

OSNABRÜCKER SCHRIFTEN ZUR MATHEMATIK

Reihe V Vorlesungsskripten

Heft 50 Sommersemester 1997

Wahrscheinlichkeitstheorie und Statistik I

H. Behncke

Fachbereich Mathematik/Informatik
Universität Osnabrück

OSM Osnabrücker Schriften zur Mathematik

Neuaufgabe April 1997

Herausgeber	Selbstverlag der Universität Osnabrück Fachbereich Mathematik/Informatik 49069 Osnabrück
Geschäftsführer	Prof. Dr. W. Bruns
Berater:	Prof. Dr. P. Brucker (Angew. Mathematik) Prof. Dr. E. Cohors-Fresenborg (Didaktik der Mathematik) Prof. Dr. V. Sperschneider (Informatik) Prof. Dr. R. Vogt (Reine Mathematik)
Druck	Hausdruckerei der Universität Osnabrück

Copyright bei den Autoren

Weitere Reihen der OSM:

- Reihe D Mathematisch-didaktische Manuskripte
- Reihe I Manuskripte der Informatik
- Reihe M Mathematische Manuskripte
- Reihe P Preprints
- Reihe U Materialien zum Mathematikunterricht

Wahrscheinlichkeitstheorie und Statistik I

Horst Behncke

Vorwort

Zu Beginn will ich Ihnen kurz ein paar generelle Bemerkungen zu dieser Vorlesung und dem Stoff geben, der hier behandelt werden soll.

Die Vorlesung zerfällt in zwei Teile. Im ersten Teil werden wir uns hauptsächlich mit Fragen der Wahrscheinlichkeitstheorie beschäftigen, während das kommende Semester der Statistik gewidmet ist. Diese beiden Vorlesungen sind in Osnabrück wie auch an den meisten anderen Universitäten Pflichtveranstaltungen, weil die Wahrscheinlichkeitsrechnung einen zentralen Forschungsgegenstand und einen der wichtigsten Anwendungsbereiche der Mathematik beschreibt. Stochastische und statistische Fragestellungen werden Ihnen in Ihrem späteren Berufsleben also noch sehr häufig begegnen. Aufgrund dieser Bedeutung wird die Wahrscheinlichkeitsrechnung schon seit Jahren in den Gymnasien üblicherweise behandelt. Von solchen Grundkenntnissen wird aber allenfalls nur der Stoff eines normalen Kurses Wahrscheinlichkeitsrechnung des Gymnasiums vorausgesetzt.

Zu jeder dieser Vorlesungen gibt es eine Übung. Der Leistungsnachweis für die erste Veranstaltung WUST I ist für Studenten des Diplomstudienganges Pflicht, weil man nur so die nötige Vertrautheit mit den Begriffen erwirbt. Der Nachweis für die Veranstaltung WUST II sollte Pflicht sein, weil man von einem Mathematiker die Kenntnis elementarer statistischer Verfahren erwartet. Ich empfehle Ihnen daher nachdrücklich:

Bearbeiten Sie regelmäßig und gründlich die Übungsaufgaben.

Wahrscheinlichkeitstheorie und Statistik haben heute einen breiten Anwendungsbereich. Kaum eine Wissenschaft kann ohne sie auskommen; man denke z.B. an die vielfältigen Anwendungen in der Physik, Biologie, Medizin, Volkswirtschaft oder Sozialwissenschaft.

Dennoch ist die Wahrscheinlichkeitstheorie eine abstrakte mathematische Theorie, auf Axiomen aufgebaut wie jede mathematische Disziplin. Genausowenig wie man z.B. in der Geometrie “Geraden” und “Punkte” definieren kann, genauso wenig kann man “Ereignisraum” und “Wahrscheinlichkeit” definieren. Man legt sie durch ihre Grundeigenschaften axiomatisch fest. Hierbei wird man natürlich von einer gewissen Anschauung, die aus einfachen Beispielen wie dem Würfelspiel gewonnen wird, geleitet; wie man auch in der Geometrie von einer intuitiven Anschauung von Geraden und Punkten geleitet wird. Aufgabe der Wahrscheinlichkeitstheorie ist es nun, aus diesen Axiomen Schlüsse zu ziehen. Obwohl diese Schlüsse selbst abstrakt sind, stammen die zu lösenden Probleme wieder aus den Anwendungen. Auf diese Art besteht eine enge Beziehung zwischen abstrakter Theorie und Anwendungen in der Praxis: Probleme außerhalb der Mathematik führen zu neuen theoretischen Fragestellungen

und Ergebnissen, neue theoretische Ergebnisse ziehen neue, vorher oft unbekannte Anwendungen nach sich. Streng genommen ist auch die Herleitung des Modells, d.h. des Ereignisraumes und der Wahrscheinlichkeiten, für einen bestimmten Vorgang nicht Aufgabe der Wahrscheinlichkeitstheorie sondern wird von außen vorgegeben. Solche Modelle stellen stets eine Idealisierung des tatsächlichen Vorgangs dar.

Diese grundlegenden Überlegungen gelten für jede mathematische Disziplin, da jede auch noch so abstrakte Theorie einen praktischen Hintergrund hat. Da bei der Wahrscheinlichkeitstheorie die Anwendungen besonders nahe liegen, ist es wichtig, auf den axiomatischen Charakter der Wahrscheinlichkeitstheorie hinzuweisen. Die besondere Anschaulichkeit hat Vor- und Nachteile: einmal erleichtert sie das Verständnis, andererseits besteht jedoch die Gefahr, sich zu sehr von der Anschauung leiten zu lassen, was leicht zu Fehlschlüssen führen kann. Es ist daher, wie auch in anderen Gebieten der Mathematik notwendig, die Bereiche Intuitiver Hintergrund der Theorie, axiomatischer Aufbau und Anwendungen, sorgfältig auseinanderzuhalten.

Auch in dieser Vorlesung werden wir uns an den axiomatischen Stil halten, wie er von Kolmogorov begründet wurde. In diesem Sinne ist diese Vorlesung daher Standard, da der Stoff weitgehend kanonisch ist. Trotzdem soll diese Vorlesung darüberhinaus eher beispiel- und anwendungsorientiert sein, d.h. der formale Rahmen wird durch die Axiome und die Theorie bestimmt, aber Beispiele, Anwendungen und Begriffsbildung stehen im Vordergrund. Aus diesem Grunde werde ich daher häufig auch auf die Beweise komplizierter maßtheoretischer Sätze verzichten und eher Anwendungen und Spezialfälle betrachten. Die Anwendungsorientierung bringt einen weiteren Gesichtspunkt mit sich, nämlich den des Wahrscheinlichkeitsmodells. Dieser Begriff und die damit verbundene Grundhaltung wird für diese Vorlesung zentral sein, so daß wir die Vorlesung auch "Wahrscheinlichkeitsmodelle" nennen könnten.

Der Aufbau dieser Vorlesung ist grob wie folgt:

1. Zufall, Wahrscheinlichkeit und Modelle
2. Diskrete und geometrische Modelle
3. Die Axiome
4. Maßtheorie
5. Grundlagen der Wahrscheinlichkeitstheorie
6. Spezielle Verteilungen und ihre Anwendungen
7. Grenzwertsätze

Die Wahrscheinlichkeitstheorie hat inzwischen vielerlei Anwendungen gefunden z.B.:

- Stochastische Prozesse (Warteschlangen, Lagerhaltung, stat. Physik, stochastische Modelle der Biologie)
- Versicherungsmathematik
- Zuverlässigkeitstheorie

- Stochastische Optimierung
- Stochastische Differentialgleichungen

Wahrscheinlich werden wir aber keines dieser Gebiete hier ansprechen können.

Auch die Statistik läßt sich als ein Anwendungsgebiet der Wahrscheinlichkeitstheorie auffassen. Im Gegensatz zur Wahrscheinlichkeitstheorie, die eine mathematisch geschlossene, einheitliche Theorie darstellt, gilt dies so nicht für die Statistik, die immer noch aus einer Vielzahl von Methoden ohne einheitliche Theorie besteht. Sie kann daher noch nicht als einheitliches Ganzes angesehen werden. Dies erging jedoch jeder mathematischen Disziplin in ihren Anfängen so. Das Axiomensystem der Wahrscheinlichkeitstheorie z.B. ist erst in unserem Jahrhundert von dem russischen Mathematiker Kolmogorov (1933) aufgestellt worden, während die Anfänge der Wahrscheinlichkeitstheorie im 17. Jahrhundert liegen. In den Grundlagen der Statistik hat sich sogar etwas wie ein Glaubensstreit entwickelt (vgl. hierzu den Artikel von B. Effron in *Am. Math. Monthly* 1978). Für die Grundlagen der Wahrscheinlichkeitstheorie gab es lange Zeit ähnliche Probleme mit der Rechtfertigung und Grundlegung der Axiome. Für Details sei auf das Buch von Schnoor verwiesen. Wir wollen hier den üblichen, naiven Zugang zur Begründung der Wahrscheinlichkeitstheorie wählen.

Literaturverzeichnis

- [1] Bauer, H.: Wahrscheinlichkeitstheorie und Maßtheorie
- [2] Breiman, L.: Probability and Stochastic Processes
- [3] Chung, K. L.: Elementare Wahrscheinlichkeitstheorie und Stochastische Prozesse
- [4] Feller, W.: An Introduction to Probability Theory and its Applications
- [5] Freudenthal, H.: Wahrscheinlichkeit und Statistik
- [6] Hinderer, K.: Grundbegriffe der Wahrscheinlichkeitstheorie
- [7] Loeve, M.: Probability Theory I
- [8] Massey, L. D.: Probability and Statistics
- [9] Lesalkin, L.: Problems in Probability
- [10] Olkin, I.; Gleser, L.; Derman, C.: Probability Models and Applications
- [11] Schnoor, C. P.: Zufälligkeit und Wahrscheinlichkeit

Inhaltsverzeichnis

1	Grundlagen der Wahrscheinlichkeitstheorie	3
1.1	Zufall, Wahrscheinlichkeit und Modelle	3
1.2	Die Algebra der Ereignisse	10
1.3	Die Wahrscheinlichkeit	13
1.4	Wahrscheinlichkeitsmodelle	17
2	Geometrische und Diskrete Modelle	19
2.1	Geometrische Modelle	19
2.2	Diskrete endliche Modelle	22
3	Einfache Konzepte der Wahrscheinlichkeitstheorie	33
3.1	Bedingte Wahrscheinlichkeit und Unabhängigkeit	33
3.2	Bayes' Satz	39
4	Maßtheorie	43
4.1	σ -Algebren	43
4.2	Meßbare Abbildungen	46
4.3	Messbare numerische Funktionen	50
4.4	Maße	52
4.5	Integrationstheorie	56
4.6	Konvergenzsätze	62
4.7	Der Satz von Radon Nikodym	64
5	Zufallsvariable	67
5.1	Zufallsvariable, Verteilungen und Momente	67
5.2	Diskrete Verteilungen	70
5.2.1	Die Binomial Verteilung	71
5.2.2	Die hypergeometrische Verteilung	74
5.2.3	Die Poisson-Verteilung	74
5.2.4	Die negative Binomial-Verteilung	76
5.2.5	Multinomial Verteilung	78
5.3	Stetige Verteilungen	81

5.3.1	Die gleichmäßige Verteilung	82
5.3.2	Die Normalverteilung	83
5.3.3	Lage- und Skalenparameter	91
5.3.4	Die Exponentialverteilung	92
5.3.5	Die Laplace-Verteilung	92
5.3.6	Die Gamma-Verteilung	93
5.3.7	Die Beta-Verteilung	93
5.3.8	Weibull-Verteilung	94
5.3.9	Lognormal-Verteilung: $LN(a, m, \sigma)$	94
6	Zufallsvektoren	99
6.1	Mehr über Zufallsvariable	99
6.2	Zufallsvektoren	101
6.3	Unabhängige σ -Algebren	108
6.4	Unabhängige Zufallsvariablen	111
6.5	Die Multivariate Normalverteilung	118
6.6	Der Poisson Prozeß	121
7	Gesetze der großen Zahlen	123
7.1	Konvergenzbegriffe in der Wahrscheinlichkeitstheorie	123
7.2	Das starke Gesetz der großen Zahlen	126
7.3	Das schwache Gesetz der großen Zahlen	131
7.4	Der zentrale Grenzwertsatz	132

Kapitel 1

Grundlagen der Wahrscheinlichkeitstheorie

1.1 Zufall, Wahrscheinlichkeit und Modelle

In der Wahrscheinlichkeitstheorie (W.Th.) hat man es mit Erscheinungen zu tun, deren Ausgang ungewiß, zufällig ist. Unter *Zufall* verstehen wir dabei das absichtslose, unbestimmte plan- und regellose Eintreffen von Vorgängen und Ereignissen. In diesem Zusammenhang sprechen wir auch von *Zufallserscheinungen* oder *zufälligen Ereignissen*. Beispielsweise ist das Ergebnis beim Münzwurf zufällig, denn wir können vorher nicht mit Sicherheit sagen, ob Kopf oder Zahl fällt. Gleichfalls ist der Ausgang beim Würfelexperiment ungewiß. Dagegen stehen Erscheinungen, die durch eine Theorie vollständig beschrieben sind und vorausberechnet werden können, wie etwa Planetenbewegung oder chemische Reaktionen.

Obwohl es im Prinzip mit Hilfe der Gesetze der klassischen Mechanik möglich ist, im voraus bei einer geworfenen Münze zu bestimmen, ob Kopf oder Zahl fällt (Determinismus), fassen wir doch den Münzwurf als ein Zufallsexperiment auf, da das Ergebnis von einer so großen Zahl von Anfangsdaten abhängt, daß eine Berechnung praktisch unmöglich ist.

Ohne näher auf die philosophischen Probleme mit Begriffen wie Zufall, Determinismus und Kausalität einzugehen, sagen wir: *Zufällige Ereignisse sind solche, die außerhalb einer bestimmten oder außerhalb jeder Gesetzmäßigkeit ablaufen*. Hier einige Beispiele.

Beispiel 1.1 Zufallsgeräte: Hierunter fallen fast alle Gerätschaften, die in Glücksspielen verwendet werden, wie etwa:

- a) Würfel
- b) Karten
- c) Roulette
- d) Lotto

Beispiel 1.2

- a) Das Ereignis, daß ein Neugeborenes ein Junge oder ein Mädchen ist, ist zufällig.

Gleiches gilt für

- b) Haarfarbe
- c) Gewicht etc.

bei Neugeborenen, innerhalb gewisser Grenzen.

Beispiel 1.3 Viele Vorgänge sind zufällig in der Zeit, wie etwa

- a) Brenndauer einer Glühbirne
- b) Zerfall eines bestimmten radioaktiven Atoms
- c) Das Platzen eines Popkorns im heißen Topf.

Alle diese Beispiele haben noch eine weitere Eigenschaft: Sie lassen sich beliebig häufig wiederholen. Solche Erscheinungen nennen wir *Zufallsexperimente*.

Zufallsexperimente sind also zufällige Ereignisse, die sich unter gleichen, wohldefinierten Bedingungen beliebig häufig wiederholen lassen.

Dabei wollen wir unter beliebig häufig wiederholen nicht unbedingt eine zeitliche Abfolge verstehen. Sondern wir meinen damit auch das gleichzeitige Nebeneinanderablaufen von Zufallserscheinungen. Beispiel hierfür sind der Fall von Regentropfen, das Keimen von Saatgut oder Entscheidungen für eine politische Partei. Wir sprechen dabei dann von *Massenerscheinungen*.

Sicherlich wäre der Ausbruch eines Krieges im nächsten Jahr ein zufälliges Ereignis, aber jedoch kein Zufallsexperiment. Ähnliches gilt für den Anstieg von Aktienkursen und ähnlichen Ereignissen in der Wirtschaft.

Das Maß für die Zufälligkeit eines Ereignisses ist seine *Wahrscheinlichkeit* (W). Ereignisse können dadurch hinsichtlich ihrer Zufälligkeit verglichen und eingeordnet werden. Die Skala reicht dabei von sicher bis unmöglich. Typische Aussagen dieser Art sind

- i) HSV wird fast sicher Deutscher Meister
- ii) Im nächsten Jahr gibt es wahrscheinlich keinen Krieg
- iii) Das Pferd „Blue Boy“ gewinnt wahrscheinlich
- iv) Der Aufschwung kommt sicher bald
- v) Die Wahrscheinlichkeit 8 Augen mit 2 Würfeln zu würfeln, ist $5/36$
- vi) Die Zuverlässigkeit einer Minuteman ist 0,8

Von diesen Aussagen sind die letzten beiden quantitativer Natur. Aussagen dieser Art macht die Wahrscheinlichkeitstheorie.

Das quantitative Maß für die Zufälligkeit eines Ereignisses ist die (mathematische) Wahrscheinlichkeit.

Man unterscheidet dabei grob die folgenden drei Unterbegriffe:

Subjektive Wahrscheinlichkeit

Die subjektive Wahrscheinlichkeit ist ein quantitatives Maß für die Möglichkeit, das Eintreffen eines Ereignisses aufgrund der Einschätzung eines Individuums (Subjektes). Sie kann von Person zu Person verschieden sein, sollte jedoch gewissen Konsistenzbedingungen genügen. Dieser Begriff wurde axiomatisch zuerst von Savage (The foundations of statistics) 1954 eingeführt. Formal wird dieser Wert über "Wetten" festgelegt. Ein Beispiel mag dies verdeutlichen. Wenn ich sage: „Morgen scheint mit der Wahrscheinlichkeit 0,4 die Sonne“, so bedeutet dies, daß ich bereit bin, eine Wette 4:6 dafür einzugehen, ungünstigere wie 5:5 aber nicht.

Die subjektive Wahrscheinlichkeit ist wegen ihrer mangelnden Objektivität etwas in Verruf geraten, wird aber doch häufiger angewandt als man wahrhaben möchte. Sie tritt meist in Wahrscheinlichkeitsmodellen auf, die keine Zufallsexperimente beschreiben. Beispiele hierzu sind mathematische Entscheidungsmodelle für die Wirtschaft, Zuverlässigkeit, Politik und das Militär.

A Priori Wahrscheinlichkeit

Diese Wahrscheinlichkeit wird theoretisch logisch, erfahrungsunabhängig begründet. Als Beispiel dazu betrachten wir einen idealen, homogenen unsymmetrischen Würfel. Bei einem solchen Würfel ist keine der Seiten ausgezeichnet, so daß beim Würfeln alle Augenwerte die gleiche Wahrscheinlichkeit, nämlich $\frac{1}{6}$ haben.

Allgemeiner gelten solche Betrachtungen für die üblichen Zufallsgeräte, die meist gerade so konstruiert sind, daß alle Ereignisse gleichwahrscheinlich sind. Hierbei sei etwa an das Lotto-Kugelgerät erinnert, das allwöchentlich im Fernsehen zu betrachten ist und wo ein Notar etwaige Manipulationen ausschließt.

In diesem Zusammenhang spricht man häufig von *fairen* Zufallsgeräten und meint damit, daß alle Ergebnisse gleichwahrscheinlich sind.

Empirische Wahrscheinlichkeit

Die empirische Wahrscheinlichkeit wird mit Hilfe einer Folge von Zufallsexperimenten bestimmt. Nehmen wir an, dieses Zufallsexperiment werde n -mal durchgeführt und das Ereignis A werde $n(A)$ mal beobachtet. Man nennt dann

$$p_n(A) = n(A)/n \quad (1.1)$$

die *relative Häufigkeit* für das Eintreffen von A . Falls dann $\lim p_n(A) = p(A)$ existiert, so wird man $p(A)$ die empirische Wahrscheinlichkeit für das Ereignis A nennen. Diese Definition setzt neben der beliebigen Wiederholbarkeit des Experimentes also auch noch die Existenz des Grenzwertes für jedes Ergebnis voraus. In der Praxis wird man $p(A)$ durch $p_n(A)$ für große n ersetzen.

Die empirische Wahrscheinlichkeit ist für die Theorie von großer Bedeutung, da ihr Wert am ehesten einen objektiven Charakter hat und sie die in den Naturwissenschaften geforderte Reproduzierbarkeit erfüllt. Auch die a priori Wahrscheinlichkeit erfüllt diese Bedingungen. Allerdings sollte man sich darüber klar sein, daß diese

Begriffe in ihrer reinen Form daher auch praktisch nur in den Naturwissenschaften auftreten.

Beispiel 1.4 Wahrscheinlichkeitsaussagen bei der Wettervorhersage sind zum Teil empirischer Natur.

Beispiel 1.5 Die Zuverlässigkeit der amerikanischen Minutemen wird mit ungefähr 0,8 angegeben. Trotzdem hat noch *keine* der Raketen die „realistische“ Polarroute geflogen. Es gab wegen eines möglichen 'bias' in der Trefferverteilung der Raketen deshalb mehrere Anhörungen.

Beispiel 1.6 Wenn ein Tennisspieler einen anderen bei 5 Spielen viermal geschlagen hat, so ist 0,8 sicher eine mögliche Gewinnwahrscheinlichkeit, empirisch wird man sie aber kaum nennen können.

Die Wahrscheinlichkeitstheorie soll verwandt werden, um Zufallserscheinungen zu beschreiben. Eine solche Beschreibung geschieht im Rahmen eines mathematischen Modells.

Aufgabe 1.1 Ist die Frage „Wie groß ist die Wahrscheinlichkeit mit einem beliebigen Würfel eine „6“ zu würfeln sinnvoll?“ Wie läßt sich diese Frage sinnvoll machen?

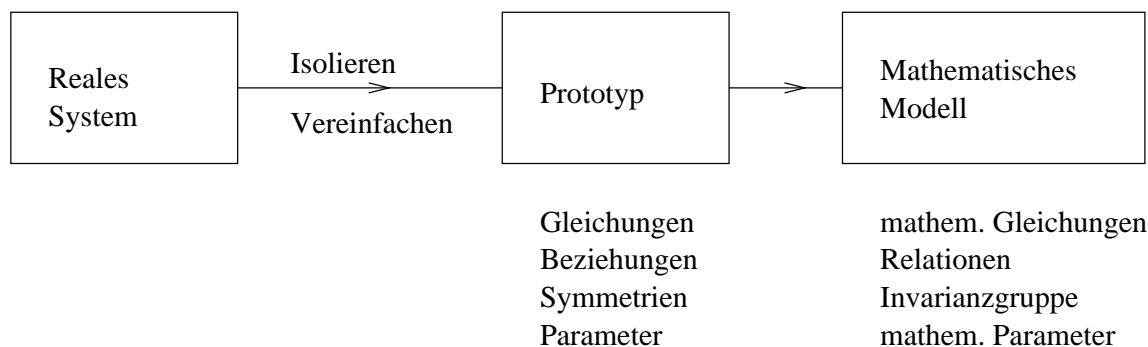
Aufgabe 1.2 Wie groß ist die a priori Wahrscheinlichkeit für das Ergebnis 1, 2, 3, 4, 5, 6 beim „6 aus 49“ Lotto?

Aufgabe 1.3 Ein Spekulant schätzt die Gewinnchancen für ein Börsengeschäft als 0,3. Wie hoch muß der zu erwartende Gewinn mindestens sein, um ihn zu veranlassen, mit DM 10.000,- einzusteigen?

Mathematisches Modell

Ein mathematisches Modell eines „realen“ Systems ist ein in sich vollständiges und widerspruchsfreies System von Gleichungen oder Relationen, die dem realen System entsprechen.

Modellieren vollzieht sich dabei meist in folgenden Schritten:



Das Attribut ‘real’ wird dabei bewußt im naiven Sinne verwandt. Dabei sollen natürlich die wesentlichen Eigenschaften des Prototyps im Modell reflektiert werden, d.h. das Modell soll die wichtigsten Strukturen des Prototyps darstellen. Das Modellieren ist also eine Art des Homomorphismus. In der Regel wird das Modell eine vereinfachte Darstellung des Systems sein, d.h. die unwesentlichen Eigenschaften werden weggelassen. Der Grund dafür ist, daß man dann an dem so vereinfachten System-Modell viele Beziehungen und Regeln leichter erkennen kann, die in dem meist sehr komplexen System nicht offensichtlich sind. Schon dies aber zeigt, daß es für ein ‘reales System’ meist viele Modelle gibt, die sich in der Genauigkeit ihrer Beschreibung unterscheiden.

Als wichtigste Typen seien hier die *diskreten Modelle* und die *kontinuierlichen Modelle* genannt. Bei einem diskreten Modell werden die Parameter des Systems durch diskrete Variable, d.h. Größen die nur abzählbar viele Werte annehmen können, beschrieben. Bei einem stetigen Modell durchlaufen die Parameter reelle Intervalle. Ein diskretes Modell heißt *endlich*, wenn alle Parameter nur endlich viele Werte annehmen können. Natürlich kann man argumentieren, daß letztlich alle Modelle endlich sein könnten oder sollten, da die Welt endlich ist. Dies ist jedoch nur vordergründig richtig, da die Güte eines Modells durch die Genauigkeit der Vorhersagen und den (mathematischen) Aufwand bestimmt wird. Ohne stetige Modelle wären etwa alle Anwendungen der Differentialgleichung etc. überflüssig, und durch entsprechende diskretisierte Varianten zu ersetzen.

Ferner sind unendliche (diskrete) Modelle notwendig, da sie als Approximation für endliche Modelle mit großen aber ungenau bekannten Wertbereichen gelten. Implizit wird dabei natürlich gefordert, daß die entsprechenden Limiten existieren, während bei kontinuierlichen Modellen häufig implizit Stetigkeitsannahmen über verwandte Funktionen gemacht werden.

Ein mathematisches Modell sollte jedoch die wichtigsten Strukturen möglichst genau widerspiegeln. Die im Modell verwandten Größen sollten daher eine direkte Bedeutung haben und entsprechend interpretierbar sein. Die Möglichkeit so etwas für die verschiedensten Strukturen zu leisten, die Flexibilität der Mathematik, ist der Hauptgrund für ihre weite Anwendbarkeit.

Mit der Hilfe eines Modells lassen sich mathematische Ergebnisse ableiten, die zu Vorhersagen, neuen Einsichten oder neuen gezielten Experimenten führen. Stimmen die Vorhersagen nicht mit beobachteten Ergebnissen überein, muß das Modell revidiert werden. Auch eine Verfeinerung oder Verbesserung aufgrund der Vorhersagen ist möglich. Dies führt dann zu einer Rückkoppelung vom Modell zum System. Diese Überlegungen gelten natürlich für alle Typen mathematischer Modelle.

Wir wollen uns nun noch überlegen, wie ein typisches Wahrscheinlichkeitsmodell (WM) auszusehen hat. Wir werden dabei schon von dem isolierten Prototyp ausgehen und auch nur ganz einfache Systeme beschreiben.

Womit können wir ein Zufallsexperiment beschreiben? Zunächst einmal müssen wir die Menge aller möglichen Ausgänge (Ergebnisse) eines Zufallsexperiments kennen. Im Fall des Würfels sind dies z.B. (1, 2, 3, 4, 5, 6) und im Fall des Geschlechtes von Kindern (J, M).

Die Menge aller Ausgänge eines Zufallsexperimentes wird der Ereignisraum genannt.

Beispiele von Ereignisräumen:

Beispiel 1.7 Würfeln: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Beispiel 1.8 Münzwerfen: $\Omega = \{K, Z\}$

Beispiel 1.9 Zweimaliges Münzwerfen: $\Omega = \{KK, KZ, ZK, ZZ\}$

Beispiel 1.10 Anzahl der Unfälle auf den Bundesautobahnen am 01.03.1984: $\Omega = \{0, 1, 2, \dots\}$

In dem letzten Beispiel wird schon eine Idealisierung deutlich. Obwohl es am 01.03.1984 sicher weniger als 61,5 Mio. Unfälle gibt, wählt man $\Omega = \mathbb{N}$. Ähnliches gilt bei

Beispiel 1.11 Bei der Bildung von Polymeren entstehen Ketten verschiedener Länge. Messen wir die Länge der Moleküleinheiten, so ist $\Omega = \{1, 2, 3, \dots\}$.

Noch komplizierter wird das System beim Auftreten kontinuierlicher Größen.

Beispiel 1.12 Brenndauer von Glühbirnen: Wählt man hier $\Omega = \{t | t \geq 0\}$ als Menge der Ergebnisse, so bedeutet dies eben auch, daß eine Aussage wie: „Die Brenndauer der Glühbirnen XYZ ist $100\pi/e^3h$ sinnvoll ist“. Neben $\Omega = [0, \infty)$ könnte man natürlich auch $\Omega = \{0, 1, 2, \dots\}$ wählen, wenn die Brenndauer nur auf 1 h genau bestimmt wird.

Ähnliches wie Beispiel 1.12 ist

Beispiel 1.13 Lassen wir eine Kugel auf einen ebenen Fußboden fallen, so ist $\Omega = \{\text{Punkte des Fußbodens}\}$.

Aufgabe 1.4 Gebe die Ereignisräume für die folgenden Zufallsexperimente an:

- a) 30 Studenten werden gefragt, ob sie regelmäßig in der Mensa essen. Die Anzahl der Studenten, die dies tun, wird notiert.
- b) Das Gewicht von Mastochsen wird bestimmt (auf 1 kg genau).
- c) Die Lebensdauer von Radoröhren wird festgehalten (die Lebensdauer kann mit perfekter Genauigkeit gemessen werden).
- d) Es wird 3mal hintereinander eine Münze geworfen.
- e) Es wird das „6 aus 49“ Lotto gespielt.

Die Wahl des Ereignisraumes ist zum Teil durch die Art der Information bestimmt, die man haben will. Interessiert beim Würfeln beispielsweise nur, ob eine „6“ gefallen ist, so wird man als Ereignisraum

$$\Omega = \{6, \text{ nicht } 6\}$$

wählen.

In der Wahrscheinlichkeitstheorie will man Aussagen über die Wahrscheinlichkeit von Ereignissen machen. Beispiele für Ereignisse sind dabei

Beispiel 1.14 Es wird eine gerade Zahl gewürfelt.

Beispiel 1.15 Es geschehen mehr als 100 Unfälle am 01.03.1997 auf den Bundesautobahnen.

Beispiel 1.16 Beim zweimaligen Würfeln ist die Augensumme 7.

Beispiel 1.17 Beim 6 aus 49 Lotto ist eine der Zahlen die 1.

Beispiel 1.18 Die Lebensdauer der Glühbirne liegt zwischen 100 h und 150 h.

Beispiel 1.19 Die Kugel liegt in dem Rechteck $[a_1, b_1] \times [a_2, b_2]$.

Ereignisse entsprechen also Teilmengen des Ereignisraumes.

Gelegentlich nennt man auch die Einpunktmengen des Ereignisraumes *Elementarereignisse*.

Aufgabe 1.5 Gebe für die folgenden Beispiele den Ereignisraum und die Ereignisse an:

- Zweimaliges Würfeln, Ereignis = Gesamtaugenanzahl ≤ 10 .
- Bei einer Untersuchung werden die Personen nach ihren Geburtstagen befragt. Ereignis = Hat im März Geburtstag.
- Es wird die Studiendauer (in Semestern) von Studenten der Universität X untersucht. Ereignis = studiert höchstens 6 Semester.

Wir haben oben Ergebnisse (Ereignisse) als Teilmengen des Ereignisraumes interpretiert.

Im Falle stetiger Modelle bedeutet dies aber noch nicht, daß auch alle Teilmengen von Ω messbaren Ereignissen entsprechen. Allerdings gibt es hier Mengen, die besonders ausgezeichnet sind. Falls $\Omega = \mathbb{R}$, so sind dies Intervalle, denn beispielsweise ist es sinnvoll zu sagen: Die Lebensdauer der Glühbirne liegt zwischen 100 h und 160 h. Falls $\Omega \subset \mathbb{R}^n$, so sind die entsprechenden Mengen n -dimensionale Rechtecke oder Kugeln. Wir wollen solche Ereignisse (Mengen) *natürliche Mengen* nennen.

Im Falle diskreter Modelle wählen wir einfach die Elementarereignisse als natürliche Mengen.

Aufgabe 1.6

- Man bestimme die relative Häufigkeit der Buchstaben A, B, \dots in einem deutschen (englischen, französischen) Text.
- Analog bestimme man die relative Häufigkeit von Worten mit $1, 2, \dots$ Buchstaben.

Aufgabe 1.7 Bestimme aus 1000 Zufallszahlen die relative Häufigkeit für:

- Ziffer ist ungerade
- Ziffer ist 7
1. Ziffer plus 2. Ziffer ≤ 6

1.2 Die Algebra der Ereignisse

Wir haben in 1.1 Ergebnisse als Teilmengen des Ereignisraumes Ω interpretiert. Dies erlaubt es uns, auf besonders einfache Weise die Beziehungen zwischen den Ereignissen durch Mengenoperationen auszudrücken.

Betrachten wir zunächst das Beispiel des zweimaligen Würfels. Der Ereignisraum Ω besteht dann aus allen Paaren von Zahlen $1, 2, \dots, 6$.

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 6)\}.$$

Betrachten wir hier die folgenden Ereignisse

- A: Die Augensumme ist gerade
 B: Die Augensumme ist ≥ 10
 C: Es wird mindestens einmal die 6 gewürfelt.

Wir können dann auch schreiben:

$$\begin{aligned}\Omega &= \{(m, n) | m, n = 1, 2, \dots, 6\} \\ A &= \{(m, n) \in \Omega | m + n \in 2\mathbb{N}\} \\ B &= \{(m, n) \in \Omega | m + n \geq 10\} \\ C &= \{(6, n), (n, 6) | n = 1, 2, \dots, 6\}\end{aligned}$$

Das Ereignis: Die Augensumme ist ungerade, ist dann genau die Menge

$$\Omega \setminus A = \{(m, n) \in \Omega | (m, n) \notin A\}.$$

Das Ereignis: Die Augensumme ist gerade und ≥ 10 ist offensichtlich genau die Menge $A \cap B$.

$$A \cap B = \{(m, n) | m + n = 10, 12\}$$

Dagegen ist $A \cup B$ das Ereignis: Die Augenzahl ist gerade oder ≥ 10 . Wie wir sehen, haben die Operationen Durchschnitt, Vereinigung und Komplement eine große Bedeutung für Ereignisse. Offensichtlich tritt das Ereignis Ω immer ein, denn beim zweimaligen Würfeln fällt genau ein Zahlenpaar.

Ω selber wird daher als das *sichere* Ereignis bezeichnet. Dagegen bezeichnet \emptyset das *unmögliche* Ereignis. Die Beziehung $A \subset B$ bedeutet also in unserer Sprache: Wann immer A eintritt, tritt auch B ein.

Wir fassen diese Regeln nun zusammen. Dabei bezeichnet Ω einen Ereignisraum und A, B, \dots seien Ereignisse.

	Bedeutung
Ω	das sichere Ereignis
\emptyset	das unmögliche Ereignis
$A \subset B$	A impliziert B ; wenn A dann B
$A \cap B$	A und B geschehen gleichzeitig
$A \cup B$	A oder B geschehen
$\Omega \setminus A = A^c$	A tritt nicht ein

Aufgabe 1.8

- a) Wie sieht die Menge aus, die das Ereignis: A oder B geschehen, aber es treten nicht A und B gleichzeitig ein, beschreibt?

- b) Beschreibe das Ereignis $A \cap (\Omega \setminus B)$ in Worten.
 c) Beschreibe das Ereignis $A \cap B \cap C$ in Worten und bestimme $A \cap B \cap C$ im obigen Beispiel.

Gilt für Ereignisse $A_i, i = 1, \dots, A_i \cap A_j = \emptyset$ für $i \neq j$, so nennt man sie *disjunkt*. Disjunkte Ereignisse können also nicht simultan eintreten, sie schließen einander aus. Die Vereinigung disjunkter Ereignisse werden wir mit $\sum A_j$ bezeichnen.

Aufgrund der oben motivierten Operation wird man nun für das System \sum aller Ereignisse fordern.

$$\sum_1 : \Omega \in \sum$$

$$\sum_2 : A \in \sum \Rightarrow A^c = \Omega \setminus A \in \sum$$

$$\sum_3 : A, B \in \sum \Rightarrow A \cup B \in \sum$$

\sum_1, \sum_2 und \sum_3 sind gerade die Axiome für eine *Mengenalgebra*. Wir nennen daher auch \sum die *Ereignisalgebra*. Eine unmittelbare Konsequenz aus \sum_1, \sum_2 und \sum_3 ist

$$A_1, \dots, A_n \in \sum \Rightarrow A_1 \cup \dots \cup A_n, A_1 \cap \dots \cap A_n \in \sum. \quad (1.2)$$

Aufgabe 1.9 Beweise (??).

Lemma 1.1 Sind $\sum_i, i \in I$ Mengenalgebren über der gleichen Grundmenge Ω , so ist auch $\sum = \bigcap_{i \in I} \sum_i$ eine Mengenalgebra.

Beweis: Offensichtlich gilt $\Omega \in \sum_i$ für alle $i \in I$. Also gilt auch $\Omega \in \sum$. Es sei nun $A, B \in \sum$, d.h. $A, B \in \sum_i \forall i \in I$. Also gilt auch $A^c, A \cup B \in \sum_i \forall i \in I$. Daher muß auch $A^c, A \cup B \in \sum$ sein.

Ist nun \mathcal{E} ein beliebiges System von Teilmengen einer nichtleeren Menge Ω , so ist offensichtlich

$$\sum(\mathcal{E}) = \bigcap \sum, \quad \sum \text{ ist Mengenalgebra, und } \mathcal{E} \subset \sum$$

die kleinste Mengenalgebra, die \mathcal{E} enthält.

Definition 1.1 $\sum(\mathcal{E})$ ist die von \mathcal{E} erzeugte Mengenalgebra.

Wir haben schon die Bedeutung der *natürlichen Ereignisse* hervorgehoben, denn es sind genau solche Ereignisse, die im Rahmen eines Modells eine besonders einfache Interpretation haben. Diese sollten daher unbedingt zu \sum gehören.

Wir fordern daher zusätzlich für das System \mathcal{E} der natürlichen Ereignisse

$$\sum_4 : \mathcal{E} \subset \sum.$$

Eine unmittelbare Konsequenz davon ist

$$\sum(\mathcal{E}) \subseteq \sum.$$

Unendliche und stetige Systeme stellen Idealisierungen tatsächlicher Gegebenheiten dar. Man wird daher auch noch zusätzliche "Stetigkeitsbedingungen" bei unendlichen Wahrscheinlichkeitsmodellen fordern. Von solcher Art ist

$$\sum_5 : A_1, A_2, \dots \in \sum, A_1 \subset A_2 \subset A_3 \dots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \sum.$$

\sum_5 bedeutet gerade, daß \sum auch stabil ist bezüglich des Grenzprozesses $\cup A_i$, wenn $A_i \uparrow$.

Lemma 1.2 Gilt für eine Mengenalgebra \sum das Axiom \sum_5 so ist \sum abgeschlossen unter abzählbaren Durchschnitten und Vereinigungen, d.h. insbesondere gilt für $A_1, A_2, \dots \in \sum$

$$\text{i) } \bigcup_{i=1}^{\infty} A_i \in \sum$$

$$\text{ii) } \bigcap_{i=1}^{\infty} A_i \in \sum$$

$$\text{iii) } \underline{\lim} A_i = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i \in \sum$$

$$\text{iv) } \overline{\lim} A_i = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \in \sum$$

Beweis: Es gilt $B_i = (A_1 \cup \dots \cup A_i) \uparrow$ d.h. $B_i \subset B_{i+1}$ und $B_i \in \sum$. Wegen \sum_5 gilt dann $\cup A_i = \cup B_i \in \sum$. Es ist $\bigcap_{i=1}^{\infty} A_i = ((\cap A_i)^c)^c = (\cup A_i^c)^c$.

Nach dem eben gezeigten, gehört aber A_i^c und $\cup A_i^c$ zu \sum . Also gilt auch $\cap A_i \in \sum$. Die restlichen Aussagen sind nun trivial.

Aufgabe 1.10 Man zeige: Für eine Mengenalgebra \sum sind äquivalent:

$$\sum_5 : A_i \in \sum, A_i \uparrow \quad \Rightarrow \quad \bigcup_i A_i \in \sum$$

$$\sum_{5_1} : A_i \in \sum \quad \Rightarrow \quad \bigcup_i A_i \in \sum$$

$$\sum_{5_2} : A_i \in \sum, A_i \downarrow \text{ d.h. } A_{i+1} \subset A_i \quad \Rightarrow \quad \bigcap_i A_i \in \sum$$

$$\sum_{5_3} : A_i \in \sum \quad \Rightarrow \quad \bigcap_i A_i \in \sum$$

$$\sum_{5_4} : A_i \in \sum, A_i \text{ disjunkt} \quad \Rightarrow \quad \bigcup_i A_i = \sum_{i=1}^{\infty} A_i \in \sum.$$

Definition 1.2 Eine Mengenalgebra \sum , die \sum_5 bzw. eine der dazu äquivalenten Bedingungen $\sum_{5_i}, i = 1, \dots, 4$, erfüllt, heißt σ -Algebra.

Aufgabe 1.11 Es sei Ω eine nichtleere Menge. Dann sind

$$\text{a) } \sum_0 = \{\Omega, \emptyset\}$$

$$\text{b) } \sum_0 = \{A \subset \Omega \mid A \text{ oder } A^c \text{ ist endlich}\}$$

Mengenalgebren.

$$\text{c) } \mathcal{P}(\Omega) = \{A \mid A \subset \Omega\} \text{ die Potenzmenge von } \Omega$$

$$\text{d) } \sum_0 = \{A \mid A \subset \Omega, A \text{ oder } A^c \text{ ist abzählbar}\}$$

sind sogar σ -Algebren.

Aufgabe 1.12 Sind $\sum_i, i \in I$, σ -Algebren über Ω , so ist auch $\sum = \bigcap_{i \in I} \sum_i$ eine σ -Algebra.

Aufgabe 1.13 Es sei Ω eine nichtleere Menge und $A_1, \dots, A_n \subset \Omega$. Für $\mathcal{E} = \{A_1, \dots, A_n\}$ zeige, daß $\sum(\mathcal{E})$ endlich ist und schätze $|\sum(\mathcal{E})|$, die Anzahl von $\sum(\mathcal{E})$ ab.

1.3 Die Wahrscheinlichkeit

In 1.2 haben wir den zufälligen Ereignissen Elemente der σ -Algebra \sum der Ereignisalgebra zugewiesen. Es gilt nun noch die Wahrscheinlichkeit für jedes dieser Ereignisse festzulegen. Wir werden uns dabei am Begriff der empirischen Wahrscheinlichkeit orientieren.

Nehmen wir also an, A sei ein Ereignis bei einem Zufallsexperiment. Führen wir dieses Experiment n -mal unabhängig durch und beobachten dabei n_A -mal A , so haben wir mit

$$P_n(A) = \frac{n_A}{n}$$

die *relative Häufigkeit* von A bezeichnet. Es gilt nun offensichtlich:

$$P_n(\Omega) = 1, \quad P_n(\emptyset) = 0, \quad 0 \leq P_n(A) \leq 1.$$

Sind A und B disjunkte Ereignisse, so sieht man auch leicht

$$P_n(A + B) = P_n(A) + P_n(B).$$

Für die empirische Wahrscheinlichkeit $P(A) = \lim_n P_n(A)$ hat dies die folgenden Konsequenzen:

$$P_1: P(A) \geq 0$$

$$P_2: P(\Omega) = 1$$

$$P_3: P(A + B) = P(A) + P(B)$$

Durch Induktion folgt hieraus ohne große Schwierigkeit

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

In Zukunft wollen wir für die Wahrscheinlichkeit stets P_1 , P_2 und P_3 voraussetzen, auch wenn es sich nicht um empirische Wahrscheinlichkeit, wie in den folgenden Beispielen handelt.

Beispiel 1.20 Bei einmaligem Würfeln mit einem „fairen“ Würfel sind alle Zahlen gleichwahrscheinlich. Da $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $P(\{1\}) = P(\{6\})$ muß also wegen P_2 und P_3 $P(\{i\}) = 1/6$ sein.

Beispiel 1.21 Zieht man aus einem Skatspiel jeweils 1 Karte, so kommt jede Karte gleichwahrscheinlich dran. Wir haben also $\Omega = \{1, 2, \dots, 32\}$, und die gleiche Überlegung, wie in Beispiel 1.20 zeigt $P(\{i\}) = \frac{1}{32}$.

Beispiel 1.22 Es wird ein Paar fairer Würfel geworfen. Wie groß ist die Wahrscheinlichkeit, daß die Augensumme 7 ist? Da wir annehmen, daß die Würfel fair sind, ist jedes Ergebnis (m, n) , $m, n = 1, \dots, 6$, gleichwahrscheinlich. Da es insgesamt 36 solcher Paare gibt, tritt jedes Paar (m, n) mit der Wahrscheinlichkeit $1/36$ ein. Gefragt aber ist nach dem Eintreffen des Ereignisses $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Wegen P_3 ist die Wahrscheinlichkeit dafür $\frac{6}{36} = \frac{1}{6}$.

Aufgabe 1.14 3 faire Münzen werden geworfen. Berechne die Wahrscheinlichkeit für die folgenden Ereignisse:

- A = Es fällt genau 1 mal Kopf
 B = Es fällt genau 2 mal Kopf
 C = Es fällt höchstens 1 mal Kopf.

Aufgabe 1.15 Wie groß ist beim 2-maligen Würfeln mit einem fairen Würfel die Wahrscheinlichkeit, daß die Augensumme ≤ 8 (9, 10, 11, 12) ist?

Beim Münzenwerfen wird man bei einer halbwegs ebenmäßigen Münze, diese als fair ansehen können. Dies ist tatsächlich der Fall, wie die Experimente der Mathematiker Buffon und Pearson gezeigt haben (k = Kopf).

	n	n_k	P_n
Buffon	4040	2048	0,5080
K. Pearson	12000	6019	0,5016
K. Pearson	24000	12012	0,5005

Das Modell, das das Münzwerfen beschreibt, hat daher die Form

$$\Omega = \{K, Z\}, \quad \sum = P(\Omega), \quad P(\{K\}) = 1/2 = P(\{Z\}).$$

Bei unendlichen Wahrscheinlichkeitsmodellen wird im Allgemeinen P_1 , P_2 und P_3 nicht ausreichen um P zu bestimmen und um für P ein Kalkül zu entwickeln. Wir brauchen also noch eine weitere Annahme, die die Stetigkeit von P bezüglich der durch \sum_5 zugelassenen Grenzoperationen garantiert. Sinnvoll ist hier

$$P_4 : A_i \in \sum, A_i \uparrow \Rightarrow P(A_i) \uparrow P(\bigcup_{i=1}^{\infty} A_i).$$

Man könnte auch

$$P_{4_1} : A_i \in \sum \text{ disjunkt} \Rightarrow \sum_{i=1}^{\infty} P(A_i) = P(\sum_{i=1}^{\infty} A_i)$$

oder

$$P_{4_2} : A_i \in \sum, A_i \downarrow \Rightarrow P(A_i) \downarrow P(\bigcap_{i=1}^{\infty} A_i)$$

fordern. Wie wir sehen werden, sind diese Bedingungen äquivalent.

Lemma 1.3 Die Bedingungen P_4, P_{4_1}, P_{4_2} sind äquivalent.

Beweis: $P_4 \Rightarrow P_{4_1}$: Es seien A_1, A_2, \dots disjunkt. Dann haben wir

$$\sum_{i=1}^n P(A_i) = P\left(\sum_{i=1}^n A_i\right) \uparrow P(\bigcup_{i=1}^{\infty} A_i)$$

und daraus folgt die Behauptung

$P_{4_1} \Rightarrow P_4$: Es sei $B_i = A_i \setminus A_{i-1}$, $A_0 = \emptyset$. Dann sind die B_i disjunkt und $A_n = \bigcup_{i=1}^n B_i$. Dies gibt

$$P(A_n) = \sum_{i=1}^n P(B_i) \uparrow \sum_{i=1}^{\infty} P(B_i) = P\left(\sum_{i=1}^{\infty} B_i\right) = P(\cup A_n).$$

Für die restliche Behauptung verwende Komplemente.

Definition 1.3 Es sei Ω eine nichtleere Menge und Σ eine σ -Algebra über Ω . Eine Funktion $P : \Sigma \rightarrow \mathbb{R}$ heißt *Wahrscheinlichkeitsmaß* (W.Maß), wenn $P1, P2, P3, P4$ gilt.

Definition 1.4 Ein Tripel (Ω, Σ, P) heißt *Wahrscheinlichkeitsraum* (W.Raum), wenn Ω eine nichtleere Menge ist, Σ eine σ -Algebra über Ω ist und P ein Wahrscheinlichkeitsmaß auf Σ ist.

Beispiel 1.23 Eine faire Münze werde so häufig geworfen, bis zum erstenmal Kopf fällt. Es ist hier

$$\Omega = \{K, ZK, ZZK, ZZZK, \dots\}.$$

Schreiben wir für

$$\underbrace{Z \dots Z}_{n\text{-mal}} K = Z^n K,$$

so gilt offensichtlich

$$P(Z^n K) = 2^{-(n+1)},$$

denn die Reihenfolge $Z \dots ZK = Z^n K$ ist eine von 2^{n+1} gleichwahrscheinlichen Möglichkeiten. Sicherlich definiert die Festlegung $P(Z^n K) = 2^{-(n+1)}$ zusammen mit P_3 ein Wahrscheinlichkeitsmaß auf Ω , wenn man

$$P(A) = \sum_{Z^n K \in A} P(Z^n K) = \sum_{Z^n K \in A} 2^{-(n+1)}$$

definiert. Denn dann gilt trivialerweise $P1$. Ferner ist

$$P(\Omega) = \sum_{n=0}^{\infty} P(Z^n K) = \sum_{n=0}^{\infty} 2^{-(n+1)} = 1.$$

Die Wahrscheinlichkeit etwa, daß das Spiel vor dem 5. Wurf abbricht, ist

$$P(\{K, ZK, ZZK, ZZZK\}) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16}.$$

Wir haben bisher die Eigenschaften von P und Σ so festgelegt, daß sie aufeinander bezogen sind. Dies betrifft insbesondere die Stetigkeitsforderungen Σ_5 und P_4 . Natürlich könnte man auch Σ_5 durch

$$\Sigma_{5'} : A_i \in \Sigma, i \in I \Rightarrow \cup A_i \in \Sigma.$$

Dann müßte man entsprechend auch

$$P_{4'} : A_i \uparrow \Rightarrow P(A_i) \uparrow P(\cup A_i)$$

fordern. Daß dies zu einem Widerspruch führt, sieht man aus folgendem

Beispiel 1.24 Drehen eines Glücksrades. Die Position eines Glücksrades wird durch den Winkel α bestimmt. Es ist also $\Omega = [0, 2\pi)$. Ist A ein Winkelintervall, so werden wir

$$P(A) = \frac{\text{Bogenlänge von } A}{2\pi} \text{ setzen.}$$

Wenn nun $\sum = \mathcal{P}(\Omega)$ und wenn P_4 für beliebige Vereinigungen gilt, so haben wir für $A \in \sum$

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} 0 = 0.$$

Das aber ist unmöglich.

Also kann man die Axiome \sum_5 und P_4 nicht durch die wesentlich stärkeren Annahmen, nämlich $\sum_{5'}$ und $P_{4'}$ ersetzen.

Aufgabe 1.16 Eine faire Münze werde solange geworfen, bis zweimal das gleiche Ergebnis hintereinander fällt. Beschreibe den Ereignisraum und gebe die Wahrscheinlichkeit für die Elementarereignisse an.

Definition 1.5 Ein Wahrscheinlichkeitsraum (Ω, \sum, P) heißt *diskret (endlich)*, wenn Ω abzählbar (endlich) ist.

Satz 1.1 Ein Wahrscheinlichkeitsmaß P eines diskreten Wahrscheinlichkeitsraumes (Ω, \sum, P) mit $\sum = \mathcal{P}(\Omega)$ ist eindeutig durch seine Werte $P_\omega = P(\{\omega\})$ auf den Elementarereignissen bestimmt.

Beweis: Es gilt für $A \subset \Omega$

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

wegen P_{4_2} .

Satz 1.2 Ist umgekehrt Ω eine abzählbare Menge und $p : \Omega \rightarrow \mathbb{R}$ eine Funktion, so daß $p(\omega) \geq 0$ und $\sum_{\omega \in \Omega} p(\omega) = 1$, so wird durch

$$P(A) = \sum_{\omega \in A} p(\omega)$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$ definiert.

Beweis: Definiert man nun $P(A)$ wie oben, so gilt offensichtlich

$$P(A) \geq 0, \quad P(\Omega) = 1 \text{ und } P(A + B) = P(A) + P(B).$$

Es seien nun A_1, \dots disjunkt. Dann gilt

$$\sum_i P(A_i) = \sum_i \sum_{\omega_j \in A_i} p(\omega_j) = \sum_{\omega_j \in \cup A_i} p(\omega_j) = P(\cup A_i),$$

da die Reihe rechts absolut konvergiert und da absolut konvergente Reihen beliebig umgeordnet werden dürfen.

Aufgabe 1.17 Es sei Σ eine Mengenalgebra und $P : \Sigma \rightarrow \mathbb{R}$ eine Funktion, die P_1, P_2 und P_3 erfüllt. Dann gilt für $A, B, \dots \in \Sigma$

- i) $A \subset B \Rightarrow P(A) \leq P(B)$
- ii) $P(A^c) = 1 - P(A)$
- iii) $P(A \cup B) + P(A \cap B) = P(A) + P(B)$
- iv) $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_i P(A_i)$

Aufgabe 1.18 Es sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und $A, B, C \in \Sigma$. Dann gilt $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

Beispiel 1.25 Um den Satz anzuwenden, berechnen wir, wie groß die Wahrscheinlichkeit ist, aus einem Skatspiel eine Kreuz-Karte oder ein As zu ziehen. Dann ist etwa $\Omega = \{1, 2, \dots, 32\}$.

$$\begin{aligned} A &= \text{Menge aller Kreuzkarten} &= \{1, \dots, 8\} \\ B &= \text{Menge aller Asse} &= \{1, 9, 17, 25\}. \end{aligned}$$

Ferner gilt $p(\{1\}) = \dots = P(\{32\}) = \frac{1}{32}$. Also haben wir

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{32}[8 + 4 - 1] = \frac{11}{32}.$$

1.4 Wahrscheinlichkeitsmodelle

Mit den in 1.1, 1.2 und 1.3 entwickelten Begriffen können wir das Modellieren von Zufallssystemen im einfachsten Fall wie folgt beschreiben.

Bestimme für das gegebene Zufallsereignis einen geeigneten Wahrscheinlichkeitsraum.

Um dies zu tun, muß man zunächst entscheiden, ob das Modell diskret oder stetig sein soll, wenn dies nicht schon ohnehin klar ist. Dann kann man sich an folgendem Schema orientieren.

System	Modell
Elementarereignisse	Punkte aus Ω , dem Ereignisraum
natürliche Ereignisse	natürliche Mengen, \mathcal{E}
Ereignisalgebra	$\Sigma = \Sigma(\mathcal{E})$
Wahrscheinlichkeit	$P(A), \quad A \in \Sigma$

Diskrete Modelle

Ω = Raum der Elementarereignisse
natürliche Ereignisse = Elementarereignisse

Σ = $\Sigma(\mathcal{E}) = \mathcal{P}(\Omega)$

P ist durch seine Werte auf den Elementarereignissen $P(\{\omega\})$ eindeutig bestimmt.

Stetige Modelle

Ω = Raum der Elementarereignisse, meist gilt $\Omega \subset \mathbb{R}^n$.

Natürliche Elementarereignisse – Quader, Kugeln im \mathbb{R}^n .

$\Sigma = \mathcal{L}(\Omega)$ = Menge der Borelmengen

Bei den stetigen Modellen ist nun die Bestimmung von P wesentlich schwieriger, da man insbesondere davon ausgehen muß, daß P unmittelbar nur auf \mathcal{E} definiert werden kann. Die Definition von P auf Σ vollzieht man allgemein in folgenden Schritten

P auf	\mathcal{E}	vorgegeben
	↓	
Festlegen von	P	auf $A(\mathcal{E})$, der von \mathcal{E} erzeugten Mengenalgebra
	↓	
Festsetzen von	P	auf $\Sigma(\mathcal{E})$, der von \mathcal{E} erzeugten σ -Algebra.

Während die Bestimmung von P auf $A(\mathcal{E})$ unproblematisch ist, erfordert die Fortsetzung von P auf $\Sigma(\mathcal{E})$ einigen mathematischen Aufwand, wie etwa einen Fortsetzungs- und Eindeutigkeitssatz (siehe Kapitel 3), insbesondere muß hier an P eine gewisse Stetigkeitsforderung gestellt werden.

An dieser Stelle erhebt sich natürlich die Frage, warum man nicht schließlich $\Sigma = \mathcal{P}(\Omega)$ wie im diskreten Fall erreichen kann. Daß dies im Allgemeinen nicht möglich ist, hängt mit der Existenz nichtmeßbarer Mengen in \mathbb{R} zusammen [vgl. hierzu H. Bauer Wahrscheinlichkeitstheorie und Maßtheorie § 8]. Nichtmeßbare Mengen sind dabei Mengen, deren Existenz mit $P4$ unverträglich ist.

Die Konstruktion eines Wahrscheinlichkeitsmodells für ein Zufallssystem reduziert sich daher im einfachsten Fall auf die Festlegung von P auf \mathcal{E} .

Kapitel 2

Geometrische und Diskrete Modelle

2.1 Geometrische Modelle

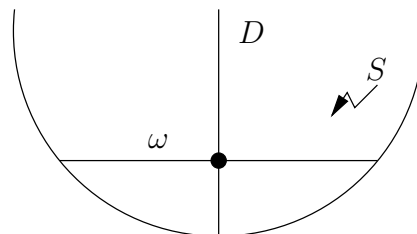
In diesem Kapitel wird das Wahrscheinlichkeitsmaß P in den Modellen als a priori Wahrscheinlichkeit festgelegt. Dies ist nicht immer unproblematisch, wie das Bertrand'sche Paradox zeigt.

Bertrand'sches Paradox:

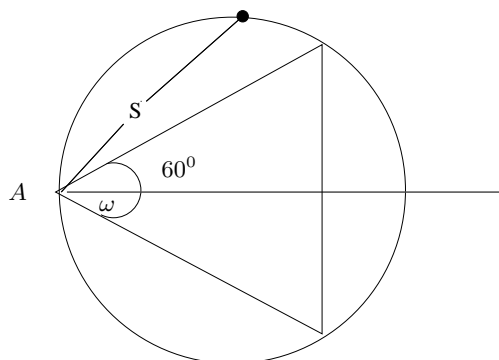
In einen Einheitskreis wird willkürlich eine Sehne S gezogen. Wie groß ist die Wahrscheinlichkeit, daß die Sehne länger ist als die Seite eines dem Kreis einbeschriebenen gleichseitigen Dreiecks?

1. *Lösung:* Aus Symmetriegründen können wir die Richtung der Sehne S als fest annehmen. Zeichne nun senkrecht zu S den Durchmesser D . Wir können dann Ω mit der Menge der Schnittpunkte von S und D identifizieren, $\Omega = [0, 2]$. Man sieht dann leicht, daß S die geforderte Bedingung erfüllt, wenn S D zwischen $\frac{1}{2}$ und $\frac{3}{2}$ schneidet. Wir haben also $P([\frac{1}{2}, \frac{3}{2}])$ zu bestimmen. Jeder Schnittpunkt kommt aber gleichmäßig häufig vor. Also gilt

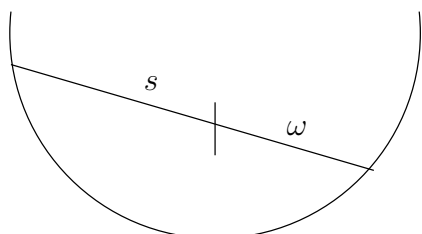
$$P\left(\frac{1}{2}, \frac{3}{2}\right) = \frac{\text{Länge von } [\frac{1}{2}, \frac{3}{2}]}{\text{Länge von } D} = \frac{1}{2}.$$



2. *Lösung:* Aus Symmetriegründen können wir einen der Schnittpunkte der Sehne mit dem Kreis A fixieren. Die Lage der Sehne wird dann eindeutig durch den Winkel $\omega \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ bestimmt. Der gewünschte Winkelsektor A wird dann gerade durch den schraffierten Bereich gegeben. Argumentiert man wie eben, sieht man $P(A) = \frac{1}{3}$, denn dieser Winkelsektor hat den Öffnungswinkel 60° .



3. Lösung:



Die Lage der Sehne wird eindeutig durch die Position des Sehnenmittelpunktes bestimmt. ω kann Werte im Einheitskreis annehmen. Das gesuchte Ereignis tritt offensichtlich dann ein, wenn ω weniger als $\frac{1}{2}$ vom Kreismittelpunkt weg ist.

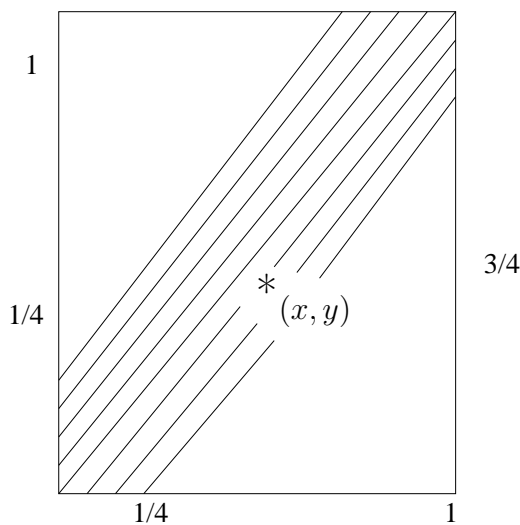
Wir haben damit

$$P(A) = \frac{\text{Fläche eines Kreises mit Radius } \frac{1}{2}}{\text{Fläche des Einheitskreises}} = \frac{1}{4}.$$

Die verschiedenen Ergebnisse haben ihren Ursprung offensichtlich darin, daß wir hier 3 *verschiedene* Modelle für den gleichen Vorgang erstellt haben. Willkürlich bzw. zufällig wurde dann jeweils so interpretiert, daß P die Gleichverteilung in dem entsprechenden Modell ist.

Das Verabredungsproblem:

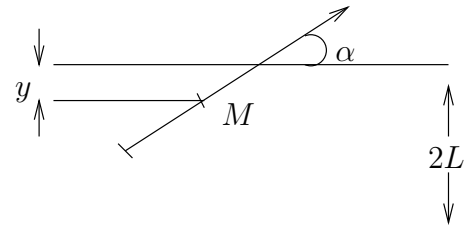
Zwei Personen A, B verabreden sich zwischen 16.00 und 17.00 Uhr am Postamt. Jeder will genau 15 Minuten warten. Wie groß ist die Wahrscheinlichkeit, daß sie sich treffen, wenn die Ankunftszeit eines jeden willkürlich ist? Jedes Elementarereignis ist hier (x, y) . Dabei bezeichnet x die Ankunftszeit von A in Stunden nach 16.00 Uhr und y die entsprechende Ankunftszeit von B . Wir können also Ω als $[0, 1] \times [0, 1]$ interpretieren. Willkürlich bedeutet in diesem Fall $P(C) = \text{Fläche von } C$, nämlich die Gleichverteilung auf Ω . A und B treffen sich, wenn ihre Ankunftszeiten in dem schraffierten Bereich liegen. Die Wahrscheinlichkeit, daß sich beide treffen, ist damit $1 - (\frac{3}{4})^2 = \frac{7}{16}$.



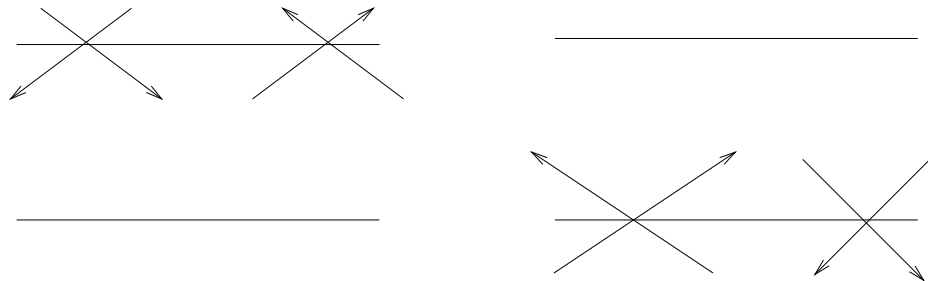
Buffon'sches Nadelproblem

Eine dünne Nadel der Länge 2ℓ werde zufällig auf einen Fußboden geworfen, der aus langen Brettern der Breite $2L$ besteht. Wie groß ist die Wahrscheinlichkeit, daß die Nadel über einer Fußbodenritze liegt?

Die Lage der Nadel in Bezug auf unser Problem wird eindeutig durch den Abstand y des Nadelmittelpunktes M von der nördlichen Ritze und dem Winkel α bestimmt (Zeichnung), da die horizontale Richtung (x Richtung) für unser Problem offensichtlich irrelevant ist.



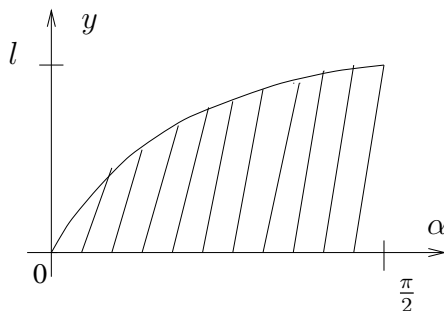
Die folgenden 8 Positionen der Nadel gehen auseinander offensichtlich durch Symmetrieabbildungen hervor. Wir haben daher



$$P = 8 \frac{\text{Fläche der Treffer, wenn } 0 \leq \alpha \leq \frac{\pi}{2} \text{ und es wird die Nordritze getroffen}}{\text{Gesamtfläche}}$$

Ein Treffer liegt offensichtlich genau dann vor, wenn $\ell \cdot \sin \alpha \geq y$.

Die dazugehörige Fläche ist schraffiert gezeichnet. Ihr Flächeninhalt ist l . Damit haben wir



$$P = 8 \frac{l}{2\pi} \cdot 2a = \frac{2\ell}{\pi a}.$$

Durch das Buffon'sche Experiment ließe sich also experimentell π bestimmen, wenn das Modell richtig ist. Tatsächlich ist die Übereinstimmung mit Experimenten recht gut.

n	Exp. Wert von π	
5000	3,1596	Wolf (1850)
3204	3,1553	Smith (1855)

Aufgabe 2.1 Eine Münze vom Durchmesser d werde willkürlich auf einen Fußboden geworfen, der aus langen Brettern der Breite L besteht. Wie groß ist die Wahrscheinlichkeit, daß die Münze auf einer Ritze liegt?

Aufgabe 2.2 Eine Münze vom Durchmesser d werde willkürlich auf einen quadratisch gefliesten Fußboden geworfen. Die Fliesen haben die Maße $L \times L$. Wie groß ist die Wahrscheinlichkeit, daß die Münze auf einer (zwei) Ritzen liegt?

2.2 Diskrete endliche Modelle

In diesem Abschnitt wollen wir uns mit endlichen Wahrscheinlichkeitsmodellen beschäftigen, die hauptsächlich zur Beschreibung von Zufallsgeräten dienen. Die meisten Zufallsgeräte sind so gebaut, daß jedes Elementarereignis die gleiche Wahrscheinlichkeit hat. Solche Zufallsgeräte werden wir als *fair* bezeichnen. Etwas allgemeiner bedeutet fair, daß a priori Wahrscheinlichkeit und empirische Wahrscheinlichkeit übereinstimmen.

Ist nun $\Omega = \{\omega_1, \dots, \omega_n\}$ der Ereignisraum für ein solches faires Zufallsgerät, so ist $\sum = \mathcal{P}(\Omega)$. Da nach Annahme $P(\{\omega_1\}) = p_1 = \dots = P(\{\omega_i\}) = p_i$ und $\sum p_i = 1$ muß $p_i = \frac{1}{n}$ sein. Damit ist aber der Wahrscheinlichkeitsraum schon vollständig beschrieben. Die Wahrscheinlichkeit $P(A)$ für $A \subset \Omega$ wird dann

$$P(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega} = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl aller Fälle}}. \quad (2.1)$$

Ein solches Modell nennen wir einen *Laplace'schen Wahrscheinlichkeitsraum* mit n Elementen. In solchen Modellen reduziert sich die Berechnung von $P(A)$ nach (??) meist auf ein kombinatorisches Problem. Solche Fragestellungen wollen wir im folgenden behandeln. Dabei bezeichne $|A|$ die Mächtigkeit der Menge A und $A \times B$ das kartesische Produkt von A und B .

Lemma 2.1 Es seien A_1, \dots, A_k endliche Mengen. Dann gilt

$$|A_1 \times \dots \times A_k| = |A_1| \cdot \dots \cdot |A_k|.$$

Beweis: Induktion nach k .

Beispiel 2.1 In einem Skatenspiel gibt es vier Farben:

$A = \{\text{Kreuz, Pik, Herz, Karo}\}$ und 8 Werte

$B = \{\text{As, 10, König, Dame, Bube, 9, 8, 7}\}$.

Also enthält ein Spiel $|A \times B| = 4 \cdot 8$ Karten.

Beispiel 2.2 Klassifiziert man Personen nach Geschlecht, Religion (3 Religionen), politischer Entscheidung (4 Möglichkeiten) und nach Beruf (15 Berufe), so gibt es $2 \cdot 3 \cdot 4 \cdot 15 = 360$ Möglichkeiten.

Aufgabe 2.3

- a) In der Stadt X haben die Autokennzeichen 2 Buchstaben und 3 Ziffern. Wiviel Autos könnte man dort registrieren lassen?

- b) Es bezeichne Ω den Ereignisraum für sechsmaliges Würfeln. Berechne $|\Omega|$.
- c) Eine Münze wird zehnmal geworfen. Wieviel Ergebnismöglichkeiten gibt es?

Aufgabe 2.4 Wieviel Spielmöglichkeiten gibt es beim 12er Toto?

Würfelt man dreimal nacheinander, so gibt es $6 \cdot 6 \cdot 6 = 216$ Möglichkeiten. Mit Hilfe dieses Satzes lassen sich bereits einfache Beispiele berechnen. Wir wollen die Beispiele durchrechnen, die zur Begründung der Wahrscheinlichkeitstheorie führten. Im 17. Jahrhundert wurde folgendes Würfelspiel gespielt: Man würfelte viermal hintereinander und die Bank gewann, wenn der Spieler mindestens eine Sechs würfelte. Lohnte sich das Geschäft für die Bank?

Als Ereignisraum Ω wählen wir die Menge der geordneten 4-Tupel der Zahlen 1, 2, 3, 4, 5, 6. Nach obigem Satz besteht er aus 6^4 Elementen. Jedes Elementarereignis hat die Wahrscheinlichkeit $\frac{1}{6^4}$. In dem Beispiel ist es einfacher, die Wahrscheinlichkeit des Gegenereignisses: „der Spieler gewinnt“ bzw. „in allen 4 Würfeln kommt keine 6 vor“ zu berechnen. Dieses Ereignis besteht aus allen 4-Tupel der Zahlen 1, 2, 3, 4, 5, enthält also 5^4 Elementarereignisse. Also ist P („der Spieler gewinnt“) $= \frac{5^4}{6^4} \approx 0,482$, P („die Bank gewinnt“) $= 1 - \frac{5^4}{6^4} \approx 0,518$. In mehr als der Hälfte aller Fälle wird also die Bank gewinnen.

Der Chevalier de Méré hatte sich ein ähnliches Spiel ausgedacht: Man würfelt 24 mal mit zwei Würfeln und gewinnt, wenn man keine doppelte Sechs wirft. Er meinte, die Chancen seien die gleichen, wie bei dem ersten Spiel. Da in Wirklichkeit aber bei diesem Spiel der Spieler öfter gewann als die Bank, wandte er sich an den Mathematiker Pascal. Wollen wir die Chancen bei diesem Spiel berechnen, so wählen wir als Ereignisraum alle 24-Tupel von Paaren der Zahlen 1, 2, 3, 4, 5, 6. Er hat $(6 \cdot 6)^{24}$ Elemente. Günstig für den Spieler sind alle 24-Tupel von Paaren außer dem Paar (6,6), also $(35)^{24}$ Fälle. Die Wahrscheinlichkeit des Gewinns des Spielers ist $\frac{35^{24}}{36^{24}} \approx 0,508$. Die Chancen sind für den Spieler also günstiger als für die Bank.

Permutation: Das Lemma läßt sich auch so interpretieren: *k aufeinanderfolgende Entscheidungen (Auswahlverfahren) mit genau n_i Möglichkeiten beim i -ten Schritt führen zu $n_1 \cdot n_2 \cdot \dots \cdot n_k$ verschiedenen Ereignissen.*

Wir wollen dieses Ergebnis auf folgenden Fall anwenden: Es sei ein Topf mit n verschiedenen Kugeln $\{a_1, \dots, a_n\}$ gegeben. Wir wollen wissen, auf wieviele Weisen man r mal, $r \leq n$, Kugeln aus diesem Topf ziehen kann. Wir müssen dabei die folgenden Fälle unterscheiden:

- i) Nachdem eine Kugel gezogen ist, wird nur das Ergebnis notiert und die Kugel wird wieder in den Topf gelegt (Auswahl mit Wiederholung)
- ii) Ist eine Kugel einmal gezogen, wird sie nicht wieder zurückgelegt (Auswahl ohne Wiederholung)

Dieses Problem führt auf folgende Definition.

Definition 2.1 Es sei $A = \{a_1, \dots, a_n\}$ eine Menge mit n Elementen. Jedes geordnete r -Tupel $(a_1, \dots, a_r) \in A \times \dots \times A = A^r$ heißt eine *r -Permutation von A mit Wiederholung*. Eine r -Permutation, die aus lauter verschiedenen Elementen besteht, heißt *r -Permutation ohne Wiederholung*.

Früher wurden r -Permutationen auch als Variationen bezeichnet.

Satz 2.1 Es sei $A = \{a_1, \dots, a_n\}$. Dann gibt es genau n^r r -Permutationen mit Wiederholung $r = 1, 2, \dots$ und $n(n-1) \cdots (n-r+1)$ r -Permutationen ohne Wiederholung $r = 1, \dots, n$.

Beweis:

- Jede r -Permutation mit Wiederholung beschreibt genau ein Element aus A^r und umgekehrt ist jedes Element aus A^r eine r -Permutation mit Wiederholung. Wegen des Satzes (2.1) gilt $|A^r| = |A|^r = n^r$.
- Jede r -Permutation ohne Wiederholung entspricht genau einem Auswahlverfahren von r -Elementen aus einem Topf, wobei die Elemente nicht zurückgelegt werden. Beim ersten Schritt hat man genau n Möglichkeiten. Beim zweiten Schritt nur noch $n - 1$ Möglichkeiten. Beim dritten Mal sind es nur noch $n - 2$ Möglichkeiten. Beim r -ten Mal haben wir noch $n - r + 1$ Möglichkeiten, so daß es wegen unserer Bemerkung oben insgesamt genau $n(n - 1) \cdots (n - r + 1)$ Möglichkeiten gibt.

Beispiel 2.3 Das Direktorium einer Gesellschaft autorisiert die Gesellschaft 3 von 10 Topmanagern zu befördern und zwar einen zum Vizepräsidenten, einen zum 1. Stellvertreter des Vizepräsidenten und einen zum 2. Stellvertreter. Wieviele Beförderungsvorschläge gibt es?

Offensichtlich entspricht ein solcher Vorschlag genau einer 3 Permutation ohne Wiederholung aus 10 Elementen. Es gibt also genau $10 \cdot 9 \cdot 8$ Möglichkeiten.

Beispiel 2.4 *Das Geburtstagsproblem* – auf einer Party sind r Personen. Wie groß ist die Wahrscheinlichkeit P_r , daß alle an verschiedenen Tagen Geburtstag haben?

Nehmen wir an, daß das Jahr 365 Tage hat, so können wir, wenn wir uns die Leute der Reihe nach aufgestellt denken, dieses Arrangement als eine r -Permutation auffassen. Davon gibt es 365^r aber nur $365 \cdot 364 \cdots (365 - r + 1)$ davon sind r -Permutationen ohne Wiederholung.

Es ist also

$$P_r = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{365 \cdot 364 \cdots (365 - r + 1)}{365^r}$$

r	5	10	20	22	23	30	50	60
P_r	0,973	0,883	0,589	0,524	0,493	0,294	0,030	0,006

Aus der Tafel sehen wir, daß man getrost wetten kann, daß auf einer Party mit r Personen zwei am gleichen Tag Geburtstag haben, sobald $r \geq 23$ ist.

Beispiel 2.5 Es gibt $n! = n(n - 1) \cdots 2 \cdot 1$ Möglichkeiten, n Elemente auf verschiedene Weise anzuordnen, denn jede solche Anordnung entspricht einer n -Permutation ohne Wiederholung.

Beispiel 2.6 Wie groß ist die Wahrscheinlichkeit beim Würfeln mit 6 fairen Würfeln lauter verschiedene Augenzahlen zu erzielen? Antwort $\frac{6!}{6^6}$.

Sieht man bei den r -Permutationen von der Anordnung ab, erhält man den Begriff der r -Kombination.

Definition 2.2 Es sei $A = \{a_1, \dots, a_n\}$ eine Menge mit n -Elementen. Eine Teilmenge von r -Elementen, $r \leq n$, von A wird auch eine r -Kombination von A genannt.

Satz 2.2 Es gibt genau

$$\binom{n}{r} = \frac{n(n-1)\dots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

r -Kombinationen einer n -elementigen Menge.

Beweis: Es sei $C_{n,r}$ die Anzahl aller r -Kombinationen. Ordnen wir noch alle dieser Kombinationen auf die $r!$ verschiedenen Weisen an, so erhalten wir alle r Permutationen ohne Wiederholung. Also gilt $n(n-1)\dots(n-r+1) = C_{n,r}r!$ und daraus folgt das Ergebnis.

Beispiel 2.7 Es gibt $\binom{49}{6}$ Spielmöglichkeiten beim „6 aus 49“ Lotto.

Beispiel 2.8 Es gibt $\binom{32}{10}$ Möglichkeiten 10 Karten aus einem Skatspiel zu ziehen. Wie groß ist die Wahrscheinlichkeit, 4 Buben zu haben? Um 4 Buben zu haben, können die anderen 6 Karten aus den restlichen 28 willkürlich gewählt werden. Davon gibt es $\binom{28}{6}$ Möglichkeiten. Die gesuchte Wahrscheinlichkeit ist also

$$\frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl aller Fälle}} = \frac{\binom{28}{6}}{\binom{32}{10}} = 0,00595.$$

Also sind etwa in jedem 60. Skatspiel alle Buben in einer Hand, da es in einem Skatspiel 3 Spieler gibt.

Aufgabe 2.5 Wie groß ist die Wahrscheinlichkeit beim Skatspiel, daß der Spieler (10 Karten + 2 Karten Skat) 4 Buben hat?

Beispiel 2.9 Ein Losverkäufer hat 100 Lose in seinem Topf. Davon sind 90 Nieten und 10 Gewinne. Wie groß ist die Wahrscheinlichkeit beim Kauf von r Losen mindestens 1 Gewinnlos zu haben? Bei diesem Beispiel ist es günstiger zu fragen, wie groß die Wahrscheinlichkeit ist, *kein* Gewinnlos zu haben. Insgesamt gibt es $\binom{100}{r}$ Möglichkeiten, r Lose zu ziehen.

Aber es gibt nur $\binom{90}{r}$ Möglichkeiten, nur Nieten zu ziehen. Also ist die Wahrscheinlichkeit, nur Nieten zu ziehen

$$\frac{\binom{90}{r}}{\binom{100}{r}} = \frac{90 \cdot 89 \dots (90-r+1)}{100 \cdot 99 \dots (100-r+1)}.$$

Beim Kauf von r Losen hat man mit der Wahrscheinlichkeit $1 - \frac{90 \dots (90-r+1)}{100 \dots (100-r+1)}$ mindestens ein Gewinnlos. Für $r = 2$ erhält man etwa $1 - \frac{90 \cdot 89}{100 \cdot 99} = 1 - \frac{89}{110} = \frac{21}{110}$.

Beispiel 2.10 Von einer Sendung von 100 Schaltern weiß man, daß sie 10 unbrauchbare enthält. Wie groß ist die Wahrscheinlichkeit, daß ein Käufer, der r Schalter kauft, genau s ($s \leq r$) heile hat? Es gibt genau $\binom{100}{r}$ Möglichkeiten, r Schalter auszuwählen. Davon gibt es $\binom{90}{s}$ und $\binom{10}{r-s}$ Möglichkeiten, s heile bzw. $r - s$ kaputte auszuwählen. Also ist die Wahrscheinlichkeit

$$\frac{\binom{90}{s} \binom{10}{r-s}}{\binom{100}{r}}.$$

Beispiel 2.11 Aus einem Schuhschrank mit r Paar Schuhen werden willkürlich r Schuhe ausgewählt. Wie groß ist die Wahrscheinlichkeit, daß mindestens 1 Paar dabei ist? Wir fragen hier besser nach der Wahrscheinlichkeit, daß kein Paar dabei ist. Es gibt insgesamt $\binom{2r}{r}$ Möglichkeiten, r Schuhe aus $2r$ Schuhen auszuwählen. Ferner gibt es 2^r Möglichkeiten, aus r Paaren r verschiedene Schuhe, nämlich von jedem Paar genau einen Schuh, auszuwählen. Also ist die Wahrscheinlichkeit unter den r Schuhen kein Paar zu finden $\frac{2^r}{\binom{2r}{r}}$. Für $r = 2$ etwa erhalten wir

$$\frac{4}{\frac{4 \cdot 3}{2}} = \frac{2}{3}.$$

Bei den r -Kombinationen sind alle Elemente verschieden. Man kann entsprechend aber auch r -Kombinationen mit Wiederholung betrachten, bei denen die Elemente nicht notwendig verschieden sein müssen.

Lemma 2.2 Es sei $A = \{a_1, \dots, a_n\}$. Dann gibt es $\binom{n+r-1}{n}$ verschiedene r -Kombinationen mit Wiederholung.

Beweis: Wir fassen jedes Element von A als einen Schubkasten auf. In jedem Schubkasten legen wir soviele „*“, wie das entsprechende Element in der r -Kombination mit Wiederholung auftritt. Eine r -Kombination wird dann durch

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & & & n-1 & n \\ ***| & *| & & |**| & & & |****| \end{array}$$

eine solche Folge von — und * eindeutig beschrieben. Es gibt insgesamt $r + (n - 1)$ Zeichen, davon r mal * und $n - 1$ mal —.

Wegen Satz (2.2) gibt es daher $\binom{n+r-1}{r}$ verschiedene Zeichenkombinationen.

Beispiel 2.12 Bose-Statistik Es sollen r ununterscheidbare Kugeln auf n Zellen verteilt werden. Aus dem Beweis des Lemmas sieht man, daß es dazu genau $\binom{n+r-1}{r}$ Möglichkeiten gibt.

Beispiel 2.13 Fermion Dirac Statistik Es sollen r ununterscheidbare Teilchen auf n Zellen verteilt werden, so daß in jeder Zelle höchstens ein Teilchen sitzt. Dann muß $n \geq r$ sein, und jedes solche Arrangement wird eindeutig durch die besetzten Zellen beschrieben. Da es $\binom{n}{r}$ solche Teilmengen einer r elementigen Menge gibt, gibt es bei der F.D. Statistik $\binom{n}{r}$ Möglichkeiten.

Beispiel 2.14 Es sei f eine analytische Funktion von n Variablen x_1, \dots, x_n . Dann hat f insgesamt $\binom{n+r-1}{r}$ verschiedene partielle Ableitungen der Ordnung r .

Die Hypergeometrische Verteilung:

In einer Urne seien N Kugeln, davon M rote und $N - M$ weiße. Es werden n Kugeln gezogen. Davon sind k rote mit der Wahrscheinlichkeit

$$p_k = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad 0 \leq k \leq n. \quad (2.2)$$

Es sei nun $\frac{M}{N} = p$, $\frac{N-M}{N} = 1 - p = q$. Dann gilt

$$p_k = \frac{M(M-1)\dots(M-k+1)(N-M)\dots(N-M-n+k+1)}{k!(n-k)!N(N-1)\dots(N-n+1)} n! = \binom{n}{k} \frac{AB}{C}$$

$$\begin{aligned} \text{mit } A &= M(M-1)\dots(M-k+1) \text{ oder } (M-k+1)^k \leq A \leq M^k \\ B &= (N-M)\dots(N-M-n+k+1) \text{ oder } (N-M-n+k+1)^{n-k} \\ &\leq B \leq (N-M)^{n-k} \\ C &= N(N-1)\dots(N-n+1) \text{ oder } (N-n+1)^n \leq C \leq N^n. \end{aligned}$$

Dies gibt

$$\binom{n}{k} \left(p - \frac{k-1}{N}\right)^k \left(q - \frac{n-k-1}{N}\right)^{n-k} \leq p_k \leq p^k q^{n-k} \binom{n}{k} \left(1 - \frac{n-1}{N}\right)^n.$$

Für $N \rightarrow \infty$, $\frac{M}{N} \rightarrow p$ gilt also $p_k \rightarrow p^k q^{n-k} \binom{n}{k}$.

Anwendungen der Hypergeometrischen Verteilung

Beispiel 2.15 *Qualitätstest* In einer Fabrik werden am Tag N Teile hergestellt. Die Anzahl der defekten Teile sei M , M unbekannt. Es werden n Teile zufällig ausgewählt. Dann ist die Wahrscheinlichkeit P_k , genau defekte bzw. $u-k$ funktionsfähige Teile zu erhalten

$$p_k = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \approx p^k (1-p)^{n-k} \binom{n}{k}.$$

Beispiel 2.16 Schätzen der Größe einer unbekanntem Tierpopulation

N Tiere leben in einer isolierten Gegend

M Tiere werden gefangen und markiert.

Nach einiger Zeit werden n Tiere gefangen und k markierte gefunden. Bestimme das wahrscheinlichste N , wenn M , n und k bekannt sind. Dann gilt

$$p_k(N) = \frac{\binom{M}{k} \binom{n-M}{n-k}}{\binom{N}{n}}.$$

Daher ist

$$\frac{p_k(N)}{p_k(N-1)} = \frac{(N-M)(N-n)}{(N-M-n+k)N} \geq 1$$

falls $Nk < Mn$ oder $Nk > Mn$.

Zunächst gilt daher $p_k(N) \nearrow$ als Funktion von N und dann \searrow . Das Maximum liegt ungefähr bei $Nk \sim Mn$ oder $\frac{N}{M} \sim \frac{n}{k}$.

Beispiel 2.17 *Das Hut – Problem* n Hüte werden an der Garderobe abgegeben und leider völlig durcheinander gebracht. Was ist die Wahrscheinlichkeit, daß *keiner* seinen Hut wiederbekommt?

Wir bezeichnen die Hüte mit $\{1, 2, \dots, n\}$. Die Vertauschung führt zu $\{\sigma(1), \dots, \sigma(n)\}$ mit $\sigma \in S_n$ der symmetrischen Gruppe von n -Elementen.

Es sei $A_n = \{\sigma \in S_n | \sigma(i) \neq i \forall i\}$ die Menge der vollständigen Permutationen

$$P(A_n) = \frac{|A_n|}{|S_n|} = \frac{|A_n|}{n!} = p_n.$$

Wir wollen $|A_n|$ rekursiv bestimmen. Dazu nehmen wir an $(n-1)$ Personen hätten verschiedene Hüte, aber eine Person hätte den eigenen. Dann läßt sich daraus eine vollständige Permutation herstellen, in dem diese Person mit irgendeiner der $(n-1)$ Personen die Hüte tauscht. Das gibt

$$(n-1)|A_{n-1}|$$

Möglichkeiten.

Es gibt aber noch andere Möglichkeiten, nämlich, daß $n-2$ Personen verschiedene Hüte haben, aber 2 Personen gleiche. Es gibt davon $(n-1)|A_{n-2}|$ Möglichkeiten.

Mehr Möglichkeiten gibt es nicht, denn geht man von einer vollständigen Permutation aus und wendet den obigen Prozeß umgekehrt an, Austausch eines Hutes, so erhält man $\sigma(i) = i$ für ein i oder $\sigma(i) = i$ für genau 2. Insgesamt haben wir

$$\begin{aligned} |A_n| &= (n-1)|A_{n-1}| + (n-1)|A_{n-2}| \quad \text{oder} \\ p_n &= \frac{n-1}{n}p_{n-1} + \frac{1}{n}p_{n-2}, \quad p_1 = 0, \quad p_2 = \frac{1}{2}. \end{aligned}$$

Behauptung:

$$p_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots \pm \frac{1}{n!}.$$

Beweis: Induktion nach n . Insbesondere gilt $p_n \rightarrow e^{-1}$ für $n \rightarrow \infty$.

Partitionen:

Wir haben oben untersucht, wieviele Möglichkeiten es gibt, um aus einer n -elementigen Menge eine r -elementige Teilmenge auszusondern. Stattdessen können wir aber auch sagen, daß bei einem solchen Aussondern die Menge in eine r - und eine $(n-r)$ -elementige Menge zerlegt wird. Allgemeiner kann man nur fragen, wieviele Möglichkeiten es gibt, eine n -elementige Menge in eine r_1 , eine r_2, \dots und eine r_k -elementige Menge zu zerlegen, wo $n = r_1 + \dots + r_k$.

Eine solche Zerlegung nennen wir eine (r_1, \dots, r_k) *Partition*.

Satz 2.3 Es gibt $\frac{n!}{r_1!r_2!\dots r_k!}(r_1, \dots, r_k)$ – Partitionen einer n -elementigen Menge, $n = r_1 + \dots + r_k$.

Beweis: Der Beweis kann durch Induktion nach k geführt werden. Wir aber machen es direkt.

Zunächst wählen wir eine r_1 -elementige Teilmenge aus. Dafür gibt es $\binom{n}{r_1}$ Möglichkeiten. Aus den restlichen $n - r_1$ Elementen wählen wir eine r_2 -elementige Menge aus. Diese gibt $\binom{n-r_1}{r_2}$ Möglichkeiten. Fahren wir so fort, erhalten wir insgesamt:

$$\binom{n}{r_1} \binom{n-r_1}{r_2} \cdots \binom{n-r_1-\dots-r_{k-2}}{r_{k-1}} = \frac{n!(n-r_1)! \cdots (n-r_1-\dots-r_{k-2})!}{r_1!(n-r_1)! r_2!(n-r_1-r_2)! r_{k-1}! \cdots r_k!}$$

$$\text{da } n-r_1-\dots-r_{k-1} = r_k \quad = \frac{n!}{r_1! \cdots r_k!}.$$

Beispiel 2.18 Es gibt also $\frac{32!}{10! \cdot 10! \cdot 10! \cdot 2!}$ verschiedene Skatspiele.

Beispiel 2.19 Beim Bridge werden 52 Karten auf 4 Spieler gleichmäßig verteilt. Also gibt es $\frac{(52)!}{(13!)^4}$ verschiedene Bridge-Kombinationen.

In vielen kombinatorischen Ausdrücken treten Fakultäten auf. Für kleine n läßt sich $n!$ noch relativ leicht berechnen. Aber für große n braucht man gute Näherungsformeln für $n!$. Für die Fakultät ist dies die sog. *Stirlingsche Formel* (J. Stirling).

$$n! = \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} \cdot e^{\frac{\Theta}{12n}} \quad 0 < \Theta < 1. \quad \text{Noch genauer ist}$$

$$\sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} \cdots\right).$$

Aufgabe 2.6

- Wie groß ist die Wahrscheinlichkeit, beim Skatspiel mindestens 3 Buben zu haben?
- Wie groß ist die Wahrscheinlichkeit, daß beim Skatspiel genau 2 Buben im Skat liegen?

Aufgabe 2.7 Wieviele Möglichkeiten gibt es beim „6 aus 49“ Lotto, wenn es auch noch auf die Reihenfolge ankommt, wie die Kugeln fallen?

Aufgabe 2.8 Ein Losverkäufer hat 50 Lose in seinem Topf. Davon sind 40 Nieten und 10 Gewinnlose. Wenn man 4 Lose kauft, wie groß ist die Wahrscheinlichkeit, genau 0(1, 2, 3, 4) Gewinnlose zu haben?

Aufgabe 2.9 Wie groß ist die Wahrscheinlichkeit, daß beim 6-maligen Münzenwerfen genau 2(3)-mal Kopf fällt?

Aufgabe 2.10 Wie groß ist die Wahrscheinlichkeit, daß r Leute an verschiedenen Wochentagen Geburtstag haben?

Aufgabe 2.11 In einem Kasten von 600 Knöpfen sind 3 % defekt. Wie groß ist die Wahrscheinlichkeit, daß unter 30 zufällig ausgewählten Knöpfen 2 defekt sind?

Aufgabe 2.12 Es wird mit 6 fairen Würfeln gewürfelt. Wie groß ist die Wahrscheinlichkeit, 3 Paare zu haben?

Aufgabe 2.13 In einem Kasten sind N_1 rote, N_2 weiße und N_3 blaue Kugeln. Es werden n Kugeln willkürlich gezogen. Wie groß ist die Wahrscheinlichkeit, k_1 rote, k_2 weiße und k_3 blaue zu haben, $k_1 + k_2 + k_3 = n$.

Aufgabe 2.14 Bei der sog. Glücksspirale der Olympialotterie 1971 wurden die 7-ziffrigen Gewinnzahlen auf die Art ermittelt, daß aus einer Trommel, welche je 7 Kugeln mit den Ziffern 0 bis 9 enthielt, nach Durchmischen 7 Kugeln entnommen und deren Ziffern zu einer Zahl angeordnet wurden.

- a) Man wiederlege die Behauptung, daß jede mögliche 7-ziffrige Zahl die gleiche Gewinnchance hatte.
- b) Man zeige, daß die Wahrscheinlichkeit, daß eine bestimmte 7-ziffrige Zahl (z_1, z_2, \dots, z_7) gezogen wird, nur davon abhängt, wieviele k -Tupel von gleichen Ziffern in (z_1, z_2, \dots, z_7) vorkommen, $1 \leq k \leq 7$.
Man zeige ferner, daß es 15 Sorten von 7-ziffrigen Zahlen mit verschiedenen Gewinnchancen gibt.

Aufgabe 2.15 Man bestimme die Wahrscheinlichkeit bei $6n$ -maligem Würfeln n mal eine „6“ zu würfeln. Was ist wahrscheinlicher:

Bei 6-maligem Würfeln einmal eine „6“ zu würfeln?

Bei 12-maligem Würfeln zweimal eine „6“ zu würfeln?

Bei 18-maligem Würfeln dreimal eine „6“ zu würfeln?

Aufgabe 2.16 Ein Würfel, der von allen Seiten angemalt ist, wird gleichmäßig in n^3 Teile zersägt. Wie groß ist die Wahrscheinlichkeit, daß ein willkürlich herausgegriffener Würfel keine (1, 2 oder 3) bemalte Seiten hat? Man bestimme die Wahrscheinlichkeit im Falle $n = 10$.

Aufgabe 2.17 Wie groß ist die Wahrscheinlichkeit, bei einem Skatblatt mindestens (genau) 2 Buben und 2 Asse zu erhalten?

Aufgabe 2.18 S. Banach pflegte in jeder Hosentasche eine Schachtel Streichhölzer zu haben. Unter der Annahme, daß jede Schachtel n Streichhölzer enthält und die Streichhölzer zufällig entnommen werden, bestimme die Verteilung der Anzahl der Streichhölzer der einen Tasche, sobald die andere leer ist.

Aufgabe 2.19

In einem Kasten liegen zehn 20-Pfennig-Stücke,

fünf 5-Pfennig-Stücke,

und zwei 10-Pfennig-Stücke.

Man nehme willkürlich 6 Münzen. Wie groß ist die Wahrscheinlichkeit dafür, daß insgesamt höchstens eine Mark genommen wurde?

Aufgabe 2.20 Man bestimme die Wahrscheinlichkeit dafür, daß eine beliebig ausgewählte ganze Zahl N

- a) bei der Quadrierung,

- b) bei Erhebung in die vierte Potenz und
- c) bei Multiplikation mit einer beliebigen Zahl

eine Zahl ergibt, die mit einer 1 endet.

Aufgabe 2.21 Auf wieviele Arten kann man n Personen an einen runden Tisch setzen?

Aufgabe 2.22 Es seien M und N endliche Mengen mit $|M| = m$ und $|N| = n$. Wieviele injektive Abbildungen gibt es von M nach N ?

Aufgabe 2.23 Es wird 3 mal gewürfelt. Wie groß ist die Wahrscheinlichkeit, daß die Augenzahl immer größer wird (Betrachte die Fälle \leq und $<$ getrennt)?

Aufgabe 2.24 Mäxchen sammelt Kronenkorkenbilder für sein Poster. Er braucht N Stück. Wieviele Flaschen muß er im Mittel kaufen, um seine Sammlung zu vervollständigen?

Aufgabe 2.25 Beim Geburtstagsproblem haben wir implizit die Annahme gemacht, daß jeder Tag im Jahr gleich häufig Geburtstag eines Menschen ist. Wie ändert sich das Problem, wenn sich die Geburtstage in bestimmten Monaten häufen?

Kapitel 3

Einfache Konzepte der Wahrscheinlichkeitstheorie

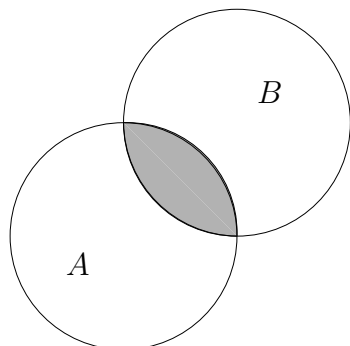
3.1 Bedingte Wahrscheinlichkeit und Unabhängigkeit

In der Wahrscheinlichkeitstheorie beschäftigen wir uns mit Experimenten, deren Ausgang zufällig ist. Allen möglichen Ausgängen ordnen wir Wahrscheinlichkeiten zu. Wenn wir über den Ausgang eines Experimentes zusätzliche Informationen erhalten, so wird es sinnvoll sein, diese zu verwerten, und die Wahrscheinlichkeiten entsprechend zu ändern. Die zusätzliche Information wird darin bestehen, daß wir wissen, daß ein bestimmtes Ereignis eingetreten ist, d.h. also, daß nur noch die Elemente einer Teilmenge des Ereignisraumes als Ausgänge des Experimentes in Frage kommen. Das Problem lautet also: Wie wird man die Wahrscheinlichkeit von Ereignissen definieren, wenn ein bestimmtes Ereignis eingetreten ist?

Beispiel 3.1 Wir würfeln blind und ein Mitspieler teilt uns mit, daß wir eine gerade Augenzahl geworfen haben. Dann werden wir der Augenzahl 6 die Wahrscheinlichkeit $\frac{1}{3}$, der Augenzahl 5 jedoch die Wahrscheinlichkeit 0 geben. Wir schreiben hierfür mit $A = \{2, 4, 6\}$

$P(\{6\} \text{ unter der Annahme } A) = \frac{1}{3}$. Es ist dies aber genau $= \frac{P(\{6\})}{P(A)}$.

Allgemeiner seien nun A und B Ereignisse eines W -Raumes (Ω, Σ, P) . Wir wollen nun die Wahrscheinlichkeit berechnen, mit der B eintritt, falls wir bereits wissen, daß A eingetreten ist.



Dabei wollen wir zunächst annehmen, daß (Ω, Σ, P) ein Laplace'scher W -Raum mit n Elementen ist. Es sei $|A| = m$, $|B| = 1$ und $|A \cap B| = k$. Da wir bereits wissen, daß A eintritt, wird das Ergebnis des Zufallexperiments einer der m Punkte in A sein. Jeder dieser Punkte ist gleichwahrscheinlich. B tritt nun genau dann ein, wenn ein Ereignis aus $A \cap B$ eintritt. Die Wahrscheinlichkeit dafür aber ist $\frac{k}{m} = \frac{k/n}{m/n} = \frac{P(A \cap B)}{P(A)}$.

Dies legt folgende Definition nahe

Definition 3.1 Es sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und A sei ein Ereignis mit $P(A) > 0$. Für ein beliebiges Ereignis B heißt dann

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

die *bedingte Wahrscheinlichkeit von B unter der Hypothese A* oder auch *die bedingte Wahrscheinlichkeit von B gegeben A*.

Beispiel 3.2 Es werden 2 faire Würfel geworfen. Wie groß ist die Wahrscheinlichkeit, eine Augensumme von 7 zu werfen, wenn man weiß, daß mindestens ein Würfel eine 4 zeigt?

Es sei nun

$A =$ ein Würfel zeigt 4 $= \{(4, 1), \dots, (4, 6); (1, 4), \dots, (6, 4)\}$

$B =$ die Augensumme ist 7 $= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$

Dann ist $A \cap B = \{(3, 4), (4, 3)\}$ und

$$P(A|B) = \frac{2/36}{11/36} = \frac{2}{11}.$$

Beispiel 3.3 Betrachten wir alle Familien mit 2 Kindern. Ordnen wir die Kinder dem Alter nach, so ergeben sich folgende Möglichkeiten: $S = \{JJ, JM, MJ, MM\}$. Nimmt man an, daß Jungen und Mädchen gleichwahrscheinlich geboren werden, so hat jedes der 4 Elementarereignisse aus Ω die Wahrscheinlichkeit $\frac{1}{4}$. Wie ist die Wahrscheinlichkeit, daß eine Familie 2 Jungen hat, unter der Voraussetzung, daß mindestens 1 Junge da ist? Es sei $B = \{JJ\}$ und $A = \{JM, MJ, JJ\}$. Dann ist $P(B|A) = \frac{1/4}{3/4} = \frac{1}{3}$ die gesuchte Wahrscheinlichkeit. Diese Ergebnis ist überraschend, da man die Wahrscheinlichkeit $\frac{1}{2}$ erwartet.

Was ist die Wahrscheinlichkeit, daß eine Familie 2 Jungen hat, wenn man weiß, daß das *älteste* Kind ein Junge ist? Hier muß man $A = \{JM, JJ\}$ wählen und erhält $P(B|A) = \frac{1/4}{2/4} = \frac{1}{2}$.

Beispiel 3.4 Die Wahrscheinlichkeit einer Familie, n Kinder zu haben sei μ_n , $\sum_{n=0}^{\infty} \mu_n = 1$. Wie groß ist die Wahrscheinlichkeit, daß die Familie k Jungen und

$n - k$ Mädchen hat? Dabei wollen wir annehmen, daß Jungen und Mädchen gleichwahrscheinlich auftreten. Unsere Frage oben können wir nicht unmittelbar lösen. Stattdessen beantworten wir zunächst die Frage: Wie groß ist die Wahrscheinlichkeit, daß eine Familie mit n Kindern genau k Jungen hat? In diesem Fall besteht Ω aus den 2^n n -Tupeln von Symbolen J bzw. M , wenn die Kinder wieder dem Alter nach geordnet werden. Eine Familien hat genau k Jungen, wenn in so einem n -Tupel $JMJJ \dots k$ mal das Symbol J auftaucht. Das Erscheinen des Symbols J definiert uns also eine k -elementige Teilmenge von $\{1, \dots, n\}$ und jede Teilmenge kennzeichnet das Auftauchen von k Jungen. Da es $\binom{n}{k}$ k -elementige Teilmengen von $\{1, \dots, n\}$ gibt, ist die Wahrscheinlichkeit, daß eine Familie mit n Kindern k Jungen hat also $\binom{n}{k} 2^{-n}$.

Es sei nun

$$\begin{aligned} A_n &= \text{Die Familie hat } n \text{ Kinder.} \\ B &= \text{Die Familie hat } k \text{ Jungen und } n - k \text{ Mädchen.} \end{aligned}$$

Dann haben wir

$$P(B) = P(B \cap A) = P(A) \cdot P(B|A) = \mu_n \cdot 2^{-n} \binom{n}{k}.$$

Beispiel 3.5 Aus dem Skatspiel wird eine Karte gezogen und nicht zurückgesteckt, sondern verdeckt weggelegt.

Dann wird eine zweite Karte gezogen. Wie groß ist die Wahrscheinlichkeit, daß die zweite Karte ein As ist? Unsere Betrachtung wird dadurch kompliziert, daß der Ausgang des ersten Ziehens das Ergebnis beim zweiten Zug beeinflusst. Wir helfen uns mit bedingten Wahrscheinlichkeiten und betrachten

$$\begin{aligned} A &= \text{Es wird ein As beim ersten Mal gezogen.} \\ B &= \text{Es wird ein As beim zweiten Mal gezogen.} \end{aligned}$$

Dann haben wegen $B = (B \cap A) \cup (B \cap (\Omega \setminus A))$

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap (\Omega \setminus A)) \\ &= P(B|A)P(A) + P(B|\Omega \setminus A)P(\Omega \setminus A) \\ &= \frac{3}{31} \cdot \frac{1}{8} + \frac{4}{31} \cdot \frac{7}{8} = \frac{31}{31 \cdot 8} = \frac{1}{8}. \end{aligned}$$

In den Beispielen (3.4) und (3.5) haben wir die bedingten Wahrscheinlichkeiten verwendet, um Wahrscheinlichkeiten zu berechnen. Maßgebend dafür war

1. $P(A \cap B) = P(A|B)P(A)$.

Die Verallgemeinerung dieser Formel ist der *Multiplikationssatz*: Es seien B_1, \dots, B_n Ereignisse. Dann gilt:

2. $P(B_1 \cap \dots \cap B_n) = P(B_1)P(B_2|B_1)P(B_3|B_1 \cap B_2) \dots P(B_n|B_1 \cap \dots \cap B_{n-1})$

Der Beweis dieses Satzes wird durch Induktion geführt (Aufgabe).

Eine weitere wichtige Beziehung zwischen Wahrscheinlichkeiten und bedingten Wahrscheinlichkeiten gibt der *Satz von der totalen Wahrscheinlichkeit*:

Satz 3.1 Es sei $A_1, A_2 \dots$ eine Zerlegung von Ω in paarweise disjunkte Teilmengen. Dann gilt für jedes Ereignis B

$$3. P(B) = \sum P(B|A_i)P(A_i)$$

Beweis: Es ist $B = (B \cap A_1) \cup \dots \cup (B \cap A_n) \cup \dots$ und alle diese Mengen sind disjunkt. Also gilt:

$$P(B) = \sum P(B \cap A_i) = \sum P(B|A_i)P(A_i).$$

Eine Anwendung von Satz (3.1) hatten wir schon in Beispiel (3.5) gesehen. Als zweite Anwendung wollen wir noch einmal zu Beispiel (3.4) zurückkehren.

Beispiel 3.6 Wie groß ist die Wahrscheinlichkeit, daß in einer Familie nur Jungen vorkommen?

Hier sei wieder

$$\begin{aligned} A_n &= \text{Die Familie hat } n \text{ Kinder.} \\ B &= \text{Die Familie hat nur Jungen.} \end{aligned}$$

Dann gilt:

$$P(B) = \sum P(B|A_n)P(A_n) = \sum_{n \geq 1} 2^{-n} \mu_n.$$

Es sei $(\Omega, \mathcal{P}(\Omega), P)$ ein (diskreter) W -Raum und $A \subset \Omega$ mit $P(A) > 0$.

Dann definiert

$$\mathcal{P}(\Omega) \ni B \rightarrow P(B|A) = \nu(B)$$

ein Maß auf Ω . Dieses ist auf A konzentriert, weil $\nu(\Omega \setminus A) = 0$.

Wenn nun $P(B|A) = P(B)$, so beeinflusst das Eintreffen von A nicht das Eintreffen von B . Wir sagen dazu: A und B sind *unabhängig*.

Wegen $P(B|A) = P(B) = \frac{P(A \cap B)}{P(A)} = P(B)$ führt dieses zu:

Definition 3.2 Zwei Ereignisse A und B heißen *unabhängig*, wenn

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

Beispiel 3.7 Es seien A, B die Ereignisse

$$\begin{aligned} A &= \text{Aus einem Skatspiel wird ein As gezogen,} \\ B &= \text{Aus dem Skatspiel wird ein Kreuz gezogen.} \end{aligned}$$

Intuitiv erwartet man, daß beide Ereignisse unabhängig sind. Dies ist auch der Fall, denn

$$P(A \cap B) = \frac{1}{32} = \frac{1}{8} \cdot \frac{1}{4} = P(A) \cdot P(B).$$

Beispiel 3.8 Die Ereignisse A, B aus Beispiel (3.5) sind nicht unabhängig, denn $P(A \cap B) = \frac{\binom{4}{2}}{\binom{32}{2}}$, da es bei den $\binom{32}{2}$ Kartenpaaren nur $\binom{4}{2}$ Paare von Assen gibt. Andererseits ist

$$P(A) = \frac{1}{8} = P(B), \text{ also } P(A)P(B) = \frac{1}{64} \neq \frac{\binom{4}{2}}{\binom{32}{2}} = \frac{4 \cdot 3}{32 \cdot 31} = \frac{2}{248}.$$

Der Begriff der Unabhängigkeit kann folgendermaßen auf beliebig viele Ereignisse verallgemeinert werden.

Definition 3.3 Eine Menge $\{A_i\}_{i \in I}$ von Ereignissen ist unabhängig, wenn für jede endliche Teilmenge $\{A_{i_1}, \dots, A_{i_n}\}$ gilt:

$$P(A_{i_1} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_n}).$$

Beispiel 3.9 Für die Unabhängigkeit einer Familie (A_i) von Ereignissen genügt *nicht* die paarweise Unabhängigkeit, d.h.

$$P(A_{i_1} \cap A_{i_2}) = P(A_{i_1}) \cdot P(A_{i_2}), \quad i_1, i_2 \in I.$$

Dazu betrachte man die Menge

$$S = \{(1, 2, 3); (1, 3, 2); (2, 1, 3); (2, 3, 1); (3, 1, 2); (3, 2, 1); (1, 1, 1); (2, 2, 2); (3, 3, 3)\}.$$

Wähle P so, daß jedes Elementarereignis die Wahrscheinlichkeit $\frac{1}{9}$ hat. Für $i = 1, 2, 3$ sei ferner $A_i = \{(n_1, n_2, n_3) | n_i = 1\}$. Dann gilt

$$P(A_i) = \frac{1}{3} \text{ und } P(A_i \cap A_j) = \frac{1}{9} = P(A_i)P(A_j).$$

Es ist aber $P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq P(A_1)P(A_2)P(A_3)$.

Beispiel 3.10 Bei der Produktion von Glühlampen werden Wendel, Birne und Gewinde in verschiedenen Arbeitsgängen hergestellt. Aus vergangenen Stichproben weiß man, daß

- i) 2% aller Wendeln Fehler haben,
- ii) 1% aller Birnen fehlerhaft sind,
- iii) 2% aller Gewinde defekt sind.

Wie groß ist die Wahrscheinlichkeit, daß eine willkürlich gewählte Birne in Ordnung ist? Da die Arbeitsgänge für die Wendel, die Birne und das Gewinde getrennt ablaufen, kann man annehmen, daß die Ereignisse:

$$\begin{aligned} A &= \text{die Wendel ist intakt,} \\ B &= \text{die Birne ist intakt,} \\ C &= \text{das Gewinde ist intakt} \end{aligned}$$

unabhängig sind. Es gilt daher

$$P(A \cap B \cap C) = P(A)P(B)P(C) = 0,98 \cdot 0,99 \cdot 0,98 = 0,958.$$

In vielen praktischen Problemen ist es häufig schwierig, zu entscheiden, ob gewisse Ereignisse unabhängig sind. In Beispiel 3.10 dürfte dies etwa nicht gelten, wenn die Wendel und das Gewinde in einem Arbeitsgang hergestellt werden würden. Im wahrscheinlichkeitstheoretischen Modellieren wird man daher immer prüfen müssen, wieweit solche Annahmen gerechtfertigt sind. Die Wahl des Modells hängt also entscheidend davon ab, wie der Mathematiker die Situation beurteilt.

Aufgabe 3.1 Eine Verkehrsstatistik über 1100 Verkehrsunfälle gibt folgende Auskünfte über das Geschlecht des Fahrers und die vermutete Fehlerquelle:

	Defekt am Fahrzeug	Alkohol	Fehler beim Fahrverhalten und Einschätzung der Lage
männlich	310	102	208
weiblich	280	45	155

Falls ein Unfall durch einen Defekt am Fahrzeug verursacht war, wie groß ist die Wahrscheinlichkeit, daß der Fahrer männlich war?

Falls der Fahrer ein Mann war, mit welcher Wahrscheinlichkeit war die Unfallursache ein Fehler am Fahrzeug?

Aufgabe 3.2 Es wird 4 mal mit einem fairen Würfel gewürfelt. Es seien dann:

A_k = Die Augensumme der ersten beiden Würfel ist k .

B_k = Die Augensumme des 2. und 3. Würfels ist k .

C_k = Die Augensumme des 3. und 4. Würfels ist k .

$k = 2, \dots, 12$. Dann sind A_k und C_l ($k, l = 2, \dots, 12$) unabhängig.

Sind auch A_k und B_l unabhängig für $k, l = 2, \dots, 12$?

Aufgabe 3.3 Es seien A und B Ereignisse eines Maßraumes $(\Omega, \mathcal{P}(\Omega), P)$. Dann gilt:

- i) A, B unabhängig $\iff A, \Omega/B$ unabhängig $\iff \Omega/A, B$ unabhängig.
- ii) Gilt $P(A) = 0$ oder 1 , so sind A und B unabhängig.
- iii) Ist $A \subset B$ und sind A und B unabhängig, so gilt $P(A) = 0$ oder $P(B) = 1$.

Aufgabe 3.4 Aus einem Skatspiel wird eine Karte gezogen. Was ist die Wahrscheinlichkeit, daß es ein As ist, unter der Hypothese, daß die Karte weder ein Bube noch eine Dame ist?

Aufgabe 3.5 Für welche Ereignisse A und B wird man sinnvollerweise annehmen können, daß sie unabhängig sind.

- i) A Die Person ist größer als 1,80 m.
 B Die Person wiegt mehr als 80 kg.

- ii) A Es regnet am 1.3.1984.
 B Es regnet am 2.3.1984.
- iii) A Die Person hat blaue Augen.
 B Die Person fährt einen Sportwagen.

3.2 Bayes' Satz

In vielen Problemen wird der Ausgang eines Zufallsexperimentes beobachtet und man versucht, daraus Rückschlüsse auf die Ursache zu ziehen.

Satz 3.2 (*Satz von Bayes:*) Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und sei A_1, A_2, \dots eine Zerlegung von Ω in disjunkte Teilmengen, $A_i \cap A_j = \emptyset, i \neq j$, $\Omega = \bigcup_{i=1}^{\infty} A_i$.

Es trete das Ereignis B ein. Dann gilt für die bedingte Wahrscheinlichkeit $P(A_j|B)$

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)} = \frac{P(A_j)P(B|A_j)}{P(B)}.$$

Man interpretiert hier $P(A_i|B)$ als Wahrscheinlichkeit, daß A_i die Ursache von B ist.

Beweis: Es ist

$$\begin{aligned} P(A_j)P(B|A_j) &= P(B \cap A_j) \\ &= P(B)P(A_j|B) = P(A_j|B) \sum_{i=1}^{\infty} P(B \cap A_i) \\ &= P(A_j|B) \sum_{i=1}^{\infty} P(A_i)P(B|A_i). \end{aligned}$$

Beispiel 3.11 Eine Karte wird aus einem Skatspiel gezogen und verdeckt beiseite gelegt. Darauf wird eine zweite Karte gezogen. Falls diese ein Kreuz ist, wie groß ist die Wahrscheinlichkeit, daß auch die erste Karte ein Kreuz war? Wir setzen

- A = Die erste Karte ist ein Kreuz.
 B = Die zweite Karte ist ein Kreuz.

$$A = A_1, A_2 = \Omega/A_1$$

Dann haben wir

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{\frac{1}{4} \cdot P(B|A)}{\frac{1}{4}} = P(B|A) = \frac{7}{31}.$$

Beispiel 3.12 In einer Schraubenfabrik werden Schrauben durch 3 Maschinen produziert und zwar produziert die erste (zweite, dritte) Maschine 20 % (30 %, 50 %) der Gesamtproduktion. Die erste (zweite, dritte) Maschine hat 2 % (3 %, 4 %) Ausschuß. Eine Schraube wird aus dem Lager willkürlich gewählt und als defekt erkannt. Wie

groß ist die Wahrscheinlichkeit, daß die Schraube von der k -ten Maschine kommt? Es sei A_k = Schraube kommt von der k -ten Maschine, $k = 1, 2, 3$, B = die Schraube ist defekt.

Dann gilt

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum P(A_i)P(B|A_i)} = \frac{P(A_k)P(B|A_k)}{0,2 \cdot 0,02 + 0,3 \cdot 0,03 + 0,5 \cdot 0,05}$$

$$P(A_1|B) = \frac{4 \cdot 10^{-3}}{33 \cdot 10^{-3}} = \frac{4}{33}$$

$$P(A_2|B) = \frac{9 \cdot 10^{-3}}{33 \cdot 10^{-3}} = \frac{3}{11}$$

$$P(A_3|B) = \frac{20 \cdot 10^{-3}}{33 \cdot 10^{-3}} = \frac{20}{33}$$

Beispiel 3.13 Die meisten Krankheiten führen zu einer Reihe von Symptomen, deren Gesamtheit als das Krankheitsbild bezeichnet wird. In seiner Diagnose versucht nun der Arzt, aus den beobachteten Symptomen auf die Krankheit zu schließen. Dies wird dadurch kompliziert, daß verschiedene Krankheiten durchaus ein gleiches Symptom, z.B. Fieber, hervorrufen können. Wir wollen nun ein einfaches wahrscheinlichkeitstheoretisches Modell der Diagnostik betrachten, das auf Bayes' Satz beruht. Dazu seien k_1, \dots, k_n Krankheiten und s_1, \dots, s_m seien beobachtete Symptome dieser Krankheiten.

Ferner sei S die Menge aller Menschen mit genau einer Krankheit. Es sei K_i dann die Teilmenge aller Menschen aus S mit der Krankheit k_i .

Ein Patient kommt zum Arzt und zeigt die Symptome $\{s_{i_1}, \dots, s_{i_r}\}$. Der Arzt möchte nun wissen, welche Krankheit K die wahrscheinlichste Ursache dafür ist. Dazu sei A die Teilmenge (von Ω) aller Leute, die die Symptome s_{i_1}, \dots, s_{i_r} zeigen: Der Arzt möchte also $P(K_1|A), \dots, P(K_n|A)$ wissen, um die wahrscheinlichste Ursache zu bestimmen. Aus der medizinischen Statistik kennt er die relativen Häufigkeiten der Krankheiten K_i , also $P(K_i)$. Ferner kennt man aus der medizinischen Statistik $P(A|K_i)$ d.h. die relativen Häufigkeiten der Symptome, wenn die Krankheit K_i vorliegt. Damit aber erlaubt uns der Satz von Bayes, die gesuchten bedingten Wahrscheinlichkeiten zu berechnen.

Als Beispiel hierzu betrachten wir folgenden fiktiven Fall: Die Krankenhausstatistiken für 50.000 Kranke mit den Krankheiten k_1 = Grippe, k_2 = Mandelentzündung und k_3 = Bronchitis ergaben folgende Daten

	Zahl mit der Krankheit K_i	Anzahl mit der Krankheit K_i und Symptomen: Fieber, Kopfschmerzen, Frösteln
K_1	30.000	25.000
K_2	8.000	5.000
K_3	12.000	6.000

Hat nun ein Patient die Symptome: Fieber, Kopfschmerzen und Frösteln, so hat er mit der Wahrscheinlichkeit

$$\begin{aligned} P(K_1|A) &= \frac{P(K_1)P(A|K_1)}{\sum P(K_i)P(A|K_i)} = \frac{\frac{30.000}{50.000} \cdot \frac{25.000}{30.000}}{\frac{3}{5} \cdot \frac{25}{30} + \frac{8}{50} \cdot \frac{5}{8} + \frac{12}{50} \cdot \frac{1}{2}} \\ &= \frac{\frac{3}{5} \cdot \frac{5}{6}}{\frac{3}{5} \cdot \frac{5}{6} + \frac{1}{10} + \frac{3}{25}} = \frac{25}{25 + 5 + 6} = \frac{25}{36} \end{aligned}$$

Grippe.

Mit der Wahrscheinlichkeit

$$P(K_2|A) = \frac{\frac{8}{50} \cdot \frac{5}{8}}{\frac{3}{5} \cdot \frac{5}{6} + \frac{1}{10} + \frac{3}{25}} = \frac{5}{25 + 5 + 6} = \frac{5}{36}$$

hat er Mandelentzündung.

Aufgabe 3.6 In einem Kasten liegen 4 Münzen. Eine davon hat auf beiden Seiten Kopf, die anderen sind fair. Eine Münze wird wahllos geworfen. Es fällt Kopf. Mit welcher Wahrscheinlichkeit haben wir die falsche Münze gezogen?

Aufgabe 3.7 Es seien 10 Urnen gegeben, die rote und weiße Kugeln enthalten. Es seien in der i -ten Urne i weiße und $10 - i$ rote Kugeln. Wie groß ist die Wahrscheinlichkeit bei willkürlicher Auswahl einer Urne eine rote Kugel zu ziehen? Wie groß ist die Wahrscheinlichkeit, daß aus der i -ten Urne gezogen wurde, wenn die Kugel rot ist?

Aufgabe 3.8 Es seien $A_1, A_2 \in \Sigma$ mit $P(A_1 \cap A_2) \neq 0$. Ferner bezeichne $P_{A_i}(B) = P(B|A_i)$. Dann gilt $P(B|A_1 \cap A_2) = P_{A_1}(B|A_2)$.

Aufgabe 3.9 An einer Kasse, an der Fahrscheine für 50 Pfennig verkauft werden, steht eine Schlange von $2n$ Personen. Wie groß ist die Wahrscheinlichkeit dafür, daß keiner der Käufer auf Wechselgeld warten muß, wenn sich vor dem Verkauf des Fahrscheins an den ersten Käufer in der Kasse $2m$ 50-Pfennig-Stücke befanden und Bezahlung gleichwahrscheinlich in 50-Pfennigstücken bzw. Markstücken erfolgt? Man berechne die Wahrscheinlichkeit speziell für $n = 100$, $m = 50$.

Aufgabe 3.10 Am Roulettetisch ist 5 mal hintereinander nicht "Rot" gefallen. Soll man nun "Rot" oder "Schwarz" setzen? Hierzu nehme man an, daß das Gerät fair sei.

Aufgabe 3.11 Bei einem Quiz kann der Gewinner eine von 3 Türen, hinter denen ein Preis verborgen ist, auswählen. Nach seiner Wahl öffnet der Quizmaster eine der anderen beiden Türen ohne Preis. Der Spieler darf nun seine Wahl noch einmal überdenken und ggf. eine andere Tür wählen. Wie würden Sie sich verhalten?

Aufgabe 3.12 Das Würfelspiel Craps wird mit 2 fairen Würfeln gespielt. Wirft der Spieler eine Augensumme 7 oder 11 hat er gewonnen. Bei einer Augensumme 2, 3 oder 12 hat er verloren. Ansonsten darf er weiter würfeln. Wie groß ist seine Gewinnchance?

Kapitel 4

Maßtheorie

4.1 σ -Algebren

Anders als bei diskreten Wahrscheinlichkeitsräumen, muß man sich bei überabzählbaren Ereignisräumen genauer überlegen, welche σ -Algebren Σ als Ereignis-Algebren in Frage kommen.

Solche überabzählbaren Ereignisräume treten hauptsächlich im Zusammenhang mit folgenden Problemen auf

- i) Der Ereignisraum ist ein Teil des \mathbb{R}^n
- ii) Man untersucht ∞ -häufiges Wiederholen eines Zufallsexperimentes.

Beispiel 4.1 Betrachten wir das Drehen eines Glücksrades. Die Elementarereignisse sind hier die Zeigerpositionen, also Winkel. Als Ereignisraum können wir also $[0, 2\pi] = \Omega$ wählen. Für einen Winkelintervall Δ gibt dann $\frac{\text{Bogenlänge von } \Delta}{2\pi}$ die Wahrscheinlichkeit an, daß der Zeiger in Δ stehen bleibt. Wir werden daher unser System Σ von Ereignissen so wählen, daß die Intervalle dazu gehören. Schwieriger wird allerdings dann die Entscheidung, welche komplizierten Mengen A dazugehören und wie groß $P(A)$ ist. Probleme dieser Art wollen wir im Folgenden untersuchen.

Definition 4.1 Es sei Ω eine Menge und $\Sigma \subset \mathcal{P}(\Omega)$. Σ heißt σ -Algebra, falls

- i) $\Omega \in \Sigma$
- ii) $A \in \Sigma \Rightarrow \Omega \setminus A = A^c \in \Sigma$
- iii) $A_1, A_2, A_3 \dots \in \Sigma \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Der Begriff σ -Algebra weist gewisse Ähnlichkeiten zum Begriff der offenen Mengen in der Topologie auf.

Beispiel 4.2

- a) $\Sigma = \{\emptyset, \Omega\}$ ist eine σ -Algebra
- b) $\Sigma = \{\emptyset, A, A^c, \Omega\}$ ist eine σ -Algebra
- c) $\Sigma = \mathcal{P}(\Omega)$ ist eine σ -Algebra.

Lemma 4.1 Es sei Σ eine σ -Algebra und $A, A_i, B \in \Sigma$. Dann gilt

- i) $A \setminus B \in \Sigma$
- ii) $\bigcap_{i=1}^{\infty} A_i \in \Sigma$
- iii) $\overline{\lim} A_i = \bigcap_{k=1}^{\infty} \bigcup_{i=k}^{\infty} A_i \in \Sigma$
- iv) $\underline{\lim} A_i = \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} A_i \in \Sigma$

Lemma 4.2 Es sei Ω eine Menge und $\Sigma_i \subset \mathcal{P}(\Omega)$ für $i \in I$ σ -Algebren. Dann ist $\bigcap_{i \in I} \Sigma_i$ eine σ -Algebra.

Definition 4.2 Es sei Ω eine Menge und $A_i \subset \Omega$ $i \in I$. Die kleinste σ -Algebra Σ , die alle A_i enthält, heißt die von den A_i erzeugte σ -Algebra und die $\{A_i\}_{i \in I}$ werden ein Erzeugendensystem von Σ genannt. Wir schreiben auch $\Sigma = \Sigma(\{A_i\}_{i \in I})$.

Ist Ω ein metrischer Raum und \mathcal{O} das System aller offenen Mengen, so nennt man $\Sigma(\mathcal{O})$ auch das System der Borelmengen und schreibt auch $\mathcal{B}(\Omega) = \Sigma(\mathcal{O})$. Die weitaus gebräuchlichsten σ -Algebren in Anwendungen sind $\mathcal{B}_n = \mathcal{B}(\mathbb{R}^n)$, $\mathcal{B}([0, 1])$, $\mathcal{B}([a, b])$ mit $a, b \in \mathbb{R}$, $a < b$. Ist allgemeiner $\Omega \subset \mathbb{R}^n$, $\Omega \in \mathcal{B}_n$, so sei $\mathcal{B}(\Omega) = \Omega \cap \mathcal{B}_n$ die σ -Algebra der Borelmengen auf Ω .

Bemerkung 4.1 $\mathcal{B}(\mathbb{R}^n)$ enthält alle Mengen, die in vernünftiger Weise in der Praxis auftreten, z.B. die abgeschlossenen Mengen im \mathbb{R}^n, \dots . Man weiß $|\mathcal{B}(\mathbb{R}^n)| = c$ mit $c = |\mathbb{R}|$ Mächtigkeit des Kontinuums. Andererseits ist $|\mathcal{P}(\mathbb{R}^n)| = 2^c > c$. Also gibt es sehr viele nicht Borelmengen.

Bemerkung 4.2 Die Definition der Borelmengen zeigt schon eine grundsätzliche Schwierigkeit bei σ -Algebren: In der Regel ist es außerordentlich schwierig, die Elemente einer σ -Algebra zu beschreiben – die σ -Algebra wird vielmehr über ihre Erzeuger definiert.

Dies hat zur Folge, daß viele Beweise für σ -Algebren nach folgendem Schema ablaufen:

- I) Die Erzeuger (A_i) von Σ haben E .
- II) Die Mengen mit der gewünschten Eigenschaft E bilden eine σ -Algebra.

Also hat jedes $A \in \Sigma$ die Eigenschaft.

Aus diesem Grunde ist es häufig wichtig, besonders einfache Erzeugendensysteme zu wählen.

Beispiel 4.3

$$\mathcal{B}(\mathbb{R}^n) = \Sigma((-\infty, a_1] \times \dots \times (-\infty, a_n] | a = (a_1, \dots, a_n) \in \mathbb{Q}^n)$$

Beweis: Man sieht leicht, daß $\sum(\dots) \subset \mathcal{B}(\mathbb{R}^n)$ und daß

$$(-\infty, a_1] \times \dots \times (-\infty, a_n] \in \sum \quad \forall a = (a_1, \dots, a_n) \in \mathbb{R}^n = (-\infty, a].$$

Dann gilt etwa $a^{(n)} = a_1, \dots, \rightarrow a$, so ist $\overline{\lim_{u \rightarrow \infty}}(-\infty, a^{(n)}) = (-\infty, a]$. Ferner ist $(-\infty, a) = \bigcup_{u=1}^{\infty} (-\infty, a - \frac{1}{n}] \in \sum$ und dann auch

$$(a, b) = (a_1, b_1) \times \dots \times (a_n, b_n) \in \sum.$$

Es sei nun $\mathcal{O} \subset \mathbb{R}^n$ offen. Dann ist $\mathcal{O} \cap \mathbb{Q}^n$ dicht in \mathcal{O} . Für jedes $c \in \mathcal{O} \cap \mathbb{Q}^n$ existiert dann ein $I_c \in \mathcal{O}$ mit $I_c = (a, b)$. Dann aber ist $\mathcal{O} = \bigcup_{c \in \mathcal{O} \cap \mathbb{Q}^n} I_c$.

Dies zeigt: Jede offene Menge gehört zu \sum . Da \sum eine σ -Algebra ist, muß also $\mathcal{B}(\mathbb{R}^n) \subset \sum$ sein.

Mit diesen Begriffen können wir jetzt auch das *Produkt* von σ -Algebren erklären.

Es seien $(\Omega_i, \sum_i) i = 1, \dots, n$ σ -Algebren. Eine Menge $A_1 \times \dots \times A_n \subset \prod_{i=1}^n \Omega_i = \Omega$ mit $A_i \in \sum_i$ heißt dann ein *Rechteck*. Naheliegender ist es, zu fordern, daß die \sum -Algebra für Ω alle Rechtecke enthält. Wir definieren daher

Definition 4.3

$$\begin{aligned} \bigotimes_{i=1}^n \sum_i &= \sum \left(A_1 \times \dots \times A_n \mid A_i \in \sum_i \right) \\ \bigotimes_{i=1}^n \left(\Omega_i, \sum_i \right) &= \left(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \sum_i \right) \end{aligned}$$

als das *Produkt der σ -Algebren*.

Beispiel 4.4 $\mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}(\mathbb{R})$, denn $\mathcal{B}(\mathbb{R}^n)$ wird von $(-\infty, a]$ erzeugt, aber diese Mengen sind alles Rechtecke.

Ist eine σ -Algebra \sum vorgegeben, so nennen wir die Elemente dann auch die \sum -*messbaren Mengen*.

Wir wollen uns nun überlegen, wie man das gleichzeitige Durchführen mehrerer Zufallsgeschehnisse vernünftig beschreibt. Wir nehmen dabei an, daß die Modelle für die einzelnen Experimente bereits erstellt sind. Solche Fragestellungen kommen in der Wahrscheinlichkeitstheorie ziemlich häufig vor, wie die folgenden Beispiele zeigen:

Beispiel 4.5 Mehrfaches Würfeln oder Würfeln mit mehreren Würfeln gleichzeitig.

Beispiel 4.6 Untersuchungen im Bereich der Sozial- oder Wirtschaftswissenschaften, wo nach mehreren Parametern gleichzeitig gefragt wird, wie etwa: Schulbildung, Einkommen, polit. Präferenz.

Es seien also (Ω_i, \sum_i, P_i) $i = 1, \dots, n$, Wahrscheinlichkeitsräume. Der Ereignisraum für das Nebeneinander Durchführen der Experimente $1, \dots, n$ ist dann offensichtlich $\Omega = \prod_{i=1}^n \Omega_i$, das kartesische Produkt der Ω_i , denn ein zusammengesetztes Elementarereignis hat offensichtlich die Form $(\omega_1, \dots, \omega_n) \in \Omega$.

Sind $E_i \in \sum_i$ die zugehörigen „natürlichen Ereignismengen“, so werden die entsprechenden Mengen genau $E_1 \times \dots \times E_n$ sein. Solche Mengen wollen wir verallgemeinerte Rechteckmengen nennen. Das System aller solcher Mengen sei E . Offensichtlich gilt dann

$$\sum(E) = \bigotimes_{i=1}^n \sum(E_i).$$

Bemerkung 4.3 Man überlegt sich leicht, daß für \otimes ein Assoziativgesetz gilt.

Mit Hilfe der endlichen Produkte können wir also Wiederholungen von Zufallsexperimenten beschreiben. Wir haben aber noch kein geeignetes Modell, das das ∞ häufige Wiederholen eines Experimentes bezeichnet. Dabei fassen wir allerdings ∞ häufig als Idealisierung einer beliebig großen Zahl von Versuchen auf.

Wir wollen die entsprechenden Begriffe nun einführen. Es seien (Ω_i, \sum_i) σ -Algebren und $A_i \in \sum_i$ für $i = 1, 2, \dots$. Für $m \in \mathbb{N}$ nennt man dann eine Menge der Form

$$A_1 \times A_2 \times \dots \times A_n \times \left(\prod_{i=n+1}^{\infty} \Omega_i \right)$$

eine *Rechteckzylindermenge*.

Definition 4.4 $\sum = \bigotimes_{i=1}^{\infty} \sum_i$, das ∞ -Produkt der \sum_i , ist die von allen Rechteckzylindermengen erzeugte σ -Algebra:

$$\sum = \bigotimes_{i=1}^{\infty} \sum_i = \sum \left(\bigcup_{n=1}^{\infty} \left(\prod_{i=1}^n \sum_i \right) \times \prod_{i=n+1}^{\infty} \Omega_i \right).$$

Zudem sei

$$\bigotimes_{i=1}^{\infty} (\Omega_i, \sum_i) = \left(\prod_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty} \sum_i \right).$$

Bemerkung 4.4 Die Verwendung der Rechteckzylindermengen entspricht genau der Definition der Produkttopologie.

Aufgabe 4.1 Welche σ -Algebra \sum wird auf $\Omega := \mathbb{Z}$ durch die Intervalle $[n, \infty) =: I_n$, $n \in \mathbb{Z}$ erzeugt?

4.2 Meßbare Abbildungen

Im Folgenden wollen wir uns mit Abbildungen zwischen Ereignisräumen beschäftigen. Solche Abbildungen spielen nicht nur grundsätzlich eine große Rolle, sondern sind auch in Anwendungen von großer Bedeutung.

Satz 4.1 Es sei $f : \Omega \rightarrow \Omega'$ eine Abbildung und Σ' eine σ -Algebra auf Ω' , d.h. $\Sigma' \subset \mathcal{P}(\Omega')$. Dann gilt

- i) $f^{-1}(\Sigma') = \{f^{-1}(A) | A \in \Sigma'\}$ ist eine σ -Algebra auf Ω , die von f und Σ' induzierte σ -Algebra.
- ii) Wird Σ' von $\{A'_i\}$ erzeugt, so wird $f^{-1}(\Sigma')$ von $\{f^{-1}(A'_i)\}$ erzeugt.
- iii) Für jede σ -Algebra Σ auf Ω ist $\{A' \in \Sigma' | f^{-1}(A') \in \Sigma\}$ eine σ -Algebra auf Ω' .

Beweis: (i) und (iii) folgt leicht aus

- a) $f^{-1}(\cap_I A'_i) = \cap_I f^{-1}(A'_i)$
- b) $f^{-1}(\cup_I A'_i) = \cup_I f^{-1}(A'_i)$
- c) $f^{-1}(\emptyset) = \emptyset, f^{-1}(\Omega') = \Omega$
- d) $f^{-1}(A'^c) = f^{-1}(A')^c$

(ii) Sicherlich gilt $f^{-1}(\Sigma') \supset \Sigma(\{f^{-1}(A'_i)\})$ da $f^{-1}(\Sigma')$ eine σ -Algebra ist, die $f^{-1}(A'_i)$ enthält. Es ist aber $\Sigma'' = \{A' \in \mathcal{P}(\Omega') | f^{-1}(A') \in \Sigma(\{f^{-1}(A'_i)\})\}$ eine σ -Algebra (nach iii) mit $A'_i \in \Sigma''$. Also $\Sigma' \subset \Sigma''$ und damit ist (ii) bewiesen.

Ist (Ω, Σ) eine σ -Algebra und $B \subset \Omega$, so kann man auf natürliche Weise eine σ -Algebra auf B bilden. Ist nämlich $f : B \rightarrow \Omega$ die natürliche Injektion $f(\omega) = \omega, \omega \in B$, so folgt wegen

$$\begin{aligned} f^{-1}(A) &= A \cap B \\ f^{-1}\left(\sum\right) &= \left\{A \cap B | A \in \sum\right\}. \end{aligned}$$

Aus Satz 3.1 folgt dann unmittelbar

Korollar 4.1 Es ist $B \cap \Sigma = \{B \cap A | A \in \Sigma\}$ eine σ -Algebra, die *Spur σ -Algebra* von Σ in B . Wird Σ von $\{A_i\}$ erzeugt, so $B \cap \Sigma$ von $\{B \cap A_i\}$.

Strukturerhaltende Abbildungen spielen in der Mathematik eine große Rolle. Der entsprechende Begriff hier ist der Begriff der messbaren Abbildung.

Definition 4.5 Es seien (Ω, Σ) und (Ω', Σ') Ereignisräume und σ -Algebren. Eine Abbildung $f : \Omega \rightarrow \Omega'$ heißt $\Sigma - \Sigma'$ *messbar*, falls $f^{-1}(A') \in \Sigma$ für alle $A' \in \Sigma'$. Dieser Begriff ist ähnlich definiert wie der Begriff stetige Funktion.

Falls $\Omega = \mathbb{R}^n, \Sigma = \mathcal{B}_n, \Omega' = \mathbb{R}^m, \Sigma' = \mathcal{B}_m$, so heißt eine $\mathcal{B}_n - \mathcal{B}_m$ messbare Funktion auch *Borel-messbar*. Entsprechendes gilt, falls

$$\Omega \subset \mathbb{R}^n, \Omega' \subset \mathbb{R}^m \text{ und } \Sigma = \mathcal{B}(\Omega), \Sigma' = \mathcal{B}(\Omega').$$

Definition 4.6 Es sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und (Ω', Σ') eine σ -Algebra. Eine $\Sigma - \Sigma'$ messbare Funktion f heißt auch eine Ω' -*Zufallsvariable*. Falls $\Omega' \subset \mathbb{R}$ und $\Sigma' = \mathcal{B}(\Omega')$, so heißt f einfach *Zufallsvariable*. Falls $\Omega' \subset \mathbb{R}^n$, so spricht man auch von einem *Zufallsvektor*.

Ist klar, was Σ und Σ' sind, so werden wir einfach von *messbaren* Funktionen reden. Falls $\Omega \subset \mathbb{R}^n$, so ist üblicherweise $\Sigma = \mathcal{B}(\Omega)$.

Satz 4.2 Eine Abbildung $f : (\Omega, \Sigma) \rightarrow (\Omega', \Sigma')$ ist genau dann meßbar, wenn $f^{-1}(A') \in \Sigma$ für alle A' eines Erzeugendensystems von Σ' gilt.

Beweis: $\{A' \in \Sigma' \mid f^{-1}(A') \in \Sigma\}$ ist eine σ -Algebra, die nach Annahme ein Erzeugendensystem von Σ' enthält. Also gilt $f^{-1}(\Sigma') \subset \Sigma$.

Korollar 4.2 Eine Abbildung $f : (\Omega, \Sigma) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$ ist genau dann $\Sigma - \mathcal{B}_n$ messbar, wenn $f^{-1}((-\infty, a]) \in \Sigma \forall a \in \mathbb{R}^n$.

Korollar 4.3 Für jede Abbildung $f : (\Omega, \Sigma) \rightarrow (\mathbb{R}, B)$ sind äquivalent:

- i) f ist $\Sigma - \mathcal{B}$ messbar
- ii) $[f \leq \alpha] = \{\omega \in \Omega \mid f(\omega) \leq \alpha\} \in \Sigma \quad \forall \alpha \in \mathbb{R}$
- iii) $[f > \alpha] \in \Sigma \quad \forall \alpha \in \mathbb{R}$
 $[f \geq \alpha] \in \Sigma \quad \forall \alpha \in \mathbb{R}$

Beispiel 4.7 Sei $\emptyset \neq B \subset \mathbb{R}^n$. $f : B \rightarrow \mathbb{R}$ heißt nach oben halbstetig, wenn $a_n, a \in B \dots$ und $\dots a_n \rightarrow a \Rightarrow \overline{\lim} f(a_n) \leq f(a)$. Für eine nach oben halbstetige Funktion f sind $[f \geq \alpha]$ abgeschlossen in B mit der relativen Topologie. Also ist $B \cap [f \geq \alpha]$ abgeschlossen, also in $B \cap \mathcal{B}_n$. Daher ist eine nach oben halbstetige Funktion $B \cap \mathcal{B}_n - \mathcal{B}$ messbar.

Beispiel 4.8 $\emptyset \neq B \subset \mathbb{R}^n, f : B \rightarrow \mathbb{R}^m$ stetig. Dann ist $f(B \cap \mathcal{B}_n - \mathcal{B}_m)$ -messbar, denn die Urbilder offener Mengen sind offen.

Beispiel 4.9 Es seien (Ω_i, Σ_i) σ -Algebren, und, $pr_i : \prod_{j=1}^n \Omega_j \rightarrow \Omega_i$ durch $pr_i((\omega)) = \omega_i$ definiert, $i = 1, \dots, n$. Dann ist die Projektion $pr_i(\prod_{j=1}^n \Sigma_j - \Sigma_i)$ messbar.

Beispiel 4.10 Es seien $(\Omega_i, \Sigma_i), i = 1, \dots, n$, σ -Algebren und $f_i : (\Omega, \Sigma) \rightarrow (\Omega_i, \Sigma_i)$ messbar. Dann ist $f : (\Omega, \Sigma) \rightarrow (\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \Sigma_i)$ mit $f(\omega) = (f_1(\omega), \dots, f_n(\omega)) \in \prod_{i=1}^n \Sigma_i$ messbar.

Beweis: Es sei $A_i \in \Sigma_i$ und $A_1 \times \dots \times A_n$ ein Rechteck. Dann ist $f^{-1}(A_1 \times \dots \times A_n) = \bigcap_{i=1}^n f_i^{-1}(A_i) \in \Sigma$. Die Rechtecke aber bilden ein Erzeugendensystem.

Die Produkte von σ -Algebren sind gerade so definiert, daß die Projektionsabbildungen meßbar sind.

Satz 4.3 Es seien $f : (\Omega, \Sigma) \rightarrow (\Omega', \Sigma')$ und $g : (\Omega', \Sigma') \rightarrow (\Omega'', \Sigma'')$ messbar. Dann ist $g \circ f : (\Omega, \Sigma) \rightarrow (\Omega'', \Sigma'')$ messbar.

Beweis: Sei $A'' \in \Sigma''$. Dann ist $(g \circ f)^{-1}(A'') = f^{-1}(g^{-1}A'') \in \Sigma, g^{-1}(A'') \in \Sigma'$.

Definition 4.7 Es sei (Ω, Σ) eine σ -Algebra. Eine Abbildung $\mu : \Sigma \rightarrow \mathbb{R}$ heißt *Maß*, falls

- i) $\mu(A) \geq 0 \quad \forall A \in \Sigma$
- ii) $\mu(\emptyset) = 0$
- iii) A_1, A_2, \dots disjunkt, $\Rightarrow \mu(\sum A_i) = \sum \mu(A_i)$

μ heißt *σ -endlich*, wenn $B_1 \subset B_2 \subset B_3 \dots$ existieren mit $\cup B_i = \Omega$ ($B_i \nearrow \Omega$) und $\mu(B_i) < \infty$. μ heißt *endlich*, falls $\mu(\Omega) < \infty$. Falls $\mu(\Omega) = 1$, so nennt man μ auch ein *Wahrscheinlichkeitsmaß*.

Satz 4.4 Es seien (Ω, Σ) und (Ω', Σ') σ -Algebren und $f : \Omega \rightarrow \Omega'$ messbar. Ferner sei μ ein Maß auf (Ω, Σ) . Dann definiert $\mu'(A') = \mu(f^{-1}(A'))$ ein Maß auf (Ω', Σ') . μ' wird das *Bildmaß* von μ unter f genannt, $\mu' = f\mu$ ist endlich (ein *Wahrscheinlichkeitsmaß*), falls μ es ist.

Beweis: Sicherlich gilt $\mu'(A') \geq 0$ und $\mu'(\emptyset) = 0$. Nun seien A'_1, A'_2, \dots disjunkt. Dann sind auch $f^{-1}(A'_1), f^{-1}(A'_2) \dots$ disjunkt und $f^{-1}(\cup A'_i) = \mu(f^{-1}(\cup A'_i)) = \mu(\cup f^{-1}(A'_i)) = \sum \mu(f^{-1}(A'_i)) = \sum \mu'(A'_i)$.

Im übernächsten Abschnitt werden wir dieses Problem wieder aufgreifen. Wir wollen hier lediglich 2 Beispiele betrachten.

Beispiel 4.11 Es sei $(\Omega = \{0, 1\}, \mathcal{P}(\Omega), P)$ der Wahrscheinlichkeitsraum für ein Bernoulliexperiment mit $P(0) = p, P(1) = q = 1 - p$. Auf $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$ betrachte die Funktion f , die die Anzahl der Nullen beschreibt: $f : \Omega^n \rightarrow \{0, 1, \dots, n\}$ mit $f(\omega) = n - \sum \omega_i$. Es sei dann $P' = f(P)$. Dann gilt $P'(k) = P(f^{-1}(k)) = b(n, k, p)$. Das Bildmaß von P unter f ist also die *Binomialverteilung*. Der Begriff *Verteilung* hat historische Gründe. *Verteilung* bedeutet dasselbe wie *Maß*. Wir sollten hier daher besser von dem *Binomialmaß* reden.

Wir können dieses Beispiel leicht verallgemeinern.

Beispiel 4.12 Es sei $\Omega = \{1, 2, \dots, l\}, P(i) = p_i, \sum p_i = 1$. Auf $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$ betrachte nun die Funktion $f : \Omega \rightarrow \{0, 1, 2, \dots, n\}^l$ mit $f(\cdot) = (\text{Anzahl } 1, \text{Anzahl } 2, \dots, \text{Anzahl } \ell), \sum \text{Anzahl } i = n$. Dann haben wir mit $P' = f(P)$

$$P'((k_1, \dots, k_l)) = \begin{cases} 0, & \text{falls } \sum k_i \neq n \\ \frac{n!}{k_1! k_2! \dots k_l!} P_1^{k_1} \dots P_l^{k_l}, & \text{sonst.} \end{cases}$$

Diese Verteilung wird die *Multinomialverteilung* genannt.

Aufgabe 4.2 Gegeben sei eine Menge von M roten und $N - M, N > M$, weißen Kugeln. Beschreibe das Ziehen von $n < N$ Kugeln und bestimme das Bildmaß der Funktion: Anzahl rote Kugeln, wenn das Maß in der ursprünglichen Menge das Zählmaß ist.

4.3 Messbare numerische Funktionen

In diesem Abschnitt wollen wir reellwertige messbare Funktionen auf σ -Algebren (Ω, Σ) betrachten. Dabei ist es zweckmäßig, auch $+\infty$ und $-\infty$ als Werte zuzulassen. Wir definieren daher

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$$

und fassen $+\infty$ als Häufungspunkt einer Folge auf, die beliebig große Werte annimmt. Entsprechend verfahren wir mit $-\infty$. Die Borelmengen $\mathcal{B}(\overline{\mathbb{R}})$ auf $\overline{\mathbb{R}}$ werden genau wie auf \mathbb{R} durch die Intervalle erzeugt.

Wir benötigen auch noch die wichtigsten Rechenregeln auf $\overline{\mathbb{R}}$. Dazu definieren wir die Addition und Multiplikation auf $\mathbb{R} \subset \overline{\mathbb{R}}$ wie gehabt und setzen

$$\begin{aligned} a + (\pm\infty) &= \pm\infty = (\pm\infty) + a & a \in \mathbb{R} \\ \pm\infty + (\pm\infty) &= \pm\infty & (\pm\infty) - (\pm\infty) \text{ undefiniert} \\ (\pm\infty) \cdot a &= \pm\infty & \text{falls } a > 0 \\ (\pm\infty)(\pm\infty) &= +\infty \\ (+\infty) \cdot 0 &= 0 & \frac{1}{\pm\infty} = 0, \frac{1}{0} = \infty. \end{aligned}$$

Es sei (Ω, Σ) eine σ -Algebra. Eine Funktion $f : \Omega \rightarrow \overline{\mathbb{R}}$ heißt *numerische Funktion*. Eine numerische Funktion heißt *messbar*, wenn sie $\Sigma - \mathcal{B}(\overline{\mathbb{R}})$ messbar ist.

Satz 4.5 Es seien f, g messbare Abbildungen von (Ω, Σ) nach $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Dann sind $\alpha f, f + g, f \cdot g$ und $\frac{f}{g}$ messbar, sofern $f + g$ überall definiert ist, d.h. es tritt für kein $\omega \in \Omega$ $f(\omega) = \pm\infty$ und $g(\omega) = \mp\infty$ ein.

Beweis: Es genügt zu zeigen, daß $[h \leq a] = \{\omega \in \Omega | h(\omega) \leq a\} \in \Sigma$ für jede der Funktionen $h = \alpha f, f + g, \dots$. Wir haben aber $[\alpha f \leq a] = [f \leq \frac{1}{\alpha} a] \in \Sigma$.

Ferner folgt gleichermaßen, daß $\alpha f + \beta$ messbar ist. Sind nun f, g messbar, so ist

$$[f < g] = \{\omega \in \Omega | f(\omega) < g(\omega)\} = \bigcup_{r \in \mathbb{Q}} [f < r] \cap [r < g]$$

messbar. Daher ist auch $[f \leq g] = [f > g]^c \in \Sigma$. Nun ist aber $[f + g \leq a] = [f \leq a - g] \in \Sigma$. Wegen $(f + g)^2 = f^2 + 2fg + g^2$ genügt es zu zeigen, daß f^2 messbar ist, denn die Ausnahmemengen können extra diskutiert werden. Es ist aber

$$[f^2 \leq a] = \begin{cases} \emptyset & a < 0 \\ [f = 0] & a = 0 \\ [f \leq \sqrt{a}] \cup [-f \leq \sqrt{a}] & \in \Sigma. \end{cases}$$

Satz 4.6

- a) Es sei $\{f_n\}_{n \in \mathbb{N}}$ eine Folge messbarer Funktionen. Dann ist jede der folgenden Funktionen messbar

$$\sup f_n, \inf f_n, \overline{\lim} f_n, \underline{\lim} f_n.$$

- b) Gilt $f_n \rightarrow f$ punktweise (in $\overline{\mathbb{R}}$), so ist auch f messbar.

Beweis: Es ist $[\sup f_n \leq a] = \bigcap_{n \in \mathbb{N}} [f_n \leq a]$. Also ist $\sup f_n$ messbar. Damit auch $\inf f_n = -\sup(-f_n)$. Nun ist aber

$$\underline{\lim} f_n = \sup_n \inf_{m \geq n} f_m \quad \text{und} \quad \overline{\lim} f_n = \inf_n \sup_{m \geq n} f_m.$$

Also sind auch diese Funktionen messbar.

Satz (3.5) und Satz (3.6) zeigen, wie groß die Klasse der messbaren Funktionen ist. Ist insbesondere $\Omega \subset \mathbb{R}^n$, so hatten wir gesehen, daß alle stetigen Funktionen messbar sind. Damit sind aber auch alle Funktionen, die daraus durch abzählbare Grenzprozesse gewonnen werden können, d.h. praktisch alle Funktionen, die in Anwendungen auftreten, messbar. Dazu gehören etwa die monotonen Funktionen, die stückweise stetigen Funktionen und solche, die nur abzählbar viele Unstetigkeitsstellen ohne Häufungspunkte haben. Bisher konnte man die Existenz nicht messbarer Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ nur mit Hilfe des Auswahlaxioms zeigen.

Die Sätze oben lassen sich leicht auf Funktionen $f : \Omega \rightarrow \overline{\mathbb{R}}^n$ erweitern, da in diesem Fall $f = (f_1, \dots, f_n)$ mit $f_i : \Omega \rightarrow \mathbb{R}$.

Unter allen messbaren Funktionen spielen die Treppenfunktionen eine besondere Rolle.

Definition 4.8 Eine Funktion $f : (\Omega, \Sigma) \rightarrow (\overline{\mathbb{R}}, B(\overline{\mathbb{R}}))$ heißt *Treppenfunktion*, wenn f messbar ist und nur endlich viele Werte annimmt.

Satz 4.7 Es sei $f : (\Omega, \Sigma) \rightarrow (\overline{\mathbb{R}}, B(\overline{\mathbb{R}}))$ messbar. Dann ist f punktweiser Grenzwert von Treppenfunktionen. Ist $f \geq 0$, so ist f sogar isotoner punktweiser Grenzwert von positiven Treppenfunktionen. Ist f beschränkt, so ist f sogar gleichmäßiger Grenzwert von Treppenfunktionen.

Beweis: Wir definieren

$$f_n = \sum_{k=-n2^n}^{n2^n} k \cdot 2^{-n} \chi_{[k2^{-n} \leq f_1 < (k+1)2^{-n}]} + n \chi_{[f \geq n]} - n \chi_{[f \leq -n]}$$

$$\text{wo} \quad \chi_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

die charakteristische Funktion von A ist. Es gilt offensichtlich $f_n(\omega) \leq f(\omega)$ für alle ω , wo $f(\omega) \geq -n$. Falls sogar $-n \leq f(\omega) \leq n$, haben wir $|f(\omega) - f_n(\omega)| \leq 2^{-n}$. Daraus ergibt sich die Behauptung leicht.

Aufgabe 4.3

- Jede monotone reellwertige Funktion ist $\mathcal{B}_1 - \mathcal{B}_1$ messbar.
- Jede stückweise stetige Funktion ist $\mathcal{B}_1 - \mathcal{B}_1$ messbar.
- Jede stetige Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ist $\mathcal{B}_n - \mathcal{B}_m$ messbar.

Aufgabe 4.4 Es seien $f_i : \Omega \rightarrow \Omega_i$ $\Sigma - \Sigma_i$ meßbare Abbildungen. Dann ist $f(\omega) = (f_1(\omega), \dots, f_n(\omega))$, $\omega \in \Omega$, als Abbildung von Ω nach $\Omega_1 \times \dots \times \Omega_n$ $\Sigma - \Sigma_1 \otimes \dots \otimes \Sigma_n$ meßbar.

4.4 Maße

Wir haben bisher die Existenz von Maßen auf σ -Algebren vorausgesetzt. Ein wichtiges praktisches Problem ist aber, die Existenz von Maßen mit bestimmten Eigenschaften zu zeigen. Da in der Regel die σ -Algebren Σ nur über ihre Erzeuger erklärt sind, geht man üblicherweise so vor:

1. Schritt: Das Maß wird auf einem kleinen System R , das die Erzeuger enthält und kontrollierbar ist, geeignet definiert.

2. Schritt: Das Maß wird von R auf $\Sigma(R)$ fortgesetzt und die Eindeutigkeit der Fortsetzung wird gezeigt.

Da in praktischen Problemen üblicherweise nur Mengen aus R vorkommen, wollen wir den zu Schritt 2 gehörigen Existenzsatz nicht beweisen, sondern nur angeben. Die entsprechenden Beweise dazu findet man im Buch von H. Bauer.

Definition 4.9 Ein System R von Teilmengen einer Menge heißt ein *Ring*, wenn

- i) $\emptyset \in R$
- ii) $A, B \in R \Rightarrow A \setminus B \in R$
- iii) $A, B \in R \Rightarrow A \cup B \in R$

R heißt *Algebra*, falls auch iv) $\Omega \in R$.

Es ist leicht zu sehen, daß ein Mengenring abgeschlossen ist unter endlichen Durchschnitten und Vereinigungen.

Ist \mathcal{E} ein System von Mengen, so kann man genau wie im Fall von σ -Algebren den von \mathcal{E} erzeugten Ring $R(\mathcal{E})$ bzw. die von \mathcal{E} erzeugte Algebra $A(\mathcal{E})$ definieren (Aufgabe).

Beispiel 4.13 Eine σ -Algebra ist eine Algebra.

Beispiel 4.14 \emptyset ist ein Ring.

Beispiel 4.15 Das System aller endlichen Teilmengen einer Menge ist ein Ring.

Beispiel 4.16 Für $a, b \in \mathbb{R}^n$ mit $a_i \leq b_i, i = 1, \dots, n$, bezeichne $[a, b) = [a_1, b_1) \times \dots \times [a_n, b_n)$. $[a, b)$ nennen wir ein halboffenes Rechteck. Das System aller endlichen Vereinigungen solcher halboffener Rechtecke ist ein Ring \mathcal{F}_n (Aufgabe) und wir nennen ihn den Ring der *n-dimensionalen Figuren*.

Definition 4.10 Eine Funktion $\mu : R \rightarrow \mathbb{R}$ auf einem Ring R heißt *Inhalt*, falls

- i) $\mu(\emptyset) = 0$
- ii) $\mu(A) \geq 0 \quad \forall A \in R$
- iii) $\mu\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$.

Gilt (iii) auch für $n = \infty$, so heißt der Inhalt ein *Maß*.

Lemma 4.3 Für einen Inhalt μ auf einem Ring R mit $\mu(A) < \infty \forall A \in R$ sind äquivalent:

- i) μ ist ein Maß.
- ii) $A_n \nearrow A, A_n, A \in R \Rightarrow \mu(A_n) \nearrow \mu(A)$, (d.h. $A_n \subset A_{n+1} \rightarrow A = \cup A_n$)
Stetigkeit von unten.
- iii) $A_n \searrow A, A, A_n \in R \Rightarrow \mu(A_n) \searrow \mu(A)$, (d.h. $A_n \supset A_{n+1}, A = \cap A_n$)
Stetigkeit von oben.
- iv) $A_n \searrow \emptyset, A_n \in R \Rightarrow \mu(A_n) \searrow 0$
 \emptyset -Stetigkeit.

Satz 4.8 Jedes Maß μ auf einem Ring $R \subset \mathcal{P}(\Omega)$ kann auf mindestens eine Weise zu einem Maß $\tilde{\mu}$ auf $\sum(R)$ fortgesetzt werden.

Ist μ sogar σ -endlich auf R , d.h. existieren $A_n \in R$ mit $A_n \nearrow \Omega$ und $\mu(A_n) < \infty$, so ist $\tilde{\mu}$ eindeutig. Wir wollen den Satz hier nicht beweisen (Den Beweis findet man im Buch von H. Bauer).

Beispiel 4.17 Es sei $\Omega = \mathbb{R}^n$ und \mathcal{F}_n der Ring aller n -dimensionalen Figuren. Dann läßt sich jedes $A \in R$ als endliche disjunkte Vereinigung halboffener Intervalle schreiben

$$A = A_1 \cup \dots \cup A_k \quad A_i = [a^{(i)}, b^{(i)}).$$

Definiert man nun $\lambda(A) = \sum_{i=1}^k \lambda(A_i)$, wo $\lambda([a, b)) = \prod_{i=1}^n (b_i - a_i)$, so ist λ ein *Inhalt* auf R . Dazu braucht man nur zu zeigen, daß $\lambda(A)$ unabhängig von der Darstellung $A = A_1 \cup \dots \cup A_k$ ist. Tatsächlich ist $\lambda(A)$ der *Riemannsche Inhalt* (Volumen) von A . Sind etwa

$$A = A_1 \cup \dots \cup A_k = B_1 \cup \dots \cup B_\ell,$$

beides disjunkte Vereinigungen von halboffenen Intervallen, so haben wir auch

$$A = \cup_{i,j} (A_i \cap B_j)$$

und es genügt $\lambda(A) = \sum_{i,j} \lambda(A_i \cap B_j)$ zu zeigen.

Dies aber gilt falls $\lambda(A_i) = \sum_j \lambda(A_i \cap B_j) \forall i$, d.h. wir können $A = [a, b)$ annehmen (Aufgabe). Dieser Inhalt ist sogar σ -endlich, denn wählt man $A_n = [-n, n) \times \dots \times [-n, n)$, so gilt $\lambda(A_n) = (2n)^n$ und $A_n \nearrow \mathbb{R}^n$. Damit also λ ein Maß auf R ist, genügt es also die \emptyset -Stetigkeit von λ zu zeigen.

Sei also $F_n \in R$ mit $F_n \searrow \emptyset$. Wir nehmen an $0 < \delta = \lim \lambda(F_n) = \underline{\lim} \lambda(F_n)$. Da jedes F_n endliche Vereinigung halboffener Intervalle ist, existieren

$$G_n \in R \text{ mit } \overline{G_n} \subset F_n \quad \text{und} \quad \lambda(F_n \setminus G_n) < \delta \cdot 2^{-n}.$$

Sei nun $H_n = G_1 \cap \dots \cap G_n$, so gilt $\overline{H_n} \subset F_n$ und $\overline{H_{n+1}} \subset \overline{H_n}$. Die $\overline{H_n}$ sind kompakt. Da $\cap F_n = \emptyset$ und $\overline{H_n} \subset F_n$, muß $\cap \overline{H_n} = \emptyset$ gelten. Dies aber ist nur möglich, wenn $\overline{H_{n_0}} = \emptyset$ für ein n_0 , denn die $\overline{H_n}$ sind kompakt. Man zeigt aber leicht durch vollständige Induktion nach n , daß

$$\lambda(H_n) \geq \lambda(F_n) - \delta(1 - 2^{-n}) \geq \delta \cdot 2^{-n}.$$

Dies aber ist unmöglich.

Definition 4.11 Die eindeutige Fortsetzung von λ auf $\mathcal{B}(\mathbb{R}^n)$ heißt das *Lebesgue-Borelmaß* λ .

Definition 4.12 Ein *Maßraum* ist ein Tripel (Ω, Σ, μ) bestehend aus einer nicht-leeren Menge Ω , einer σ -Algebra Σ auf Ω und einem Maß μ auf Σ . Der Maßraum heißt σ -endlich (endlich), wenn μ σ -endlich (endlich, d.h. $\mu(\Omega) < \infty$) ist.

Beispiel 4.18 $(\mathbb{R}^n, \mathcal{B}_n, \lambda)$ ist ein σ -endlicher aber nicht endlicher Maßraum.

Beispiel 4.19 Es sei $A \in \mathcal{B}_n$ eine Borelmenge mit $0 < \lambda(A) < \infty$. Dann ist $(A, \mathcal{B}(A), \lambda_A)$ mit $\lambda_A(B) = \lambda(A \cap B)$, $B \in \mathcal{B}(A)$ ein endlicher Maßraum.

Beispiel 4.20 Es sei Ω eine abzählbare Menge und $p : \Omega \rightarrow [0, \infty)$ eine nichttriviale Funktion. Dann ist $(\Omega, \mathcal{R}(\Omega), \mu)$ ein Maßraum, mit

$$\mu(A) = \sum_{\omega \in A} p(\omega).$$

Beispiel 4.21 Es sei Ω eine Menge und Σ eine σ -Algebra auf Ω . Ferner sei δ_{ω_0} das *Dirac Maß* in ω_0 durch

$$\delta_{\omega_0}(A) = \begin{cases} 0, & \omega_0 \notin A \\ 1, & \omega_0 \in A \end{cases}$$

definiert. Dann ist $(\Omega, \Sigma, \delta_{\omega_0})$ ein Wahrscheinlichkeitsraum.

Es sei (Ω, Σ, μ) ein Maßraum. Dann nennen wir $A \in \Sigma$ eine μ -Nullmenge, falls $\mu(A) = 0$. Offensichtlich gilt für $\mu(A) = 0$ und $B \in \Sigma$ auch $\mu(B \cap A) = 0$, d.h. jede messbare Teilmenge einer μ -Nullmenge ist wieder eine μ -Nullmenge. Es ist nun naheliegend, zu versuchen, die Einschränkung messbar hier aufzuheben. Dazu wollen wir das μ auf eine σ -Algebra $\tilde{\Sigma}$ fortzusetzen, die Σ sowie *alle* Teilmengen von μ -Nullmengen enthält.

Dazu sei $\mathcal{N} = \{A \subset \Omega \mid \exists B \in \Sigma \text{ mit } A \subset B, \mu(B) = 0\}$. \mathcal{N} ist das System der *verallgemeinerten Nullmengen*. \mathcal{N} ist abgeschlossen unter abzählbarer Vereinigung und mit $A \in \mathcal{N}$ und $B \in \Sigma$ ist auch $A \cup B \in \mathcal{N}$. Es bezeichne nun $\tilde{\Sigma} = \{A \cup N \mid A \in \Sigma, n \in \mathcal{N}\}$.

Dann ist $\tilde{\Sigma}$ eine σ -Algebra, denn

- i) $\Omega \in \Sigma \subset \tilde{\Sigma}$
- ii) Ist $A \cup N \in \tilde{\Sigma}$ und $N \subset B$ mit $B \in \Sigma$ und $\mu(B) = 0$, so haben wir $(A \cup N)^c = [(A \cup B) \cap (B \setminus N)^c]^c = (A \cup B)^c \cup (B \setminus N) \in \tilde{\Sigma}$.
- iii) Es sei $A_1 \cup N_n, A_2 \cup N_2, \dots \in \tilde{\Sigma}$. Dann ist

$$\bigcup_{i=1}^{\infty} (A_i \cup N_i) = \bigcup_{i=1}^{\infty} A_i \cup \left(\bigcup_{i=1}^{\infty} N_i \right) \in \tilde{\Sigma}$$

$\tilde{\Sigma}$ wird die μ -Vervollständigung von Σ genannt, denn:

Definition 4.13 Ein Maßraum (Ω, Σ, μ) heißt *vollständig*, falls $A \in \Sigma, \mu(A) = 0$ und $B \subset A \Rightarrow B \in \Sigma$ d.h. falls $\Sigma = \tilde{\Sigma}$.

Satz 4.9 Es sei (Ω, Σ, μ) ein Maßraum. Dann ist $(\Omega, \tilde{\Sigma}, \tilde{\mu})$ mit $\tilde{\mu}(A \cup N) = \mu(A)$ ein vollständiger Maßraum. $(\Omega, \tilde{\Sigma}, \tilde{\mu})$ wird die *Vervollständigung* von (Ω, Σ, μ) genannt. Die Vervollständigung ist σ -endlich (endlich), falls der Ausgangsmaßraum es war.

Beispiel 4.22 $(\mathbb{R}^n, \mathcal{B}_n, \lambda)$ ist *nicht* vollständig. Dazu sei $N_1 = \{x \in \mathbb{R} \mid 0 \leq x \leq 1, \text{ in der Dezimalentwicklung von } x \text{ treten nur } 0 \text{ und } 1 \text{ auf}\}$. Dann ist $N_1 \in \mathcal{B}_1$ und N_1 ist überabzählbar mit $\lambda(N_1) = 0$. Dann ist auch $N = N_1 \times \{0\} \times \dots \times \{0\} \in \mathcal{B}_n$ und N ist überabzählbar. Da $|\mathcal{B}_n| =$ die Kardinalzahl von $\mathcal{B}_n = c =$ Mächtigkeit des Kontinuums und da $|\mathcal{P}(N_1)| = 2^c > c$ (Cantor), gibt es verallgemeinerte Nullmengen, die nicht Borelmengen sind. Eine Menge, die bezüglich $\mathcal{B}(\tilde{\mathbb{R}}^n)$ messbar ist, heißt *Lebesgue messbar* und $\tilde{\lambda}$ heißt das *Lebesgue Maß*.

Bemerkung 4.5 Mit Hilfe des Auswahlaxioms kann man die Existenz nicht Lebesgue messbarer Mengen zeigen.

Beispiel 4.23 Die Vervollständigung von $(\Omega, \Sigma, \delta_{\omega_0})$ ist $(\Omega, \mathcal{P}(\Omega), \delta_{\omega_0})$, falls $\{\omega_0\} \in \Sigma$.

Bemerkung 4.6 Während die Konstruktion der Borel σ -Algebren unabhängig vom gegebenen Maß ist, hängt die Vervollständigung vom vorgegebenen Maß ab.

Bei der Vervollständigung und auch sonst in der Theorie spielen die Mengen vom Maß 0 eine große Rolle. Wir definieren daher:

Definition 4.14 Es sei (Ω, Σ, μ) ein Maßraum. Eine Eigenschaft E gilt μ *fast überall* auf Ω , wenn es eine μ -Nullmenge A gibt, so daß E auf $\Omega \setminus A$ gilt.

Aufgabe 4.5 Man zeige, daß $A \subset \mathbb{R}^n$ eine Lebesgue Nullmenge ist, falls es zu jedem $\varepsilon > 0$ halboffene Intervalle I_n gibt mit $\sum_{n=1}^{\infty} \lambda(I_n) \leq \varepsilon$ und $A \subset \bigcup_{n=1}^{\infty} I_n$.

Aufgabe 4.6 Jede Hyperebene im \mathbb{R}^n ist eine λ Nullmenge.

Aufgabe 4.7 Zeige, daß das Lebesgue Maß der Cantormenge 0 ist.

Aufgabe 4.8 Sei Ω eine überabzählbare Menge und Σ das System aller Mengen A mit A oder $\Omega \setminus A$ abzählbar. Man zeige: Σ ist eine σ -Algebra.

Weiter sei $\mu : \Sigma \rightarrow \mathbb{R}$ mit

$$\mu(A) = \begin{cases} 0 & \text{falls } A \text{ abzählbar ist} \\ 1 & \text{falls } \Omega \setminus A \text{ abzählbar ist.} \end{cases}$$

Man zeige: μ ist ein Wahrscheinlichkeitsmaß auf (Ω, Σ) . Ist (Ω, Σ, μ) vollständig?

Aufgabe 4.9 Man zeige:

- Jede abzählbare Menge in \mathbb{R}^n ist borelsch und hat das Maß 0.
- Sei λ^n das Lebesgue-Borelmaß auf $B(\mathbb{R})^n$. Dann gilt: $\lambda^n(A) < \infty$ für alle $A \in B(\mathbb{R}^n)$ mit A beschränkt.

Aufgabe 4.10 Es sei A die Menge aller reellen Zahlen aus $[0, 1]$, deren Dezimalentwicklung keine 9 enthält. Man zeige:

- $A \in \mathcal{B}(\mathbb{R})$
- $\lambda^1(A) = 0$

Aufgabe 4.11 Wähle speziell $\Omega := [0, 1]$ und μ wie in Aufgabe 3.8. Bestimme das Bildmaß der Funktion

$$f(x) = \begin{cases} 1 & x \text{ rational} \\ 0 & \text{sonst} \end{cases}$$

Aufgabe 4.12 Man zeige: Eine Menge $M \subset \mathbb{R}^N$ ist eine Lebesgue-Nullmenge, wenn für jedes $\varepsilon > 0$ eine Folge $(I_n)_{n \in \mathbb{N}} \subset \mathbb{R}^N$ von offenen (bzw. kompakten bzw. halboffenen) Intervallen existiert mit

$$\sum_{n=1}^{\infty} \lambda^N(I_n) \leq \varepsilon \text{ und } M \subset \bigcup_{n=1}^{\infty} I_n.$$

Aufgabe 4.13 Mit Hilfe von Aufgabe 3.12 zeige man: Sei $f : \mathbb{R}^{N-1} \rightarrow \mathbb{R}$ stetig ($N \in \mathbb{N}, N \geq 2$) und $M := \{(x, y) \in \mathbb{R}^{N-1} \times \mathbb{R} : y = f(x)\}$, dann gilt: $M \in \mathcal{L}(\mathbb{R}^N)$ und $\lambda^N(M) = 0$.

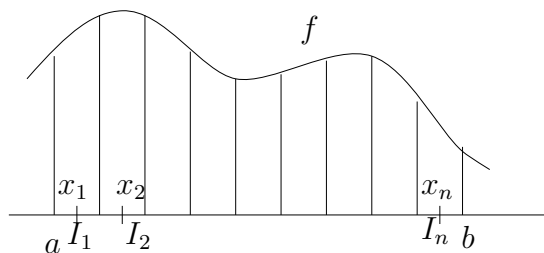
Aufgabe 4.14 Es seien (Ω, Σ, μ) ein Maßraum und (Ω', Σ') ein Maßraum. Weiter sei $T : \Omega \rightarrow \Omega'$ eine meßbare Abbildung mit $\mu' := T(\mu)$ das Bildmaß von μ unter T . Mit $(\Omega, \Sigma_0, \mu_0)$ bzw. $(\Omega', \Sigma'_0, \mu'_0)$ werde die Vervollständigung von (Ω, Σ, μ) bzw. (Ω', Σ', μ') bezeichnet. Man zeige:

- T ist $\Sigma_0 - \Sigma'_0$ meßbar.
- $T(\mu_0) = \mu'_0$.

4.5 Integrationstheorie

Um das Riemann Integral einer Funktion f zu bestimmen, zerlegt man das Integrationsintervall in kleinere Intervalle und bestimmt die Riemannsche Zwischensumme

$$\begin{aligned} Z(f) &= \sum_{i=1}^n f(x_i) |I_i|, \\ |I_i| &= \text{Volumen des Intervalls } I_i, \\ x_i &\in I_i. \end{aligned} \quad (4.1)$$



Für feiner werdende Zerlegungen strebt dann

$$Z(f) \rightarrow \int f \, dx$$

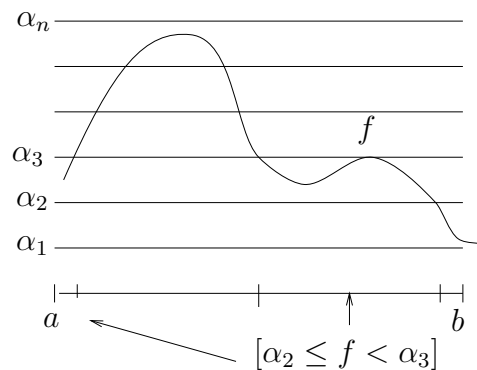
sofern f R -integrierbar ist. Die Bestimmung von $Z(f)$ nach diesem Verfahren ist etwas umständlich, da Werte $f(x_i)$ mehrfach auftreten können.

Diesen Nachteil beseitigt das Lebesgue Summationsverfahren. Bei diesem Verfahren wird der Wertebereich der Funktion unterteilt. Die entsprechende Zwischen-summe ist dann

$$\tilde{Z}(f) = \sum_{i=1}^n \alpha'_i \mu([\alpha_i \leq f < \alpha_{i+1}]), \tag{4.2}$$

wo $\alpha_1 \leq f(\omega) < \alpha_n$

$\alpha'_i \in [\alpha_i, \alpha_{i+1})$.



Der Vorteil bei diesem Verfahren besteht in

- i) Das Verfahren ist auf beliebige Maße auf beliebigen Maßräumen anwendbar, d.h. man kann in einem Zug die Integration auf dem \mathbb{R}^1 und \mathbb{R}^n mit erledigen.
- ii) Das Summationsverfahren ist ökonomischer
- iii) Der so entstehende Integralbegriff ist allgemeiner und hat bessere Konvergenzeigenschaften
- iv) Uneigentliche Integrale werden sofort mitbehandelt.

Erreicht wird das dadurch, daß man nicht mehr von Intervallen (Rechtecken) als Grundbausteinen der Theorie, sondern von allgemeinen messbaren Mengen der Form $[\alpha \leq f < \beta]$ ausgeht.

Nun kann man (??) bzw. (??) auch anders interpretieren und dies wird üblicherweise auch so gemacht. Betrachten wir in (1) nämlich die Rechteck-Treppenfunktion, d.h. Treppenfunktionen, die mit charakteristischen Funktionen von Rechtecken gebildet werden,

$$F = \sum_{i=1}^n f(x_i) \chi_{I_i} \quad \chi_{I_i}(\omega) = \begin{cases} 1 & \omega \in I_i \\ 0 & \omega \notin I_i \end{cases} \tag{4.3}$$

so ist

$$Z(f) = R - \int F.$$

Dabei bezeichnet $R - \int$ das Riemann Integral.

Die Grundbausteine des Riemann Integrals sind also Treppenfunktionen über Rechtecken (Intervallen). Es gilt sogar

Lemma 4.4 Eine Funktion $f : A \rightarrow \mathbb{R}$, $A = [a, b] \subset \mathbb{R}^n$ ist genau dann R -integrierbar, wenn zu jedem $\varepsilon > 0$ Rechteck-Treppenfunktionen f_1, f_2 existieren mit $f_2 \geq f \geq f_1$ und $\int_A f_2 - f_1 < \varepsilon$.

Dagegen wird man in der allgemeinen Maß- und Integrationstheorie für (??) die Treppenfunktionen

$$\tilde{F} = \sum \alpha'_i \chi_{[\alpha_i \leq f \leq \alpha_{i+1}]} \quad (4.4)$$

zugrunde legen. Dies geht, weil, wie wir in Kapitel 4.3 bereits gesehen hatten, jede messbare Funktion f geeignet durch Treppenfunktionen approximiert werden kann. Beginnen wir also mit dem Integralbegriff für Treppenfunktionen. Da die Addition auf $\overline{\mathbb{R}}$ für $\pm\infty + (\mp\infty)$ nicht erklärt ist, schränken wir uns zunächst nur auf positive Funktionen ein.

Sei also (Ω, Σ, μ) ein Maßraum. Die Menge aller meßbaren Treppenfunktionen u , d.h. Funktionen mit einer Darstellung der Form $u = \sum_{i=1}^n a_i \chi_{A_i}$, $A_i \in \Sigma$, bezeichnen wir mit $T(\Omega, \Sigma) = T$. Die Menge aller solcher Funktionen mit $a_i \geq 0$ bezeichnen wir mit T_+ . Es ist leicht zu sehen, daß eine Funktion $u \in T$ in der Regel mehrere Darstellungen

$$u = \sum_{i=1}^n a_i \chi_{A_i} \quad \text{mit} \quad A_i \in \Sigma \quad \text{disjunkt}$$

hat. Trotzdem aber gilt

Lemma 4.5 Es sei $u = \sum_{i=1}^n a_i \chi_{A_i} = \sum_{j=1}^m b_j \chi_{B_j}$, $A_i \in \Sigma$ disjunkt, $a_i, b_j \geq 0$. Dann gilt

$$\sum_{i=1}^n a_i \mu(A_i) = \sum_{j=1}^m b_j \mu(B_j) \in \overline{\mathbb{R}}.$$

Der Beweis des Lemmas sei als Übungsaufgabe gestellt.

Definition 4.15 Für ein $u = \sum_{i=1}^n a_i \chi_{A_i} \in T_+$, $a_i \geq 0$ nennen wir die von der speziellen Darstellung von u unabhängige Zahl $\sum a_i \mu(A_i) \in \overline{\mathbb{R}}$ das μ -Integral von u über Ω und bezeichnen es mit $\int u d\mu$.

Lemma 4.6 Es seien $u, v \in T_+$ und $a \geq 0$. Dann gilt

- i) $u + av \in T_+$ und $\int (av + u) d\mu = a \int v d\mu + \int u d\mu$
- ii) $u \leq v \Rightarrow \int u d\mu \leq \int v d\mu$
- iii) $\int \chi_A d\mu = \mu(A)$.

Wir sind bisher immer von der Darstellung $u = \sum_{i=1}^n a_i \chi_{A_i}$, A_i disjunkt ausgegangen.

Es ist aber eine einfache Konsequenz von Lemma 4.5, daß für jede Folge $A_1, \dots, A_n \in \Sigma$ und $a_i \in \mathbb{R}_+$ stets $u = \sum a_i \chi_{A_i} \in T_+$ und $\int u d\mu = \sum a_i \mu(A_i)$.

Das folgende Lemma ermöglicht es uns, den Integralbegriff auf eine größere Klasse von Funktionen auszudehnen.

Lemma 4.7 Für jede monoton wachsende Folge $(u_n)_{n \in \mathbb{N}}$ von positiven Treppenfunktionen und jedes $v \in T_+$ mit $v \leq \sup u_n$ gilt

$$\int v d\mu \leq \sup \int u_n d\mu.$$

Beweis: Es sei $v = \sum_{i=1}^n a_i \chi_{A_i}$, $A_i \in \Sigma$ disjunkt und $1 > a > 0$. Ferner sei $B_n = [u_n \geq av]$. Dann gilt nach Annahme $B_n \nearrow \Omega$ also $B_n \cap A_j \nearrow A_j$ und $\int v d\mu = \sum_{i=1}^n a_i \mu(A_i) = \lim_n \sum a_i \mu(A_i \cap B_n) = \lim_n \int v \cdot \chi_{B_n} d\mu$.

Wegen Lemma 3.6 ii) gilt aber $av \chi_{B_n} \leq u_n \chi_{B_n} \leq u_{n+k} \chi_{B_n}$ und

$$a \int v \chi_{B_n} d\mu \leq \int u_{n+k} \chi_{B_n} d\mu \leq \int u_n d\mu.$$

Also gilt dies $a \int v d\mu < \sup \int u_n d\mu$. Da $0 < a < 1$ beliebig war, folgt die Behauptung.

Korollar 4.4 Für je zwei monoton wachsende Folgen (u_n) und (v_n) aus T_+ mit $\sup_n u_n = \sup_m v_m$ gilt

$$\sup_n \int u_n d\mu = \sup_m \int v_m d\mu.$$

Beweis: Es gilt $v_m \leq \sup u_n$ oder $\int v_m d\mu \leq \sup \int u_n d\mu$. Dies zeigt $\sup_m \int v_m d\mu \leq \sup \int u_n d\mu$. Vertauscht man die Rolle von u_n und v_n , erhält man das Ergebnis.

Wir haben schon früher gesehen, daß jede positive messbare numerische Funktion punktweiser Grenzwert einer monoton wachsenden Folge positiver Treppenfunktionen ist. Umgekehrt ist klar, daß jede Einhüllende einer Folge positiver Treppenfunktionen messbar ist. Ist daher $M_+(\Omega, \Sigma) = M_+$ das System aller messbaren positiven Funktionen auf Ω , und $f \in M_+$ mit $f = \sup_n u_n$, und $u_n \in T_+$ monoton wachsend, so nennen wir die Zahl $\sup \int f u_n d\mu \in \overline{\mathbb{R}}$ das μ -Integral über f und bezeichnen es mit

$$\int f d\mu = \sup \int u_n d\mu \quad \text{wo } u_n \nearrow f. \quad (4.5)$$

Wegen des Korollars zu Lemma 4.7 ist $\int f d\mu$ unabhängig von der gewählten Folge u_n .

Aus Lemma 4.5 erhält man sofort für diesen Integralbegriff und $f, g \in M_+$

$$\begin{aligned} \int f + ag d\mu &= \int f d\mu + a \int g d\mu \\ f \leq g &\Rightarrow \int f d\mu \leq \int g d\mu. \end{aligned} \quad (4.6)$$

Satz 4.10 (*B. Levi*): Es sei f_n eine monoton wachsende Folge von Funktionen aus M_+ mit $f_n \nearrow f$. Dann gilt

$$\int \sup_n f_n d\mu = \sup \int f_n d\mu. \quad (4.7)$$

Beweis: Wähle Funktionen $u_{n,m} \in T_+$ mit $u_{n,m} \nearrow f_n$ und definiere $v_m = \sup(u_{1,m}, \dots, u_{m,m})$ (Cantorsches Diagonalverfahren). Es ist $v_m \in T_+$ und da die $(u_{n,m})_{m \in \mathbb{N}}$ monoton wachsen, gilt dies auch für die v_m . Es gilt sicher $v_m \leq f$ und damit $\sup v_m \leq f$. Ferner haben wir offensichtlich $f_n \leq \sup f_m$ und somit $f \leq \sup v_m$. Insgesamt gibt dies $\sup v_m = f$ und aus (??) folgt $\int f_n d\mu \leq \int f d\mu$ oder $\sup \int f_n d\mu \leq \int f d\mu$. Es ist aber $u_{n,m} \leq f_n \leq f_m$ für $m \leq n$. Also gilt

$$\int v_m d\mu \leq \int f_m d\mu \leq \sup \int f_m d\mu \text{ oder } \int f d\mu \leq \sup \int f_n d\mu.$$

Korollar 4.5 Für jede Folge $g_n \in M_+$ gilt

$$\sum_{n=1}^{\infty} \int g_n d\mu = \int \left(\sum g_n \right) d\mu.$$

Beweis: Wende den Satz von B. Levi auf $f_n = \sum_{k=1}^n g_k$ an.

Nach diesen Vorarbeiten können wir nun das μ -Integral auf einer größeren Klasse von Funktionen definieren. Ist f eine messbare Funktion, so sei

$$f^+ = \max(f, 0) \quad \text{und} \quad f^- = -\min(f, 0)$$

der *Positivteil* bzw. *Negativteil* von f . Es gilt $f = f^+ - f^-$, $f^+ \cdot f^- = 0$ und $|f| = f^+ + f^-$ und f^+ und f^- sind messbar. Ferner gilt $f_+, f_- \geq 0$. Damit aber können wir definieren:

Definition 4.16 f heißt μ -integrierbar, wenn sowohl f^+ als auch f^- μ -integrierbar und $\int f^+ d\mu < \infty$ und $\int f^- d\mu < \infty$ sind. In diesem Fall bezeichnet

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu$$

das μ -Integral von f .

Lemma 4.8 Für eine messbare numerische Funktion f sind äquivalent:

- i) f ist μ integrierbar
- ii) Es gibt integrierbare Funktionen $u, v \geq 0$ mit $f = u - v$
- iii) $|f|$ ist integrierbar
- iv) Es existiert eine integrierbare Funktion g mit $g \geq |f|$.

Das Lemma sei als Übungsaufgabe gestellt.

Satz 4.11 Es seien f, g integrierbar und $a \in \mathbb{R}$. Dann ist $f + ag$ integrierbar und

$$\int (f + ag) d\mu = \int f d\mu + a \int g d\mu.$$

Ist ferner $f \leq g$, so gilt $\int f d\mu \leq \int g d\mu$. Speziell haben wir

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Der Satz ist eine einfache Konsequenz von Lemma 4.7.

Wir wollen uns einige Beispiele dazu betrachten.

Beispiel 4.24 Es sei $(\Omega, \Sigma, \mu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ mit $\mu(\{n\}) = 1$, das *Zählmaß* auf \mathbb{N} . Eine Funktion $f : \mathbb{N} \rightarrow \mathbb{R}$ (reelle Zahlenfolge) ist genau dann integrierbar, wenn die Reihe $\sum f(n)$ absolut konvergiert. In diesem Fall ist

$$\sum_{n=1}^{\infty} f(n) = \int f d\mu.$$

Beispiel 4.25 Wegen Lemma (4.4) ist jede über $[a, b] \subset \mathbb{R}^n$ eigentlich R -integrierbare Funktion integrierbar bezüglich des Lebesgue Borelmaßes λ auf $[a, b]$. In diesem Fall gilt

$$R - \int f dx = \int f d\lambda, \quad (4.8)$$

denn nach der Situation aus Lemma 4.1 gilt: f ist messbar und beschränkt.

Wegen Lemma (4.8) und $\lambda([a, b]) = \prod_{i=1}^n (b_i - a_i) < \infty$ ist damit f -Lebesgue Borel integrierbar. Da für Rechtecktreppenfunktionen g gilt $R - \int g dx = \int g d\lambda$, haben wir somit

$$R = \int f_1 dx = \int f_1 d\lambda \leq \int f d\lambda \leq \int f_2 d\lambda = R - \int f_2 dx$$

und daraus folgt (??). Das Integral $\int f d\lambda$ nennt man das Lebesgue Integral.

Beispiel (4.25) gibt uns eine effektive Methode an, Integrale zu berechnen (Hauptsatz der Differential- und Integralrechnung). In Beispiel 4.24 haben wir gesehen, daß jede R -integrierbare Funktion f auf einer beschränkten Menge auch Lebesgue integrierbar ist. Daß das L -Integral (Lebesgue Integral) eine echte Verallgemeinerung des R -Integrals ist, zeigt

Beispiel 4.26 Es sei $A = [0, 1] \cap \mathbb{Q}$ und $f = \chi_A$. Da A abzählbar ist, gilt $A \in \mathcal{B}(\mathbb{R})$ und f ist Borelsch und $\int f d\lambda = \lambda(A) = 0$. Andererseits ist f nicht R -integrierbar.

Über den Zusammenhang zwischen R -Integral und Lebesgue Integral (L -Integral) können wir somit sagen:

- i) Jede eigentlich R -integrierbare Funktion ist messbar
- ii) Der Begriff L -Integral entspricht der absoluten Integrierbarkeit bei R -Integralen. In der Tat ist eine R - und L -integrale Funktion absolut R -integrel.
- iii) Jede absolut R -integrale Funktion ist L -integrel. Die Umkehrung aber gilt nicht notwendig. (Beispiel 4.25).

Es sei (Ω, Σ, μ) ein Maßraum. Die Menge aller messbaren numerischen Funktionen f , für die $\int f d\mu$ existiert, bezeichnen wir mit $L^1(\Omega, \Sigma, \mu) = L^1$.

Satz 4.12

- a) $f, g \in L^1 \Rightarrow a \in \mathbb{R} f + ag \in L^1$

- b) Für $f \in L^1$ bezeichne $|f|_1 = \int |f| d\mu$. Dann gilt
- i) $|f + ag|_1 \leq |f|_1 + |a| |g|_1$
 - ii) $|f|_1 \Leftrightarrow 0 = f = 0 \mu$ - fast überall, d.h. es existiert eine Menge $A \in \Sigma$ mit $\mu(A) = 0$ und $f|_{A^c} = 0$
- c) Es sei $(f_n) \subset L^1$ eine Folge von Funktionen mit $|f_n - f_m|_1 \rightarrow 0 (n, m \rightarrow \infty)$. Dann existiert ein $f \in L^1$ und eine Teilfolge (f_{n_k}) mit $(f_{n_k}) \rightarrow f \mu$ -f.ü. d.h. $\exists A \in \Sigma$ mit $\mu(A) = 0$ und $(f_{n_k})(\omega) \rightarrow f(\omega) \forall \omega \notin A$.

Diesen Satz wollen wir hier nicht beweisen. Wir bezeichnen mit \mathcal{N} die Menge aller Funktionen, die fast überall 0 sind. Dann ist \mathcal{N} ein Vektorraum und $L^1(\Omega, \Sigma, \mu)/\mathcal{N} = L^1/\mathcal{N} = \mathcal{L}^1$ ist ein reeller Vektorraum, auf dem eine Norm

$$|f + \mathcal{N}|_1 = |f|_1$$

erklärt ist. $(\mathcal{L}^1, |\cdot|_1)$ ist vollständig (Satz 4.12 c).

Es sei (Ω, Σ, μ) ein Maßraum und $A \in \Sigma$. Dann definieren wir

$$\int_A (f|_A) d\mu = \int f \cdot \chi_A d\mu, \quad \chi_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

und nennen dies das μ -Integral von f über A . Dies ist sinnvoll, weil jede Σ -messbare numerische Funktion auch $A \cap \Sigma = \{A \cap B | B \in \Sigma\}$ messbar ist, wenn sie auf A eingeschränkt wird.

4.6 Konvergenzsätze

Wir haben bisher zwei Konvergenzsätze für numerische Funktionen auf Maßräumen kennengelernt. Dies waren:

Satz 4.13 (*Satz von Beppo Levi*): Es sei $(f_n)_{n \in \mathbb{N}}$ eine monoton wachsende Folge von Funktionen aus $T_+(\Omega, \Sigma, \mu)$. Dann gilt

$$\int_{\Omega} (\sup f_n) d\mu = \sup_n \int_{\Omega} f_n d\mu.$$

Satz 4.14 Es sei (f_n) eine Folge messbarer numerischer Funktionen auf (Ω, Σ, μ) mit $|f_n|_1 < \infty$ und $|f_n - f_m|_1 = \int |f_n - f_m| d\mu \rightarrow 0$ für $n, m \rightarrow \infty$.

Dann existiert eine messbare Funktion f mit $|f|_1 < \infty$ und eine Teilfolge (f_{n_k}) mit $f_{n_k} \rightarrow f \mu$ -f.ü.

Für die Anwendungen brauchen wir noch

Lemma 4.9 (*Lemma von Fatou*): Für jede Folge (f_n) messbarer Funktionen gilt:

$$\int \underline{\lim} f_n d\mu \leq \underline{\lim} \int f_n d\mu.$$

Beweis: Es ist $f = \underline{\lim} f_n$ messbar. Es ist auch $g_n = \inf_{m \geq n} f_m$ messbar. Offensichtlich ist g_n monoton wachsend mit $g_n \nearrow f$. Also gilt wegen Satz 4.14 $\int f d\mu = \sup_n \int g_n d\mu \leq \int f_m d\mu$ für $m \geq n$. Also $\int f d\mu \leq \underline{\lim} \int f_n d\mu$.

Der bei weitem wichtigste Satz in dieser Theorie ist der Satz über die *dominierte Konvergenz von H. Lebesgue*.

Satz 4.15 Es sei $(f_n) \subset L^1$ μ -fast überall konvergent auf Ω und es existiere eine messbare numerische Funktion g auf Ω mit

- i) $|f_n| \leq g$ und
- ii) $\int g d\mu < \infty$

Dann existiert eine messbare reelle Funktion f auf Ω mit $f_n \rightarrow f \mu - f.\ddot{u}$. und $\int |f| d\mu < \infty$ und $\int |f_n - f| d\mu \rightarrow 0$ für $n \rightarrow \infty$.

Beweis: Nach Annahme existiert eine μ -Nullmenge $A_1 \in \Sigma$ mit $\lim_n f_n(\omega)$ existiert für alle $\omega \in A_1^c$. Ferner existiert eine μ -Nullmenge $A_2 \in \Sigma$ mit $g(\omega) < \infty$ für $\omega \in A_2^c$. Nun definiere f durch

$$f(\omega) = \begin{cases} \lim_n f_n(\omega) & \omega \in (A_1 \cup A_2)^c \\ 0 & \text{sonst.} \end{cases}$$

Dann ist f messbar und $f_n \rightarrow f \mu - f.\ddot{u}$. Ferner gilt $|f| \leq g \mu - f.\ddot{u}$. oder $f \in L^1(\Omega, \Sigma, \mu)$. Setze nun $g_n = |f_n - f|$. Dann gilt

$$0 \leq g_n \leq |f_n| + |f| \leq |f| + g$$

und mit $h = |f| + g$ haben wir

$$\int \underline{\lim} (h - g_n) d\mu \leq \underline{\lim} \int h - g_n d\mu = \int h d\mu - \overline{\lim} \int g_n d\mu.$$

Da $f_n \rightarrow f \mu - f.\ddot{u}$. gilt $h - g_n \rightarrow h \mu - f.\ddot{u}$. Also gilt $\mu - f.\ddot{u} \underline{\lim} (h - g_n) = h$ oder $\int \underline{\lim} (h - g_n) d\mu = \int h d\mu$. Also folgt $\overline{\lim} \int g_n d\mu = 0$. Da $g_n \geq 0$ ist der Beweis erbracht.

Aufgabe 4.15 Seien $f_n, g_n : \mathbb{R} \rightarrow \mathbb{R} (n \in \mathbb{N})$ mit

$$f_n(x) := \begin{cases} 0 & x \leq 0, x \geq \frac{2}{n} \\ n & x = \frac{1}{n} \\ \text{stetig linear} & \text{sonst} \end{cases} \quad \text{und} \quad g_n(x) := \begin{cases} 0 & x \leq 0, x \geq \frac{2}{n} \\ 1 & x = \frac{1}{n} \\ \text{stetig linear} & \text{sonst} \end{cases}$$

Man zeige:

- a) $f_n \rightarrow 0, g_n \rightarrow 0 (n \rightarrow \infty)$ punktweise
- b) $1 = \int f_n d\lambda^1$ für alle $n \in \mathbb{N}$. Warum ist der Lebesguesche Konvergenzsatz hier nicht anwendbar?

- c) $\int g_n d\lambda^1 \rightarrow 0 (n \rightarrow \infty)$. Warum ist hier der Lebesguesche Konvergenzsatz anwendbar?

Aufgabe 4.16

- a) Eine auf \mathbb{R} definierte, Borel-meßbare, reelle Funktion $f \geq 0$ sei Riemann-integrierbar auf jedem kompakten Intervall. Man zeige: Es gilt $f \in L^1(\lambda^1)$ genau dann, wenn das uneigentliche Riemann-Integral $\int_{\mathbb{R}} f dx$ existiert. In diesem Fall gilt dann $\int_{\mathbb{R}} f dx = \int f d\lambda^1$.

- b) Sei f eine Borel-meßbare, reelle Funktion auf \mathbb{R} mit absolut konvergentem uneigentlichem Riemann-Integral. Man zeige:

$$f \in L^1(\lambda^1) \quad \text{und} \quad \int_{\mathbb{R}} f dx = \int f d\lambda^1.$$

Aufgabe 4.17 Man zeige mit Hilfe von Aufgabe 4.16:

- a) $x \rightarrow \exp(-x^\alpha)$ ist λ^1 -integrierbar über \mathbb{R}^+ für alle $\alpha > 0$.
 b) $x \rightarrow x \exp(-x^2)$ ist λ^1 -integrierbar über \mathbb{R}^+ . Man berechne

$$\int_0^{\infty} x e^{-x^2} d\lambda^1.$$

Aufgabe 4.18 Ein Spieler wettet mit Ihnen um einen festen Betrag, daß Sie bei 20-maligem Münzenwerfen 9, 10 oder 11 mal Kopf werfen. Würden Sie die Wette annehmen?

Aufgabe 4.19 Es wird 12 mal gewürfelt. Wie groß ist die Wahrscheinlichkeit genau 2 mal eine „1“, 2 mal eine „2“, ..., 2mal eine „6“ zu würfeln?

Aufgabe 4.20 Gegeben sei der Maßraum $(\Omega, \Sigma, \delta_\omega)$ für unfestes $\omega \in \Omega$. Dabei ist δ_ω das Dirac-Maß in ω . Man zeige: Sei f meßbar, dann gilt:

- a) $f \in L^1(\delta_\omega)$ genau dann, wenn $|f(\omega)| < \infty$.
 b) Für alle $f \in L^1(\delta_\omega)$ gilt: $\int f d\delta_\omega = f(\omega)$.

4.7 Der Satz von Radon Nikodym

Im Folgenden sei (Ω, Σ, μ) ein σ -endlicher Maßraum. Ferner sei f eine nicht negativ meßbare Funktion auf Ω . Für $A \in \Sigma$ definiere dann

$$\nu(A) = \int \chi_A f d\mu \tag{4.9}$$

ein Maß ν auf (Ω, Σ) , das auch mit $\nu = f \cdot \mu$ bezeichnet wird. Da offensichtlich $\nu(\emptyset) = 0$ und $\nu(A) \leq 0$ genügt es

$$\nu\left(\sum A_i\right) = \sum \nu(A_i)$$

zu zeigen. Dies aber folgt aus dem Satz von B. Levi, wenn man dort $f_n = f \cdot \chi_{\sum A_i}$ setzt.

Falls $\nu = f \cdot \mu$, so nennt man f die *Dichte von ν bezüglich μ* und schreibt auch $f = \frac{d\nu}{d\mu}$. Diese Bezeichnung wurde gewählt, weil sie einen bekannten Sachverhalt der reellen Analysis verallgemeinert.

Beispiel 4.27 Es sei f eine stetige nichtnegative Funktion auf einem Intervall $[a, b]$. Für $a \leq x \leq b$ ist dann

$$f \cdot \lambda([0, x]) = \int_a^x f(t) d\lambda = \int_a^x f(t) dt$$

und

$$\frac{d}{dx}(f \cdot \lambda([0, x]))(x) = f(x)$$

nach dem Hauptsatz der Differential- und Integralrechnung. Mit dieser Konstruktion ist es nun leicht aus einem gegebenen Maß μ ein Wahrscheinlichkeitsmaß zu konstruieren. Ist nämlich $f \in L^1(\mu)_+$ mit $\int f d\mu = 1$, so ist $P = f \cdot \mu$ ein Wahrscheinlichkeitsmaß auf Ω . Diese Konstruktion ist deswegen wichtig, weil eine sehr große Klasse von Wahrscheinlichkeitsmaßen vom Typ $f \lambda^n$ ist.

Maße mit Dichten haben noch eine besondere Eigenschaft.

Lemma 4.10 Es sei $\nu = f \cdot \mu$ und es sei $A \in \Sigma$ mit $\mu(A) = 0$. Dann gilt $\nu(A) = 0$.

Beweis: f kann durch Treppenfunktionen f_n von unten approximiert werden. Also gilt $f_n \cdot \chi_A \nearrow \chi_A f$. Nun wende den Satz von B. Levi an und verwende $\int f_n \chi_A d\mu = 0$.

Definition 4.17 Ein Maß von ν auf einem Maßraum (Ω, Σ) heißt *absolut stetig* bezüglich μ wenn für jedes $A \in \Sigma$ gilt

$$\mu(A) = 0 \Rightarrow \nu(A) = 0.$$

Von großer theoretischer Bedeutung für die Wahrscheinlichkeitstheorie ist der nachfolgende Satz von Radon Nikodym.

Satz 4.16 Es sei (Ω, Σ, μ) ein σ -endlicher Maßraum und es sei ν absolut stetig bezüglich μ . Dann existiert eine nicht negative Σ meßbare Funktion f , so daß $\nu = f \cdot \mu$.

Da der Beweis dieses Satzes recht aufwendig ist, wollen wir ihn hier nicht beweisen. Einen Beweis kann man etwa in dem Buch von H. Bauer finden.

Aufgabe 4.21 Es seien $f, g \geq 0$ meßbar, und es gelte $f \cdot \mu = g \cdot \mu$. Dann gilt $f = g \mu$ -f.ü.. Die Umkehrung gilt auch.

Aufgabe 4.22 Für die Maße μ, ν auf (Ω, Σ) gelte $\mu(A) \leq \nu(A)$ für $A \in \Sigma$. Was läßt sich über ν sagen?

Der Satz von Radon Nykodym ist in soweit für die Anwendungen wichtig, da fast alle in den Anwendungen vorkommenden Maße diskret oder absolut stetig bezüglich des Lebesgue-Maßes sind. In 99 % aller Fälle haben die Maße die Form

$$\mu = \sum_{i=1}^{\infty} a_i \delta_{\omega_i} \quad \text{bzw.} \quad \mu = f(x) \cdot \lambda^n.$$

Kapitel 5

Zufallsvariable

5.1 Zufallsvariable, Verteilungen und Momente

Es sei (Ω, Σ, μ) ein Maßraum und (Ω', Σ') eine σ -Algebra über Ω' . Ferner sei $X : \Omega \rightarrow \Omega'$ eine $\Sigma - \Sigma'$ messbare Abbildung und $A' \in \Sigma'$. Dann definiert

$$\mu'(A') = \mu(X^{-1}A') = \mu(\{\omega \in \Omega | X(\omega) \in A'\}) \quad (5.1)$$

ein Maß μ' auf Σ' . μ' wird das *Bildmaß* von μ unter X genannt und mit $\mu' = X\mu$ bezeichnet. Man spricht auch von der *Verteilung von X bezüglich μ* , wenn μ ein Wahrscheinlichkeitsmaß ist. Offensichtlich ist $X\mu = \mu'$ σ -endlich (endlich, ein Wahrscheinlichkeitsmaß) genau, wenn μ es ist.

In Anwendungen spielen Bildmaße eine große Rolle. Daher wollen wir hier einiges über die Integration bezüglich Bildmaßen beweisen.

Satz 5.1 (*Transformationssatz*)

- a) Für jede $\Sigma' - \mathcal{B}(\overline{\mathbb{R}})$ messbare, positive numerische Funktion F auf Ω' gilt

$$\int F d(X\mu) = \int F \circ X d\mu. \quad (5.2)$$

- b) Für eine $\Sigma' - \mathcal{B}(\overline{\mathbb{R}})$ messbare, positive numerische Funktion F auf Ω' sind äquivalent: $F \in L^1(\Omega', \Sigma', X\mu)$ und $F \circ X \in L^1(\Omega, \Sigma, \mu)$. In diesem Fall gilt (5.2) für F .

Beweis:

- a) Da die $\Sigma' - \mathcal{B}(\overline{\mathbb{R}})$ messbaren positiven Funktionen isotone Limiten von Treppenfunktionen sind, und das Integral über positive Treppenfunktionen definiert ist, können wir $F = \sum_{i=1}^n a_i \chi_{A'_i}$, mit $A'_i \in \Sigma'$ disjunkt annehmen. Dann aber ist

$$\begin{aligned} \int F d(X\mu) &= \sum_{i=1}^n a_i X\mu(A'_i) = \sum_{i=1}^n a_i \mu(X^{-1}A'_i) \\ &= \int \sum_{i=1}^n a_i \chi_{X^{-1}A'_i} d\mu = \int F \circ X d\mu. \end{aligned}$$

- b) Nun zerlege F in Positiv- und Negativteil $F = F^+ - F^-$ und wende a) auf F^+ und F^- an und verwende die Linearität des Integrals.

Definition 5.1 Es sei X eine numerische $\Sigma - \mathcal{B}(\overline{\mathbb{R}})$ messbare Funktion auf dem Maßraum (Ω, Σ, μ) . X heißt μ -quasiintegrierbar, falls $\int X_+ d\mu$ oder $\int X_- d\mu$ existieren und endlich sind. In diesem Fall bezeichnet $\int X d\mu = \int X_+ d\mu - \int X_- d\mu \in \overline{\mathbb{R}}$ das μ -Integral von X .

Definition 5.2 Es sei X eine numerische Zufallsvariable auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) . Ist dann X P -quasiintegrierbar, so nennt man $E(X) = \int X dP$ den *Erwartungswert* von X . Ist allgemeiner $X^k P$ -quasiintegrierbar, so heißt

$$E(X^k) = \int X^k dP$$

das k -te *Moment* von X .

Ist nun X eine reelle Zufallsvariable auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) und F eine numerische $\mathcal{B} - \mathcal{B}(\overline{\mathbb{R}})$ messbare Funktion auf \mathbb{R} , so ist $F \circ X : \Omega \rightarrow \overline{\mathbb{R}}$ $\Sigma - \mathcal{B}(\overline{\mathbb{R}})$ messbar und wegen Satz (5.1) gilt, falls $F \circ X$ P -quasiintegrierbar ist:

$$E_P(F \circ X) = \int F \circ X dP = \int F d(XP) = E_{XP}(F). \quad (5.3)$$

Insbesondere haben wir also für das k -te Moment

$$E_P(X^k) = \int X^k dXP = E_{XP}(x^k) \quad (5.4)$$

natürlich nur, wenn diese Ausdrücke existieren.

Definition 5.3 Für eine reelle integrierbare Zufallsvariable X heißt

$$\sigma^2(X) = E((X - E(X))^2)$$

die *Varianz* von X . Die Zahl $(E((X - E(X))^2))^{1/2}$ wird auch *Streuung* oder *Standardabweichung* von X genannt.

Satz 5.2 Für eine reelle Zufallsvariable X auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) ist X^2 genau dann integrierbar, wenn X integrierbar ist und $\sigma^2(X)$ existiert. In diesem Fall gilt

$$\sigma^2(X) = E(X^2) - E(X)^2 = \int X^2 dP - \left(\int X dP \right)^2. \quad (5.5)$$

Beweis: Es sei X^2 integrierbar. Da $1 + X^2 \geq |X|$ ist X integrierbar und damit ist auch $(X - E(X))^2 = X^2 - 2XE(X) + E(X)^2$ integrierbar. Insbesondere gilt

$$\begin{aligned} \sigma^2(X) &= \int (X - E(X))^2 dP = \int (X^2 - 2E(X)X + E(X)^2) dP \\ &= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2. \end{aligned}$$

Die Umkehrung folgt analog.

Neben den Momenten $E(X^k)$ einer reellen Zufallsvariable betrachtet man häufig noch die Funktion $t \rightarrow E(e^{tX})$. Dazu sei wieder (Ω, Σ, P) ein Wahrscheinlichkeitsraum und X eine reelle Zufallsvariable. Dann ist für jedes $t \in \mathbb{R}$ die Funktion e^{tX} messbar. Wir nehmen nun an, daß für ein $r > 0$ die Zufallsvariable e^{-rX} P -integrierbar ist. Da dann $|X|^k \leq \text{konst.} \cdot (e^{rX} + e^{-rX})$ und

$$e^{t|X|} \leq \text{konst.} \cdot (e^{rX} + e^{-rX}) \text{ für } |t| \leq r \quad (5.6)$$

existieren alle Momente von X . Da

$$e^{tX} = \sum_{k=0}^{\infty} \frac{t^k}{k!} X^k$$

gilt wegen (??) und des Satzes über die dominierte Konvergenz von Lebesgue für alle $|t| \leq r$

$$E(e^{tX}) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k). \quad (5.7)$$

Dabei konvergiert die rechte Seite absolut. Es sei nun

$$M(t) = E(e^{tX}) \quad |t| \leq r. \quad (5.8)$$

M nennen wir die *momentenerzeugende Funktion* von X . Da die Reihe (??) für r konvergiert, ist $M(t)$ analytisch mit Konvergenzradius $\geq r$. Wir dürfen daher (??) gliedweise differenzieren und erhalten

$$\left(\frac{d^k}{dt^k} M \right) (0) = E(X^k). \quad (5.9)$$

Diese Formel erklärt den Begriff momentenerzeugende Funktion. Wir haben für $|t| \leq r$

$$M(t) = \int e^{tX} dP = \int e^{tx} d(XP)$$

d.h. $M(t)$ ist die Laplace Transformierte von XP .

Im nächsten Abschnitt werden wir explizit einige Beispiele von solchen Funktionen berechnen.

Definition 5.4 Es sei X eine reelle Zufallsvariable auf dem Wahrscheinlichkeitsraum Ω . Dann ist $[X < x] \in \Sigma$ für alle $x \in \mathbb{R}$ und $F_X(x) = P([X < x])$ definiert eine reellwertige Funktion F_X auf \mathbb{R} . F_X wird die *Verteilungsfunktion* von X genannt.

Der Begriff Verteilungsfunktion sollte nicht mit dem Begriff Verteilung verwechselt werden. Für eine Zufallsvariable X hat F_X die Eigenschaften

- i) F_X ist monoton wachsend
 - ii) $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow \infty} F_X(x) = 1$
 - iii) $x_n \rightarrow x- \Rightarrow F_X(x_n) \rightarrow F_X(x)$ (linksseitige Stetigkeit)
- (5.10)

Beweis:

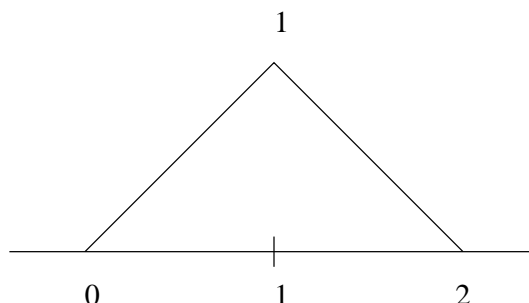
- i) Sei $x \leq y$. Dann gilt $(-\infty, x) \subset (-\infty, y)$ oder $[X < x] \subset [X < y]$. Also $F_X(x) = P([X < x]) \leq P([X < y]) = F_X(y)$.
- ii) $[X < x] \searrow \emptyset$ für $x \rightarrow -\infty$ also $F_X(x) \searrow 0$ $x \rightarrow -\infty$
 $[X < x] \nearrow \Omega$ für $x \rightarrow \infty$ also $F_X(x) \nearrow P(\Omega) = 1$ für $x \rightarrow \infty$.
- iii) Es sei nun $X_n \nearrow x$. Dann gilt $(-\infty, x_n) \nearrow (-\infty, x)$ oder $F_X(x_n) = P([X < x_n]) \nearrow P([X < x]) = F_X(x)$.

Satz 5.3 Eine Funktion F ist genau dann Verteilungsfunktion einer Zufallsvariablen X , falls F (??) erfüllt.

Beweis: Eine Richtung ist bereits gezeigt. Umgekehrt erfülle F (?? i, ii, iii). Auf dem Ring \mathcal{R} aller 1-dimensionalen Figuren, d.h. endlichen Vereinigungen von halboffenen Intervallen, definiere dann einen Inhalt μ durch $\mu([a, b]) = F(b) - F(a)$. Genau wie bei der Einführung des Lebesgueschen Integrals sieht man, daß μ \emptyset -stetig ist. Da $[-n, \cdot) \nearrow \mathbb{R}$ und $\mu([-n, n]) = (F+n) - F(-n) \leq 1$, so ist μ σ -endlich. Also hat μ eine eindeutige Fortsetzung zu μ auf \mathcal{B}_1 . Dies ist sogar ein Wahrscheinlichkeitsmaß, da $\mu(\mathbb{R}) = \lim \mu([-n, n]) = 1$. Man zeigt nun ohne Mühe, daß F die Verteilungsfunktion der Zufallsvariablen $X(x) = x$ auf $(\mathbb{R}, \mathcal{B}_1, \mu)$ ist.

Aufgabe 5.1 Es sei X eine Zufallsvariable. Bestimme die Verteilung von e^X und X^n .

Aufgabe 5.2 Die Dichte einer Zufallsvariable X sei durch das nebenstehende Dreieck bestimmt. Bestimme die Verteilungsfunktion von X^n .



Aufgabe 5.3 Berechne die momenterzeugende Funktion und die Momente für die Dreiecksverteilung.

5.2 Diskrete Verteilungen

Eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum (Ω, Σ, P) heißt *diskret*, falls der Wertebereich von X abzählbar ist. In diesem Fall können wir schreiben

$$X = \sum a_i \chi_{[X=a_i]}. \quad (5.11)$$

Dabei durchläuft a_1, a_2, \dots den gesamten Wertebereich von X . Für das Bildmaß $P' = XP$ gilt dann

$$P' = \sum P([X = a_i]) \delta_{a_i}. \quad (5.12)$$

Dabei bezeichnet δ_{a_i} das Dirac-Maß (Einheitsmasse) im Punkte a_i . Insbesondere bei diskreten Zufallsvariablen und ihren Verteilungen (Bildmaßen) haben wir es mit Maßen zu tun, die nur auf einem Teil von \mathbb{R} konzentriert sind, d.h. ihren Träger nur auf einer Teilmenge haben.

Definition 5.5 Der Träger eines Maßes μ auf \mathcal{B}_1 ist die kleinste abgeschlossene Menge A mit $\mu(A^c) = 0$. Man bezeichnet A mit $\text{supp } \mu$.

Beispiel 5.1 Der Träger von δ_a ist $\{a\}$, $a \in \mathbb{R}$. Ist P' ein Maß wie in (??), so ist

$$\text{supp } P' = \overline{\{a_i | P([X = a_i]) > 0\}}.$$

Beispiel 5.2 Der Träger des Lebesguemaßes ist \mathbb{R} .

Bei reellen Zufallsvariablen X , deren Wertebereich nur einen Teil von \mathbb{R} umfaßt, d.h. deren Bildmaße (Verteilungen) einen Träger haben, der eine echte Teilmenge von \mathbb{R} ist, betrachtet man manchmal auch die Verteilungen eingeschränkt auf den Träger des Bildmaßes. Da es sich hierbei nur um das Vernachlässigen von Mengen vom $P' = XP$ -Maß 0 handelt, werden wir beide Standpunkte einnehmen und die Verteilungen von reellen Zufallsvariablen als Menge auf \mathbb{R} bzw. auf dem Träger von P' betrachten.

5.2.1 Die Binomial Verteilung

Es sei $\Omega_1 = \{0, 1\}$, $\sum_1 = \mathcal{P}(\Omega)$ und $P_1(0) = p$, $P_1(1) = 1 - p = q$. Dann beschreibt (Ω_1, \sum_1, P_1) ein Bernoulliexperiment. Das n -fach unabhängige Wiederholen davon wird durch $(\Omega, \sum, P) = X(\Omega_1, \sum_1, P_1)^n$ beschrieben. Wie wir früher schon gesehen haben, gilt für die Zufallsvariable $X = \text{Anzahl der Nullen}$

$$P([X = k]) = b(n, k, p) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (5.13)$$

Das Bildmaß von X ist damit

$$\sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k = XP = P'. \quad (5.14)$$

Dieses Maß wird auch die *Binomialverteilung* genannt. Betrachten wir P' eingeschränkt auf den Wertebereich $\{0, 1, 2, \dots, n\}$, so ordnet P' dem Punkt k gerade das Maß $b(n, k, p)$ zu. Berechnen wir noch die momenterzeugende Funktion der Binomialverteilung.

Nach (??) gilt

$$\begin{aligned} M(t) &= E(e^{tX}) = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} e^{tk} \\ &= \sum_{k=0}^n \binom{n}{k} (p \cdot e^t) (1 - p)^{n-k} = (1 - p + pe^t)^n. \end{aligned}$$

Es ist also

$$M(t) = (q + pe^t)^n. \quad (5.15)$$

Wir erhalten damit für den Erwartungswert und das 2. Moment

$$\begin{aligned} E(X) &= \frac{d}{dt} M|_0 = e^t n (q + pe^t)^{n-1} P|_0 = np \\ E(X^2) &= \frac{d^2}{dt^2} M|_0 = e^t np(n-1)(q + pe^t)^{n-2} pe^t + e^t n (q + pe^t)^{n-1} P|_0 \\ &= n(n-1)p^2 + np. \end{aligned}$$

Damit erhalten wir für die Varianz

$$E((X - E(X))^2) = n(n-1)p^2 + np - n^2p^2 = -np^2 + np = npq.$$

Aus (??) bzw. (??) erhalten wir

$$\frac{b(n, k, p)}{b(n, k-1, p)} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p - k}{kq}$$

$$\begin{aligned} \text{d.h. } b(n, k, p) &\geq b(n, k-1, p), & \text{ falls } (n+1)p \geq k \\ \text{und } b(n, k, p) &\leq b(n, k-1, p), & \text{ falls } (n+1)p \leq k. \end{aligned}$$

Es sei daher $m \in \mathbb{N}$ mit $(n+1)p - 1 < m \leq (n+1)p$. Die $b(n, k, p)$ wachsen als Funktionen von k solange $k \leq m$ und nehmen in m ihren größten Wert an. Für $k > m$ fallen sie. Falls $(n+1)p = m$, so gilt $b(n, m, p) = b(n, m-1, p)$.

Häufig kommt es in Anwendungen vor, daß n groß ist und p sehr klein, so daß $a = np$ immer noch nicht zu groß ist. In diesem Fall haben wir

$$b(n, 0, p) = \left(1 - \frac{p \cdot n}{n}\right)^n = \left(1 - \frac{a}{n}\right)^n \rightarrow e^{-a}.$$

Ferner ist dann

$$\frac{b(n, k, p)}{b(n, k-1, p)} = \frac{np - (k-1)p}{kq} \approx \frac{a}{k}$$

oder

$$\begin{aligned} b(n, 1, p) &\approx ae^{-a} \\ b(n, 2, p) &\approx \frac{a^2}{2} e^{-a}. \end{aligned}$$

Allgemeiner erhält man durch Induktion

$$b(n, k, p) \approx \frac{a^k}{k!} e^{-a} \quad (5.16)$$

die sog. *Poisson Approximation* der Binomialverteilung. Sie gilt, solange k nicht zu groß ist, und solange a nicht zu groß ist.

Binomialverteilung $n = 10$
 Poisson-Approximation für $p = 0.1$ und $p = 0.2$
 Normal-Approximation für $p = 0.5$ und $p = 0.6$

Anstelle von (??) gilt sogar genauer die Abschätzung

$$\frac{A^k}{k!} e^{-a} e^{\frac{ka}{n}} > b(n, k, p) > \frac{a^k}{k!} e^{-a} \frac{e^{-k^2}}{(n-k)^{-a^2}} (n-a). \quad (5.17)$$

Wir wollen (??) allerdings nicht beweisen.

Beispiel 5.3 Wie groß ist die Wahrscheinlichkeit, daß von 2.500 Osnabrücker Studenten genau k am 24. Dezember Geburtstag haben?

Hier ist $n = 2.500$, $p = \frac{1}{365}$ und gefragt ist nach $b(n, k, p)$. Mit $a = \frac{2.500}{365} = \frac{500}{73} \approx 7$ sieht man leicht, daß man für $k \leq 12$ ohne weiteres die Poisson Approximation anwenden kann. Also ist $b(n, k, p) \approx \frac{a^k}{k!} e^{-a}$.

k	0	1	2	3	4	5	6	7	8	9
$\frac{a^k}{k!} e^{-a}$	0,00106	0,0073	0,0249	0,0568	0,0972	0,133	0,152	0,149	0,127	0,097

k	10	11	12	13	14	15	16	17
$\frac{a^k}{k!}e^{-a}$	0,066	0,0413	0,0236	0,0124	0,00608	0,00278	0,0012	0,00048

Der Fehlerfaktor für $k = 10$ ist hier $e^{\frac{k-a}{n}} = 1,0278$. Die Wahrscheinlichkeit, daß mehr als 13 (\geq) am 24. Dezember Geburtstag haben ist dann

$$\begin{aligned} \sum_{k=13}^{\infty} b(n, k, p) &\approx \sum_{k=13}^{\infty} \frac{a^k}{k!} e^{-a} \approx \frac{a^{13}}{13!} e^{-a} \left[1 + \frac{a}{14} + \frac{a^2}{14 \cdot 15} + \frac{a^3}{14 \cdot 15 \cdot 16} + \dots \right] \\ &= 1,87 \cdot \frac{a^{13}}{13!} e^{-a} = 0,0232. \end{aligned}$$

Aufgabe 5.4 Zeige, daß die Verteilungsfunktion einer diskreten Zufallsvariable stückweise konstant ist. Wo liegen die Sprungstellen und wie hoch sind die Sprünge.

5.2.2 Die hypergeometrische Verteilung

Eng verwandt mit der Binomialverteilung ist die hypergeometrische Verteilung. Diese beschreibt im wesentlichen die Stichprobenentnahme ohne Zurücklegen.

Dazu betrachte folgende Situation: Es seien N Objekte vorgegeben. Davon seien M defekt. Es wird eine Stichprobe von n ($n \leq N$) Objekten entnommen. Dann ist die Wahrscheinlichkeit, daß davon k defekt sind, gegeben durch

$$h(k, n, N, M) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = h_k \quad 0 \leq k \leq n. \quad (5.18)$$

Der Mittelwert der hypergeometrischen Verteilung ist

$$E(X) = \sum_{k=0}^n \binom{M}{k} \binom{N-M}{n-k} k = n \frac{M}{N} \quad (5.19)$$

und für das 2. Moment gilt

$$E(X^2) = \frac{n(n-1)(M(M-1))}{N(N-1)} + \frac{M_n}{N}, \quad \text{so daß} \quad (5.20)$$

$$\sigma^2(X) = n \frac{M}{N} \frac{(N-M)}{N} \frac{(N-n)}{(N-1)}. \quad (5.21)$$

Der Beweis von (??) – (??) sei als Übungsaufgabe gestellt.

Aufgabe 5.5 Beweise (??), (??) und (??).

5.2.3 Die Poisson-Verteilung

Wir betrachten hier das zeitliche Geschehen eines Prozesses (etwa radioaktiver Zerfall, Unfälle, Telefonanrufe). Über das Geschehen der Ereignisse machen wir die folgenden Annahmen:

- a) Das Geschehen eines solchen Ereignisses in einem Zeitintervall hängt nicht ab vom Auftreten bzw. nicht Auftreten in der Vergangenheit.
- b) Die Wahrscheinlichkeit, daß in einem Zeitintervall $(t, t + \Delta t)$ das Ereignis einmal auftritt, ist von der Form $a \Delta t + \Delta t \cdot o(\Delta t)$.
- c) Die Wahrscheinlichkeit, daß in $(t, t + \Delta t)$ mehr als ein Ereignis eintritt, ist Δt .

Es bezeichne nun $f(x, t)$ die Wahrscheinlichkeit, daß x Ereignisse in $(0, t)$ aufgetreten sind. Dann haben wir auf Grund unserer Annahmen

$$f(x, t + \Delta t) = f(x, t)(1 - a \Delta t - o(\Delta t)) + f(x - 1, t)(a \Delta t + \dots)$$

denn entweder fanden die x Ereignisse schon in $(0, t)$ statt, oder ein Ereignis fand genau in $(t, t + \Delta t)$ statt. Das Auftreten von mehreren Ereignissen in $(t, t + \Delta t)$ haben wir bereits ignoriert. Wir erhalten dann mit $\Delta t \rightarrow 0$.

$$\frac{df(x, t)}{dt} = -af(x, t) + af(x - 1, t). \quad (5.22)$$

Da $f(-1, t) = 0$, gibt dies $\frac{df(0, t)}{dt} = -af(0, t)$ oder

$$f(0, t) = e^{-at} \quad (5.23)$$

da $f(0, 0) = 1$ eine sinnvolle Annahme ist. Durch Induktion nach x zeigt man nun

$$f(x, t) = \frac{e^{-at}}{x!} (at)^x. \quad (5.24)$$

Diese Verteilung hat große Ähnlichkeit mit (??).

Definition 5.6 Die *Verteilung* auf $\mathbb{N} = \{0, 1, \dots\}$ mit den Werten

$$p(k) = \frac{a^k}{k!} e^{-a} \quad (5.25)$$

wird *Poisson Verteilung* zum Parameter $a (a > 0)$ genannt. Oben haben wir also die Poisson Verteilung zum Parameter at kennengelernt.

Bestimmen wir zunächst die Moment erzeugende Funktion der Poisson-Verteilung:

$$\begin{aligned} M(t) &= E(e^{tx}) = \sum_{k=0}^{\infty} e^{tk} e^{-a} \frac{a^k}{k!} = \sum_{k=0}^{\infty} e^{-a} \frac{(e^t a)^k}{k!}. \\ &\Rightarrow M(t) = e^{-a} e^{e^t a}. \end{aligned} \quad (5.26)$$

In 5.2.1 haben wir schon die Poisson-Verteilung als Approximation der Binomialverteilung kennengelernt, wenn n groß, p klein und $n \cdot p = a$ nicht zu groß ist ($n \geq 50, p \leq 0.05$). Es ist also p , die Wahrscheinlichkeit für das Auftreten eines Einzelereignisses, klein. Das Einzelereignis ist also selten. Aus diesem Grunde benutzt man die Poissonverteilung auch zur Beschreibung seltener Ereignisse.

Um noch einen besseren Eindruck von der Poisson-Approximation der Binomialverteilung zu erhalten, hier noch ein paar Beispiele dazu:

$p = 0.1$		$k = 0$	1	2	3	4	5	≥ 6
$n = 10$	Binomialvert.	0.349	0.387	0.194	0.057	0.011	0.002	0
	Poissonappr.	0.368	0.368	0.184	0.061	0.015	0.003	0.001

		0	1	2	3	4	5	6	≥ 7
$p = 0.05$	B.V.	0.3585	0.3774	0.1887	0.0596	0.0133	0.0022	0.0003	0.000
$n = 20$	P.V.	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001
$p = 0.01$	B.V.	0.3660	0.3697	0.1849	0.0610	0.0149	0.0029	0.0005	0.0001
$n = 100$	P.V.	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001

Ferner auch noch einige Illustrationen zur Binomialverteilung für $n = 20$. Man beachte die Lage des Maximums (s. Abbildung S. 73).

Beispiel 5.4 Die Blutgruppe AB kommt normalerweise mit einer Häufigkeit von 6% vor. In einer Gruppe von $n = 80$ Personen findet man also k AB -ler mit einer Häufigkeit von

$$p_k = e^{-80 \cdot 0.06} \cdot (80 \cdot 0.06)^k \frac{1}{k!}.$$

Die Poisson-Verteilung kann also mit Erfolg bei seltenen Ereignissen wie Aberrationen, Mutationen, Unglücke etc. angewandt werden. Auch der radioaktive Zerfall gehört dazu, denn hier liegt n meist in der Größenordnung $6 \cdot 10^{23}$ (Avogadrosche Zahl). Hier noch ein weiteres Beispiel.

Beispiel 5.5 Im 2. Weltkrieg wurde London wiederholt mit der V1 bombardiert. Damals hat man London in 580 etwa gleichgroße Flächen aufgeteilt, und die Anzahl der Treffer notiert. Es stellte sich heraus, daß zur Beschreibung der Anzahl der Treffer eine Poisson-Verteilung mit $\lambda = 0.943$ verwendet werden kann, d.h. die theoretische Zahl k ist $580 \cdot e^{-\lambda} \frac{\lambda^k}{k!}$.

Anzahl der Treffer	0	1	2	3	4	≥ 6
Beobachtete Zahlen	229	211	93	39	7	1
Theor. Zahl	225.62	212.86	100.34	31.32	7.54	2.32

5.2.4 Die negative Binomial-Verteilung

Wir betrachten nun die beliebige Wiederholung eines Bernoulli Experimentes, bis zum ersten Mal „1“ auftritt. Der Ereignisraum ist dann $\Omega = \{1, 01, 001, \dots, 0^n 1, \dots\}$ und kann mit $\{0, 1, 2, \dots\}$ identifiziert werden. Offensichtlich gilt $P(\{0^n 1\}) = P(\{n\}) = q^n p$. Man nennt dies die geometrische Verteilung.

Allgemeiner als im vorangegangenen Fall wird jetzt nach der Anzahl der Wiederholungen eines Bernoulli Experimentes gefragt, bis r -mal „1“ aufgetreten ist. Unser Ereignisraum Ω besteht jetzt also aus allen $r + k$ Tupeln, $k \geq 0$, von 0 und 1, wo die letzte Zahl eine 1 ist.

Damit der r -te Erfolg genau beim $(r + k - 1)$ -ten Zug auftritt, muß unter den vorangegangenen $r + k - 1$ Zügen k -mal die „0“ aufgetreten sein. Die Wahrscheinlichkeit dafür ist dann nach 5.2.1

$$f(k, p, r) = \binom{r + k - 1}{k} p^r q^k. \quad (5.27)$$

Wir schreiben dafür

$$\begin{aligned} f(k, p, r) &= \frac{(r + k - 1)(r + k - 2) \dots r}{k!} p^r q^k \\ &= \frac{-r(-r - 1) \dots (-r - k + 1)}{k!} p^r (-q)^k \\ &= \binom{-r}{k} p^r (-q)^k \quad k = 0, 1, \dots \end{aligned}$$

Daher der Ausdruck *negative Binomialverteilung*.

Dies gibt leicht für den Erwartungswert und das 2. Moment

$$\begin{aligned} E(X) &= \frac{d}{dt} M|_0 = a \\ E(X^2) &= \frac{d^2}{dt^2} M|_0 = a^2 + a \quad \text{und} \\ \sigma^2(X) &= a. \end{aligned} \quad (5.28)$$

Diese Identität zeigt man durch Reihenentwicklung nach Taylor. Für die momenterzeugende Funktion der negativen Binomialverteilung gilt:

$$M(t) = \sum_{k=0}^{\infty} e^{kt} \binom{-r}{k} p^r (-q)^k = \sum_{k=0}^{\infty} \binom{-r}{k} p^r (-e^t q)^k = p^r (1 - e^t q)^{-r}.$$

Dies gibt:

$$M'(t) = r p^r q e^t (1 - q e^t)^{-r-1}$$

oder

$$M'(0) = E(X) = r p^r q (1 - q)^{-r-1} = \frac{r q}{p}.$$

Entsprechend

$$M''(0) = \frac{r q}{p} + \frac{r(r + 1) q^2}{p^2}.$$

Die negative Binomial-Verteilung wurde häufig bei der Beschreibung von Zähldaten verwandt, z.B. bei Unfallstatistiken, Marktforschung (Häufigkeit mit der ein Produkt gekauft wird) oder in der Psychologie. Genauere Hinweise findet man in dem Buch von Johnson, Kotz.

Bemerkung 5.1 In den vorangegangenen Beispielen haben wir uns um eine vernünftige Definition des zugrunde liegenden Wahrscheinlichkeitsraumes und der Zufallsvariable herumgedrückt und uns hauptsächlich um die Kombinatorik gekümmert. Dieses Versäumnis wollen wir jetzt nachholen.

Da es sich bei unserer Fragestellung auch um beliebig häufige Wiederholungen von Bernoulli Experimenten handelt, legen wir von vornherein die σ -Algebra $(\Omega_1^\infty, \Sigma_1^\infty) = \bigotimes_{k=1}^\infty (\Omega_1, \mathcal{P}(\Omega_1))$ zugrunde. Auf dieser σ -Algebra müssen wir nun ein Wahrscheinlichkeitsmaß definieren. Dazu verwenden wir wieder unsere alte Prozedur, d.h. wir definieren das Maß zunächst auf einem geeigneten Ring und setzen es dann fort. Es sei also \mathcal{R} der Ring aller Mengen der Form $A \times \bigotimes_{k=n+1}^\infty \Omega_1$ mit $A \in \Sigma_1^n = \mathcal{P}(\Omega_1^n)$. Solche Mengen, die nur in ihren ersten n Koordinaten (n ist variabel, aber endlich!) festgelegt sind, nennen wir *Zylindermengen*.

Aufgabe 5.6 $\mathcal{R} =$ Menge aller Zylindermengen ist eine Algebra. Auf \mathcal{R} definiere nun einen Inhalt durch

$$P\left(A \times \bigotimes_{k=n+1}^\infty \Omega_1\right) = P_1^n(A). \quad (5.29)$$

Aufgabe 5.7

- a) P ist ein wohldefinierter Inhalt
- b) P ist ein Maß

(Hinweis: Für (b) gehe analog zum Lebesgue Borelmaß vor und verwende, daß in $\bigotimes_{n=1}^\infty \Omega_1$ mit der Produkttopologie die Zylindermengen gerade offene und abgeschlossene Mengen sind.)

Aufgrund unseres Fortsetzungs- und Eindeutigkeitsatzes gibt es also eine eindeutige Fortsetzung P_1^∞ von P zu einem Wahrscheinlichkeitsmaß auf $(\Omega_1^\infty, \Sigma_1^\infty)$. Auf dem Wahrscheinlichkeitsraum $(\Omega_1^\infty, \Sigma_1^\infty, P_1^\infty)$ betrachte nun die Zufallsvariable X , wo

$$X(\omega) = \begin{cases} k & \text{falls in } (\omega_1, \dots, \omega_{r+k}) \text{ genau } r \text{ mal } 0 \text{ vorkommt} \\ & \text{und } \omega_{r+k} = 0 \text{ ist.} \\ \infty & \text{falls jeder endliche Abschnitt von } \omega \\ & \text{weniger } (<) \text{ als } r \text{ „Nullen“ enthält.} \end{cases}$$

X ist messbar, denn $[X = k]$ ist sogar eine Zylindermenge und $[X = \infty] = \bigcap_{k=1}^\infty [X \geq k]$. Ferner haben wir

$$P([X = k]) = P\left([\tilde{X} = k] \times \bigotimes_{n+1}^\infty \Omega_1\right) = P_1^{r+n}([\tilde{X} = k]) = f(k, p, r).$$

Hierbei ist \tilde{X} die Einschränkung von X auf die ersten $r + k$ Koordinaten.

5.2.5 Multinomial Verteilung

Bei der Multinomial Verteilung interessiert man sich nicht nur für die Verteilung einer einzelnen Zufallsvariable, sondern von mehreren Zufallsvariablen gleichzeitig.

Ausgangspunkt der Betrachtung ist dabei der Wahrscheinlichkeitsraum

$$\Omega_1 = \{1, 2, \dots, \ell\}, \quad \sum_1 = \mathcal{P}(\Omega_1), \quad P_1(\{i\}) = p_i \geq 0$$

wo $\sum_{i=1}^{\ell} p_i = 1$. Betrachte nun das n -fache Wiederholen dieses Experimentes. Dies wird durch den Wahrscheinlichkeitsraum $(\Omega_1^n, \mathcal{P}(\Omega_1^n), P_1^n)$ beschrieben. Es sei nun X_i die Zufallsvariable Anzahl von „i“. Wir haben dann

$$P([X_1 = k_1, \dots, X_\ell = k_\ell]) = \begin{cases} 0 & \text{falls } k_1 + \dots + k_\ell \neq n \\ \frac{n!}{k_1! k_2! \dots k_\ell!} p_1^{k_1} \dots p_\ell^{k_\ell} & \text{falls } k_1 + \dots + k_\ell = n. \end{cases} \quad (5.30)$$

Dies ist die Multinomial Verteilung. Anders als die vorangegangenen Verteilungen ist diese Verteilung mehrdimensional, d.h. eine Verteilung auf \mathbb{R}^ℓ .

Beispiel 5.6 Wie groß ist die Wahrscheinlichkeit, daß beim $6n$ -maligen Würfeln jede Zahl genau n mal auftritt? In diesem Fall ist $\ell = 6$ und $p_1 = \dots = p_6 = \frac{1}{6}$ und $k_1 = \dots = k_6 = n$. Also suchen wir

$$\frac{(6n)!}{(n!)^6} \left(\frac{1}{6}\right)^{6n}.$$

Für $n = 1, 2$ erhalten wir 0,0154 bzw. 0,0034.

Natürlich gibt es noch weitere diskrete Verteilungen. Mehr dazu findet man in dem Buch „Discrete Distributions von Johnson, Kotz (TKC Joh)“.

Wir wollen nun noch einige Anwendungen betrachten:

Beispiel 5.7

1. Für die Montage gewisser Maschinen werden 200 Schrauben benötigt. Man weiß, daß von diesen Schrauben im Mittel 1% defekt sind. Wieviel Schrauben müssen zur Montage mitgesandt werden, damit weniger als 2% Reklamationen wegen nicht passender Schrauben zu erwarten sind?

Die Sendung wird $200+k$ Schrauben enthalten. Davon sind mit der Wahrscheinlichkeit $b_r = \binom{200+k}{r} p^r (1-p)^{200+k-r}$ genau r defekt, wobei $P = 0,01$ ist. Da zu erwarten ist, daß k nicht zu groß ist, können wir die Poisson Approximation anwenden. Dann wird

$$b_r \approx \frac{\left(2 + \frac{k}{100}\right)^r}{r!} e^{-(2 + \frac{k}{100})}.$$

Da zur Montage genau 200 Schrauben benötigt werden, und weniger als 2% Reklamationen kommen sollen, haben wir

$$0,02 \geq \sum_{r \geq k+1} b_r \approx \sum_{r \geq k+1} \frac{\left(2 + \frac{k}{100}\right)^r}{r!} e^{-(2 + \frac{k}{100})}. \quad (5.31)$$

Für eine erste Approximation lösen wir $\frac{2^{k+1}}{(k+1)!}e^{-2} \leq 0,02$ und erhalten $k = 5$. Berechnen wir nun die Summe in (??) mit $k = 5$, so sehen wir, daß $k = 5$ in der Tat die Lösung ist. Will man sogar weniger als 1% Reklamationen haben, muß man 206 Schrauben zupacken.

2. Packungen mit 50 Stück eines Produktes sind in Ordnung, wenn weniger als 5 Stück davon defekt sind. Die Herstellerfirma testet die Packungen auf folgende Weise:

Es werden nacheinander Teile aus der Packung genommen und getestet. Wenn die ersten 7(8) Stück in Ordnung sind, wird die Packung versandt, andernfalls zurückgestellt. Mit wieviel Reklamationen muß die Firma rechnen? Wieviel Prozent der zurückgestellten Packungen sind in Ordnung?

In einer Packung seien M defekte Stücke. Die Wahrscheinlichkeit, 7 intakte Teile auszuwählen, ist dann

$$\frac{\binom{50-M}{7}}{\binom{50}{7}} = \frac{(50-M) \dots (44-M)}{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = W_M.$$

Offensichtlich ist $W_M > W_{M+1}$. Die gesuchte Wahrscheinlichkeit ist daher offensichtlich W_6 , d.h. die Wahrscheinlichkeit, bei der 6 Stück defekt sind, die ersten 7 ausgewählten Stück aber intakt. Es ist

$$W_6 = \frac{44 \cdot 43 \cdot 42 \cdot 41 \cdot 40 \cdot 39 \cdot 38}{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44} = \frac{43 \cdot 41 \cdot 13 \cdot 19}{47 \cdot 46 \cdot 15 \cdot 35} = 0,384.$$

Wählt man 8 Stück aus, erhält man entsprechend

$$W'_6 = \frac{44 \cdot 43 \cdot 42 \cdot 41 \cdot 40 \cdot 39 \cdot 38 \cdot 37}{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 \cdot 43} = \frac{6 \cdot 41 \cdot 4 \cdot 39 \cdot 38 \cdot 37}{5 \cdot 7 \cdot 48 \cdot 47 \cdot 46 \cdot 45} = 0,330.$$

Eine Packung, die an sich in Ordnung ist, wird zurückgestellt, wenn unter den ersten 7 (8) Stück eines defekt ist. Die Wahrscheinlichkeit dafür ist

$$\frac{\binom{45}{6} \binom{5}{1}}{\frac{50}{7}} = \frac{7 \cdot 45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40 \cdot 5}{50 \cdot 49 \cdot \dots \cdot 44 \cdot 43} = \frac{43 \cdot 41}{47 \cdot 46} = 0,401$$

bzw.

$$\frac{\binom{45}{7} \binom{5}{1}}{\binom{50}{8}} = \frac{4 \cdot 41 \cdot 36}{7 \cdot 47 \cdot 46} = 0,422.$$

3. Ein Meteoriten Schauer trifft die Erde mit einer Treffwahrscheinlichkeit von 1^{-5} Treffern pro m^2 . Wie groß ist die Wahrscheinlichkeit, daß eine Fläche von $1(\text{km})^2$ von k Meteoriten getroffen wird?

Für dieses Problem ist die Poissonverteilung zuständig. Da $1(\text{km})^2 = 10^6 \text{ m}^2$ sind, ist die zugehörige Konstante $a = 10$

k	0	1	2	3	4	5	6	7	8	9
$\frac{10^k}{k!e^{10}}$	0	0	0,002	0,008	0,019	0,038	0,063	0,090	0,1125	0,125

k	10	11	12	13
$\frac{10^k}{k!e^{10}}$	0,125	0,114	0,094	0,072

4. *Tischtennis*: 2 Personen A und B spielen Tischtennis. In der Regel gewinnt A den Anteil p der Spiele und B den Anteil $1 - p$ der Spiele. Beim Tischtennis gewinnt der Spieler das Spiel, der vorher $2l$ (allgemeiner $2n + 1$) Punkte gesammelt hat. Insgesamt gibt es also pro Spiel höchstens $4n + 1$ Punkte und mindestens $2n + 1$ Punkte. Es sei a_r die Wahrscheinlichkeit, daß A bei einem Punktgesamtstand von $4n + 1 - r$ Punkten gewinnt. Dies geschieht, wenn A vorher $2n$ Punkte gesammelt hat. Also

$$a_r = \binom{4n - r}{2n} p^{2n+1} q^{2n-r}.$$

Die entsprechende Wahrscheinlichkeit b_r für B erhält man, indem man p und q vertauscht. Die Wahrscheinlichkeit, daß A gewinnt, ist dann $a_0 + a_1 + \dots + a_{2n}$ und die Wahrscheinlichkeit, daß B gewinnt, ist $b_0 + b_1 + \dots + b_{2n}$. Die Wahrscheinlichkeit, daß das Spiel mit einem Gesamtpunktstand von $4n - r + 1$ endet, ist $a_r + b_r$.

5. In einem großen Kasten sind rote, grüne und weiße Gummibärchen im Verhältnis 5:4:3 enthalten. Klein Erna greift sich 10 Gummibärchen willkürlich heraus. Mit welcher Wahrscheinlichkeit hat sie dann 5 rote, 3 weiße und 2 grüne?

Es handelt sich hier um die Multinomial-Verteilung mit $\ell = 3$ und der individuellen Wahrscheinlichkeit $\frac{5}{12}, \frac{1}{3}, \frac{1}{4}$. Die gesuchte Wahrscheinlichkeit ist also

$$\frac{10!}{5!2!3!} \left(\frac{5}{12}\right)^5 \left(\frac{1}{3}\right)^3 \left(\frac{1}{4}\right)^2 = 0,073.$$

5.3 Stetige Verteilungen

Es sei (Ω, Σ, μ) ein Maßraum und f eine nicht-negative μ -integrierbare Funktion mit

$$\int f d\mu = 1. \quad (5.32)$$

Dann definiert

$$\Sigma \ni A \rightarrow \int \chi_A f d\mu = P(A) \quad (5.33)$$

ein Wahrscheinlichkeitsmaß auf (Ω, Σ) .

Man sagt, dieses Wahrscheinlichkeitsmaß P habe die *Dichte* f bezüglich μ und schreibt auch

$$P = f \cdot \mu \quad \text{oder} \quad \frac{dP}{d\mu} = f. \quad (5.34)$$

Alle Maße, die wir jetzt betrachten wollen, sind von dieser Form. Allerdings werden wir uns im wesentlichen auf Maße auf \mathbb{R} beschränken.

Man sagt, die Verteilung (Bildmaß) der Zufallsvariablen X auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) habe die *Dichte* f_X , wenn die Verteilung von X von der Form

$f_X \cdot \lambda$ (??) ist. Hat die Verteilung von X die Dichte f_X , so gilt für die Verteilungsfunktion F_X von X

$$F_X(x) = P([X < x]) = P(X^{-1}(-\infty, x)) = \int_{-\infty}^x f_X(t) d\lambda(t), \quad (5.35)$$

d.h. F_X ist die Stammfunktion von f_X . Wir haben uns bisher hauptsächlich für die Verteilungen XP der Zufallsvariable X gekümmert, und der ursprüngliche Maßraum (Ω, Σ, μ) trat in den Hintergrund. Dies ist in der Wahrscheinlichkeitsrechnung durchaus möglich, da die gesamte Information über eine Zufallsvariable X in ihrer Verteilung liegt, während der zugrunde gelegte Wahrscheinlichkeitsraum nicht immer eindeutig ist. Betrachte etwa das Beispiel geometrischer Verteilungen.

5.3.1 Die gleichmäßige Verteilung

Wählt man oben etwa $f = \frac{1}{\mu(A)} \chi_A$, wo $A \in \Sigma$ mit $0 < \mu(A) < \infty$, so erhält man das Wahrscheinlichkeitsmaß

$$P_A(B) = \frac{1}{\mu(A)} \mu(A \cap B) \quad B \in \Sigma. \quad (5.36)$$

Dieses Maß kommt relativ häufig in Anwendungen vor. Früher hatten wir den Ausdruck (??) schon bei der bedingten Erwartung kennengelernt. Ein Spezialfall von (??) ist die sog. *gleichmäßige Verteilung* auf \mathbb{R} . Diese wird auf dem Maßraum $\mathbb{R}, \mathcal{B}_1, \lambda$ durch

$$f(x) = \frac{1}{b-a} \chi_{[b,a]}, \quad b > a \quad (5.37)$$

gemäß (??) beschrieben. Das entsprechende Integral hat die Form

$$\int g d(f \cdot \lambda) = \frac{1}{b-a} \int_a^b g d\lambda. \quad (5.38)$$

Definition 5.7 Die Zufallsvariable X auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) heißt *gleichmäßig verteilt*, wenn die Verteilung (Bildmaß) XP von X von der Form $(\frac{1}{b-a} \chi_{[a,b]}) \cdot \lambda$ ist.

Für die momenterzeugende Funktion der gleichmäßigen Verteilung haben wir

$$M(t) = \frac{1}{b-a} \int_a^b e^{tx} d\lambda(x) = \frac{1}{b-a} (e^{tb} - e^{ta}) \frac{1}{t} \quad (5.39)$$

und für das k -te Moment gilt

$$E(X^k) = \frac{1}{b-a} \int_a^b X^k d\lambda(x) = \frac{1}{b-a} \frac{1}{k+1} (b^{k+1} - a^{k+1}). \quad (5.40)$$

5.3.2 Die Normalverteilung

Die *Normalverteilung* auf \mathbb{R} wird durch die Dichte

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = f_{\mu,\sigma}(x) \quad x \in \mathbb{R} \quad (5.41)$$

bezüglich des Lebesgue Maßes λ beschrieben. Die Normalverteilung spielt, wie wir noch sehen werden, eine außerordentlich große Rolle in der Wahrscheinlichkeitstheorie und Statistik. $f_{\mu,\sigma}$ ist tatsächlich die Dichte eines Wahrscheinlichkeitsmaßes, da $f > 0$ und $\int_{-\infty}^{\infty} f d\lambda = 1$. Für $\mu = 0, \sigma = 1$ erhält man die *Standard-Normalverteilung*.

Definition 5.8 Eine Zufallsvariable X ist normalverteilt, wenn f die Dichte der Verteilung von X ist. Für die momenterzeugende Funktion erhalten wir

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{xt} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\lambda = e^{\mu t} \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{(x-\mu)t} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d\lambda \\ &= e^{\mu t} \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{xt} e^{-\frac{x^2}{2\sigma^2}} d\lambda = e^{\mu t} \frac{1}{\sqrt{2\pi}} \int e^{\frac{zt\sigma - z^2}{2}} d\lambda(z) \\ &= \int e^{\mu t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2zt\sigma + t^2\sigma^2) + t^2\sigma^2 \frac{1}{2}} d\lambda(z) \end{aligned}$$

$$M(t) = e^{\mu t} e^{t^2 \frac{\sigma^2}{2}}. \quad (5.42)$$

Dies gibt dann sofort

$$E(X) = \mu, \quad \sigma^2(X) = \sigma^2. \quad (5.43)$$

Die Bedeutung der Normalverteilung ist eine Folge des zentralen Grenzwertsatzes und allgemeiner ihres häufigen Auftretens in den Anwendungen. Wir wollen uns nun einen Spezialfall des zentralen Grenzwertsatzes ansehen. Entscheidend dabei benötigen wir die Stirlingsche Formel

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left[1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51.840n^3} + \dots\right]. \quad (5.44)$$

Für die Multinomialverteilung hatten wir

$$P_n(k_1, \dots, k_l) = \frac{n!}{k_1! \dots k_l!} p_1^{k_1} \dots p_l^{k_l}, \quad k_1 + \dots + k_l = n. \quad (5.45)$$

Statt der Zufallsvariable $X_i = \text{Anzahl } i, i = 1, \dots, l$ sollte man zweckmäßigerweise die zentrierte und in der Varianz normierte Zufallsvariable

$$\frac{X_i - np_i}{\sqrt{np_i q_i}} \quad \text{mit} \quad q_i = 1 - p_i$$

betrachten. Wir führen daher die Variablen

$$X_i = \frac{k_i - np_i}{\sqrt{np_i q_i}} \quad i = 1, \dots, l \quad (5.46)$$

ein. Die Approximationsformel, die wir herleiten werden, soll für kleine x_i gelten, d.h. für k_i , die nahe beim Erwartungswert np_i von X_i liegen. Wir verlangen daher

$$a_i \leq x_i \leq b_i. \quad (5.47)$$

Dies zeigt

$$k_i \geq np_i + a\sqrt{np_i q_i}. \quad (5.48)$$

Setzen wir in (??) die Formel (??) ein, erhalten wir

$$\begin{aligned} P_n &= \frac{\sqrt{2\pi n} n^n e^{-n} e^{\theta_n}}{\prod_{i=1}^l \sqrt{2\pi k_i} k_i^{k_i} e^{-k_i} e^{\theta_{k_i}}} \prod_{i=1}^l P_i^{k_i} \quad \text{mit} \quad |\theta_n| \leq \frac{1}{12n} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{l-1} \sqrt{\frac{n}{k_1 \dots k_l}} \prod_{i=1}^l \left(\frac{np_i}{k_i}\right)^{k_i} e^{\theta}, \quad \theta = \theta_n - \theta_{k_1} - \dots - \theta_{k_l}. \end{aligned}$$

Wegen (??) wird θ für große n beliebig klein. Betrachte nun

$$\begin{aligned} \ln \prod_{i=1}^l \left(\frac{np_i}{k_i}\right)^{k_i} - \sum_{i=1}^l k_i \ln \frac{k_i}{np_i} &= - \sum (np_i + x_i \sqrt{np_i q_i}) \ln \left(1 + x_i \sqrt{\frac{q_i}{np_i}}\right) \\ &\approx - \sum (np_i + x_i \sqrt{np_i q_i}) \left[x_i \sqrt{\frac{q_i}{np_i}} - \frac{1}{2} x_i^2 \frac{q_i}{np_i} + \frac{1}{3} x_i^3 \left(\sqrt{\frac{q_i}{np_i}}\right)^3 - \dots \right] \\ &= - \sum_{i=1}^l x_i \sqrt{np_i q_i} + \frac{1}{2} x_i^2 q_i + x_i^3 \left(\frac{1}{3} q_i \sqrt{\frac{q_i}{np_i}} - \frac{1}{2} q_i \sqrt{\frac{q_i}{np_i}}\right) + x_i^4 o\left(\frac{1}{n}\right). \end{aligned}$$

Es ist aber $\sum x_i \sqrt{np_i q_i} = \sum k_i - np_i = n - n = 0$ wegen (??). Also erhalten wir

$$\prod_{i=1}^l \left(\frac{np_i}{k_i}\right)^{k_i} \approx \exp\left(-\frac{1}{2} \sum x_i^2 q_i + \sum \frac{1}{6} x_i^3 q_i \sqrt{\frac{q_i}{np_i}} + x_i^4 o\left(\frac{1}{n}\right)\right). \quad (5.49)$$

Ferner haben wir

$$\sqrt{\frac{n}{k_1 \dots k_l}} = \sqrt{\frac{n}{n^l p_1 \dots p_l}} \sqrt{\prod \frac{np_i}{k_i}} = \sqrt{\frac{n}{n^l p_1 \dots p_l}} \frac{1}{\sqrt{\prod (1 + x_i \sqrt{\frac{q_i}{np_i}})}}.$$

Offensichtlich strebt der 2. Term gegen 1 für große n . Wir haben damit

Satz 5.4 Unter den oben gemachten Annahmen gilt

$$P_n(k_1, \dots, k_l) = \frac{n!}{k_1! \dots k_l!} p_1^{k_1} \dots p_l^{k_l} \rightarrow \left(\frac{1}{\sqrt{2\pi}}\right)^{l-1} \sqrt{\frac{n}{n^\ell p_1 \dots p_\ell}} e^{-\frac{1}{2} \sum_1^l x_i^2 q_i} \quad (5.50)$$

für $n \rightarrow \infty$.

Natürlich ist in Satz 5.4 die Approximation besonders gut, wenn die x_i klein sind, d.h. k_i nahe bei np_i liegt. Für $l = 2$ erhalten wir für die Binomialverteilung

Satz 5.5 (*Lokaler Satz von Moivre Laplace*): Mit

$$x = \frac{k - np}{\sqrt{npq}} \quad \text{und} \quad a \leq x \leq b \quad \text{gilt}$$

$$b(n, k, p) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}x^2} F(n, k, p) \quad \text{wo } F(n, k, p) \rightarrow 1 \quad (5.51)$$

gleichmäßig in n .

Beweis: Es ist in diesem Fall $k = k_1, k_2 = n - k$ und

$$x_1 = \frac{k - np}{\sqrt{npq}} = x, \quad x_2 = \frac{n - k - np}{\sqrt{npq}} = -x.$$

Damit erhalten wir für $a \leq x \leq b$

$$b(n, k, p) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}x^2} F.$$

Der Korrekturfaktor F hat hier die Form

$$\begin{aligned} F &= \frac{[1 + \frac{1}{12n} + \dots + \frac{1}{288n^2} + \dots]}{[1 + \frac{1}{12k} + \dots][1 + \frac{1}{12(n-k)} + \dots]} e^{\frac{1}{6}x^3(q\sqrt{\frac{q}{np}} - \sqrt{\frac{q}{nq}})} \\ &+ \frac{x^4}{12} \left(\frac{q^2}{np} + \frac{p^2}{nq} \right) + \dots \cdot \frac{1}{\sqrt{(1 + x\sqrt{\frac{q}{np}})(1 - x\sqrt{\frac{p}{nq}})}} \\ &\approx \left(1 + \frac{1}{12n}\right) \left(1 - \frac{1}{12k}\right) \left(1 - \frac{1}{12(n-k)}\right) \\ &\quad \left(1 + \frac{1}{6}x^3 \left(q\sqrt{\frac{q}{np}} - p\sqrt{\frac{p}{np}}\right) + \frac{x^4}{12} \left(\frac{q^2}{np} + \frac{p^2}{nq}\right)\right) \\ &\quad \cdot \left(1 - \frac{1}{2}x \left(\sqrt{\frac{q}{np}} - \sqrt{\frac{p}{nq}}\right) + \dots\right). \end{aligned}$$

Bemerkung 5.2 Für $p = q = \frac{1}{2}$ ist

$$F = \left(1 + \frac{1}{12n} - \frac{1}{12k} - \frac{1}{12(n-k)} + \frac{x^4}{12n} + \frac{x^2}{2n} + o\left(\frac{1}{n^{\frac{3}{2}}}\right)\right).$$

Für große n ist tatsächlich F sehr nahe bei 1. Da man die $b(n, k, p)$ für $n \leq 100$ tabelliert hat, und in Anwendungen ohnehin meist nur eine Genauigkeit bis 2 oder 3 Stellen hinter dem Komma erforderlich ist, ist

$$b(n, k, p) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

eine recht gute Näherung, solange

$$\left| \frac{k - np}{\sqrt{npq}} \right| \leq K.$$

Dies gilt auch, weil der wesentliche Teil der $b(n, k, p)$ in der Nähe von $k = np$ konzentriert ist und meist schnell abfällt. Ein expliziter Vergleich für verschiedene n und p bestätigt dies. Satz 5.5 können wir nun verwenden, um einen weiteren Grenzwertsatz herzuleiten.

Satz 5.6 (DeMoivre-Laplace): Es sei X die Zufallsvariable Anzahl der „0“ in einem n -fach wiederholten Bernoulli-Experiment. Dann gilt

$$P \left[a \leq \frac{k - np}{\sqrt{npq}} \leq b \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \quad (5.52)$$

Hierbei ist $P[\]$ die Wahrscheinlichkeit, daß $\frac{k-np}{\sqrt{npq}}$ zwischen a und b liegt.

Beweis: Es ist wegen Satz 5.5

$$\sum' b(n, k, p) \approx \frac{1}{\sqrt{2\pi}} \sum' \frac{1}{\sqrt{npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

Dabei bedeutet \sum' , daß die Summe über alle k zu erstrecken ist mit $np + a\sqrt{npq} \leq k \leq np + b\sqrt{npq}$. Das kleinste zulässige k sei α und das größte β . Dann gilt

$$\alpha - 1 < np + a\sqrt{npq} \leq \alpha \quad \text{oder} \quad \frac{\alpha - 1 - np}{\sqrt{npq}} < a < \frac{\alpha - np}{\sqrt{npq}}.$$

Dies zeigt

$$\left| a - \frac{\alpha - np}{\sqrt{npq}} \right| < \frac{1}{\sqrt{npq}}.$$

Entsprechend sieht man

$$\left| b - \frac{\beta - np}{\sqrt{npq}} \right| < \frac{1}{\sqrt{npq}}.$$

Es sei nun $\alpha = k_1 < k_1 + 1 = k_2 < \dots < k_l = \beta$ und $x_i = \frac{k_i - np}{\sqrt{npq}}$. Dann ist

$x_{i+1} - x_i = \frac{1}{\sqrt{npq}}$ und \sum' ist offensichtlich gerade die Riemannsche Zwischensumme für $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$, denn $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ ist glatt.

Bezeichnet nun $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{\lambda^2}{2}} d\lambda$ die Verteilungsfunktion der Normalverteilung mit $\mu = 0$ und $\sigma = 1$ (Standard-Normalverteilung), so läßt sich (??) auch als

$$P \left[a \leq \frac{k - np}{\sqrt{npq}} \leq b \right] = \phi(b) - \phi(a). \quad (5.53)$$

schreiben. Die Werte der Funktion $\phi(x)$ sind tabelliert. Man sieht leicht, daß

$$1 - \phi(x) = \phi(-x) \quad \forall x \in \mathbb{R}. \quad (5.54)$$

Aus diesem Grund benötigt man Tabellen nur für $x \geq 0$. Ferner gilt

$$1 - \phi(x) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \left(\frac{1}{x} - \frac{1}{x^3} + \frac{1 \cdot 3}{x^5} - \frac{1 \cdot 3 \cdot 5}{x^7} + \dots \right). \quad (5.55)$$

Daher ist ϕ meist nur für $0 \leq x \leq 4$ tabelliert, denn wir haben etwa $1 - \phi(3) = 0,0014$, $1 - \phi(4) = 0,000033$.

Betrachten wir nun einige Beispiele zur Normalverteilung.

Beispiel 5.8 Wie groß ist die Wahrscheinlichkeit beim $2n$ -maligen Münzwerfen mit einer fairen Münze, daß die Anzahl Kopf um weniger als k von n differiert? Hier ist $p = \frac{1}{2}$ und $n \rightarrow 2n$. Als kleinster Wert ist $n - k$ zugelassen. Diesem Wert entspricht der Wert

$$a = \frac{n - k - 2n \cdot \frac{1}{2}}{\sqrt{2n \cdot \frac{1}{2} \cdot \frac{1}{2}}} = -\frac{k\sqrt{2}}{\sqrt{n}}.$$

Analog erhält man $b = +\frac{k\sqrt{2}}{\sqrt{n}}$ und damit ist die gesuchte Wahrscheinlichkeit

$$\phi\left(\frac{k\sqrt{2}}{\sqrt{n}}\right) - \phi\left(-\frac{k\sqrt{2}}{\sqrt{n}}\right) = 2\phi\left(\frac{k\sqrt{2}}{\sqrt{n}}\right) - 1.$$

Für $n = 100$ und $k = 10$ gibt dies $2\phi(\sqrt{2}) - 1 = 0,814$ und für $n = 1000$ und $k = 10$ erhält man $2\phi(\sqrt{\frac{2}{10}}) - 1 = 0,3448$.

Beispiel 5.9 *Stichproben:* Die Wahlchancen der Partei XYZ sollen durch eine Befragung von n Personen mit einer Genauigkeit von $\alpha\%$ ermittelt werden. Es sei p' der Anteil der XYZ Wähler in dieser Stichprobe und p der Anteil der XYZ Wähler in der Bevölkerung. Natürlich wird man nicht $p = p'$ erwarten können, denn durch Zufall könnte man ja nur auf XYZ Wähler stoßen. Das Beste, was wir hier machen können, ist, daß die Wahrscheinlichkeit, daß $|p - p'| \leq \frac{\alpha}{100}$ ist, sehr groß wird. Diese Wahrscheinlichkeit wird der *Konfidenzgrad* genannt. Bezeichnen wir ihn mit β . Üblich gilt etwa $\beta = 0,95$. Wegen der Größe der Bevölkerung in der Bundesrepublik können wir die Stichprobenentnahme als ein n -faches Bernoulli-Experiment mit der Wahrscheinlichkeit p auffassen. Ist hier X die Zufallsvariable Anzahl XYZ Wähler, so wollen wir

$$P\left[|p - p'| \leq \frac{\alpha}{100}\right] = P\left[|X - np| \leq \frac{\alpha}{100}n\right] \geq \beta$$

machen. Es ist aber wegen Satz (5.6)

$$\begin{aligned} P\left(|X - np| \leq \frac{\alpha}{100}n\right) &= \phi\left(\frac{\alpha n}{100\sqrt{npq}}\right) - \phi\left(-\frac{\alpha n}{100\sqrt{npq}}\right) \\ &= 2\phi\left(\frac{\alpha}{100}\sqrt{\frac{n}{p \cdot q}}\right) - 1 \geq \beta. \end{aligned}$$

Wählen wir etwa $\beta = 0,95(0,99)$, so finden wir aus der Tafel: $1,96 = \frac{\alpha}{100}\sqrt{\frac{n}{p \cdot q}}$ bzw. $2,575 = \frac{\alpha}{100}\sqrt{\frac{n}{p \cdot q}}$, denn $2\phi(1,96) - 1 = 0,95$ und $2\phi(2,575) - 1 = 0,99$. Wir erhalten

somit

$$\frac{(196)^2 \cdot p \cdot q}{\alpha^2} \leq n \quad \text{bzw.} \quad n \geq \frac{(257,5)^2 pq}{\alpha^2}.$$

Hierin ist immer noch die Unbekannte $p(1-p) = pq$ enthalten. Da stets $p \cdot q \leq \frac{1}{4}$ erhalten wir als Abschätzung

$$n \geq \frac{1}{\alpha^2} 98^2 \quad \text{bzw.} \quad n \geq (129)^2 \cdot \frac{1}{\alpha^2}.$$

Bei einer 1% Genauigkeit und 95% Konfidenzgrad muß man also immerhin fast 10.000 Personen befragen. Man vergleiche dies mit den üblichen Befragungen, die von Zeitungen oder kleinen Instituten veranstaltet werden, wobei meist etwa 2.000 Personen befragt werden.

Wir wollen nun Satz 5.6 noch etwas anders interpretieren.

Es sei wieder $(\Omega_1^n, \sum_1^n, P_1)$ der Wahrscheinlichkeitsraum für das n -malige unabhängige Wiederholen eines Bernoulli-Experimentes und es sei Y_n die Zufallsvariable

$$Y_n = \frac{X - np}{\sqrt{npq}}.$$

Dann strebt Y_n gegen eine standard-normalverteilte Zufallsvariable (Satz 5.6). Hierbei ist noch einiges zu klären, denn

- i) Der hier verwendete Konvergenzbegriff ist noch unklar.
- ii) Alle Zufallsvariablen Y_n sind auf verschiedenen Wahrscheinlichkeitsräumen erklärt.

Der 2. Einwand läßt sich leicht beheben, wenn man den Wahrscheinlichkeitsraum $(\Omega_1^\infty, \sum_1^\infty, P_1^\infty) = (\Omega, \sum, P)$, den wir in 5.2.5. schon eingeführt haben, betrachtet. Auf diesem Wahrscheinlichkeitsraum sei X_i die Zufallsvariable

$$X_i(\omega) = \begin{cases} 1, & \omega_i = 0 \\ 0, & \omega_i = 1 \end{cases} \quad \text{wo} \quad \omega = (\omega_i) \in \Omega. \quad (5.56)$$

Dann ist $\sum_{i=1}^n X_i$ die Zufallsvariable „Anzahl der „0“ in den ersten n Versuchen“. Dann definiere

$$Y_n = \frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - p}{\sqrt{pq}}. \quad (5.57)$$

Satz 5.6 zeigte dann:

Y_n strebt im Sinne der Verteilung gegen eine standard-normalverteilte Zufallsvariable, d.h. für jedes Intervall (Borelmenge) A gilt

$$(Y_n \cdot P)(A) \rightarrow \frac{1}{\sqrt{2\pi}} \int \chi_{AE}^{-\frac{1}{2}x^2} d\lambda. \quad (5.58)$$

Hierbei bedeutet $Y_n P$ natürlich wieder das Bildmaß von P unter Y_n .

In dieser Form ist Satz 5.6 Spezialfall des allgemeinen *Zentralen Grenzwertsatzes*, den wir nun nach einigen Vorbereitungen, allerdings zunächst ohne Beweis, angeben wollen.

Es sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und X eine Zufallsvariable auf Ω . Dann bezeichnen wir mit $\Sigma(X)$ die σ -Algebra $\{X^{-1}(A) | A \in \mathcal{B}_1\}$. Wir nennen $\Sigma(X)$ die von X erzeugte σ -Algebra. Entsprechend definiert man die σ -Algebra $\Sigma(X_{i|i \in I})$, die von einer Familie von Zufallsvariablen $X_i, i \in I$ erzeugt wird.

Definition 5.9 Die Zufallsvariable $X_i, i \in I$ heißen *unabhängig*, wenn für jede Auswahl $A_i \in \Sigma(X_i)$ die Mengen (A_i) unabhängig sind.

Diese Begriffe werden wir später wieder aufgreifen (nächstes Kapitel), zunächst aber wollen wir den Zentralen Grenzwertsatz von Lindeberg (1922) und Levy (1925) angeben.

Satz 5.7 Zentraler Grenzwertsatz: Es sei $X_i, i = 1, 2, 3 \dots$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) (alle haben die gleiche Verteilung) mit endlicher positiver Varianz. Dann strebt

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - a}{\sigma}, \quad \text{wo } a = E(X_i) \quad \text{und} \quad \sigma = \sqrt{\sigma^2(X_i)}$$

gegen eine standard-normalverteilte Zufallsvariable, d.h.

$$Y_n P([a, b]) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} d\lambda.$$

In Satz 5.6 war $X_i = \text{Anzahl „0“ im } i\text{-ten Wurf}$. Da dieser Satz in der Wahrscheinlichkeitsrechnung von großer Bedeutung ist, gibt es natürlich inzwischen eine große Zahl von Verallgemeinerungen davon.

5.3.3 Lage- und Skalenparameter

Obwohl die Normalverteilung von zwei Parametern μ, σ abhängt, genügt es doch, eine Tabelle für die Standard-Normalverteilung $N(0, 1)$ zu haben, weil man aus ihr leicht die entsprechende Information für $N(\mu, \sigma)$ gewinnen kann. Der Grund liegt darin, daß $N(\mu, \sigma)$ durch eine einfache Transformation aus $N(0, 1)$ hervorgeht. Da dieser Prozeß häufig verwandt wird, wollen wir ihn nun etwas allgemeiner betrachten.

Es sei X eine stetig verteilte Zufallsvariable mit Dichte f_X . Ferner sei $b \in \mathbb{R}$ und $a > 0$. Dann ist auch $Y = aX + b$ stetig verteilt mit Dichte

$$f_Y(t) = \frac{1}{a} f_X\left(\frac{t-b}{a}\right).$$

Dazu betrachte die Verteilungsfunktionen

$$F_X \text{ bzw. } F_Y(t) = P(aX + b \leq t) = P\left(X \leq \frac{t-b}{a}\right) = F_X\left(\frac{t-b}{a}\right)$$

und differenziere diesen Ausdruck nach t .

Bei diesem Verfahren nennt man b einen *Lageparameter* und a einen *Skalenparameter*. Insbesondere ist $Y = \sigma X + \mu \sim N(\mu, \sigma)$ verteilt, wenn $X \sim N(0, 1)$ verteilt ist. Auch die allgemeine Gleichverteilung geht auf diese Weise aus der Gleichverteilung auf $[0, 1]$ hervor. Weitere Beispiele erkennt man unschwer an den nachfolgenden Verteilungsklassen.

5.3.4 Die Exponentialverteilung

Wir hatten bei der Diskussion der Poissonverteilung gesehen, daß die Wahrscheinlichkeit für den Zerfall von einem radioaktiven Atom durch e^{-at} gegeben ist. Die

Wahrscheinlichkeit, daß mindestens ein Zerfall in der Zeit $[0, t]$ registriert wird, ist daher

$$1 - e^{-at} = \int_0^t e^{-ax} d\lambda. \quad (5.59)$$

$1 - e^{-at}$ ist daher die Verteilungsfunktion für die „Wartezeit“ auf einem Zerfall. Die Verteilung mit der Dichte

$$f(x) = \begin{cases} ae^{-ax}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (5.60)$$

wird *Exponentialverteilung* genannt. Entsprechend heißt die Zufallsvariable X auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) exponential verteilt, wenn XP die Dichte f hat. Für die momenterzeugende Funktion der Exponentialverteilung haben wir für $|t| < a$

$$M(t) = a \int_0^{\infty} e^{tx-ax} d\lambda(x) = a \int_0^{\infty} e^{-(a-t)x} d\lambda = \frac{a}{a-t}. \quad (5.61)$$

Daraus erhalten wir

$$\begin{aligned} E(X) &= \frac{d}{dt} M|_0 = + \frac{a}{(a-t)^2} |_0 = \frac{1}{a} \\ E(X^2) &= \frac{d^2}{dt^2} M|_0 = \frac{2a}{(a-t)^3} |_0 = \frac{2}{a^2}. \end{aligned} \quad (5.62)$$

Die Exponentialverteilung spielt eine wichtige Rolle bei vielen Prozessen, insbesondere der Zuverlässigkeitstheorie.

5.3.5 Die Laplace-Verteilung

Die Laplace Verteilung wird durch die Dichte

$$f(x) = \frac{a}{2} e^{-|x-b|} \quad (5.63)$$

beschrieben. Man zeigt leicht

$$E(X) = b \quad \text{und} \quad E(X^2) = b^2 + \frac{2}{a}.$$

5.3.6 Die Gamma-Verteilung

Eine weitere Verteilung, die die Exponentialverteilung verallgemeinert, ist die Gamma Verteilung. Sie spielt in der Statistik eine Rolle. Ihre Dichte ist

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x \geq 0 \quad \beta, \alpha > 0 \\ 0, & x < 0. \end{cases} \quad (5.64)$$

Dies ist in der Tat eine Wahrscheinlichkeitsdichte, da

$$\int_0^{\infty} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^{\alpha} \int_0^{\infty} u^{\alpha-1} e^{-u} du = \beta^{\alpha} \Gamma(\alpha).$$

Die Momente der Γ -Verteilung können wir vollkommen entsprechend leicht berechnen, denn

$$\begin{aligned} E(X^k) &= \frac{1}{\beta^{\alpha} \Gamma(\alpha)} \int_0^{\infty} x^{\alpha+k-1} e^{-\frac{x}{\beta}} d\lambda = \frac{\beta^{\alpha+k} \Gamma(\alpha+k)}{\beta^{\alpha} \Gamma(\alpha)} \\ &= \beta^k \alpha(\alpha+1) \dots (\alpha+k-1). \end{aligned} \quad (5.65)$$

Für die momenterzeugende Funktion der Γ -Verteilung erhalten wir

$$\begin{aligned} M(t) &= 1 + \beta \alpha t + \alpha(\alpha+1) \beta^2 \frac{t^2}{2!} + \dots + \alpha(\alpha+1) \dots (\alpha+k-1) \beta^k \frac{t^k}{k!} + \dots \\ &= (1 - \beta t)^{-\alpha}. \end{aligned}$$

Dies hätten wir natürlich auch direkt zeigen können. Die Gammaverteilung wird häufig bei Lebensdauerproblemen und in der Zuverlässigkeitstheorie verwandt. Für $\alpha = 1$ erhält man die Exponentialverteilung und für ganzzahlige α die Erlangverteilung. Sie ist konjugiert zur Poisson-Verteilung.

Eine Summe endlich vieler unabhängiger $\Gamma(\alpha_i, \beta)$ verteilter Zufallsvariablen X_1, \dots, X_n ist $\Gamma(\alpha_1 + \dots + \alpha_n, \beta)$ verteilt.

5.3.7 Die Beta-Verteilung

Die Dichte der *Beta-Verteilung* ist

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha, \beta > 0. \quad (5.66)$$

Diese Verteilung hat Anwendungen in der Bayes-Statistik.

Schaut man sich die Definition der Beta-Funktion an, sieht man sofort, daß f eine Wahrscheinlichkeitsdichte ist, denn es ist nach einem bekannten klassischen Ergebnis

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} d\lambda(x) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (5.67)$$

(siehe etwa Fichtenholz Bd. III).

Für die Momente erhalten wir damit sofort

$$\begin{aligned} E(X^k) &= \int_0^1 x^{k+\alpha-1} (1-x)^{\beta-1} d\lambda(x) \cdot \frac{\Gamma(\alpha + \beta) \Gamma(k + \alpha) \Gamma(\beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(k + \alpha + \beta)} \\ &= \frac{(\alpha + k - 1) \dots \alpha}{(\alpha + \beta + k - 1) \dots (\alpha + \beta)}. \end{aligned} \quad (5.68)$$

Es gibt noch weitere wichtige stetige Verteilungen, die besonders in der Statistik angewendet werden. Dies sind die sogenannten Testverteilungen, die wir behandeln werden, wenn wir Tests diskutieren.

5.3.8 Weibull-Verteilung

Es bezeichne $h(t)$ die Risikorate einer Komponente, deren Lebensdauer durch die Zufallsvariable X beschrieben wird. Dann ist

$$h(t) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P(t \leq X \leq t + \delta | X \geq t).$$

Mit der Verteilungsfunktion $P(X \leq t) = 1 - F(t)$ ergibt dies

$$h(t) = \lim_{\delta \rightarrow 0} \frac{1 - F(t + \delta) - (1 - F(t))}{\delta \cdot F(t)} = -\frac{F'}{F}$$

oder

$$F(t) = \exp \left(- \int_0^t h(s) ds \right).$$

Die Dichte für X ist dann

$$f(t) = h(t) \exp \left(- \int_0^t h(s) ds \right) \quad \text{auf } [0, \infty].$$

Für $h(t) = \beta^{-\gamma} \gamma t^{\gamma-1}$ erhält man die Weibullverteilung mit der Dichte $\beta^{-\gamma} \gamma t^{\gamma-1} \exp(-(\frac{t}{\beta})^\gamma)$ auf $(0, \infty)$.

Erwartungswert = $\beta \Gamma(\frac{1}{\alpha})$, Varianz = $\beta^2 [\Gamma(\frac{2}{\alpha} + 1) - \Gamma^2(\frac{1}{\alpha} + 1)]$.

Anwendung: Zuverlässigkeit, Risikotheorie, Lebensdauer, Wartezeiten, Inkubationszeit bei Krankheiten.

5.3.9 Lognormal-Verteilung: $LN(a, m, \sigma)$

Wenn für ein a die Zufallsvariable $\ln(X - a)$ -normalverteilt mit Parametern μ und σ ist, so ist X *lognormalverteilt* mit der Dichte

$$\frac{1}{(x - a)\sigma\sqrt{2\pi}} \exp \left(-\frac{(\ln \frac{x-a}{m})^2}{2\sigma^2} \right) \quad x > a, m = e^\mu.$$

Erwartungswert = $a + m \exp(\frac{1}{2}\sigma^2)$, Varianz = $m^2 \exp(\sigma^2)(\exp(\sigma^2) - 1)$.

Anwendung: Verteilung von Einkommen und Wohlstand, Größenverteilung bei Bruch-

stücken, Verteilung von Umweltgiften, Inkubationszeit bei Krankheiten.

Es gibt ein multiplikatives Analogon des zentralen Grenzwertes.

Aufgabe 5.8 Beim Kniffelwürfeln wird mit 5 Würfeln 3mal gewürfelt. Ziel ist es, 5 gleiche Zahlen zu erreichen. Dabei kann man nach jedem Wurf eine beliebige Anzahl von Würfeln „liegen lassen“. Wie groß ist die Wahrscheinlichkeit, einen Kniffel zu erreichen?

Aufgabe 5.9 Die nachstehende Tabelle ist eine Statistik über 647 Unfälle. Beschreiben Sie diese durch eine negative Binomialverteilung mit $r = .96$ und $p = .67$.

Anzahl von Unfällen k	0	1	2	3	4	5	≥ 6
Beobachtete Häufigkeit	447	132	42	21	3	2	0

Warum sollte man gerade diese Werte wählen?

Aufgabe 5.10 Ein Marktforschungsunternehmen hat untersucht, wieviele Einheiten k eines bestimmten Produktes gekauft wurden. Es wurden insgesamt 1.750 Käufer beobachtet

k	0	1	2	3	4	5	6	7	8	9	≥ 10
Häufigkeit	1 671	43	19	9	2	3	1	0	0	2	0

Beschreiben Sie diese Daten durch eine negative Binomialverteilung mit $r = 0.041$ und $p = 0.315$.

Aufgabe 5.11 Bei einer Verkehrsstudie von 1.011 Personenfahrzeugen wurde die Anzahl k der Beifahrer bestimmt

k	0	1	2	3	4	≥ 5
Häufigkeit	678	227	56	28	8	14

Bestimmen Sie den Mittelwert und die Varianz und benutzen Sie dies um r, p für eine negative Binomialverteilung zu bestimmen. Beschreiben Sie damit diese Daten.

Aufgabe 5.12 Richardson hat die Häufigkeit von Kriegen studiert und eine Statistik über den Ausbruch von Kriegen pro Jahr für den Zeitraum 1500 bis 1931 erstellt. Er fand

Anzahl k von Kriegen, die pro Jahr ausbrechen	0	1	2	3	4	≥ 5
Häufigkeit	223	142	48	15	4	0

Wenn Kriege zufällig ausbrechen, sollten diese Daten durch welche Verteilung beschrieben werden? Überprüfen Sie dies und interpretieren Sie das Ergebnis.

Aufgabe 5.13 In einem Gebäude der New Yorker Stadtverwaltung wurden vom 01.11.23 bis zum 30.09.25 insgesamt 423 Dinge verloren. Die Tageshäufigkeit k gefundene Stücke war

k	0	1	2	3	4	5	6	7
Häufigkeit	161	134	74	32	11	2	0	1

Welche Verteilung paßt dazu? Warum?

Aufgabe 5.14 Nachstehend ist eine Statistik über die Anzahl k der jährlichen Arztbesuche von 2.810 Personen

k	0	1	2	3	4	5	6	7	8	9	10	11	12
Häufigkeit	850	535	361	283	201	149	106	76	77	54	32	31	27

k	13	14	15	16	17	18	19
Häufigkeit	14	3	6	8	3	2	14

Aufgabe 5.15 Die Wahrscheinlichkeit eines Ereignisses A sei gleich p . Man führe n unabhängige Versuche durch und bezeichne mit h die relative Häufigkeit von A in dieser Versuchsreihe.

- Es sei $p = 0,4$ und $n = 1500$. Wie groß ist die Wahrscheinlichkeit, daß h zwischen $0,40$ und $0,44$ liegt?
- Es sei $n = 14400$. Für welche Werte von p wird die Wahrscheinlichkeit für $|h - p| < 0,01$ mindestens $0,99$?

(Satz von Moivre-Laplace)

Aufgabe 5.16 Man beweise den lokalen Satz von Moivre-Laplace.

Aufgabe 5.17 Die stetige Funktion f sei die Wahrscheinlichkeitsdichte einer nicht-negativen Zufallsvariablen X mit $E(X) \in \mathbb{R}$. Man zeige: $E(X) = \int_0^{\infty} [1 - F(x)] dx$. Dabei ist F die Verteilungsfunktion von X .

Aufgabe 5.18 Die Zufallsvariable X sei normalverteilt. Die vorgegebene Normalverteilung soll durch die Gleichverteilung im Intervall $[\alpha, \beta]$ approximiert werden. Man bestimme α und β so, daß der Erwartungswert und die Varianz von X unverändert bleibt.

Aufgabe 5.19 Bei der Messung der Entfernung eines Objektes treten systematische und zufällige Fehler auf. Der systematische Fehler betrage 50 m, wobei er die Entfernung verkürzt. Die zufälligen Fehler seien normalverteilt mit einer Standardabweichung von $\sigma = 100$ m. Man bestimme

- Die Wahrscheinlichkeit dafür, daß der absolute Fehler der Entfernungsmessung 150 m nicht überschreitet;
- die Wahrscheinlichkeit dafür, daß die gemessene Entfernung die exakte Entfernung nicht übertrifft.

Aufgabe 5.20 Etwa 6% aller Wähler wählen die FDP. Wieviele Wähler mußten in einer Umfrage befragt werden, um das Ergebnis bis auf 5% genau zu wissen?

Aufgabe 5.21 Die Verteilungsfunktion von Weibull

$$F(x) = \begin{cases} 1 - e^{-\frac{x^m}{x_0}} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \text{mit } m > 0, x_0 > 0$$

charakterisiert in einer Reihe von Fällen die Lebensdauer von Elementen einer elektrischen Apparatur. Man bestimme

-
- a) die Wahrscheinlichkeitsdichte $f(x)$;
 - b) den Quantil der Verteilung der Ordnung P ;
 - c) den Quartilenabstand;
 - d) den Modalwert der Verteilung.

Kapitel 6

Zufallsvektoren

6.1 Mehr über Zufallsvariable

Es sei wieder (Ω, Σ, P) ein Wahrscheinlichkeitsraum, den wir uns als fest vorgegeben vorstellen wollen. Ferner seien X, X_1, X_2, \dots Zufallsvariable darauf. Dann haben wir wegen der Linearität des Integrals

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i). \quad (6.1)$$

Neben dem Erwartungswert als Lageparameter der Verteilung einer Zufallsvariablen X sind noch die Quantile und der Modalwert gebräuchlich. Dabei ist das *Quantil der Ordnung p* der Wert x_p , so daß links von x_p der Anteil p und rechts davon der Anteil $1 - p$ liegt. Für stetige Verteilungen gilt also

$$P([X < x_p]) = F_x(x_p) = p, \quad P([X > x_p]) = 1 - p. \quad (6.2)$$

Das Quantil der Ordnung $\frac{1}{2}$ wird *Median* genannt. Der Median wird besonders dann verwendet, wenn $E(X)$ nicht existiert (Beispiel 6.2), denn die Quantile existieren stets. Ein weiterer Grund, den Median statt des Erwartungswertes zu verwenden, ist die starke Abhängigkeit des Erwartungswertes von extrem großen bzw. kleinen Werten von X . Bei einer diskreten Zufallsvariable mit n Werten, die alle gleich wahrscheinlich sind, ist der Median der $\frac{n+1}{2}$ Wert, wenn n ungerade ist, und das Mittel des $\frac{n}{2}$ und $\frac{n+2}{2}$ Wertes, wenn n gerade ist (wenn man sich die Werte der Größe nach geordnet denkt).

Der Wert $x_{\frac{1}{4}}$ ($x_{\frac{3}{4}}$) heißt auch unteres (oberes) *Quantil* und man nennt $x_{\frac{3}{4}} - x_{\frac{1}{4}}$ den *Quartilenabstand* von x . Falls $E(X^2)$ nicht existiert, wird der Quartilenabstand auch als Maß für die Streuung (Dispersion) von X verwendet. Ein weiterer Lageparameter von X ist der Modalwert. Der *Modalwert* ist der wahrscheinlichste Wert von X , also das Maximum der Dichte von X , wenn X stetig verteilt ist und dieses Maximum existiert.

Beispiel 6.1 Für die gleichmäßige und die Normalverteilung stimmen Mittelwert und Median überein. Aber die gleichmäßige Verteilung hat keinen eindeutigen Modalwert, während dieser bei der Normalverteilung gleich dem Erwartungswert ist.

Bei der Exponential-Verteilung ist $\frac{1}{a}$ der Mittelwert, 0 der Modalwert und der Median berechnet sich aus

$$a \int_0^m e^{-ax} d\lambda(x) = 1 - e^{-am_x} = \frac{1}{2} \quad \text{oder} \quad 2 = e^{am_x} \quad \text{oder} \quad m_x = \frac{\ln 2}{a}.$$

Beispiel 6.2 Eine Zufallsvariable heißt *Cauchy* verteilt, wenn ihre Dichte durch $f(x) = \frac{k}{a^2+x^2}$ $x \in \mathbb{R}$ mit

$$\frac{1}{k} = \int_{-\infty}^{\infty} \frac{d\lambda(x)}{a^2+x^2} = \frac{1}{a} \int \frac{du}{1+u^2} = \frac{\pi}{a} \text{ ist.}$$

Der Median- und Modalwert der Cauchy-Verteilung sind 0, während der Mittelwert nicht existiert.

Beispiel 6.3 In einer Schulklasse von 20 Schülern wurde die Körpergröße gemessen. Dabei wurden die folgenden Werte erhalten (in cm):

147, 149, 151, 148, 146, 143, 158, 159, 152, 141, 149, 157, 152, 149, 146, 145, 147, 152, 153, 156.

Der zugehörige Wahrscheinlichkeitsraum ist hier $\{1, 2, \dots, 20\} = \Omega$ mit $P(i) = \frac{1}{20}$ und $X(1) = 147, X(2) = 149, \dots$. Dann ist der Erwartungswert $E(X) = \frac{1}{20}(147 + \dots + 156) = 150$ und $\sigma(X) = \sqrt{\frac{464}{20}} = 4,82$. Der Median ist 149, der untere Quartil $x_{\frac{1}{4}} = 146,25$ und der obere Quartil 152,75. Damit erhält man als Quartilenabstand $6,5$.

Satz 6.1 Es sei X eine Zufallsvariable für die $E(|X - a|^r)$ existiert, $r > 0$. Dann gilt für jedes $\lambda > 0$

$$P(|X - a| \leq \lambda) \geq 1 - \frac{E(|X - a|^r)}{\lambda^r}. \quad (6.3)$$

Beweis: Es ist

$$\begin{aligned} E(|X - a|^r) &= \int |x - a|^r dXP = \int_{|x-a| \leq \lambda} + \int_{|x-a| > \lambda} \geq \int_{|x-a| \leq \lambda} |x - a|^r dXP + \lambda^r \\ &\int_{|x-a| > \lambda} dXP \geq \lambda^r P(|x - a| > \lambda). \end{aligned}$$

Es ist aber $P(|x - a| \leq \lambda) = 1 - P(|x - a| > \lambda) \geq 1 - \frac{E(|x - a|^r)}{\lambda^r}$. Für $r = 2, a = E(X)$ und $\lambda = k, \sigma = \sigma(X)$ erhält man aus dem Satz die *Chebyscheffsche Ungleichung*

$$P(|X - EX| \leq k\sigma) \geq 1 - \frac{1}{k^2}. \quad (6.4)$$

Zum Vergleich: Wir haben für die Normalverteilung

$$P(|X - EX| \leq k\sigma) = \phi(k) - \phi(-k) = 2\phi(k) - 1 = \begin{cases} 0,6826 & k = 1 \\ 0,9546 & k = 2 \\ 0,9972 & k = 3 \\ 0,9993 & k = 4. \end{cases}$$

Wegen des Zentralen Grenzwertsatzes gilt diese Abschätzung auch für die Binomialverteilung für große n .

6.2 Zufallsvektoren

Es seien $(\Omega_i, \sum_i, \mu_i)$ σ -endliche Maßräume, $i = 1, 2$. In $\sum_1 \otimes \sum_2$ betrachte dann das System R aller endlichen Vereinigungen von Rechtecken der Form $A \times B$ mit $A \in \sum_1, B \in \sum_2$ mit $\mu_1(A) < \infty$ und $\mu_2(B) < \infty$. Dann ist R ein Ring und jede Menge aus R ist disjunkte Vereinigung von Rechtecken. Für $\bigcup_{i=1}^n A_i \times B_i \in R$ mit $A_i \times B_i \cap A_j \times B_j = \emptyset, i \neq j$, definiere nun

$$\mu(A_i \times B_i) = \sum_{i=1}^n \mu_1(A_i) \cdot \mu_2(B_i). \quad (6.5)$$

Dann ist (??) unabhängig von der gewählten Darstellung von $\bigcup_{i=1}^n A_i \times B_i$ als disjunkte Vereinigung und μ ist ein Inhalt auf R . μ ist σ -endlich, wenn μ_1 und μ_2 es sind, denn $A_n \nearrow \Omega_1, B_n \nearrow \Omega_2$ und $\mu_1(A_n) < \infty, \mu_2(B_n) < \infty$ gibt $\mu(A_n \times B_n) < \infty$ und $A_n \times B_n \nearrow \Omega_1 \times \Omega_2$. μ ist sogar \emptyset -stetig. Man nennt μ das *Produktmaß* der μ_i und schreibt auch $\mu = \mu_1 \otimes \mu_2$. Es gilt dann der wichtige *Satz von Fubini* (Für den Beweis sei auf das Buch von H. Bauer Kapitel 22 verwiesen).

Satz 6.2

a) Es sei $f : \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}_+$ $\sum_1 \otimes \sum_2 - \mathcal{B}_1$ messbar. Dann sind die Funktionen

$$\begin{aligned} \omega_2 &\rightarrow \int f(\omega_1, \omega_2) d\mu_1(\omega_1) \quad \text{und} \\ \omega_1 &\rightarrow \int f(\omega_1, \omega_2) d\mu_2(\omega_2). \end{aligned}$$

$\sum_2 - \mathcal{B}_1$ bzw. $\sum_1 - \mathcal{B}_1$ messbar und es gilt

$$\begin{aligned} \int f d\mu_1 \otimes \mu_2 &= \int \left(\int f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \\ &= \int \left(\int f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \end{aligned} \quad (6.6)$$

- b) Ist $f \mu_1 \otimes \mu_2$ integrierbar, so sind $\omega_2 \rightarrow \int f(\omega_1, \omega_2) d\mu_1(\omega_1)$ und $\omega_1 \rightarrow \int f(\omega_1, \omega_2) d\mu_2(\omega_2)$ bzw. μ_1 integrierbar, und es gilt (??).
- c) für eine beliebige $\sum_1 \otimes \sum_2$ meßbare Funktion f sind ferner äquivalent:

- i) $f \in \mathcal{L}^1(\mu), \mu = \mu_1 \otimes \mu_2$.
- ii) $f(\cdot, \omega_2) \in \mathcal{L}^1(\mu_1)$ für μ_2 fast alle $\omega_2 \in \Omega_2$ und $\int |f(\omega_1, \dots)| d\mu_1(\omega_1) \in \mathcal{L}^1(\mu_2)$.
- iii) $f(\omega_1, \cdot) \in \mathcal{L}^1(\mu_2)$ für μ_1 fast alle $\omega_1 \in \Omega_1$ und $\int |f(\omega_2, \dots)| d\mu_2(\omega_2) \in \mathcal{L}^1(\mu_1)$.

Ist eine dieser Aussagen erfüllt, so gilt

$$\begin{aligned} \int f d\mu &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \end{aligned}$$

Der Satz von Fubini besagt also, daß es unter der Voraussetzung der $\mu_1 \otimes \mu_2$ Integrierbarkeit nicht auf die Reihenfolge der Faktoren ankommt.

Für den Satz von Fubini ist die σ -Endlichkeit *beider* Maße entscheidend. Natürlich läßt sich der Satz durch Induktion sofort auf n Faktoren verallgemeinern.

Aufgabe 6.1 Formuliere den entsprechenden Satz für Doppelreihen.

Aufgabe 6.2 Es sei h radialsymmetrisch $h(x) = h_1(|x|)$, $|x| = (x_1^2 + \dots + x_n^2)^{\frac{1}{2}}$ λ^n -integrierbar. Dann gilt

$$\int_{\mathbb{R}^n} h(x) d\lambda^n = \mathcal{O}_n \int_0^\infty r^{n-1} h_1(r) dr.$$

Dabei ist \mathcal{O}_n die Oberfläche der Einheitskugel im \mathbb{R}^n .

Aufgabe 6.3 Wähle insbesondere $h(x) = \exp(-|x|^2)$ und zeige mit Hilfe des Satzes von Fubini

$$I^n = \int h d\lambda_n = \mathcal{O}_n \int r^{n-1} e^{-r^2} dr \quad (6.7)$$

wo

$$I = \int e^{-x^2} d\lambda(x).$$

Für $n = 2$ läßt sich das Integral rechts in (6.7) berechnen, und damit läßt sich I und \mathcal{O}_n bestimmen.

Bezeichnet λ_n das Lebesgue Maß auf dem \mathbb{R}^n , so sieht man leicht aufgrund des oben Gesagten

$$\lambda_n \otimes \lambda_m = \lambda_{n+m}. \quad (6.8)$$

Es sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und $X = (X_1, \dots, X_n)$ ein Zufallsvektor auf Ω . Ferner sei $F : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathcal{B}_n \rightarrow \mathcal{B}_1$ messbar. Dann ist $F \circ X$ eine Zufallsvariable auf (Ω, Σ, P) und

$$\int F \circ X dP = \int F dXP.$$

Das Bildmaß XP auf \mathcal{B}_n heißt die *gemeinsame Verteilung* der X_1, \dots, X_n .

Beispiel 6.4 2 Punkte werden unabhängig aus $[0,1]$ ausgewählt, so daß die Verteilung für jeden gleichmäßig ist. Was ist ihre erwartete Entfernung? Als Ω wähle hier $[0,1] \times [0,1]$ und $X(x,y) = |x-y|$ und $P = \lambda_1 \otimes \lambda_1 = \lambda_2$. Dann ist wegen

$$\begin{aligned} EX &= \int |x-y| d\lambda_2 = 2 \int_{\Delta} (x-y) dx dy = 2 \int_0^1 \left(\int_0^x (x-y) dy \right) dx \\ &= 2 \int_0^1 x^2 - \frac{1}{2}x^2 = \frac{2}{6}. \end{aligned}$$

Definition 6.1

$$\begin{aligned} F_X(x_1, \dots, x_n) &= P([X \in (-\infty, x_1) \times \dots \times (-\infty, x_n)]) \\ &= P([X_i < x_i | i = 1, \dots, n]) \end{aligned}$$

heißt die *gemeinsame Verteilungsfunktion* der X_1, \dots, X_n . Man zeigt leicht:

Satz 6.3 F_X hat die Eigenschaften

- i) $0 \leq F(x_1, \dots, x_n) \leq 1$.
- ii) F ist in jeder Variablen isoton.
- iii) $\lim_{\substack{x_k \rightarrow -\infty \\ k=1, \dots, n}} F(x_1, \dots, x_n) = 0$, $\lim_{\substack{x_k \rightarrow \infty \\ k=1, \dots, n}} F(x_1, \dots, x_n) = 1$.
- iv) F ist in jeder Variablen linksseitig stetig.

Der Beweis verläuft analog zum eindimensionalen Fall. Anders als im eindimensionalen Fall aber charakterisieren die Eigenschaften i) ii) iii) und iv) nicht die Verteilungsfunktion eines Vektors, wie das folgende Beispiel zeigt.

Beispiel 6.5

$$F(x,y) = \begin{cases} 0 & x+y \leq 0 \\ 1 & x+y > 0. \end{cases}$$

Wäre dieses F die Verteilungsfunktion eines Zufallsvektors, so hätten wir

$$P(X_1 \in [a,b], X_2 \in [c,d]) = F(b,d) - F(a,d) + F(a,c) \geq 0.$$

Für $a, c = 0$ und $b, d = 1$ aber haben wir

$$F(1,1) - F(1,0) + F(0,0) = -1,$$

dies aber ist unmöglich.

Definition 6.2 Ein Zufallsvektor $X = (X_1, \dots, X_n)$ heißt diskret, wenn er nur abzählbar viele Werte annehmen kann. X heißt *stetig* mit der *Dichte* f ($f \geq 0$, $f \mathcal{B}_n$ - \mathcal{B}_1 messbar und $\int f d\lambda_n = 1$) falls für alle $A \in \mathcal{B}_n$

$$XP(A) = \int \chi_A f d\lambda_n. \quad (6.9)$$

Im letzteren Fall gilt allgemeiner für ein XP -quasiintegrierbares F

$$\int F \circ X dP = \int F(X_1, \dots, X_n) f(X_1, \dots, X_n) d\lambda_n. \tag{6.10}$$

Ein Beispiel für einen diskreten Zufallsvektor hatten wir schon kennengelernt, nämlich die Multinomialverteilung.

Ist $\tilde{X} = (X, Y)$ ein zweikomponenten Zufallsvektor, so interessiert man sich häufig für die Wahrscheinlichkeiten, die durch eine Komponente bestimmt werden, wobei die andere Komponente beliebige Werte annehmen kann. Die so entstehenden eindimensionalen Verteilungen heißen *Randverteilungen*. Sie sind durch

$$P_X(A) = \tilde{X}P(A \times \mathbb{R}), P_Y(B) = \tilde{X}P(\mathbb{R} \times B) \quad A, B \in \mathcal{B}_1 \tag{6.11}$$

festgelegt.

Hat $\tilde{X}P$ die Dichte f , so haben auch P_X und P_Y Dichten, nämlich f_x, f_y . Diese sind durch

$$\begin{aligned} f_x(x) &= \int f(x, y) dy \\ f_y(y) &= \int f(x, y) dx \end{aligned} \tag{6.12}$$

gegeben. Wegen des Satzes von Fubini sind diese Funktionen messbar und Wahrscheinlichkeitsdichten, d.h.

$$\int f_x(x) dx = 1 = \int f_y(y) dy.$$

Beispiel 6.6 Ist (X, Y) ein diskreter Zufallsvektor mit

$P_{i,k} = P([X = x_i, Y = y_k])$	x^y	y_1	y_2	y_3	\dots	
$x_i \in$ Wertebereich X	x_1	p_{11}	p_{12}	p_{13}	\dots	p_1
$y_k \in$ Wertebereich Y	x_2	p_{21}	p_{22}	p_{23}	\dots	p_2
So läßt sich die Verteilung mit den Randverteilungen in dem nebenstehenden Schema besonders übersichtlich angeben.	x_3	p_{31}	p_{32}	p_{33}	\dots	p_3
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
		q_1	q_2	q_3	\dots	

Dabei gibt $\{p_i\}$ mit $p_i = \sum_k p_{ik}$ und $\{q_k\} = \sum_i p_{ik}$ die *Randverteilung* von X bzw. Y an.

Allgemeiner läßt sich, genau wie oben, die Randverteilung eines n -Komponenten Zufallsvektors nach k vorgegebenen Variablen definieren.

Will man 2 Zufallsvariablen X, Y vergleichen, so kann man sich dazu des Zufallsvektors (X, Y) bedienen. Für jede $\mathcal{B}_2 - \mathcal{B}_1$ messbare Funktion F haben wir dann wegen (??)

$$\int F \circ (X, Y) dP = \int F(x, y) d(X, Y)P.$$

Wählt man insbesondere $F(x, y) = G(x)$, so haben wir

$$\begin{aligned} E(G \circ X) &= \int G(x)d(X, Y)P \\ &= \int G(x)dXP. \end{aligned} \quad (6.13)$$

Entsprechendes gilt für $G(y)$. So gibt beispielsweise

$$\begin{aligned} E(X) &= \int xd(X, Y)P, & E(Y) &= \int yd(X, Y)P \\ \sigma^2(X) &= \int (x - EX)^2d(X, Y)P, & \sigma^2(Y) &= \int (y - EY)^2d(X, Y)P. \end{aligned}$$

Neben den quadratischen Ausdrücken $\sigma^2(X), \sigma^2(Y)$ kann man natürlich auch noch das gemischte Moment

$$\sigma_{x,y} = E((X - EX)(Y - EY)) = \int (x - EX)(y - EY)d(X, Y)P \quad (6.14)$$

betrachten. Man nennt diesen Ausdruck die *Kovarianz* von X und Y und schreibt auch $\sigma_{x,y} = \text{cov}(X, Y)$. Die Kovarianz ist ein Maß für die Abhängigkeit von X und Y . Wegen der Cauchy-Schwartz'schen-Ungleichung haben wir

$$\begin{aligned} |\sigma_{x,y}| &= \int |(x - EX)(y - EY)d(X, Y)P| \leq \left(\int (x - EX)^2d(X, Y)P \right)^{\frac{1}{2}} \\ &\quad \left(\int (y - EY)^2d(X, Y)P \right)^{\frac{1}{2}} \leq \sigma_x \sigma_y. \end{aligned}$$

Daher gilt

$$-1 \leq \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} = P_{x,y} \leq 1. \quad (6.15)$$

Falls $P_{x,y}$, der *Korrelationskoeffizient* von X und Y gleich ± 1 ist, so sind X und Y proportional.

Die Kovarianz bzw. der Korrelationskoeffizient ist ein Maß für die lineare Abhängigkeit.

Beispiel 6.7 Auf $\Omega = [0, 1]$, $\Sigma = B([0, 1])$ und $P = \lambda$ betrachte die Zufallsvariable

X und Y , wo $X(x) = x$ und $Y(x) = x^2$. Dann haben wir

$$E(X) = \int_0^1 x dx = \frac{1}{2}$$

$$E(Y) = \int_0^1 x^2 dx = \frac{1}{3}$$

$$\sigma^2(X) = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$$

$$\sigma^2(Y) = \int_0^1 \left(x^2 - \frac{1}{3}\right)^2 dx = \frac{1}{5} - \frac{2}{9} + \frac{1}{9} = \frac{4}{45}$$

$$\sigma_{x,y} = \int_0^1 \left(x - \frac{1}{2}\right) \left(x^2 - \frac{1}{3}\right) dx = \frac{1}{4} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{1}{12}.$$

Dies gibt

$$P_{x,y} = \frac{1}{12 \sqrt{\frac{1}{12}} \sqrt{\frac{4}{45}}} = \sqrt{\frac{45}{48}} = \sqrt{\frac{15}{16}} = \frac{1}{4} \sqrt{15}.$$

Beispiel 6.8 Ein Punkt ist gleichmäßig über $\Omega = [0, 1] \times [0, 1]$ verteilt. Es seien X und R die Zufallsvariable x -Koordinate und Entfernung von 0. Dann ist

$$E(X) = \int x dx dy = \frac{1}{2}$$

$$\sigma^2(X) = \int x^2 dx dy - \frac{1}{4} = \frac{1}{12}$$

$$\begin{aligned} E(R) &= \int \sqrt{x^2 + y^2} dx dy = 2 \int_0^1 \left(\int_0^x \sqrt{y^2 + x^2} dy \right) dx \\ &= 2 \int_0^1 \int_0^1 \left(\int_0^1 x^2 \sqrt{1 + \left(\frac{y}{x}\right)^2} d\frac{y}{x} \right) dx = 2 \int_0^1 \int_0^1 \left(\int_0^1 x^2 \sqrt{1 + u^2} du \right) dx \\ &= \frac{2}{3} \int_0^1 \sqrt{1 + u^2} du \\ &= \frac{1}{3} (\sqrt{2} + \ln(1 + \sqrt{2})) = 0,7652 = \alpha \\ \sigma^2(R) &= \int (x^2 + y^2) dx dy - \alpha^2 = \frac{2}{3} - \alpha^2 = 0,081142. \end{aligned}$$

Ferner ist

$$\begin{aligned}
 \sigma_{x,R} &= \int_0^1 \int_0^1 \left(x - \frac{1}{2}\right) (\sqrt{x^2 + y^2} - \alpha) dx dy \\
 &= \int_0^1 \int_0^1 x \sqrt{x^2 + y^2} dx dy - \frac{1}{2} \int_0^1 \int_0^1 \sqrt{x^2 + y^2} dx dy \\
 &= \frac{1}{3} \int_0^1 (1 + y^2)^{\frac{3}{2}} - y^3 dy - \int_0^1 \int_0^x \sqrt{x^2 + y^2} dx dy \\
 &= \frac{1}{3} \int_0^1 (1 + y^2)^{\frac{3}{2}} dy - \frac{1}{12} - \int_0^1 x^2 \int_0^1 \sqrt{1 + u^2} du = \\
 &= \frac{1}{12} \left\{ \sqrt{8} + \frac{3}{2} \sqrt{2} + \frac{3}{2} \ln(1 + \sqrt{2}) \right\} - \frac{1}{12} - \frac{1}{3} \frac{1}{2} (\sqrt{2} + \ln(1 + \sqrt{2})) \\
 &= \frac{1}{12} \left\{ \frac{3}{2} \sqrt{2} - 1 - \frac{1}{2} \ln(1 + \sqrt{2}) \right\} = 0,0567194625 \\
 P_{x,R} &= 0,68976.
 \end{aligned}$$

Aufgabe 6.4 Die Dichte der Zufallsvariablen X sei $f(x)$. Sei $Y = X^2$, dann bestimme man die Verteilungsfunktion $F(x, y)$ des Vektors (X, Y) .

Aufgabe 6.5 Der Zufallsvektor (X, Y) habe die Wahrscheinlichkeitsdichte

$$f(x, y) = \frac{A}{\pi^2(16 + x^2)(25 + y^2)}.$$

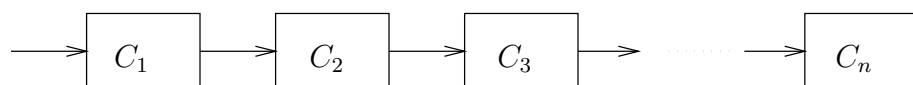
Man bestimme:

- die Größe A ;
- die Verteilungsfunktion $F(x, y)$.

Aufgabe 6.6 Der Zufallsvektor (X, Y, Z) ist innerhalb der Kugel S vom Radius r gleichverteilt. Man bestimme die Verteilungsdichte $f(x, y, z)$ des Vektors und die Verteilungsdichten $f_1(x), f_2(y), f_3(z)$ der Komponenten. Weiter bestimme man $E(\sqrt{x^2 + y^2 + z^2})$.

Aufgabe 6.7

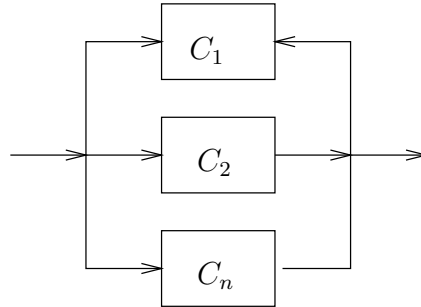
- n -Komponenten C_1, \dots, C_n eines Systems seien in Reihe geschaltet.



Es sei X_i die Lebensdauer Zufallsvariable der i -ten Komponente. Beschreibe die Lebensdauer X des Gesamtsystems.

- b) Unter der Annahme der Unabhängigkeit der X_i bestimme X für den Fall der Exponentialverteilung mit Parametern $\lambda_1, \dots, \lambda_n$ bzw. der Gleichverteilung in $[0, a_i]$.

Aufgabe 6.8 Unter denselben Bedingungen wie in Aufgabe 6.7 behandle das Problem, wo die Komponenten parallel geschaltet sind.



Aufgabe 6.9 Die Lebensdauer eines Stromaggregats ist exponentialverteilt mit Erwartungswert $150d$. Wie groß ist die Wahrscheinlichkeit, daß die Arktisforschungsgruppe mit 2 Aggregaten mindestens $200d$ operationsfähig ist.

6.3 Unabhängige σ -Algebren

Es sei (Ω, Σ, P) ein fest gewählter Wahrscheinlichkeitsraum und $(\sum_i)_{i \in I}$ eine Familie von Unter- σ -Algebren von Σ . Die Familie $(\sum_i)_{i \in I}$ heißen *unabhängig*, wenn für jede endliche Teilmenge $\{i_1, \dots, i_n\} \subset I$ und jede mögliche Auswahl von Mengen $A_{ij} \in \sum_{i_j}$ die Beziehung

$$P(A_{i_1} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n}) \quad (6.16)$$

gilt.

Es ist unmittelbar klar, daß die (\sum_i) genau dann unabhängig sind, wenn jede endliche Teilfamilie der (\sum_i) unabhängig ist.

Dieser Begriff der Unabhängigkeit verallgemeinert den früher eingeführten Begriff der Unabhängigkeit, denn:

Die Mengen $\{A_i\}_{i \in I}$ sind genau dann unabhängig, wenn dies für die $\sum_i = \sum(A_i)$ gilt.

Satz 6.4 Die σ -Algebren $(\sum_i)_{i \in I}$ seien unabhängig und es sei $(I_j)_{j \in J}$ eine disjunkte Zerlegung von I . Dann sind die σ -Algebren $(\sum^j) = \sum(\bigcup_{i \in I_j} \sum_i)_{j \in J}$ unabhängig.

Beweis: Wegen der Bemerkung oben, können wir J als endlich ansehen. Falls die Behauptung für $J = \{1, 2\}$ schon gilt, so haben wir, wenn wir Induktion nach $|J|$ anwenden, für A_1, \dots, A_{n+1} mit $A_i \in \sum^i$:

$$\begin{aligned} P(A_1 \cap \dots \cap A_n \cap A_{n+1}) &= P(A_1 \cap \dots \cap A_n) \cdot P(A_{n+1}), \\ \left(A_1 \cap \dots \cap A_n \in \sum \left(\bigcup_{j=1}^n \bigcup_{i \in I_j} \sum_i \right) \right) &= P(A_1) \cdots P(A_{n+1}) \end{aligned}$$

nach Induktionsannahme.

Also können wir jetzt $|J| = 2$ annehmen.

Es sei nun R_j der von $\sum_i, i \in I_j$ erzeugte Ring. Man sieht leicht, daß jedes $A_j \in R_j$ disjunkte Vereinigung von Mengen der Form $B_k^{(j)} = C_{k_1} \cap \dots \cap C_{k_r}$ mit $C_{k_1} \in \sum_{k_i}$ und $k_i \in I_j$ ist. Es gilt aber für $A_j \in R_j, j = 1, 2$,

$$\begin{aligned} P(A_1 \cap A_2) &= P((\cup_k B_k^{(1)}) \cap (\cup_l B_l^{(2)})) \\ &= \sum_{k,l} P(B_k^{(1)} \cap B_l^{(2)}) = \sum_{k,l} P(C_{k_1} \cap \dots \cap C_{k_r} \cap C_{l_1} \cap \dots \cap C_{l_s}) \\ &= \sum_{k,l} P(C_{k_1}) \cdots P(C_{k_r}) \cdot P(C_{l_1}) \cdots P(C_{l_s}) \end{aligned}$$

wegen der Unabhängigkeit der \sum_i . Dies wiederum gibt

$$\begin{aligned} &= \sum_{k,l} P(C_{k_1} \cap \dots \cap C_{k_r}) \cdot P(C_{l_1} \cap \dots \cap C_{l_s}) = \sum_{k,l} P(B_k^{(1)}) P(B_l^{(2)}) \\ &= P(\cup_k B_k^{(1)}) P(\cup_l B_l^{(2)}) = P(A_1) \cdot P(A_2). \end{aligned}$$

$B \rightarrow P_1(B) = P(A \cap B)$ ist ein \emptyset -stetiger Inhalt auf R_2

$B \rightarrow P_2(B) = P(A)P(B)$ ist ein \emptyset -stetiger Inhalt auf R_2

P_1 und P_2 lassen sich also auf $\sum(R_2) = \sum^{(2)}$ eindeutig fortsetzen. Da $P_1 = P_2$ auf R_2 , so gilt danach $P_1 = P_2$ auf $\sum^{(2)}$.

Es sei nun $A \in R_1$ fest und S_A das System aller $B \in \sum$, so daß $P(A \cap B) = P(A)P(B)$. Wegen der Stetigkeit von P ist S_A abgeschlossen unter monotonen Limiten. Ferner ist S_A abgeschlossen unter Komplementbildung und es gilt $R_2 \subset S_A$. Daher ist mit $A_i, B_j \in R_2, i, j \in \mathbb{N}$ auch $\bigcap_{i=1}^{\infty} A_i, \bigcup_{i=1}^{\infty} A_i$ und $\bigcap_{i=1}^{\infty} A_i \cup \bigcap_{j=1}^{\infty} B_j \in S_A$.

Das System \tilde{R}_2 aller solcher Mengen bildet offensichtlich eine σ -durchschnittsstabile Algebra. Offensichtlich ist $\tilde{R}_2 = \sum(R_2) \subset S_A$.

Nun ersetze oben R_2 durch $\sum(R_2)$ und vertausche die Rollen von R_2 und R_1 .

Beispiel 6.9 Es sei $(\Omega, \sum, P) = (\Omega_1, \sum_1, P_1) \otimes (\Omega_2, \sum_2, P_2)$ und $A = A_1 \times \Omega_2$ und $B = \Omega_1 \times B_2$ mit $A_1 \in \sum_1, B_2 \in \sum_2$. Dann sind A und B unabhängig.

Definition 6.3 Es seien $(\sum_i)_{i \in \mathbb{N}}$ σ -Unteralgebren von \sum . Dann ist $\sum^{(n)} = \sum(\bigcup_{k=n}^{\infty} \sum_k)$ eine σ -Algebra. Gleichfalls ist daher auch

$$\sum_{\infty} = \bigcap_n \sum^{(n)}$$

eine σ -Algebra. \sum_{∞} ist die *terminale σ -Algebra* der $(\sum_i)_{i \in \mathbb{N}}$ und die Elemente von \sum_{∞} heißen *terminale Ereignisse*.

Satz 6.5 (*Null-Eins-Gesetz von Kolmogoroff*) Es sei $(\sum_i)_{i \in \mathbb{N}}$ eine Folge unabhängiger σ -Unteralgebren von \sum und es sei \sum_∞ die terminale σ -Algebra der (\sum_i) . Dann gilt für jedes $A \in \sum_\infty$

$$P(A) = 0 \quad \text{oder} \quad P(A) = 1. \quad (6.17)$$

Beweis: Wegen Satz 6.4 ist jedes $A \in \sum^{(n+1)} = \sum(\bigcup_{k=n+1}^{\infty} \sum_k)$ unabhängig von jedem \sum_i mit $i \leq n$. Daher ist jedes $A \in \sum_\infty$ unabhängig von jedem $\sum_n, n \in \mathbb{N}$. Wiederum wegen Satz 6.4 ist daher A unabhängig von $\sum(\bigcup_{n=1}^{\infty} \sum_n) = \sum^{(1)}$. Offensichtlich ist $\sum_\infty \subset \sum^{(1)}$ und wegen (??) gilt daher $P(A \cap A) = P(A) = P(A)^2$, also $P(A)$ gleich 0 oder 1.

Wir wollen aus Kolmogoroffs Null-Eins-Gesetz noch einige Folgerungen herleiten. Zuvor aber wollen wir noch das *Borel-Cantelli Lemma* beweisen.

Lemma 6.1 Es sei $A_1, A_2, \dots \in \sum$ mit $\sum_{i=1}^{\infty} P(A_i) < \infty$ und $A_\infty = \overline{\lim} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$. Dann gilt $P(A_\infty) = 0$.

Beweis: Es ist A_∞ die Menge aller $\omega \in \Omega$, die in unendlich vielen A_n liegen. Trivialerweise gilt dann $A_\infty \subset \bigcup_{m=n}^{\infty} A_n \forall m \in \mathbb{N}$ oder $P(A_\infty) \leq P(\bigcup_{m=n}^{\infty} A_n) \leq \sum_{m=n}^{\infty} P(A_n) \rightarrow 0$ für $m \rightarrow \infty$.

Die Umkehrung diese Lemmas gilt nicht. Dazu betrachte $A_n = (0, \frac{1}{n}) \subset [0, 1]$. Dann ist $A_\infty = \emptyset$ aber $\sum P(A_n) = \sum \frac{1}{n} = \infty$. Für unabhängige (A_n) aber gilt die Umkehrung.

Korollar 6.1 Es seien $A_1, A_2, \dots \in \sum$ unabhängige Ereignisse und $A_\infty = \overline{\lim} A_n$. Dann gilt

- i) $P(A_\infty) = 0 \Leftrightarrow \sum P(A_n) < \infty$
- ii) $P(A_\infty) = 1 \Leftrightarrow \sum P(A_n) = \infty$

Beweis: Es sei $\sum_i = \{A_i, A_i^c, \emptyset, \Omega\}$. Dann ist $\{\sum_i\}_{i \in \mathbb{N}}$ eine Familie unabhängiger σ -Algebren. Ferner ist $A^{(n)} = \bigcup_{i=n}^{\infty} A_i \in \sum^{(m)} = \sum(\bigcup_{i=m}^{\infty} \sum_i)$ für $n \geq m$. Also ist wegen $A^{(n)} \searrow$ auch $A_\infty = \bigcap_{n \geq m} A^{(n)} \in \sum^{(m)}$ für alle m . Daher ist $A_\infty \in \sum_\infty$ und wegen des Kolmogoroffschen Null-Eins-Gesetzes gilt $P(A_\infty) = 0$ oder 1. Es genügt nun offensichtlich, zu zeigen, daß aus $\sum P(A_n) = \infty$ die Gleichheit $P(A_\infty) = 1$ folgt. Aus der Definition von A_∞ folgt sofort

$$\begin{aligned} P(A_\infty) &= \lim_{n \rightarrow \infty} P(\bigcup_{i=n}^{\infty} A_i) \quad \text{oder} \quad 1 - P(A_\infty) = \lim_{n \rightarrow \infty} P((\bigcup_{i=n}^{\infty} A_i)^c) \\ &= \lim_{n \rightarrow \infty} P(\bigcap_{i=n}^{\infty} A_i^c) = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \prod_{i=n}^{n+m} P(A_i^c) = \lim_{n \rightarrow \infty} \prod_{i=n}^{\infty} (1 - P(A_i)) \end{aligned}$$

Da aber $\sum P(A_n) = \infty$ muß $\prod_{i=n}^{\infty} (1 - P(A_i)) \rightarrow 0$ oder $P(A_\infty) = 1$.

Korollar 6.2 Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen. Dann gilt für die Menge

$$C = \left\{ \omega \in \Omega \mid \sum_{n=1}^{\infty} X_n(\omega) \text{ konvergiert} \right\}$$

- i) $C \in \Sigma$ und
- ii) $P(C) = 0, 1$.

Beweis: Es sei $C_n = \{\omega \in \Omega \mid \sum_{n=N}^{\infty} X_n(\omega)\}$ konvergiert. Dann gilt $C_N = C$ und

$$C_n = \bigcap_{n=1}^{\infty} \bigcup_{m=N}^{\infty} \bigcap_{k=m}^{\infty} \bigcap_{l=k}^{\infty} \left[|X_k + \dots + X_l| < \frac{1}{n} \right] \subset \sum^{(N)} = \sum(X_N, X_{N+1}, \dots).$$

Also gilt $C \in \sum_{\infty}$ und wegen des Null-Eins-Gesetzes folgt die Behauptung.

Aufgabe 6.10 Es sei $(\sum_i)_{i \in I}$ eine unendliche Familie von unabhängigen σ -Algebren und F das System aller endlichen Teilmengen von I . Dann ist $\sum_{\infty} = \bigcup_{J \in F} \sum(\sum_i : i \notin J)$ eine σ -Algebra, die terminale σ -Algebra. Ferner gilt für jedes $A \in \sum_{\infty}$, $P(A) = 0, 1$.

Aufgabe 6.11 Auf dem Wahrscheinlichkeitsraum $([0, 1], B([0, 1]), \lambda)$ betrachte die σ -Algebra \sum_n , die von $\{[0, \frac{1}{2} + \frac{1}{2^n}], [\frac{1}{2} + \frac{1}{2^n}, 1]\}$ erzeugt werden. Bestimme \sum_{∞} .

6.4 Unabhängige Zufallsvariablen

Definition 6.4 Eine Familie von Zufallsvariablen $(X_i)_{i \in I}$ heißt *unabhängig*, wenn die σ -Algebren $\sum(X_i) = X_i^{-1}(B_1)$ unabhängig sind.

Die $(X_i)_{i \in I}$ sind offensichtlich genau dann unabhängig, wenn für jede Auswahl von endlich vielen Indizes $i_1, \dots, i_n \in I$ und jede Wahl von Borelmengen $A_1, \dots, A_n \in B_1$ gilt

$$P(X_{i_1}^{-1}A_1 \cap X_{i_2}^{-1}A_2 \cap \dots \cap X_{i_n}^{-1}A_n) = P(X_{i_1}^{-1}A_1) \cdots P(X_{i_n}^{-1}A_n). \quad (6.18)$$

Es sei $(F_i)_{i \in I}$ eine Familie von Borelfunktionen. Da $\sum(F_i \circ X_i) \subset \sum(X_i)$, folgt aus der Unabhängigkeit der (X_i) sofort die Unabhängigkeit der $(F_i \circ X_i)_{i \in I}$.

Es seien $(X_i)_{i \in I}$ unabhängige Zufallsvariable und $(I_j)_{j \in J}$ eine disjunkte Zerlegung von I . Ferner seien $F_j: B_{n_j} - B_1$ messbare Funktionen mit $|n_j| \leq |I_j|$. Dann sind auch die Zufallsvariablen $(F_j(X_{i_1(j)}, \dots, X_{i_n(j)}))$ unabhängig, wo $i_k(j) \in I_j$. Dies folgt unmittelbar aus Satz (6.4) über das Zusammensetzen unabhängiger Zufallsvariable.

Satz 6.6 Es sei X_1, \dots, X_n ein System unabhängiger Zufallsvariablen und F_1, \dots, F_n eine Familie von Borelfunktionen, so daß $F_i \circ X_i$ P -integrierbar ist. Dann ist $(F_i \circ X_i)$ P -integrierbar und es gilt

$$\int (F_i \circ X_i \cdots F_n \circ X_n) dP = \left(\int F_1 \circ X_1 dP \right) \cdots \left(\int F_n \circ X_n dP \right). \quad (6.19)$$

Beweis: Durch Induktion nach n führt man den Beweis auf den Fall $n = 2$ zurück. Wegen der Linearität des Integrals können wir uns sogar auf positive Funktionen beschränken. Aufgrund der Stetigkeitseigenschaften des Integrals genügt es ferner, die Behauptung für Treppenfunktionen F_i zu beweisen. Die Linearität des Integrals führt das Problem schließlich auf den Fall $F_1 = \chi_A$ und $F_2 = \chi_B$, mit $A, B \in \mathcal{B}_1$ zurück. Dann aber haben wir

$$\begin{aligned} \int (\chi_A \circ X_1, \chi_B \circ X_2) dP &= \int \chi_{X_1^{-1}A} \cdot \chi_{X_2^{-1}B} dP \\ &= P(X_1^{-1}A \cap X_2^{-1}B) \\ &= P(X_1^{-1}A)P(X_2^{-1}B) \quad \text{wegen (??)} \\ &= \left(\int \chi_A \circ X_1 dP \right) \left(\int \chi_B \circ X_2 dP \right). \end{aligned}$$

Korollar 6.3 Es seien X, Y unabhängige P -integrierbare Zufallsvariablen. Dann gilt

$$E(XY) = E(X)E(Y) \quad (6.20)$$

$$\sigma_{x,y} = E((X - EX)(Y - EY)) = 0 \quad (6.21)$$

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y). \quad (6.22)$$

Beweis: (??) folgt unmittelbar aus (??). Gleichfalls wegen (??) gilt $\sigma_{x,y} = E(X - EX)(Y - EY) = 0$. Ferner haben wir:

$$\begin{aligned} \sigma^2(X + Y) &= E((X + Y)^2 - E(X + Y))^2 = EX^2 + 2E(XY) + EY^2 \\ &\quad - (EX)^2 - E(Y)^2 - 2(EX)(EY) \\ &= EX^2 - (EX)^2 + (E^2) - (EY^2). \end{aligned}$$

Allgemeiner gilt für unabhängige Zufallsvariablen X_1, \dots, X_n

$$\sigma^2(X_1 + \dots + X_n) = \sigma^2(X_1) + \dots + \sigma^2(X_n). \quad (6.23)$$

Man zeigt dies durch Induktion nach n unter Verwendung von (??).

Beispiel 6.10 Ein Punkt sei gleichmäßig auf dem Einheitskreis Ω der Ebene verteilt. Sind seine beiden Koordinaten unabhängig? Da (X, Y) die Identität auf Ω ist, ist die gemeinsame Verteilung von (X, Y) die gleichmäßige Verteilung auf Ω . Es sei nun $A = [0, a]$ und $B = [0, b]$. Wenn X und Y unabhängig sind, so gilt wegen (??)

$$\frac{1}{\pi} \int_{\Omega} \chi_{[0,a] \times [0,b]} dx dy = \frac{1}{\pi} \int_{\Omega} \chi_{[0,a]}(x) dx dy \cdot \frac{1}{\pi} \int_{\Omega} \chi_{[0,b]}(y) dx dy$$

oder

$$\lambda(\Omega \cap [0, a] \times [0, b]) \frac{1}{\pi} F_a F_b.$$

Mit $a = \sin \phi$ wird

$$F_a = 2\phi \frac{1}{\pi^2} \pi + \cos \phi \sin \phi$$

$$a = \frac{1}{\sqrt{2}}.$$

Setzen wir $b = \sin \psi$, so erhalten wir schließlich

$$\lambda(\Omega \cap [0, \sin \phi] \times [0, \sin \psi]) = \frac{1}{\pi} (\phi + \cos \phi \sin \phi) (\psi + \cos \psi \sin \psi).$$

Für $\phi, \psi \leq \frac{\pi}{4}$ aber bedeutet dies

$$\pi \sin \phi \sin \psi = \phi \psi + \phi \cos \psi + \psi \cos \phi \sin \phi + \cos \phi \sin \phi \sin \psi \cos \psi.$$

Dies aber ist sicher unmöglich. Also sind X und Y nicht unabhängig, was übrigens auch klar ist, denn wenn X -Werte nahe bei 1 annimmt, so muß Y -Werte nahe bei 0 annehmen. Dennoch aber sind X und Y unkorreliert, denn

$$E(X) = 0 = E(Y). \text{ Ferner ist } \frac{1}{\pi} \int_{\Omega} xy dx dy = 0 = \sigma_{xy}.$$

Dies zeigt, daß Unkorreliertheit nicht die Unabhängigkeit zur Folge hat.

Satz 6.7 Es seien X_1, \dots, X_n Zufallsvariablen und $X = (X_1, \dots, X_n)$ der zugehörige Zufallsvektor. Die Zufallsvariable X_1, \dots, X_n sind genau dann unabhängig, wenn

$$XP = X_1P \otimes \dots \otimes X_nP. \quad (6.24)$$

Beweis: Sowohl XP als auch $X_1P \otimes \dots \otimes X_nP$ sind Wahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathcal{B}_n)$. Wegen des Fortsetzungssatzes sind diese Maße durch ihre Werte auf dem Ring der n -dimensionalen Figuren R_n eindeutig festgelegt, d.h. wenn beide Maße auf Mengen der Form $A_1 \times \dots \times A_n, A_i \in \mathcal{B}_1$ die gleichen Werte annehmen, so sind sie gleich. Es gilt aber

$$\begin{aligned} XP(A_1 \times \dots \times A_n) &= P(X_1^{-1}A_1 \cap \dots \cap X_n^{-1}A_n) = X_1P(A_1) \cdots X_nP(A_n) \\ &= (X_1P) \otimes \dots \otimes (X_nP)(A_1 \times \dots \times A_n) \end{aligned}$$

wegen (??), wenn die X_i unabhängig sind. Die umgekehrte Behauptung wird entsprechend bewiesen.

Korollar 6.4 Die X_1, \dots, X_n sind genau dann unabhängig, wenn die gemeinsame Verteilungsfunktion F_x von $X = (X_1, \dots, X_n)$ das Produkt der Einzelverteilungsfunktion ist

$$F_x = (x_1, \dots, x_n) = F_{x_1}(x_1) \cdots F_{x_n}(x_n). \quad (6.25)$$

Beweis: Die (X_1, \dots, X_n) seien unabhängig. Dann gilt wegen (??) $F_x(x_1, \dots, x_n) = XP((-\infty, x_1) \times \dots \times (-\infty, x_n)) = X_1P((-\infty, x_1)) \cdots X_nP((-\infty, x_n)) = F_{x_1}(x_1) \cdots F_{x_n}(x_n)$. Die Umkehrung folgt analog.

Korollar 6.5 Es seien X_1, \dots, X_n unabhängige stetig verteilte Zufallsvariablen mit der Dichte f_i für $i = 1, \dots, n$. Dann ist $X = (X_1, \dots, X_n)$ stetig verteilt mit der Dichte

$$f_x(x_1, \dots, x_n) = f_{x_1}(x_1) \cdots f_{x_n}(x_n). \quad (6.26)$$

Die Umkehrung gilt ebenfalls.

Beweis: Dieses Korollar folgt unmittelbar aus (??).

In Anwendungen kommt es immer wieder vor, daß man die Verteilung von Funktionen von Zufallsvariablen bestimmen muß. Am einfachsten läßt sich dies mit Hilfe der Verteilungsfunktion bewerkstelligen. Es seien (X_1, \dots, X_n) Zufallsvariablen und H eine $\mathcal{B}_n - \mathcal{B}_1$ messbare Funktion. Ferner sei G eine $\mathcal{B}_1 - \mathcal{B}_1$ messbare Funktion. Wegen Satz (5.1) gilt dann

$$\int G \circ H \circ X dP = \int G \circ H(x_1, \dots, x_n) dXP. \quad (6.27)$$

Wählen wir hier $G = \chi_{(-\infty, x)}$, so erhalten wir

$$F_{H(X_1, \dots, X_n)}(x) = \int \chi_{(-\infty, x)} \circ H(x_1, \dots, x_n) dXP = \int_{H(x_1, \dots, x_n) < x} dXP. \quad (6.28)$$

Bei stetigen Verteilungen können wir dann die Dichte f_H von $H(X_1, \dots, X_n)$ bestimmen, falls diese existiert, indem wir $f_H = F'_H$ benutzen.

Ist $H \circ X$ diskret verteilt, so folgt aus (??) mit $G = \chi_{\{x\}}$

$$P([H(X_1, \dots, X_n) = x]) = HP(\{x\}) = \int_{H(x_1, \dots, x_n) = x} dXP. \quad (6.29)$$

Sind hierbei die X_1, \dots, X_n noch unabhängig, so können wir in (??), (??) und (??) XP durch $X_1P_1 \otimes \dots \otimes X_nP_n$ wegen (??) ersetzen. Als Beispiel wollen wir diese Formel nun auf den Fall $n = 2$ und $H(X_1, X_2) = X_1 + X_2$ anwenden.

Satz 6.8

a) Es seien μ_1 und μ_2 (endliche) Maße auf \mathcal{B}_n . Dann definiert

$$\mathcal{B}_1 \ni A \rightarrow \nu(A) = \int \chi_A(x_1 + x_2) d\mu_1(x_1) d\mu_2(x_2) \quad (6.30)$$

ein (endliches) Maß auf \mathcal{B}_n . ν wird die *Faltung* von μ_1 und μ_2 genannt und mit $\nu = \mu_1 * \mu_2$ bezeichnet.

- b) Sind μ_1 und μ_2 Wahrscheinlichkeitsmaße, so ist $\mu_1 * \mu_2$ ebenfalls ein Wahrscheinlichkeitsmaß.
- c) Die Faltung ist kommutativ und assoziativ.
- d) Sind X und Y unabhängige Zufallsvariablen auf (Ω, Σ, P) , so ist $X + Y$ nach $XP * YP$ verteilt.

Beweis:

- a) Daß $\nu(A) = \int \chi_A(x_1+x_2)d(\mu_1 \otimes \mu_2)(x_1+x_2)$ ein Maß definiert, folgt unmittelbar aus der Linearität und Stetigkeit des Integrals. Sind μ_1 und μ_2 endlich mit $\mu_1(\mathbb{R}^n) = |\mu_1|$ und $\mu_2(\mathbb{R}^n) = |\mu_2|$, so folgt

$$\mu_1 * \mu_2(\mathbb{R}^n) = \int_{\mathbb{R}^n} d(\mu_1 \otimes \mu_2) = |\mu_1| |\mu_2|.$$

Dies zeigt auch (b).

- c) Es ist

$$\begin{aligned} (\mu_1 * \mu_2)(A) &= \int \chi_A(x_1 + x_2) d\mu_1(x_1) d\mu_2(x_2) \\ &= \int \chi_A(x_2 + x_1) d\mu_2(x_2) d\mu_1(x_1) = (\mu_2 * \mu_1)(A) \end{aligned}$$

wegen des Satzes von Fubini. Ferner folgt aus der Linearität und Stetigkeit des Integrals mit Hilfe von (??)

$$\int f d\mu_1 * \mu_2 = \int f_1(x_1 + x_2) d\mu_1(x_1) d\mu_2(x_2). \quad (6.31)$$

Man zeigt dies in der üblichen Weise zunächst für Treppenfunktionen und dehnt dies Ergebnis dann auf beliebige $d\mu_1 * \mu_2$ integrable Funktionen aus. Wir haben nun für Borelmaße μ_1, μ_2 und μ_3

$$[(\mu_1 * \mu_2) * \mu_3](A) = \int \chi_A(z_1 + x_3) d(\mu_1 * \mu_2)(z_1) d\mu_3(x_3).$$

Dies ist wegen

$$\begin{aligned} &= \int \chi_A((x_1 + x_2)x_3) d\mu_1(x_1) d\mu_2(x_2) d\mu_3(x_3) \\ &= \int \chi_A(x_1 + (x_2 + x_3)) d\mu_1(x_1) d\mu_2(x_2) d\mu_3(x_3) \\ &= \dots = [\mu_1 * (\mu_2 * \mu_3)](A). \end{aligned} \quad (6.32)$$

- d) Wegen Satz 6.7 gilt

$$\begin{aligned} (X + Y)P(A) &= \int \chi_A(X + Y) dP = \int \chi_A(x + y) d(X, Y)P \\ &= \int \chi_A(x + y) dXP(x) dYP(y) = (XP * YP)(A). \end{aligned}$$

Wir wollen nun explizitere Formeln für die Faltung bestimmen.

Korollar 6.6

- a) Ist $\mu_1 = \sum_i b_i \delta_{a_i}$ und $\mu_2 = \sum_j d_j \delta_{c_j}$, wobei a_i alle verschieden, c_j alle verschieden $b_i \geq 0$, $d_j \geq 0$, $\sum b_i < \infty$ und $\sum d_j < \infty$, so gilt

$$\mu_1 * \mu_2 = \sum_{i,j} b_i d_j \delta_{a_i + c_j}. \quad (6.33)$$

- b) Sind μ_1, μ_2 stetige Maße mit den Dichten f_1, f_2 mit $f_1, f_2 \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}_1, \lambda)$, so gilt

$$\begin{aligned} \mu_1 * \mu_2 \text{ ist stetig mit der Dichte } f_1 * f_2, \text{ wo} \\ f_1 * f_2(y) = \int f_1(x)f_2(y-x)d\lambda. \end{aligned} \quad (6.34)$$

Beweis:

- a) Offensichtlich ist die Faltung linear. Daher genügt es, $\delta_a * \delta_c = \delta_{a+c}$ zu zeigen. Dies aber ist trivial, denn $\int f d\delta_a * \delta_c = \int f(x+y)\delta_a(a)\delta_b(y) = f(a+c)$.
- b) Wegen Satz 6.8 a) ist $f_1 \cdot \lambda * f_2 \cdot \lambda$ ein endliches Maß. Für eine beliebig beschränkte Borelfunktion g gilt daher

$$\begin{aligned} \int g d(f_1 \cdot \lambda) * (f_2 \cdot \lambda) &= \int g(x+y)f_1(x)f_2(y)d\lambda_2(x,y) \\ &= \int \left(\int g(x+y)f_2(y)d\lambda(y) \right) f_1(x)d\lambda(x) \\ &= \int \left(\int g(z)f_2(z-x)d\lambda(z) \right) f_1(x)d\lambda(x). \end{aligned}$$

Hier haben wir die Translationsinvarianz von λ verwendet, d.h.

$$\begin{aligned} \lambda(A) &= \lambda(A+x) \forall A \in \mathcal{B}_1 \\ &= \int g(z) \left(\int f_2(z-x) \right) f_1(x)d\lambda(x)d\lambda(z). \end{aligned}$$

Dies aber zeigt das Ergebnis.

Beispiel 6.11 Es seien X und Y unabhängige Zufallsvariablen und beide seien binomial verteilt mit Parametern n, p bzw. m, p . Dann ist $X + Y$ binomial verteilt mit Parametern $m + n$ und p .

Beweis: Es ist

$$\begin{aligned} XP * YP &= \sum_{k=0}^n \sum_{l=0}^m \binom{n}{k} \binom{m}{l} p^k q^{n-k} p^l q^{m-l} \delta_{k+l} \\ &= \sum_{i=0}^{m+n} \left(\sum_{k+l=i} \binom{n}{k} \binom{m}{l} \right) p^i q^{m+n-i} \delta_i \\ &= \sum_{i=0}^{m+n} \binom{m+n}{i} p^i q^{m+n-i} \delta_i \end{aligned}$$

weil

$$\sum_{k=0}^n \binom{n}{k} \binom{m}{i-k} = \binom{m+n}{i}.$$

Beispiel 6.12 Es seien X und Y unabhängige Poisson verteilte Zufallsvariablen mit Parametern a bzw. b . Dann ist $X + Y$ ebenfalls Poisson verteilt mit Parameter $a + b$.

Beweis: Es ist

$$\begin{aligned} XP * YP &= \sum_{k,l=0}^{\infty} e^{-a} \frac{a^k}{k!} e^{-b} \frac{b^l}{l!} \delta_{l+k} \\ &= \sum_{i=0}^{\infty} e^{-(a+b)} \left(\sum_{k+l=i} \frac{a^k}{k!} \frac{b^l}{l!} \delta_i \right) \\ &= \sum_{i=0}^{\infty} e^{-(a+b)} \frac{(a+b)^i}{i!} \delta_i, \end{aligned}$$

denn

$$\begin{aligned} \sum_{i=l+k} \frac{a^k b^l}{k! l!} &= \sum_{k=0}^i \frac{a^k}{k!} \frac{b^{i-k}}{(i-k)!} \frac{i!}{i!} \\ &= \sum_{k=0}^i \binom{i}{k} a^k b^{i-k} \frac{1}{i!} = (a+b)^i \frac{1}{i!}. \end{aligned}$$

Beispiel 6.13 Es seien X und Y unabhängige normalverteilte Zufallsvariablen mit Parametern μ_1, σ_1^2 bzw. μ_2, σ_2^2 . Dann ist $X + Y$ normalverteilt mit Parametern $\mu_1 + \mu_2$ und $\sigma^2 = \sigma_1^2 + \sigma_2^2$.

Beweis: Wir haben

$$\begin{aligned} &\frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \int e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} e^{-\frac{(y-x-\mu_2)^2}{2\sigma_2^2}} d\lambda(x) \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(y-\mu_1-\mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}} \cdot I \end{aligned}$$

wo mit $x - \mu_1 = u$ und $y - \mu_1 - \mu_2 = v$

$$\begin{aligned} I &= \frac{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}{2\pi\sigma_1\sigma_2} \int e^{-\frac{u^2}{2\sigma_1^2}} e^{-\frac{(v-u)^2}{2\sigma_2^2}} e^{-\frac{v^2}{2(\sigma_1^2 + \sigma_2^2)}} du \\ &= \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2\pi\sigma_1^2 + \sigma_2^2}} \int \exp - \left\{ \frac{u^2(\sigma_2^2 + \sigma_1^2)}{2\sigma_1^2 + \sigma_2^2} - \frac{2uv}{2\sigma_2^2} - \frac{v^2(\sigma_2^2 - \sigma_1^2 - \sigma_2^2)}{2(\sigma_1^2 + \sigma_2^2)\sigma_2^2} \right\} \\ &= \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2\pi\sigma_1^2 + \sigma_2^2}} \int \exp - \left\{ \frac{\sigma_2^2 + \sigma_1^2}{2\sigma_1^2\sigma_2^2} \left[u^2 - 2\frac{uv\sigma_2^2}{\sigma_1^2\sigma_2^2} + \frac{v^2\sigma_1^4}{2(\sigma_1^2 + \sigma_2^2)^2} \right] \right\} = 1 \end{aligned}$$

Beispiel 6.14 Es seien 2 Stück radioaktiven Materials gegeben, die radioaktive Strahlen mit der Intensität (Zerfälle pro Zeiteinheit) a_1 bzw. a_2 aussenden. Dann registriert ein Geigenzähler die Verteilung von $X + Z$ wo X und Y die Verteilungen der Einzelquellen sind. Da $X(Y)$ Poisson verteilt ist, mit Parameter $a_1(a_2)$, ist $X + Z$ auch Poisson verteilt mit Parameter $a_1 + a_2$. Dies ist natürlich unmittelbar einsichtig.

6.5 Die Multivariate Normalverteilung

Sind zwei oder mehrere Zufallsvariablen nicht unabhängig, wird man ihre gemeinsame Verteilung, d.h. also eine Verteilung im \mathbb{R}^n untersuchen. Die weitaus wichtigste dieser Verteilungen ist die Verallgemeinerung der eindimensionalen Normalverteilung. Um diese Verteilung einzuführen, benötigen wir ein paar Grundbegriffe der linearen Algebra.

Im \mathbb{R}^n führen wir durch $\langle x, x' \rangle = \sum_{i=1}^n x_i x'_i$ ein Skalarprodukt ein. Ist $A = (a_{ij})$ eine

Matrix so definiert $\langle Ax, x' \rangle = \sum_{i,j=1}^n a_{ij} x_i x'_j$ eine quadratische Form, auf dem \mathbb{R}^n .

Eine solche Form heißt positiv definit, wenn für alle $x \in \mathbb{R}^n$ $\langle Ax, x \rangle \in \mathbb{R}$ ist.

$$(\langle Ax, x \rangle \geq \varepsilon \langle x, x \rangle = \varepsilon |x|^2 \text{ für ein } \varepsilon > 0).$$

Lemma 6.2 Ist A hermitesch, so existiert eine Drehung (des Raumes $x \rightarrow Dx$), so daß DAD^{-1} Diagonalgestalt hat.

Ist A positiv definit, so hat DAD^{-1} die Form $\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$ mit $a_i > 0$.

Definition 6.5 Die Dichtefunktion einer n -dimensionalen Normalverteilung hat die Form

$$f(x_1, \dots, x_n) = K e^{-\frac{1}{2} \langle A(x-b), (x-b) \rangle}$$

wo K eine Normierungskonstante, A positiv definite Matrix und $b \in \mathbb{R}^n$ ein Vektor ist.

Um K zu bestimmen, berechnen wir zunächst allgemeiner

$$I = \int g(x_1, \dots, x_n) f(x_1, \dots, x_n) d\lambda_n = K \int g(x) e^{-\frac{1}{2} \langle A(x-b), (x-b) \rangle} d\lambda_n.$$

Wegen der Translationsinvarianz von λ_n gibt dies

$$\begin{aligned} I &= K \int g(z+b) e^{-\frac{1}{2} \langle Az, z \rangle} d\lambda_n \\ &= K \int g(z+b) e^{-\frac{1}{2} \langle D^{-1} D A D^{-1} D z, z \rangle} d\lambda_n \\ &= K \int g(z+b) e^{-\frac{1}{2} \langle D A D^{-1} (D z), (D z) \rangle} d\lambda_n(z) \end{aligned}$$

Da eine Drehung volumentreu ist, erhalten wir mit $Dz = w$

$$I = K \int g(D^{-1}w + b) e^{\frac{1}{2} \langle (D A D^{-1}) w, w \rangle} d\lambda_n(w).$$

Nehmen wir nun an, D sei so gewählt, daß

$$D A D^{-1} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}, \quad a_i > 0$$

Dann ist $\langle DAD^{-1}w, w \rangle = \sum a_i w_i^2 = \sum (a_i^{\frac{1}{2}} w_i)^2$.

Setzen wir daher $Sw = (a_i^{\frac{1}{2}} w_i) = u$, so erhalten wir

$$\begin{aligned}
 I &= K \int g(D^{-1}w + b) e^{\frac{1}{2}\langle Sw, Sw \rangle} d\lambda_n(w) \\
 &= \frac{K}{\prod a_i^{\frac{1}{2}}} \int g(D^{-1}S^{-1}u + b) e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u) \\
 &= (\text{Det}S)^{-1} K \int g(D^{-1}S^{-1}u + b) e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u) \\
 &= (\text{Det}DAD^{-1})^{-\frac{1}{2}} K \int g(D^{-1}S^{-1}u + b) e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u) \\
 &= (\text{Det}A)^{-\frac{1}{2}} K \int g(D^{-1}S^{-1}u + b) e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u)
 \end{aligned}$$

Insgesamt also haben wir

$$\begin{aligned}
 &K \int g(x) e^{-\frac{1}{2}\langle A(x-b), (x-b) \rangle} d\lambda_n(x) \\
 &= (\det A)^{-\frac{1}{2}} K \int g(D^{-1}S^{-1}u + b) e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u)
 \end{aligned} \tag{6.35}$$

Für $g = 1$ erhalten wir daraus mit der Normierungsbedingung

$$1 = (\det A)^{-\frac{1}{2}} K \int e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u) = K (\det A)^{-\frac{1}{2}} (2\pi)^{\frac{11}{2}}$$

oder

$$K = (\det A)^{\frac{1}{2}} (2\pi)^{-\frac{11}{2}}. \tag{6.36}$$

Die Dichtefunktion der n -dimensionalen Normalverteilung ist damit

$$f(x) = (\det A)^{\frac{1}{2}} (2\pi)^{-\frac{n}{2}} e^{-\langle A(x-b), (x-b) \rangle \frac{1}{2}}. \tag{6.37}$$

Mit Hilfe von (??) können wir jetzt die Erwartungswerte der einzelnen Koordinaten bestimmen.

$$\begin{aligned}
 E(X_i) &= (2\pi)^{-\frac{n}{2}} \int (D^{-1}S^{-1}u + b)_i e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u) \\
 &= (2\pi)^{-\frac{n}{2}} \int b_i e^{-\frac{1}{2}\langle u, u \rangle} d\lambda_n(u) = b_i.
 \end{aligned} \tag{6.38}$$

Schreiben wir als $X = (X_1, \dots, X_n)$, so sehen wir

$$E(X) = b. \tag{6.39}$$

Entsprechend sehen wir für die Kovarianzmatrix

$$\begin{aligned}
E((X_i - b_i)(X_j - b_j)) &= K \int (x - b)_i (x - b)_j e^{-\frac{1}{2} \langle A(x-b), (x-b) \rangle} d\lambda_n(x) \\
&= K \int x_i x_j e^{-\frac{1}{2} \langle Ax, x \rangle} d\lambda_n(x) \\
&= K (\det A)^{-\frac{1}{2}} \int (D^{-1}S^{-1}u)_j e^{-\frac{1}{2} \langle u, u \rangle} (D^{-1}S^{-1}n)_i d\lambda_n(u) \\
&= (2\pi)^{-\frac{n}{2}} \sum_{k,l} (D^{-1}S^{-1})_{ik} (D^{-1}S^{-1})_{jl} \int u_k u_l e^{-\frac{1}{2} \langle u, u \rangle} d\lambda_n(u) \\
&= \sum_{k=1}^n (D^{-1}S^{-1})_{ik} (D^{-1}S^{-1})_{jk} = (D^{-1}S^{-1}S^{-1}D)_{i,j} = (A_{i,j}^{-1})
\end{aligned}$$

Bezeichnet daher C die Kovarianzmatrix, so schreibt man für die Dichte der multivariaten Normalverteilung auch

$$f(x) = (2\pi)^{-\frac{n}{2}} (\det C)^{-\frac{1}{2}} e^{-\frac{1}{2} \langle C^{-1}(x-b), (x-b) \rangle} \quad (6.40)$$

und bezeichnet sie mit $N(b, C)$.

Satz 6.9 Es sei $X = (X_1, \dots, X_n)$ $N(b, C)$ verteilt. Dann ist $BX = Y$ $N(Bb, BCB^t)$ verteilt, B invertierbar.

Beweis: Die Dichte von Y erhält man, wenn man x durch $B^{-1}y$ ersetzt. Für die quadratische Form im Exponenten gilt dann

$$\begin{aligned}
\langle C^{-1}B^{-1}(y - Bb), B^{-1}(y - Bb) \rangle &= \langle (B^+)^{-1}C^{-1}B^{-1}(y - Bb), y - Bb \rangle \\
&= \langle (BCB^t)^{-1}(y - Bb), (y - Bb) \rangle
\end{aligned}$$

Bei dieser linearen Transformation $x = B^{-1}y$ ist das $d\lambda_n(x)$ Volumenelement gleich dem $|\det B|^{-1}d\lambda_n(y)$ Volumenelement. Nun ist aber $(\det C)^{-\frac{1}{2}}|\det B|^{-1} = (\det BCB^t)^{-\frac{1}{2}}$, so daß die Behauptung bewiesen ist.

Satz 6.10 Es sei X $N(b, C)$ verteilt und $X^{(1)} = (X_1, \dots, X_m)$ und $X^{(2)} = (X_{m+1}, \dots, X_n)$. Dann ist die Randverteilung von $X^{(2)}$ gerade $N(b^{(2)}, C_2)$.

Dabei erhält man $b^{(2)}$ und $C_{2,2}$ als Einschränkung von b bzw. C auf die Koordinaten $m+1, \dots, n$.

Beweis: Wir führen neue unkorrelierte Variable $Y^{(1)}$ und $Y^{(2)}$ durch

$$Y^{(1)} = X^{(1)} + MX^{(2)}, \quad Y^{(2)} = X^{(2)}$$

ein, wobei M so zu bestimmen ist, daß $Y^{(1)}$ und $Y^{(2)}$ verschwindende Kovarianz haben, d.h.

$$\begin{aligned}
0 &= E((Y^{(1)} - EY^{(1)})(Y^{(2)} - EY^{(2)})) = E(((X^{(1)} - b^{(1)}) + M(X^{(2)}))(X^{(2)} - b^{(2)})) \\
&= C_{1,2} + MC_{2,2} \text{ oder } M = -C_{1,2}C_{2,2}^{-1}.
\end{aligned}$$

Nun ist aber

$$Y = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} 1 - C_{1,2}C_{2,2}^{-1} \\ 1 \end{pmatrix} \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

das Bild von X unter einer invertierbaren Abbildung. Also ist Y normalverteilt mit Mittelwert

$$EY = \begin{pmatrix} 1 - C_{1,2}C_{2,2}^{-1} \\ 1 \end{pmatrix} \begin{pmatrix} b^{(1)} \\ b^{(2)} \end{pmatrix} = b'$$

und Kovarianz

$$\begin{aligned} E((Y - b')(Y - b')') &= E \begin{pmatrix} (Y^{(1)} - b'^{(1)})(Y^{(1)} - b'^{(1)})' & (Y^{(2)} - b'^{(2)})(Y^{(1)} - b'^{(1)})' \\ (Y^{(1)} - b'^{(1)})(Y^{(2)} - b'^{(2)})' & (Y^{(2)} - b'^{(2)})(Y^{(2)} - b'^{(2)})' \end{pmatrix} \\ &= \begin{pmatrix} C_{1,1} - C_{1,2}C_{2,2}^{-1}C_{2,1} & 0 \\ 0 & C_{2,2} \end{pmatrix} \end{aligned}$$

da

$$\begin{aligned} E((Y^{(1)} - b'^{(1)})(Y^{(1)} - b'^{(1)})') &= E([(X^{(1)} - b^{(1)} - C_{1,2}C_{2,2}^{-1}(X^{(2)} - b^{(2)}))][(X^{(1)} - b^{(1)}) \\ &\quad - C_{1,2}C_{2,2}^{-1}(X^{(2)} - b^{(2)})']) = C_{1,1} - C_{1,2}C_{2,2}^{-1}C_{2,2}. \end{aligned}$$

Damit sind $Y^{(1)}$ und $Y^{(2)}$ unkorreliert, also unabhängig. Die Randverteilung von $Y^{(2)}$ ist also $N(b^{(2)}C_{2,2})$.

6.6 Der Poisson Prozeß

Die folgenden Überlegungen sind für die Anwendungen, aber auch für die Fortführung der Theorie von großer Wichtigkeit. Wir betrachten eine Folge T_1, T_2, T_3, \dots von unabhängigen, identisch exponentialverteilten Zufallsvariablen mit Parameter λ . Diese T_i können etwa die Zwischenankunftszeiten von Fahrzeugen auf einer Straße oder von Kunden an einem Schalter, die Zeitintervalle zwischen Zerfällen einer Menge radioaktiven Materials oder ähnliches beschreiben. $T = T_1 + \dots + T_n$ beschreibt dann die Wartezeit bis zum Eintreffen n solcher Ereignisse. Genauer gilt also

$$P(T \leq t) = \int_{\substack{t_1+t_2+\dots+t_n \leq t \\ t_i \geq 0}} \dots \int \lambda^n e^{-\lambda t_1} e^{-\lambda t_2} \dots e^{-\lambda t_n} dt_1 \dots dt_n$$

die Wahrscheinlichkeit, daß n Ereignisse in $[0, t]$ stattfinden. Um dieses Integral zu berechnen, führen wir die neuen Variablen $u_1 = t_1, u_2 = t_1 + t_2, \dots, u_n = t_1 + t_2 + \dots + t_n$ ein. Da die entsprechende Funktionaldeterminante 1 ist, haben wir sofort

$$\begin{aligned} P(T \leq t) &= \int \dots \int \lambda^n e^{-\lambda u_n} \\ &\quad 0 \leq du_1 \leq \dots \leq du_n \leq t. \end{aligned}$$

Das Integral über die ersten $(n-1)$ Variablen ist aber nichts anderes, als das Integral über den $(n-1)$ Simplex mit Kantenlänge u_n . Dieser hat das Volumen

$$\frac{1}{(n-1)!} u_n^{n-1}.$$

Wir haben also

$$P(T \leq t) = \int_0^t \frac{\lambda^n}{(n-1)!} u_n^{n-1} e^{-\lambda u_n} du_n$$

d.h. die Dichte von T ist

$$\frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}.$$

Dieses Ergebnis erhält man natürlich auch sofort aus der bekannten Wahrscheinlichkeit, in $[0, t]$ n Ereignisse zu beobachten (Poisson-Verteilung mit Parameter $t \cdot \lambda$)

$$\frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Beschreibt T_i die Lebensdauer einer Maschine, so kann man $T_1 + \dots + T_n$ auch als einen *Erneuerungsprozeß* interpretieren, bei der die Maschinen nacheinander durch neue ersetzt werden. T_i beschreibt dabei die Lebensdauer der i -ten Maschine. Dies läßt sich auch auf andere Verteilungen verallgemeinern. Interessiert man sich oben nicht primär für Zeitpunkte des Eintreffens der Ereignisse, sondern eher für die Anzahl der Ereignisse, so wird man auf eine Familie $X_t, t \geq 0$ von Zufallsvariablen geführt, deren Verteilung durch

$$P(X_t \geq n) = P(T_1 + \dots + T_n \leq t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}$$

gegeben ist. Dabei bedeutet $P(X_t \geq n)$ die Wahrscheinlichkeit, daß in $[0, t]$ mehr als n Ereignisse stattfinden, während $P(T_1 + \dots + T_n \leq t)$ als Wahrscheinlichkeit, daß n Ereignisse vor t stattgefunden haben, zu interpretieren ist. Dies gibt dann

$$\begin{aligned} P(X_t = n) &= P(X_t \geq n) - P(X_t \geq n+1) \\ &= \int_0^t \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} - \int_0^t \frac{\lambda^{n+1}}{n!} t^n e^{-\lambda t} dt. \end{aligned}$$

Partielle Integration des ersten Integrals ergibt

$$\begin{aligned} &= \frac{\lambda^n}{(n-1)!} \left\{ \left[\frac{1}{n} t^n e^{-\lambda t} \right]_0^t + \frac{\lambda}{n} \int_0^t t^n e^{-\lambda t} \right\} - \int_0^t \frac{\lambda^{n+1}}{n!} t^n e^{-\lambda t} dt \\ &= \frac{\lambda^n}{n!} t^n e^{-\lambda t}, \end{aligned}$$

da sich die Integrale aufheben. X_t ist also Poisson verteilt.

Kapitel 7

Gesetze der großen Zahlen

7.1 Konvergenzbegriffe in der Wahrscheinlichkeitstheorie

Wir haben schon einmal kurz mit Konvergenz in der Wahrscheinlichkeitstheorie zu tun gehabt, nämlich bei der Betrachtung des Satzes von Moivre Laplace. In diesem Abschnitt wollen wir die verschiedenen Konvergenzarten in der Wahrscheinlichkeitstheorie untersuchen und ihre Beziehung feststellen.

Definition 7.1 Eine Folge von Zufallsvariablen (X_n) konvergiert *P fast sicher gegen* X , wenn $P([X_n \rightarrow X]) = P(\{\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)\}) = 1$. Wir schreiben dafür $X_n \xrightarrow{f.s.} X$ bzw. $X_n \rightarrow X$ P-f.s.

Satz 7.1 $X_n \xrightarrow{f.s.} X \iff \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} [|X_k - X| \geq \varepsilon]\right) = 0$ für alle $\varepsilon > 0$.

Beweis: Es sei $A_k = [|X_k - X| \geq \varepsilon]$ und $B_n = \bigcup_{k=n}^{\infty} A_k$. Ferner sei

$$C = [X_n \rightarrow X] \text{ und } D_\varepsilon = [\overline{\lim} |X_n - X| \geq \varepsilon].$$

Dann gilt $B_{n+1} \subset B_n$ und daher $B_{n+1} \cap C \subset B_n \cap C$ und $\lim P(C \cap B_n) = P(C \cap \bigcap_n B_n)$.

- Ist nun $X_n \xrightarrow{f.s.} X$, so gilt $P(C) = 1$ und $\bigcap_{n=1}^{\infty} (B_n \cap C) = \emptyset$. Also ist $\lim P(B_n \cap C) = 0$ oder $\lim P(B_n) = 0$.
- Umgekehrt gelte nun $\lim P(B_n) = 0$. Es ist $D_\varepsilon \subset B_n$ für alle n . Also ist $P(D_\varepsilon) = 0$. Daher ist $P(\bigcup_{k=1}^{\infty} D_{\frac{1}{k}}) = 0$. Da aber $C^c \subset \bigcup_{k=1}^{\infty} D_{\frac{1}{k}}$, ist $P(C^c) = 1 - P(C) = 0$ und $X_n \rightarrow X$ f.s..

Definition 7.2 X_n konvergiert *stochastisch gegen* X , wenn für jedes $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1.$$

Wir schreiben dafür $X_n \xrightarrow{p} X$. Man sieht leicht, daß $X_n \xrightarrow{p} X \not\Rightarrow X_n \rightarrow X$ f.s. gilt.

Beispiel 7.1 Jede natürliche Zahl n kann in eindeutiger Weise als $n = 2^{k(n)} + r(n)$ mit $0 \leq r(n) < 2^{k(n)}$ geschrieben werden. Definiere dann $X_n = \chi_{[r(n), 2^{-k(n)}, (r+1)2^{-k(n)}]}$ auf $([0, 1], \mathcal{B}_1, \lambda)$. Dann gilt $X_n \xrightarrow{p} 0$ aber nicht $X_n \xrightarrow{f.s.} 0$.

Satz 7.2 $X_n \xrightarrow{f.s.} X \implies X_n \xrightarrow{p} X$.

Beweis: Sei $\varepsilon > 0$. Definiere dann wie im Beweis zu Satz (7.1), B_n, A_n, \dots . Dann gilt $\lim P(B_n) = 0$ und daher $\lim P(A_n) = 0$, da $A_n \subset B_n$. Daher gilt $\lim P(|X_n - X| \geq \varepsilon) = 0$ oder $\lim P(|X_n - X| < \varepsilon) = 1$. Dies aber bedeutet $X_n \xrightarrow{p} X$.

Definition 7.3 Es seien X_n und X Zufallsvariablen und $p \geq 1$, so daß $E(|X_n|^p) < \infty$ und $E(|x|^p) < \infty$. Dann konvergiert die Folge X_n gegen X im p -ten Mittel, wenn $E(|X_n - X|^p) \rightarrow 0$, wir schreiben dafür $X_n \xrightarrow{p.M.} X$.

Satz 7.3 $X_n \xrightarrow{p.M.} X \implies X_n \xrightarrow{p} X, \quad 1 \leq p < \infty$.

Beweis: Es sei $\varepsilon > 0$. Dann gilt wegen der Chebyscheffschen Ungleichung

$$P(|X_n - X| \geq \varepsilon) = P(|X_n - X|^p \geq \varepsilon^p) \leq \frac{1}{\varepsilon^p} E(|X_n - X|^p).$$

Da nun $X_n \rightarrow X$ im p -ten Mittel, so folgt $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$, d.h. $X_n \xrightarrow{p} X$.

Man sieht leicht, daß die Folge aus Beispiel 1 im p -ten Mittel gegen Null konvergiert, für alle $\infty > p \geq 1$. Wir können dann auch nur erwarten, daß eine Teilfolge f.ü. punktweise gegen 0 strebt. Als Teilfolge können wir etwa (X_{2^n}) wählen.

Noch schwächer als der Begriff der stochastischen Konvergenz ist der Begriff der schwachen Konvergenz.

Definition 7.4 X_n konvergiert schwach gegen X , wenn für jedes $f \in \mathcal{C}^b(\mathbb{R})$, der Menge aller stetigen beschränkten Funktionen auf \mathbb{R} , gilt

$$\int f dX_n P \rightarrow \int f dX P.$$

Wir schreiben dafür $X_n \xrightarrow{w} X$.

Lemma 7.1 Für eine Folge von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ auf dem Wahrscheinlichkeitsraum (Ω, Σ, P) sind äquivalent

- i) $X_n \xrightarrow{w} X$
- ii) $\int f dX_n P \rightarrow \int f dX P \quad \forall f \in \mathcal{K}(\mathbb{R}) =$ Menge aller stetigen Funktionen mit beschränktem Träger
- iii) $F_{X_n}(x) \rightarrow F_X(x)$ für alle x in denen F_X stetig ist.

Beweis: ii) Da $\mathcal{K}(\mathbb{R}) \subset \mathcal{C}^b(\mathbb{R})$, so folgt ii) aus i).

ii) \rightarrow i): Da $X P$ ein Wahrscheinlichkeitsmaß ist, existiert zu $\varepsilon > 0$ ein n mit

$XP([-n, n]) > 1 - \varepsilon$. Es sei nun

$$g(x) = \begin{cases} 1 & |x| \leq n \\ 0 & |x| \geq n+1 \\ \text{linear} & \text{dazwischen} \end{cases} .$$

Dann gilt nach Annahme

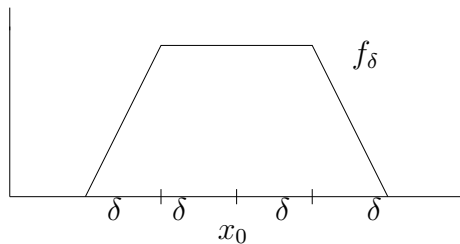
$$\int g dXP \geq 1 - \varepsilon \text{ und } \int g dX_nP \geq 1 - 2\varepsilon$$

für alle $n \geq n_0(\varepsilon)$ für ein geeignetes n_0 . Es sei nun $f \in \mathcal{C}^b(\mathbb{R})$. Wir haben dann

$$\begin{aligned} \left| \int f dX_nP - \int f dXP \right| &\leq \left| \int fg dX_nP - \int fg dXP \right| \\ &\quad + \left| \int f(1-g) dX_nP \right| + \left| \int f(1-g) dXP \right|. \end{aligned}$$

Da $fg \in \mathcal{K}(\mathbb{R})$, konvergiert der erste Term für $n \rightarrow \infty$ gegen 0. Den zweiten Term können wir durch $|f|_\infty \cdot \varepsilon$ und den dritten durch $|f|_\infty \cdot \varepsilon$ abschätzen, falls $n \geq n_0$. Dabei ist $|f|_\infty = \sup |f(x)|$. Daraus aber folgt dann ii \rightarrow i.

i) \rightarrow iii): Es sei nun x_0 ein Stetigkeitspunkt von F_X und $\varepsilon > 0$. Dann existiert ein $\delta > 0$ mit $\int f_\delta dXP < \varepsilon$. Nach Annahme existiert dann ein n_0 mit $\int f_\delta dX_nP < 2\varepsilon$ für alle $n \geq n_0$.



Es sei nun

$$f(x) = \begin{cases} 1 & |x - x_0| \leq \delta \\ 0 & |x - x_0| \geq 2\delta \\ \text{linear} & \text{dazwischen.} \end{cases}$$

Dann gilt $\int f dX_nP \rightarrow \int f dXP$ nach Annahme und es existiert ein n_1 mit $\left| \int f dX_nP - \int f dXP \right| < \varepsilon \quad \forall n \geq n_1$. Für $n \geq n_0, n_1$ haben wir dann

$$\begin{aligned} |F_{X_n}(x_0) - F_X(x_0)| &= \left| \int \chi_{(-\infty, x_0)} dX_nP - \int \chi_{(-\infty, x_0)} dXP \right| \\ &\leq \left| \int f dX_nP - \int f dXP \right| \\ &\quad + \left| \int f_\delta dX_nP \right| + \left| \int f_\delta dXP \right| < 4\varepsilon \end{aligned}$$

iii) \rightarrow i) Dieser Teil des Lemmas sei als Übungsaufgabe gestellt. Hierbei genügt es, die gegebene Funktion $f \in \mathcal{K}(\mathbb{R})$ durch Treppenfunktionen zu approximieren, deren Sprungstellen in den Stetigkeitsstellen von F_x liegen.

Schwache Konvergenz von Zufallsvariablen bedeutet also im wesentlichen Konvergenz in der Verteilungsfunktion.

Es sei hier daran erinnert, daß wir den Begriff der schwachen Konvergenz schon im Zusammenhang mit dem Satz von Moivre Laplace bzw. dem zentralen Grenzwertsatz kennengelernt hatten. Denn wir hatten:

Satz 7.4 Zentraler Grenzwertsatz: Es sei (X_n) eine Folge von unabhängigen identisch verteilter Zufallsvariablen mit endlicher positiver Varianz, dann gilt

$$Y_n = \frac{1}{\sqrt{n\sigma^2}} \left(\sum_{k=1}^n X_k - \mu \right) \xrightarrow{w} X \quad \begin{array}{l} \sigma = \sqrt{\sigma^2(X_n)} \\ \mu = E(X_n) \end{array}$$

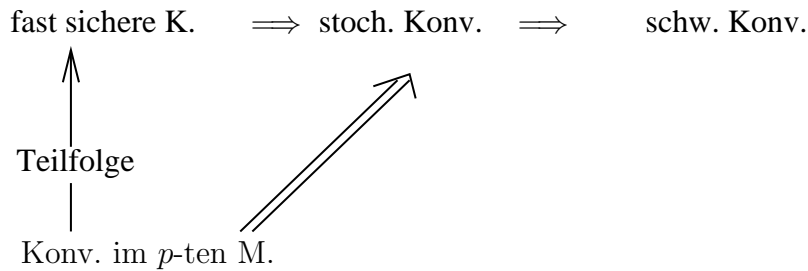
und X ist standard-normalverteilt.

Satz 7.5 $X_n \xrightarrow{p} X \implies X_n \xrightarrow{w} X$

Beweis: Es sei $\varepsilon > 0$ und $A_n = [|X_n - X| < \varepsilon]$. Dann gilt nach Annahme $P(A_n) \rightarrow 1$. Dann haben wir

$$\begin{aligned} F_{X_n}(x) &= P([X_n < x] | A_n)P(A_n) + P([X_n < x] | A_n^c)P(A_n^c) \\ &\leq P([X < x + \varepsilon]) + P(A_n^c) \end{aligned}$$

Dies zeigt: $\overline{\lim} F_{X_n}(x) \leq F_X(x + \varepsilon)$ und $\underline{\lim} F_{X_n}(x) \geq F_X(x - \varepsilon)$, da $P(A_n^c) \rightarrow 0$. Ist nun x eine Stetigkeitsstelle von F_X , so sehen wir $F_{X_n}(x) \rightarrow F_X(x)$. Wir fassen alle diese Sätze zusammen:



7.2 Das starke Gesetz der großen Zahlen

Betrachten wir die n -fache Wiederholung eines Bernoulli Experimentes mit der Ausgangswahrscheinlichkeit p für das Eintreten des Ereignisses „0“. Ist dann $X^{(n)}$ die Zufallsvariable Anzahl der „0“, so ist

$$Y_n = \frac{1}{n} X^{(n)}$$

die relative Häufigkeit des Eintreffens von „0“. Aufgrund der Motivierung des Begriffs Wahrscheinlichkeit über die relative Häufigkeit erwarten wir

$$Y_n \rightarrow p.$$

Wir haben hierbei wieder mehrere Probleme:

- i) Gibt es einen Wahrscheinlichkeitsraum auf dem alle Zufallsvariablen Y_n erklärt werden können?
- ii) Was bedeutet p ?
- iii) Um welche Art von Konvergenz handelt es sich?

Zu i) Bisher haben wir immer in einer solchen Situation immer den Wahrscheinlichkeitsraum $(\Omega_1, \sum_1, P_1)^\infty$ wo $\Omega_1 = \{0, 1\}$ und $P(0) = p, P(1) = 1 - p$ betrachten. Dann ist $Y_n(\omega)$ gerade die relative Häufigkeit der „0“ in den ersten n (!) Koordinaten von ω .

Zu ii) Wir erwarten, daß bei beliebig häufiger Wiederholung die relative Häufigkeit genau p ist, d.h. die Grenzverteilung nimmt sicher den Wert p an. Dies bedeutet aber, daß $\lim Y_n$ die Verteilung δ_p hat. Daher muß $\lim Y_n$ den Wert p fast sicher annehmen.

Zu iii) Wie man leicht sieht, hat Y_n den Erwartungswert p und die Varianz $\sqrt{\frac{pq}{n}}$, denn es gilt $E(X^{(n)}) = np$ und $\sigma^2(X^{(n)}) = npq$.

Der Satz von Moivre Laplace sagt uns nun, daß

$$P\left(\left[p - r_n\sqrt{\frac{pq}{n}} \leq Y_n \leq p + r_n\sqrt{\frac{pq}{n}}\right]\right) \approx \phi(r_n) - \phi(-r_n) \approx 1 - \frac{2}{r_n\sqrt{2\pi}}e^{-\frac{r_n^2}{2}}.$$

Falls nun $y = \lim Y_n$ existiert, so zeigt uns Satz (7.2), daß $Y_n \rightarrow Y$ f.s.

Wir wollen diese Überlegungen nun noch etwas anders fassen. Dazu sei X_k die Zufallsvariablen-Anzahl der „0“ beim k -ten Zug. Dann sind alle X_k unabhängig und $X^{(n)} = \sum_{k=1}^n X_k$. Für unseren Fall des Bernoulli Experimentes haben wir dann gezeigt

$$\frac{1}{n} \sum_{k=1}^n (X_k - EX_k) \rightarrow 0 \text{ f.s.} \quad (7.1)$$

Definition 7.5 Eine Folge P -integrierbarer Zufallsvariablen (X_k) genügt dem *starken Gesetz der großen Zahlen*, falls (??) gilt.

Unser Ziel in diesem Abschnitt ist der Beweis des *Satzes von Kolmogoroff*.

Satz 7.6 Es sei (X_n) eine Folge unabhängiger Zufallsvariablen mit den Varianzen $\sigma^2(X_n)$. Gilt dann

$$\sum_{n=1}^{\infty} \frac{\sigma^2(X_n)}{n^2} < \infty, \quad (7.2)$$

so gilt für (X_n) das starke Gesetz der großen Zahlen.

Der Beweis dieses Satzes basiert entscheidend auf der *Kolmogoroffschen Ungleichung*, die eine Verallgemeinerung der Tschebyscheffschen Ungleichung ist. Wir stellen diese Ungleichung mit Beweis daher voran.

Lemma 7.2 (*Ungleichung von Kolmogoroff*): Es seien (X_n) unabhängige Zufallsvariable mit endlichen Varianzen. Ferner sei $Y_n = \sum_{k=1}^n (X_k - EX_k)$. Dann gilt

$$P([\max_{1 \leq k \leq n} |Y_k| \geq \varepsilon]) \leq \frac{1}{\varepsilon^2} \sigma^2(Y_n) \quad (7.3)$$

Beweis: Es sei $A_0 = [|Y_k| < \varepsilon, k = 1, \dots, n]$, $A_1 = [|Y_1| \geq \varepsilon]$ und allgemeiner $A_m = [|Y_k| < \varepsilon, k = 1, \dots, m-1 \text{ und } |Y_m| \geq \varepsilon]$. Dann sind die $A_k, k = 0, \dots, n$,

disjunkt und $[\max |Y_k| \geq \varepsilon] = \sum_{m=1}^n A_m$ oder $P([\]) = \sum_{m=1}^n P(A_m)$. Es genügt also

$$\sum_{m=1}^n P(A_m) < \frac{\sigma^2(Y_n)}{\varepsilon^2} \text{ zu zeigen.}$$

Es seien nun P_m die Wahrscheinlichkeitsmaße $P(A|A_m)$. Dann haben wir wegen $E(Y_n) = 0$

$$\sigma^2(Y_n) = \int_{-\infty}^{\infty} x^2 dY P = \sum_{m=0}^n P(A_m) \int x^2 dY_n P_m$$

da für P wegen des Satzes der totalen Wahrscheinlichkeit gilt

$$P = \sum_{m=0}^n P(A_m) \cdot P_m.$$

Es ist dann also

$$\sigma^2(Y_n) \geq \sum_{m=1}^n P(A_m) \int x^2 dY_n P_m.$$

Ferner gilt

$$\begin{aligned} \int x^2 dY_n P_m &= \int Y_n^2 dP_m = \int \left\{ Y_m^2 + 2 \sum_{k>m} Y_m (X_k - EX_k) + 2 \sum_{k>m} (X_k - EX_k)^2 \right. \\ &\quad \left. + 2 \sum_{k>\ell>m} (X_k - EX_k)(X_\ell - EX_\ell) \right\} dP_m \\ &\geq \int \left\{ Y_m^2 + 2 \sum_{k>m} Y_m (X_k - EX_k) + 2 \sum_{k>\ell>m} (X_k - EX_k)(X_\ell - EX_\ell) \right\} dP_m. \end{aligned}$$

Das Ereignis A_m ist nur durch X_1, \dots, X_m bestimmt. Da die (X_i) unabhängig sind, sind es auch Y_m und $(X_k - EX_k)$ für $k > m$. Sind nun U und V Treppenfunktionen

$$U = \sum_i a_i \chi_{C_i}, \quad C_i \in \sum (Y_m) \text{ und } V = \sum b_j \chi_{B_j}, \quad B_j \in \sum (X_k)$$

mit $k > m$, so haben wir

$$\begin{aligned} \int U(V - EV) dP_m &= \sum_{i,j} a_i b_j \int \chi_{C_i} (\chi_{B_j} - E\chi_{B_j}) dP_m \\ &= \sum_{i,j} a_i b_j \frac{1}{P(A_m)} [P(A_m \cap C_i \cap B_j) - P(B_j)P(C_i \cap A_m)] \\ &= \sum_{i,j} a_i b_j \frac{1}{P(A_m)} [P(A_m \cap C_i)P(B_j) - P(B_j)P(C_i \cap A_m)] = 0. \end{aligned}$$

Offensichtlich gilt diese Aussage auch für Grenzwerte solcher Funktionen. Daher haben wir für $k > m$

$$\int Y_m (X_k - EX_k) dP_m = 0.$$

Entsprechend zeigt man $\int (X_k - EX_k)(X_\ell - EX_\ell)dP_m = 0$ für $k > \ell > m$. Damit erhalten wir

$$\int Y_n^2 dP_m \geq \int Y_m^2 dP_m, \quad m \leq n.$$

Es ist aber $\int Y_m^2 dP_m \geq \varepsilon^2$ da Y_m auf $A_m \geq \varepsilon$ ist.

Wir erhalten also

$$\sigma^2(Y_n) \geq \varepsilon^2 \sum_{m=1}^n P(A_m)$$

und hieraus folgt unmittelbar die gesuchte Ungleichung.

Beweis des Satzes von Kolmogoroff:

Es sei $Z_n = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k)$ und es seien N und m_0 natürliche Zahlen, so daß $2^{m_0} < N \leq 2^{m_0+1}$. Es sei $P_N = P([\sup_{n>N} |Z_n| > \varepsilon])$. Wir wollen dann zeigen, daß zu $\varepsilon > 0$ und $\eta > 0$ stets ein N existiert mit $P_N < \eta$. Dazu definieren wir

$$Q_m = P\left(\left[\max_{2^m < n \leq 2^{m+1}} |Z_n| > \varepsilon\right]\right) \text{ für } m \geq m_0.$$

Ersetzen wir in der Kolmogoroffschen Ungleichung noch $\sigma^2(X_k)$ durch $\sigma^2\left(\frac{X_k}{2^{m+1}}\right)$, so sehen wir

$$Q_m \leq P\left(\left[\max_{2^m < n \leq 2^{m+1}} |Z_n| > \varepsilon\right]\right) \leq \frac{1}{\varepsilon^2 2^{2m+2}} \sum_{k=2^m}^{2^{m+1}} \sigma^2(X_k).$$

Es ist aber

$$P_N \leq \sum_{m=m_0}^{\infty} Q_m \leq \sum_{m=m_0}^{\infty} \frac{1}{\varepsilon^2 2^{2m+2}} \sum_{k=2^m}^{2^{m+1}} \sigma^2(X_k).$$

Setzen wir nun

$$m(k) = \begin{cases} m_0 & \text{für } 1 \leq k \leq 2^{m_0+1} \\ m_0 + j & \text{für } 2^{m_0+j} + 1 \leq k \leq 2^{m_0+j+1}, \end{cases}$$

so können wir die Reihenfolge der Summation in der Summe rechts umkehren. Man erhält

$$P_N \leq \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma^2(X_k) \sum_{m=m(k)}^{\infty} \frac{1}{2^{2m+2}}.$$

Nun ist aber $\sum_{m=m(k)}^{\infty} 2^{-(2m+2)} = \frac{1}{3} 2^{-m(k)} \leq \frac{4}{3k^2}$. Also erhalten wir

$$P_N \leq \frac{1}{3\varepsilon^2} \left[\frac{1}{2^{2m_0}} \sum_{k=1}^{m_0} \sigma^2(X_k) + 4 \sum_{k=m_0+1}^{\infty} \frac{\sigma^2(X_k)}{k^2} \right].$$

Ist nun $\eta > 0$ und $\varepsilon > 0$ vorgegeben, so existiert stets ein m_0 bzw. N , so daß

$$\frac{4}{3\varepsilon^2} \sum_{m_0+1}^{\infty} \frac{\sigma(X_k)}{k^2} < \frac{\eta}{2}.$$

Da ferner $\sum_{k=1}^{\infty} \frac{\sigma^2(X_k)}{k^2} < \infty$, folgt leicht $\lim_{k \rightarrow \infty} \frac{1}{k^2} \sum_{i=1}^k \sigma^2(X_i) = 0$ und dies zeigt, daß auch der erste Summand für große m_0 durch $\frac{\eta}{2}$ abgeschätzt werden kann.

Zusammen gibt dies $P_N < \eta$ und der Satz ist damit bewiesen.

Korollar 7.1 Die (X_i) seien unabhängig und es existiere eine Konstante K mit $K > \sigma^2(X_i)$. Dann gilt das starke Gesetz der großen Zahlen für die Folge (X_i) .

Korollar 7.2 Die (X_i) seien unabhängig und identisch verteilt und es existiere $E(X_i^2)$. Dann gilt

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow E(X) \quad \text{f.s.} \quad (7.4)$$

Korollar (7.2) liefert uns im Spezialfall also gerade das eingangs bewiesene Ergebnis über die Binomialverteilung.

Man kann nun zeigen, daß es in Korollar 2 nicht auf die Existenz von $E(X_i^2)$ ankommt, d.h. es gilt

Satz 7.7 (Kolmogoroff): Es sei (X_i) eine Folge von integrierbaren, identisch verteilten Zufallsvariablen. Dann gilt (??).

Den Beweis für diesen Satz wollen wir hier nicht ausführen. Wir bemerken lediglich, daß man den Beweis dieses Satzes auf den Satz von Kolmogoroff (7.7) zurückspielt, indem man die X_i geeignet abschneidet.

Wir wollen nun Korollar 7.2 auf ein einfaches Beispiel anwenden.

Beispiel 7.2 Es sei $g \geq 2$ eine natürliche Zahl. Betrachte nun den Wahrscheinlichkeitsraum $([0, 1], \mathcal{B}_1([0, 1]), \lambda)$. Es sei A_g die Menge aller Zahlen $x \in [0, 1)$, deren g -adische Entwicklung abbricht. Dann ist A_g abzählbar und jedes $x \in [0, 1) \setminus A_g$ hat eine *eindeutige* Darstellung

$$X = \sum_{k=1}^{\infty} x(k)g^{-k} \quad \text{mit } x(k) \in \{0, 1, \dots, g-1\}.$$

Es sei $\ell \in \{0, 1, \dots, g-1\}$ fest. Definiere dann

$$X_k^{(\ell)}(x) = \begin{cases} 0 & x \in A_g \\ 0 & \text{falls } x(k) \neq \ell \\ 1 & \text{falls } x(k) = \ell. \end{cases}$$

Dann sind die $(X_k^{(\ell)})_{k \in \mathbb{N}}$ unabhängig, denn die Mengen

$$B_k = \{x \in [0, 1) \setminus A_g \mid x(k) = \ell\}$$

sind unabhängig. Letzteres gilt, weil

$$\lambda(B_{k_1} \cap \dots \cap B_{k_r}) = g^{-r} = \lambda(B_{k_1}) \cdots \lambda(B_{k_r}).$$

Die X_k sind identisch verteilt, $X_k \lambda = \frac{g-1}{g} \delta_0 + \frac{1}{g} \delta_1$ mit $EX_k = \frac{1}{g}$ und $EX_k^2 = \frac{1}{g}$. Also gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \frac{1}{g} \quad \text{f.s.}$$

Bezeichnet $C_g^{(\ell)}$ die Menge, auf der $\frac{1}{n} \sum X_k$ konvergiert, so ist $C_g^{(\ell)}$ genau die Menge aller $x \in [0, 1)$, für die die Zahl ℓ in g -adischen Entwicklung mit der mittleren Häufigkeit $\frac{1}{g}$ vorkommt. Wir haben also gezeigt $\lambda(C_g^{(\ell)}) = 1$.

$\bigcap_{\ell=0}^{g-1} C_g^{(\ell)}$ ist dann die Menge aller reellen Zahlen, in deren g -adischer Entwicklung jede der Zahlen $\{0, \dots, g-1\}$ mit der relativen Häufigkeit $\frac{1}{g}$ vorkommt.

Es gilt dann $\lambda(C_g) = 1$ und sogar $\lambda(\bigcap_{g \geq 2} C_g) = 1$, d.h. fast alle Zahlen in $[0, 1]$ haben eine reguläre Darstellung bezüglich der Basis g . Gleichwohl ist offen, ob π oder e diese Eigenschaft haben.

7.3 Das schwache Gesetz der großen Zahlen

Analog zum starken Gesetz der großen Zahlen definieren wir nun

Definition 7.6 Eine Folge integrierbarer Zufallsvariable (X_i) genügt dem schwachen Gesetz der großen Zahlen, wenn

$$\frac{1}{n} \sum_{i=1}^n (X_i - EX_i) \xrightarrow{P} 0. \quad (7.5)$$

Als erstes zeigen wir das Markovsche Gesetz der großen Zahlen.

Satz 7.8 Es sei (X_k) eine Folge von Zufallsvariable, so daß

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sigma^2 \left(\sum_{k=1}^n X_k \right) = 0. \quad (7.6)$$

Dann erfüllt (X_k) die Bedingung (??).

Beweis: Aus der Tschebyscheffschen Ungleichung folgt mit

$$X = \frac{1}{n} \sum_{k=1}^n X_k$$

$$P\left(\left|\frac{1}{n} \sum X_k - E\left(\frac{1}{n} \sum X_k\right)\right| \geq \varepsilon\right) \leq \frac{\sigma^2\left(\sum_{k=1}^n X_k\right)}{n^2 \varepsilon}.$$

Für $n \rightarrow \infty$ gibt dies das gewünschte Ergebnis.

Korollar 7.3 Die (X_k) seien paarweise unkorreliert und es gelte

$$\lim \frac{1}{n^2} \sum_{i=1}^n \sigma^2(X_i) = 0.$$

Dann genügt die Folge dem schwachen Gesetz der großen Zahlen.

Beweis: Es ist

$$\begin{aligned} \sigma^2 \left(\sum_{k=1}^n X_k \right) &= E \left(\sum_{k=1}^n (X_k - EX_k)^2 \right) = \sum_{k=1}^n E(X_k - EX_k)^2 \\ &\quad + 2 \sum_{k < \ell} E((X_k - EX_k)(X_\ell - EX_\ell)) = \sum_{k=1}^n \sigma^2(X_k) \end{aligned}$$

und damit ist der Satz von Markov anwendbar.

Korollar 7.4 Die (X_k) seien paarweise unkorreliert und es gelte $\sigma^2(X_k) \leq K \quad \forall k$. Dann genügt die Folge dem schwachen Gesetz der großen Zahlen.

7.4 Der zentrale Grenzwertsatz

Bevor wir mit dem Beweis des zentralen Grenzwertsatzes beginnen, benötigen wir noch einige Vorbereitungen. Es sei X eine Zufallsvariable. Wir hatten dann die momenterzeugende Funktion $M_X(t)$ durch

$$M_X(t) = E(e^{tX}) = \int e^{tX} dXP(x) \quad (7.7)$$

definiert. Dabei hatten wir aber auch gesehen, daß $M_X(t)$ nicht notwendig existieren muß.

Aus diesem Grunde betrachtet man häufig statt der momenterzeugenden Funktion die *charakteristische Funktion*, die durch

$$\phi_X(t) = E(e^{itX}) = \int e^{itx} dXP(x) \quad (7.8)$$

definiert ist. Dabei wird das Integral über eine komplexwertige Funktion f so definiert, daß es wieder linear ist. Also: Es sei (Ω, Σ, μ) ein Maßraum $f : \Omega \rightarrow \mathbb{C}$ heißt messbar (integrierbar), wenn dies für $Re f, Im f$ gilt. In diesem Fall definiert man

$$\int f d\mu = \int Re f d\mu + i \int Im f d\mu.$$

Wir sehen dann sofort: Ist X eine Zufallsvariable, so existiert $t \rightarrow \phi_X(t)$, denn $|Re e^{itx}|, |Im e^{itx}| \leq |e^{itx}| = 1$. Ist allgemeiner μ ein endliches Maß auf \mathbb{R} , so bezeichnet

$$F_{\pm}(\mu)(t) = \int e^{\pm itx} d\mu \quad (7.9)$$

die *Fourier-Transformierte* von μ . Die charakteristische Funktion von X ist also die Fourier-Transformierte der Verteilung von X . Ist $f \in L^1(\mathbb{R})$, so bezeichnet $(F_{\pm}f)(t) = \int f(x)e^{\pm itx} d\lambda(x)$ ebenfalls die Fourier-Transformierte von f .

Lemma 7.3

- a) Es sei μ ein endliches Borel-Maß auf \mathbb{R} . Dann ist $t \rightarrow F_{\pm}(\mu)(t)$ stetig und beschränkt.
- b) Sind μ, ν endliche Borel-Maße, so gilt

$$F_{\pm}(\mu * \nu) = F_{\pm}(\mu) \cdot F_{\pm}(\nu). \quad (7.10)$$

Beweis:

- a) Es sei $\varepsilon > 0$. Dann existiert ein $N > 0$ mit $\int_{|x| \geq N} d\mu < \varepsilon$. Es sei I ein endliches, abgeschlossenes Intervall. Dann ist $[-N, N] \times I \ni (x, t) \rightarrow e^{itx}$ gleichmäßig stetig. Also existiert ein $\delta > 0$ mit $|e^{\pm itx} - e^{\pm isx}| < \varepsilon$ für alle $x \in [-N, N]$ und $s, t \in I$ mit $|s - t| < \delta$. Für $s, t \in I$ und $|s - t| < \delta$ erhalten wir damit

$$\begin{aligned} |F_{\pm}(\mu)(t) - F_{\pm}(\mu)(s)| &\leq \left| \int_{-N}^N (e^{\pm isx} - e^{\pm itx}) d\mu \right| + \left| \int_{|x| > N} e^{isx} d\mu \right| \\ &\quad + \left| \int_{|x| > N} e^{\pm itx} d\mu \right| \leq \varepsilon \int_{-N}^N d\mu + 2\varepsilon \\ &\leq (\mu(\mathbb{R}) + 2)\varepsilon. \end{aligned}$$

- b)

$$\begin{aligned} F_{\pm}(\mu * \nu)(t) &= \int e^{\pm i(x+y)t} d\mu(x) d\nu(y) \\ &= \int e^{\pm itx} d\mu(x) \int e^{\pm iyt} d\nu(y) \\ &= F_{\pm}(\mu)(t) \cdot F_{\pm}(\nu)(t) \end{aligned}$$

Aufgabe 7.1 Ist X eine Zufallsvariable für die das k -te Moment existiert, so ist die charakteristische Funktion k -mal stetig differenzierbar.

Aufgabe 7.2 Existieren $E(e^{\pm tX})$ für ein $t > 0$, so ist $s \rightarrow M_X(s)$ analytisch in einer Umgebung der 0 und $\phi_X(t) = M_X(it)$.

Lemma 7.4 Es sei X standard normalverteilt (Mittelwert 0, Varianz 1). Dann ist $\phi_X(t) = e^{-t^2/2}$

Beweis: Es ist

$$\begin{aligned} \phi_X(t) &= \frac{1}{\sqrt{2\pi}} \int e^{itx} e^{-\frac{x^2}{2}} dx \\ &= \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{k!} \int (itx)^k e^{-\frac{x^2}{2}} dx \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} t^{2k} \frac{1}{\sqrt{2\pi}} \int x^{2k} e^{-\frac{x^2}{2}} dx, \end{aligned}$$

$$= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} t^{2k} \frac{(2k)!}{2^k k!} = e^{-\frac{t^2}{2}}$$

denn die ungeraden Terme fallen weg.

Lemma 7.5

- a) Es sei f eine beliebig oft differenzierbare, reellwertige Funktion mit kompaktem Träger. Dann gilt der Fouriersche Inversionsatz

$$f = \frac{1}{2\pi} F_{\pm}(F_{\mp}f) = \frac{1}{2\pi} F_{\mp}(F_{\pm}f). \quad (7.11)$$

- b) Es gilt ferner das Riemann-Lebesgue Lemma: Ist $f \in K(\mathbb{R})$, so verschwinden $F_{+}f$ und $F_{-}f$ im Unendlichen.

Beweis: a) und b) werden im Rahmen der Fourierintegrale (Analysis) bewiesen.

Aufgabe 7.3 Beweise das Riemann-Lebesgue Lemma. (Führe den Beweis auf Intervalltreppenfunktionen zurück.)

Satz 7.9 Zentraler Grenzwertsatz (Lindeberg-Levy): Es sei $(X_i)_{i=1,2,\dots}$ eine Familie unabhängiger, identisch verteilter Zufallsvariablen mit Mittelwert μ und Varianz σ . Dann gilt

$$Y_n = \sum_{i=1}^n \frac{(X_i - \mu)}{\sigma\sqrt{n}} \xrightarrow{w} Y \quad (7.12)$$

und Y ist standard normalverteilt.

Beweis:

- a) Der Beweis wird in 2 Teilen geführt. Zunächst zeigen wir, daß ϕ_{Y_n} auf jedem Kompaktum gleichmäßig gegen ϕ_y strebt.

Es bezeichne ϕ die charakteristische Funktion von $\frac{(X_j - \mu)}{\sigma}$

$$\phi(t) = \int e^{i\frac{1}{\sigma}(X_j - \mu)t} dP.$$

Dann haben wir

$$\phi_{\frac{(X_j - \mu)}{\sigma\sqrt{n}}}(t) = \phi\left(\frac{t}{\sqrt{n}}\right).$$

Wegen Lemma (7.3) erhalten wir damit

$$\phi_{Y_n}(t) = \phi\left(\frac{t}{\sqrt{n}}\right)^n.$$

Das Ergebnis von Aufgabe 7.3 zeigt uns, daß ϕ 2-mal stetig differenzierbar ist. Wenden wir den Satz von Taylor an, so sehen wir

$$\begin{aligned} \phi(t) &= \phi(0) + \phi'(0)t + \phi''(t')\frac{t^2}{2}, \quad t' \in (0, t) \\ &= 1 + \phi''(t')\frac{t^2}{2}, \end{aligned}$$

denn $\phi(0) = \int dP = 1$ und

$$\phi'(0) = \frac{d}{dt} \int e^{it(X_j - \mu)/\sigma} dP|_{t=0} = i E\left(\frac{X_j - \mu}{\sigma}\right) = 0.$$

Damit wird

$$\phi_{Y_n}(t) = \left(1 + \frac{t^2}{2n} \phi''(t')\right)^n, \quad t' \in \left(0, \frac{t}{\sqrt{n}}\right).$$

Es ist aber

$$\begin{aligned} \phi''(t) &= \frac{d^2}{dt^2} \int e^{it(X_j - \mu)/\sigma} dP \\ &= \frac{-1}{\sigma^2} \int (X_j - \mu)^2 e^{it(x_j - \mu)/\sigma} dP \rightarrow -1 \end{aligned}$$

für $t \rightarrow 0$. Dies zeigt, daß für ein festes, beschränktes abgeschlossenes Intervall I

$$\phi_{Y_n}(t) \rightarrow e^{-\frac{t^2}{2}}$$

gleichmäßig für alle $t \in I$, denn $a_n \rightarrow a$ gibt $(1 + \frac{a_n}{n})^n \rightarrow e^a$. (Beweis?)

b) Daher haben wir

$$\int e^{itx} dY_n P(x) \rightarrow e^{-\frac{t^2}{2}}$$

für alle $t \in I$. Ist nun g eine beliebige, gleichmäßig stetige Funktion mit kompaktem Träger I , so folgt durch Integration

$$\int \int e^{itx} dY_n P(x) g(t) dt \rightarrow \int e^{-\frac{t^2}{2}} g(t) dt.$$

Wegen des Satzes von Fubini und Lemma (7.4) können wir dafür schreiben

$$\begin{aligned} \int \left(\int e^{itx} g(t) dt \right) dY_n P(x) &\rightarrow \frac{1}{\sqrt{2\pi}} \int \int e^{itx} e^{-\frac{x^2}{2}} dx g(t) dt \\ &= \frac{1}{\sqrt{2\pi}} \int \int e^{itx} g(t) dt e^{-\frac{x^2}{2}} dx \end{aligned}$$

oder

$$\int F_+ g dY_n P \rightarrow \int (F_+ g) d\nu$$

wo $\nu = \frac{1}{\sqrt{2\pi}} e^{-x^2} \cdot \lambda$.

Durch ein einfaches Grenzargument läßt sich dieses Ergebnis auch noch auf stetige, integrierbare g ausdehnen. Wegen Lemma 7.5 gilt daher

$$\int f dY_n P \rightarrow \int f d\nu \quad (7.13)$$

für alle f , die beliebig oft differenzierbar sind und kompakte Träger haben. Da sich stetige Funktionen mit kompakten Träger gleichmäßig durch solche Funktionen approximieren lassen, gilt (??) auch für alle $f \in K(\mathbb{R})$. Damit aber gilt $Y_n \xrightarrow{w} Y$.

Der Zentrale Grenzwertsatz hat eine große Reihe von Verallgemeinerungen, siehe z.B. das Buch von H. Bauer.

Satz 7.10 Es sei X_n eine unabhängige Folge reeller quadratisch integrierbarer Zufallsvariablen mit Varianzen σ_n^2 . Es sei ferner

$$S_n = \sum_{i=1}^n X_i - E(X_i) \text{ und } s_n = \sigma(S_n) = \left(\sum_{i=1}^n \sigma_i^2 \right)^{1/2}.$$

Gilt dann die Fellerbedingung

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{\sigma_i}{s_n} = 0,$$

so folgt $\frac{s_n}{\sigma_n} \xrightarrow{w}$ Standard-Normalverteilung.

Für die Anwendungen von großer Bedeutung ist aber auch eine genauere Beschreibung der Konvergenzrate im Zentralen Grenzwertsatz, weil dadurch die Einsatzmöglichkeiten des Zentralen Grenzwertsatzes besser gewährleistet sind. Stellvertretend für eine Reihe allgemeiner Ergebnisse (siehe z.B. P. Hall Rates of Convergence in the Central Limit Theorem) mag hier das klassische Ergebnis von Berry und Esseen 1945 stehen.

Satz 7.11 Es seien X_1, X_2, \dots unabhängige Zufallsvariable mit Mittelwert 0 und $E(|X_i|^3) < \infty$. Ferner sei $S_n = \sum_{i=1}^n X_i$ und $s_n = \sigma(S_n)$. Dann gibt es eine universelle Konstante $C, C \leq 0,7975$, so daß

$$\sup_x |P(S_n \leq s_n x) - \phi(x)| \leq C \cdot s_n^{-3} \sum_{i=1}^n E(|X_i|^3).$$

Dabei ist ϕ die Verteilungsfunktion der Standard Normalverteilung,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Falls die X_i identisch verteilt sind, mit $E(X_i^2) = s^2$ und $E(|X_i|^3) = d$, so bedeutet diese Ungleichung gerade

$$\sup_x |P(S_n \leq s_n x) - \phi(x)| \leq C n^{-1/2} d / s^3.$$

Daß dieses Ergebnis ziemlich optimal ist, hatten wir schon beim Satz von Moivre Laplace gesehen. Für größere $|x|$ kann man in der Abschätzung oben C durch $30,54/(1 + |x|^3)$ ersetzen. Diese Abschätzung ist also für $|x| \geq 3,5$ besser als die Abschätzung des Satzes von Berry-Esseen.

Aufgabe 7.4 Beim Kniffelwürfeln wird mit 5 Würfeln 3mal gewürfelt. Ziel ist es, 5 gleiche Zahlen zu erreichen. Dabei kann man nach jedem Wurf eine beliebige Anzahl von Würfeln "liegen lassen". Wie groß ist die Wahrscheinlichkeit, einen Kniffel zu erzielen?

Aufgabe 7.5 Die nachstehende Tabelle ist eine Statistik über 647 Unfälle. Beschreiben Sie diese durch eine negative Binomialverteilung mit $r = .96$ und $p = .67$.

Anzahl von Unfällen k	0	1	2	3	4	5	≥ 6
Beobachtete Häufigkeit	447	132	42	21	3	2	0

Warum sollte man gerade diese Werte wählen?

Aufgabe 7.6 Ein Marktforschungsunternehmen hat untersucht, wie viele Einheiten k eines bestimmten Produktes gekauft werden. Es wurden insgesamt 1750 Käufer beobachtet

k	0	1	2	3	4	5	6	7	8	9	≥ 10
Häufigkeit	1671	43	19	9	2	3	1	0	0	2	0

Beschreiben Sie diese Daten durch eine negative Binomialverteilung mit $r = 0.041$ und $p = 0.315$.

Aufgabe 7.7 Bei einer Verkehrsstudie von 1011 Personenfahrzeugen wurde die Anzahl k der Beifahrer bestimmt

k	0	1	2	3	4	≥ 5
Häufigkeit	678	227	56	28	8	14

Bestimmen Sie den Mittelwert und die Varianz und benutzen Sie dies um r, p für eine negative Binomialverteilung zu bestimmen. Beschreiben Sie damit diese Daten.

Aufgabe 7.8 Richardson hat die Häufigkeit von Kriegen studiert und eine Statistik über den Ausbruch von Kriegen pro Jahr für den Zeitraum von 1500 bis 1931 erstellt. Er fand

Anzahl k von Kriegen die pro Jahr ausbrachen	0	1	2	3	4	≥ 5
Häufigkeit	223	142	48	15	4	0

Wenn Kriege zufällig ausbrechen, sollten diese Daten durch welche Verteilung beschrieben werden? Überprüfen Sie dies und interpretieren Sie das Ergebnis.

Aufgabe 7.9 In einem Gebäude der New Yorker Stadtverwaltung wurden vom 1.11.23 bis 30.9.25 insgesamt 423 Dinge verloren. Die Tageshäufigkeit k gefundener Stücke war

k	0	1	2	3	4	5	6	7
	169	134	74	32	11	2	0	1

Welche Verteilung paßt dazu? Warum?

Aufgabe 7.10 Nachstehend ist eine Statistik über die Anzahl k der jährlichen Arztbesuche von 2810 Personen

k	0	1	2	3	4	5	6	7	8	9	10
Häufigkeit	820	535	369	283	201	149	106	76	77	54	32

k	11	12	13	14	15	16	17	18	≥ 19
	31	27	14	3	6	8	3	2	14

Aufgabe 7.11 Die nachstehende Tabelle gibt die Häufigkeit an, mit der in 3 Bundesliga-Saisons k Tore fielen

k	0	1	2	3	4	5	≥ 6
	0.0507	0.0974	0.2137	0.1905	0.1951	0.1196	0.1331

Versuchen Sie eine geeignete Verteilung dafür zu finden.