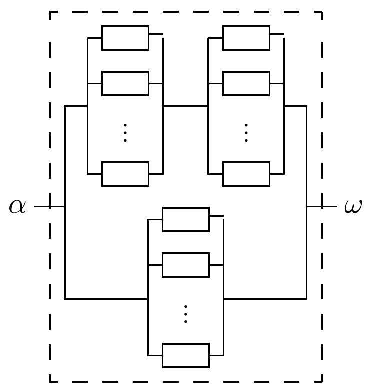
Erhöhung der Systemzuverlässigkeit (Erhöhung der Wahrscheinlichkeit, nicht auszufallen).

i. Variante A: **Redundanz auf der Systemebene**



n komplette Systeme werden parallel geschaltet.

ii. Variante B: **Redundanz auf der Komponentenebene**



Jede Komponente wird n-fach parallel geschaltet.

53. An einer Straßenecke wird Ihnen folgendes Würfelspiel mit zwei symmetrischen Tetraederwürfeln (4 Seiten, die mit gleicher Wahrscheinlichkeit jeweils die Augenzahlen 1, 2, 3 oder 4 zeigen) vorgeschlagen. Zeigen beide Würfel die gleiche Augenzahl, erhalten Sie das 5-fache Ihres Einsatzes zurück. Unterscheiden sich die Augenzahlen um i Punkte, verlieren Sie das i-fache Ihres Einsatzes. Der Einsatz geht in jedem Fall an den Anbieter des Spiels.

   (a) Geben sie einen geeigneten Laplaceraum an.

   (b) Berechnen Sie den Erwartungswert des Gewinns (negativer Gewinn = Verlust). Wer profitiert im Mittel von diesem Spiel?

   (c) Wie hoch muss die Auszahlung in diesem Spiel bei gleicher Augenzahl sein,damit das Spiel bei sonst unveränderten Konditionen "fair" ist, also der "mittlere" Gewinn gleich Null ist.

A.3 Allgemeine Wahrscheinlichkeitsräume

1. Bestimmen Sie die Wahrscheinlichkeit $P(-0.5 \leq X \leq 1.5)$ falls
   (a) $X \sim \mathcal{N}(0, 1^2)$ –verteilt,
   (b) $X \sim \mathcal{N}(0.5, 2^2)$ – verteilt ist.

   Für welche x – Werte gilt

   (c) $P(X < x) \geq 0.99$, falls $X \sim \mathcal{N}(1, 2^2)$ – verteilt ist,
   (d) $P(-x < X < x) \geq 0.95$, falls $X \sim \mathcal{N}(0, 1^2)$ ist?

2. Eine Abfüllmaschine füllt X Gramm eines Produktes in Y Gramm schwere Dosen. Dann werden jeweils 20 Dosen in eine Z Gramm schwere Kiste verpackt. X, Y, Z seien unabhängige normalverteilte Zufallsvariable mit
$$X \sim \mathcal{N}(155, 4^2) \quad , \quad Y \sim \mathcal{N}(45, 3^2) \quad \text{und} \quad Z \sim \mathcal{N}(500, 10^2) \ .$$
   (a) Bestimmen Sie den Erwartungswert und die Streuung des Inhalts einer zufällig aus der Produktion herausgegriffenen Dose.
   (b) Bestimmen Sie den Erwartungswert und die Streuung des Gesamtgewichts einer zufällig aus der Produktion herausgegriffenen Dose.
   (c) Bestimmen Sie die Wahrscheinlichkeit, dass eine zufällig aus der Produktion herausgegriffene gefüllte Kiste schwerer als 4550 Gramm ist.

3. Für das Funktionieren einer bestimmten Maschine ist es erforderlich, dass ein bestimmtes auswechselbares Teil intakt ist. Über seine Lebensdauer in Betriebsstunden ist bekannt, dass sie $\mathcal{N}(\mu, \sigma^2)$–verteilt ist mit $\mu = 108$ und $\sigma = 10$. An jedem Arbeitstag laufen gleichzeitig 10 dieser Maschinen während 16 Stunden. Die Zeit für das Auswechseln defekter Teile soll nicht berücksichtigt werden und der Ausfall defekter Teile erfolge bei allen Maschinen unabhängig von den übrigen.
   (a) X sei die Lebensdauer eines Vorrats von $n = 150$ Stück dieser Teile in Arbeitstagen. Berechnen Sie den Erwartungswert und die Streuung der Lebensdauer dieses Vorrats.
   (b) Ermitteln Sie die Wahrscheinlichkeit, dass der obige Vorrat mindestens 100 Arbeitstage ausreicht.
   (c) Es wird erwogen, jedes Teil nach maximal 100 Betriebsstunden auszuwechseln, auch wenn es noch intakt ist. Wie groß ist die Wahrscheinlichkeit, dass von 10 Teilen keines vorzeitig, d. h. vor Ablauf von 100 Betriebsstunden ausgewechselt werden muss?
   Nach wie viel Betriebsstunden müsste man die Teile auswechseln, damit die Wahrscheinlichkeit für ein vorzeitiges Auswechseln eines einzelnen Teils höchstens 5% betrüge?

4. Ein bestimmtes technisches Gerät wird aus drei Einzelteilen A, B und C zusammengesetzt, wobei das Gerät genau dann funktioniert, wenn alle drei Teile intakt sind und kein Fehler beim Zusammensetzen passiert.
   Die Wahrscheinlichkeit, dass die Einzelteile defekt sind, betragen jeweils 1%, 1% und 5%, während die Wahrscheinlichkeit für einen Fehler beim Zusammenbau 2% betrage. Ferner ist bekannt, dass alle Fehlertypen unabhängig voneinander auftreten.

(a) Berechnen Sie den Erwartungswert und die Streuung der Anzahl defekter Exemplare des Geräts, die sich in einer Lieferung von 1000 Exemplaren befinden.

(b) Die Firma will für eine zufällig zusammengestellte Lieferung von 1000 Exemplaren eine Garantie geben, dass sich höchstens 110 defekte Geräte in der Lieferung befinden. Mit welcher Wahrscheinlichkeit trifft die Garantieaussage zu?

(c) Es sei bekannt, dass jedes defekte Exemplar des Gerätes der Herstellerfirma insgesamt DM 100 an Kosten verursacht. Aufgrund dieser Tatsache soll die Frage entschieden werden, ob von der Möglichkeit Gebrauch gemacht werden soll, die Einzelteile des Typs C zu einem höheren Preis zu beziehen, so dass auch sie nur noch mit einer Wahrscheinlichkeit von 1% defekt sind. Wie hoch darf der Aufpreis pro Stück des Teils C sein, damit es sich im Erwartungswert gerade noch lohnt, das Teil C der höheren Preisklasse zu beziehen?

Hinweis: Benutzen Sie in (b) eine geeignete Näherung.

5. Ein Flugzeug habe 300 Sitzplätze. Wie viele Reservierungen darf eine Fluggesellschaft für einen Flug damit vornehmen, wenn erfahrungsgemäß eine Reservierung mit einer Wahrscheinlichkeit von 10% nicht genutzt wird und die Wahrscheinlichkeit für eine Überbuchung höchstens 2,5% betragen soll. (Benutzen Sie die Normalapproximation.)

6. Ein Softwarehaus beschäftigt 200 Mitarbeiter, von denen jeder einen Teil seiner Arbeitszeit vor einem Bildschirmgerät verbringt. Es ist aber nicht erforderlich, dass jeder Mitarbeiter ein eigenes Terminal zur Verfügung hat. Über einen langen Zeitraum wurde ermittelt, dass der Spitzenbedarf täglich um 8 Uhr erreicht wird und dass jeder Mitarbeiter unabhängig von den übrigen Mitarbeitern an 30% der Arbeitstagen um diese Zeit an einem Bildschirm arbeiten muss.

   Wie groß muss die Zahl der aufgestellten Bildschirmgeräte mindestens sein, damit an einem festen Tag jeder Mitarbeiter, der um 8 Uhr ein Bildschirmgerät benötigt, mit einer Wahrscheinlichkeit von mindestens 95% ein freies Gerät vorfindet?

7. Die stetige Zufallsvariable X_n besitzt für jedes feste $n = 1, 2, 3, \ldots$ die Wahrscheinlichkeitsdichte f_{X_n} mit

$$f_{X_n}(x) \;=\; \begin{cases} c_n \cdot \frac{1}{x} & \text{falls} \quad \frac{1}{n} \le x \le n \;, \\ 0 & \text{sonst} \;. \end{cases}$$

   Bestimmen Sie c_n und die Wahrscheinlichkeit $P(X < 1)$!
   Welche Bedeutung hat der Wert x=1 für die Lage der Realisierungen der Zufallsvariablen X_n?

8. Eine Pumpe sei ununterbrochen in Betrieb, bis sie ausfällt. Die Zufallsvariable X, die die zufällige Dauer der Funktionstüchtigkeit der Pumpe beschreibt, habe die Wahrscheinlichkeitsdichte

$$f_X(x) \;=\; \begin{cases} \lambda^2 x e^{-\lambda x} \,, & x > 0 \;, \\ 0 \,, & x \le 0 \;. \end{cases}$$

   Weiter sei bekannt, dass Pumpen dieser Art im Mittel 100 Stunden laufen, bis sie ausfallen.

(a) Zeigen Sie, dass die *Ausfallrate* $g(t) := \dfrac{f_X(t)}{1 - F_X(t)}$ für $t > 0$ streng monoton wachsend ist.

(b) Wie ist der Parameter λ zu wählen, damit der Erwartungswert von X gleich der obigen *mittleren Laufzeit* dieser Pumpen ist? Bestimmen Sie für diesen Parameter die folgenden Wahrscheinlichkeiten:

$$P(X \leq 100) \quad , \quad P(X \leq 200 \mid X \geq 100) \ .$$

Hinweis: Die unbestimmten Integrale können einer Formelsammlung entnommen oder durch partielle Integration ermittelt werden.

9. Die Zufallsvariable X beschreibe den Durchmesser (in mm) maschinell gefertigter Unterlegscheiben. Der Erwartungswert μ der Zufallsvariablen X hänge von der Maschineneinstellung ab und kann sich im Laufe der Zeit ändern. Die Varianz $\mathrm{Var}\{X\} = 0.04$ sei als feste Maschinengröße bekannt und von der Maschineneinstellung unabhängig. Zur Schätzung des unbekannten Erwartungswertes μ werden n Scheiben zufällig aus der Produktion ausgewählt. Dabei beschreibe

$$\overline{X}_n \ \frac{1}{n}\sum_{i=1}^{n} X_i$$

den mittleren Durchmesser. Wie groß muss n mindestens sein, damit $P\left(|\overline{X}_n - \mu| < 0.1\right) \geq 0.999$ erfüllt ist,

(a) falls über die Verteilung von X nichts bekannt ist,

(b) falls X näherungsweise normalverteilt ist.

10. Die Zufallsvariable X habe die Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & \text{falls} \quad x < 1, \\ \frac{1}{6} + \frac{2}{3}(x-1) & \text{falls} \quad 1 \leq x < 2, \\ 1 & \text{falls} \quad 2 \leq x \end{cases}$$

Berechnen Sie $P(X \leq 5/3)$, $P(X > 3/2)$ und $P(4/3 < X \leq 5/3)$.
(Hinweis: Es handelt sich um eine gemischte Zufallsvariable.)

A.4 Multivariate Verteilungen

1. Ein Kind nimmt eine Handvoll Steine, d.h. mit Wahrscheinlichkeit 1/3 ergreift es drei Steine, mit W. 1/2 zwei Steine und mit W. 1/6 einen Stein. Diese Steine wirft das Kind nun nacheinander gegen eine Dose. Die Dose wird bei jedem Wurf unabhängig von der Anzahl der ergriffenen Steine und den übrigen Würfen mit einer W. von 2/3 getroffen. Die Zufallsvariable Y gibt die Anzahl der Treffer an.

(a) Berechnen Sie die Wahrscheinlichkeitsverteilung, den Erwartungswert und die Varianz der Zufallsvariablen Y.

(b) Das Kind hat die Dose nicht getroffen. Wie groß ist die Wahrscheinlichkeit, dass das Kind höchstens 2 Steine ergriffen hatte?

2. Das Intervall $[0,1]$ werde durch eine auf $(0,1)$ gleichverteilte Zufallsvariable U in zwei Teile geteilt. Berechnen Sie

   (a) die mittlere Länge des linken Teilstücks,

   (b) die mittlere Länge des kürzeren Teilstücks und

   (c) die mittlere Fläche des Rechtecks, dessen Kanten von beiden Teilstücken gebildet werden.

3. Auf einer Baustelle vollenden die Maurer ihre Tätigkeit am Rohbau zu einem Zeitpunkt X. Die Zimmerleute beginnen anschließend ihre Tätigkeit zum Zeitpunkt Y. Beide Zeitpunkte sind für den Bauherrn nicht explizit vorhersehbar und werden daher von ihm als Zufallsvariable aufgefasst. Dabei nimmt er folgendes an:

$$X \quad \text{ist gleichverteilt auf dem Intervall } [0,1).$$

   Beenden die Maurer ihre Arbeit zu einem Zeitpunkt $0 \le x < 1$, so ist der Arbeitsbeginn der Zimmerleute auf dem Restintervall $[x,1]$ gleichverteilt.

   (a) Bestimmen Sie die gemeinsame Dichte von X und Y !

   (b) Bestimmen Sie die Dichte von Y !

   (c) Wie groß ist der Erwartungswert von Y ?

   (d) Sind X und Y stochastisch unabhängig ?

4. Der Zufallsvektor (X,Y) besitzt die Wahrscheinlichkeitsdichte

$$f_{X,Y}(x,y) \;=\; \begin{cases} c \cdot x \cdot |y| & \text{falls} \quad 0 \le |y| \le x^2 < 1\ ,\ x \ge 0\ , \\ 0 & \text{sonst}\ . \end{cases}$$

   (a) Bestimmen Sie die Konstante c.

   (b) Ermitteln Sie die Marginaldichten f_X und f_Y.

   (c) Sind X und Y unabhängig? Begründen Sie Ihre Aussage.

   (d) Berechnen Sie $\mathrm{E}\{X\}$ und $\mathrm{Var}\{X\}$.

   (e) Bestimmen Sie die bedingte Dichte $f_{X|Y=y}(x)$.

   Geben Sie bei (c) und (e) auch die zulässigen Werte von x und y an.

5. Ein Zufallsvektor $\begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$ heißt genau dann zweidimensional exponentialverteilt mit den Parametern $\lambda_1, \lambda_2, \lambda_3$, wenn seine Verteilungsfunktion F die Gestalt

$$F(t_1,t_2) \;=\; \begin{cases} 1 - e^{-(\lambda_1+\lambda_3)t_1} - e^{-(\lambda_2+\lambda_3)t_2} + e^{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max\{t_1,t_2\}} & , \quad \text{für } t_1, t_2 > 0 \\ 0 & , \quad \text{sonst} \end{cases}$$

   besitzt mit $\lambda_1, \lambda_2 > 0$ und $\lambda_3 \ge 0$.

(a) Bestimmen Sie die Randverteilungen von F.

(b) Zeigen Sie, dass T_1 und T_2 genau dann unabhängig exponentialverteilt sind, wenn $\lambda_3 = 0$ gilt.

(c) Berechnen Sie
$$P\big(T_1 > t_1\ ,\ T_2 > t_2\big) \qquad \forall\ t_1, t_2 > 0$$
und
$$P\big(\min\{T_1, T_2\}\ >\ t\big) \qquad \forall\ t > 0\ .$$

6. Eine Urne enthält 3 Kugeln, die mit den Ziffern 1, 4 bzw. 7 beschriftet sind. Nacheinander wird dreimal eine Kugel gezogen, ihre Aufschrift notiert und die Kugel wieder in die Urne zurückgelegt (Ziehen mit Zurücklegen). Mit den Zufallsvariablen X wird das Minimum und Y das Maximum der aufgeschriebenen Ziffern bezeichnet.

   (a) Bestimmen Sie die gemeinsame Verteilung von X und Y.

   (b) Ermitteln Sie die Marginalverteilung der Zufallsvariablen X.

   (c) Wie groß sind der Erwartungswert und die Varianz von X.

7. Es seien X und Y zwei nichtnegative Zufallsvariable mit
$$P(X > x, Y > y) = e^{-\lambda x - \mu y - axy}, \qquad x \geq 0, y \geq 0,$$
wobei $\lambda > 0, \mu > 0$ und $a \geq 0$ fest vorgegebene Parameter sind.

   (a) Man bestimme die Verteilungsfunktionen F_X und F_Y von X bzw. Y. Was ergibt sich für die Erwartungswerte von X und Y?

   (b) Man berechne die Dichte $f_{X,Y}$ des Zufallsvektors (X, Y).

   (c) Für welches a sind X und Y unabhängig? (Hinweis: Man benutze Satz 8.1.)

8. Die Dichte eines zweidimensionalen stetigen Zufallsvektors (X, Y) sei
$$f_{X,Y}(x, y) = \begin{cases} c\,(\mid x \mid + \mid y \mid) & \text{für} \quad 0 \leq\mid x \mid + \mid y \mid< 1 \\ 0 & \text{sonst} \end{cases} .$$

   (a) Bestimmen Sie die Konstante $c \in \mathbb{R}$.

   (b) Ermitteln Sie die bedingte Dichte $f_{Y|X=x}(y)$. Für welche $x, y \in \mathbb{R}$ ist sie definiert?

   (c) Berechnen Sie $P(0 < Y < \frac{1}{2})$.

9. Wie im Beispiel 8.5 der Vorlesung betrachte man einen Zufallsvektor (X, Y) mit der Dichte
$$f_{X,Y}(x, y) = \begin{cases} 2 & \text{falls} \quad 0 < x < y < 1, \\ 0 & \text{sonst.} \end{cases}$$
Man berechne die Kovarianz von X und Y.

10. Die stetige Zufallsvariable X besitze die Wahrscheinlichkeitsdichte
$$f_X(x) \ = \ \begin{cases} \frac{c}{x^2} & \text{falls } x \geq 2\ , \\ 0 & \text{sonst}\ . \end{cases}$$
Unter der Bedingung $X = x$ sei die Zufallsvariable Y gleichverteilt im Intervall $(x, x + 1]$.

(a) Bestimmen Sie die Konstante $c \in \mathbb{R}$!

(b) Sind X und Y stochastisch unabhängig (Begründung!)?

(c) Bestimmen Sie die gemeinsame Wahrscheinlichkeitsdichte von X und Y und skizzieren Sie den Wertebereich des Zufallsvektors $(X, Y)^T$!

(d) Bestimmen Sie die Wahrscheinlichkeitsdichte von Y und den Erwartungswert von X unter der Bedingung $Y = 3$!

11. (a) Bestimmen Sie $c \geq 0$, so dass die Funktion

$$f(x,y,z) = \begin{cases} c \cdot x^2 \cdot y \cdot z & \text{falls } |x| \leq 2 \;,\; 0 \leq y < 1 \;,\; 0 < z < 1 \;, \\ 0 & \text{sonst} \;, \end{cases}$$

die Wahrscheinlichkeitsdichte eines Zufallsvektors $(X, Y, Z)^{\mathrm{T}}$ ist.

(b) Ermitteln Sie die Marginaldichten f_X und f_Y.

(c) Wie groß ist der Erwartungsvektor von $(X, Y, Z)^{\mathrm{T}}$.

(d) Sind X, Y und Z vollständig unabhängig?

(e) Geben Sie die bedingte Dichte von Y unter der Bedingung $X = 0,5$ an.