Maß- und Wahrscheinlichkeitstheorie

Karl Grill
Institut für Stochastik und Wirtschaftsmathematik
TU Wien

Bierkrug:

29. Juni 2023
Version 0.2023.3

Inhaltsverzeichnis

1 Einleitung **4**
- 1.1 Messen und Zählen . . . 4
- 1.2 Das Grab des Archimedes . . . 5
- 1.3 Wahrscheinlichkeiten . . . 8
- 1.4 Voraussetzungen und Notation . . . 8
  - 1.4.1 Notation . . . 8
  - 1.4.2 Voraussetzungen . . . 10
- 1.5 Genderwahnsinn . . . 11

2 Mengensysteme und Maßfunktionen **12**
- 2.1 Eigenschaften von Maßfunktionen und Inhalten . . . 23
- 2.2 Der Fortsetzungssatz . . . 34
- 2.3 Ergänzungen . . . 41
  - 2.3.1 Topologische Räume und Borelmengen . . . 41
  - 2.3.2 Alternativer Beweis des monotone class theorems . . . 44
  - 2.3.3 Semiregularität . . . 44
  - 2.3.4 Äußere Maße und die Borelmengen . . . 45
  - 2.3.5 Hausdorffmaße . . . 46
  - 2.3.6 Der Satz von Vitali-Hahn-Saks . . . 46
  - 2.3.7 Der Witz mit den Eiern (huevos, nicht cojones) . . . 47
- 2.4 Wiederholungsfragen . . . 48

3 Lebesgue-Stieltjes Maße **50**
- 3.1 Eindimensionale Lebesgue-Stieltjes Maße . . . 50
- 3.2 Mehrdimensionale Lebesgue-Stieltjes Maße . . . 54
- 3.3 Regularität . . . 55
- 3.4 Das Lebesguemaß . . . 57
- 3.5 Spezielle Verteilungen . . . 58
  - 3.5.1 Diskrete Verteilungen . . . 59
  - 3.5.2 Stetige Verteilungen . . . 61
- 3.6 Ergänzungen . . . 63
  - 3.6.1 Borelmaße . . . 63
- 3.7 Wiederholungsfragen . . . 63

4 Messbare Funktionen **64**
- 4.1 (Erweitert) reellwertige Funktionen . . . 65
- 4.2 Wahrscheinlichkeitsräume . . . 68
- 4.3 Maße und messbare Funktionen . . . 68
- 4.4 Konvergenz von messbaren Funktionen . . . 69
- 4.5 Ergänzungen . . . 73
  - 4.5.1 Folgenkonvergenz und Teilfolgenkonvergenz . . . 73
- 4.6 Wiederholungsfragen . . . 74

5 Das Integral **76**
5.1 Definition . . . 76
5.2 Grenzwertsätze . . . 79
5.3 Riemann-Integrale . . . 82
5.4 Komplexe Integrale . . . 83
5.5 Erwartungswerte . . . 84
5.6 Die Ungleichungen von Markov und Chebychev . . . 86
5.7 Gesetze der großen Zahlen . . . 87
5.7.1 Maximalungleichungen . . . 87
5.7.2 Reihen mit unabhängigen Summanden . . . 88
5.7.3 Das starke Gesetz der großen Zahlen . . . 92
5.8 Spezielle Verteilungen . . . 93
5.8.1 Diskrete Verteilungen . . . 94
5.8.2 Stetige Verteilungen . . . 95
5.9 Ergänzungen . . . 96
5.9.1 Alternativer Beweis des starken Gesetzes der großen Zahlen . . . 96
5.9.2 Das Stieltjes-Integral . . . 96
5.10 Wiederholungsfragen . . . 97

6 Signierte Maße **99**
6.1 Ringe . . . 103
6.2 Signierte Maße auf $\mathbb{R}$ . . . 104
6.3 Wiederholungsfragen . . . 105

7 Der Satz von Radon-Nikodym **106**
7.1 In den reellen Zahlen . . . 109
7.1.1 Kochrezept für reelle Verteilungsfunktionen . . . 113
7.2 Bedingte Erwartungen . . . 113
7.3 Ergänzungen . . . 118
7.3.1 Hauptsatz-Variationen . . . 118
7.4 Wiederholungsfragen . . . 118

8 Integralungleichungen und L_p-Räume **120**
8.1 Wiederholungsfragen . . . 123

9 Produkträume **124**
9.1 Das Produkt von zwei sigmaendlichen Maßräumen . . . 124
9.2 Markovkerne . . . 129
9.3 Unendliche Produkträume . . . 130
9.4 Wiederholungsfragen . . . 132

10 Transformationssätze **133**
10.1 Wiederholungsfragen . . . 137

11 Null-Eins Gesetze **138**
11.1 Wiederholungsfragen . . . 140

12 Charakteristische Funktionen **141**
12.1 Definition und Eigenschaften . . . 141
12.2 Beispiele . . . 144
12.2.1 Diskrete Verteilungen . . . 144
12.2.2 Stetige Verteilungen . . . 144
12.2.3 Der Darstellungssatz von Riesz (Aldi-Version) . . . 145
12.3 Konvergenz von Verteilungen . . . 146
12.4 Der zentrale Grenzwertsatz . . . 150
12.4.1 Lokale Grenzwertsätze . . . 152
12.5 Mehrdimensionale charakteristische Funktionen . . . 154

12.6 Wiederholungsfragen 155

13 Momente und die Momentenerzeugende **156**
13.1 Definitionen 156
13.2 Große Abweichungen 157
13.3 Das Gesetz vom iterierten Logarithmus 161
13.4 Wiederholungsfragen 164

14 Martingale **165**
14.1 Wiederholungsfragen 168

A Tables **169**

Kapitel 1

Einleitung

> For he, by Geometrick scale, Could take the Size of Pots of Ale
>
> Samuel Butler

Ich hoffe, meine Leserinnen und Leser verzeihen mir den kleinen Kalauer mit "der Maß" (es wird nicht der letzte sein). Bei näherer Betrachtung stellt sich heraus, dass man damit nicht allzu weit vom Thema abkommt — schließlich kommt beides von "messen". Außerdem sind wir nicht in schlechter Gesellschaft — kein geringerer als Hilbert hat ja gesagt, dass man genausogut mit Tischen, Stühlen und Bierkrügen Geometrie betreiben kann wie mit Ebenen, Geraden und Punkten (vorausgesetzt, es herrschen dazwischen dieselben Beziehungen — eine allerliebste Vorstellung, das Verbindungsseidel).

Für uns Mathematiker ist die Maß (zumindest auf dem entsprechenden Abstraktionsniveau) einfach eine Punktmenge im Raum $\mathbb{R}^3$. Das Maß, im Alltag eher Inhalt oder Volumen genannt, einer solchen Menge ist eine Zahl, die vernünftigerweise zumindest die folgenden Eigenschaften haben sollte:

1. Nichtnegativität,
2. Additivität: das Maß der Vereinigung zweier disjunkter Mengen ist die Summe der Maße der einzelnen Mengen,
3. Bewegungsinvarianz: das Maß einer Menge ändert sich nicht, wenn man sie dreht oder verschiebt.

Diese Forderungen sind in der elementaren Geometrie recht erfolgreich, und wir werden sie als Ausgangspunkt für die Entwicklung der weiteren Theorie nehmen. Wir werden allerdings allgemeinere Maße zulassen (also auf die Bewegungsinvarianz verzichten) und dafür den zweiten Punkt ein wenig erweitern.

1.1 Messen und Zählen

Bevor wir uns mit dem Messen von Flächen und Volumina befassen, wollen wir an die Urgründe des Messens zurückkehren: letzten Endes messen wir Längen, Flächen und Volumina, indem wir zählen, wie oft ein (kleines) Referenzmaß hineinpasst (und Archimedes, von dem noch die Rede sein wird, hat den Euklidischen Axiomen eines hinzugefügt, das besagt, dass auch von einer sehr kurzen Strecke nur endlich viele Kopien in eine lange hineinpassen).

Das Zählen wiederum besteht, wenn man keine Abkürzungen verwendet, aus einem fortwährenden Um-Eins-Weiterzählen, 0–1–2–.... Alle Zahlen, die wir so erhalten, bilden die natürlichen Zahlen. In der Theorie der Ordinalzahlen wird diese Menge der natürlichen Zahlen gerne mit ω bezeichnet, und das Standardmodell für die natürlichen Zahlen verwendet $0 = \emptyset$ und $\alpha + 1 = \alpha \cup \{\alpha\}$. Mit diesem Modell kann man nach ω weiterzählen: $\omega + 1, \omega + 2, \ldots$, und mit entsprechend viel (unendlicher) Zeit kommt man zu Zahlen wie $\omega + \omega$, $\omega * \omega$ usw. (das Produkt kann man als die Paare

von natürlichen Zahlen mit der lexikographischen Ordnung ansehen). Alle die Mengen, die wir so erhalten, haben eine natürliche Ordnung, die eine besondere Eigenschaft haben: jede Teilmenge hat ein kleinstes Element. Eine solche Ordnung heißt *Wohlordnung*. Es ist nicht schwer zu sehen, dass von zwei wohlgeordneten Mengen eine zu einem Anfangsstück der anderen (Ordnungs-)isomorph ist, und dass eine solche Einbettung eindeutig bestimmt ist.

Wenn man an das Auswahlaxiom glaubt, dann kann jede Menge wohlgeordnet werden, und unter unseren "immer eins weiter"-Ordinalzahlen gibt es genau eine, die zu dieser Wohlordnung isomorph ist. Das gibt eine Möglichkeit, schwierige Dinge zu beweisen, die transfinite Induktion. Die verläuft ähnlich wie die Induktion in den natürlichen Zahlen, nur gibt es jetzt neben den Zahlen, die einen unmittelbaren Vorgänger haben (wie die endlichen Zahlen) noch die Limeszahlen (wie ω), die nicht als Vorgänger+1, sondern als Vereinigung aller kleineren erhalten werden, und in der transfiniten Induktion gibt es dann zwei Induktionsschritte für diese beiden Fälle.

Etwas einfacher und mit weniger Bedarf an Struktur kommt das Konzept der Mächtigkeit aus, sogar ohne Zählen, wie in der Geschichte von dem Räuber und der Räuberin, die eine Anzahl von Dublonen erbeutet haben und dann beim Teilen feststellen, dass sie nicht weiter als bis 3 zählen können und dividieren schon gar nicht. Sie kommen zu einer gerechten Aufteilung, indem sie die Münzen nach dem Prinzip "eine für dich, eine für mich" gegenüberlegen.

Diese Methode gibt uns die Möglichkeit, zwei Mengen zu vergleichen, indem wir sie als gleich groß (gleich mächtig) ansehen, wenn es eine Bijektion zwischen ihnen gibt, und andernfalls die Menge als kleiner ansehen, die in die andere injektiv abgebildet werden kann. Es könnte natürlich sein, dass zwei Mengen gar nicht in diesem Sinn vergleichbar sind, aber das Auswahlaxiom hilft uns, auch diese Hürde zu überwinden. Die Äquivalenzklassen von gleichmächtigen Mengen nennen wir Kardinalzahl. Die bekannte Tatsache, dass eine Potenzmenge stets größere Mächtigkeit hat als ihre Grundmenge, sagt uns, dass es neben den unendlich vielen endlichen Kardinalzahlen auch unendlich viele unendliche (transfinite) gibt. Wir können die Kardinalzahlen mit Hilfe der Ordinalzahlen durchnumerieren. Traditionellerweise werden dafür hebräische Buchstaben verwendet, $\aleph_0$ (sollte eigentlich ${}_0\aleph$ geschrieben werden) ist die Mächtigkeit der natürlichen Zahlen, $\aleph_1$ die erste überabzählbare, usw. Eine zweite Hierarchie ergibt sich durch bilden der Potenzmenge: $\beth_0 = \aleph_0$, und $\beth_{\alpha+1}$ ist die Mächtigkeit der Potenzmenge von $\beth_\alpha$. Insbesondere ist $\beth_1$ die Mächtigkeit der reellen Zahlen. An dieser Stelle erhebt sich natürlich die Frage, ob sich die beiden Hierarchien unterscheiden. Der Fall $\aleph_1 = \beth_1$ ist als die "Kontinuumshypothese" bekannt. Sie ist — wie auch die verallgemeinerte Kontinuumshypothese $\beth_\alpha = \aleph_\alpha$ für alle α — unabhängig von den übrigen Axiomen der Mengenlehre (ZFC).

Wir werden zwei Schreibweisen für die Mächtigkeit von Mengen verwenden: $\mathrm{card}(A)$ für die Mächtigkeit im Sinne der Kardinalzahlen, und $|A|$ für die etwas gröbere Version, die die verschiedenen Unendlichkeiten nicht unterscheidet und unendlichen Mengen den Wert ∞ zuordnen. Für die oft gebrauchte Bedingung "höchstens abzählbar", $\mathrm{card}(A) \leq \aleph_0$ werden wir einfach $A \leq \aleph_0$ schreiben.

1.2 Das Grab des Archimedes

Nach dem kurzen mengentheoretischen Vorgeplänkel (das hier nicht nur aus Jux und Tollerei steht — in der Maßtheorie gibt es einige Fragen, deren Antwort sehr davon abhängt, welche Axiome der Mengentheorie zugrunde liegen), wollen wir uns mit einigen Fragen der Flächen- und Volumsberechnung beschäftigen und folgen dabei Archimedes, der auf seine Berechnung von Kugelvolumen und -oberfläche so stolz war, dass er verfügte, dass sein Grabmal mit einer Darstellung von Kugel und Zylinder geschmückt werden soll. Abbildung 1.2 zeigt die Wiederentdeckung dieses Grabes durch Cicero 78 v.C., gesehen durch die Augen des frühen 19. Jahrhunderts.

Bevor wir uns aber mit dieser dreidimensionalen Frage herumschlagen, sehen wir uns an, wie Archimedes zu seinen Näherungen für die Kreiszahl π gekommen ist, bekanntlich

$$3\frac{10}{71} \leq \pi \leq 3\frac{1}{7}.$$

Die Methode besteht darin, dem Kreis Vielecke mit wachsender Seitenzahl einzuschreiben und umzuschreiben. Der Umfang des Kreises liegt dann zwischen den Umfängen der beiden Vielecke

Abbildung 1.1: Cicero entdeckt das Grab des Archimedes, Benjamin West, 1804
In der rechten obere Ecke ist die Kugel zu erkennen

(Archimedes hat dafür praktisch das Konzept der Konvexität erfunden), etwas leichter ist das für die Flächen zu sehen. Archimedes begann mit dem Sechseck und gelangte durch Verdoppeln der Seitenzahlen bis zum 96-Eck. Das kann man mit heutigen Mitteln etwa so machen, dass wir den Umfang des eingeschriebenen $3 \cdot 2^n$-Ecks mit u_n und den des umgeschriebenen mit U_n bezeichnet. Mit der Notation $a_n = 1/U_n$, $b_n = 1/u_n$ ergeben sich die Rekursionsformeln

$$a_{n+1} = \frac{a_n + b_n}{2}, \; b_{n+1} = \sqrt{a_{n+1} b_n}$$

(wenn man unter der Wurzel a_{n+1} durch a_n ersetzt, dann konvergieren übrigens die Folgen a_n und b_n gegen einen gemeinsamen Grenzwert, das *arithmetisch-geometrische Mittel* von a_0 und b_0; und — Überraschung — auch das kann man zur Berechnung von π verwenden, und hier ist die Konvergenz sehr viel schneller als mit der Archimedes-Methode oder der beliebten Arcustangensreihe).

Für uns ist hier allerdings nur die Methode wichtig: es werden zu einer Menge A Teil- und Obermengen gesucht, deren Inhalt hinreichend leicht zu bestimmen ist; das Supremum der Inhalte der eingeschriebenen Mengen ist das innere Maß von A, das Infimum der Inhalte der umgeschriebenen Mengen das äußere Maß. Wenn inneres und äußeres Maß übereinstimmen, dann nennen wir die Menge A messbar. Die Details hängen davon ab, wie "hinreichend leicht" interpretiert wird. Verwendet man wie Archimedes Vielecke, führt das zum Jordan-Maß, verwendet man abzählbare Vereinigungen von Vielecken für außen und die Komplemente von solchen Vereinigungen für innen, dann ergibt sich das Lebesgue-Maß (das sich vom Jordan-Maß nur durch seinen größeren Definitionsbereich unterscheidet). Die Details dazu werden wir im nächsten Kapitel behandeln.

Doch nun zu Archimedes' erstem Zugang zum Volumen der Kugel (den er selbst als zu wenig geometrisch empfunden hat, deswegen hat er einen zweiten, "geometrischeren" Beweis geliefert): wir stellen eine Kugel mit Radius r und, einen Kegel und einen Zylinder (jeweils mit Radius und Höhe $2r$) nebeneinander und schneiden in Höhe x — vom Scheitel der Kugel bzw. der Spitze des Kegels nach unten gemessen. Die Schnitte sind jeweils Kreise, beim Kegel mit Radius x, beim Zylinder mit Radius $2r$ und bei der Kugel mit Radius $\sqrt{x(2r-x)}$. Die Summe der Flächen der Schnitte von Kreis und Zylinder ergeben

$$x^2\pi + x(2r-x)\pi = 2rx\pi.$$

Das ist nicht dasselbe wie die Fläche des Zylinderschnitts ($4r^2\pi$) sondern steht dazu im Verhältnis $x : 2r$. Und hier kommt das ebenfalls von Archimedes entdeckte Hebelgesetz zu Hilfe: wir stecken den Kreis (und betrachten jetzt die Kreise als sehr dünne aber doch materielle Scheiben) aus dem Zylinder in Entfernung x vom Drehpunkt (hochkant) auf einen Hebel, die beiden anderen Kreise hängen wir an eine Schnur, die in Entfernung $2r$ auf der anderen Seite am Hebel befestigt ist. Die Scheiben sind dann im Gleichgewicht, und das bleibt auch so, wenn wir alle anderen Scheibchen dazugeben. Am Ende steckt der ganze Zylinder horizontal auf dem Hebel, sein Schwerpunkt liegt in Entfernung r vom Drehpunkt, die Kugel und der Kegel in Entfernung $2r$ wiegen ihn auf. Also betragen die vereinigten Volumina von Kugel und Kegel genau die Hälfte des Zylindervolumens,

und weil der Zylinder genau ein Drittel des Volumens des Zylinders hat (was schon vor Archimedes bekannt war), ergibt sich das Kugelvolumen als die bekannten $4r^3\pi/3$. Der heute natürlicher erscheinende Weg, mithilfe des Prinzips von Cavalieri zu zeigen, dass eine Halbkugel und ein Zylinder mit Radius und Höhe r und kegelförmiger Bohrung gleiches Volumen haben, weil die Schnitte in gleichen Höhen gleiche Flächen haben, ist von Galileo Galilei erstmals beschritten worden.

Für uns ist es mit der Infinitesimalrechnung ein leichtes, Flächen von einfachen Figuren zu bestimmen. Wenn etwa f eine stetige nichtnegative Funktion ist, dann können wir die Fläche der Menge

$$A = \{(x, y) : a \leq x \leq b, 0 < y < f(x)\}$$

als das Integral

$$F(A) = \int_a^b f(x)dx$$

bestimmen. Das läuft darauf hinaus, dass wir diese Fläche in vertikale Streifen zerlegen (Abb. 1.2).

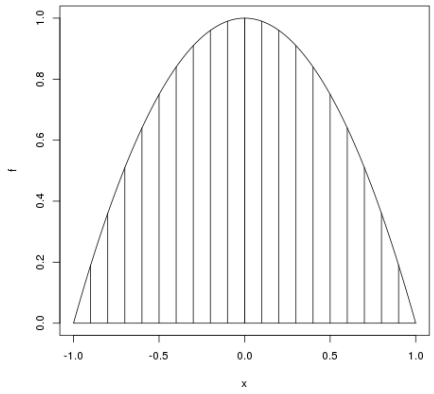


Abbildung 1.2: Das Riemann-Integral

Diese Methode zeigt einen wesentlichen Unterschied zu den Überlegungen von Archimedes, Galileo und Cavalieri: dort wurden nicht vertikale, sondern horizontale Scheibchen verwendet. Warum sollen wir das nicht auch für unser Integral verwenden? Wir zerlegen die Fläche unter der Kurve also jetzt so:

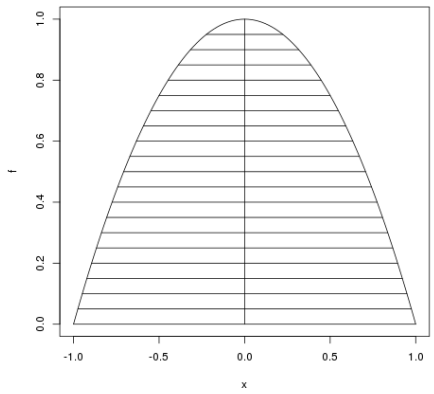


Abbildung 1.3: Das Lebesgue-Integral

Wir bestimmen also die Fläche jetzt als

$$F(A) = \int_0^\infty \text{Länge von} \{x : a \leq x \leq b, f(x) > y\} dy.$$

Es ist nicht gleich auf den ersten Blick erkennbar, dass diese Idee wirklich einen Vorteil bringt, aber es wird sich herausstellen, dass sie einen Integralbegriff liefert, der viele angenehme Eigenschaften hat: es können wesentlich mehr Funktionen integriert werden als mit dem Riemann-Integral, die Bedingungen für Grenzübergänge unter dem Integral werden sehr viel einfacher und allgemeiner, und ganz nebenbei erhalten wir auch ein Kriterium für die Riemann-Integrierbarkeit einer Funktion.

Fürs erste müssen wir aber feststellen, dass wir für dieses Integral noch ein wenig an unserem Konzept der Längenmessung arbeiten müssen. Wenn wir nämlich die Parabel aus den vorigen Abbildungen auf den Kopf stellen, zerfällt der horizontale Schnitt in zwei Teile. Das ist nicht wirklich ein Problem, weil wir natürlich die Gesamtlänge als die Summe der Längen der Teilintervalle bestimmen. Etwas komplizierter ist die Funktion

$$f(x) = \begin{cases} \sin(1/x) & \text{für } 0 < x \leq 1, \\ 0 & \text{für } x = 0. \end{cases}$$

Diese ist im Intervall $[0, 1]$ Riemann-integrierbar (die eine Unstetigkeit bei 0 stört nicht wirklich), wenn wir aber die horizontalen Schnitte betrachten, dann zerfallen diese in (abzählbar) unendlich viele Teilintervalle. Auch in diesem Fall werden wir die Gesamtlänge als die (jetzt unendliche) Summe der Längen der Teilintervalle berechnen wollen. Dieses Beispiel zeigt schon den schlimmsten Fall, den es zumindest für stetige Funktionen geben kann: die Schnittmenge in Höhe y ist ja als Urbild der offenen Menge (y, ∞) offen, und bekanntlich kann jede offene Teilmenge von $\mathbb{R}$ als abzählbare Vereinigung von disjunkten offenen Intervallen dargestellt werden.

1.3 Wahrscheinlichkeiten

Eine weitere wichtige Klasse von Maßen kommt aus der Wahrscheinlichkeitsrechnung: betrachten wir etwa das zweitliebste Spielzeug des Probabilisten, den Würfel. Die möglichen Ausgänge ("Elementarereignisse") sind die Zahlen von 1 bis 6, von denen man meistens annimmt, dass sie alle mit der gleichen Wahrscheinlichkeit angenommen werden. Ereignisse wie "die Augenzahl ist gerade" kann man als Menge von Elementarereignissen, bei deren Auftreten dieses Ereignis eintritt, schreiben (in unserem Beispiel also $\{2, 4, 6\}$). Diesen Ereignissen werden nichtnegative Zahlen, ihre Wahrscheinlichkeiten, zugeordnet. Wieder ist die Wahrscheinlichkeit der Vereinigung von zwei disjunkten Ereignissen die Summe der einzelnen Wahrscheinlichkeiten. Wir haben also auch hier ein Maß mit der zusätzlichen Einschränkung, dass die Funktionswerte höchstens gleich 1 sein dürfen. Ein solches Maß heißt dann "Wahrscheinlichkeitsmaß".

1.4 Voraussetzungen und Notation

1.4.1 Notation

> I am the one who loves changing from nothing to one.
>
> Leonard Cohen, You know who I am

Wir verwenden die üblichen Symbole:

- $\mathbb{N} = \{1, 2, \ldots\}$, die natürlichen Zahlen
- $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$, die natürlichen Zahlen
- $\mathbb{Z}$, die ganzen Zahlen,
- $\mathbb{Q}$, die rationalen Zahlen,

- $\mathbb{R}$, die reellen Zahlen,
- $\mathbb{C}$, die komplexen Zahlen.

Eine komplexe Zahl ist von der Form $z = a+bi$ mit $i^2 = -1$. a ist der Realteil, b der Imginärteil von z, in Zeichen

$$a = \Re(z) = \Re(a + bi), b = \Im(z) = \Im(a + bi).$$

Wenn wir von einer der drei Mengen $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ nur die positiven (negativen) Elemente betrachten, dann zeigen wir das durch ein hochgestelltes +(-) an. Alle Mengen können wir durch die Unendlichkeiten $\pm\infty$ ergänzen, etwa $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ und $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Für die komplexe Zahl $z = a + bi$ bezeichnen wir Real- und Imaginärteil mit

$$\Re(z) = \Re(a + bi) = a, \ \Im(z) = \Im(a + bi) = b.$$

In den meisten Fällen haben wir es mit Mengen zu tun, die Teilmengen einer vorher festgelegten Grundmenge Ω sind. Für diese Mengen können wir die üblichen Operationen Vereinigung, Durchschnitt und Differenz ($A \cup B, \bigcup_{i\in I} A_i, A \cap B, \bigcap_{i\in I} A_i, A \setminus B$) und auch das Komplement $A^C = \Omega \setminus A$ bilden. Gelegentlich verwenden wir auch die symmetrische Differenz

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B).$$

Das Produkt von zwei Mengen ist

$$A \times B = \{(x, y) : x \in A, y \in B\}.$$

Diese Multiplikation ist eigentlich nicht assoziativ, wir tun aber so und setzen

$$A \times B \times C = \{(x, y, z) : x \in A, y \in B, z \in C\}.$$

Unendliche Produkte definieren wir als die Menge aller Auswahlfunktionen:

$$\bigtimes_{i\in I} A_i = \{f : I \to \bigcup_I A_i : f(i) \in A_i\}.$$

Wenn alle A_i dieselbe Menge A sind, dann schreiben wir für das Produkt A^I. Es ist also etwa $\mathbb{R}^{\mathbb{R}}$ die Menge aller reellen Funktionen auf den reellen Zahlen.

Die Indikatorfunktion einer Menge A ist die Funktion (mit Definitionsbereich Ω)

$$A(x) = \begin{cases} 1 & \text{wenn } x \in A, \\ 0 & \text{wenn } x \notin A. \end{cases}$$

Wir bezeichnen also die Indikatorfunktion mit demselben Buchstaben wie die Menge selbst. Das spart ein wenig Schreibarbeit und sollte in den meisten Fällen keine Verwirrung stiften. Diese Idee steckt etwa auch hinter der Notation 2^A für die Potenzmenge von A, die auch wir verwenden werden: im Standardmodell für die natürlichen Zahlen ist ja $2 = \{0, 1\}$ und damit 2^A die Menge aller Indikatorfunktionen auf A.

Diese Identifikation von Mengen mit ihren Indikatorfunktionen gibt uns die Möglichkeit, disjunkte Vereinigungen sparsam aufzuschreiben: $A = B + C$ bzw. $A = \sum_I A_i$ soll bedeuten, dass die Mengen B und C bzw. $A_i, i \in I$ disjunkt sind und Vereinigung A haben, denn das ist genau das, was man erhält, wenn man die Gleichungen als Gleichungen zwischen den Indikatorfunktionen interpretiert.

Sehr oft kommen Mengen der Form $\{x \in \Omega : P(x)\}$ vor, wobei P irgendeine Aussage über x ist. Dafür verwenden wir die Kurznotation $[P]$. Die Schnittmengen aus dem vorigen Kapitel könnten wir also einfach als $[f > y]$ schreiben. Natürlich dürfen wir eine solche Menge auch wieder als ihre eigene Indikatorfunktion interpretieren, und wir werden auch Notationen wie $[x > 0]$ verwenden, die wir so interpretieren, dass dieser Ausdruck den Wert 1 hat, wenn die Aussage in der Klammer zutrifft und sonst den Wert 0. Damit sind etwa $A(x)$ und $[x \in A]$ gleichbedeutend und geben Wert 1, wenn $x \in A$, und sonst 0.

Funktionen schreiben wir einfach mit ihrem Namen, etwa f oder — gleichbedeutend — mit den Notationen $f()$ und $f(.)$ (speziell wenn wir *sehr* deutlich machen wollen, dass wir nicht eine Menge selber, sondern ihre Indikatorfunktion meinen). Die letzte Notation ist besonders nützlich, wenn wir bei einer Funktion von zwei Variablen eine festhalten wollen, etwa

$$f(x_1, .) : x_2 \mapsto f(x_1, x_2) \ (x_1 \text{ fix, Funktion des zweiten Arguments}).$$

Mit unserer Identifikation von Mengen und ihren Indikatoren können wir dieselbe Notation für Mengen verwenden:

$$A(x_1, .) = \{x_2 : (x_1, x_2) \in A\}.$$

Eine Funktion $f : \Omega_1 \to \Omega_2$ induziert (durch ihre elementweise Anwendung) eine Funktion von 2^{Ω_1} nach 2^{Ω_2}, die wir wieder mit f bezeichnen, also für $A \subseteq \Omega_1$:

$$f(A) = \{f(x), x \in A\}.$$

Das könnte man natürlich weiterspinnen...

Wenn $f : \Omega_1 \to \Omega_2$ injektiv ist, dann gibt es die Umkehrfunktion $f^{-1} : f(\Omega_1) \to \Omega_1$. Mit derselben Notation bezeichnen wir auch die Urbildfunktion, die auf der Potenzmenge von Ω_2 definiert ist:

$$f^{-1}(B) = \{\omega \in \Omega_1 : f(\omega) \in B\}.$$

Eine Quelle der Zweideutigkeit bei unserer Schreibweise der Indikatoren sind Ausdrücke wie $A + 1$ oder $2A$. Der letzte Ausdruck kann entweder als die Funktion interpretiert werden, die allen Elementen von A den Wert 2 zuordnet und sonst 0 ist, oder als die Menge, deren Elemente durch Verdoppeln der Elemente von A erhalten werden. Wir werden die erste Interpretation verwenden, für den anderen Fall führen wir neue Operatoren ein:

$$a \oplus A = \{a + x, x \in A\}, A \oplus B = \{x + y : x \in A, y \in B\},$$

und analog für $\ominus, \odot, \oslash$.

Für Maxima und Minima haben wir auch eine sparsame Notation:

$$x \vee y := \max(x, y), x \wedge y = \min(x, y).$$

Das Supremum und das Infimum einer Menge sind die kleinste obere und die größte untere Schranke (wir nehmen natürlich an, dass es sich um eine Teilmenge einer Grundmenge Ω — meist der reellen Zahlen — handelt, in der diese Operationen möglich sind). Mit der entsprechenden Interpretation von "ex falso quodlibet" kommt man zu den Interpretationen

$$\inf \emptyset = \sup \Omega, \sup \emptyset = \inf \Omega$$

(jedes Element von Ω ist ja eine obere Schranke für $\emptyset$). Speziell in den reellen Zahlen haben wir also $\inf \emptyset = \infty$.

1.4.2 Voraussetzungen

Es gibt einige Dinge aus der Analysis, die hier als bekannt vorausgesetzt werden; einiges davon wird leider erst parallel zu dieser Vorlesung entwickelt. Speziell sind das das Riemann-Integral (fürs erste sollte das genügen, was sie aus der Schule darüber wissen) und die Topologie der reellen Zahlen:

- offene/abgeschlossene Mengen
- die topologische Definition der Stetigkeit: f ist genau dann stetig, wenn für jede offene Menge U $f^{-1}(U)$ offen ist
- der Satz von Heine-Borel: eine Teilmenge von $\mathbb{R}^d$ ist genau dann kompakt, wenn sie beschränkt und abgeschlossen ist (eine Menge wird kompakt genannt, wenn zu jeder Überdeckung dieser Menge durch offene Mengen eine eine endliche Teilüberdeckung existiert).

  Weil hier mit Analysis eine klassische *race condition* besteht, ein quick and dirty Beweis der Version, die wir benötigen:

Satz 1.1: Heine-Borel für Arme

$a < b$ und $a_n < b_n$ für $n \in \mathbb{N}$ seien reelle Zahlen und

$$[a, b] \subseteq \bigcup_{n \in \mathbb{N}}]a_n, b_n[.$$

Dann gibt es ein $N \in \mathbb{N}$ mit

$$[a, b] \subseteq \bigcup_{n=1}^{N}]a_n, b_n[.$$

Wenn das nicht der Fall ist, dann gibt es zu jedem $N \in \mathbb{N}$ ein $x_N \in [a, b] \setminus \bigcup_{n=1}^{N}]a_n, b_n[$. Die Folge (x_N) hat einen Häufungspunkt $x \in [a, b]$. Wir können ein K finden, für das $x \in]a_K, b_K[$ gilt. Weil x ein Häufungspunkt von (x_N) ist, gibt es unendlich viele N mit $x_N \in]a_K, b_K[$, und das führt zu einem Widerspruch, sobald $N > K$ ist.

Darüber hinaus sind folgende Dinge wichtig:

- monotone Funktionen, speziell, dass sie abzählbar viele Unstetigkeitsstellen haben.
- Folgen und Reihen, speziell dass Reihen (und Doppelreihen) mit positiven Gliedern beliebig umgeordnet werden dürfen, ohne die Summe (endlich oder unendlich) zu ändern
- Grenzwerte, Häufungspunkte, Limes inferior/superior

1.5 Genderwahnsinn

Momentan (2023) ist wie so vieles auch das Verhältnis der Geschlechter in Frage gestellt, in einem Ausmaß, dass sich zu den klassischen zwei biologisch motivierten ein ganzes Spektrum von benutzerdefinierten Geschlechtern gesellt hat, und es gibt Bestrebungen, diese Vielfalt oder auch nur das Ungleichgewicht zwischen den "biologischen" Geschlechtern (vergleiche auch die lustige Erklärung zum equal pay day auf Wikipedia: *"Kritiker bemängeln, der Equal Pay Day vermittele —- wider besseres Wissen der Organisatoren –- den Eindruck, dass Frauen bei gleicher Arbeit hauptsächlich aufgrund von geschlechtsspezifischer Diskriminierung schlechter bezahlt würden. Tatsächlich ist der Lohnunterschied aber zum Großteil damit zu erklären, dass Frauen öfter in Teilzeit und in eher schlechter bezahlten Berufen (z. B. im sozialen und Bildungsbereich) arbeiten."* [https://de.wikipedia.org/wiki/Equal_Pay_Day, abgerufen am 12. März 2023]). Auf der anderen Seite lassen zumindest gängige Werbesujets ("Danke Mama!") eine Rückbesinnung auf das Männer- und Frauenbild der frühen Nachkriegszeit erkennen, und wenn das noch nicht genügt, sollten eine/einen letztlich die illustren Gegner (und Gegnerinnen) dieses "Genderwahnsinns" davon überzeugen können, dass es sich beim Gendern um eine wertvolle Sache handelt. Etwas weniger klar ist, welche konkrete Ausprägung des Genderns zur Anwendung kommen soll — da hat sich vom alten Binnen-I über die Nennung beider (?) Geschlechter, Unterstrich, Sternchen und Doppelpunkt eine reiche Auswahl an Möglichkeiten entwickelt. Es gibt nun in einem mathematischen Text wie diesem nicht allzuviel Gelegenheit, diese zur Anwendung zu bringen, aber an den seltenen Stellen, wo Personen in motivierenden Geschichten auftauchen, soll doch in irgendeiner Form gegendert werden, und ich habe mich dazu entschlossen, im Sinne der Vielfalt jeweils eine der Möglichkeiten zufällig auszuwählen oder auch einmal das Gendern bleiben zu lassen, oder im Sinne eines bequemeren Leseflusses auch nur die vorhandenen Rollen mit unterschiedlichen Geschlechtern zu besetzen, wobei im Zweifelsfall die bessere Rolle weiblich besetzt wird. Dieser Zugang ist zugegebenermaßen etwas naiv, aber letzten Endes geht es darum, dass niemand wegen seiner/ihrer/**er Herkunft, religiösen Überzeugung oder sexuellen Ausrichtung diskriminiert wird, und das darf ruhig auch etwas plakativ passieren und in der unbeholfenen Art und Weise, die uns jetzt zur Verfügung steht. In einer Zeit, wo Casus gewürfelt werden und niemand den korrekten Plural von "Finale" bilden kann, sollte auch diese kleine Vergewaltigung des Sprache drin sein — "Das wird man ja noch sagen dürfen!"

Kapitel 2

Mengensysteme und Maßfunktionen

"Yes, indeed," said the Unicorn,... "What can we measure?...
We are experts in the theory of measurement, not its practice."

J.L. Synge

Wir fassen zuerst zusammen, was wir von unseren Maßfunktionen erwarten:

- Das Maß ordnet Teilmengen einer Menge Ω Zahlenwerte zu.
- Das Maß einer Vereinigung (endlich oder abzählbar) von disjunkten Mengen soll gleich der Summe der einzelnen Maße sein.
- Die Werte des Maßes sollen nichtnegativ sein (reell oder auch ∞)

Der dritte Punkt in dieser Aufzählung ist verhandelbar — in der Tat werden wir uns im Laufe dieser Vorlesung auch mit "Maßen" beschäftigen, die negative oder komplexe Werte annehmen können.

Wenden wir uns dem ersten Punkt zu: unser erstes Interesse gilt dem Definitionsbereich unserer zukünftigen Maßfunktionen. Wir haben bereits bemerkt, dass wir eine Menge von Teilmengen einer bestimmten Menge dafür verwenden, und definieren gleich:

Definition 2.1

Ω sei eine beliebige Menge. Eine Teilmenge $\mathfrak{C}$ der Potenzmenge 2^Ω heißt *Mengensystem*. Die Menge Ω wird auch als die Grundmenge des Mengensystems bezeichnet, und wenn es notwendig ist, sie zu spezifizieren, sprechen wir genauer von einem Mengensystem über Ω.

Definition 2.2

$\mathfrak{C}$ sei ein Mengensystem. Eine Funktion

$$\mu : \mathfrak{C} \to \bar{\mathbb{R}} (\text{bzw. } \mathbb{C})$$

nennen wir Mengenfunktion.

Anmerkungen:

- Wir verzichten darauf, die komplexen Zahlen so wie die reellen mit Unendlichkeiten zu garnieren. Solange man nicht zu wild damit rechnet (multipliziert), kann man eine solche komplexe Mengenfunktion einfach als ein Paar — Real- und Imaginärteil — von reellen Mengenfunktionen betrachten.

- Es macht keine großen konzeptuellen Probleme, Mengenfunktionen mit Werten in $\mathbb{R}^d$ oder in einem unendlichdimensionalen Vektorraum zu betrachten (der, um den Umgang mit unendlichen Summen nicht allzu schwer zu machen, ein Banachraum sein sollte). Für Mengenfunktionen wären natürlich beliebige Wertebereiche möglich, aber wir sind hauptsächlich an Maßen interessiert, und für die muss man zumindest addieren können.

  Solche Banachwertigen Maße werden als "Vektormaße" bezeichnet.

Bei den Mengensystemen, die wir betrachten wollen, gibt es zwei Ziele: einerseits soll eine Maßfunktion für möglichst viele Mengen definiert sein, also wollen wir als Definitionsbereich ein möglichst reichhaltiges Mengensystem. Andererseits wollen wir nicht für jede einzelne Menge dieses großen Systems den Wert der Maßfunktion angeben müssen. Unser Ziel ist, ein vergleichsweise überschaubares System zu finden, auf dem wir leicht ein Maß definieren können, und dann von dort die Definition (möglichst eindeutig) auf ein umfangreicheres System fortzusetzen.

Das ist ja genau das, was wir in der Schule mit der Fläche getan haben: Diese war zuerst auch nur für Rechtecke definiert, und die Eigenschaften der Additivität, der Bewegungsinvarianz und der Nichtnegativität haben geholfen, schwierigeren Mengen eine Fläche zu geben. Wir verzichten im Sinne der Allgemeinheit auf die Bewegungsinvarianz und vorerst auch auf die Nichtnegativität und definieren:

Definition 2.3

Die Mengenfunktion μ auf dem Mengensystem $\mathfrak{C}$ heißt

1. additiv, wenn für alle $n \in \mathbb{N}$ und alle disjunkten Folgen $(A_1, \dots, A_n)$ mit $A_i \in \mathfrak{C}$ und $A = \sum_{i=1}^{n} A_i \in \mathfrak{C}$
$$\mu(A) = \sum_{i=1}^{n} \mu(A_i).$$
Wenn $\emptyset \in \mathfrak{C}$, dann soll $\mu(\emptyset) = 0$ gelten.

2. sigmaadditiv (auch abzählbar additiv), wenn sie additiv ist und zusätzlich für alle und $A, A_n \in \mathfrak{C}, n \in \mathbb{N}$ mit $A = \sum_{i=1}^{\infty} A_i$
$$\mu(A) = \sum_{i \in \mathbb{N}} \mu(A_i)$$
gilt.

Auch hier sind einige Anmerkungen fällig:

1. Die Summen, die in diesen Definitionen vorkommen, müssen einen bestimmten Wert haben. Bei Funktionen, die nur nichtnegative Werte annehmen, ist das kein Problem, im anderen Fall dürfen jedenfalls $+\infty$ und $-\infty$ nicht beide als Summanden auftreten. Die unendliche Summe darf nicht bedingt konvergieren, sonst könnte man ihr durch Umordnen beliebige Werte geben. Dies lässt sich so umformulieren: wir teilen sie in die Summe der positiven Glieder und die der negativen Glieder, und eine der beiden Summen muss einen endlichen Wert haben.

2. In den meisten Mengensystemen, die wir verwenden, ist die leere Menge enthalten, dann muss man bei der Sigmaadditivität die endliche Additivität nicht explizit fordern.

3. Wenn die leere Menge in $\mathfrak{C}$ enthalten ist und es eine Menge $A \in \mathfrak{C}$ mit $-\infty < \mu(A) < \infty$ gibt, dann folgt aus der Additivität, dass $\mu(\emptyset) = 0$ gelten muss. Mit unserer expliziten Forderung $\mu(\emptyset) = 0$ schließen wir also nur die trivialen Fälle $\mu \equiv \infty$ und $\mu \equiv -\infty$ aus.

4. Solange wir nichts Genaueres über das Mengensystem $\mathfrak{C}$ wissen, müssen wir explizit verlangen, dass die Vereinigungen in den Definitionen von Additivität und Sigmaadditivität wieder in diesem Mengensystem liegen — für Vereinigungen, die nicht in $\mathfrak{C}$ legen, ist ja der Funktionswert des Maßes (noch) nicht definiert.

5. Wir haben keine eigene Definition für die Beziehung $\mu(A+B) = \mu(A)+\mu(B)$. Diese hat wenig Wert, wenn man sie nicht zur Additivität erweitern kann, was natürlich ein wenig Struktur auf dem zugrundeliegenden Mengensystem erfordert.

Definition 2.4

Eine Mengenfunktion heißt *Maßfunktion* oder kurz *Maß*, wenn sie sigmaadditiv ist und nur nichtnegative Werte annimmt.

Diese Definition wird nicht in allen Texten in dieser Allgemeinheit verwendet. Oft werden nur Maße, die auf einer Sigmaalgebra (s. Seite15) definiert sind, als "Maß" anerkannt, Maße auf allgemeineren Mengensystemen heißen dann "Prämaße", und auch für diese wird oft gefordert, dass sie zumindest auf einem Ring (S. 15) oder Semiring (S. 16) definiert sind.

Definition 2.5

Eine Mengenfunktion heißt *Inhalt*, wenn sie additiv ist und nur nichtnegative Werte annimmt.

In Anmerkung 4 haben wir festgestellt, dass wir in der Definition der Additivität bzw. Sigmaadditivität verlangen müssen, dass neben den disjunkten Mengen A_i auch ihre Vereinigung A in dem Mengensystem $\mathfrak{C}$ liegt. Wenn das nicht der Fall ist, dann können wir versuchen, die Gleichung $\mu(A) = \sum_i \mu(A_i)$ zur Bestimmung von $\mu(A)$ zu verwenden. Wir können also endlichen bzw. abzählbaren Vereinigungen von disjunkten Mengen ein Maß zuordnen. Dasselbe gilt für Differenzen $A \setminus B$ mit $B \subseteq A$ und $A, B \in \mathfrak{C}$ Aus der Additivität folgt dann $\mu(A \setminus B) = \mu(A) - \mu(B)$ (wobei wir natürlich annehmen, dass diese Differenz definiert ist, etwa dass μ nur endliche Werte annimmt). Es kann allerdings passieren, dass wir durch unterschiedliche Darstellungen derselben Menge A zu unterschiedlichen Werten für $\mu(A)$ kommen. Wenn wir dieses Problem der Wohldefiniertheit kurz ignorieren, dürfen wir annehmen, dass das System der Mengen, deren Maß wir berechnen können, bezüglich endlicher und abzählbarer disjunkter Vereinigungen und bezüglich Differenzen von Mengen, von denen eine in der anderen enthalten ist, abgeschlossen ist.

Das führt uns zu der ersten Definition eines speziellen Typs von Mengensystemen:

Definition 2.6

Ein Dynkin-System (im weiteren Sinn) ist ein nichtleeres Mengensystem $\mathfrak{D}$, das folgende Bedingungen erfüllt

1. Wenn $A, B \in \mathfrak{D}$ mit $B \subseteq A$, dann ist auch $A \setminus B \in \mathfrak{D}$.
2. Wenn $A_n \in \mathfrak{D}, n \in \mathbb{N}$ und $A = \sum_{n\in\mathbb{N}} A_n$, dann ist $A \in \mathfrak{D}$.

Ein Dynkin-System $\mathfrak{D}$ mit $\Omega \in \mathfrak{D}$ nennen wir Dynkin-System im engeren Sinn.

Anmerkung: in der Literatur werden meist nur Dynkin-Systeme im engeren Sinn betrachtet (und natürlich wird die Unterscheidung enger/weiter gar nicht gemacht); die doppelte Definition hier ist durch einen Lesefehler des Autors entstanden, hat sich aber als gar nicht so unpraktisch erwiesen.

Dynkin-Systeme sind unser erstes Beispiel für Mengensysteme, die durch Abgeschlossenheit gegenüber gewissen Mengenoperationen gekennzeichnet sind. Eine gemeinsame Eigenschaft dieser Systeme ist, dass ein beliebiger Durchschnitt von Systemen eines bestimmten Typs (über derselben Grundmenge) wieder vom selben Typ ist: ein Durchschnitt von Dynkin-Systemen ist wieder ein Dynkin-System (wenn die disjunkte Folge A_n im Durchschnitt liegt, liegt sie in jedem der einzelnen Systeme, damit ihre Vereinigung, und diese ist dann auch im Durchschnitt; dasselbe Argument funktioniert für die Differenz), und die analoge Aussage gilt auch für die Ringe, Algebren, Sigmaringe, Sigmaalgebren und monotonen Systeme, die wir gleich definieren werden. Das hat zur Folge, dass es zu jedem Mengensystem $\mathfrak{C}$ ein kleinstes Dynkin-System (analog einen kleinsten Ring, eine kleinste Algebra,...) gibt, das $\mathfrak{C}$ enthält, dieses bestimmt sich als der Durchschnitt aller Dynkin-Systeme, in denen $\mathfrak{C}$ enthalten ist. Wir bezeichnen es mit $\mathfrak{D}(\mathfrak{C})$ bzw. $\mathfrak{D}^*(\mathfrak{C})$, je nachdem, ob wir es im engeren oder im weiteren Sinn verstehen wollen.

Die Einschränkung, die uns die Dynkin-Systeme bei der Bildung von Vereinigungen und Differenzen (nur disjunkt/nur Teilmengen) auferlegt, ist nicht schön, wir entfernen sie in der nächsten Definition.

Definition 2.7

Ein *Sigmaring* ist ein Mengensystem, das nichtleer und gegenüber der Bildung von Differenzen und abzählbaren Vereinigungen abgeschlossen ist. Ein Sigmaring, der Ω enthält, heißt *Sigmaalgebra*.

Abbildung 2.1: Schloss Sigmaringen, aus Wikimedia von Gerhard Werner, CC-BY-SA

Anmerkungen:

1. Diese Definition hat natürlich zur Folge, dass ein Sigmaring bezüglich jeder endlich- oder abzählbarstelligen Operation abgeschlossen ist, außer bezüglich der Komplementbildung, eine Sigmaalgebra auch bezüglich dieser.
2. Den von einem Mengensystem $\mathfrak{C}$ erzeugten Sigmaring bzw. die davon erzeugte Sigmaalgebra bezeichnen wir mit $\mathfrak{R}_\sigma(\mathfrak{C})$ bzw. $\mathfrak{A}_\sigma(\mathfrak{C})$
3. Wenn eine Sigmaalgebra (ein Ring, eine Algebra, ein Sigmaring) von dem Mengensystem $\mathfrak{C}$ erzeugt wird, nennen wir $\mathfrak{C}$ ein Erzeugendensystem für diese Sigmaalgebra (...).
4. Ein Dynkin-System (im weiteren Sinn), das bezüglich der Durchschnittsbildung abgeschlossen ist, ist ein Sigmaring, und wenn $\mathfrak{C}$ bezüglich der Durchschnittsbildung abgeschlossen ist, dann ist $\mathfrak{R}_\sigma(\mathfrak{C}) = \mathfrak{D}^*(\mathfrak{C})$ und $\mathfrak{A}_\sigma(\mathfrak{C}) = \mathfrak{D}(\mathfrak{C})$. Das dürfen Sie in den Übungen beweisen.
5. Eine Sigmaalgebra ist natürlich das Optimum an Mengensystem, mit dem wir arbeiten können, nur noch übertroffen von der Potenzmenge. Solange Ω höchstens abzählbar ist, kann man mit der Potenzmenge relativ problemlos arbeiten, aber etwa bei den reellen Zahlen ist das etwas schwieriger. Es gibt zum Beispiel den Satz von Ulam, der besagt, dass es (unter Voraussetzung der Kontinuumshypothese) auf der Potenzmenge von $[0,1]$ keine Maßfunktion geben kann, die die Bedingungen $\mu(\{x\}) = 0$ für alle $x \in [0,1]$ und $\mu([0,1]) = 1$ erfüllt. Speziell lässt sich also die Länge nicht zu einem Maß auf der ganzen Potenzmenge fortsetzen.

Ein wenig bescheidener ist es, statt der abzählbaren nur Abgeschlossenheit bezüglich endlicher Mengenoperationen zu verlangen:

Definition 2.8

Ein nichtleeres Mengensystem $\mathfrak{R}$ (über der Menge Ω) heißt *Ring*, wenn gilt: aus $A, B \in \mathfrak{R}$ folgt

1. $A \cup B \in \mathfrak{R}$

2. $A \setminus B \in \mathfrak{R}$

Ein Ring, der Ω enthält, heißt Algebra.

Wir bezeichnen den von $\mathfrak{C}$ erzeugten Ring mit $\mathfrak{R}(\mathfrak{C})$, die erzeugte Algebra mit $\mathfrak{A}(\mathfrak{C})$. Der Name "Ring" erklärt sich aus dem folgenden

Satz 2.1

Für einen Ring $\mathfrak{R}$ gilt:

1. $\emptyset \in \mathfrak{R}$,
2. $A, B \in \mathfrak{R} \Rightarrow A \cap B \in \mathfrak{R}$,
3. $A, B \in \mathfrak{R} \Rightarrow A \triangle B \in \mathfrak{R}$, und sogar
4. Ein Ring $\mathfrak{R}$ bildet mit der symmetrischen Differenz als Addition und dem Durchschnitt als Multiplikation einen Ring (im Sinne der Algebra). Umgekehrt ist $\mathfrak{R}$ genau dann ein (Mengen-) Ring, wenn $(\mathfrak{R}, \triangle, \cap)$ ein Ring (im Sinn der Algebra) ist.

Der Beweis erfolgt in den Übungen.

Ringe sind natürlich bezüglich aller endlichen Mengenoperationen abgeschlossen, so, wie wir es bei den Sigmaringen für die abzählbaren festgestellt haben. Dass in der Definition nur Vereinigung und Differenz erwähnt werden, liegt daran, dass daraus schon alle anderen abgeleitet werden können und dass man bestrebt ist, mit möglichst wenigen Axiomen auszukommen — je weniger Axiome man hat, umso weniger muss man überprüfen.

Beispiele für Ringe:

1. $\mathfrak{R} = 2^{\Omega}$,
2. $\mathfrak{R} = \{A \subseteq \Omega : A \text{ endlich}\}$
3. $\mathfrak{R} = \{\emptyset, \Omega\}$.

Beispiele für Sigmaringe:

1. $\mathfrak{R} = 2^{\Omega}$,
2. $\mathfrak{R} = \{A \subseteq \Omega : A \text{ höchstens abzählbar}\}$
3. $\mathfrak{R} = \{\emptyset, \Omega\}$.

Für den nächsten Typ von Mengensystem nehmen wir uns die Intervalle in den reellen Zahlen zum Vorbild, die den Ausgangspunkt für die Längenmessung bilden: eine schnelle Überprüfung zeigt, dass diese gegenüber der Durchschnittsbildung abgeschlossen sind, bezüglich der Vereinigung und der Differenz sieht die Sache auf den ersten Blick triste aus. Eine nähere Betrachtung zeigt, dass man die Differenz $B \setminus A$ mit $A \subseteq B$ als Vereinigung von disjunkten Intervallen schreiben kann, und bei ganz genauem Hinsehen ergibt sich, dass wir diese Intervalle einzeln zu A hinzugeben können und bei jedem Zwischenschritt wieder ein Intervall bekommen. Diese Eigenschaften verwenden wir für eine neue Definition:

Definition 2.9

Ein Mengensystem $\mathfrak{C}$ heißt Semiring (im weiteren Sinn), wenn gilt:

1. Falls $A, B \in \mathfrak{C}$, dann ist auch $A \cap B \in \mathfrak{C}$.
2. Falls $A, B \in \mathfrak{C}$ und $A \subseteq B$, dann gibt es endlich viele disjunkte Mengen $C_1 \dots C_n$ in $\mathfrak{C}$, sodass
$$B \setminus A = \bigcup_{i=1}^{n} C_i.$$

Wenn man zusätzlich fordert:

3. Für die Mengen C_i aus Punkt 2 gilt für alle $k \leq n$:

$$A \cup \bigcup_{i=1}^{k} C_i \in \mathfrak{C},$$

dann heißt $\mathfrak{C}$ ein Semiring im engeren Sinn.

Die etwas spröde Definition in Punkt 3 (eine Erfindung von John von Neumann) hat den Vorteil, dass wir so wie bei Ringen die Additivität nur für die Vereinigung von zwei disjunkten Mengen fordern müssen, der allgemeine Fall lässt sich daraus ableiten:

Satz 2.2

$\mathfrak{T}$ sei ein Semiring im engeren Sinn. Dann ist eine Mengenfunktion μ auf $\mathfrak{T}$ genau dann additiv, wenn für disjunkte Mengen $A, B \in \mathfrak{T}$, deren Vereinigung ebenfalls in $\mathfrak{T}$ liegt,

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Den Beweis führen wir mit vollständiger Induktion nach n in der Beziehung

$$\mu(A) = \sum_{i=1}^{n} \mu(A_i)$$

wenn $A_i, A \in \mathfrak{T}$, $A = \sum_{i=1}^{n} A_i$.

Es sei also die Behauptung des Satzes für ein bestimmtes n wahr, und $A_1, \ldots, A_{n+1}$ disjunkte Mengen aus $\mathfrak{T}$, deren Vereinigung $A = A_1 \cup \ldots \cup A_{n+1}$ auch in $\mathfrak{T}$ liegt. Wir setzen wie in der Definition des Semirings

$$A \setminus A_{n+1} = \bigcup_{j=1}^{m} C_j$$

mit disjunkten $C_j \in \mathfrak{T}$, und für alle $k \leq m$ ist

$$D_k = A_{n+1} \cup \bigcup_{j=1}^{k} C_j \in \mathfrak{C}.$$

Klarerweise gilt

$$D_k = D_{k-1} \cup C_k,$$

D_{k-1} und C_k sind disjunkt, also gilt

$$\mu(D_k) = \mu(D_{k-1}) + \mu(C_k),$$

und schließlich

$$\mu(A) = \mu(D_m) = \mu(D_0) + \sum_{j=1}^{m} \mu(C_j) = \mu(A_{n+1}) + \sum_{j=1}^{m} \mu(C_j).$$

Dasselbe Argument kann man auf die Mengen $A_i \cap D_k$ anwenden, und man erhält

$$\mu(A_i) = \sum_{j=1}^{m} \mu(A_i \cap C_j).$$

Andererseits gilt

$$C_j = \bigcup_{i=1}^{n} (C_j \cap A_i),$$

und aus der Induktionsvoraussetzung folgt

$$\mu(C_j) = \sum_{i=1}^{n} \mu(C_j \cap A_i).$$

Wir fassen zusammen:

$$\mu(A) = \mu(A_{n+1}) + \sum_{j=1}^{m} \mu(C_j) = \mu(A_{n+1}) + \sum_{j=1}^{m} \sum_{i=1}^{n} \mu(A_i \cap C_j) =$$

$$\mu(A_{n+1}) + \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(A_i \cap C_j) = \mu(A_{n+1}) + \sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{n+1} \mu(A_i).$$

Damit ist der Satz bewiesen, und wir sehen, dass wir mit unserer etwas komplizierteren Definition des Semirings eine wesentliche Vereinfachung bei der Überprüfung der Additivität gewinnen. Wir werden deshalb ab jetzt, wenn nicht konkret Gegenteiliges gesagt wird, Semiringe stets im engeren Sinn verstehen.

Beispiele für Semiringe:

1. $\mathfrak{T} = \{\emptyset\}$,
2. $\mathfrak{T} = \{\emptyset\} \cup \{\{x\} : x \in \Omega\}$,
3. $\Omega = \mathbb{R}$, $\mathfrak{T} = \{]a, b] : a \leq b\}$.

Der von dem Semiring in Punkt 3. erzeugte Sigmaring (der eigentlich eine Sigmaalgebra ist), heißt die Sigmaalgebra der Borelmengen $\mathfrak{B}$ und wird in dieser Vorlesung eine zentrale Rolle spielen. Diese Sigmaalgebra ist ein sehr reichhaltiges System, sie enthält etwa alle offenen und abgeschlossenen Mengen, alle abzählbaren Durchschnitte von solchen Mengen, usw. (allerdings kann man — mit Hilfe des Auswahlaxioms — zeigen, dass die Kardinalität von $\mathfrak{B}$ gleich der von $\mathbb{R}$ ist). Diese Sigmaalgebra gibt es auch in mehr als einer Dimension:

Definition 2.10

Es sei $\Omega = \mathbb{R}^d$ und $\mathfrak{T} = \{]a, b] : a, b \in \mathbb{R}^d, a \leq b\}$, wobei wir die Ungleichung koordinatenweise verstehen, und $]a, b]$ als den Hyperquader $]a_1, b_1] \times]a_2, b_2] \times \ldots \times]a_d, b_d]$ definieren. Dann heißt die von $\mathfrak{T}$ erzeugte Sigmaalgebra $\mathfrak{B}_d$ die Sigmaalgebra der (d-dimensionalen) Borelmengen.

Die halboffenen Intervalle sind ein praktisches Erzeugendensystem für die Borelmengen. Indem wir $\mathbb{C}$ mit $\mathbb{R}^2$ identifizieren, können wir auch die komplexen Zahlen $\mathbb{C}$ (und $\mathbb{C}^d$) mit Borelmengen versehen.

Das System $\mathfrak{T}$ in dieser Definition ist ein Semiring. Dies folgt aus dem folgenden allgemeinen

Satz 2.3

Ω_1 und Ω_2 seien zwei Mengen, über denen jeweils ein Semiring $\mathfrak{T}_1$ bzw. $\mathfrak{T}_2$ definiert sei. Dann ist

$$\mathfrak{T}_1 \otimes \mathfrak{T}_2 := \{A_1 \times A_2 : A_1 \in \mathfrak{T}_1, A_2 \in \mathfrak{T}_2\}$$

ein Semiring über $\Omega_1 \times \Omega_2$.

Dieser folgt wiederum aus dem noch allgemeineren

Satz 2.4

$\mathfrak{T}_1$ und $\mathfrak{T}_2$ seien Semiringe über derselben Menge Ω. Dann ist auch

$$\mathfrak{T} = \{A_1 \cap A_2 : A_1 \in \mathfrak{T}_1, A_2 \in \mathfrak{T}_2\}$$

ein Semiring über Ω.

Die Abgeschlossenheit bezüglich der Durchschittsbildung ist offensichtlich. Wir wählen $A_1, B_1 \in \mathfrak{T}_1$ und $A_2, B_2 \in \mathfrak{T}_2$ mit $A_1 \cap A_2 \subseteq B_1 \cap B_2$. Wir können annehmen, dass $A_1 \subseteq B_1$ und $A_2 \supseteq B_2$ gilt (sonst können wir A_i durch $A_i \cap B_i$ ersetzen). Dann gibt es disjunkte Mengen $C_1, \ldots, C_n \in \mathfrak{T}_1$ mit

$$B_1 \setminus A_1 = \bigcup_{i=1}^{n} C_i$$

und $D_1, \ldots D_m \in \mathfrak{T}_2$ mit

$$B_2 \setminus A_2 = \bigcup_{i=1}^{m} D_i$$

Wir setzen

$$E_i = \begin{cases} C_i \cap A_2 & \text{für } i = 1, \ldots, n, \\ B_1 \cap D_{i-n} & \text{für } i = n+1, \ldots, m+n. \end{cases}$$

$E_1 \ldots E_{m+n}$ ist dann eine disjunkte Zerlegung von $(B_1 \cap B_2) \setminus (A_1 \cap A_2)$, und wenn $\mathfrak{T}_1$ und $\mathfrak{T}_2$ Semiringe im engeren Sinn sind, dann ist für $1 \leq k \leq m+n$

$$(A_1 \cap A_2) \cup \bigcup_{i=1}^{k} E_i \in \mathfrak{T},$$

also auch $\mathfrak{T}$ ein Semiring im engeren Sinn.

Der vorige Satz ergibt sich aus diesem mit den Semiringen über $\Omega_1 \times \Omega_2$

$$\mathfrak{T}_1^* = \{A_1 \times \Omega_2 : A_1 \in \mathfrak{T}_1\}$$

und

$$\mathfrak{T}_2^* = \{\Omega_1 \times A_2 : A_2 \in \mathfrak{T}_2\}.$$

In Satz 2.3 wird die Aussage nicht besser, wenn man die Semiringe über den Faktormengen durch bessere Mengensysteme ersetzt — auch wenn $\mathfrak{T}_1$ und $\mathfrak{T}_2$ Sigmaalgebren sind, ist $\mathfrak{T}_1 \otimes \mathfrak{T}_2$ im allgemeinen nur ein Semiring. Wir können aber definieren:

Definition 2.11

Ω_1 und Ω_2 seien zwei Mengen, $\mathfrak{S}_i, i = 1, 2$ jeweils eine Sigmaalgebra über Ω_i. Wir nennen

$$\mathfrak{S}_1 \times \mathfrak{S}_2 = \mathfrak{A}_\sigma(\mathfrak{S}_1 \otimes \mathfrak{S}_2)$$

das Produkt (genauer: die Produktsigmaalgebra) von $\mathfrak{S}_1$ und $\mathfrak{S}_2$.

Anmerkungen:

1. Wir haben wie schon in anderen Fällen das Symbol "$\times$" mit einer neuen Bedeutung versehen — in objektorientierten Programmiersprachen wird das "Überladen" genannt, und wie dort verlassen wir uns darauf, dass die Bedeutung aus dem Kontext ersichtlich ist (sprich: wir werden das Produkt von zwei Sigmaalgebren stets in dem Sinn interpretieren, der eben definiert wurde, und nicht als das gewöhnliche Mengenprodukt).

2. Wenn es auf den beiden Sigmaalgebren Maße μ_1 und μ_2 gibt, dann ist es eine natürliche Idee, auf der Produktsigmaalgebra auch ein Produktmaß zu definieren, durch die Vorschrift $\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ — das ist ja genau der Weg, auf dem wir in der Elementargeometrie von der Längen- zur Flächenmessung gekommen sind. Dass dadurch tatsächlich ein Maß definiert wird, werden wir später behandeln.

3. Speziell ist das Produkt von zwei Borel-Sigmaalgebren wieder Borel:

$$\mathfrak{B}_n \times \mathfrak{B}_m = \mathfrak{B}_{m+n}.$$

Die folgenden Sätze zeigen, wie man aus jedem der Mengensysteme, die wir bis jetzt kennengelernt haben, neue Systeme vom selben Typ erhalten kann (wir formulieren und beweisen die Sätze für Sigmaalgebren, aber sie gelten — mit analogen Beweisen — auch für Ringe, Sigmaringe und Algebren):

Satz 2.5

$\mathfrak{S}$ sei eine Sigmaalgebra über der Grundmenge Ω_2 und $f : \Omega_1 \to \Omega_2$ eine Abbildung. Dann ist

$$f^{-1}(\mathfrak{S}) := \{f^{-1}(A) : A \in \mathfrak{S}\}$$

eine Sigmaalgebra über Ω_1, und wird als Urbild (Urbildsigmaalgebra) von $\mathfrak{S}$ bezüglich f bezeichnet.

Ist $\mathfrak{C}$ ein Erzeugendensystem für $\mathfrak{S}$, dann ist $f^{-1}(\mathfrak{C})$ ein Erzeugendensystem für $f^{-1}(\mathfrak{S})$, also

$$f^{-1}(\mathfrak{A}_\sigma(\mathfrak{C})) = \mathfrak{A}_\sigma(f^{-1}(\mathfrak{C})).$$

Zum Beweis seien $A_n, n \in \mathbb{N}$ Mengen aus $f^{-1}(\mathfrak{S})$. Dann gibt es Mengen C_n aus $\mathfrak{S}$, sodass $A_n = f^{-1}(C_n)$, und

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} f^{-1}(C_n) = f^{-1}(\bigcup_{n\in\mathbb{N}} C_n) \in f^{-1}(\mathfrak{S}),$$

und ebenso folgert man $A^C \in f^{-1}(\mathfrak{S})$, wenn $A \in f^{-1}(\mathfrak{S})$.

Für den zweiten Teil verwenden wir

Trick 1: Deal: zwei Ungleichungen für eine Gleichung

Wir gehen zu unserem Dealer und sagen: "Jimmy, ich brauch' eine Gleichung!" Jimmy sieht uns mit sinistrem Lächeln in die Augen und antwortet: "Du willst eine Gleichung? Das kostet dich zwei Ungleichungen."

Generell besteht jede Gleichung aus zwei Ungleichungen: für zwei Mengen A und B ist $A = B$ äquivalent zu $A \subseteq B$ und $B \subseteq A$, für zwei Zahlen x und y ist $x = y$ äquivalent zu $x \leq y$ und $y \leq x$. Oft ist es einfacher, die beiden Ungleichungen einzeln nachzuweisen.

Wir stellen erst einmal fest, dass die linke Seite eine Sigmaalgebra ist, die $f^{-1}(\mathfrak{C})$ enthält und deswegen die rechte Seite umfassen muss. Die umgekehrte Inklusion zeigen wir mit der Methode der guten Mengen:

Trick 2: Prinzip der guten Mengen

Um zu zeigen, dass alle Elemente x einer Menge X eine gewisse Eigenschaft $P(x)$ haben, kann man die "gute Menge"

$$G = \{x \in X : P(x)\}$$

ansetzen und nachweisen, dass $G = X$ gilt. Das klingt nach einer Zirkeldefinition, macht aber in Situationen wie in Satz 2.5 Sinn: die Menge X ist die kleinste Menge, die eine gewisse Menge X_0 enthält und bezüglich gewisser Operationen abgeschlossen ist, und die Elemente von X_0 haben die gewünschte Eigenschaft $P()$. Dann genügt es zu zeigen, dass G ebenfalls bezüglich dieser Operationen abgeschlossen ist.

Wir setzen also

$$\mathfrak{G} = \{A \in \mathfrak{A}_\sigma(\mathfrak{C}) : f^{-1}(A) \in \mathfrak{A}_\sigma(f^{-1}(\mathfrak{C}))\}.$$

$\mathfrak{G}$ ist eine Sigmaalgebra (aufmerksame Leser:innen werden hier bemerken, dass diese Behauptung einen Beweis verlangt, und dürfen diesen zur Übung auch ausformulieren) und enthält $\mathfrak{C}$, also auch $\mathfrak{A}_\sigma(\mathfrak{C})$.

Satz 2.6

Es sei $\mathfrak{S}$ eine Sigmaalgebra über Ω und $A \subseteq \Omega$. Dann ist

$$\mathfrak{S} \cap A = \{A \cap B : B \in \mathfrak{S}\},$$

die Restriktion (oder Spur) von $\mathfrak{S}$ auf A, ebenfalls eine Sigmaalgebra über A und für ein beliebiges Mengensystem $\mathfrak{C}$ über Ω gilt

$$\mathfrak{A}_\sigma(A \cap \mathfrak{C}) = A \cap \mathfrak{A}_\sigma(\mathfrak{C}).$$

Dieser Satz lässt sich ähnlich wie der vorige mit dem Prinzip der guten Mengen beweisen (was ich Ihnen als Übung ans Herz legen möchte).

Wir können ihn aber auch auf den vorhergehenden Satz zurückführen: wir setzen $\Omega_1 = A$, $\Omega_2 = \Omega$ und f gleich der Identität (genauer: der Einbettung von A in Ω).

Die letzten beiden Sätze gelten sinngemäß für Ringe, Algebren und Sigmaringe. Für Semiringe funktioniert zumindest unser Beweis nicht (und der "erzeugte" Semiring zu einem Mengensystem existiert im allgemeinem nicht, also ist die zweite Hälfte beider Sätze ohne Sinn). Ob Urbilder oder Spuren von Semiringen Semiringe sind, und die analogen Fragen für Dynkin-Systeme und die monotonen Systeme, die wir weiter unten definieren, werden in den Übungen ergründet.

Als nächstes wollen wir darüber nachdenken, wie man aus einem einfachen Mengensystem ein komplizierteres erhalten kann, also aus einem Semiring den erzeugten Ring bzw. aus einem Ring den erzeugten Sigmaring.

Für den von einem Semiring erzeugten Ring gibt es eine relativ einfache Darstellung, nämlich

Satz 2.7

$\mathfrak{T}$ sei ein Semiring. Dann gilt für den von T erzeugten Ring

$$\mathfrak{R}(\mathfrak{T}) = \{\bigcup_{i=1}^{n} A_i : n \in \mathbb{N}, A_i \in \mathfrak{T}, i = 1, \ldots, n\} =$$

$$\{\sum_{i=1}^{n} A_i : n \in \mathbb{N}, A_i \in \mathfrak{T}, i = 1, \ldots, n\}.$$

$\mathfrak{R}(\mathfrak{T})$ kann also als die Menge aller endlichen Vereinigungen bzw. aller *disjunkten* endlichen Vereinigungen dargestellt werden.

Offensichtlich sind beide Mengensysteme, die wir im obigen Satz definiert haben, in $\mathfrak{R}(\mathfrak{T})$ enthalten. Wir brauchen also nur noch zu zeigen, dass das zweite System ein Ring ist. Es ist also für zwei Mengen

$$A = \sum_{i=1}^{n} A_i$$

und

$$B = \sum_{j=1}^{m} B_j,$$

wobei sowohl die A_i als auch die B_j aus dem Semiring sind, zu zeigen, dass sowohl ihre Vereinigung als auch ihre Differenz wieder von derselben Form sind. Dabei können wir uns auf den Fall $B \in \mathfrak{T}$ beschränken, weil wir ja die Vereinigung und die Differenz von A und B erhalten können, indem wir erst B_1, dann B_2,... zu A hinzufügen bzw. von A abziehen. Aus der Definition des Semirings folgt, dass

$$A_i \setminus B = A_i \setminus (A_i \cap B) = \sum_{j=1}^{n_i} C_{ij}$$

mit Mengen $C_{ij}, j = 1 \dots n_i$ aus dem Semiring. Weil die Mengen $C_{ij}, i = 1, \dots, n, j = 1, \dots, n_i$ disjunkt sind, ist

$$A \setminus B = \bigcup_{i=1}^{n} \bigcup_{j=1}^{n_i} C_{ij}$$

die gesuchte Darstellung. Für die Vereinigung brauchen wir nur noch festzuhalten, dass

$$A \cup B = B \cup (A \setminus B)$$

ist.

Für den erzeugten Sigmaring gibt es leider keine so schöne konstruktive Darstellung wie für den Ring, aber es gibt zumindest ein Resultat, das gelegentlich die Argumentation ein wenig leichter macht. Dazu definieren wir zuerst in Analogie zu den Grenzwerten von Zahlenfolgen:

Definition 2.12

$A_1, A_2, \dots$ sei eine nichtfallende Mengenfolge (d.h., für alle n gilt $A_n \subseteq A_{n+1}$), dann nennen wir

$$\lim_n A_n = \bigcup_{n \in \mathbb{N}} A_n$$

den Grenzwert der Folge A_n. Analog soll der Grenzwert einer nichtwachsenden Mengenfolge gleich dem Durchschnitt der Folge sein. Außerdem setzen wir

$$\limsup_n A_n = \lim_n \bigcup_{k \geq n} A_k = \bigcap_n \bigcup_{k \geq n} A_k$$

und

$$\liminf A_n = \lim_n \bigcap_{k \geq n} A_k = \bigcup_n \bigcap_{k \geq n} A_k.$$

Anschaulich gesprochen ist der limes superior die Menge aller Punkte, die in unendlich vielen der Mengen A_n enthalten sind, und der limes inferior ist die Menge aller Punkte, die in fast allen A_n enthalten sind.

Mit diesem Rüstzeug kommen wir zu

Definition 2.13

Ein Mengensystem heißt monoton, wenn es gegenüber der Grenzwertbildung für monotone Mengenfolgen abgeschlossen ist, also wenn für jede monotone Mengenfolge A_n aus diesem Mengensystem auch $\lim_n A_n$ im Mengensystem enthalten ist.

Man kann leicht zeigen (und tut es in den Übungen):

Satz 2.8

Jeder monotone Ring ist ein Sigmaring.

Es gilt aber ein noch schönerer Satz:

Satz 2.9: monotone class theorem

Der von einem Ring $\mathfrak{R}$ erzeugte Sigmaring $\mathfrak{S}$ stimmt mit dem von diesem Ring erzeugten monotonen System $\mathfrak{M}$ überein.

Zuerst sei bemerkt, dass die Anmerkung zu Definition 2.6 auch für monotone Systeme gilt, und es also Sinn macht, vom erzeugten monotonen System zu sprechen.

Dann ist, weil ja jeder Sigmaring monoton ist, jedenfalls $\mathfrak{M}$ in $\mathfrak{S}$ enthalten, also bleibt noch die umgekehrte Inklusion zu zeigen. Dazu zeigen wir natürlich, dass $\mathfrak{M}$ ein Sigmaring ist, und weil

$\mathfrak{M}$ ja monoton ist, können wir uns darauf beschränken, die Ringeigenschaft von $\mathfrak{M}$ zu beweisen. Es ist also zu beweisen, dass für zwei Mengen A und B aus $\mathfrak{M}$ auch ihre Vereinigung und ihre Differenz in $\mathfrak{M}$ liegen. Für diesen Beweis bemühen wir wieder das (in vielen ähnlichen Fällen günstig anzuwendende) "Prinzip der guten Mengen". Bevor wir das angehen, sei erwähnt, dass dies auch "zu Fuß" gemacht werden könnte, mit einer Form der transfiniten Induktion. Diese Argumentation hat zwei Nachteile: einerseits ist sie nicht besonders anschaulich, und andererseits verwendet sie das Auswahlaxiom, das wir nach Möglichkeit vermeiden wollen. Die Details werden im letzten Abschnitt des Kapitels besprochen.

Wir wollen also lieber das "Prinzip der guten Mengen" nun auch zum Beweis des monotone class Theorems einsetzen, also um zu zeigen, dass das von einem Ring erzeugte monotone System bezüglich der Vereinigung und der Mengendifferenz abgeschlossen ist. Das ist nicht ganz so einfach, weil wir es jetzt nicht wie beim Komplement mit einstelligen, sondern mit zweistelligen Verknüpfungen zu tun haben. Dieses Dilemma können wir lösen, indem wir eine der Mengen festhalten: wir definieren für jede Menge A aus $\mathfrak{M}$ die Menge der der Elemente von $\mathfrak{M}$, die für A "gut" sind, also bei Vereinigung und Differenzbildung mit A nicht aus $\mathfrak{M}$ herausführen:

$$\mathfrak{G}(A) = \{B \in \mathfrak{M} : A \cup B, A \setminus B, B \setminus A \in \mathfrak{M}\}.$$

Wir haben diese Definition bewusst symmetrisch gestaltet, daher gilt:

$$B \in \mathfrak{G}(A) \Rightarrow A \in \mathfrak{G}(B).$$

Außerdem ist $\mathfrak{G}(A)$ für jedes A ein monotones System, da für eine monotone Folge $B_n \in \mathfrak{G}(A)$

$$A \setminus \lim B_n = \lim(A \setminus B_n),$$

$$A \cup \lim B_n = \lim(A \cup B_n)$$

und

$$(\lim B_n) \setminus A = \lim(B_n \setminus A)$$

gilt, und die Folgen auf der linken Seite ebenfalls monoton sind.

Nun beginnen wir mit einer Menge A aus dem Ring $\mathfrak{R}$. Für eine solche Menge ist offensichtlich jede Ringmenge in $\mathfrak{G}(A)$ enthalten, also gilt $\mathfrak{R} \subseteq \mathfrak{G}(A)$. Da $\mathfrak{G}(A)$ monoton ist und $\mathfrak{R}$ umfasst, ist $\mathfrak{G}(A) = \mathfrak{M}$. Für ein allgemeines $A \in \mathfrak{M}$ und $B \in \mathfrak{R}$ gilt, wie wir gerade gezeigt haben, $A \in \mathfrak{G}(B)$, und daher $B \in \mathfrak{G}(A)$, also ist auch in diesem Fall $\mathfrak{R} \subseteq \mathfrak{G}(A)$ und schließlich $\mathfrak{G}(A) = \mathfrak{M}$. Wir haben also gezeigt, dass für alle A und B in $\mathfrak{M}$ $B \in \mathfrak{G}(A)$ gilt, also dass $A \cup B$ und $A \setminus B$ in $\mathfrak{M}$ liegen; also ist $\mathfrak{M}$ wie behauptet ein Ring, und alles ist gut.

Mit derselben Methode kommen wir zu

Satz 2.10

Wenn $\mathfrak{C}$ bezüglich endlicher Durchschnitte abgeschlossen ist, dann gilt

$$\mathfrak{R}_\sigma(\mathfrak{C}) = \mathfrak{D}^*(\mathfrak{C})$$

und

$$\mathfrak{A}_\sigma(\mathfrak{C}) = \mathfrak{D}(\mathfrak{C}).$$

Es stimmt also in diesem Fall die erzeugte Sigmaalgebra bzw. der erzeugte Sigmaring mit dem erzeugten Dynkinsystem im engeren bzw. weiteren Sinn überein.

Das wurde bereits erwähnt, und so wie vorher delegieren wir den Beweis an die Übungen.

2.1 Eigenschaften von Maßfunktionen und Inhalten

> They are giving you a number and take away your name.
>
> Johnny Rivers, Secret Agent Man

Im vorigen Abschnitt haben wir Maße und Inhalte (als nichtnegative sigmaadditive bzw. addi-

tive Mengenfunktionen) definiert und dazu einige spezielle Mengensysteme, die geeignete Definitionsbereiche dafür darstellen.

Hier sind einige Beispiele für Maßfunktionen:

1. $\mathfrak{C} = 2^\Omega$, $\mu(A) = 0$.
2. $\mathfrak{C} = 2^\Omega$, $x \in \Omega$ beliebig,
$$\delta_x(A) = A(x) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst,} \end{cases}$$
das Punktmaß (oder Dirac'-Maß) bei x.
3. $\mu(A) = |A|$, das Zählmaß.
4. Die Länge von Intervallen: $\Omega = \mathbb{R}$, $\mathfrak{C} = \{]a, b] : a, b \in \mathbb{R}, a \leq b\}$, $\lambda(]a, b]) = b - a$. Hier ist leicht zu sehen, dass λ ein Inhalt ist, also additiv. Dass es sich tatsächlich um ein Maß handelt, wird im nächsten Kapitel gezeigt.

Wir definieren noch:

Definition 2.14

μ sei eine Maßfunktion auf einem Mengensystem $\mathfrak{C}$.

1. $A \in \mathfrak{C}$ ist von endlichem Maß, wenn $\mu(A) < \infty$,
2. $A \subseteq \Omega$ ist von sigmaendlichem Maß (kurz "sigmaendlich"), wenn eine Folge (A_n) von Mengen aus $\mathfrak{C}$ gibt mit $\mu(A_n) < \infty$ und $A \subseteq \bigcup_n A_n$.
3. μ heißt *endlich* bzw. *sigmaendlich*, wenn jedes $A \in \mathfrak{C}$ endliches bzw. sigmaendliches Maß hat.
4. μ heißt *totalsigmaendlich* bzw. *totalendlich*, wenn Ω sigmaendliches bzw. endliches Maß hat.
5. Ein endliches Maß mit $\mu(\Omega) = 1$ (dazu muss natürlich $\Omega \in \mathfrak{C}$ gelten) heißt *Wahrscheinlichkeitsmaß.*

Dass das Zählmaß ein Maß darstellt, folgt aus dem folgenden Satz über Grenzwerte von Maßfunktionen bzw. Inhalten:

Satz 2.11

1. μ_n sei eine Folge von Inhalten auf dem Mengensystem $\mathfrak{C}$. Wenn für alle $A \in \mathfrak{C}$ $\mu(A) = \lim \mu_n(A)$ existiert, dann ist μ ein Inhalt.
2. μ_n sei eine nichtfallende Folge von Maßen auf dem Mengensystem $\mathfrak{C}$ bzw. allgemeiner $\mathcal{M}$ eine aufwärts gerichtete Menge von Maßen auf $\mathfrak{C}$ (d.h., für $\mu_1, \mu_2 \in \mathcal{M}$ gibt es ein $\mu_3 \in \mathcal{M}$ mit $\max(\mu_1, \mu_2) \leq \mu_3$). Dann ist auch $\mu = \lim \mu_n$ bzw. $\mu = \sup \mathcal{M}$ ein Maß. Insbesondere ist für eine Familie $(\mu_\alpha, \alpha \in I)$ die Summe $\sum_{\alpha \in I} \mu_\alpha$ ein Maß.

Der erste Teil (der sich ähnlich wie der zweite auch allgemeiner für Netze formulieren lässt) ist trivial zu beweisen, für den zweiten wählen wir eine Folge A_n von disjunkten Mengen aus $\mathfrak{C}$, deren Vereinigung A ebenfalls in $\mathfrak{C}$ liegt. Für jedes $M < \mu(A)$ können wir ein Maß $\nu \in \mathcal{M}$ finden, sodass $\nu(A) > M$ gilt. Daraus folgt

$$\sum_n \mu(A_n) \geq \sum_n \nu(A_n) = \nu(A) > M,$$

und da M beliebig gewählt werden kann, folgt daraus

$$\sum_n \mu(A_n) \geq \mu(A).$$

Die andere Richtung ist trivial erfüllt, wenn $\mu(A) = \infty$ gilt. Im anderen Fall gilt für jedes n $\mu(A_n) \leq \mu(A) < \infty$. Für jedes $\epsilon > 0$ und jedes n gibt es ein Maß $\nu_{n,\epsilon} \in \mathcal{M}$ mit $\nu_{n\epsilon}(A_n) \geq \mu(A_n) - \epsilon/2^n$. Weil $\mathcal{M}$ aufwärts gerichtet ist, gibt es ein $\nu \in \mathcal{M}$ mit $\nu \geq \nu_{n,\epsilon}$ für alle $n = 1, \ldots N$. Damit ist

$$\mu(A) \geq \nu(A) \geq \sum_{n=1}^{N} \nu(A_n) \geq \sum_{n=1}^{N} \mu(A_n) - \epsilon.$$

Weil $\epsilon > 0$ beliebig gewählt werden kann, gilt auch

$$\mu(A) \geq \sum_{n=1}^{N} \mu(A_n)$$

und für $N \to \infty$

$$\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n).$$

Der punktweise Grenzwert einer Folge von Maßfunktionen, die nicht monoton wächst, ist im allgemeinen keine Maßfunktion (aber immer noch ein Inhalt). Es gilt aber der bemerkenswerte

Satz 2.12: Vitali-Hahn-Saks

μ_n sei eine Folge von endlichen Maßfunktionen auf dem Sigmaring $\mathfrak{R}$. Wenn für jedes $A \in \mathfrak{R}$ der Grenzwert $\mu(A) = \lim_n \mu_n(A)$ existiert, dann ist μ ein Maß.

Der Beweis wird im letzten Abschnitt dieses Kapitels geführt.

Wir sind natürlich vor allem an Maßfunktionen interessiert, die auf Sigmaringen definiert sind, vor allem auch deshalb, weil dann die Vereinigung aus der Definition der Sigmaadditivität automatisch im Mengensystem liegt. Um eine Maßfunktion festzulegen, wollen wir aber lieber ein etwas weniger umfangreiches System benutzen, und Semiringe sind hier sehr praktisch, besonders weil (für Semiringe im engeren Sinn) die Additivität nur für zwei Mengen überprüft werden muss.

Wir beginnen nun mit einer Maßfunktion auf einem Semiring und wollen daraus eine Maßfunktion auf dem erzeugten Sigmaring gewinnen. Auf dem Weg zu diesem Ziel werden wir vorerst den erzeugten Ring betrachten:

Satz 2.13: Fortsetzungssatz I

μ sei ein Inhalt auf dem Semiring $\mathfrak{T}$. Dann gilt

1. μ lässt sich in eindeutiger Weise zu einem Inhalt auf dem von $\mathfrak{T}$ erzeugten Ring $\mathfrak{R}$ fortsetzen (wir werden diese Fortsetzung ebenfalls mit dem Buchstaben μ bezeichnen).
2. Falls μ ein Maß ist, dann ist auch die Fortsetzung auf den Ring ein Maß.

Wir wissen, dass der von $\mathfrak{T}$ erzeugte Ring die Menge aller disjunkten Vereinigungen von Mengen aus dem Semiring ist. Falls

$$A = \sum_{i=1}^{n} A_i$$

eine solche Darstellung von $A \in \mathfrak{R}$ ist, dann muss gelten

$$\mu(A) = \sum_{i=1}^{n} \mu(A_i).$$

Dadurch ist μ auf $\mathfrak{R}$ auch schon festgelegt, und es ist nur nachzuprüfen, dass diese Fortsetzung wohldefiniert und additiv bzw. sigmaadditiv ist.

Für die Wohldefiniertheit sei

$$A = \sum_{j=1}^{m} B_j$$

eine weitere Darstellung von A als Vereinigung von disjunkten Mengen aus $\mathfrak{T}$. Dann gilt wegen der Additivität von μ auf $\mathfrak{T}$

$$\mu(A_i) = \sum_{j=1}^{m} \mu(A_i \cap B_j)$$

und

$$\mu(B_j) = \sum_{i=1}^{n} \mu(A_i \cap B_j).$$

und schließlich

$$\sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(A_i \cap B_j) =$$

$$\sum_{j=1}^{m} \sum_{i=1}^{n} \mu(A_i \cap B_j) = \sum_{j=1}^{m} \mu(B_j).$$

Für die Additivität brauchen wir nur festzustellen, dass, falls

$$A = \sum_i A_i$$

und

$$B = \sum_j B_j$$

zwei disjunkte Mengen sind, dann

$$A \cup B = \sum_i A_i + \sum_j B_j$$

schon eine Darstellung der Vereinigung durch disjunkte Mengen aus dem Semiring ist.

Die Sigmaadditivität ist etwas schwieriger, aber inzwischen haben wir ja schon etwas Übung: zuerst stellen wir fest, dass eine abzählbare Vereinigung von disjunkten Mengen aus dem Ring sich als abzählbare Vereinigung von Mengen aus dem Semiring schreiben lässt, indem wir jede der Mengen aus dem Ring durch eine endliche Vereinigung von Mengen aus dem Semiring ersetzen. Es bleibt also zu zeigen: wenn A_n eine Folge von disjunkten Mengen aus dem Semiring ist, deren Vereinigung B im Ring liegt, dann ist

$$\mu(B) = \sum_n \mu(A_n).$$

Wir können B wieder als endliche Vereinigung von disjunkten Mengen B_j aus dem Semiring schreiben, und wegen der Sigmaadditivität von μ auf $\mathfrak{T}$ gilt:

$$\mu(B) = \sum_j \mu(B_j) = \sum_j \sum_n \mu(B_j \cap A_n) = \sum_n \sum_j \mu(B_j \cap A_n) = \sum_n \mu(A_n).$$

Bevor wir den nächsten Schritt, nämlich die Fortsetzung auf den Sigmaring, wagen, sammeln wir ein paar Eigenschaften von Maßen auf Semiringen. Für die Beweise können wir annehmen, dass der Definitionsbereich des Maßes ein Ring ist, weil wir ja wissen, dass wir das Maß immer auf den erzeugten Ring fortsetzen können.

Satz 2.14: Eigenschaften von Maßen (Inhalten) I

μ sei ein Inhalt auf dem Semiring $\mathfrak{T}$. Dann gilt:

1. (Monotonie) Falls A, B Mengen aus $\mathfrak{T}$ sind mit $A \subseteq B$, dann gilt

$$\mu(A) \leq \mu(B).$$

2. (Subadditivität) Falls A und $B_n,\ n \leq N$ Mengen aus $\mathfrak{T}$ sind mit $A \subseteq \bigcup_{n=1}^{N} B_n$, dann

gilt

$$\mu(A) \leq \sum_{n=1}^{N} \mu(B_n).$$

3. Ist μ ein Maß, dann gilt auch (Sigmasubadditivität, abzählbare Subadditivität) Falls A und $B_n,\ n \in \mathbb{N}$ Mengen aus $\mathfrak{T}$ sind mit $A \subseteq \bigcup B_n$, dann gilt

$$\mu(A) \leq \sum_{n\in\mathbb{N}} \mu(B_n).$$

4. Ohne die Sigmaadditivität kommt die umgekehrte Richtung aus, die schon eine Hälfte der Sigmaadditivität liefert: es seien die disjunkten Mengen $A_n, n \in \mathbb{N}$ und die Menge A aus $\mathfrak{T}$ und $\sum_n A_n \subseteq A$. Dann gilt

$$\sum_{n\in\mathbb{N}} \mu(A_n) \leq \mu(A).$$

Der erste Teil ist einfach:

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A).$$

Für den zweiten Teil verwenden wir

Trick 3: Monoton/disjunkt machen

Aus einer Folge A_n erhalten wir durch

$$B_n = \bigcup_{i=1}^{n} A_i$$

eine monotone Folge, die

$$\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$$

erfüllt (und natürlich auch $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$). Aus der monotonen Folge B_n erhalten wir mit

$$C_1 = B_1 = A_1,$$

$$C_n = B_n \setminus B_{n-1} = A_n \setminus \bigcup_{i=1}^{n-1} A_i$$

eine disjunkte Folge mit

$$C_n \subseteq A_n,$$

die ebenfalls die Beziehungen

$$\bigcup_{i=1}^{n} C_i = \bigcup_{i=1}^{n} A_i$$

und

$$\bigcup_{n\in\mathbb{N}} C_i = \bigcup_{n\in\mathbb{N}} A_n$$

erfüllt.

In diesem Sinn setzen wir

$$C_n = A \cap B_n \setminus \bigcup_{k<n} B_k.$$

Diese Mengen sind disjunkt, und es gilt $C_n \subseteq B_n$ und

$$\bigcup_{n\in\mathbb{N}} C_n = A,$$

und aus dem ersten Teil unseres Satzes folgt

$$\mu(A) = \sum \mu(C_n) \leq \sum \mu(B_n).$$

Die letzte Eigenschaft ergibt sich direkt aus der endlichen Version: $\sum_{n=1}^{N} A_n \subseteq A$ liegt ja im Ring, daher gilt

$$\sum_{n=1}^{N} \mu(A_n) = \mu(\sum_{n=1}^{N} A_n) \leq \mu(A),$$

und $N \to \infty$ liefert die Behauptung.

Der letzte Satz besagt, dass Maßfunktionen in einem gewissen Sinn monoton sind. Im nächsten Satz geht es um eine Art Stetigkeit:

Satz 2.15: Eigenschaften von Maßen II

μ sei eine Maßfunktion auf dem Semiring $\mathfrak{T}$. Dann gilt

1. (Stetigkeit von unten) Falls A_n eine nichtfallende Mengenfolge ist, dann gilt
$$\mu(\lim A_n) = \lim \mu(A_n).$$
2. (Stetigkeit von oben) Falls A_n eine nichtwachsende Mengenfolge ist und zusätzlich $\mu(A_1) < \infty$, dann gilt
$$\mu(\lim A_n) = \lim \mu(A_n).$$

Für den ersten Teil dieses Satzes setzen wir

$$B_n = A_n \setminus A_{n-1},$$

(mit $A_0 = \emptyset$). Diese Mengen sind disjunkt, und

$$A_n = \bigcup_{k=1}^{n} B_k$$

und

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

Damit erhalten wir

$$\mu(\lim A_n) = \mu(\bigcup A_n) = \mu(\bigcup B_n) =$$

$$\sum_{n\in\mathbb{N}} \mu(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu(B_n) = \lim_{N\to\infty} \mu(A_N).$$

Den zweiten Teil können wir leicht aus dem ersten erhalten, wenn wir beachten, dass

$$\lim A_n = A_1 \setminus (\lim(A_1 \setminus A_n)).$$

Dieser Satz hat auch eine Umkehrung, für die wir allerdings annehmen müssen, dass der Definitionsbereich ein Ring ist:

Satz 2.16: Kriterien für Sigmaadditivität

μ sei eine additive Funktion auf einem Ring $\mathfrak{R}$. Dann gilt:

1. Falls für jede nichtfallende Mengenfolge A_n mit $\lim A_n \in \mathfrak{R}$ gilt, dass
$$\mu(\lim A_n) = \lim \mu(A_n)$$
(also wenn μ stetig von unten ist), dann ist μ sigmaadditiv.
2. Falls μ endlich ist und für jede nichtwachsende Mengenfolge A_n mit $\lim A_n = \emptyset$ gilt,

dass

$$\lim \mu(A_n) = 0$$

(also wenn μ bei $\emptyset$ von oben stetig ist), dann ist μ sigmaadditiv.

Wir müssen zeigen, dass für disjunkte Mengen A_n aus $\mathfrak{R}$ mit

$$A = \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{R}$$

gilt, dass

$$\mu(A) = \sum \mu(A_n).$$

Im ersten Fall folgt das leicht aus der Tatsache, dass

$$A = \lim_n \bigcup_{k\le n} A_k$$

und der Additivität von μ. Den zweiten Fall reduzieren wir (wie im Witz mit den Eiern) auf den ersten, indem wir feststellen, dass für eine nichtfallende Folge A_n die Folge

$$B_n = (\lim_k A_k) \setminus A_n$$

eine nichtwachsende Folge mit Durchschnitt $\emptyset$ ist.

Als nächstes betrachten wir die Vereinigung von zwei Mengen mit endlichem Maß, die nicht unbedingt leeren Durchschnitt haben müssen (wieder in einem Ring):

$$\mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A) =$$

$$\mu(A) + \mu(B \setminus (A \cap B)) = \mu(A) + \mu(B) - \mu(A \cap B).$$

Aus dieser Gleichung erhält man mittels vollständiger Induktion den folgenden Satz:

Satz 2.17: Additionstheorem

$A_1, \ldots, A_n$ seien Mengen aus einem Ring mit einem Inhalt μ, und $\mu(A_i)$ sei endlich für alle $i = 1, \ldots, n$. Dann gilt:

$$\mu(\bigcup_{i=1}^{n} A_n) = \sum_{I\subseteq\{1,\ldots,n\}, I\neq\emptyset} (-1)^{|I|-1} \mu(\bigcap_{i\in I} A_i) =$$

$$\mu(A_1) + \ldots + \mu(A_n) - \mu(A_1 \cap A_2) - \ldots \mu(A_{n-1} \cap A_n) + \mu(A_1 \cap A_2 \cap A_3) + + - - \ldots +$$
$$(-1)^{n-1} \mu(A_1 \cap \ldots \cap A_n).$$

Zum Abschluss zeigen wir noch einen Satz für Maße auf Sigmaringen, der uns später bei Fragen der Konvergenz immer wieder nützlich sein wird:

Satz 2.18: Lemma von Borel-Cantelli

μ sei eine Maßfunktion auf dem Sigmaring $\mathfrak{S}$. Falls die Mengen A_n aus $\mathfrak{S}$ die Beziehung

$$\sum \mu(A_n) < \infty$$

erfüllen, dann gilt

$$\mu(\limsup A_n) = 0.$$

Zum Beweis verwenden wir die Stetigkeit von μ:

$$\mu(\limsup A_n) = \mu(\lim_n \bigcup_{k\ge n} A_k) = \lim_n \mu(\bigcup_{k\ge n} A_k) \le \lim_n \sum_{k\ge n} \mu(A_k) = 0.$$

Ebenfalls aus der Stetigkeit von Maßfunktionen folgt

Satz 2.19

μ sei ein Maß auf einem Sigmaring. Dann gilt

$$\mu(\liminf A_n) \leq \liminf \mu(A_n).$$

Ist μ endlich, dann gilt

$$\mu(\limsup A_n) \geq \limsup \mu(A_n).$$

Zum Borel-Cantelli Lemma gibt es eine Umkehrung, für deren Formulierung wir uns ein wenig näher mit Wahrscheinlichkeitsräumen auseinandersetzen müssen.

Zuerst drehen wir ein wenig an der Notation. Um darauf hinzuweisen, dass wir es mit einem Wahrscheinlichkeitsmaß zu tun haben, verwenden wir statt μ den Buchstaben $\mathbb{P}$ als Bezeichnung. Die Elemente des Sigmarings, auf dem $\mathbb{P}$ definiert ist, nennen wir *Ereignisse* und definieren:

Definition 2.15

A und B seien zwei Ereignisse mit $\mathbb{P}(B) > 0$. Dann ist die *bedingte Wahrscheinlichkeit* von A unter (der Bedingung) B

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

Zur Motivation dieser Definition kann die frequentistische Betrachtungsweise dienen: wir glauben vorübergehend daran, dass, wenn wir eine große Zahl N von gleichartigen unabhängigen Versuchen durchführen, die Anzahl N_A der Versuche, bei denen das Ereignis A eintritt, sehr nahe an $N\mathbb{P}(A)$ liegt (so, dass der Quotient N_A/N gegen $\mathbb{P}(A)$ konvergiert). Wir fragen uns, wie sich unsere Einschätzung der Wahrscheinlichkeit ändert, wenn wir zusätzlich wissen, dass das Ereignis B eingetreten ist. Das bedeutet, dass der boebachtete Versuch zu den (ungefähr) $N\mathbb{P}(B)$ Versuchen gehört, bei denen $\mathbb{B}$ eintritt. Dass dann auch noch A beobachtet wird, ist bei den Versuchen der Fall, bei denen A und B gleichzeitig eintreten, und davon gibt es (ungefähr) $N\mathbb{P}(A \cap B)$ Stück. Das gibt uns den Quotienten der beiden Zahlen $N\mathbb{P}(A \cap B)$ und $N\mathbb{P}(B)$ als Wahrscheinlichkeit für A unter der Bedingung B (diese Herleitung erhebt natürlich keinerlei Anspruch darauf, als mathematisches Argument durchzugehen).

Ist die bedingte Wahrscheinlichkeit gleich der "unbedingten" $\mathbb{P}(A)$, dann heißt A unabhängig von B. Ausmultiplizieren der entsprechenden Gleichung ergibt eine symmetrische Beziehung, in der wir auch die Bedingung $\mathbb{P}(B) > 0$ weglassen können:

Definition 2.16

Die Ereignisse A und B heißen unabhängig, wenn

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Für mehr als zwei Ereignisse haben wir

Definition 2.17

1. Die Ereignisse $(A_i, i \in I)$ heißen *(vollständig) unabhängig*, wenn für alle $J \subseteq I$ mit $|J| < \infty$
$$\mathbb{P}(\bigcap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i).$$
2. Die Ereignisse $(A_i, i \in I)$ heißen *paarweise unabhängig*, wenn für alle $i \neq j \in I$
$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

Auch für Ereignisse, die nicht unabhängig sind, lässt sich die Wahrscheinlichkeit des Durchschnitts als Produkt darstellen (zumindest, wenn sie positiv ist):

Satz 2.20: Multiplikationstheorem

$A_1, \ldots, A_n$ seien Ereignisse mit $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. Dann gilt

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\ldots\mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Der Beweis ergibt sich aus der Definition der bedingten Wahrscheinlichkeit durch Ausmultiplizieren und wiederholtes Einsetzen.

In der ausmultiplizierten Form $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$ kann im Gegensatz zur Definition der bedingten Wahrscheinlichkeit auch $\mathbb{P}(A) = 0$ eingesetzt werden, dabei ist sogar egal, was für die bedingte Wahrscheinlichkeit eingesetzt wird. Damit dieses Vorgehen Sinn macht, muss ein sinnvoller Wert für die bedingte Wahrscheinlichkeit aus anderen Überlegungen stammen, wie im folgenden Beispiel

Beispiel 2.1

Ein Würfel wird einmal geworfen, anschließend werden eine schwarze und so viele weiße Kugeln, wie die Zahl auf dem Würfel angibt, in eine Urne gelegt und eine Kugel zufällig gezogen. Wir fragen nach der Wahrscheinlichkeit, dass die schwarze Kugel gezogen wird.

Wir bezeichnen mit A_n das Ereignis, dass n gewüfelt wird, und mit B das Ereignis, dass die schwarze Kugel gezogen wird. Unsere Beschreibung ergibt die bedingte Wahrscheinlichkeit

$$\mathbb{P}(B|A_n) = \frac{1}{n+1},$$

weil wir annehmen, dass jede der $n+1$ Kugeln in der Urne mit gleicher Wahrscheinlichkeit gezogen wird. Wir benötigen nur die Fälle $n = 1, \ldots, 6$, aber wir können auch mit $n > 6$ arbeiten, wenn wir dafür $\mathbb{P}(A_n) = 0$ (oder genauer $A_n = \emptyset$, das "unmögliche Ereignis") setzen.

Die Wahrscheinlichkeit von B erhalten wir als

$$\mathbb{P}(B) = \mathbb{P}(\Omega \cap B) = \mathbb{P}(\bigcup_{n=2}^{6}(A_n \cap B)) = \sum_{n=2}^{6}\mathbb{P}(A_n \cap B) = \sum_{n=2}^{6}\mathbb{P}(A_n)\mathbb{P}(B|A_n) =$$

$$\frac{1}{6}\cdot\frac{1}{2} + \frac{1}{6}\cdot\frac{1}{3} + \frac{1}{6}\cdot\frac{1}{4} + \frac{1}{6}\cdot\frac{1}{5} + \frac{1}{6}\cdot\frac{1}{6} + \frac{1}{6}\cdot\frac{1}{7} = \frac{223}{840}.$$

Das Vorgehen, das wir zur Berechnung von $\mathbb{P}(B)$ verwendet haben, lässt sich auch in allgemeiner Form anwenden: wir nehmen an, dass (endlich oder abzählbar viele) Ereignisse B_i gegeben sind, von denen genau eines eintritt, und dass wir sowohl die Wahrscheinlichkeiten der Ereignisse B_i und die bedingten Wahrscheinlichkeiten eines weiteren Ereignisses A bezüglich jedes dieser Ereignisse kennen (oder leicht berechnen können). Dann gilt

Satz 2.21: Satz von der vollständigen Wahrscheinlichkeit

B_i seien disjunkte Ereignisse mit $\mathbb{P}(B_i) > 0$ und $\bigcup_i B_i = \Omega$ und A ein beliebiges Ereignis. Dann gilt

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i).$$

Wir können dann umgekehrt die bedingte Wahrscheinlichkeit ausrechnen, dass eines der Ereignisse B_i eingetreten ist, wenn A beobachtet wurde:

Satz 2.22: Satz von Bayes

Unter denselben Voraussetzungen wie im vorigen Satz gilt (wenn der Nenner positiv ist)

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\sum_i \mathbb{P}(B_i)\mathbb{P}(A|B_i)}.$$

Beispiel 2.2: Blutgruppen

Wir wenden diesen Satz auf eine Frage an, die mich einige Zeit beschäftigt hat: ich habe lange Zeit meine Blutgruppe nicht gekannt, aber die meiner Frau (A) und die meines Sohnes (0). Wir wollen also nun die bedingte Wahrscheinlichkeit für die möglichen Ausprägungen meiner Blutgruppe bestimmen. Dazu müssen wir zuerst einige Informationen zu den Blutgruppen einholen: wie jede genetisch bedingte Eigenschaft wird auch die Blutgruppe durch zwei Gene (eins vom Vater, eins von der Mutter) bestimmt. Diese können die Ausprägungen a, b und o besitzen. Bei einem zufällig gewählten Menschen nehmen die beiden Gene unabhängig voneinander die einzelnen Werte mit Wahrscheinlichkeiten p_a, p_b und p_0 an. Die Kombinationen aa, ao und oa resultieren in Blutgruppe A, bb, $b0$ und $0b$ ergeben B, ab und ba AB und schließlich oo 0. Pschyrembels medizinisches Wörterbuch liefert, dass in Europa 47% der Bevölkerung Blutgruppe A haben, 9% B, 4% AB und 40% 0. Daraus ergibt sich (ungefähr, die Gleichungen sind überbestimmt und nicht exakt zu lösen)

$$p_a = 0.300, p_b = 0.067, p_0 = 0.633.$$

Aus den Informationen über die Blutgruppen meiner Frau und meines Sohnes können wir folgern, dass mein Sohn von mir ein Gen o haben muss. Meine Ausstattung muss also entweder oa/ao (wir bezeichnen dieses Ereignis mit A_a), ob/bo (Ereignis A_b) oder oo (Ereignis A_0) sein. Die a-priori-Wahrscheinlichkeiten dafür sind $2p_ap_0$, $2p_bp_0$ und p_0^2. In den ersten beiden Fällen ist die bedingte Wahrscheinlichkeit für die Weitergabe einer o (Ereignis B) 1/2, im letzten 1.

Jetzt haben wir alles beisammen, was in den Satz von Bayes einzusetzen ist, und erhalten

$$\mathbb{P}(A_a|B) = \frac{\frac{1}{2}2p_ap_0}{\frac{1}{2}2p_ap_0 + \frac{1}{2}2p_bp_0 + 1p_0^2} = .300$$

und analog

$$\mathbb{P}(A_b|B) = .067$$

und

$$\mathbb{P}(A_0|B) = .633.$$

Wir würden also am ehesten darauf wetten, dass ich Blutgruppe 0 habe. Inzwischen habe ich mich natürlich pieksen lassen (und für die Bestimmung geblecht), und es wäre schön zu berichten, dass unsere Analyse uns zur korrekten Vermutung geführt hat, aber solche Bilderbuchenden gibt es nicht immer — ich darf mich über Blutgruppe A freuen.

Für unabhängige Ereignisse gilt

Satz 2.23: Zweites Lemma von Borel-Cantelli

$\mathbb{P}$ sei ein Wahrscheinlichkeitsmaß auf der Sigmaalgebra $\mathfrak{S}$ und $(A_n \in \mathfrak{S}, n \in \mathbb{N})$ unabhängig mit

$$\sum_n \mathbb{P}(A_n) = \infty.$$

Dann gilt $\mathbb{P}(\limsup_n A_n) = 1$.

Zum Beweis stellen wir fest, dass

1. die Familie $(A_i, i \in I)$ genau dann unabhängig ist, wenn $(A_i^C, i \in I)$ unabhängig ist, und

2. Aus der Stetigkeit von oben folgt, dass sich auch die Wahrscheinlichkeit eines abzählbaren Durchschnitts von unabhängigen Ereignissen als das Produkt der einzelnen Wahrscheinlichkeiten schreiben lässt.

Damit ergibt sich

$$\mathbb{P}(\limsup A_n) = 1 - \mathbb{P}(\liminf A_n^C) = 1 - \lim_{n\to\infty} \prod_{k=n}^{\infty} (1 - \mathbb{P}(A_k)) = 1 - 0 = 1.$$

Der Begriff der Unabhängigkeit lässt sich auf andere Objekte übertragen. Insbesondere können wir die Unabhängigkeit von Sigmaalgebren definieren:

Definition 2.18

$(\Omega, \mathfrak{S}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, I eine beliebige Indexmenge, $(\mathfrak{S}_i, i \in I)$ eine Familie von Untersigmaalgebren von $\mathfrak{S}$. Diese Familie heißt unabhängig, wenn für $n \in \mathbb{N}$, $\{i_1, \ldots, i_n\} \subseteq I$ und $A_j \in \mathfrak{S}_{i_j}, j = 1, \ldots, n$

$$\mathbb{P}(\bigcap_{j=1}^{n} A_j) = \prod_{j=1}^{n} \mathbb{P}(A_j).$$

Zum Nachweis der Unabhängigkeit muss man nicht alle Elemente der Sigmaalgebren untersuchen. Dazu zeigen wir erst einen allgemeinen Satz zum Vergleich von Maßen:

Satz 2.24

$(\Omega, \mathfrak{S})$ sei ein Messraum, $\mathfrak{C}$ ein durchschnittstabiles Erzeugendensytem von $\mathfrak{S}$, μ_1 und μ_2 Maße auf $\mathfrak{S}$.

1. Wenn die Wahrscheinlichkeitsmaße μ_1 und μ_2 auf $\mathfrak{C}$ übereinstimmen, dann gilt $\mu_1 = \mu_2$.
2. Wenn die endlichen Maße μ_1 und μ_2 auf $\mathfrak{C}$ übereinstimmen und es $A_n \in \mathfrak{C}, n \in \mathbb{N}$ gibt mit $\bigcup_n A_n = \Omega$ gibt, dann gilt $\mu_1 = \mu_2$.

Es ist leicht einzusehen (das entsprechende Argument hat uns zur Definition von Dynkin-Systemen geführt), dass die Menge

$$\mathfrak{D} = \{A \in \mathfrak{S} : \mu_1(A) = \mu_2(A)\}$$

bezüglich der Bildung von disjunkten Vereinigungen und Differenzen von Mengen, bei denen eine in der anderen enthalten ist, abgeschlossen ist, also zumindest ein Dynkin-System im weiteren Sinn. Sind μ_1 und μ_2 Wahrscheinlichkeitsmaße, dann gilt $\Omega \in \mathfrak{D}$, und $\mathfrak{D}$ ist Dynkin im engeren Sinn. Als Dynkin-System, das von einem durchschnittstabilen System erzeugt wird, ist $\mathfrak{D}$ ein Sigmaring; in beiden Fällen ist $\Omega \in \mathfrak{D}$, im ersten Fall wegen $\mu_1(\Omega) = \mu_2(\Omega) = 1$, im zweiten als $\bigcup_n A_n$. $\mathfrak{D}$ ist also die von $\mathfrak{C}$ erzeugte Sigmaalgebra und stimmt daher mit $\mathfrak{S}$ überein. Es gilt also $\mu_1(A) = \mu_2(A)$ für alle $A \in \mathfrak{S}$, und μ_1 und μ_2 sind identisch.

Diese Bedingung gibt uns die folgenden Kriterien fur die Unabhängigkeit von Sigmaalgebren:

Satz 2.25

$(\Omega, \mathfrak{S}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathfrak{S}_i, i \in I)$ eine Familie von Untersigmaalgebren von $\mathfrak{S}$. Für jedes $i \in I$ sei $\mathfrak{C}_i$ ein durchschnittstabiles Erzeugendensystem von $\mathfrak{S}_i$. Wenn für jedes endliche $J \subseteq I$ und $A_j \in \mathfrak{C}_j, j \in J$

$$\mathbb{P}(\bigcap_{j\in J} A_j) = \prod_{j\in J} \mathbb{P}(A_j),$$

dann ist die Familie $(\mathfrak{S}_i)$ unabhängig.

Satz 2.26

Die Untersigmaalgebren $\mathfrak{S}_1, \ldots, \mathfrak{S}_N$ sind unabhängig, wenn für jedes $i = 1, \ldots, n$ ein durchschnittstabiles Erzeugendensystem $\mathfrak{C}_i$ von $\mathfrak{S}_i$ existiert, das eine Folge von Ereignissen enhält, deren Vereinigung Ω ist, und wenn für alle $A_i \in \mathfrak{C}_i, i = 1, \ldots, N$ gilt, dass

$$\mathbb{P}(\bigcap_{i=1}^{N} A_i) = \prod_{i=1}^{N} \mathbb{P}(A_i).$$

Im ersten Satz genügt es, den Fall zu betrachten, dass I endlich ist. In Satz 2.25 können wir $\mathfrak{C}_i$ durch $\tilde{\mathfrak{C}}_i = \mathfrak{C}_i \cup \{\Omega\}$ ersetzen, dann folgt seine Aussage aus Satz 2.26. Um diesen zu beweisen, führen wir die Gute Menge

$$\mathfrak{G} = \{A \in \mathfrak{S}_1 : \mathbb{P}(A \cap \bigcap_{i=2}^{N} A_i) = \mathbb{P}(A) \prod_{i=2}^{N} \mathbb{P}(A_i).$$

Man überzeugt sich leicht, dass $\mathfrak{G}$ Dynkin im weiteren Sinn ist und $\mathfrak{C}_1$ enthält. Deshalb stimmt $\mathfrak{G}$ mit $\mathfrak{S}_1$ überein. Wir können also $\mathfrak{C}_1$ durch $\mathfrak{S}_1$ ersetzen, und es gilt immer noch, dass für alle $A_1 \in \mathfrak{S}_1$ und $A_i \in \mathfrak{C}_i$ für $i = 2, \ldots, N$

$$\mathbb{P}(\bigcap_{i=1}^{N} A_i) = \prod_{i=1}^{n} \mathbb{P}(A_i)$$

gilt. Mit demselben Argument können wir jetzt der Reihe nach auch $\mathfrak{C}_2$, $\mathfrak{C}_3$ usw. durch $\mathfrak{S}_2$ etc. ersetzen, und damit ist der Satz bewiesen.

2.2 Der Fortsetzungssatz

In diesem Abschnitt wollen wir zeigen, dass wir ein Maß auf einem Ring zu einem Maß auf dem erzeugten Sigmaring fortsetzen können. Die zentrale Idee ist eine Variation des Ausschöpfungsprinzips, das schon Archimedes verwendet hat. Wir werden also eine Menge mit Mengen überdecken, deren Maß wir leicht bestimmen können, allerdings verwenden wir im Gegensatz zum klassischen Ausschöpfungsprinzip statt endlichen abzählbare Vereinigungen von Mengen aus dem Ring:

Definition 2.19

μ sei ein Maß auf dem Ring $\mathfrak{R}$. Für jede Menge A definieren wir das von μ erzeugte (induzierte) äußere Maß als

$$\mu^*(A) = \inf\{\sum_{n\in\mathbb{N}} \mu(E_n) : E_n \in \mathfrak{R}, A \subseteq \bigcup_{n\in\mathbb{N}} E_n\}.$$

Dabei setzen wir $\inf \emptyset = \infty$, damit diese Definition für alle Mengen funktioniert.

Das so definierte äußere Maß hat folgende Eigenschaften:

Satz 2.27

1. $\mu^*(\emptyset) = 0$.
2. Nichtnegativität: $\mu^*(A) \geq 0$.
3. Monotonie: Falls $A \subseteq B$, dann gilt $\mu^*(A) \leq \mu^*(B)$.
4. Sigmasubadditivität: Falls $A \subseteq \bigcup_{n\in\mathbb{N}} B_n$, dann gilt

$$\mu^*(A) \leq \sum_{n\in\mathbb{N}} \mu^*(B_n).$$

5. Falls $A \in \mathfrak{R}$, dann ist $\mu^*(A) = \mu(A)$.

Die ersten drei Eigenschaften sind offensichtlich; für die vierte genügt es, den Fall zu betrachten, dass die Reihe auf der rechten Seite endlich ist. Dann ist jedes $\mu^*(B_n)$ endlich, und wir können für jedes $\epsilon > 0$ Mengen E_{nk} aus dem Ring finden, für die

$$\sum_{k\in\mathbb{N}} \mu(E_{nk}) \leq \mu^*(B_n) + \frac{\epsilon}{2^n}$$

und

$$B_n \subseteq \bigcup_{k\in\mathbb{N}} E_{nk}.$$

Dann gilt natürlich

$$A \subseteq \bigcup_{n,k\in\mathbb{N}} E_{nk}$$

und daher

$$\mu^*(A) \leq \sum_{n,k\in\mathbb{N}} \mu(E_{nk}) \leq \sum_{n\in\mathbb{N}} (\mu^*(B_n) + \frac{\epsilon}{2^n}) = \sum_{n\in\mathbb{N}} \mu^*(B_n) + \epsilon.$$

Da ϵ beliebig klein gewählt werden kann, ist die Behauptung bewiesen.

Für die letzte Behauptung des Satzes stellen wir einerseits fest, dass aus Satz 2.14 folgt, dass für $A \in \mathfrak{R}$

$$\mu^*(A) \geq \mu(A)$$

sein muss, und andererseits erhalten wir aus der speziellen Überdeckung

$$E_1 = A,\ E_n = \emptyset,\ n > 1,$$

dass

$$\mu^*(A) \leq \mu(A)$$

ist.

Man muss nicht nur äußere Maßfunktionen betrachten, die von Maßfunktionen auf Ringen erzeugt werden. Dazu definieren wir:

Definition 2.20

Jede Funktion μ^* die für alle $A \subseteq \Omega$ definiert ist und die Eigenschaften 1. bis 4. von Satz 2.27 erfüllt, heißt eine äußere Maßfunktion.

Die erzeugte äußere Maßfunktion μ^* wird natürlich die gesuchte Fortsetzung des Maßes μ sein. Leider ist eine äußere Maßfunktion nur in Ausnahmefällen ein Maß. Die Lösung dieses Problems besteht darin, dass wir den Definitionsbereich einschränken. Wir wollen also gewisse Teilmengen von Ω auszeichnen. die wir "messbar" nennen werden.

Im Sinne des Ausschöpfungsprinzips wollen wir dazu neben dem äußeren Maß auch ein inneres Maß haben und eine Menge messbar nennen, wenn ihr äußeres und ihr inneres Maß übereinstimmen. Die analoge Idee zur Definition des äußeren Maßes, nämlich das innere Maß als Supremum aller Summen der Maße von abzählbar vielen disjunkten Mengen aus dem Ring zu definieren, die in einer Menge enthalten sind, funktioniert nicht so gut, weil diese Funktion mit der übereinstimmt, bei der nur endlich viele Mengen verwendet werden. Die bessere Idee, zumindest wenn μ totalendlich ist, ist

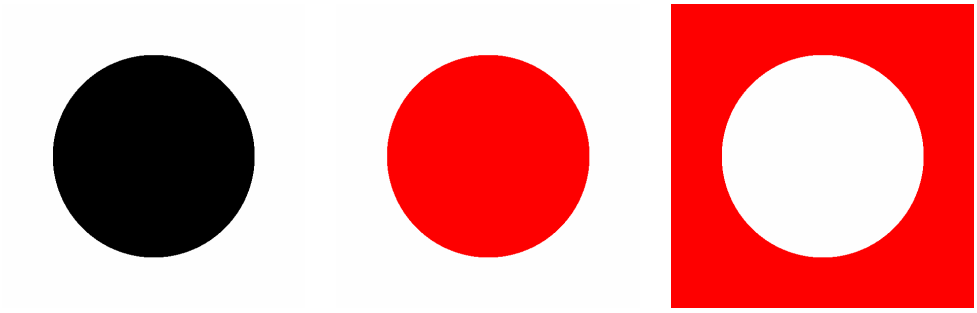
$$\mu_*(A) = \mu(\Omega) - \mu^*(A^C)$$

zu setzen. In diesem Fall reduziert sich die Bedingung

$$\mu^*(A) = \mu_*(A)$$

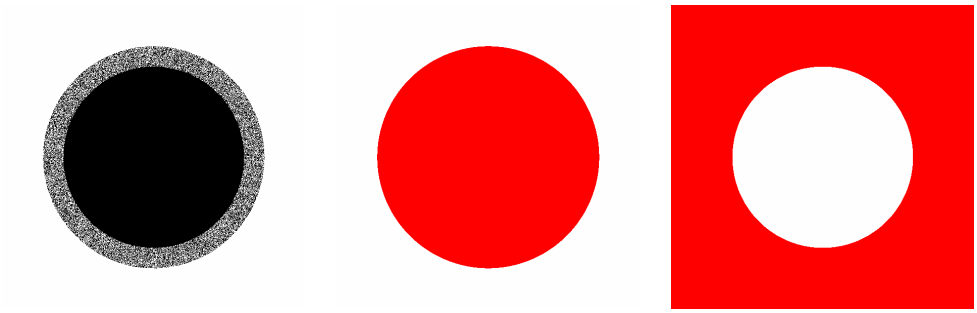
auf

$$\mu^*(A) + \mu^*(A^C) = \mu(\Omega).$$



A $\mu^*(A)$ $\mu^*(A^C)$

Abbildung 2.2: $\mu(\Omega) < \infty$, eine messbare Menge $\mu^*(A) + \mu^*(A^C) = \mu(\Omega)$.



A $\mu^*(A)$ $\mu^*(A^C)$

Abbildung 2.3: $\mu(\Omega) < \infty$, eine nicht messbare Menge $\mu^*(A) + \mu^*(A^C) > \mu(\Omega)$.

Zur Veranschaulichung stellen wir unsere Mengen als Punktmengen in der Ebene dar; das äußere Maß einer Menge A veranschaulichen wir uns als die Fläche einer Menge mit "glattem Rand", die A einschließt. In Abbildung 2.2 ist unsere Vorstellung einer messbaren Menge (mit "glattem Rand") dargestellt, in Abbildung 2.3, wie wir uns eine nicht messbare Menge vorstellen.

Wenn Ω nicht im Ring enthalten ist oder $\mu(\Omega) = \infty$, dann versagt diese Definition der Messbarkeit. In unserer Veranschaulichung geht es darum, dass die Menge A glatten Rand haben soll, also Ω sauber zerschneidet (der "glatte Schnitt", der angeblich nicht wehtut, wie Politiker gerne behaupten) — wir sagen, dass Ω messbar zerlegt wird. Das müssen wir nicht unbedingt für den ganzen Rand überprüfen, sondern nur für ein Stück, das wir von einer Menge aus dem Ring ausschneiden lassen (Abbildung 2.4).

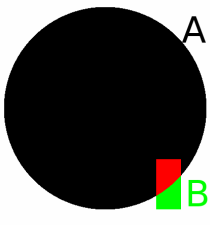


Abbildung 2.4: Eine messbare Menge: eine Testmenge aus dem Ring wird messbar geteilt

Bei näherer Betrachtung stellen wir fest, dass der Rand der Testmenge nicht unbedingt glatt sein muss, uns geht es ja nur darum, dass der Schnitt durch diese Menge sauber ist (Abbildung

2.5).

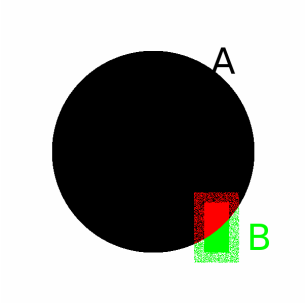


Abbildung 2.5: Eine messbare Menge: eine beliebige Testmenge wird messbar geteilt

Diese sehr heuristischen Überlegungen führen uns zu der folgenden Definition:

Definition 2.21: Messbarkeit nach Carathéodory

Eine Menge $A \subseteq \Omega$ heißt messbar bezüglich der äußeren Maßfunktion μ^*, wenn für jedes $B \subseteq \Omega$ gilt:

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A).$$

Wir sagen, dass A die Menge B messbar zerlegt.

Wir bezeichnen mit $\mathfrak{M}$ das System der messbaren Mengen und zeigen:

Satz 2.28

$\mathfrak{M}$ ist eine Sigmaalgebra, und μ^* ist ein Maß auf $\mathfrak{M}$. Falls darüber hinaus μ^* von dem Maß μ auf dem Ring $\mathfrak{R}$ erzeugt wurde, dann gilt $\mathfrak{R} \subseteq \mathfrak{M}$.

Wir stellen als erstes fest, dass wir für die Messbarkeit einer Menge A nur zeigen müssen, dass für alle $B \subseteq \Omega$

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \setminus A)$$

gilt, weil die umgekehrte Ungleichung aus der Sigmasubadditivität folgt. Insbesondere sind alle Mengen mit $\mu^*(A) = 0$ messbar. Außerdem können wir uns auf Mengen B mit $\mu^*(B) < \infty$ beschränken.

Zuerst zeigen wir, dass $\mathfrak{M}$ eine Algebra ist. Aus der Definition der Messbarkeit folgt unmittelbar, dass für eine messbare Menge A auch ihr Komplement A^C messbar ist. Wir brauchen also nur noch zu zeigen, dass etwa die Vereinigung von zwei messbaren Mengen A_1 und A_2 wieder messbar ist. Wegen der Messbarkeit von A_1 gilt für jedes B

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \setminus A_1),$$

und wegen der Messbarkeit von A_2 gilt weiter

$$\mu^*(B \cap A_1) = \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^C)$$

und

$$\mu^*(B \cap A_1^C) = \mu^*(B \cap A_1^C \cap A_2) + \mu^*(B \cap A_1^C \cap A_2^C)$$

(wir haben die Mengendifferenzen als Durchschnitt mit dem Komplement geschrieben, um uns Klammern zu ersparen).

Wegen der Beziehungen

$$B \cap A_1^C \cap A_2^C = B \setminus (A_1 \cup A_2)$$

und

$$(B \cap A_1 \cap A_2) \cup (B \cap A_1^C \cap A_2) \cup (B \cap A_1 \cap A_2^C) = B \cap (A_1 \cup A_2)$$

erhalten wir aus der Sigmasubadditivität

$$\mu^*(B \cap (A_1 \cup A_2)) \leq \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1^C \cap A_2) + \mu^*(B \cap A_1 \cap A_2^C)$$

und schließlich

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \setminus A_1) =$$
$$\mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^C) + \mu^*(B \cap A_1^C \cap A_2) + \mu^*(B \cap A_1^C \cap A_2^C) \geq$$
$$\mu^*(B \cap (A_1 \cup A_2)) + \mu^*(B \setminus (A_1 \cup A_2)).$$

$A_1 \cup A_2$ ist also messbar.

Als nächstes wollen wir zeigen, dass $\mathfrak{M}$ eine Sigmaalgebra ist. Dazu zeigen wir die Abgeschlossenheit bezüglich einer Vereinigung von abzählbar vielen disjunkten Mengen. Wir beginnen mit zwei disjunkten messbaren Mengen A_1 und A_2 und stellen als direkte Folgerung aus dem Messbarkeitskriterium fest, dass

$$\mu^*(B \cap (A_1 \cup A_2)) = \mu^*(B \cap A_1) + \mu^*(B \cap A_2).$$

Durch vollständige Induktion erhalten wir daraus, dass für eine Folge A_n von disjunkten Mengen und jedes $n > 0$

$$\mu^*(B \cap \bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mu^*(B \cap A_k)$$

gilt. Daraus erhalten wir unter Verwendung der Monotonie von μ^*

$$\mu^*(B) = \mu^*(B \cap \bigcup_{k=1}^{n} A_k) + \mu^*(B \setminus \bigcup_{k=1}^{n} A_k) \geq \sum_{k=1}^{n} \mu^*(B \cap A_k) + \mu^*(B \setminus \bigcup_{k \in \mathbb{N}} A_k).$$

Wir lassen jetzt n gegen Unendlich gehen und erhalten aus der Sigmasubadditivität

$$\mu^*(B) \geq \sum_{k \in \mathbb{N}} \mu^*(B \cap A_k) + \mu^*(B \setminus \bigcup_{k \in \mathbb{N}} A_k) \geq \mu^*(B \cap \bigcup_{k \in \mathbb{N}} A_k) + \mu^*(B \setminus \bigcup_{k \in \mathbb{N}} A_k).$$

Es ist also auch die Vereinigung von abzählbar vielen disjunkten messbaren Mengen eine messbare Menge und somit $\mathfrak{M}$ eine Sigmaalgebra.

Wenn wir in dem obigen Beweis für B die Vereinigung der A_n einsetzen, erhalten wir unmittelbar, dass auch

$$\mu^*(\bigcup_n A_n) = \sum_n \mu^*(A_n)$$

gilt, also dass μ^* auf $\mathfrak{M}$ ein Maß ist.

Wir brauchen also nur noch zu zeigen, dass alle Mengen A aus dem Ring $\mathfrak{R}$ messbar sind, wenn μ^* von dem Maß μ auf $\mathfrak{R}$ erzeugt wird. Dazu sei B eine beliebige Teilmenge von Ω und E_n eine Überdeckung von B durch Mengen aus dem Ring. Dann ist $E_n \cap A$ eine Überdeckung von $B \cap A$ und $E_n \setminus A$ eine Überdeckung von $B \setminus A$ und daher

$$\mu^*(B \cap A) \leq \sum_n \mu(E_n \cap A)$$

und

$$\mu^*(B \setminus A) \leq \sum_n \mu(E_n \setminus A),$$

und wenn wir diese beiden Ungleichungen addieren, erhalten wir

$$\mu^*(B \cap A) + \mu^*(B \setminus A) \leq \sum_n \mu(E_n).$$

Weil diese Ungleichung für alle Überdeckungen gilt, gilt sie auch, wenn wir rechts das Infimum $\mu^*(B)$ über alle Überdeckungen einsetzen, und damit ist gezeigt, dass A messbar ist.

Aus diesem Satz erhalten wir unmittelbar

Satz 2.29: Fortsetzungssatz für Maßfunktionen

μ sei ein Maß auf einem Semiring $\mathfrak{T}$. Dann kann man auf dem von $\mathfrak{T}$ erzeugten Sigmaring ein Maß finden, dass auf $\mathfrak{T}$ mit μ übereinstimmt.

Wenn μ auf $\mathfrak{T}$ sigmaendlich ist, dann ist die Fortsetzung auf den erzeugten Sigmaring eindeutig bestimmt.

Wir haben im vorigen Abschnitt gesehen, dass μ auf den erzeugten Ring fortgesetzt werden kann. Satz 2.28 zeigt, dass durch das erzeugte äußere Maß μ^* eine Fortsetzung auf das System der messbaren Mengen gegeben wird. Dieses ist eine Sigmaalgebra, die den Ring umfasst, also auch den davon erzeugten Sigmaring. Die Einschränkung von μ^* auf den von $\mathfrak{T}$ erzeugten Sigmaring ist also genau die gesuchte Fortsetzung.

Bevor wir die Aussage über die Eindeutigkeit der Fortsetzung beweisen, ein Beispiel dafür, dass diese Eindeutigkeit nicht immer besteht:

Es sei $\Omega = \mathbb{R}$ und $T = \{]a, b] : a \leq b\}$. Der erzeugte Sigmaring ist das System der Borelmengen $\mathfrak{B}$. Wir definieren auf $\mathfrak{B}$ zwei Maßfunktionen:

$$\mu_1(A) = |A|,$$

$$\mu_2(A) = \begin{cases} 0 & \text{falls } A = \emptyset, \\ \infty & \text{sonst.} \end{cases}$$

μ_1 und μ_2 stimmen auf $\mathfrak{T}$ überein, also ist die Fortsetzung der Einschränkung von μ_1 auf $\mathfrak{T}$ zu einer Maßfunktion auf $\mathfrak{B}$ nicht eindeutig bestimmt.

Wir wollen zeigen, dass für jede Fortsetzung $\tilde{\mu}$ von μ auf den Sigmaring $\tilde{\mu} = \mu^*$ gilt.

Da μ sigmaendlich ist, kann jede Menge A aus dem von $\mathfrak{R}$ erzeugten Sigmaring durch abzählbar viele Mengen $E_n \in \mathfrak{R}$ überdeckt werden (man kann sich leicht überzeugen, dass das System aller Mengen, für die es eine solche Überdeckung gibt, einen Sigmaring bildet, der $\mathfrak{R}$ umfasst). Wir können annehmen, dass diese Mengen E_n disjunkt sind, indem wir E_n durch $E_n \setminus \bigcup_{k<n} E_k$ ersetzen.

Wegen

$$\tilde{\mu}(A) = \sum_n \tilde{\mu}(A \cap E_n)$$

brauchen wir nur zu zeigen, dass für jede Menge A aus dem Sigmaring, für die es eine Menge E aus dem Ring mit $A \subseteq E$ und $\mu(E) < \infty$ gibt, $\tilde{\mu}(A) = \mu^*(A)$ gilt. Dazu stellen wir fest, dass jedenfalls

$$\tilde{\mu}(A) \leq \mu^*(A)$$

gelten muss, da ja nach Satz 2.14 für jede Überdeckung von A durch Mengen E_n aus dem Ring

$$\tilde{\mu}(A) \leq \sum_n \tilde{\mu}(E_n) = \sum_n \mu(E_n)$$

gilt.

Daraus folgern wir

$$\mu(E) = \tilde{\mu}(E) = \tilde{\mu}(A) + \tilde{\mu}(E \setminus A) \leq \mu^*(A) + \mu^*(E \setminus A) = \mu^*(E) = \mu(E).$$

In dieser Ungleichungskette muss also überall Gleichheit gelten, und daher gilt, wie behauptet $\tilde{\mu}(A) = \mu^*(A)$.

Es sei hier noch einmal darauf hingewiesen, dass nur die Fortsetzung auf den erzeugten Sigmaring und nicht die auf die erzeugte Sigmaalgebra eindeutig bestimmt ist: in den Übungen wird gezeigt, dass für ein endliches Maß auf einem Sigmaring, der Ω nicht enthält, die Fortsetzung auf die erzeugte Sigmaalgebra niemals eindeutig bestimmt ist.

Nun wollen wir den Zusammenhang zwischen dem System der messbaren Mengen und dem vom Ring $\mathfrak{R}$ erzeugten Sigmaring näher untersuchen (Von jetzt an werden wir die Fortsetzung auf das System der messbaren Mengen ebenfalls mit μ bezeichnen).

Das erste Ergebnis in dieser Hinsicht ist der

Satz 2.30: Approximationssatz

Falls A eine messbare Menge mit endlichem Maß ist, dann gibt es zu jedem $\epsilon > 0$ eine Menge $E \in \mathfrak{R}$, sodass $\mu(A \triangle E) < \epsilon$.

Nach der Definition von μ^* gibt es eine Folge E_n von Mengen aus $\mathfrak{R}$ mit $\sum_{n\in\mathbb{N}} \mu(E_n) \leq \mu(A)+\epsilon$. Da diese Reihe endlich ist, können wir ein N finden, sodass $\sum_{n>N} \mu(E_n) < \epsilon$ ist. Wir setzen $E = \bigcup_{n\leq N} E_n$. Einerseits ist

$$\mu(E \setminus A) \leq \mu(\bigcup_{n\in\mathbb{N}} E_n \setminus A) = \mu(\bigcup_{n\in\mathbb{N}} E_n) - \mu(A) \leq \sum_n \mu(E_n) - \mu(A) \leq \epsilon$$

und

$$\mu(A \setminus E) \leq \mu(\bigcup_{n>N} E_n) \leq \sum_{n>N} \mu(E_n) \leq \epsilon.$$

Insgesamt ergibt sich

$$\mu(A \triangle E) = \mu(A \setminus E) + \mu(E \setminus A) \leq 2\epsilon.$$

Wenn wir durch Mengen aus dem von $\mathfrak{R}$ erzeugten Sigmaring $\mathfrak{S}$ approximieren, ist die Sache noch einfacher:

Satz 2.31

Zu jeder messbaren Menge A mit endlichem Maß gibt es Mengen B und C aus $\mathfrak{S}$ mit $B \subseteq A \subseteq C$ und $\mu(B) = \mu(C) = \mu(A)$.

Wir beginnen wieder mit der Definition des äußeren Maßes, und diesmal wählen wir zu jedem k eine Überdeckung E_{kn} von A mit $\sum_n \mu(E_{kn}) \leq \mu(A) + \frac{1}{k}$. Es ist leicht zu sehen, dass

$$C = \bigcap_k \bigcup_n E_{kn}$$

die gewünschte obere Approximation liefert. Für die Approximation von unten setzen wir $B = C \setminus D$, wobei D eine obere Approximation für $C \setminus A$ ist.

Aus diesem Satz folgt, dass jede messbare Menge mit endlichem Maß sich als

$$A = B \cup N$$

schreiben lässt, wobei B aus dem Sigmaring $\mathfrak{S}$ und N eine Teilmenge einer Menge mit Maß 0 ist. Dasselbe gilt für sigmaendliche Mengen (die man ja in abzählbar viele Mengen mit endlichem Maß zerlegen kann). Umgekehrt ist jede solche Menge messbar, denn N hat äußeres Maß 0, und ist messbar, und B ist als Element von $\mathfrak{S}$ messbar.

Wenn nun das Maß μ auf $\mathfrak{R}$ totalsigmaendlich ist (d.h., wenn Ω durch abzählbar viele Mengen aus dem Ring mit endlichem Maß überdeckt werden kann), dann ist jede messbare Menge in dieser Weise darstellbar. Wir haben dafür einen eigenen Namen:

Definition 2.22

μ sei eine Maßfunktion auf dem Sigmaring $\mathfrak{S}$. Dann heißt $\mathfrak{S}$ vollständig bezüglich μ, wenn jede Teilmenge einer Nullmenge in $\mathfrak{S}$ enthalten ist, also, wenn aus $A \subseteq B \in \mathfrak{S}$ mit $\mu(B) = 0$ $A \in \mathfrak{S}$ folgt.

Wenn der Maßraum $(\Omega, \mathfrak{S}, \mu)$ nicht vollständig ist, kann man ihn vollständig machen:

$$\bar{\mathfrak{S}} = \{B \cup N : B \in \mathfrak{S}, N \subseteq M \in \mathfrak{S}, \mu(M) = 0\}$$

die Vervollständigung von $\mathfrak{S}$ bezüglich μ.

Aus den obigen Ausführungen folgt, dass jedenfalls dann, wenn Ω eine sigmaendliche Menge ist, das System der messbaren Mengen mit der Vervollständigung von $\mathfrak{S}$ bezüglich μ übereinstimmt, also

Satz 2.32

μ sei ein totalsigmaendliches Maß auf dem Ring $\mathfrak{R}$. Dann ist die Fortsetzung von μ auf die erzeugte Sigmaalgebra $\mathfrak{A}_\sigma(\mathfrak{R})$ eindeutig bestimmt und

$$\mathfrak{M}_{\mu^*} = \bar{\mathfrak{A}}_\sigma(\mathfrak{R}).$$

Und das führt uns zurück zu unserer etwas heuristischen Veranschaulichung: diese ist nun gar nicht mehr so weit von der Realität entfernt. Wenn wir die "schönen" Mengen, deren Maß mit dem äußeren Maß einer Menge übereinstimmen soll, aus dem vom ursprünglichen Ring erzeugten Sigmaring wählen. Insbesondere sind die Schritte, die uns zur Definition der Carathéodory-Messbarkeit geführt haben, in den entsprechenden Situationen durchaus korrekt (und haben in der historischen Entwicklung ihren Platz):

Satz 2.33

Wenn μ^* von dem Maß μ auf dem Ring $\mathfrak{R}$ erzeugt wird, dann ist A genau dann messbar, wenn für alle $B \in \mathfrak{R}$

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A).$$

Wenn in dieser Situation (also μ^* ein induziertes äußeres Maß) $\mu^*(\Omega)$ endlich ist, dann ist A genau dann messbar, wenn

$$\mu^*(\Omega) = \mu^*(A) + \mu^*(A^C).$$

Der Beweis wird in den Übungen geführt.

Wir sind nun dort angelangt, wo wir hinwollten: zumindest für ein sigmaendliches Ausgangsmaß können wir ein (eindeutig bestimmtes) Maß auf dem ganzen Sigmaring (der hoffentlich auch Ω enthält und somit eine Sigmaalgebra ist) erhalten, wenn wir es nur auf einem Semiring festlegen, der diesen Sigmaring erzeugt.

Weil wir Maße vorzugsweise auf Sigmaalgebren verwenden wollen und wir von einem Semiring oder Ring immer zur erzeugten Sigmaalgebra kommen können, haben wir die Definitionen:

Definition 2.23

- Ω sei eine Menge, $\mathfrak{S}$ eine Sigmaalgebra über Ω. Dann heißt das Paar $(\Omega, \mathfrak{S})$ ein *Messraum* (oder *messbarer Raum*).
- $(\Omega, \mathfrak{S})$ sei ein Messraum, und zusätzlich sei auf $\mathfrak{S}$ ein Maß μ definiert. Dann heißt das Tripel $(\Omega, \mathfrak{S}, \mu)$ ein *Maßraum*. Ist das Maß μ endlich oder sigmaendlich bzw. ein Wahrscheinlichkeitsmaß, dann sprechen wir von einem endlichen oder sigmaendlichen Maßraum bzw. einem Wahrscheinlichkeitsraum.

2.3 Ergänzungen

2.3.1 Topologische Räume und Borelmengen

Für die Borelmengen gibt es neben den halboffenen Intervallen einige andere interessante Erzeugendensysteme, etwa

- $\mathfrak{T} = \{]-\infty, a], a \in \mathbb{R}\}$,
- $\mathfrak{T} = \{]-\infty, a], a \in \mathbb{Q}\}$,
- $\mathfrak{T} = \{]a, b[: a, b \in \mathbb{R}, a < b\}$,
- $\mathfrak{T} = \{U \subseteq \mathbb{R} : U \text{ offen}\}$.

Der letzte Punkt erlaubt es uns, Borelmengen für allgemeine topologische Räume zu definieren:

Definition 2.24

Ω sei eine beliebige Menge. Eine Topologie über Ω ist ein Mengensystem $\tau \subseteq 2^\Omega$ mit den Eigenschaften

1. $\{\emptyset, \Omega\} \subseteq \tau$.
2. Wenn $A, B \in \tau$, dann ist $A \cap B \in \tau$.
3. Wenn $\sigma \subseteq \tau$, dann ist $\bigcup_{A\in\sigma} A \in \tau$.

Die Elemente von τ werden als die offenen Mengen dieser Topologie bezeichnet.

Ein topologischer Raum ist ein Paar (Ω, τ) aus einer Grundmenge Ω und einer Topologie τ über Ω.

Die bekannten metrischen Räume (speziell die reellen Zahlen) passen in diese Definition, indem wie gewohnt die Mengen U als offen bezeichnet werden, die zu jedem $x \in U$ eine ganze Kugel

$$B(x, \epsilon) = \{y : d(x, y) < \epsilon\}$$

mit einem $\epsilon = \epsilon(x) > 0$ enthalten.

Bekannte Definitionen aus der Welt der metrischen Räume haben Entsprechungen in dieser allgemeineren Theorie:

Definition 2.25

(Ω, τ) sei ein topologischer Raum.

- $U \subseteq \Omega$ heißt Umgebung von $x \in \Omega$, wenn es $V \in \tau$ gibt mit $x \in V \subseteq \Omega$ (d.h., wenn U eine offene Menge enthält, in der x enthalten ist).
- Eine Menge $A \subseteq \Omega$ heißt abgeschlossen, wenn ihr Komplement offen ist, d.h., wenn $A^C \in \tau$.
- Der Abschluss $\bar{A}$ der Menge $A \subseteq \Omega$ ist die kleinste abgeschlossene Menge, die A enthält:

$$\bar{A} = \bigcap_{B\supseteq A : B^C\in\tau} B.$$

- (Ω', τ') sei ein weiterer topologischer Raum. $f : \Omega \to \Omega'$ heißt stetig, wenn die Urbilder von offenen Mengen offen sind, also wenn $f^{-1}(\tau') \subseteq \tau$.
- Die Folge (x_n), $x_n \in \Omega$, konvergiert gegen $x \in \Omega$, wenn in jeder Umgebung U von x fast alle Elemente der Folge (x_n) liegen (d.h., alle Elemente bis auf endlich viele, was wieder bedeutet, dass es ein n_0 gibt, sodass $x_n \in U$ für alle $n \geq n_0$ gilt).

Definition 2.26

(Ω, τ) sei ein topologischer Raum. Die Borelsche Sigmaalgebra $\mathfrak{B}(\Omega, \tau)$ (auch $\mathfrak{B}(\Omega)$, $\mathfrak{B}(\tau)$ oder einfach nur $\mathfrak{B}$, wenn keine Verwechslungsgefahr besteht) ist die kleinste Sigmaalgebra, die τ umfasst.

$$\mathfrak{B}(\Omega, \tau) = \mathfrak{A}_\sigma(\tau).$$

Leider sind in allgemeinen topologischen Räumen konvergente Folgen nicht ausreichend, um ihre Topologie zu beschreiben. Dazu muss zuerst definiert werden, wie wir dieses "Beschreiben" verstehen wollen. Dazu versehen wir die entsprechenden Definitionen mit neuen Namen:

Definition 2.27

(Ω, τ) sei ein topologischer Raum.

1. $A \subseteq \Omega$ heißt folgenoffen, wenn für jedes $x \in A$ und jede Folge (x_n) mit $x_n \to x$ fast alle Elemente von (x_n) in A liegen.
2. $A \subseteq \Omega$ heißt folgenabgeschlossen, wenn für jedes $x \in \Omega$, für das es eine Folge (x_n) mit $x_n \in A$ und $x_n \to x$ gibt, auch $x \in A$ gilt.
3. (Ω', τ') sei ein weiterer topologischer Raum. Dann heißt $f : \Omega \to \Omega'$ folgenstetig in $x \in \Omega$, wenn für jede Folge x_n mit $x_n \to x$ auch $f(x_n) \to f(x)$ (in Ω') gilt.
4. (Ω', τ') sei ein weiterer topologischer Raum. Dann heißt $f : \Omega \to \Omega'$ folgenstetig, wenn f in allen $x \in \Omega$ folgenstetig ist.
5. $A \subseteq \Omega$ sei beliebig. Der Folgenabschluss $(A)_{seq}$ ist die Menge aller Grenzwerte von Folgen in A:
$$(A)_{seq} = \{x \in \Omega : \exists x_n \in A : x_n \to x\}.$$

Das gibt uns die beiden Möglichkeiten:

Definition 2.28

Ein topologischer Raum (Ω, τ) heißt

- folgenbestimmt, wenn jede folgenabgeschlossene Menge abgeschlossen ist (das ist äquvialent dazu, dass jede folgenoffene Menge offen bzw. jede folgenstetige Funktion stetig ist).
- ein Fréchet-Urysohn Raum, wenn für jedes $A \subseteq \Omega$ ihr Folgenabschluss mit dem topologischen Abschluss übereinstimmt, also wenn $\bar{A} = (A)_{seq}$ git.

Ein Beispiel für einen nicht folgenbestimmten Raum liefert der Raum $\mathbb{R}^{\mathbb{R}}$ der reellen Funktionen mit der Produkttopologie:

Definition 2.29

$((\Omega_i, \tau_i), i \in I)$ sei eine Familie von topologischen Räumen. Die Produkttopologie ist die kleinste Topologie bezüglich der alle Projektionen

$$\pi_i : \bigtimes_{j \in I} \Omega_j$$

mit

$$\pi_i(\alpha) = \alpha(i)$$

stetig sind. In dieser Topologie ist eine Menge U genau dann offen, wenn zu jedem $\alpha \in U$ Mengen $A_i \in \tau_i, i \in I$ gibt mit $\alpha(i) \in A_i$ für alle $i \in I$ und $A_i = \Omega_i$ für alle $i \in I$ bis auf endlich viele.

In dieser Topologie konvergiert eine Folge α_n genau dann gegen α, wenn für jedes i $\alpha_n(i) \to \alpha(i)$ (in (Ω_i, τ_i)). Die Produkttopologie heißt auch die Topologie der punktweisen Konvergenz.

Später werden wir uns mit der punktweisen Konvergenz von Funktionenfolgen noch näher beschäftigen. Hier ist interessant, dass dieser Raum nicht folgenbestimmt ist. Es ist nämlich die Menge

$$A = \{f \in \mathbb{R}^{\mathbb{R}} : [f \neq 0] \text{ ist höchstens abzählbar}\}$$

folgenabgeschlossen, aber nicht abgeschlossen. Ist nämlich (f_n) eine Folge von Funktionen aus A, dann ist zu jedem n die Menge

$$N_n = [f_n \neq 0]$$

höchstens abzählbar, damit auch $N = \bigcup_{n\in\mathbb{N}} N_n$. Wenn nun f_n punktweise gegen f konvergiert, dann ist auf N^C jedes f_n gleich 0, also auch der Grenzwert f. $[f \neq 0]$ ist also auch höchstens abzählbar und daher $f \in A$ und A folgenabgeschlossen.

Andererseits enthält jede nichtleere offene Menge U eine Menge der Form

$$\bigtimes_{i\in I} A_i$$

mit $J = \{i \in I : A_i \neq \mathbb{R}\}$ endlich. Dann ist α mit

$$\alpha(i) = \begin{cases} a_i \in A_i & \text{wenn } i \in J \\ 0 & \text{sonst} \end{cases}$$

in $A \cap U$ enthalten. Deswegen ist $\emptyset$ die einzige offene Menge, die in A^C enthalten ist, und somit $\bar{A} = \mathbb{R}^{\mathbb{R}}$ und A nicht abgeschlossen.

2.3.2 Alternativer Beweis des monotone class theorems

Es soll gezeigt werden, dass das von einem Ring erzeugte monotone System mit dem erzeugten Sigmaring übereinstimmt.

Wir können das erzeugte monotone System erzeugen, indem wir schrittweise die Limiten von monotonen Folgen zu $\mathfrak{R}$ hinzufügen:

$$\mathfrak{M}_0 = \mathfrak{R}, \mathfrak{M}_{2n+1} = \{B \subseteq \Omega : \exists B_n \in \mathfrak{M}_{2n} : B_n \uparrow B\}, \mathfrak{M}_{2n+2} = \{B \subseteq \Omega : \exists B_n \in \mathfrak{M}_{2n+1} : B_n \downarrow B\}.$$

Es ist leicht einzusehen, dass jedes $\mathfrak{M}_n$ bezüglich endlicher Durchschnitte und Vereinigungen abgeschlossen ist. Für die Differenz ist das nicht ganz so einfach, aber es lässt sich zeigen, dass $\mathfrak{M}_{2n}$ alle Differenzen von Mengen aus $\mathfrak{M}_n$ enthält. Jedenfalls ist

$$\mathfrak{M}_\omega = \bigcup_{n\in\mathbb{N}} \mathfrak{M}_n$$

ein Ring. Leider ist noch nicht garantiert, dass dieses System monoton ist. Aber wir machen unerschrocken weiter, geben jeweils monotone Limiten zu unseren Mengensystemen hinzu, und wenn wir unendlich viele beisammen haben, vereinigen wir sie...

Irgendwann kommen wir mit diesem Schema (einer transfiniten Induktion) zu der ersten überabzählbaren Ordinalzahl ω_1. Dann sind wir fertig, und unsere Konstruktion garantiert, dass das so erzeugte monotone System ein Ring ist. Diese Argumentation hat zwei Nachteile: einerseits ist sie nicht besonders anschaulich, und andererseits wollen wir das Auswahlaxiom nach Möglichkeit vermeiden.

Aber dieser Beweis hat auch einen Vorteil: wenn $\mathfrak{R}$ unendlich aber höchstens von der Mächtigkeit des Kontinuums ist, dann hat das erzeugte monotone System die Mächtigkeit des Kontinuums (Jedes C_α hat eine Mächtigkeit, die nicht größer ist als die des Kontinuums: falls α eine Nachfolgerzahl ist, als Bild von $C_{\alpha-1}^{\mathbb{N}}$, falls α eine Limeszahl ist, als Vereinigung von höchstens Kontinuum vielen Mengen mit Mchtigkeit höchstens gleich dem Kontinuum). Insbesondere gilt das für die Borelmengen.

2.3.3 Semiregularität

Endlichkeitseigenschaften werden an vielen Stellen in unserer weiteren Diskussion eine Rolle spielen; manchmal müssen wir Endlichkeit verlangen, sehr oft kommen wir mit der Sigmaendlichkeit aus; nicht sigmaendliche Maße haben hier hauptsächlich eine Rolle als Gegenbeispiele. Extreme Beispiele sind Maße, die nur die Werte 0 und ∞ annehmen. Als Antithese dazu dient die folgende Abschwächung von Endlichkeit bzw. Sigmaendlichkeit:

Definition 2.30: Semiregularität

Das Maß μ auf dem Mengensystem $\mathfrak{C}$ heißt *semiregulär*, wenn es zu jedem $A \in \mathfrak{C}$ mit $\mu(A) > 0$ ein $B \in \mathfrak{C}$ gibt mit $B \subseteq A$ und $0 < \mu(B) < \infty$ (also, wenn jede Menge mit positivem Maß

eine Menge mit positivem endlichem Maß enthält).

2.3.4 Äußere Maße und die Borelmengen

Im nächsten Kapitel werden wir uns mit dem Spezialfall von Maßen auf den Borelmengen befassen. Diese wollen wir eindeutig festlegen, indem wir sie auf den Intervallen definieren. Dazu sollen sie auf dem System der (linkshalboffenen) Intervalle sigmaendlich sein. Wir werden sogar annehmen, dass sie dort endlich sind. Vorher wollen wir uns aber noch überlegen, unter welchen Bedingungen für ein frei definiertes äußeres Maß μ^* auf den reellen Zahlen alle Borelmengen μ^*-messbar sind.

Dazu definieren wir

Definition 2.31

Der Abstand zweier Mengen $A, B \in \mathbb{R}$ ist

$$d(A,B) = \inf\{|x-y| : x \in A, y \in B\}.$$

und

Definition 2.32

Das äußere Maß μ^* heißt arithmetisch, wenn für alle $A, B \in \mathbb{R}$ mit $d(A,B) > 0$ $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ gilt.

Es gilt der Satz:

Satz 2.34: Carathéodory

Für ein äußeres Maß μ^* auf $\mathbb{R}$ gilt genau dann $\mathfrak{B} \subseteq \mathfrak{M}_{\mu^*}$, wenn μ^* arithmetisch ist.

Eine Richtung ist einfach: wenn alle Borelmengen messbar sind, dann jedenfalls die offenen Mengen. Wenn A und B positiven Abstand $d(A,B)$ haben, dann ist

$$U = \bigcup_{x \in A} \{y : |y-x| < d(A,B)\}$$

offen, $A = (A \cup B) \cap U$, $B = (A \cup B) \setminus U)$ und daher, wegen der Messbarkeit von U, $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$.

Für die andere Richtung nehmen wir an, dass μ^* arithmetisch ist, und zeigen, dass dann jede abgeschlossene Menge A messbar ist.

Wir wählen also eine Menge B mit endlichem äußerem Maß und wollen zeigen, dass B durch A messbar zerlegt wird. Wir definieren

$$C_n = \{x : d(\{x\}, A) > 1/n\}$$

und stellen fest, dass wegen der Abgeschlossenheit von A

$$A^C = \bigcup_n C_n$$

gilt.

Weil der Abstand von A und C_n positiv ist, gilt

$$\mu^*(B) \geq \mu * (B \cap (A \cup C_n)) \geq \mu^*(B \cap A) + \mu^*(B \cap C_n).$$

Wenn wir hier $n \to \infty$ gehen lassen könnten und dabei $\mu^*(B \cap C_n)$ gegen $\mu^*(B \cap A^C)$ ginge, wäre alles bewiesen. Wir setzen $C_0 = \emptyset$ und

$$D_n = B \cap (C_n \setminus C_{n-1}).$$

Da $d(D_n, D_{n+2}) > 0$ ist, erhalten wir

$$\sum_{n=1}^{N} \mu^*(D_{2n}) \leq \mu^*(B) < \infty$$

und

$$\sum_{n=1}^{N} \mu^*(D_{2n-1}) \leq \mu^*(B) < \infty$$

wenn wir beides addieren und $N \to \infty$ gehen lassen, ergibt sich

$$\sum_{n\in\mathbb{N}} \mu^*(D_n) < \infty.$$

Damit erhalten wir

$$\mu(B \cap C_N) \leq \mu^*(B \cap A^C) = \mu^*((B \cap C_N) \cup \bigcup_{n>N} D_n) \leq \mu^*(B \cap C_N) + \sum_{n>N} \mu^*(D_n).$$

Die letzte Summe geht als Reihenrest einer konvergenten Reihe für $N \to \infty$ gegen 0, also ist tatsächlich

$$\mu^*(B \cap A^C) = \lim_{n\to\infty} \mu^*(B \cap C_n),$$

und der Satz bewiesen.

2.3.5 Hausdorffmaße

In $\mathbb{R}^d$ können wir den Durchmesser einer Menge $A \subseteq \mathbb{R}^d$ als

$$d(A) = \sup\{|x - y| : x, y \in A\}$$

definieren. Für $\alpha > 0$ ist

$$\eta^*_\alpha(A) = \lim_{\epsilon\to 0} \inf\{\sum_{n\in\mathbb{N}} d(A_n)^\alpha : A_n \subseteq \mathbb{R}^d, d(A_n) < \epsilon, n \in \mathbb{N}, A \subseteq \bigcup_{n\in\mathbb{N}} A_n\}.$$

Ein äußeres Maß (das wird in den Übungen gezeigt). Die Einschränkung η_α dieses äußeren Maßes auf seine messbaren Mengen nennen wir das Hausdorffmaß. Der vorige Abschnitt zeigt, dass zu diesen messbaren Mengen in jedem Fall die Borelmengen gehören.

2.3.6 Der Satz von Vitali-Hahn-Saks

Satz 2.35: Vitali-Hahn-Saks

μ_n sei eine Folge von endlichen Maßfunktionen auf dem Sigmaring $\mathfrak{R}$. Wenn für jedes $A \in \mathfrak{R}$ der Grenzwert $\mu(A) = \lim_n \mu_n(A)$ existiert, dann ist μ ein Maß.

Wir zeigen diesen Satz indirekt: wir zeigen: wenn die Sigmaadditivität der Grenzfunktion verletzt ist, dann gibt es eine Menge $B \in \mathfrak{R}$, für die die Folge $\mu_n(B)$ nicht konvergiert. Dazu betrachten wir eine Folge A_n von disjunkten Mengen, für die

$$\mu(\sum_n A_n) \neq \sum_n \mu(A_n)$$

gilt. Aus der Additivität von μ folgt

$$\sum_n \mu(A_n) \leq \mu(\sum_n A_n),$$

und weil die Gleichheit nicht gelten darf, ist

$$\sum_n \mu(A_n) < \mu(\sum_n A_n)$$

und insbesondere

$$\sum_n \mu(A_n) < \infty.$$

Wir wählen $\epsilon > 0$ so, dass

$$\sum_n \mu(A_n) + 5\epsilon < \mu(\sum_n A_n)$$

Wir können ohne Beschränkung der Allgemeinheit (indem wir aus den Folgen A_k und μ_n notfalls endlich viele Elemente entfernen) annehmen, dass mit

$$A = \sum_k A_k$$

für alle k bzw. n

$$\sum_k \mu(A_k) < \epsilon$$

und

$$\mu_n(A) > 5\epsilon.$$

Zu jedem n können wir ein $k(n)$ finden, sodass

$$\sum_{k>k_n} \mu_n(A_k) < \epsilon,$$

und zu jedem k ein $n(k)$, sodass für $n \geq n(k)$

$$\sum_{i \leq k} \mu_n(A_i) < \epsilon.$$

Wir beginnen mit $n_1 = 1$ und setzen für $l \geq 1$

$$k_l = k(n_l), n_{l+1} = n(k_l).$$

Jetzt können wir

$$B = \bigcup_{l \in \mathbb{N}} \bigcup_{k_{2l-1} < k \leq k_{2l}} A_k \in \mathfrak{R}$$

setzen und erhalten

$$\mu_{n_{2l}}(B) > 3\epsilon, \mu_{n_{2l-1}}(B) < 2\epsilon,$$

also konvergiert die Folge $\mu_n(B)$ nicht, im Widerspruch zur Voraussetzung.

Dieser Satz gilt auch für signierte, komplexe und Vektormaße, mit einem sehr ähnlichen Beweis. In diesen Fällen muss man aber zusätzlich annehmen, das die Grenzwerte endlich sind (bei Maßfunktionen ist das nicht notwendig).

2.3.7 Der Witz mit den Eiern (huevos, nicht cojones)

Auch wenn letztere von Matthias Strolz fast salonfähig gemacht wurden, soll hier nicht der Eindruck entstehen, dass wir sexistische Metaphern verwenden. Also, hier in aller Harmlosigkeit:

Eine Physikerin und ein Mathematiker wollen Eier kochen.

Fall 1: Die Eier sind Im Kühlschrank.

- Die Physikerin geht zum Kühlschrank, öffnet die Kühlschranktür, nimmt zwei Eier heraus, schließt die Kühlschranktür, geht zum Küchenschrank, nimmt einen Topf heraus, legt die Eier hinein, geht zur Spüle, lässt Wasser in den Topf laufen, geht zum Herd, stellt den Topf darauf, dreht die Platte auf.

– Der Mathematiker geht zum Kühlschrank, öffnet die Kühlschranktür, nimmt zwei Eier heraus, schließt die Kühlschranktür, geht zum Küchenschrank, nimmt einen Topf heraus, legt die Eier hinein, geht zur Spüle, lässt Wasser in den Topf laufen, geht zum Herd, stellt den Topf darauf, dreht die Platte auf.

Fall 2: Die Eier sind Im Keller.

– Die Physikerin geht zur Kellertür, öffnet die Kellertür, geht die Stufen hinunter, nimmt zwei Eier aus dem Kellerregal, geht die Stufen hinauf, schließt die Kellertür, geht in die Küche zum Küchenschrank, nimmt einen Topf heraus, legt die Eier hinein, geht zur Spüle, lässt Wasser in den Topf laufen, geht zum Herd, stellt den Topf darauf, dreht die Platte auf.

– Der Mathematiker geht zur Kellertür, öffnet die Kellertür, geht die Stufen hinunter, nimmt zwei Eier aus dem Kellerregal, geht die Stufen hinauf, schließt die Kellertür, geht in die Küche zum Kühlschrank, öffnet die Kühlschranktür, legt die Eier hinein, schließt die Kühlschranktür und hat so Fall 1 hergestellt.

2.4 Wiederholungsfragen

1. Definieren Sie Ring, Sigmaring, Algebra, Sigmaalgebra, Semiring, monotones System, Dynkin-System.
2. Definieren Sie Mengenfunktion, Additivität, Sigmaadditivität, Maß, Inhalt.
3. Wie erhält man aus einem Semiring den erzeugten Ring?
4. Wie erhält man aus einem Ring den erzeugten Sigmaring?
5. Definieren Sie: Menge mit endlichem Maß, Menge mit sigmaendlichem Maß, endliches Maß, sigmaendliches Maß, Wahrscheinlichkeitsmaß.
6. Was kann man über punktweise Grenzwerte bzw. Summen von Maßen sagen?
7. Was sagt der Fortsetzungssatz für Maßfunktionen?
8. Was ist ein äußeres Maß?
9. Wann heißt eine Menge messbar bezüglich eines äußeren Maßes?
10. Wann heißt ein Maßraum vollständig?
11. Wie ist die Vervollständigung einer Sigmaalgebra bezüglich eines Maßes definiert?
12. Wie hängen die μ^*-messbaren Mengen und die erzeugte Sigmaalgebra $\mathfrak{A}_\sigma(\mathfrak{R})$ für ein total-sigmaendliches Maß auf dem Ring $\mathfrak{R}$ zusammen?
13. Welche Eigenschaften haben Inhalte bzw. Maßfunktionen?
14. Das Additionstheorem
15. Welche Kriterien für die Sigmaadditivität einer additiven Funktion auf einem Ring gibt es?
16. Wie hängt (für sigmaendliche Maßräume) die (vom Ring erzeugte) Sigmaalgebra mit dem System der messbaren Mengen zusammen? zwischen
17. Wie ist die bedingte Wahrscheinlichkeit definiert?
18. Wann heißen zwei Ereignisse unabhängig?
19. Wie kann man die Unabhängigkeit von mehreren Ereignissen definieren?
20. Was besagt der Satz von der vollständigen Wahrscheinlichkeit?

21. Was besagt der Satz von Bayes?
22. Die Lemmata von Borel-Cantelli.
23. Unabhängigkeit von Sigmaalgebren

Kapitel 3

Lebesgue-Stieltjes Maße

3.1 Eindimensionale Lebesgue-Stieltjes Maße

Jetzt ist es an der Zeit, dass wir uns etwas eingehender mit den reellen Zahlen und allgemeiner mit endlichdimensionalen Räumen und den darauf definierten Borelmengen befassen. Insbesondere werden wir zeigen dass mit der Länge (bzw. Fläche und Volumen in höheren Dimensionen) tatsächlich ein Maß auf den Borelmengen gegeben ist. Fürs erste beschränken wir uns auf den eindimensionalen Fall. In diesem werden die Borelmengen $\mathfrak{B}$ vom Semiring der linkshalboffenen Intervalle $\mathfrak{T}$ erzeugt, und wir können jedes sigmaendliche Maß auf $\mathfrak{T}$ in eindeutiger Weise zu einem (ebenfalls sigmaendlichen) Maß auf $\mathfrak{B}$ fortsetzen. Allerdings sind nicht alle sigmaendlichen Maße auf $\mathfrak{B}$ in dieser Weise zu erhalten (man setze beispielsweise als Maß einer Menge die Anzahl der darin enthaltenen rationalen Zahlen), und andererseits gibt es wieder unter den Maßen, die auf $\mathfrak{T}$ sigmaendlich sind, recht unangenehme Zeitgenossen, also werden wir die schlimmsten Anomalien vermeiden, indem wir gleich annehmen, dass unsere Maße auf $\mathfrak{T}$ endlich sind, und definieren:

Definition 3.1

Ein Lebesgue-Stieltjes Maß ist ein Maß auf $(\mathbb{R}, \mathfrak{B})$, das für beschränkte Intervalle endliche Werte annimmt. (Diese Definition funktioniert wortgleich auch in d Dimensionen, wenn wir die Intervalle als die entsprechenden "Hyperquader" interpretieren).

Diese Definition ist natürlich äquivalent zu der Forderung, dass alle beschränkten Borelmengen endliches Maß besitzen. Insbesondere sind alle endlichen Maße Lebesgue-Stieltjes Maße.

Offensichtlich sind alle Lebesgue-Stieltjes-Maße auf $\mathfrak{T}$ eingeschränkt sigmaendlich (sogar endlich), und daher sind sie durch die Angabe ihrer Werte auf $\mathfrak{T}$ eindeutig festzulegen.

Wir müssen also für jedes Intervall $]a, b]$ den Wert des Maßes festlegen. Das bedeutet im wesentlichen, dass wir eine Funktion der beiden Variablen a und b angeben müssen. Diese Aufgabe können wir aber noch weiter vereinfachen (und werden noch dazu feststellen, dass wir damit auch die Überprüfung, ob das Ergebnis ein Maß ist, leichter machen), indem wir die Tatsache ausnutzen, dass für $a < b < c$ sowohl

$$]a, b] \cup]b, c] =]a, c]$$

als auch

$$]a, b] \cap]b, c] = \emptyset.$$

Für ein Lebesgue-Stieltjes Maß μ muss jetzt gelten

$$\mu(]a, c]) = \mu(]a, b]) + \mu(]b, c]),$$

und weil $\mu(]a, b])$ endlich ist,

$$\mu(]b, c]) = \mu(]a, c]) - \mu(]a, b]).$$

Wir werden jetzt kurz zusätzlich annehmen, dass das Maß μ endlich ist (auf $\mathfrak{B}$!). In diesem Fall dürfen wir nämlich oben für a den Wert $-\infty$ einsetzen, und die entsprechende Gleichung gilt dann für alle $b < c$. Wir setzen

$$F(x) = \mu(]-\infty, x])$$

und erhalten

$$\mu(]b, c]) = F(c) - F(b).$$

Zum Festlegen eines endlichen Maßes ist also nur noch die Funktion F von einer reellen Variable anzugeben. Damit hat sich die Dimensionalität unseres Problems noch einmal halbiert, und wir sind endlich zufrieden, abgesehen davon, dass wir uns dasselbe für alle Lebesgue-Stieltjes Maßfunktionen wünschen, weshalb wir zunächst definieren:

Definition 3.2

μ sei ein (eindimensionales) Lebesgue-Stieltjes Maß. Eine reelle Funktion F heißt *Verteilungsfunktion* von μ, wenn für alle $a < b$

$$\mu(]a, b]) = F(b) - F(a).$$

Diese Definition wirft natürlich eine Reihe von Fragen auf:

1. Gibt es zu einer Lebesgue-Stieltjes Maßfunktion immer eine Verteilungsfunktion?
2. Vorausgesetzt, die Antwort auf die vorige Frage ist "ja", wie weit ist dann die Verteilungsfunktion eindeutig bestimmt (keine sehr wichtige Frage, zugegeben, aber nicht uninteressant, und wegen der Vollständigkeit...)?
3. Kann man leicht feststellen, ob eine bestimmte Funktion eine Verteilungsfunktion ist (diese Frage ist die wichtigste — schließlich wollen wir diese Verteilungsfunktionen dazu verwenden, dass wir Maße definieren)?
4. Schließlich: Ist durch eine Verteilungsfunktion das zugehörige Maß eindeutig bestimmt?

Die letzte Frage sollte eigentlich klar sein: durch die Definition der Verteilungsfunktion (in die andere Richtung gelesen) ist das Maß μ auf dem Semiring $\mathfrak{T}$ bestimmt, und den Rest erledigt der Fortsetzungssatz.

Um die erste Frage zu lösen, brauchen wir nur eine Verteilungsfunktion F anzugeben, und wir setzen

$$F(x) = \begin{cases} \mu(]0, x]) & \text{falls } x > 0, \\ -\mu(]x, 0]) & \text{sonst.} \end{cases}$$

Man überzeugt sich leicht, dass dies wirklich eine Verteilungsfunktion für μ ist.

Die Eindeutigkeitsfrage für F ist nicht ganz so leicht zu beantworten. Aus der Definition der Verteilungsfunktion ist nämlich zu erkennen, dass man zu F eine beliebige Konstante addieren kann und damit wieder eine Verteilungsfunktion für dieselbe Maßfunktion erhält; es gibt also zu einer Maßfunktion mehr als eine Verteilungsfunktion; andererseits gilt für zwei Verteilungsfunktionen F, G zu derselben Maßfunktion

$$\mu(]x, y]) = F(y) - F(x) = G(y) - G(x),$$

und wir können die Variablen trennen:

$$F(y) - G(y) = F(x) - G(x).$$

Dieser Ausdruck darf also weder von x noch von y abhängen und ist daher konstant. Zwei Verteilungsfunktionen zu derselben Maßfunktion können sich also auch nicht durch mehr als eine additive Konstante unterscheiden.

Letztlich bleibt uns also die Aufgabe, zu entscheiden, ob eine Funktion eine Verteilungsfunktion ist. Dazu müssen wir die zwei charakteristischen Eigenschaften einer Maßfunktion — Nichtnegativität und Sigmaadditivität — auf unsere Verteilungsfunktionen übertragen.

Die Nichtnegativität liefert für $y > x$

$$F(y) - F(x) = \mu(]x, y]) \geq 0$$

oder

$$F(y) \geq F(x).$$

F ist also monoton nichtfallend.

Die Sigmaadditivität haben wir schon in Satz 2.15 mit Stetigkeitseigenschaften in Verbindung gebracht. Wir wählen eine monoton nichtwachsende Folge x_n mit Grenzwert x. Dann fällt die Folge der Intervalle $]x, x_n]$ gegen die leere Menge, und daher gilt

$$0 = \mu(\emptyset) = \lim \mu(]x, x_n]) = \lim_{n\to\infty} (F(x_n) - F(x)).$$

Daraus folgt natürlich, dass $F(x_n)$ gegen $F(x)$ konvergiert, und F ist rechtsstetig.

Jede Verteilungsfunktion ist also nichtfallend und rechtsstetig. Dass auch die Umkehrung gilt, formulieren wir in einem Satz, in dem wir auch alles festhalten, was wir bisher festgestellt haben:

Satz 3.1

Eine reelle Funktion ist genau dann eine Verteilungsfunktion, wenn sie monoton nichtfallend und rechtsstetig ist, d.h., zu jeder nichtfallenden und rechtsstetigen Funktion gibt es ein Lebesgue-Stieltjes Maß mit dieser Funktion als Verteilungsfunktion. Überdies ist ein Lebesgue-Stieltjes Maß durch eine Verteilungsfunktion eindeutig bestimmt. Andererseits gibt es zu jedem Lebesgue-Stieltjes Maß mindestens eine Verteilungsfunktion, und zwei verschiedene Verteilungsfunktionen zu demselben Maß unterscheiden sich durch eine additive Konstante.

Der einzige Punkt, der noch zu klären ist, ist die Tatsache, dass eine nichtfallende rechtsstetige Funktion F durch die Formel

$$\mu(]a, b]) = F(b) - F(a)$$

tatsächlich eine Maßfunktion auf $\mathfrak{T}$ festlegt.

Da F monoton ist, erhalten wir unmittelbar, dass μ nichtnegativ ist.

Als nächstes zeigen wir die Additivität von μ. Da $\mathfrak{T}$ ein Semiring ist, brauchen wir uns nur um die Vereinigung von zwei disjunkten Mengen zu kümmern. Es seien also $]a, b]$ und $]c, d]$ zwei nichtleere Intervalle, deren Vereinigung wieder ein Intervall ist (wenn eines der beiden Intervalle leer ist, ist die Frage noch einfacher zu lösen). Dann muss, wie man sich leicht überlegt, entweder $b = c$ oder $a = d$ gelten. Wir nehmen ohne Beschränkung der Allgemeinheit den ersten Fall an und erhalten

$$\mu(]a, d]) = F(d) - F(a) = F(d) - F(b) + F(b) - F(a) = \mu(]a, b]) + \mu(]b, d]).$$

Es gilt für μ also auch die Additivität.

Nun wollen wir die Sigmaadditivität zeigen. Es sei $]a_n, b_n]$ eine Folge von disjunkten Intervallen, deren Vereinigung ein Intervall $]a, b]$ darstellt.

Aus der Additivität und Nichtnegativität von μ erhalten wir

$$\mu(]a, b]) \geq \mu(\bigcup_{n=1}^{N}]a_n, b_n]) = \sum_{n=1}^{N} \mu(]a_n, b_n]),$$

und durch den Grenzübergang $N \to \infty$ erhalten wir

$$\mu(]a, b]) \geq \sum_{n=1}^{\infty} \mu(]a_n, b_n]).$$

Es bleibt uns also noch die umgekehrte Ungleichung. Auch diese wollen wir auf eine endliche Vereinigung zurückführen, denn dann genügt uns die einfache Additivität von μ, die wir schon bewiesen haben. Das Mittel dazu ist der Satz von Heine-Borel, denn schließlich haben wir hier die Situation, dass ein Intervall von anderen überdeckt wird. Allerdings sollte das überdeckte Intervall abgeschlossen sein und die überdeckenden offen, und nicht wie hier alle Intervalle halboffen. Dieses Problem ist aber leicht zu lösen: wir schneiden einfach vom Intervall $]a, b]$ links ein kleines Stück ab (und machen es zu einem abgeschlossenen), und dehnen die Intervalle $]a_n, b_n]$ nach rechts aus. Wir müssen dabei nur aufpassen, dass wir nicht zuviel wegnehmen bzw. dazugeben. Wir setzen

$$\tilde{a} > a$$

und

$$\tilde{b}_n > b_n$$

so, dass

$$F(\tilde{a}) < F(a) + \epsilon$$

und

$$F(\tilde{b}_n) < F(b_n) + \frac{\epsilon}{2^n}.$$

Jetzt ist $]a_n, \tilde{b}_n[$ eine offene Überdeckung des abgeschlossenen Intervalls $[\tilde{a}, b]$. Dafür funktioniert der Satz von Heine-Borel, wir können also eine endliche Teilüberdeckung finden, und weil wir an linkshalboffenen Intervallen interessiert sind, schreiben wir gleich

$$]\tilde{a}, b] \subseteq [\tilde{a}, b] \subseteq \bigcup_{n=1}^{N}]a_n, \tilde{b}_n[\subseteq \bigcup_{n=1}^{N}]a_n, \tilde{b}_n].$$

Jetzt verwenden wir die Additivität und Nichtnegativität von μ und erhalten

$$F(b) - F(\tilde{a}) \leq \sum_{n=1}^{N} (F(\tilde{b}_n) - F(a_n)) \leq \sum_{n=1}^{\infty} (F(\tilde{b}_n) - F(a_n))$$

und schließlich

$$F(b) - F(a) \leq F(b) - F(\tilde{a}) + \epsilon \leq \sum_{n=1}^{\infty} (F(\tilde{b}_n) - F(a_n)) + \epsilon$$

$$\leq \sum_{n=1}^{\infty} (F(b_n) + \frac{\epsilon}{2^n} - F(a_n)) + \epsilon \leq \sum_{n=1}^{\infty} (F(b_n) - F(a_n)) + 2\epsilon.$$

Da ϵ beliebig war, haben wir gezeigt, dass

$$\mu(]a, b]) \leq \sum_{n=1}^{\infty} \mu(]a_n, b_n]),$$

und die Sigmaadditivität von μ ist gezeigt.

Mithilfe der Verteilungsfunktion lassen sich Wahrscheinlichkeiten berechnen: Aus der Definition folgt direkt

$$\mu_F(]a, b]) = F(b) - F(a).$$

Die Stetigkeit der Maßfunktion μ liefert daraus

$$\mu_F(]a, b[) = F(b - 0) - F(a),$$

$$\mu_F([a, b[) = F(b - 0) - F(a - 0),$$

$$\mu_F([a, b]) = F(b) - F(a - 0).$$

In der letzten Gleichung können wir $a = b$ setzen und erhalten

$$\mu_F(\{x\}) = F(x) - F(x - 0).$$

Daraus folgt, $\mu_F(\{x\}) = 0$ für alle $x \in R$ genau dann gilt, wenn F stetig ist, also keine Sprünge hat. Die Antithese dazu ist eine Verteilungsfunktion, die nur Sprünge hat, also

Definition 3.3

Die Verteilungsfunktion F heißt diskret, wenn es eine abzählbare Menge D und für jedes $x \in D$ ein $p(x) > 0$ gibt, sodass

$$F(b) - F(a) = \sum_{x \in D \cap]a,b]} p(x).$$

Jede Verteilungsfunktion lässt sich in eindeutiger Weise (bis auf additive Konstante) in eine stetige und eine diskrete Verteilungsfunktion zerlegen. Das folgt aus dem folgenden allgemeineren Satz:

Definition 3.4

Das Maß μ auf dem Messraum $(\Omega, \mathfrak{S}, \mathbb{P})$ mit $\{\omega\} \in \mathfrak{S}$ für alle $\omega \in \Omega$ heißt

- stetig, wenn für alle $\mu(\{\omega\}) = 0$ für alle $\omega \in \Omega$,
- diskret, wenn es eine höchstens abzählbare Menge D gibt mit $\mu(D^C) = 0$.

Satz 3.2

Jedes sigmaendliche Maß μ auf dem Messraum $(\Omega, \mathfrak{S}, \mathbb{P})$ mit $\{\omega\} \in \mathfrak{S}$ für alle $\omega \in \Omega$ lässt sich in eindeutiger Weise in ein diskretes und ein stetiges Maß zerlegen.

Der Beweis wird in den Übungen geführt.

3.2 Mehrdimensionale Lebesgue-Stieltjes Maße

Auch im mehrdimensionalen Fall $\Omega = \mathbb{R}^n$, $\mathfrak{S} = \mathfrak{B}_n$ definieren wir ein Lebesgue-Stieltjes Maß als ein Maß, das für beschränkte Borelmengen endliche Werte ergibt. Wir wollen auch hier eine Verteilungsfunktion finden, die das Maß festlegt. Dazu gehen wir wie im eindimensionalen Fall von einem endlichen Maß μ aus und setzen

$$F(x) = F(x_1, \ldots, x_n) = \mu(]-\infty, x]) = \mu(]-\infty, x_1] \times \ldots \times]-\infty, x_n]).$$

Wir betrachten zunächst den Fall $n = 2$ und versuchen, das Maß des Rechtecks $]a_1, b_1] \times]a_2, b_2]$ aus den Funktionswerten von F zu bestimmen. Das funktioniert natürlich so, dass man vom Funktionswert der rechten oberen Ecke erst einmal die Werte der beiden angrenzenden Ecken abzieht. Dabei wird allerdings etwas doppelt abgezogen, und deswegen muss man den Funktionswert für die linke untere Ecke wieder dazuzählen.

Also:

$$\mu(]a, b]) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Für $n > 2$ Dimensionen ist es nicht ganz leicht, das Äquivalent dazu lesbar anzuschreiben, etwa:

$$\mu(]a, b]) = \sum_{e \in \{0,1\}^n} (-1)^{\sum e_i} F(ae + b(1 - e)).$$

Dabei sind alle Operationen mit den Vektoren a, b und e komponentenweise zu verstehen.

Oder man definiert Differenzoperatoren

$$\Delta_i^{(a,b)} f(x) = f(z) - f(y),$$

wobei $z_i = b$, $y_i = a$ und $y_j = z_j = x_j$ für $j \neq i$.

Damit ist

$$\mu(]a, b]) = \Delta^{(a,b)} F = \Delta_1^{(a_1,b_1)} \Delta_2^{(a_2,b_2)} \ldots \Delta_n^{(a_n,b_n)} F.$$

Die Rechtsstetigkeit der Verteilungsfunktion bleibt genauso wie im eindimensionalen Fall gültig, die Monotonie (die ja aus der Nichtnegativität des Maßes folgt) müssen wir durch die Forderung ersetzen, dass der obige Ausdruck für $\mu(]a, b])$ nichtnegativ sein soll.

Wenn diese Forderungen erfüllt sind, dann kann man so wie im eindimensionalen Fall zeigen, dass dazu auch ein Lebesgue-Stieltjes Maß existiert:

Satz 3.3

Die Funktion $F : \mathbb{R}^n \to \mathbb{R}$ ist genau dann Verteilungsfunktion eines Lebesgue-Stieltjes Maßes, wenn

1. F rechtsstetig ist und
2. $\Delta^{(a,b)}F \geq 0$ für alle $a, b \in \mathbb{R}^n$ mit $a \leq b$.

Umgekehrt kann man zu jedem Lebesgue-Stieltjes Maß eine Verteilungsfunktion finden, etwa:

$$F(x) = \text{sig}(\prod_{i=1}^{n} x_i)\mu(]\min(x, 0), \max(x, 0)]).$$

Die Eindeutigkeitsfrage für die Verteilungsfunktion ist nicht ganz so einfach zu beantworten. Es gilt: man kann jede weitere Verteilungsfunktion zum selben Maß μ in der Form

$$F(x) + \sum_{i=1}^{n} f_i(x)$$

schreiben, wobei die Funktion f_i von x_i nicht abhängt.

Zum Abschluss sei noch bemerkt, dass eine Verteilungsfunktion nicht monoton sein muss, und dass umgekehrt eine Funktion, die in ihren Argumenten monoton ist, keine Verteilungsfunktion sein muss. Ein Beispiel dafür bildet die Verteilungsfunktion des beliebtesten Maßes, des Lebesgue-Maßes:

$$F(x) = x_1 x_2 \dots x_n.$$

Andere Beispiele dazu werden in den Übungen besprochen. Allerdings werden wir für endliche Maße, und besonders für Wahrscheinlichkeitsmaße als "die Verteilungsfunktion" immer die Funktion

$$F(x) = \mu(]-\infty, x])$$

verstehen. Eine solche Verteilungsfunktion nennen wir Verteilungsfunktion im engeren Sinn. Eine solche hat zusätzlich zu den schon bekannten noch die folgenden Eigenschaften:

1. F ist in jeder Argumentvariable monoton nichtfallend.
2. $\lim_{x_i \to -\infty} F(x_1, \dots, x_d) = 0$.
3. $\lim_{\min(x_1,\dots,x_d)\to\infty} F(x_1, \dots, x_d) = 1$.

3.3 Regularität

Wir haben im Zusammenhang mit dem Fortsetzungssatz festgestellt, dass jede Menge mit endlichem Maß aus dem erzeugten Sigmaring durch Mengen aus dem Ring approximiert werden kann. Auf den reellen Zahlen können wir die Approximation mit topologischen Eigenschaften in Verbindung bringen:

Definition 3.5

μ sei ein Maß auf $(\mathbb{R}^d, \mathfrak{B}^d)$. $A \in \mathfrak{B}^d$ heißt

1. regulär von oben, wenn
$$\mu(A) = \inf\{\mu(U) : A \subseteq U, U \text{ offen}\},$$
2. regulär von unten, wenn
$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ kompakt}\},$$
3. regulär, wenn A sowohl von unten als auch von oben regulär ist.

4. abgeschlossen-regulär, wenn

$$\mu(A) = \sup\{\mu(F) : F \subseteq F, K \text{ abgeschlossen}\},$$

Ist jede Menge $A \in \mathfrak{B}^d$ (von oben/von unten) regulär, dann heißt μ (von oben/unten) regulär.

Satz 3.4

Jedes Lebesgue-Stieltjes Maß ist regulär von oben.

Da jede Menge mit unendlichem Maß in trivialer Weise von oben regulär ist, genügt es, den Fall $\mu(A) = \mu^*(A) < \infty$ zu betrachten (hier ist μ^* das äußere Maß, das von der Einschränkung von μ auf die linkshalboffenen Intervalle erzeugt wird). Für jedes $\epsilon > 0$ gibt es nach der Definition des äußeren Maßes eine Folge von Intervallen $]a_n, b_n]$ mit

$$A \subseteq \bigcup_n]a_n, b_n]$$

und

$$\sum_n \mu(]a_n, b_n]) < \mu(A) + \epsilon/2.$$

Dann können wir $\hat{b}_n > b_n$ finden, sodass

$$\mu(]a_n, \hat{b}_n[) \leq \mu(]a_n, b_n]) + \epsilon/2^{n+1}$$

gilt. Die Menge

$$U = \bigcup_n]a_n, \hat{b}_n]$$

ist dann offen, umfasst A und erfüllt $\mu(U) \leq \mu(A) + \epsilon$.

Satz 3.5

Jedes sigmaendliche Maß auf $\mathfrak{B}^n$ ist regulär von unten.

Es sei zunächst μ ein endliches Maß und A eine beliebige Borelmenge. Nach dem vorigen Satz können wir eine offene Menge U finden mit $A^C \subseteq U$ und $\mu(U) < \mu(A^C) + \epsilon$. Dann ist $F = U^C$ eine abgeschlossenen Menge mit $F \subseteq A$ und $\mu(F) > \mu(A) - \epsilon$. Die Mengen

$$F_n = F \cap [-n, n]^d$$

sind abgeschlossen und beschränkt, also kompakt, und wegen $F_n \uparrow F$ gilt

$$\lim_{n\to\infty} \mu(F_n) = \mu(F) > \mu(A) - \epsilon.$$

Es gibt daher ein n_0 mit

$$\mu(F_{n_0}) > \mu(A) - \epsilon.$$

Das zeigt die Regularität von unten für endliche Maße. Im sigmaendlichen Fall können wir eine Folge von Borelmengen $B_n \uparrow \Omega$ mit $\mu(B_n) < \infty$ finden. Wir wählen $A \in \mathfrak{B}$ und $M < \mu(A)$. Wegen $\mu(B_n \cap A) \uparrow \mu(A)$ können wir ein n finden, sodass $\mu(B_n \cap A) > M$ gilt. Jetzt nützen wir die Regularität von unten des endlichen Maßes μ_n aus, das durch

$$\mu_n(A) = \mu(A \cap B_n)$$

definiert ist. Es gibt also eine kompakte Menge $K \subseteq A$ mit

$$\mu_n(K) > M.$$

Dann gilt auch

$$\mu(K) \geq \mu(K \cap B_n) = \mu_n(K) > M.$$

Zu jeder Zahl $M < \mu(A)$ gibt es also eine kompakte Menge mit größerem Maß, und die Regularität von unten ist bewiesen.

3.4 Das Lebesguemaß

Jetzt sind wir endlich in der Lage, unsere klassischen Längenmaße und Volumina unterzubringen. Dazu setzen wir für die Verteilungsfunktion F einfach $F(x) = x$ ein bzw. in d Dimensionen $F(x) = x_1 \dots x_d$. Das zugehörige Lebesgue-Stieltjes Maß (das Intervallen tatsächlich ihre Länge, ihre Fläche,... zuordnet: es ist nicht schwer nachzurechnen, dass

$$\Delta^{(a,b)} x_1 \dots x_d = \prod_{i=1}^{d} (b_i - a_i)).$$

bezeichnen wir mit λ (bzw. mit λ_d, wenn die Dimension nicht aus dem Kontext klar ist) und nennen es das Lebesguemaß.

Die Vervollständigung der Borelmengen bezüglich des Lebesguemaßes (oder mit anderen Worten, das System der λ^*-messbaren Mengen) nennen wir das System der Lebesguemengen (oder der Lebesgue-messbaren Mengen).

Das Lebesguemaß ist translationsinvariant:

Satz 3.6

$A \subseteq \mathbb{R}^d$ sei Borel-messbar (bzw. Lebesgue-messbar), $y \in \mathbb{R}^d$. Dann ist auch $A \oplus y$ Borel- bzw. Lebesgue-messbar und

$$\lambda_d(A \oplus y) = \lambda_d(A).$$

Wir bezeichnen die Translation um $-y$ mit T und den Semiring der linkshalboffenen Intervalle mit $\mathfrak{T}$. Dann ist offensichtlich $T^{-1}(\mathfrak{T}) = \mathfrak{T}$, und Satz 2.5 liefert

$$\mathfrak{B} = \mathfrak{A}_\sigma(\mathfrak{T}) = \mathfrak{A}_\sigma(T^{-1}(\mathfrak{T})) = T^{-1}(A_\sigma(\mathfrak{T})) = T^{-1}(\mathfrak{B}).$$

Borelmengen werden also durch Verschiebungen auf Borelmengen abgebildet. Außerdem ist λ auf $\mathfrak{T}$ translationsinvariant, und diese Invarianz überträgt sich auf das erzeugte äußere Maß und damit auf die messbaren Mengen.

Bis auf einen konstanten Faktor ist das Lebesguemaß das einzige translationsinvariante Lebesgue-Stieltjes Maß:

Satz 3.7

μ sei ein translationsinvariantes Lebesgue-Stieltjes Maß auf $(\mathbb{R}^d, \mathfrak{B})$. Dann gibt es eine Konstante $c \in [0, \infty[$, sodass

$$\mu = c\lambda_d.$$

Wir führen den Beweis hier nur für $d = 1$. Wir setzen

$$c = \mu(]0, 1]).$$

Dann ist für $n \in \mathbb{N}$

$$c = \mu(\bigcup_{i=1}^{n}]\frac{i-1}{n}, \frac{i}{n}]) = \sum_{i=1}^{n} \mu(]\frac{i-1}{n}, \frac{i}{n}]) = n\mu(]0, \frac{1}{n}]),$$

also

$$\mu(]0, \frac{1}{n}]) = \frac{c}{n}$$

und für $m \in \mathbb{N}$

$$\mu(]0, \frac{m}{n}]) = c\frac{m}{n}.$$

Das gibt die Beziehung

$$\mu(]0, x]) = cx$$

für alle rationalen $x \geq 0$. Für beliebiges x ergibt sich dieselbe Gleichung durch Approximation von oben mit rationalen Zahlen; schließlich ist

$$\mu(]a, b]) = \mu(]0, b - a]) = c(b - a).$$

μ stimmt also auf $\mathfrak{T}$ mit $c\lambda$ überein, und der Fortsetzungssatz ergibt, dass die beiden Maße auch auf $\mathfrak{B}$ und sogar auf $\bar{\mathfrak{B}}$, den Lebesguemengen, übereinstimmen.

Es gilt sogar etwas allgemeiner, dass ein sigmaendliches translationsinvariantes Maß auf $(\mathbb{R}^d, \mathfrak{B})$ proportional zum Lebesguemaß ist. Das wird im Kapitel über Produktmaße gezeigt werden. Wenn die Sigmaendlichkeit nicht gefordert wird, dann lassen sich viele translationsinvariante Maße finden. Es sind etwa die Hausdorffmaße η_α translationsinvariant; sigmaendlich sind sie nur für $\alpha = d$, und dann natürlich proportional zum Lebesguemaß.

Die Translationsinvarianz gibt uns die Möglichkeit, eine Menge zu finden, die nicht Lebesgue-messbar (und also auch nicht Borel) ist:

Definition 3.6

Eine Vitalimenge ist eine Teilmenge, die aus jeder Äquivalenzklasse $(x)_\sim$ zu der Äquivalenzrelation

$$a \sim b \Leftrightarrow b - a \in \mathbb{Q}$$

genau einen Vertreter enthält.

Anmerkung: die Existenz solcher Mengen ergibt sich aus dem Auswahlaxiom. Dass wir hier auf das Auswahlaxiom zurückgreifen kommt nicht von ungefähr: es gibt Alternativen zum Auswahlaxiom, unter denen alle Teilmengen der reellen Zahlen Lebesgue-messbar sind.

Satz 3.8

Vitalimengen sind nicht Lebesgue-messbar.

Wir können o.B.d.A. annehmen, dass $V \subseteq]0, 1]$ gilt, weil

$$V' = \bigcup_{n \in \mathbb{N}} (V \cap [n, n + 1]) \ominus n$$

eine messbare Menge ist, wenn V messbar ist.

Gemäß der Definition von V sind für $q, r \in \mathbb{Q}, q \neq r$ die Mengen $V \oplus q$ uns $V \oplus r$ disjunkt. Andererseits gibt es zu jedem $x \in]0, 1]$ ein $y \in V$ mit $x - y \in \mathbb{Q}$, und es gilt natürlich $|x - y| \leq 1$. Daraus folgt, dass

$$]0, 1] \subseteq \bigcup_{q \in \mathbb{Q} \cap [-1,1]} (V \oplus q) \subseteq [-1, 2].$$

Wäre V messbar, dann wäre

$$\lambda(\bigcup_{q \in \mathbb{Q} \cap [-1,1]} (V \oplus q)) = \infty \cdot \lambda(V).$$

Ist $\lambda(V) = 0$, dann ist das 0, ist $\lambda(V) > 0$, dann ergibt sich ∞. In beiden Fällen ist das ein Widerspruch dazu, dass das Lebesguemaß dieser Menge zwischen 1 und 3 liegen sollte.

3.5 Spezielle Verteilungen

Wahrscheinlichkeitsmaße (die als endliche Maße zu den Lebesgue-Stieltjes-Maßen zählen) werden gerne als *Verteilungen* bezeichnet. Einige häufig auftretenden Verteilungen haben es zu eigenen Namen gebracht. Dabei werden diskrete (auf einer abzählbaren Menge, in diesem Zusammenhang durchwegs die ganzen Zahlen, konzentrierte) und stetige (solche, bei denen die Verteilungsfunktion stetig ist, sogar stückweise stetig differenzierbar ist) unterschieden. Wir beginnen mit den diskreten Exemplaren

3.5.1 Diskrete Verteilungen

Diese Verteilungen werden durch ihre *Wahrscheinlichkeitsfunktion*

$$p(x) = \mathbb{P}(\{x\})$$

festgelegt. Die Verteilungsfunktion ergibt sich daraus natürlich als $F(x) = \sum_{y \leq x} p(y)$.

Die deterministische (degenerierte, Dirac-) Verteilung $D(a)$

Diese Verteilung lässt dem Zufall nicht viel Raum: der Wert a wird mit Wahrscheinlichkeit 1 angenommen.

Die Alternativverteilung $A(p)$

Diese einfachste nichtdegenerierte Verteilung hat nur zwei Werte mit positiver Wahrscheinlichkeit, 0 und 1:

$$\mathbb{P}(\{1\}) = p, \mathbb{P}(\{0\}) = 1 - p.$$

Die diskrete Gleichverteilung $D(a, b)$

Diese entspricht einem Laplace-Raum mit $\Omega = \{a, a+1, \dots b\}$ mit Parametern $a \leq b \in mathbbZ$, also

$$\mathbb{P}(\{x\}) = \frac{1}{b-a+1}, a \leq x \leq b, x \in \mathbb{Z}.$$

Die Binomialverteilung $B(n, p)$

Wir betrachten n unabhängige Ereignisse (die oft als "Experimente" bezeichnet werden, das Eintreten eines Ereignisses wird dann auch als "Erfolg" bezeichnet) $A_1, \dots, A_n$ mit $\mathbb{P}(A_i) = p, i = 1, \dots, n$. Wir zählen, wie viele dieser Ereignisse eintreten (die Anzahl der "Erfolge"). $p(k)$ ist die Wahrscheinlichkeit, dass diese Anzahl $= k$ ist. Diese ergibt sich folgendermaßen: zuerst wählen wir die k Ereignisse aus, die eintreten sollen, die anderen $n-k$ dürfen nicht eintreten. Das geht auf $\binom{n}{k}$ Arten. Jedes Ereignis, das eintreten soll, tut das mit Wahrscheinlichkeit p, eines, das nicht eintreten soll, hat dafür Wahrschinlichkeit $1-p$. Weil die Ereignisse unabhängig sind, ist die Wahrscheinlichkeit, dass all das gleichzeitig passiert, das Produkt dieser Wahrscheinlichkeiten, also $p^k(1-p)^{n-k}$. Insgesamt haben wir

$$\mathbb{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \dots n.$$

Die hypergeometrische Verteilung $H(N, A, n)$

Ähnlich wie bei der Binomialverteilung wird gezählt, wie viele von n Ereignissen eintreten, nur sind jetzt die einzelnen Ereignisse nicht mehr unabhängig sind. Dahinter steht ein Urnenmodell: in einer Urne befindet sich eine gewisse Anzahl von Kugeln, daraus wird eine Kugel gezogen, und jede Kugel hat die gleiche Wahrscheinlichkeit, gezogen zu werden. Konkret sollen $N \in \mathbb{N}$ Kugeln in der Urne liegen, davon sind $A \leq N$ weiß und $N - A$ schwarz. Es wird $n \leq N$ mal ohne Zurücklegen gezogen (bei jeder Ziehung ist eine Kugel weniger in der Urne), und A_i ist das Ereignis, dass bei der i-ten Ziehung eine weiße Kugel gezogen wird. Wir zählen wieder die Ereignisse, die eintreten, also die Anzahl der weißen Kugeln unter den gezogenen. Die entsprechende Verteilung lässt sich nach dem Vorbild der Binomialverteilung mit dem Multiplikationssatz berechnen, was

$$\mathbb{P}(\{k\}) = \binom{n}{k} \frac{A(A-1)\dots(A-k+1)(N-A)(N-A-1)\dots(N-A-n+k+1)}{N(N-1)\dots(N-n+1)}$$

ergibt. Alternativ kann man aus Symmetrieüberlegungen ableiten, dass alle $\binom{N}{n}$ möglichen Auswahlen von n Kugeln aus den N vorhandenen die gleiche Wahrscheinlichkeit haben, als Ergebnis

der n Ziehungen zu erscheinen. Davon sind die $\binom{A}{k}\binom{N-A}{n-k}$ solche, bei denen k weiße Kugel gezogen wurden. Das gibt die alternative Form

$$\mathbb{P}(\{k\}) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}.$$

Hier wäre noch nötig, die Werte von k einzuschränken; das lässt sich vermeiden, wenn wir die Konvention $\binom{n}{k} = 0$ für $k < 0$ und $k > n$ verwendet.

Man kann auch *mit* Zurücklegen ziehen, dann sind die Ereignisse A_i unabhängig, und es ergibt sich eine Binomialverteilung $B(n, A/N)$.

Die Poissonverteilung $P(\lambda)$

Diese Verteilung ergibt sich aus der Binomialverteilung durch den Grenzübergang $n \to \infty$, $np \to \lambda \geq 0$. Die Wahrscheinlichkeitsfunktion ist

$$\mathbb{P}(k) = \frac{\lambda^k}{k!} e^{-\lambda} (k = 0, 1, \ldots).$$

Die geometrische Verteilung $G(p)/G^*(p)$

Hier sind wir wieder bei einer Folge (A_n) von unabhängigen Ereignissen mit $P(A_n) = p \in]0, 1]$. Diesmal zählen wir die Anzahl der Versuche bis zum ersten Erfolg. Das gibt die erste Version $G^*(p)$ der geometrischen Verteilung:

$$\mathbb{P}(\{k\}) = p(1-p)^{k-1}, k = 1, 2, \ldots$$

(es müssen ja $k-1$ Misserfolge und danach ein Erfolg auftreten).

Weniger intuitiv, aber von größerer Bedeutung ist die Version $G(p)$, bei der nicht die Versuche, sondern die Misserfolge gezählt werden. Das gibt

$$\mathbb{P}(\{k\}) = p(1-p)^{k}, k = 0, 1, \ldots$$

Die negative Binomialverteilung $NB(\alpha, p)/NB^*(n, p)$

Ähnlich wie bei der geometrischen Verteilung werden von der negativen Binomialverteilung NB^* unabhängige Versuche gezählt, allerdings nicht bis zum ersten Erfolg, sondern bis zum n-ten. Das Ereignis $\{k\}$ erfordert, dass unter den ersten $k-1$ Versuchen $n-1$ Erfolge sind, und dass der k-te Versuch ebenfalls ein Erfolg ist. Das gibt

$$\mathbb{P}(\{k\}) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, k = n, n+1, \ldots$$

NB zählt wieder die Anzahl der Misserfolge, die in Kauf zu nehmen sind, bis man n Erfolge beisammen hat. Das ergibt

$$\mathbb{P}(\{k\}) = \binom{n+k-1}{n-1} p^n (1-p)^k = \binom{n+k-1}{k} p^n (1-p)^k, k = 0, 1, \ldots$$

Darin können wir für n eine beliebige reelle Zahl einsetzen, die wir zur Unterscheidung vom ganzzahligen n α nennen wollen. Dabei geht natürlich die anschauliche Interpretation verloren, und der Binomialkoeffizient wird als

$$\binom{x}{k} = \frac{x(x-1)\ldots(x-k+1)}{k!}$$

interpretiert.

3.5.2 Stetige Verteilungen

Es wurde bereits erwähnt, dass diese Verteilungsfunktionen nicht nur stetig, sondern sogar stkkweise stetig differenzierbar sind. Diese Ableitung $f = F'$ nennen wir die *Dichte* der Verteilung. Mit ihr kann die Verteilungsfunktion als

$$F(x) = \int_{-\infty}^{x} f(u)\,du$$

dargestellt werden (an den endlich vielen stellen, an denen F' nicht definiert ist, können wir f beliebige Werte geben; das wird meist so getan, dass f dort zumindest einseitig stetig ist. Oft ist es einfacher, die Dichte anzugeben, weil das Integral nicht geschlossen dargestellt werden kann.

Die stetige Gleichverteilung $U(a, b)$

Die Dichte ist im Intervall $[a, b]$ konstant, also

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \le x \le b, \\ 0 & \text{sonst.} \end{cases}$$

Für diese Verteilung können wir auch die Verteilungsfunktion schön angeben:

$$F(x) = \begin{cases} 0 & \text{für } x < a, \\ \frac{x-a}{b-a} & \text{für } a \le x \le b, \\ 1 & \text{für } x \ge b. \end{cases}$$

Die Exponentialverteilung $E(\lambda)$

Diese hat die Dichte ($\lambda > 0$)

$$f(x) = \lambda e^{-\lambda x}[x \ge 0]$$

und die Verteilungsfunktion

$$F(x) = (1 - e^{-\lambda x})[x \ge 0].$$

Die Gammaverteilung $\Gamma(\alpha, \lambda)$

Wir definieren

Definition 3.7

Die Gammafunktion ist für $\alpha > 0$ gegeben durch

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx,$$

die Betafunktion durch ($\alpha > 0, \beta > 0$)

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

Es bestehen die Beziehungen

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$
$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$$

und speziell für $n \in \mathbb{N}$

$$\Gamma(n) = (n-1)!$$

Damit können wir die Dichte der Gammaverteilung angeben ($\alpha, \lambda > 0$):

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}[x > 0].$$

Daraus ergibt sich die Exponentialverteilung als Spezialfall $\alpha = 1$.

Die Betaverteilung erster Art $B_1(\alpha, \beta)$

Diese hat die Dichte $(\alpha, \beta > 0)$

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}[0 < x < 1].$$

Die Betaverteilung zweiter Art $B_2(\alpha, \beta)$

Hat Dichte $(\alpha, \beta > 0)$

$$f(x) = \frac{1}{B(\alpha, \beta)} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}}[x > 0].$$

Die Normalverteilung $N(\mu, \sigma^2)$

Diese Verteilung ist eine der wichtigsten, wenn nicht *die* wichtigste. Sie hat die Dichte $(\mu \in \mathbb{R}, \sigma^2 > 0)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Die Verteilung $N(0,1)$ heißt *Standard-Normalverteilung*, und wir haben eine eigene Bezeichnung für ihre Dichte

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Ihre Verteilungsfunktion

$$\Phi(x) = i = \int_{-\infty}^{x} \phi(u)du$$

kann nicht in geschlossener Form dargestellt werden; mit Ihrer Hilfe ergibt sich die Verteilungsfunktion der allgemeinen Normalverteilung $N(\mu, \sigma^2)$ als

$$F(x) = \Phi(\frac{x-\mu}{\sigma}).$$

Für ihre Berechnung gibt es in allen Statistikpaketen entsprechende Routinen (in R macht das die Funktion `pnorm()`), oder man verwendet eine Tabelle wie die in Anhang A.

Beispiel 3.1

Um etwa $\Phi(0.57)$ zu bestimmen, wird in der Zeile "0.5" die Spalte "7" aufgesucht, dort findet sich der Funktionswert 0.716.

Etwas umfangreicher ist die Berechnung der Wahrscheinlichkeit des Intervalls $[1, 4]$ für die Verteilung $N(2, 49)$. Diese ergibt sich zu

$$\Phi(\frac{4-2}{7}) - \Phi(\frac{1-2}{7}) = \Phi(0.2857\ldots) - \Phi(-0.1428\ldots) \approx 0.614 - (1 - 0.556) = 0.170.$$

Wir haben die Argumente einfach auf zwei Stellen gerundet (man könnte auch zur Verbesserung der Genauigkeit interpolieren) und für das negative Argument die Identität

$$\Phi(-x) = 1 - \Phi(x)$$

verwendet.

Die Cauchyverteilung

Diese Verteilung fungiert oft als Gegenbeispiel, weil sie sehr viel Masse in den Enden hat. Ihre Dichte ist

$$\frac{1}{\pi(1+x^2)}.$$

3.6 Ergänzungen

3.6.1 Borelmaße

So, wie man die Borelmengen in allgemeinen topologischen Räumen definieren kann, kann man auch das Konzept der Lebesgue-Stieltjes Maße verallgemeinern. Der entsprechende Begriff ist die Lokalendlichkeit:

Definition 3.8

(Ω, τ) sei ein topologischer Raum, $\mathfrak{S}$ eine Sigmaalgebra über Ω, die die offenen Mengen enthält (und damit natürlich die Borelmengen). Dann heißt das Maß μ auf $\mathfrak{S}$ lokalendlich, wenn jeder Punkt eine Umgebung mit endlichem Maß besitzt, also wenn es zu jedem $\omega \in \Omega$ eine offene Menge $U \in \tau$ gibt mit $x \in U$ und $\mu(U) < \infty$.

Ein Borelmaß ist ein lokalendliches Maß auf der Sigmaalgebra der Borelmengen.

Mit dieser Definition sind die Lebesgue-Stieltjes Maße genau die Borelmaße auf $\mathbb{R}^d$.

Die Definitionen der Regularität von unten und von oben übertragen sich direkt auf diesen allgemeinen Fall. Ein von unten reguläres Borelmaß heißt "Radonmaß", und ein reguläres Borelmaß heißt, wie man auf Wikipedia nachlesen kann, "reguläres Borelmaß".

Wir haben oben bewiesen, dass in den reellen Zahlen alle Borelmaße regulär sind. Dieser Satz gilt im allgemeinen nicht mehr. Ein einfacher Fall ist der folgende:

Satz 3.9

Wenn jede offene Menge als abzählbare Vereinigung von kompakten Mengen darstellbar ist, dann ist jedes endliche Borelmaß regulär.

3.7 Wiederholungsfragen

1. Wie ist ein Lebesgue-Stieltjes Maß definiert?
2. Wann nennen wir F Verteilungsfunktion des Lebesgue-Stietjes Maßes μ?
3. Welche Eigenschaften hat eine Verteilungsfunktion?
4. Wann ist F Verteilungsfunktion des mehrdimensionalen Lebesgue-Stieltjes Maßes μ?
5. Welche Eigenschaften hat eine mehrdimensionale Verteilungsfunktion?
6. Wann sprechen wir von einer Verteilungsfunktion im engeren Sinn? Welche zusätzlichen Eigenschaften gelten für diese Funktionen?
7. Welche speziellen diskreten Wahrscheinlichkeitsverteilungen haben wir definiert?
8. Welche speziellen stetigen Wahrscheinlichkeitsverteilungen haben wir definiert?
9. Wann heißt ein Maß regulär von oben?
10. Wann heißt ein Maß regulär von unten?
11. Wann heißt ein Maß regulär?
12. Was kann man über die Regularitätseigenschaften von Maßen auf $\mathbb{R}^d$ sagen?

Kapitel 4

Messbare Funktionen

Wir haben uns bisher mit Maßen auseinandergesetzt, und dabei schon einiges erreicht (und dürfen uns dafür auf die Schulter klopfen), jetzt steuern wir auf unseren neuen Integralbegriff zu. Da wir Integrale nicht definieren können, ohne dafür Funktionen zur Verfügung zu haben, wollen wir uns nun mit den Funktionen befassen, die wir später in unser Integral stecken werden.

Wir wollen das neue Integral ja durch das Riemann-Integral

$$\int_0^\infty \mu([f > y])dy$$

definieren (wir erlauben uns, als x-"Achse" allgemeinere Mengen als die reellen Zahlen zuzulassen, damit ist etwa auch die Volumenmessung á la Archimedes/Galileo/Cavalieri inkludiert), und damit das Sinn macht, muss die Menge $[f > y]$ ein Maß haben, also messbar sein. Satz 2.5 (in der Version für Sigmaalgebren) impliziert, dass auch alle Urbilder von Mengen aus der Sigmaalgebra, die von dem System der Intervalle $]y, \infty[, y \in \mathbb{R}$ erzeugt wird, also für alle Borelmengen, messbar sein müssen.

Wenn wir auch als Bildraum nicht nur die reellen Zahlen zulassen, kommen wir so zu der Definition

Definition 4.1

Es seien $(\Omega_1, \mathfrak{S}_1)$ und $(\Omega_2, \mathfrak{S}_2)$ zwei Messräume. Die Funktion $f : \Omega_1 \to \Omega_2$ heißt messbar (genauer $\mathfrak{S}_1$-$\mathfrak{S}_2$-messbar, wenn keine Gefahr von Verwechslungen besteht, können wir die genaue Spezifikation weglassen), wenn für jede Menge $A \in \mathfrak{S}_2$ das Urbild $f^{-1}(A)$ in $\mathfrak{S}_1$ liegt.

Wir schreiben dafür

$$f : (\Omega_1, \mathfrak{S}_1) \to (\Omega_2, \mathfrak{S}_2).$$

Ist insbesondere $\Omega_2 = \mathbb{R}^n$ und $\mathfrak{S}_2 = \mathfrak{B}_n$, dann nennen wir f kurz $\mathfrak{S}_1$-messbar. Ist zusätzlich $\Omega_1 = \mathbb{R}^m$ und $\mathfrak{S}_1 = \mathfrak{B}_m$, dann heißt f Borel-messbar, und falls $\mathfrak{S}_1$ das System der Lebesguemengen ist, dann heißt f Lebesgue-messbar.

Anmerkung: wie so oft können wir $\mathbb{C}$ mit $\mathbb{R}^2$ identifizieren und so auch komplexwertige messbare Funktionen definieren.

Diese Definition erinnert stark an die stetigen Funktionen: eine Funktion ist bekanntlich genau dann stetig, wenn die Urbilder von offenen Mengen offen sind.

Um die Messbarkeit einer Funktion festzustellen, müssen wir theoretisch alle möglichen Mengen $A \in \mathfrak{S}_2$ daraufhin überprüfen, ob ihr Urbild in $\mathfrak{S}_1$ liegt. Wir würden lieber mit weniger auskommen, und dazu verhilft uns der

Satz 4.1: Messbarkeitskriterium

$\mathfrak{C}$ sei ein beliebiges Erzeugendensystem von $\mathfrak{S}_2$. Dann ist $f : \Omega_1 \to \Omega_2$ genau dann $\mathfrak{S}_1$-$\mathfrak{S}_2$-messbar, wenn für jedes $A \in \mathfrak{C}$ das Urbild $f^{-1}(A)$ in $\mathfrak{S}_1$ liegt.

Das ist eine direkte Folgerung aus Satz 2.5, der besagt, dass die vom Urbild eines Mengensystems erzeugte Sigmaalgebra mit dem Urbild der erzeugten Sigmaalgebra übereinstimmt.

Wenn wir für $\mathfrak{S}_2$ die Borelmengen einsetzen, dann können wir in diesem Satz als Erzeugendensystem etwa die offenen Intervalle oder die abgeschlossenen Intervalle oder die einseitig unendlichen Intervalle verwenden, und können so feststellen (wenn wir auch für $\mathfrak{S}_1$ die Borelmengen setzen), dass alle stetigen und auch alle monotonen Funktionen Borel-messbar sind.

Noch schnell ein allgemeiner Satz, bevor wir uns den reellwertigen Funktionen zuwenden:

Satz 4.2

Es sei $f_1 : (\Omega_1, \mathfrak{S}_1) \to (\Omega_2, \mathfrak{S}_2)$ und $f_2 : (\Omega_2, \mathfrak{S}_2) \to (\Omega_3, \mathfrak{S}_3)$. Dann ist $f_2 \circ f_1 : (\Omega_1, \mathfrak{S}_1) \to (\Omega_3, \mathfrak{S}_3)$.

Das folgt direkt aus der Definition der Messbarkeit.

4.1 (Erweitert) reellwertige Funktionen

Wir haben vorhin schon gesehen, dass die messbaren Funktionen im engeren Sinn (d.h., die, bei denen der Bildraum aus den reellen Zahlen mit ihren Borelmengen besteht) die stetigen und auch die monotonen Funktionen einschließen. Wir wollen uns nun etwas näher mit den Eigenschaften dieser Klasse von Funktionen beschäftigen.

Wir wollen auch Funktionen zulassen, die unendliche Werte annehmen. Dazu ergänzen wir $\mathbb{R}$ zu $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ und $\mathfrak{B}$ zu $\mathfrak{B}(\bar{\mathbb{R}}) = \mathfrak{A}_\sigma(\mathfrak{B} \cup \{\{-\infty\}, \{\infty\}\})$. In mehreren Dimensionen setzen wir $\mathfrak{B}(\bar{\mathbb{R}}^d) = \mathfrak{B}(\bar{\mathbb{R}})^d$ (wir verwenden nicht die natürlicher erscheinende Notation $\bar{\mathfrak{B}}$, um Verwechslungen mit der Vervollständigung zu vermeiden).

Als erstes wollen wir die Messbarkeit n-dimensionaler Funktionen auf den eindimensionalen Fall zurückführen:

Satz 4.3

Es sei $f = (f_1, \dots, f_n) : \Omega \to \bar{\mathbb{R}}^n$. Dann ist f genau dann messbar, wenn f_i für alle $i = 1, \dots, n$ messbar ist.

Einerseits folgt aus der Messbarkeit von f für jede eindimensionale Borelmenge A:

$$f_i^{-1}(A) = f^{-1}(\bar{\mathbb{R}}^{i-1} \times A \times \bar{\mathbb{R}}^{n-i}) \in \mathfrak{S},$$

also ist auch f_i messbar. Umgekehrt folgt aus der Messbarkeit aller f_i die Messbarkeit von f. Dazu zeigen wir, dass das Urbild jedes (n-dimensionalen) Intervalls messbar ist:

$$f^{-1}(]a_1, b_1] \times \ldots \times]a_n, b_n]) = \bigcap_{i=1}^{n} f_i^{-1}(]a_i, b_i]).$$

Dieser Durchschnitt ist messbar, weil alle f_i messbare Funktionen sind. Weil die Intervalle ein Erzeugendensystem für die Borelmengen sind, ist der Satz bewiesen.

Generell (und mit praktisch demselben Beweis) ist

$$f = (f_1, f_2) : \Omega \to \Omega_1 \times \Omega_2$$

genau dann $\mathfrak{S} - \mathfrak{S}_1 \times \mathfrak{S}_2$-messbar ist, wenn $f_i : (\Omega, \mathfrak{S}) \to (\Omega_i, \mathfrak{S}_i)$, $i = 1, 2$.

Als nächstes stellen wir fest, dass wir auf die messbaren Funktionen die Grundrechenarten anwenden können:

Satz 4.4

Es seien $f, g : (\Omega, \mathfrak{S}) \to (\bar{\mathbb{R}}, \mathfrak{B}(\bar{\mathbb{R}}))$. Dann sind auch $f+g, f-g, fg, f/g$ alle messbar, vorausgesetzt, dass diese Operationen überall definiert sind (etwas allgemeiner ist die Einschränkungen einer solchen Verknüpfung auf die Menge aller $\omega \in \Omega$, für die sie definiert ist, dort messbar; etwa ist f/g auf der Menge $[g \neq 0]$ messbar, oder $f + g$ auf dem Komplement von

$[f = \infty, g = -\infty] \cup [f = -\infty, g = \infty]).$

Zum Beweis setzen wir $f_1 = (f, g)$ und $f_2(x, y) = x + y$. Die erste Funktion ist nach dem eben bewiesenen Satz messbar, die zweite, weil sie stetig ist, und daher ist nach Satz 4.2 auch $f_2 \circ f_1 = f + g$ messbar. Dieselbe Idee funktioniert natürlich für alle anderen Fälle, wobei wir natürlich dafür sorgen müssen, dass die Ergebnisse für alle Argumente definiert sind (d.h., dass keine unbestimmte Form $\infty - \infty$ oder $0 * \infty$ Auftritt); wenn man sich auf reelle Funktionen beschränkt (also die Unendlichkeiten ausschließt), reduziert sich diese Forderung darauf, dass bei der Division der Divisor nicht null sein darf.

Mit derselben Argumentation kann man erkennen, dass auch Minimum und Maximum von zwei messbaren Funktionen wieder messbar sind. Es gilt aber etwas mehr:

Satz 4.5

$f_n, n \in \mathbb{N}$ seien messbare Funktionen. Dann sind auch die folgenden Funktionen messbar:

$$\sup_n f_n, \inf_n f_n, \liminf_n f_n, \limsup_n f_n,$$

und auch

$$\lim_n f_n,$$

wenn dieser Grenzwert existiert (d.h., wenn die Folge f_n punktweise konvergiert).

Für $f = \sup_n f_n$ gilt

$$f^{-1}(]-\infty, x]) = \{\omega \in \Omega : \sup f_n \leq x\} = \{\omega \in \Omega : f_n \leq x, n \in \mathbb{N}\} = \bigcap_n f_n^{-1}(]-\infty, x]).$$

Also ist das Supremum messbar, und analog beweist man die Messbarkeit des Infimums. Der Rest folgt aus den Darstellungen

$$\limsup_n f_n = \inf_n \sup_{k \geq n} f_k$$

und

$$\liminf_n f_n = \sup_n \inf_{k \geq n} f_k.$$

Jetzt wollen wir uns einer Klasse von besonders einfachen messbaren Funktionen zuwenden. Auf englisch heißen sie auch so ("simple functions"), aber auf deutsch haben wir einen eigenen Namen dafür:

Definition 4.2

Eine messbare Funktion f heißt Treppenfunktion, wenn sie nur endlich viele verschiedene Werte $a_1, \ldots, a_n$ annimmt. Eine solche Funktion kann man natürlich in der Form

$$f = \sum_{i=1}^{n} a_i A_i$$

darstellen, wobei $A_1, \ldots, A_n$ eine messbare Zerlegung von Ω ist, also $A_i \in \mathfrak{S}$ disjunkt mit $\bigcup A_i = \Omega$.

Auf die Forderung, dass alle a_i verschieden sind, kann verzichtet werden und auch darauf, dass die Vereinigung aller A_i gleich Ω ist. Die (eindeutig bestimmte) Darstellung mit verschiedenen a_i und $\sum A_i = \Omega$ nennen wir die kanonische Darstellung von f.

Die kanonische Darstellung erhält man offensichtlich, wenn man einfach $A_i = f^{-1}(\{a_i\})$ setzt und a_i durch alle möglichen Werte von f (inklusive der 0) laufen lässt. Umgekehrt ist jede Funktion, die man in dieser Form darstellen kann, messbar (als Urbilder können ja nur irgendwelche Vereinigungen der Mengen A_i auftreten, und die sind alle messbar).

Was die Treppenfunktionen für uns interessant macht, ist die Tatsache, dass sich durch sie jede messbare Funktion approximieren lässt:

Satz 4.6: Approximationssatz für messbare Funktionen

Jede messbare Funktion f kann als punktweiser Grenzwert einer Folge von Treppenfunktionen f_n dargestellt werden. Ist zusätzlich f nichtnegativ, dann kann die Folge f_n nichtfallend gewählt werden.

Zum Beweis setzen wir einfach

$$f_n = \begin{cases} 2^n & \text{wenn } f(x) \geq 2^n, \\ k/2^n & \text{wenn } k/2^n \leq f(x) < (k+1)/2^n,\ k = -4^n, \ldots, 4^n - 1, \\ -2^n & \text{wenn } f(x) < -2^n. \end{cases}$$

An diesem Punkt können wir einen Satz beweisen, der eine Umkehrung von Satz 4.2 ist:

Satz 4.7: Faktorisierungssatz für messbare Funktionen

$(\Omega_1, \mathfrak{S}_1)$ und $(\Omega_2, \mathfrak{S}_2)$ seien zwei Messräume und $h : (\Omega_1, \mathfrak{S}_1) \to (\Omega_2, \mathfrak{S}_2)$ eine messbare Funktion. Die Funktion $f : (\Omega_1, \mathfrak{S}_1) \to (\mathbb{R}, \mathfrak{B})$ lässt sich genau dann in der Form $g \circ h$ darstellen, wenn sie bezüglich $h^{-1}(\mathfrak{S}_2)$ messbar ist.

Der Beweis dieses Satzes verläuft nach einem Schema, das immer wieder beim Beweis von Sätzen über messbare Funktionen auftauchen wird: zuerst wird die Behauptung für Indikatorfunktionen bewiesen, dann für Treppenfunktionen, und schließlich der allgemeine Fall (oft kommt noch ein zusätzlicher Schritt für nichtnegative Funktionen dazu, aber das brauchen wir hier nicht).

Die Notwendigkeit der Bedingung ist eine unmittelbare Folge von Satz 4.2, und für die Suffizienz verwenden wir die Vorgangsweise, die wir gerade beschrieben haben:

1. $f = A()$ mit $A \in h^{-1}(\mathfrak{S}_2)$. Dann gibt es ein $B \in \mathfrak{S}_2$ mit $A = h^{-1}(B)$, und wir können $g = B()$ setzen.
2. f eine Treppenfunktion: $f = \sum_{i=1}^n a_i A_i()$. Wir setzen natürlich $g = \sum_{i=1}^n a_i B_i()$, wobei B_i wie im vorigen Punkt so gewählt wird, dass $A_i = h^{-1}(B_i)$ gilt..
3. Im allgemeinen Fall stellen wir f als Limes einer Folge f_n von Treppenfunktionen dar; g_n seien die entsprechenden Funktionen mit $f_n = g_n \circ h$. Die Folge g_n muss nicht konvergieren (obwohl sie mit etwas mehr Arbeit so gewählt werden kann), aber $g = \limsup g_n$ existiert auf jeden Fall und leistet auch das Gewünschte (zumindest, wenn wir auch $\pm\infty$ als Werte zulassen; wenn das nicht der Fall ist, also wenn f nur reelle Werte annimmt, kann man unendliche Werte von g einfach durch 0 ersetzen)..

Wir haben gesehen, dass alle stetigen Funktionen Borel-messbar sind und dass die Menge der Borel-messbaren Funktionen bezüglich der Bildung von punktweisen Grenzwerten abgeschlossen ist. Das gibt uns eine Charakterisierung der Borel-messbaren Funktionen:

Satz 4.8

Die Menge $\mathcal{B}$ der Borel-messbaren Funktionen ist die kleinste Menge von Funktionen, die die stetigen Funktionen enthält und gegenüber punktweiser Grenzwertbildung abgeschlossen ist (soll heißen: wenn eine Folge (f_n) von Funktionen aus $\mathcal{B}$ punktweise gegen eine reellwertige Funktion f konvergiert, dann liegt auch f in $\mathcal{B}$).

Wir skizzieren hier den Beweis, die Details sind der Leserin/dem Leser überlassen: wir bezeichnen mit $\mathcal{F}$ die kleinste Menge, die die stetigen Funktionen enthält und gegenüber der punktweisen Konvergenz abgeschlossen ist. Klarerweise gilt $\mathcal{F} \subseteq \mathcal{B}$. Für die andere Richtung zeigt man:

1. $\mathcal{F}$ ist bezüglich Addition und Multiplikation abgeschlossen (Prinzip der guten Mengen).
2. $\{A \subseteq \mathbb{R} : A() \in \mathcal{F}\}$ ist eine Sigmaalgebra (ditto).
3. Für $a < b$ ist $]a, b]() \in \mathcal{F}$ (diese Indikatorfunktion ist nicht schwer als Grenzwert einer Folge von stetigen Funktionen darzustellen).

Wenn diese Punkte gezeigt sind, ist klar, dass alle Indikatoren von Borelmengen in $\mathcal{F}$ liegen, ebenso alle messbaren Treppenfunktionen, und als Grenzwerte davon alle messbaren Funktionen.

4.2 Wahrscheinlichkeitsräume

An diesem Punkt ist es an der Zeit, einige spezielle Bezeichnungen einzuführen, die sich auf Wahrscheinlichkeitsräume beziehen.

Definition 4.3

$(\Omega, \mathfrak{S}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum. Die Elemente von Ω heißen Elementarereignisse, die Elemente der Sigmaalgebra $\mathfrak{S}$ heißen Ereignisse, die leere Menge wird als das unmögliche Ereignis bezeichnet und Ω als das sichere Ereignis. Eine messbare Funktion auf einem Wahrscheinlichkeitsraum heißt Zufallsvariable (meistens ist dann der Wertebereich reell, auch mehrdimensional).

Wir werden in Hinkunft immer wieder spezielle Bezeichnungen für Wahrscheinlichkeitsräume haben.

4.3 Maße und messbare Funktionen

Wir haben bislang nur Messräume betrachtet. Nun soll zusätzlich zu dieser Struktur auch ein Maß definiert sein. Unsere erste Definition betrifft die Verträglichkeit einer messbaren Funktion mit Maßen auf dem Urbild- und Bildraum:

Definition 4.4

$(\Omega_1, \mathfrak{S}_1, \mu_1)$ und $\Omega_2, \mathfrak{S}_2, \mu_2)$ seien zwei Maßräume. Die Funktion $f : (\Omega_1, \mathfrak{S}_1) \to (\Omega_2, \mathfrak{S}_2)$ heißt maßtreu, wenn für alle $A \in \mathfrak{S}_2$

$$\mu_1(f^{-1}(A)) = \mu_2(A).$$

Wenn wir jetzt nur auf dem Urbildraum ein Maß haben, können wir auf dem Bildraum ein Maß definieren, das f zu einer maßtreuen Abbildung macht:

Definition 4.5

$(\Omega, \mathfrak{S}, \mu)$ sei ein Maßraum und $f : (\Omega, \mathfrak{S}) \to (\Omega_2, \mathfrak{S}_2)$ eine messbare Abbildung das Maß

$$\mu_f(A) = \mu(f^{-1}(A))$$

heißt das von f (auf $\mathfrak{S}_2$) induzierte Maß. Ist $(\Omega, \mathfrak{S}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable, dann nennen wir das von X induzierte Maß $\mathbb{P}_X$ die Verteilung von X. Dieses induzierte Maß hat, wenn X Werte in $\mathbb{R}^d$ annimmt, als Lebesgue-Stieltjes-Maß eine Verteilungsfunktion, die wir mit F_X bezeichnen.

Diese Definition macht es uns möglich, die Unabhängigkeit von Ereignissen auf Zufallsvariable zu übertragen:

Definition 4.6

Die Zufallsvariablen $X_1, \ldots, X_n$ $(X_i : (\Omega, \mathfrak{S}) \to (\Omega_i, \mathfrak{S}_i))$ heißen unabhängig, wenn für alle $A_i \in \mathfrak{S}_i$

$$\mathbb{P}(X_i \in A_i, i = 1, \ldots, n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

Für reellwertige Zufallsvariable ist das äquivalent dazu, dass sich die Verteilungsfunktion von $(X_1, \ldots, X_n)$ als Produkt der einzelnen Verteilungsfunktionen schreiben lässt:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \ldots F_{X_n}(x_n).$$

Ein unendliches System von Zufallsvariablen heißt unabhängig, wenn jedes endliche Teilsystem

unabhängig ist.

4.4 Konvergenz von messbaren Funktionen

In diesem Abschnitt studieren wir einige Möglichkeiten, die Konvergenz einer Folge von messbaren Funktionen zu definieren, und die Zusammenhänge, die zwischen diesen Konvergenzarten bestehen. Dazu beginnen wir mit einer allgemeinen Definition:

Definition 4.7

$(\Omega, \mathfrak{S}, \mu)$ sei ein Maßraum und $P = P(\omega)$ eine Aussage über die Elemente $\omega \in \Omega$. Wir sagen, dass P fast überall gilt (bzw. fast sicher, wenn μ ein Wahrscheinlichkeitsmaß ist), wenn es eine Menge $A \in \mathfrak{S}$ gibt, sodass $\mu(A) = 0$ und $P(\omega)$ für alle $\omega \in A^C$ gilt.

Mit dieser Definition können wir schon zwei neue Konvergenzarten definieren, indem wir die entsprechenden Definitionen aus der Analysis "aufweichen":

Definition 4.8

Die Folge f_n von messbaren Funktionen (die auch die Werte $\pm\infty$ annehmen dürfen) heißt fast überall (fast sicher, wenn μ ein Wahrscheinlichkeitsmaß ist) konvergent gegen die Funktion f, wenn es eine Menge A mit $\mu(A) = 0$ gibt, sodass f_n auf dem Komplement von A punktweise gegen f konvergiert.

Die Folge f_n von messbaren reellwertigen Funktionen heißt fast überall gleichmäßig konvergent gegen die messbare Funktion f, wenn es eine Menge $A \in \mathfrak{S}$ gibt, sodass $\mu(A) = 0$ und f_n auf dem Komplement von A gleichmäßig gegen f konvergiert.

Da es für diese Definitionen unerheblich ist, wie sich die entsprechenden Funktionen auf der jeweiligen Ausnahmemenge vom Maß 0 verhalten, können wir auch darauf verzichten, dass die Funktionen dort messbar oder auch nur definiert sind. Das führt zu den Definitionen:

Definition 4.9

$(\Omega_i, \mathfrak{S}_i)$ seien zwei Messräume, μ ein Maß auf $\mathfrak{S}_1$. Wir nennen die Funktion $f : \Omega_1^* \to \Omega_2$ fast überall (auf Ω_1) definiert, wenn es eine Menge $N \in \mathfrak{S}_1$ gibt mit $\mu(N) = 0$ und $\Omega_1^* \subseteq N^C \subseteq \Omega_1$, und fast überall messbar, wenn zusätzlich die Einschränkung von f auf N^C messbar ist.

Wir werden ein wenig schlampig von einer fast überall definierten bzw. fast überall messbaren Funktion von Ω_1 nach Ω_2 sprechen. Zur Rechtfertigung können wir uns vorstellen, dass wir Ω_2 durch ein zusätzliches Symbol "☠" erweitern, das wir als Wert für die Funktion f an den Stellen verwenden, an denen sie ursprünglich nicht definiert ist. Andererseits gibt es zu jeder fast überall messbaren Funktion f eine messbare Funktion $\tilde{f}$, die fast überall mit f übereinstimmt. Auf jeden Fall können wir die beiden Konvergenzdefinitionen auf fast überall messbare (und im Fall der fast überall gleichmäßigen Konvergenz fast überall endliche) Funktionen ausdehnen. Die Limiten für diese Konvergenzen sind offensichtlich nur fast überall eindeutig bestimmt. Generell wird sich hier und weiter unten an vielen Definitionen nichts ändern, wenn wir eine Funktion durch eine andere ersetzen, die mit ihr fast überall übereinstimmt.

Für die gleichmäßige Konvergenz fast überall gibt es eine Norm, sodass die Konvergenz bezüglich dieser Norm genau die gleichmäßige Konvergenz fast überall ist. Dazu definieren wir:

Definition 4.10

Das essentielle Supremum einer messbaren Funktion f ist die kleinste obere Schranke, die f fast überall übertrifft:

$$\operatorname{ess\,sup} f = \inf\{\lambda \in \mathbb{R} : \mu(\{x : f(x) > \lambda\}) = 0\}.$$

Das essentielle Supremum des Betrags von f nennen wir die ∞-Norm:

$$\|f\|_\infty = \operatorname{ess\,sup}|f|.$$

Das Infimum in der Definition ist tatsächlich ein Minimum, also

$$\mu([f > \operatorname{ess\,sup} f]) = 0.$$

Es gilt:

Satz 4.9

Auf der Menge

$$\mathcal{L}_\infty = \{f : \|f\|_\infty < \infty\}$$

ist $\|.\|_\infty$ eine Seminorm (zu einer Norm fehlt nur die Bedingung, dass aus $\|f\| = 0$ $f = 0$ folgt. Wir können nur folgern, dass f fast überall gleich 0 ist). Die Konvergenz bezüglich dieser Seminorm ist genau die gleichmäßige Konvergenz fast überall.

Wenn wir fast überall gleiche Funktionen identifizieren, also die Äquivalenzrelation

$$f \sim g \Leftrightarrow f = g \text{ fast überall}$$

einführen, dann ist

$$L_\infty = \mathcal{L}_\infty / \sim$$

mit $\|.\|_\infty$ ein vollständiger normierter Vektorraum.

Der einzige Teil dieses Satzes, der (vielleicht) interessant ist, ist die Behauptung, dass die Konvergenz bezüglich der ∞-Norm die gleichmäßige Konvergenz fast überall impliziert. Wenn aber $\delta_n = \|f_n - f\|_\infty$ gegen 0 konvergiert, dann gibt es zu jedem n eine Menge A_n mit $\mu(A_n) = 0$, sodass für alle $x \notin A_n$ $|f_n(x) - f(x)| \leq \delta_n$. Die Vereinigung aller A_n hat dann auch das Maß 0, und auf ihrem Komplement konvergiert f_n gleichmäßig gegen f.

Die ∞-Norm spielt also für die gleichmäßige Konvergenz fast überall dieselbe Rolle wie die gewöhnliche Supremumsnorm für die gleichmäßige Konvergenz.

Wir definieren noch zwei Konvergenzarten, die in der Maßtheorie von Bedeutung sind:

Definition 4.11

Die Folge f_n von messbaren reellwertigen Funktionen heißt μ-fast gleichmäßig konvergent, wenn es zu jedem $\epsilon > 0$ eine Menge A_ϵ mit $\mu(A_\epsilon) < \epsilon$ existiert, sodass die Folge f_n auf A^C gleichmäßig konvergiert.

Die Folge f_n von messbaren reellwertigen Funktionen konvergiert im Maß (bzw. in Wahrscheinlichkeit) gegen f, wenn für alle $\epsilon > 0$

$$\lim_{n\to\infty} \mu(\{x : |f_n(x) - f(x)| > \epsilon\}) = 0.$$

Die Konvergenz im Maß ist ein wichtiges Konzept: vom wahrscheinlichkeitstheoretischen (bzw. statistischen) Standpunkt betrachtet ist diese Konvergenz oft genau das, was man haben möchte: wenn man etwa f_n als (aus einer zufällig gewählten Stichprobe der Größe n gewonnene) Schätzwerte für f auffasst, dann garantiert die Konvergenz in Wahrscheinlichkeit, dass die Chance, einen Fehler von einer bestimmten Größe zu erhalten, beliebig klein gemacht werden kann, wenn nur n groß genug gewählt wird.

So wie bei der Konvergenz fast überall können wir auch die Definitionen der Konvergenz fast gleichmäßig und im Maß auf fast überall messbare Funktionen ausdehnen. Bei der fast gleichmäßigen Konvergenz sollte das offensichtlich sein, bei der Konvergenz im Maß ist das nicht so offensichtlich, weil in der Definition das Maß der Menge $[|f_n - f| > \epsilon]$ steht. Der einfachste Weg ist, f_n und f durch messbare reellwertige Funktionen $\tilde{f}_n$ und $\tilde{f}$ zu ersetzen, die fast überall mit f_n bzw. f übereinstimmen. Die Folge (f_n) soll dann im Maß konvergent gegen f heißen, wenn die Folge $(\tilde{f}_n)$ im Maß gegen $\tilde{f}$ konvergiert.

Die fast gleichmäßige Konvergenz ist hauptsächlich von beweistechnischem Interesse. Sie steht zwischen der fast überall gleichmäßigen Konvergenz und den anderen Konvergenzarten dieses Kapitels:

Satz 4.10

- Die fast überall gleichmäßige Konvergenz impliziert die fast gleichmäßige.
- Die fast gleichmäßige Konvergenz impliziert die Konvergenz fast überall.
- Die fast gleichmäßige Konvergenz impliziert die Konvergenz im Maß.

Die (offensichtlichen) Beweise sind der Leserin/dem Leser überlassen.

In endlichen Maßräumen wird der zweite Punkt aus dem letzten Satz zu einer Äquivalenz:

Satz 4.11: Egorov

Wenn μ ein endliches Maß ist, dann folgt aus der Konvergenz μ-fast überall die μ-fast gleichmäßige Konvergenz.

Zum Beweis dieses Satzes setzen wir

$$A_{n,\epsilon} = \{x : |f_k(x) - f(x)| > \epsilon \text{ für ein } k \geq n\}.$$

Da f_n fast überall gegen f konvergiert, gilt

$$\lim \mu(A_{n,\epsilon}) = \mu(\bigcap A_{n,\epsilon}) \leq \mu(\{x : f_n(x) \not\to f(x)\}) = 0.$$

Wir wählen n_k so, dass $\mu(A_{n_k,1/k}) < 2^{-k}$. Die Menge

$$B_k = \bigcup_{i>k} A_{n_i,1/i}$$

hat Maß $< 2^{-k}$, und auf ihrem Komplement konvergiert f_n gleichmäßig gegen f. Damit ist die fast gleichmäßige Konvergenz bewiesen.

Die Konvergenz im Maß ist für endliche Maßräume die schwächste aller Konvergenzarten. In jedem Fall (auch für unendliche Maße) impliziert die fast gleichmäsige Konvergenz die Konvergenz im Maß. Der Satz von Egoroff garantiert dann, dass in endlichen Maßräumen die Konvergenz fast überall die Konvergenz im Maß impliziert. Mit einem kleinen Kunstgriff geht es auch in der umgekehrten Richtung. Dazu definieren wir

Definition 4.12

Die Folge f_n ist eine Cauchy-Folge im Maß, wenn für alle $\epsilon > 0$

$$\lim_{m,n\to\infty} \mu(\{x : |f_n - f_m| > \epsilon\}) = 0.$$

Wir werden beweisen:

Satz 4.12

Falls f_n eine Cauchy-Folge im Maß ist, dann gibt es eine Teilfolge von f_n, die fast gleichmäßig konvergiert (und damit fast überall).

Wir können für jedes k ein N_k finden, sodass für $n, m > N_k$

$$\mu(\{x : |f_n - f_m| > 2^{-k}\}) < 2^{-k}.$$

Wir setzen

$$A_k = \{x : |f_{n_k} - f_{n_{k+1}}| > 2^{-k}.$$

Die Menge $B_n = \bigcup_{k>n} A_{N_k}$ erfüllt offensichtlich, dass $\mu(B_n) < 2^{-n}$, und auf dem Komplement von B_n ist die Folge f_{N_k} eine Cauchyfolge bezüglich der gleichmässigen Konvergenz, und konvergiert daher gleichmäßig.

Insbesondere folgt aus diesem Satz

Satz 4.13

Jede Cauchy-Folge im Maß konvergiert im Maß.

Die Konvergenz im Maß lässt sich metrisieren:

Definition 4.13

Die Lévy-Distanz (Lévy-Metrik) zwischen zwei endlichen messbaren Funktionen ist gegeben durch

$$d(f,g) = \inf\{\epsilon > 0 : \mu([|f-g| > \epsilon]) < \epsilon\}$$

(mit der üblichen Konvention $\inf \emptyset = \infty$).

Satz 4.14

$d(.,.)$ ist eine Pseudometrik auf der Menge $\mathcal{L}_0$ der reellwertigen messbaren Funktionen. f_n konvergiert genau dann im Maß gegen f, wenn $d(f_n, f) \to 0$.

Auf der Menge $L_0 = \mathcal{L}_0/\sim$ der Äquivalenzklassen fast überall übereinstimmender Funktionen aus $\mathcal{L}_0$ ist d eine Metrik.

(L_0, d) ist ein vollständiger metrischer Raum.

Die Beziehungen $d(f,g) \geq 0$ und $d(f,f) = 0$ sollten aus der Definition von d klar sein. Die Dreiecksungleichung

$$d(f,h) \leq d(f,g) + d(g,h)$$

müssen wir nur für den Fall zeigen, dass die rechte Seite endlich ist. Wir wählen $\epsilon > 0$ und a, b so, dass $d(f,g) < a < d(f,g) + \epsilon$ und $d(g,h) < b < d(g,h) + \epsilon$). Dann folgt aus der Definition von d

$$\mu([|f-g|] > a) < a$$

und

$$\mu([|g-h|] > b) < b.$$

Das gibt

$$\mu([|f-h| > a+b]) \leq \mu([|f-g| + |g-h| > a+b) \leq \mu([|f-g| > a] \cup [|g-h| > b]) \leq$$

$$\mu([|f-g| > a]) + \mu([|g-h| > b]) < a + b.$$

Es ist also

$$d(f,h) \leq a + b < d(f,g) + d(g,h) + 2\epsilon,$$

und, weil $\epsilon > 0$ beliebig war,

$$d(f,h) \leq a + b < d(f,g) + d(g,h).$$

Als nächstes wählen wir eine Folge f_n, die gegen eine Funktion f im Maß konvergiert. Für $\epsilon > 0$ geht $\mu([|f_n - f| > \epsilon])$ gegen 0, also gibt es ein n_0, sodass für $n \geq n_0$

$$\mu([|f_n - f| > \epsilon]) < \epsilon.$$

Deshalb ist für $n \geq n_0$

$$d(f_n, f) \leq \epsilon.$$

Weil das für jedes $\epsilon > 0$ funktioniert, geht $d(f_n, f)$ gegen 0.

Geht umgekehrt $d(f_n, f)$ gegen 0, dann gibt es zu jedem $\epsilon > 0$ und $0 < \delta < \epsilon$ ein n_0, sodass für $n \geq n_0$

$$d(f_n, f) < \delta,$$

und daher

$$\mu([|f_n - f| > \epsilon]) < \mu([|f_n - f| > \delta) < \delta),$$

also geht $\mu([|f_n - f| > \epsilon])$ gegen 0, und f_n konvergiert im Maß gegen f.

Als nächstes überlegen wir uns, wann $d(f, g) = 0$ gilt. Dazu muss für jedes $\epsilon > 0$

$$\mu([|f - g| > \epsilon]) < \epsilon$$

gelten, und daher für $0 < \delta < \epsilon$)

$$\mu([f - g| > \epsilon]) \leq \mu([|f - g| > \delta]) < \delta.$$

Daraus folgt

$$\mu([f - g| > \epsilon]) = 0.$$

Schließlich ergibt sich, dass $f = g$ fast überall gelten muss. Das bedeutet natürlich auch, dass die Grenzfunktion bezüglich der Konvergenz im Maß fast überall eindeutig bestimmt ist.

In endlichen Maßräumen ist die Konvergenz im Maß die schwächste aller Konvergenzen, in Räumen, die nicht endlich sind, kann es vorkommen, dass die Konvergenz im Maß sogar stärker ist als die Konvergenz fast überall, im allgemeinen Fall kann man Folgen finden, die im Maß aber nicht fast überall konvergieren, und auch solche, die es umgekehrt halten. Das bringt uns (insbesondere wenn wir im nächsten Kapitel über die Konvergenz von Integralen sprechen) in die Verlegenheit, unsere Sätze jeweils in zwei Versionen zu formulieren, einmal für die Konvergenz im Maß, einmal für die fast überall. Besser wäre es, einen kleinsten gemeinsamen Nenner für alle Konvergenzarten zu finden, und Satz 4.12 zeigt uns den Weg dazu:

Definition 4.14

Die Folge f_n von fast überall messbaren Funktionen auf dem Maßraum $(\Omega, \mathfrak{S}, \mu)$ konvergiert gegen die Funktion f fast überall entlang Teilfolgen, wenn es zu jeder Teilfolge f_{n_k} eine (Unter-) Teilfolge $f_{n_{k_l}}$ gibt, die μ-fast überall gegen f konvergiert.

Diese Konvergenz ist nicht nur als kleinster gemeinsamer Nenner der anderen Konvergenzarten nützlich. Die Konvergenz im Maß bleibt nämlich im allgemeinen unter stetigen Transformationen nicht erhalten, wenn also $f_n \to f$ im Maß und $g : \mathbb{R} \to \mathbb{R}$ stetig ist, dann konvergiert im allgemeinen $g \circ f_n \to g \circ f$ nicht im Maß. Es folgt aber offensichtlich aus der Teilfolgenkonvergenz von f_n to f die von $g \circ f_n$ gegen $g \circ f$.

Abschließend zusammengefasst ergibt sich also folgendes Bild:

Fast überall gleichmäßig impliziert fast gleichmäßig impliziert sowohl im Maß als auch fast überall, und jede der beiden die fast überall Konvergenz entlang Teilfolgen.

Die fast gleichmäßige Konvergenz entlang Teilfolgen ist dasselbe wie die Konvergenz im Maß.

In endlichen Maßräumen impliziert die fast überall Konvergenz die fast gleichmäßige (und ist daher dazu äquivalent) und die im Maß. Dort ist auch die Konvergenz im Maß äquivalent zur Konvergenz fast überall entlang Teilfolgen.

Die Konvergenz im Maß ist in endlichen Maßräumen die schwächste, aber (Stichwort Teilfolgenkonvergenz) man kann aus jeder im Maß konvergenten Folge eine Teilfolge aussondern, die fast gleichmäßig konvergiert.

4.5 Ergänzungen

4.5.1 Folgenkonvergenz und Teilfolgenkonvergenz

Die Konvergenzen fast überall und fast gleichmäßig können im Gegenssatz zu der im Maß oder der fast überall gleichmäßigen Konvergenz nicht durch eine Metrik (oder allgemeiner eine Topologie) induziert werden. Das hat Fréchet dazu inspiriert, Folgenkonvergenz unabhängig von einer Topologie zu definieren:

Definition 4.15

X sei eine beliebige nichtleere Menge Eine C-Konvergenzstruktur ist ein System von Teilmengen $(\Lambda(x), x \in X)$ von $X^{\mathbb{N}}$ mit den Eigenschaften

$C1$ $(x, x, \ldots) \in \Lambda(x)$ (konstante Folgen konvergieren),

$C2$ Falls $(x_n)_{n\in\mathbb{N}} \in \Lambda(x)$ und (y_n) eine Teilfolge von (x_n), dann ist auch $(y_n) \in \Lambda(x)$.

Gilt zusätzlich

L $\Lambda(x) \cap \Lambda(y) = \emptyset$ für $x \neq y$ (Grenzwerte sind eindeutig bestimmt),

dann heißt Λ eine L-Konvergenz, gilt darüber hinaus

L^* Falls jede Teilfolge (y_n) eine (Unter-) Teilfolge $(z_n) \in \Lambda(x)$ enthält, dann ist $(x_n) \in \Lambda(x)$.

dann nennen wir Λ eine L^*-Konvergenz.

$\Lambda(x)$ versteht sich natürlich als die Kollektion aller Folgen, die gegen x konvergieren. Statt $(x_n) \in \Lambda(x)$ schreibt man auch $(x_n) \to_\Lambda x$.

Definition 4.16

Λ sei eine C-Konvergenz. Eine Menge $A \subseteq X$ heißt $(\Lambda-)$ folgenoffen, wenn für jedes $x \in A$ und $(x_n) \to x$ ein n_0 existiert, sodass für $n \geq n_0$ $x_\in A$ gilt.

$A \subseteq X$ heißt folgenabgeschlossen, wenn für jede Folge $(x_n) \in \Lambda(x)$ mit $x_n \in A, n \in \mathbb{N}$ auch $x \in A$ gilt (das sind die klassischen Definitionen).

Jede topologische Konvergenz ist eine L^*-Konvergenz. Umgekehrt gibt es zu jeder Folgenkonvergenz eine natürliche Topologie $\tau(\Lambda)$, nämlich die feinste Topologie, in der alle Λ-konvergenten Folgen konvergieren. In dieser sind genau die folgenabgeschlossenen Mengen abgeschlossen bzw. alle folgenoffenen Mengen offen.

Definition 4.17

Die Folge f_n von fast überall messbaren und fast überall endlichen Funktionen auf dem Maßraum $(\Omega, \mathfrak{S}, \mu)$ konvergiert lokal im Maß gegen die fast überall messbare und fast überall endliche Funktion f_n, wenn f_n auf jeder Menge $A \in \mathfrak{S}$ mit $\mu(A) < \infty$ im Maß gegen f konvergiert.

Wenn eine Folge f_n die Eigenschaft hat, dass jede Teilfolge eine fast überall gegen f konvergente Teilfolge besitzt, dann konvergiert f_n lokal im Maß; die Umkehrung gilt in sigmaendlichen Maßräumen; in nicht sigmaendlichen Räumen hat die lokale Konvergenz im Maß nicht allzu großen Wert (etwa wenn μ nur die Werte 0 und ∞ annimmt).

4.6 Wiederholungsfragen

1. Wann heißt eine Funktion messbar?
2. Wie kann man (einfach) feststellen, ob eine Funktion messbar ist?
3. Wann heißt eine Funktion Borel-messbar?
4. Wann heißt eine Funktion Lebesgue-messbar?
5. Welche Operationen kann man mit messbaren Funktionen durchführen?
6. Was besagt der Approximationssatz für messbare Funktionen?
7. Welche Konvergenzarten gibt es in der Maßtheorie?

8. Welche Beziehungen bestehen zwischen den einzelnen Konvergenzarten?
9. Wann heißt eine Funktion maßtreu?
10. Was ist das von einer messbaren Funktion induzierte Maß?
11. Erklären Sie die Räume L_0 und L_∞.

Kapitel 5

Das Integral

5.1 Definition

Wir sind jetzt endlich in der Lage, das "neue, bessere" Integral zu definieren, das am Beginn versprochen wurde.

Wir haben schon gesagt, dass wir für nichtnegative messbare Funktionen f das Integral als

$$\int f d\mu = \int_0^\infty \mu([f > y])dy$$

definieren wollen. Wenn f auch negative Werte annimmt, wird das Integral natürlich als "Fläche über der X-Achse"-"Fläche unter der X-Achse" definiert, also

Definition 5.1

$(\Omega, \mathfrak{S}, \mu)$ sei ein Maßraum. Das Integral der Funktion $f : (\Omega, \mathfrak{S}) \to (\bar{\mathbb{R}}, \mathfrak{B}(\mathbb{R}))$ ist gegeben durch

$$\int f d\mu = \int_0^\infty \mu([f > y])dy - \int_0^\infty \mu([f < -y])dy,$$

wobei wir die Integrale auf der rechten Seite als uneigentliche Riemannintegrale ansehen, deren Definition wir noch dadurch ergänzen, dass wir sie gleich ∞ setzen, wenn der Integrand für ein $y > 0$ unendlich ist. Wenn beide Integrale auf der rechten Seite den Wert ∞ haben, ist das Integral von f nicht definiert, ist mindestens eines endlich, dann hat das Integral einen bestimmten Wert, wir sagen dann, dass das Integral von f existiert, und nennen f quasiintegrierbar. Sind beide Integrale auf der rechten Seite endlich und damit auch $\int f d\mu$, dann nennen wir f integrierbar.

Wir werden zuerst die Eigenschaften von Integralen nichtnegativer Funktionen betrachten; für diese brauchen wir nicht über die Existenz des Integrals nachzudenken:

Satz 5.1

Eigenschaften des Integral von nichtnegativen Funktionen

1. Wenn $\int f = 0$, dann ist fast überall $f = 0$.
2. Falls $f \geq g$, dann $\int f \geq \int g$.
3. $\int cf = c \int f$ ($c > 0$ konstant).
4. $\int (f + g) = \int f + \int g$.

5. Das sogenannte unbestimmte Integral

$$\int_A f(x)d\mu(x) = \int A(x)f(x)d\mu(x)$$

ist ein Maß.

6. Falls μ sigmaendlich ist, folgt aus $\int_A f \geq \int_A g$ für alle A, dass μ-fast überall $f \geq g$ gilt.

Beweis:

1. Damit das Integral gleich 0 ist, muss für jedes $\epsilon > 0$ $\mu([f > \epsilon]) = 0$ gelten. Dann hat aber auch $[f > 0] = \bigcup_n [f > 1/n]$ Maß 0.
2. Das folgt direkt aus $[g > y] \subseteq [f > y]$.
3. $\int cf = \int_0^\infty \mu([f > y/c])dy$ und mit der Substitution $x = cy$ folgt die Behauptung.
4. Das ist mit einer direkten Anwendung unserer Integraldefinition nicht so einfach zu zeigen. Deshalb betrachten wir zuerst das Integral einer Treppenfunktion:

Satz 5.2

Wenn $f = \sum_{i=1}^n a_i A_i$ eine nichtnegative Treppenfunktion ist, dann ist

$$\int f = \sum_{i=1}^n a_i \mu(A_i).$$

Das ergibt sich aus der Tatsache, dass

$$\mu([f > y]) = \mu(\bigcup_{i:a_i > y} A_i) = \sum_{i:a_i > y} \mu(A_i)$$

gilt. Damit ergibt sich, dass das Integral einer Summe von zwei nichtnegativen Treppenfunktionen gleich der Summe der einzelnen Integrale ist (die Summe der Treppenfunktionen $\sum_i a_i A_i()$ und $\sum_j b_j B_j()$ können wir als $\sum_{i,j}(a_i + b_j)(A_i \cap B_j)()$ schreiben).

Der allgemeine Fall folgt dann daraus, dass jede nichtnegative messbare Funktion als Grenzwert einer nichtfallenden Folge von Treppenfunktionen dargestellt werden kann, und aus

Satz 5.3: monotone Konvergenz, Beppo Levi-Theorem

Falls $f_n \geq 0$ eine monoton nichtfallende Folge ist, dann gilt

$$\lim \int f_n = \int \lim f_n.$$

Die Ungleichung $\lim \int f_n \leq \int \lim f_n$ folgt aus Punkt 2. Für die umgekehrte Ungleichung setzen wir $f = \lim f_n$ und wählen $M < \int \lim f_n$. Weil das uneigentliche Riemannintegral $\int_0^\infty \mu([f > y])dy$ der Limes der Integrale $\int_a^b \mu([f > y])dy$ für $a \to 0$ und $b \to \infty$ ist und dieses Integral wiederum der Grenzwert von entsprechenden Untersummen, können wir Zahlen $0 \leq t_0 < \ldots < t_N < \infty$ finden, sodass

$$\sum_{i=1}^N (t_i - t_{i-1})\mu([f > t_i]) > M$$

gilt. Weil $\mu([f > t_i]) = \lim_n \mu([f_n > t_i])$ gilt, haben wir für hinreichend großes n

$$\int f_n \geq \sum_{i=1}^N (t_i - t_{i-1})\mu([f_n > t_i]) > M,$$

und daher

$$\lim_n \int f_n > M,$$

und weil $M < \int f$ beliebig gewählt werden kann, auch $\lim \int f_n \geq \int f$.

5. Die Additivität von

$$\nu(A) = \int_A f$$

ergibt sich aus der Additivität des Integrals, die Sigmaadditivität folgt daraus durch den Satz von der monotonen Konvergenz.

6. Wir setzen für zwei rationale Zahlen $0 \leq a < b$

$$A_{ab} = \{x : f(x) \leq a, g(x) \geq b\}.$$

Falls diese Menge positives Maß hat, gibt es wegen der Sigmaendlichkeit von μ eine Teilmenge B davon, die endliches positives Maß hat. Für diese Menge ist

$$\int_B g \geq b\mu(B) > a\mu(B) \geq \int_B f,$$

im Widerspruch zur Voraussetzung. Daher ist $\mu(A_{ab}) = 0$ und

$$\mu(\{x : f(x) < g(x)\}) = \mu(\bigcup_{a,b\in\mathbb{Q},a<b} A_{ab}) = 0,$$

also ist μ-fast überall $f \geq g$.

Das Integral einer messbaren Funktion f kann als

$$\int f = \int f_+ - \int f_-$$

geschrieben werden (wieder vorausgesetzt, dass die rechte Seite definiert ist). Damit erhalten wir

Satz 5.4

Das Integral hat folgende Eigenschaften:

1. Falls $f \geq g$, dann ist $\int f \geq \int g$.
2. $\int(f + g) = \int f + \int g$, falls die rechte Seite definiert ist.
3. Für $c \in \mathbb{R}$ gilt $\int cf = c \int f$.
4. Das unbestimmte Integral $\int_A f$ ist eine sigmaadditive Funktion.
5. Falls μ sigmaendlich ist folgt aus $\int_A f \geq \int_A g$ für alle A, dass μ-fast überall $f \geq g$ gilt.

Diese Eigenschaften ergeben sich unmittelbar aus den analogen Eigenschaften für nichtnegative Funktionen. Nur die Additivität ist möglicherweise nicht sofort erkennbar. Dafür nehmen wir ohne Beschränkung der Allgemeinheit an, dass $\int f_-$ und $\int g_-$ endlich sind. Dann ist wegen $(f + g)_- \leq f_- + g_-$ auch das Integral von $f + g$ definiert. Wegen

$$f + g = (f + g)_+ - (f + g)_- = f_+ - f_- + g_+ - g_-$$

gilt

$$(f + g)_+ + f_- + g_- = (f + g)_- + f_+ + g_+,$$

und weil hier nur nichtnegative Funktionen addiert werden,

$$\int (f + g)_+ + \int f_- + \int g_- = \int (f + g)_- + \int f_+ + \int g_+,$$

und weil die Integrale der Negativteile endlich sind, können wir sie auf beiden Seiten abziehen und erhalten die Behauptung.

Schließlich betrachten wir den Fall, dass wir zwei Funktionen verknüpfen:

Satz 5.5: Transformationssatz

$(\Omega_i, \mathfrak{S}_i), i = 1, 2$ seien zwei Messräume, $g : (\Omega_1, \mathfrak{S}_1) \to (\Omega_2, \mathfrak{S}_2)$, $f : (\Omega_2, \mathfrak{S}_2) \to (\mathbb{R}, \mathfrak{B})$ und μ ein Maß auf $\mathfrak{S}_1$. Dann gilt

$$\int f \circ g d\mu = \int f d\mu_g.$$

Wir müssen nur feststellen, dass $\mu([f \circ g > y]) = \mu_g([f > y])$ und $\mu([f \circ g < y]) = \mu_g([f < y])$.

5.2 Grenzwertsätze

Wir wollen jetzt untersuchen, wann wir Integration und Grenzübergang vertauschen können. Einen Satz dazu haben wir schon kennengelernt, nämlich den Satz von der monotonen Konvergenz, den wir etwas allgemeiner formulieren:

Satz 5.6: monotone Konvergenz 2

Falls f_n eine monoton nichtfallende Folge von messbaren Funktionen ist und eine Funktion g existiert mit $f_n \geq g$ und $\int g > -\infty$, dann ist

$$\lim \int f_n = \int \lim f_n.$$

Zum Beweis braucht man nur die ursprüngliche Version des Satzes auf die Funktionen $f_n - g$ anzuwenden.

Aus diesem Satz erhalten wir

Satz 5.7: Lemma von Fatou

1. Falls $f_n \geq g$ mit $\int g > -\infty$, dann ist
$$\int \liminf f_n \leq \liminf \int f_n.$$
2. Falls $f_n \leq g$ mit $\int g < \infty$, dann ist
$$\int \limsup f_n \geq \limsup \int f_n.$$
3. Falls $|f_n| \leq g$ mit $\int g < \infty$, dann ist
$$\int \liminf f_n \leq \liminf \int f_n \leq \limsup \int f_n \leq \int \limsup f_n.$$

Zum Beweis von 1. stellen wir fest, dass $\liminf f_n = \lim g_n$, wobei $g_n = \inf_{k \geq n} f_k$. g_n ist eine nichtfallende Folge, und es gilt $g \leq g_n \leq f_n$. Wir können also auf die Folge g_n den Satz von der monotonen Konvergenz anwenden, und erhalten

$$\int \liminf f_n = \int \lim g_n = \lim \int g_n = \liminf \int g_n \leq \liminf \int f_n.$$

Punkt 2. folgt aus Punkt 1., wenn wir die Folge $-f_n$ betrachten, und Punkt 3. ist eine Kombination aus den ersten beiden. Wenn zusätzlich f_n konvergiert, dann ergibt sich

Satz 5.8: Satz von der dominierten Konvergenz

Falls $|f_n| \leq g$ mit $\int g < \infty$ und f_n μ-fast überall entlang Teilfolgen gegen f konvergiert, dann

gilt

$$\int f = \lim \int f_n.$$

Wir nehmen zuerst an, dass f_n punktweise gegen f konvergiert. In diesem Fall brauchen wir nur festzustellen, dass die äußeren Glieder im Lemma von Fatou übereinstimmen.

Wenn f_n fast überall gegen f konvergiert, dann gibt es eine Menge $A \in \mathfrak{S}$ mit $\mu(A^C) = 0$ und $f_n(\omega) \to f(\omega)$ für alle $\omega \in A$, also $A()f_n \to A()f$ punktweise. Dann ist

$$\int f_n = \int_A f_n \to \int_A f = \int f.$$

Im allgemeinen Fall können wir eine Teilfolge f_{n_k} finden, für die

$$\lim_{k\to\infty} |\int f_{n_k} - \int f| = \limsup_n |\int f_n - \int f|.$$

Aus dieser Folge können wir eine (Unter-)Teilfolge $f_{n_{k_l}}$ aussuchen, die fast überall gegen f konvergiert. Wieder folgt aus dem, was wir schon bewiesen haben, dass $\int f_{n_{k_l}} \to \int f$ gilt. Deshalb ist

$$\limsup_n |\int f_n - \int f| = 0,$$

also gilt auch jetzt

$$\int f_n \to \int f.$$

Das Lemma von Fatou und der Satz von der dominierten Konvergenz lassen sich sehr schön auf Reihen anwenden und können dabei einiges an ϵ-Akrobatik ersparen. Das funktioniert so gut, dass hier eigentlich der Schlachtruf "Nie wieder Epsilon!" stehen könnte, wäre nicht...

Die Bedingungen dieses Satzes von der dominierten Konvergenz lassen sich noch etwas abschwächen. Dazu definieren wir

Definition 5.2: Gleichmäßige Integrierbarkeit

Die Folge f_n heißt gleichmäßig integrierbar wenn zu jedem $\epsilon > 0$ eine integrierbare Funktion $g_\epsilon \geq 0$ existiert, sodass für alle n

$$\int |f_n|[|f_n| > g_\epsilon] \leq \epsilon. \tag{5.1}$$

Allgemeiner heißt eine Menge $\mathcal{F}$ von integrierbaren Funktionen gleichmäßig integrierbar, wenn (5.1) für alle $f \in \mathcal{F}$ gilt.

Im wichtigen Fall, dass μ endlich ist, gilt

Satz 5.9

Wenn μ endlich ist, dann sind die folgenden Aussagen äquivalent:

1. $\mathcal{F}$ ist gleichmäßig integrierbar.
2. Zu jedem $\epsilon > 0$ gibt es ein $M = M(\epsilon)$, sodass für alle $f \in \mathcal{F}$

   $$\int |f|[|f| > M] < \epsilon$$

   (wir können in diesem Fall also g_ϵ konstant wählen).
3. $\{\int |f| : f \in \mathcal{F}\}$ ist beschränkt, und zu jedem $\epsilon > 0$ gibt es ein $\delta = \delta(\epsilon) > 0$, sodass aus $\mu(A) < \delta$

   $$\int_A |f| < \epsilon$$

für alle $f \in \mathcal{F}$ folgt.

4. Es gibt eine Funktion $h : \mathbb{R}_+ \to \mathbb{R}_+$ mit $\lim_{t\to\infty} h(t)/t = \infty$, sodass

$$\{\int h(|f(\omega)|)\mu(d\omega) : f \in \mathcal{F}\}$$

beschränkt ist.

Bemerkungen:

1. Im Sinne der Definition für endliche Maße könnten wir $g_\epsilon = M_\epsilon h$ verlangen, mit Konstanten M_ϵ und einer integrierbaren Funktion h, die für alle ϵ dieselbe ist.
2. Wegen $2(|f|-g/2)_+ \geq |f|[|f| > g] \geq (|f|-g)_+$ ist die gleichmäßige Integrierbarkeit äquivalent dazu, dass es zu jedem $\epsilon > 0$ ieine integrierbare Funktion g_ϵ gibt, sodass

$$\int (|f_n| - g_\epsilon)_+ \leq \epsilon.$$

3. Es genügt, eine Funktion g_ϵ zu finden, die (5.1) ab einem Index erfüllt (der von ϵ abhängen darf).
4. Die Funktion h in Punkt 4 des Satzes kann nichtfallend und konvex gewählt werden.

Aus Punkt 2. des Satzes folgt unmittelbar die gleichmäßige Integrierbarkeit, weil die konstante Funktion $g = M$ integrierbar ist. F'ur die Gegenrichtung verwenden wir Bemerkung 2 und die Subbadditivität von $(.)_+$:

$$\int (|f| - M)_+ \leq \int (|f| - g_{\epsilon/2})_+ + \int (g - M)_+ \leq \frac{\epsilon}{2} + \int (g - M)_+.$$

Das letzte Integral kann kleiner als $\epsilon/2$ gemacht werden, indem M hinreichend groß gewählt wird.

Aus der gleichmäßigen Integrierbarkeit folgt die Beschränktheit der Integrale. Der zweite Teil von Punkt 3 ergibt sich für

$$\delta = \frac{\epsilon}{2M_{\epsilon/2}}.$$

Für die umgekehrte Richtung wählen wir $K > 0$ so, dass

$$\int |f| \leq K$$

für alle $f \in \mathcal{F}$ gilt. Dann können wir $M(\epsilon) = K/\delta(\epsilon)$ setzen.

Wenn Punkt 4 gilt, dann können wir K so wählen, dass für alle $f \in \mathcal{F}$

$$\int h \circ |f| \leq K.$$

Dann genügt es, M so zu wählen, dass für $t \geq M$

$$h(t) > \frac{tK}{\epsilon}$$

gilt. Umgekehrt funktioniert

$$h(t) = t \sum_{n\in\mathbb{N}} [M_{2^{-n}}, \infty[(t).$$

Satz 5.10

Wenn $f_n \to f$ fast überall entlang Teilfolgen und (f_n) gleichmäßig integrierbar ist, dann gilt

$$\lim \int f_n = \int f.$$

Offensichtlich bleibt $\int |f_n|$ beschränkt, und das Lemma von Fatou impliziert, dass auch f integrierbar ist.

Wir können annehmen, dass $g_\epsilon \geq |f|$ gilt (wir können ja g_ϵ durch $\max(|f|, g_\epsilon)$ ersetzen).

Wir setzen $h_n = \max(-g_\epsilon, \min(f_n, g_\epsilon))$ und stellen fest, dass

$$|\int f_n - \int h_n| \leq \int |f_n - h_n| = \int (|f_n| - g_\epsilon)_+ \leq \epsilon.$$

$\int h_n$ konvergiert aber nach dem Satz von der dominierten Konvergenz gegen $\int f$, also unterscheidet sich $\int f_n$ für hinreichend großes n um weniger als 2ϵ von $\int f$.

Anmerkung: in den beiden letzten Sätzen ergibt sich durch Anwendung desselben Satzes auf $|f_n - f|$, dass auch $\int |f_n - f|$ gegen 0 geht.

Die gleichmäßige Integrierbarkeit ist die schwächste "vernünftige" Bedingung, die die Vertauschbarkeit von Integral und Grenzübergang sicherstellt. In der Tat gilt

Satz 5.11

Falls $f_n \geq 0$, $f_n \to f$ fast überall entlang Teilfolgen und $\int f_n \to \int f < \infty$, dann ist (f_n) gleichmäßig integrierbar.

Beweis:

$$\inf f_n = \int (f_n - f)_+ + \int \min(f_n, f).$$

Die linke Seite (laut Voraussetzung) und der zweite Summand auf der rechten (wegen dominierter Konvergenz) konvergieren gegen $\int f$. Daher geht der erste Summand auf der rechten Seite gegen 0, ist also für hinreichend großes n kleiner ϵ, und nach Anmerkung 3 nach der Definition ist (f_n) gleichmäßig integrierbar.

5.3 Riemann-Integrale

In diesem Abschnitt wollen wir den Zusammenhang zwischen der Integration im Riemannschen Sinn und dem Lebesgue-Integral untersuchen. Dazu haben wir

Satz 5.12

Die Funktion $f : [a, b] \to \mathbb{R}$ ist genau dann Riemann-integrierbar, wenn sie beschränkt und fast überall stetig ist (also, wenn das Lebesguemaß der Menge ihrer Unstetigkeitsstellen 0 ist).

f ist dann Lebesgue-messbar, und das Riemann-Integral stimmt mit dem Lebesgue-Integral überein:

$$\int_a^b f(x)dx = \int_{[a,b]} f d\lambda.$$

Die Beschränktheit ist jedenfalls notwendig, da sonst die Obersummen (bzw. die Untersummen, wenn f nach unten unbeschränkt ist) unendlich werden.

Wir wählen die Teilungsfolge

$$t_{n,i} = a + \frac{(b-a)i}{2^n}, \ i = 0, \ldots 2^n.$$

Damit definieren wir die Funktionen

$$\bar{f}_n = \sum_{i=1}^{2^n}]t_{n,i-1}, t_{n,i}] () \sup f([t_{n,i-1}, t_{n,i}])$$

und

$$\underline{f}_n = \sum_{i=1}^{2^n}]t_{n,i-1}, t_{n,i}] () \inf f([t_{n,i-1}, t_{n,i}]).$$

$\int_{]a,b]} \bar{f}_n d\lambda$ und $\int_{]a,b]} \underline{f}_n d\lambda$ sind genau die n-te Obersumme und Untersumme für das Riemannintegral.

Die Folge $\bar{f}_n$ ist monoton nichtwachsend, $\underline{f}_n$ nichtfallend und $\underline{f}_n \leq f \leq \bar{f}_n$ auf $]a,b]$. Es existieren also $\underline{f} = \lim_n \underline{f}_n$ und $\bar{f} = \lim_n \bar{f}_n$. Nach dem Satz von der monotonen Konvergenz ist

$$\int_{]a,b]} \underline{f} d\lambda = \lim_n \int_{]a,b]} \underline{f}_n d\lambda$$

und

$$\int_{]a,b]} \bar{f} d\lambda = \lim_n \int_{]a,b]} \bar{f}_n d\lambda.$$

Die rechten Seiten sind die Grenzwerte der Obersummen und der Untersummen, und das Riemannintegral existiert genau dann, wenn sie übereinstimmen. Das passiert genau dann, wenn

$$\int_{]a,b]} \bar{f} d\lambda = \int_{]a,b]} \underline{f} d\lambda.$$

Das wiederum ist genau dann der Fall, wenn $\bar{f}$ und $\underline{f}$ λ-fast überall übereinstimmen. Es ist leicht einzusehen, dass für ein $x \in]a,b]$, dass nicht mit einem der abzählbar vielen Punkte $t_{n,i}$ übereinstimmt, $\bar{f}(x) = \underline{f}(x)$ genau dann gilt, wenn f in x stetig ist. Die Menge der Unstetigkeitspunkte von f unterscheidet also nur um eine abzählbare Menge von der Menge der Punkte, in denen sich $\bar{f}$ und $\underline{f}$ unterscheiden, und daher haben beide Mengen dasselbe Lebesguemaß. Es existiert also das Riemannintegral genau dann, wenn die Menge der Unstetigkeitsstellen von f Lebesguemaß 0 hat. f muss dann nicht unbedingt Borel-messbar sein, aber weil f fast überall mit den messbaren Funktionen $\bar{f}$ und $\underline{f}$ übereinstimmen muss, ist f fast überall Borel-messbar (also Lebesgue-messbar), $\int_{]a,b]} f d\lambda$ existiert also und stimmt mit dem Riemannintegral überein.

Uneigentliche Riemannintegrale lassen sich nicht immer direkt in Lebesgueintegrale übersetzen — wenn das Integral absolut konvergiert, dann existiert auch das Lebesgue-Integral und stimmt mit dem Riemannintegral überein (das folgt aus den Sätzen von der monotonen und dominierten Konvergenz). Im anderen Fall existiert das Lebesgueintegral nicht. Wir können dann aber immer noch das uneigentliche Riemannintegral als Limes von Lebesgueintegralen darstellen, etwa

$$\int_0^\infty \frac{\sin(x)}{x} dx = \lim_{T\to\infty} \int_{[0,T]} \frac{\sin(x)}{x} d\lambda(x).$$

Wenn man die Überlegungen von vorhin weitertreibt, kommt man zum folgenden

Satz 5.13

F sei eine Verteilungsfunktion, die in den Punkten $a_1 < a_2 < \ldots < a_n$ Sprünge hat und dazwischen stetig differenzierbar ist, und $f \geq 0$ sei eine stetige Funktion. Dann ist

$$\int f d\mu_F = \sum_{i=1}^{n+1} \int_{a_{i-1}}^{a_i} f(x)F'(x)dx + \sum_{i=1}^{n} f(a_i)(F(a_i) - F(a_i - 0),$$

wobei wir $a_0 = -\infty$ und $a_{n+1} = \infty$ setzen.

5.4 Komplexe Integrale

Wir haben schon früher festgestellt, dass wir eine komplexwertige Funktion $f = f_1 + if_2$ als Paar (f_1, f_2) von reellen Funktionen ansehen können, und dass wir mit dieser Interpretation die komplexen Borelmengen mit den zweidimensionalen Borelmengen identifizieren. Die komplexwertige Funktion f ist also genau dann messbar, wenn ihr Real- und Imaginärteil messbar sind.

Damit kommen wir zu der natürlichen Definition

Definition 5.3

$f = f_1 + i\omega f_2$ sei messbar auf $(\Omega, \mathfrak{S})$, und $|f|$ sei integrierbar. Dann definieren wir das Integral von f als

$$\int f d\mu = \int f_1 d\mu + i \int f_2 d\mu.$$

Anmerkung: Diese Definition ließe sich allgemeiner gestalten, wenn man nur annimmt, dass Real- und Imaginärteil von f quasiintegrierbar sind. Wir beschränken uns hier auf integrierbare Funktionen, weil wir ja auch bei komplexwertigen sigmaadditiven Funktionen nur solche zulassen, die nur endliche Werte annehmen. Das hat einerseits ästhetische Gründe, andererseits ersparen wir uns dadurch die Peinlichkeit, dass der Fall eintreten kann, dass das Integral einer Funktion f existiert, aber nicht das von cf mit einer komplexen Konstante c.

Für das Integral von komplexen Funktionen gelten analoge Eigenschaften wie für das reelle Integral:

Satz 5.14

1. $\int (af + bg) = a \int f + b \int g$,
2. $|\int f| \leq \int |f|$,
3. wenn μ ein Wahrscheinlichkeitsmaß ist und f und g unabhängig, dann ist $\int fg d\mu = \int f d\mu \int g d\mu$.
4. $\nu(A) = \int_A f d\mu$ ist sigmaadditiv.

Auch die Sätze von der dominierten Konvergenz und von der Konvergenz durch gleichmäßige Integrierbarkeit bleiben für Integrale von komplexwertigen Funktionen gültig.

Diese Eigenschaften folgen direkt aus denen für reelle Integrale, nur (2) ist vielleicht nicht ganz offensichtlich. Wir können ein z mit $|z| = 1$ finden, sodass

$$|\int f| = z \int f$$

gilt. Dann ist

$$|\int f| = \Re(|\int f|) = \Re(z \int f) = \int \Re(zf) \leq \int |zf| = \int |f|.$$

Wir haben gesehen, dass für eine (quasi-)integrierbare reelle oder komplexe Funktion ihr unbestimmtes Integral eine sigmaadditive Mengenfunktion darstellt. Eine solche Funktion werden wir auch ein signiertes bzw. komplexes Maß nennen. Die Frage liegt nahe, ob jede solche reell- oder komplexwertige Maßfunktion sich als unbestimmtes Integral einer geeignet gewählten Funktion bezüglich eines geeignet gewählten Maßes darstellen lässt. Im nächsten Kapitel werden wir diese Frage für reelle signierte Maße (mit ja) beantworten; im komplexen Fall müssen wir uns dazu zuerst mit der generellen Frage beschäftigen, wann eine Maßfunktion als Integral bezüglich einer anderen darstellbar ist, und das wird im übernächsten Kapitel geschehen.

5.5 Erwartungswerte

$(\Omega, \mathfrak{S}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, X eine reellwertige Zufallsvariable. Dann nennen wir

$$\mathbb{E}(X) = \int X d\mathbb{P}$$

den Erwartungswert von X. Wenn es nötig ist, das Maß, bezüglich dem wir integrieren, zu spezifizieren, schreiben wir $\mathbb{E}_{\mathbb{P}}$.

Neben den bekannten Eigenschaften des Integrals (Linearität, Additivität, etc.) können wir feststellen

Satz 5.15

X und Y seien unabhängig und integrierbar (oder beide nichtnegativ). Dann gilt

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Wir können uns auf den Fall beschränken, dass X und Y nichtnegativ sind, weil $(XY)_+ = X_+Y_+ + X_-Y_-$ gilt. Wir können X und Y von unten durch Treppenfunktionen der Form $\sum a_i[X \leq x_i]$ bzw. $\sum b_j[Y \leq y_j]$ approximieren, deshalb genügt es, den Satz für Treppenfunktionen zu beweisen.

Dafür gilt

$$\mathbb{E}(XY) = \sum_{i,j} a_i b_j \mathbb{P}(X \leq x_i, Y \leq y_j) = \sum_{i,j} a_i b_j \mathbb{P}(X \leq x_i)\mathbb{P}(Y \leq y_j) = \mathbb{E}(X)\mathbb{E}(Y).$$

Der Erwartungswert kann als "typischer Wert" oder "Lageparameter" der Verteilung von X angesehen werden. Wenn man zusätzlich wissen will, wie gut dieser Wert die Zufallsvariable repräsentiert, kann man etwa das Mittel der (absoluten) Abstände betrachten. Da der Absolutbetrag nicht so leicht zu behandeln ist, wählen wir zum Entfernen des Vorzeichens lieber das Quadrat:

Definition 5.4

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

heißt die Varianz von X. Die Quadratwurzel der Varianz heißt Streuung oder Standardabweichung.

Wir stellen fest:

Satz 5.16: Eigenschaften der Varianz

1. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$,
2. $\mathbb{V}(X) = \mathbb{E}((X - a)^2) - (\mathbb{E}(X) - a)^2$ (Steiner'scher Verschiebungssatz),
3. $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$, wenn X und Y unabhängig sind.

Die erste Behauptung folgt direkt aus der Linearität des Integrals, weil

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

und daher

$$aX + b - \mathbb{E}(aX + b) = a(X - \mathbb{E}(X)).$$

Diese Beziehung erlaubt es uns, für die beiden anderen Behauptungen o.B.d.A. $\mathbb{E}(X) = 0$ anzunehmen.

Dann ist

$$\mathbb{E}((X - a)^2) = \mathbb{E}(X^2 - 2aX + a^2) = \mathbb{E}(X^2) + a^2 = \mathbb{V}(X) + a^2.$$

und

$$\mathbb{V}(X + Y) = \mathbb{E}((X + Y)^2) = \mathbb{E}(X^2) + 2\mathbb{E}(XY) + E(Y^2) =$$
$$\mathbb{V}(X) + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{V}(X) = \mathbb{V}(X) + \mathbb{V}(Y).$$

Der Steinersche Verschiebungssatz für $a = 0$ gibt uns die Formel

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2,$$

die oft günstig für die Berechnung der Varianz ist.

Die Summenformel für die Varianzen lässt sich im allgemeinen Fall so schreiben:

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbf{Cov}(X,Y),$$

wobei

$$\mathbf{Cov}(X,Y) = \mathbb{E}((X - E(X))(Y - E(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

die *Kovarianz* von X und Y genannt wird. Wenn die Kovarianz gleich 0 ist, heißen X und Y *unkorreliert*.

Für eine Folge X_n von reellen Zufallsvariablen definieren wir die Partialsummen

$$S_n = \sum_{i=0}^{n} X_i \quad (S_0 = 0).$$

Wenn die Zufallsvariablen unkorreliert sind und endliche Varianzen haben, dann gilt

$$\mathbb{V}(S_n) = \sum_{i=1}^{n} \mathbb{V}(X_i).$$

Haben alle X_i dieselbe Varianz σ^2, dann ist

$$\mathbb{V}(S_n) = n\sigma^2.$$

Für $\bar{X}_n = S_n/n$, das Stichprobenmittel, ergibt sich damit

$$\mathbb{V}(\bar{X}_n) = \sigma^2/n.$$

Diese Feststellung führt uns zu einigen zentralen Sätzen der Wahrscheinlichkeitstheorie, den Gesetzen der großen Zahlen. Dazu benötigen wir zuerst einige Sätze zur Abschätzung von Wahrscheinlichkeiten mit Hilfe der Varianz bzw. Erwartungswerte.

5.6 Die Ungleichungen von Markov und Chebychev

Der erste Satz gilt nicht nur für Wahrscheinlichkeitsmaße:

Satz 5.17: Ungleichung von Markov

Falls $f \geq 0$, dann gilt für jedes $c > 0$

$$\mu(\{x : f(x) \geq c\}) \leq \frac{1}{c} \int f d\mu.$$

Zum Beweis setzen wir

$$\int f d\mu = \int_{\{x:f(x)\geq c\}} f d\mu + \int_{\{x:f(x)<c\}} f d\mu \geq$$

$$\int_{\{x:f(x)\geq c\}} f d\mu \geq \int_{\{x:f(x)\geq c\}} c d\mu = c\mu(\{x : f(x) \geq c\}).$$

Satz 5.18: Ungleichung von Chebychev

Falls die Varianz von X endlich ist, dann gilt für $c > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{\mathbb{V}(X)}{c^2}.$$

Zum Beweis wenden wir die Ungleichung von Markov auf die Zufallsvariable $Y = (X - E(X))^2$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) = \mathbb{P}(Y > c^2) \leq \frac{\mathbb{E}(Y)}{c^2} = \frac{\mathbb{V}(X)}{c^2}.$$

Diese Ungleichung gibt uns

Satz 5.19: schwaches Gesetz der großen Zahlen

Sind die Zufallsvariablen $(X_n, n \in \mathbb{N})$ unabhängig und identisch verteilt mit endlicher Varianz, dann gilt

$$\bar{X}_n \to \mathbb{E}(X_1) \text{ in Wahrscheinlichkeit.}$$

Beweis: Die Ungleichung von Chebychev impliziert, dass

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| > \epsilon) \leq \frac{\mathbb{V}(X_1)}{n\epsilon^2} \to 0.$$

Dieses Gesetz lässt einigen Spielraum für Verbesserungen: Wenn man auf der linken Seite $\mathbb{E}(X_1)$ durch das Mittel der Erwartungswerte ersetzt, braucht man nicht mehr anzunehmen, dass alle Variablen identisch verteilt sind, die Unabhängigkeit lässt sich durch die Unkorreliertheit ersetzen, und auch die Gleichheit der Varianzen lässt sich durch die Forderung ersetzen, dass $\sum_{i\leq n} \mathbb{V}(X_i)/n^2 \to 0$. Ja, man kann sogar auf die Unkorreliertheit verzichten: etwa wenn die Varianzen beschränkt sind und

$$\lim_{n\to\infty} \sup_i \mathrm{Cov}(X_i, X_{i+n}) \to 0.$$

5.7 Gesetze der großen Zahlen

5.7.1 Maximalungleichungen

Wenn man weiterhin die Unabhängigkeit der einzelnen Zufallsvariablen annimmt, kann man natürlich Konvergenz in einem stärkeren Sinn erreichen. Dazu ist es notwendig, nicht nur eine Summe zu kontrollieren, sondern Abschätzungen für Wahrscheinlichkeiten, in denen mehrere Summen vorkommen, zu finden, also für Wahrscheinlichkeiten der Form

$$\mathbb{P}(\max_{i\leq n} |S_i| > c).$$

Eine der einfachsten und dennoch (oder gerade deswegen) sehr allgemein verwendbaren Ungleichungen ist die folgende

Satz 5.20: Ungleichung von Lévy-Ottaviani

$X_1, \ldots, X_n$ seien unabhängig, $S_i = X_1 + \ldots + X_i$, $a, b > 0$. Dann gilt

$$\mathbb{P}(\max_{i\leq n} |S_i| \geq a + b) \leq \frac{\mathbb{P}(|S_n| \geq b)}{1 - \max_{i\leq n} \mathbb{P}(|S_n - S_i| > a)}.$$

Anmerkung: Dieser Satz verwendet keine Erwartungswerte oder Varianzen der Zufallsvariablen und benötigt daher keine Integrabilitätsbedingungen. Es gibt von diesem Satz unzählige Varianten, wobei auch (wenn rechts im Nenner entsprechende bedingte Wahrscheinlichkeiten eingesetzt werden) die Forderung der Unabhängigkeit abgeschwächt werden kann oder die Zufallsvariablen X_i Werte in allgemeineren (normierten) Räumen annehmen können.

Für den Beweis verwenden wir den immer wieder gern gebrauchten

Trick 4: Das erste Mal

Für viele Ereignisse A, die durch die Zufallsvariablen $X_1, \ldots, X_n$ definiert werden können, gibt es eine (zufällige) Zahl $T \leq n$, sodass die Kenntnis von $X_1, \ldots, X_T$ schon genügt, um festzustellen, dass A eintritt (und dass die Variablen X_i mit $i > T$ darauf keinen weiteren Einfluss haben). Dann ist es oft hilfreich, das Ereignis A in die Ereignisse $[T = t]$, $t = 1, \ldots, n$ zu zerlegen.

Für den Beweis der Ungleichung von Ottaviani setzen wir natürlich T gleich dem ersten Index t, für den $|S_t| \geq a + b$ gilt (mit $T = n + 1$, wenn für alle $t \leq n$ $|S_t| < a + b$ gilt). Wenn für $i \leq n$ $T = i$ gilt, dann ist jedenfalls $|S_i| \geq a + b$. Gilt zusätzlich $|S_n - S_i| \leq a$, dann ergibt sich $|S_n| \geq b$. Wir haben also

$$\mathbb{P}(T = i, |S_n| \geq b) \geq \mathbb{P}(T = i, |S_n - S_i| \leq a) = \mathbb{P}(T = i)\mathbb{P}(|S_n - S_i| \leq a) \geq$$

$$\mathbb{P}(T = i)(1 - \max_{i \leq n} \mathbb{P}(|S_n - S_i| > a)).$$

Über i summiert ergibt das

$$\mathbb{P}(|S_n| \geq b) \geq (1 - \max_{i \leq n} \mathbb{P}(|S_n - S_i| > a)) \sum_{i \leq n} \mathbb{P}(T = i) = (1 - \max_{i \leq n} \mathbb{P}(|S_n - S_i| > a))\mathbb{P}(\max_{i \leq n} |S_i| \geq a + b).$$

Durchdividieren vollendet den Beweis.

Etwas näher an den Ungleichungen von Markov und Chebychev ist die folgende, deren Beweis ebenfalls den "erstes Mal" Trick benützt.

Satz 5.21: Ungleichung von Kolmogorov

$X_1, \ldots, X_n$ seien unabhängige Zufallsvariable mit $\mathbb{E}(X_i) = 0$ und $\mathbb{V}(X_i) = \sigma_i^2 < \infty$. Dann gilt für $c > 0$

$$\mathbb{P}(\max_{0 \leq i \leq n} |S_i| \geq c) \leq \frac{\mathbb{V}(S_n)}{c^2}.$$

Zum Beweis setzen wir

$$Y_i = [|S_i| \geq c, |S_j| < c, j < i].$$

Y_i ist genau dann 1, wenn S_n beim Index i zum ersten Mal c erreicht (oder überschreitet). Offensichtlich gilt $\sum Y_i \leq 1$ und

$$[\max_{i \leq n} |S_i \geq c] = \sum_{i \leq n} Y_i.$$

Wir erhalten

$$\mathbb{V}(S_n) = \mathbb{E}(S_n^2) \geq \mathbb{E}(S_n^2 \sum i \leq nY_i).$$

Für die Summanden der letzten Summe gilt

$$\mathbb{E}(Y_i(S_n)^2) = \mathbb{E}(Y_i S_i^2) + 2\mathbb{E}(Y_i S_i(S_n - S_i)) + \mathbb{E}(Y_i(S_n - S_i)^2).$$

Im ersten Summanden können wir S_i^2 mit c^2 nach unten abschätzen, im zweiten ist $S_n - S - i$ von den anderen beiden Faktoren unabhängig, damit ist der Erwartungswert des Produkts gleich dem Produkt der Erwartungswerte und somit 0, und der letzte Summand ist nicht negativ. Damit haben wir insgesamt

$$\mathbb{V}(S_n) \geq c^2 \mathbb{E}(\sum_{i \leq n} Y_i) = c^2 \mathbb{P}(\max_{i \leq n} |S_i| \geq c).$$

5.7.2 Reihen mit unabhängigen Summanden

Bevor wir uns wieder der Konvergenz der Stichprobenmittel zuwenden, sehen wir uns Reihen von unabhängigen Zufallsvariablen an. Zuerst kommen ein paar neue Definitionen:

Definition 5.5

Die Folge X_n von Zufallsvariablen heißt beschränkt in Wahrscheinlichkeit, wenn es zu jedem $\epsilon > 0$ ein $M = M(\epsilon)$ gibt, sodass für alle n

$$\mathbb{P}(|X_n| > M) < \epsilon$$

gilt.

Definition 5.6

Die Folge X_n von Zufallsvariablen konvergiert gegen die Zufallsvariable X in Verteilung, wenn für jede stetige beschränkte Funktion h

$$\lim_{n\to\infty} \mathbb{E}(h(X_n)) = \mathbb{E}(h(X))$$

gilt.

Die Konvergenz in Verteilung wird uns später noch eingehender beschäftigen. An dieser Stelle wollen wir festhalten, dass die Konvergenz in Wahrscheinlichkeit die in Verteilung impliziert und diese die Beschränktheit in Wahrscheinlichkeit. Die erste Feststellung ergibt sich daraus, dass auch die Folge $h(X_n)$ in Wahrscheinlichkeit konvergiert (das wird in den Übungen gezeigt) und aus dem Satz von der dominierten Konvergenz, die zweite, indem man die Funktion $h(x) = \min((|x| - M)_+, 1)$ betrachtet.

Für die nächsten Sätze sei (X_n) eine Folge von unabhängigen Zufallsvariablen, $S_n = \sum_{i=1}^n X_i$.

Satz 5.22: Äquivalenzsatz von Lévy

Die Folge S_n ist genau dann fast sicher beschränkt, wenn sie in Wahrscheinlichkeit beschränkt ist.

Die Folge S_n konvergiert genau dann fast sicher, wenn sie in Verteilung konvergiert (weil die Konvergenz in Wahrscheinlichkeit zwischen diesen beiden steht, sind also für Summen von unabhängigen Zufallsvariablen die drei Konvergenzen fast sicher, im Maß und in Verteilung äquivalent).

Wir zeigen an dieser Stelle, dass die Beschränktheit bzw. Konvergenz in Wahrscheinlichkeit die entsprechende Eigenschaft mit Wahrscheinlichkeit eins impliziert. Dass auch die Konvergenz in Verteilung zu den anderen beiden Konvergenzarten äquivalent ist, wird üblicherweise mit Hilfe der Theorie der charakteristischen Funktionen gezeigt, die wir an dieser Stelle nicht in allen Details ausbreiten wollen. Wir werden diesen Teil in den Beweis des nächsten Satzes (des Dreireihensatzes von Kolmogorov) integrieren. Dieses Vorgehen verschleiert die Tatsache, dass der Äquivalenzsatz von Lévy unter allgemeineren Bedingungen (X_n darf Werte in einem beliebigen Banachraum annehmen) gilt als der Dreireihensatz (geht noch in Hilberträumen).

Wir führen den Beweis, dass die Konvergenz im Maß die Konvergenz fast sicher impliziert, im Detail, der Beweis für die Beschränktheit verläuft analog und ist sogar etwas einfacher.

Da S_n im Maß konvergiert, gibt es zu allen $\epsilon > 0$ und $\delta > 0$ ein n_0, sodass für $n, m \geq n_0$

$$\mathbb{P}(|S_n - S_m| \geq \epsilon) < \delta$$

gilt. Für $\delta < 1/2$ folgt daraus mit Hilfe der Ungleichung von Lévy-Ottaviani für jedes $n > n_0$

$$\mathbb{P}(\max_{n_0 \leq i \leq n} |S_i - S_{n_0}| \geq 2\epsilon) \leq \frac{\delta}{1-\delta} \leq 2\delta.$$

Für $n \to \infty$ ergibt sich daraus

$$\mathbb{P}(A_{n_0}) \leq 2\delta$$

mit

$$A_{n_0} = [\sup_{n \geq n_0} |S_n - S_{n_0}| \geq \epsilon].$$

Weil $\delta \in]0, 1/2[$ beliebig gewählt werden kann, ergibt sich

$$\mathbb{P}(\bigcap A_n) = \lim_n \mathbb{P}(A_n) = 0;$$

$\bigcap A_n$ enthält das Ereignis, dass sich über jeder Grenze Indizes m und n finden lassen mit $|S_n - S_m| \geq 4\epsilon$. Dieses Ereignis hat also für jedes $\epsilon > 0$ Wahrscheinlichkeit 0, und dasselbe gilt für die Vereinigung dieser Ereignisse über eine Folge von Werten für ϵ, die gegen 0 konvergiert; insgesamt bildet S_n mit Wahrscheinlichkeit 1 eine Cauchyfolge und konvergiert daher.

Der nächste Satz gibt ein Kriterium für die Konvergenz bzw. Beschränktheit der Folge S_n:

Satz 5.23: Dreireihensatz von Kolmogorov

(X_n) sei eine Folge von unabhängigen Zufallsvariablen. Die Reihe

$$\sum_n X_n$$

(d.h, die Folge ihrer Partialsummen S_n) konvergiert genau dann, wenn die folgenden drei Reihen konvergieren (für irgendein $\epsilon > 0$ und damit für alle):

$$P_n(\epsilon) = \sum_{i \leq n} \mathbb{P}(|X_i| > \epsilon)$$

$$E_n(\epsilon) = \sum_{i \leq n} \mathbb{E}(X_i[|X_i| \leq \epsilon])$$

$$V_n(\epsilon) = \sum_{i \leq n} \mathbb{V}(X_i[|X_i| \leq \epsilon]).$$

Die Folge S_n ist genau dann beschränkt, wenn für ein $\epsilon > 0$ die drei Reihen $P_n(\epsilon)$, $E_n(\epsilon)$ und $V_n(\epsilon)$ beschränkt bleiben.

Es ist leicht zu zeigen, dass die Dreireihenbedingungen in beiden Fällen hinreichend ist. Aus der Konvergenz der Reihe P_n und dem Borel-Cantelli Lemma folgt, dass mit Wahrscheinlichkeit 1 nur endlich viele der Ereignisse $[|X_n| > \epsilon]$ eintreten. Deshalb stimmen die beiden Folgen (X_n) und $Y_n = X_n[|X_n| \leq \epsilon]$ ab einem Index überein, und daher ist die Konvergenz bzw. Beschränktheit von $\sum X_n$ äquivalent zu der von $\sum Y_n$. Die Ungleichung von Chebychev impliziert dann die Konvergenz bzw. Beschränktheit von $\sum(Y_n - \mathbb{E}(Y_n))$ in Wahrscheinlichkeit, und weil auch $E_n(\epsilon)$ konvergent bzw. beschränkt ist, gilt dasselbe für $\sum Y_n$.

Die Notwendigkeit der Dreireihenbedingung benötigt ein paar zusätzliche Tricks. Der erste Schritt ist einfach: wenn die erste Reihe $P_n(\epsilon)$ für ein $\epsilon > 0$ divergiert, ist nach dem zweiten Lemma von Borel-Cantelli fast sicher für unendlich viele n $|X_n| > \epsilon$, und deshalb kann die Reihe $\sum X_n$ nicht konvergieren. Wenn $\sum X_n$ konvergiert, muss also die erste Reihe für jedes $\epsilon > 0$ konvergieren. Analog ergibt sich, dass die Folge S_n mit Wahrscheinlichkeit 1 unbeschränkt ist, wenn $P_n(\epsilon)$ für jedes $\epsilon > 0$ divergiert. Für die Beschränktheit von S_n ist also notwendig, dass $P_n(\epsilon)$ für ein $\epsilon > 0$ konvergiert. Dann müssen auch die abgeschnittenen Zufallsvariablen Y_n eine konvergente bzw. beschränkte Reihe bilden, wie wir schon beim Beweis der Suffizienz festgestellt haben.

Der nächste Schritt besteht darin, die Notwendigkeit der Konvergenz der Reihe $V_n(\epsilon)$ zu zeigen. Dazu ist es notwendig, die Erwartungswerte von Y_n zu eliminieren, was meist dadurch gelöst wird, dass man eine unabhängige Kopie (Y_n^*) der Folge (Y_n) verwendet und die Differenzen $Y_n - Y_n^*$ betrachtet, die Erwartungswert 0 und die doppelte Varianz von Y_n haben. Dass das möglich ist, folgt aus der Theorie der Produkträume, die noch vor uns liegt. Wir gehen daher einen anderen Weg. Unser Ausgangspunkt ist die schwächste der Bedingungen, die Beschränktheit im Maß. Wir führen zur Vereinfachung der Notation die Summen

$$T_n = \sum_{i \leq n} Y_i$$

und

$$Z_n = T_n - \mathbb{E}(T_n) = \sum_{i \leq n} (Y_n - \mathbb{E}(Y_n))$$

ein. Außerdem sei

$$U_n = Y_n - \mathbb{E}(Y_n).$$

Aus der Theorie der charakteristischen Funktionen benötigen wir nur ein paar grundlegende Dinge (später werden wir uns im Zusammenhang mit einer eingehenden Diskussion der Konvergenz in Verteilung sehr viel genauer damit beschäftigen):

Definition 5.7

Die charakteristische Funktion ϕ_X der reellen Zufallsvariable X ist gegeben durch

$$\phi_X(t) = \mathbb{E}(e^{iXt}) = \mathbb{E}(\cos(Xt)) + i\mathbb{E}(\sin(Xt)), t \in \mathbb{R}.$$

Da Sinus und Kosinus beschränkt und stetig sind, ist klar, dass aus der Konvergenz in Verteilung einer Folge von Zufallsvariablen die punktweise Konvergenz ihrer charakteristischen Funktionen folgt. Dies (und die Tatsache, dass auch die Umkehrung gilt, was wir aber jetzt nicht brauchen) macht die charakteristische Funktion zu *dem* Werkzeug für das Studium der Verteilungskonvergenz.

Wir fixieren nun ein $\delta > 0$ und wählen nun M so, dass $\mathbb{P}(|X| > M) < \delta$. Für $|t| < 1/M$ erhalten wir

$$\mathbb{E}(\cos(Xt)) = \mathbb{E}(\cos(Xt)[|X| \leq 1/M]) + \mathbb{E}(\cos(Xt)[|X| > 1/M]) \geq \cos(1)(1-\delta) - \delta \geq (1-3\delta)/2.$$

Für $\delta < 1/3$ ergibt sich daraus für $|t| \leq 1/M$

$$|\phi_X(t)| \geq (1-3\delta)/2.$$

Wenn wir diese Überlegung auf die Folge T_n anwenden, die im Maß beschränkt ist, stellen wir fest, dass für hinreichend kleines t die Folge

$$\phi_{T_n}(t)$$

von 0 weg beschränkt ist. Es ist aber

$$\phi_{T_n}(t) = \mathbb{E}(e^{itT_n}) = \mathbb{E}(e^{it(Z_n + \mathbb{E}(T_n))}) = e^{it\mathbb{E}(T_n)}\phi_{Z_n}(t).$$

Somit gilt

$$|\phi_{T_n}(t)| = |\phi_{Z_n}(t)| = \prod_{j=1}^{n} |\phi_{U_j}(t)|.$$

Wir wissen, dass U_j beschränkt ist — aus $|Y_j| \leq \epsilon$ folgt $|U_j| \leq 2\epsilon$. Für hinreichend kleines t ($|t| \leq 1/(2\epsilon)$) ergibt sich aus den Ungleichungen

$$\cos(x) \leq 1 - \frac{x^2}{2} + \frac{x^4}{24}$$

und

$$|\sin(x) - x| \leq \frac{|x|^3}{6},$$

die aus den Taylorentwicklungen folgen, zusammen mit $\mathbb{E}(U_j) = 0$ die (grobe) Abschätzung

$$|\phi_{U_j}(t)| \leq 1 - \frac{1}{10}t^2\mathbb{E}(U_j^2) = 1 - \frac{1}{10}t^2\mathbb{V}(Y_j) \leq e^{-t^2\mathbb{V}(Y_j)/10}.$$

Da

$$\prod_{j=1}^{n} |\phi_{U_j}(t)|$$

von 0 weg beschränkt bleibt, muss

$$V_n(\epsilon) = \sum_{j=1}^{n} \mathbb{V}(Y_j)$$

beschränkt bleiben, also konvergieren. Die Folge U_n erfüllt also alle drei Reihenbedingungen, daher konvergiert die Reihe $\sum U_n$ und ist somit auch beschränkt. Da auch die Reihe $\sum Y_n$ beschränkt sein soll, gilt das auch für die Differenz

$$\sum Y_n - \sum U_n = \sum \mathbb{E}(Y_n).$$

Das beweist die Notwendigkeit der Dreireihenbedingung für die Beschränktheit. Ebenso folgt aus der fast sicheren Konvergenz von $\sum Y_n$ die Konvergenz von $\sum \mathbb{E}(Y_n)$, also die Notwendigkeit der

Dreireihenbedingung für die Konvergenz fast sicher/in Wahrscheinlichkeit. Es bleibt uns also der Fall, dass wir nur wissen, dass S_n in Verteilung konvergiert. Weil die Verteilungskonvergenz die Beschränktheit in Wahrscheinlichkeit nach sich zieht, können wir aus dem gerade bewiesenen folgern, dass für ein $\epsilon > 0$ die drei Reihen beschränkt bleiben. Wir wissen außerdem, dass die Folge der charakteristischen Funktionen ϕ_{S_n} punktweise konvergiert, und die erste Reihenbedingung impliziert dasselbe für die charakteristischen Funktionen ϕ_{T_n}. Da die zentrierten Variablen U_n die Dreireihenbedingung erfüllen, konvergiert die Folge Z_n mit Wahrscheinlichkeit 1 und daher auch in Verteilung, also haben wir auch punktweise Konvergenz der Folge ϕ_{Z_n}. Für hinreichend kleines t ist $\phi_{Z_n}(t)$ von 0 weg beschränkt, und daher muss

$$e^{itE_n(\epsilon)} = \phi_{T_n}(t)/\phi_{Z_n}(t)$$

konvergieren. Wir wissen schon, dass $E_n(\epsilon)$ beschränkt ist; wenn wir t so klein wählen, dass für alle n die Ungleichung $|tE_n| < \pi/2$ gilt, dann können wir folgern, dass auch E_n konvergiert, und der Beweis des Dreireihensatzes ist komplett.

5.7.3 Das starke Gesetz der großen Zahlen

Nun wenden wir uns wieder den Mittelwerten zu:

Satz 5.24: Starkes Gesetz der großen Zahlen von Kolmogorov

X_n seien unabhängige Zufallsvariable mit $\mathbb{E}(X_n) = 0$ und

$$\sum_{n\in\mathbb{N}} \frac{\mathbb{V}(X_n)}{n^2} < \infty.$$

Dann gilt

$$\lim_{n\to\infty} \bar{X}_n = 0$$

fast sicher.

Wir definieren

$$A_k = [\max_{j\leq 2^k} |S_j| \geq \epsilon 2^{k-1}].$$

Die Ungleichung von Kolmogorov liefert

$$\mathbb{P}(A_k) \leq \frac{\mathbb{V}(S_{2^k})}{(\epsilon 2^{k-1})^2},$$

und

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) \leq \sum_k \frac{\sum_{n\leq 2^k} \mathbb{V}(X_n)}{(\epsilon 2^{k-1})^2} =$$

$$\frac{4}{\epsilon^2} \sum_n \mathbb{V}(X_n) \sum k : 2^k \geq n \frac{1}{2^{2k}} \leq \frac{16}{3\epsilon^2} \sum_n \frac{\mathbb{V}(X_n)}{n^2} < \infty.$$

Borel-Cantelli sagt uns jetzt, dass mit Wahrscheinlichkeit 1 für hinreichend großes k

$$\max_{n\leq 2^k} |S_n| \leq \epsilon 2^{k-1}$$

gilt. Für großes n können wir ein k finden, das $2^{k-1} \leq n \leq 2^k$ erfüllt, und

$$|S_n| \leq \max_{n\leq 2^k} |S_n| \leq \epsilon 2^{k-1} \leq \epsilon n$$

und $\bar{X}_n = S_n/n$ geht tatsächlich gegen 0.

Für identisch verteilte Zufallsvariable können wir auf die Varianzen verzichten:

Satz 5.25: starkes Gesetz der großen Zahlen von Kolmogorov 2

(X_n) sei eine Folge von unabhängig identisch verteilten Zufallsvariablen mit endlichem Erwartungswert. Dann gilt fast sicher

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k\leq n} X_k = \mathbb{E}(X_1).$$

Wir setzen $Y_n = X_n[|X_n| \leq n]$. Wegen

$$\mathbb{E}(|X_1|) = \int_0^\infty \mathbb{P}(|X_1| > x)dx \geq \sum_{n=1}^\infty \mathbb{P}(|X_n| > n) = \sum_{n=1}^\infty \mathbb{P}(Y_n \neq X_n)$$

folgt aus dem Borel-Cantelli Lemma, dass mit Wahrscheinlichkeit 1 für fast alle n $X_n = Y_n$ gilt, also konvergieren die Mittelwerte der X_n genau dann, wenn die Mittelwerte der Y_n konvergieren, und die beiden Grenzwerte stimmen überein. Da $\mathbb{E}(Y_n) \to \mathbb{E}(X_1)$ gilt, gilt auch

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}(Y_i) = \mathbb{E}(X_1).$$

Wir müssen also nur noch zeigen, dass die Y_n die Voraussetzungen des ersten Gesetzes der großen Zahlen erfüllen. Es ist

$$\sum_{n\in\mathbb{N}} \frac{\mathbb{V}(Y_n)}{n^2} \leq \sum_{n\in\mathbb{N}} \frac{\mathbb{E}(Y_n^2)}{n^2} = \sum_{n\in\mathbb{N}} \frac{\mathbb{E}(X_1^2[|X_1| \leq n])}{n^2} = \mathbb{E}(X_1^2 \sum_{n\geq|X_1|} \frac{1}{n^2}) \leq$$

$$1 + \mathbb{E}(X_1^2 \sum_{n\geq|X_1|+1} \frac{1}{n^2}) \leq 1 + \mathbb{E}(X_1^2 \frac{1}{|X_1|} = 1 + \mathbb{E}(|X_1|) < \infty.$$

Damit ist das erste Gesetz der großen Zahlen auf die Y_n anwendbar, und unser zweiter Satz bewiesen.

Mit etwas mehr Arbeit lassen sich die Voraussetzungen abschwächen, etwa bleibt der letzte Satz gültig, wenn man nur annimmt, dass die Zufallsvariablen X_n paarweise unabhängig sind. Dazu gibt es auch eine Umkehrung:

Satz 5.26

Die Folge (X_n) sei (paarweise) unabhängig und identisch verteilt. Dann konvergiert $\bar{X}_n = S_n/n$ genau dann mit Wahrscheinlichkeit 1 gegen einen endlichen Grenzwert, wenn X_n integrierbar ist.

Eine Richtung haben wir schon gezeigt, das ist genau das Gesetz der großen Zahlen. Für die andere Richtung nehmen wir an, dass $E(|X_n|) = \infty$ gilt. Dann ist

$$\sum_n \mathbb{P}(|X| > n) = \infty,$$

und nach dem Borel-Cantelli Lemma gibt es mit Wahrscheinlichkeit 1 unendlich viele n mit $|X_n| > n$; wenn $\bar{X}_n$ konvergiert, dann müsste aber

$$|\frac{X_n}{n}| = |\bar{X}_n - \frac{n-1}{n}\bar{X}_{n-1}| \to 0$$

gelten.

5.8 Spezielle Verteilungen

Wir sammeln jetzt Formeln für Erwartungswerte und Varianzen der bekannten Verteilungen.

5.8.1 Diskrete Verteilungen

Mit der Wahrscheinlichkeitsfunktion $p(x) = \mathbb{P}(X = x)$ gilt

$$\mathbb{E}(X) = \sum_x xp(x), \mathbb{E}(X^2) = \sum_x x^2 p(x).$$

- Die deterministische Verteilung $D(a)$ hat $\mathbb{E}(X) = a$ und $\mathbb{V}(X) = 0$.
- Für die Alternativverteilung $A(p)$ ist

  $$\mathbb{E}(X) = \mathbb{E}(X^2) = p$$

  und

  $$\mathbb{V}(X) = p - p^2 = p(1-p).$$

- Für die diskrete Gleichverteilung $D(a, b)$ ergibt sich mit den Summenformeln

  $$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}, \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

  $$\mathbb{E}(X) = \frac{1}{b-a+1} \sum_{i=a}^{b} i = \frac{1}{b-a+1} \left(\frac{b(b+1)}{2} - \frac{a(a-1)}{2} \right) = \frac{a+b}{2}.$$

  Dies ist ein allgemeines Phänomen: wenn eine symmetrische Verteilung endlichen Erwartungswert hat, dann stimmt dieser mit dem Symmetriezentrum überein.

  Für die Varianz ergibt sich (das wird in den Übungen berechnet)

  $$\mathbb{V}(X) = \frac{(b-a+1)^2 - 1}{12}.$$

- Für die Binomialverteilung $B(n, p)$ gilt

  $$\mathbb{E}(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

  $$\sum_{x=1}^{n} n \binom{n-1}{x-1} p^x (1-p)^{n-x} = \sum_{k=0}^{n} n \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} = np.$$

  $$\mathbb{E}(X^2) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} =$$

  $$\sum_{x=1}^{n} nx \binom{n-1}{x-1} p^x (1-p)^{n-x} = \sum_{k=0}^{n} n(x+1) \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} = np(1 + (n-1)p).$$

  $$\mathbb{V}(X) = np(1-p).$$

- Für die Hypergeometrische Verteilung $H(N, A, n)$ ergibt eine ähnliche Rechnung

  $$\mathbb{E}(X) = \frac{nA}{N}, \mathbb{V}(X) = n \frac{A}{N} \frac{N-A}{N} \frac{N-n}{N-1}.$$

- Bei der Poissonverteilung $P(\lambda)$ funktioniert dieselbe Methode, und es ergibt sich

  $$\mathbb{E}(X) = \mathbb{V}(X) = \lambda.$$

- Die geometrische Verteilung: für $X \sim G^*(p)$ ist

$$\mathbb{E}(X) = \frac{1}{p}, \mathbb{V}(X) = \frac{1-p}{p^2},$$

für $X \sim G(p)$

$$\mathbb{E}(X) = \frac{1-p}{p}, \mathbb{V}(X) = \frac{1-p}{p^2}.$$

- Negative Binomialverteilung: für $X \sim NB^*(n,p)$ ist

$$\mathbb{E}(X) = \frac{n}{p}, \mathbb{V}(X) = \frac{n(1-p)}{p^2},$$

für $X \sim NB(\alpha, p)$

$$\mathbb{E}(X) = \frac{\alpha(1-p)}{p}, \mathbb{V}(X) = \frac{\alpha(1-p)}{p^2}.$$

5.8.2 Stetige Verteilungen

Hier haben wir mit der Dichte f

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx, \mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx.$$

- $X \sim U(a,b)$:

$$\mathbb{E}(X) = \frac{a+b}{2}, \mathbb{V}(X) = \frac{(b-a)^2}{12}.$$

- $X \sim E(\lambda)$:

$$\mathbb{E}(X) = \frac{1}{\lambda}, \mathbb{V}(X) = \frac{1}{\lambda^2}.$$

- $X \sim \Gamma(\alpha, \lambda)$:

$$\mathbb{E}(X) = \frac{\alpha}{\lambda}, \mathbb{V}(X) = \frac{\alpha}{\lambda^2}.$$

item $X \sim B_1(\alpha, \beta)$:

$$\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}, \mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

- $X \sim B_2(\alpha, \beta)$:

$$\mathbb{E}(X) = \frac{\alpha}{\beta-1}(\beta > 1),$$

$$\mathbb{V}(X) = \frac{\alpha(\alpha+\beta-1)}{(\beta-1)^2(\beta-2)}(\beta > 2).$$

Für $\beta \leq 1$ ist der Erwartungswert ∞ und die Varianz existiert nicht, für $1 < \beta \leq 2$ ist die Varianz unendlich.

- $X \sim N(\mu, \sigma^2)$:

$$\mathbb{E}(X) = \mu, \mathbb{V}(X) = \sigma^2.$$

- Die Cauchyverteilung hat keinen Erwartungswert, weil $\mathbb{E}(X_+) = \mathbb{E}(X_-) = \infty$.

5.9 Ergänzungen

5.9.1 Alternativer Beweis des starken Gesetzes der großen Zahlen

Die Reihenbedingung für die Varianzen impliziert zusammen mit der Ungleichung von Chebychev, dass

$$\sum_n \frac{X_n}{n}$$

in Wahrscheinlichkeit konvergiert und damit nach dem Äquivalenzsatz von Lévy auch fast sicher. Von dieser Feststellung zur Aussage unseres Satzes verhilft uns

Satz 5.27: Kronecker-Lemma

a_n sei eine Folge von positiven Zahlen mit $a_n \uparrow \infty$. Wenn $\sum_n \frac{x_n}{a_n}$ konvergiert, dann gilt

$$\lim_{n\to\infty} \frac{\sum_{k\le n} x_k}{a_n} = 0.$$

Zum Beweis dieses Lemmas setzen wir

$$s_n = \sum_{k=n}^{\infty} \frac{x_k}{a_k}.$$

Laut Voraussetzung konvergieren diese Reihenreste gegen 0, also gibt es zu jedem $\epsilon > 0$ ein n_0, sodass für $n \ge n_0$ $|s_n| < \epsilon$ gilt. Es gilt

$$x_n = a_n(s_n - s_{n+1}).$$

Damit erhalten wir

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} a_k(s_k - s_{k+1}) = a_1 s_1 - a_n s_{n+1} + \sum_{k=2}^{n} s_k(a_k - a_{k-1}).$$

Wenn wir das durch a_n dividieren, gehen die ersten beiden Terme auf der rechten Seite gegen 0, und für die Summe erhalten wir

$$|\sum_{k=2}^{n} s_k(a_k = a_{k-1})| \le \sum_{k=2}^{n} |s_k|(a_k - a_{k-1}) \le$$

$$\sum_{k=2}^{n_0-1} |s_k|(a_k - a_{k-1}) + \epsilon(a_n - a_{n_0}).$$

Der erste Term bleibt beschränkt, deshalb geht die obere Schranke für $n \to \infty$ gegen ϵ, wenn man sie durch a_n dividiert. Da $\epsilon > 0$ beliebig gewählt werden kann, ergibt sich die gewünschte Konvergenz der Mittelwerte gegen 0. Damit ist das Kronecker-Lemma bewiesen.

Seine Anwendung mit $a_n = n$ gibt uns, dass mit Wahrscheinlichkeit 1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = 0$$

ist, und der Beweis des Gesetzes der großen Zahlen ist ebenfalls abgeschlossen.

5.9.2 Das Stieltjes-Integral

Das Riemann-Stieltjes-Integral der Funktion $f : [a, b] \to \mathbb{R}$ (des "Integranden") bezüglich der Funktion $F : [a, b] \to \mathbb{R}$ (des "Integrators") ist der Grenzwert der Stieltjes-Summen

$$\sum_{i=1}^{n} f(x_i)(F(t_i) - F(t_{i-1}))$$

mit $a = t_0 < t_1 < \ldots < t_n = b$, $t_{i-1} \leq x_i \leq t_i$, $i = 1, \ldots, n$ für $\max_{1 \leq i \leq n}(t_i - t_{i-1}) \to 0$, wenn er existiert.

Wir nehmen an, dass F monoton nichtfallend ist. Dann können wir die Ober- und Untersummen

$$\sum_{i=1}^{n} \sup f([t_{i-1}, t_i])(F(t_i) - F(t_{i-1}))$$

bzw.

$$\sum_{i=1}^{n} \inf f([t_{i-1}, t_i])(F(t_i) - F(t_{i-1}))$$

betrachten. Damit kann das obere Integral als Infimum der Obersummen und das untere Integral als Supremum der Obersummen definieren, und die Existenz der Riemann-Stieltjes Integrals ist äquivaent dayu, dass oberes und unteres Integral übereinstimmen. Es ist leicht zu sehen, dass es für die Existenz des Integrals notwendig ist, dass f und F keine gemeinsamen Unstetigkeiten haben. Deshalb können wir im Inneren des Intervalls $[a, b]$ F durch die rechtsseitigen Grenzwerte ersetzen. Genauer können wir die Verteilungsfunktion

$$\tilde{F}(x) = \begin{cases} F(a) & \text{für } x < a, \\ F(x+0) & \text{für } a \leq x < b, \\ F(b) & \text{für } x \geq b \end{cases}$$

definieren. Ober- und Untersumme lassen sich als Integrale von geeigneten Treppenfunktionen bezüglich $\mu_{\tilde{F}}$ schreiben, wie wir es für das gewöhnliche Riemann-Integral getan haben. Es ergibt sich, dass die Funktion f genau dann bezüglich F Stieljes-integrierbar ist, wenn sie beschränkt und $\mu_{\tilde{F}}$-fast überall stetig ist. f ist dann $\mu_{\tilde{F}}$-fast überall messbar, und

$$\int_a^b f(x) dF(x) = \int_{[a,b]} f d\mu_{\tilde{F}}.$$

wenn Ober- und Untersummen gegen

5.10 Wiederholungsfragen

1. Wie ist das Integral definiert?
2. Welche Eigenschaften hat das Integral?
3. Wann dürfen Integration und Grenzübergang vertauscht werden?
4. Was sagt der Satz von der monotonen Konvergenz?
5. Was sagt der Satz von der dominierten Konvergenz?
6. Wie ist die gleichmäßige Integrierbarkeit definiert?
7. Wie kann die gleichmäßige Integrierbarkeit für endliche Maße charakterisiert werden?
8. Wann ist eine Funktion Riemann-integrierbar?
9. Was besagt der Transformationssatz für Integrale?
10. Was ist der Erwartungswert und welche Eigenschaften hat er?
11. Wie ist die Varianz definiert und welche Eiganschften hat sie?
12. Wie ist die Kovarianz von zwei Zufallsvariablen definiert?
13. Wann heißen zwei Zufallsvariable unkorreliert?
14. Wie ist das Integral einer komplexwertigen Funktion definiert?

15. Die Ungleichung von Markov.
16. Die Ungleichung von Chebychev.
17. Die Ungleichung von Lévy-Ottaviani.
18. Die Ungleichung von Kolmogorov.
19. Was besagt das schwache Gesetz der großen Zahlen?
20. Was besagt der Äquivalenzsatz von Lévy?
21. Was besagt der Dreireihensatz von Kolmogorov?
22. Was besagt das starke Gesetz der großen Zahlen?
23. Das starke Gesetz der großen Zahlen für identisch verteilte Zufallsvariable und seine Umkehrung.

Kapitel 6

Signierte Maße

In diesem Kapitel wollen wir uns mit sigmaadditiven Funktionen beschäftigen, die nicht notwendig nichtnegativ sind. Als Motivation betrachten wir ein Integral

$$\nu(A) = \int_A f d\mu.$$

Wir haben schon früher festgestellt, dass dadurch eine sigmaadditive Funktion definiert ist, wenn das Integral von f existiert. Dieses Integral können wir zerlegen:

$$\nu(A) = \int_A f_+ d\mu - \int_A f_- d\mu.$$

Wenn wir $\nu_1(A) = \int_A f_+ d\mu$ und $\nu_2(A) = \int_A f_- d\mu$ setzen, dann haben wir dadurch eine Darstellung

$$\nu(A) = \nu_1(A) - \nu_2(A)$$

von ν als Differenz zweier Maße gefunden. Wenn wir noch

$$P = \{x : f(x) \geq 0\}$$

und

$$N = P^C = \{x : f(x) < 0\}$$

setzen, dann ist für jede Teilmenge A von P $\nu(A) \geq 0$, und für $A \subseteq N$ ist $\nu(A) \leq 0$. Außerdem gilt $\nu_1(N) = \nu_2(P) = 0$.

Wir definieren nun

Definition 6.1

Eine Funktion μ heißt signierte Maßfunktion (oder signiertes Maß), wenn

1. sie auf einer Sigmaalgebra $\mathfrak{S}$ definiert ist,
2. ihr Wertebereich entweder $(-\infty, \infty]$ oder $[-\infty, \infty)$ ist (d.h., es können nicht sowohl $+\infty$ und $-\infty$ als Werte auftreten)
3. sie sigmaadditiv ist, also für disjunkte Mengen $A_n \in \mathfrak{S}$

   $$\mu(\bigcup A_n) = \sum \mu(A_n),$$

   wobei in der rechten Reihe entweder die Summe der positiven Glieder oder die Summe der negativen Glieder endlich sein muss (diese Bedingung und die aus Punkt 2. stellen sicher, dass bei der Addition keine unbestimmten Formen auftreten).

Die Eigenschaften, die wir oben für Integrale gefunden haben, wollen wir auf diesen allgemeineren Fall übertragen. Dazu definieren wir:

Definition 6.2

μ sei ein signiertes Maß auf $(\Omega, \mathfrak{S})$. Eine Menge $A \in \mathfrak{S}$ heißt positiv (negativ) wenn für jede messbare Teilmenge B von A $\mu(B) \geq 0$ (≤ 0) ist. Wenn für jede messbare Teilmenge B von A $\mu(B) = 0$ ist, dann heißt A eine Nullmenge.

Definition 6.3

μ sei ein signiertes Maß auf $(\Omega, \mathfrak{S})$. Das Paar (P, N) heißt eine Hahn-Zerlegung, wenn

1. $P, N \in \mathfrak{S}$,
2. $P \cap N = \emptyset$,
3. $P \cup N = \Omega$,
4. P positiv, N negativ.

Definition 6.4

Zwei Maßfunktionen μ_1 und μ_2 auf demselben Maßraum $(\Omega, \mathfrak{S})$ heißen singulär zueinander ($\mu_1 \perp \mu_2$), wenn es eine messbare Zerlegung (A, B) von Ω (d.h., $A, B \in \mathfrak{S}$, $A \cap B = \emptyset$, $A \cup B = \Omega$) gibt, sodass $\mu_1(A) = \mu_2(B) = 0$.

Definition 6.5

μ sei eine signierte Maßfunktion. Eine Darstellung

$$\mu(A) = \mu_+(A) - \mu_-(A)$$

mit zwei singulären Maßfunktionen μ_+ und μ_- heißt eine Jordan-Zerlegung von μ.

Der zentrale Satz dieses Kapitels ist der

Satz 6.1: Zerlegungssatz von Hahn

Zu jeder signierten Maßfunktion gibt es mindestens eine Hahn-Zerlegung.

Wir stellen als erstes fest, dass es zu jeder Menge A aus $\mathfrak{S}$ mit $\mu(A) < 0$ eine negative Teilmenge gibt, deren Maß nicht größer ist als das von A. Dazu gehen wir vor wie ein Inuitbildhauer, wenn er ein Walross schnitzt — wir schneiden einfach alles weg, was nicht wie eine negative Menge aussieht. Genauer gesagt, wollen wir annehmen, dass μ den Wert $-\infty$ nicht annimmt. Dann setzen wir $A_1 = A$, und für $n \geq 1$

$$k_n = \inf\{k \in \mathbb{N} : \exists B \subset A_n : \mu(B) \geq 1/k\},$$

(mit der Konvention $\inf \emptyset = \infty$) und wählen B_n so, dass $B_n \subset A$ und $\mu(B_n) \geq 1/k_n$, und schließlich $A_{n+1} = A_n \setminus B_n$ (natürlich nur falls k_n endlich ist; falls $k_n = \infty$, dann ist A_n aber schon eine positive Menge).

Wir zeigen, dass

$$C = A \setminus \bigcup B_n = \bigcap A_n$$

eine negative Menge ist. Da $\mu(C) > -\infty$ ist, gilt

$$\sum \frac{1}{k_n} \leq \sum \mu(B_n) = \mu(A) - \mu(C) < \infty,$$

also konvergiert diese Reihe, und insbesondere gilt $k_n \to \infty$. Dann folgt aus $D \subset C$, dass $D \subset A_n$ und daher $\mu(D) < \frac{1}{k_n - 1}$, und für $n \to \infty$ ergibt sich $\mu(D) \leq 0$, also ist C negativ.

Als nächstes zeigen wir, dass für eine Folge A_n von negativen Mengen die Vereinigung

$$A = \bigcup A_n$$

eine negative Menge ist und dass für jedes n $\mu(A) \leq \mu(A_n)$ gilt. Dazu setzen wir

$$B_n = A_n \setminus \bigcup_{i<n} A_i.$$

Die Mengen B_n sind disjunkt, und es gilt $A = \bigcup B_n$. Für eine Menge $C \subset A$ gilt

$$\mu(C) = \mu(\bigcup C \cap B_n) = \sum \mu(C \cap B_n),$$

und weil $C \cap B_n \subset A_n$ gilt, ist $\mu(C \cap B_n) \leq 0$ und daher $\mu(C) \leq 0$.

Nun setzen wir

$$K = \inf\{\mu(A) : A \in \mathfrak{S}, A \text{ negativ}\}.$$

Wir können eine Folge A_n von negativen Mengen finden, für die

$$\lim \mu(A_n) = K$$

gilt. Dann ist aber, wie wir gerade gesehen haben,

$$A = \bigcup A_n$$

auch eine negative Menge, und

$$\mu(A) \leq K.$$

Nach der Definition von K muss aber auch die umgekehrte Ungleichung gelten, und daher ist

$$\mu(A) = K.$$

Wir schließen daraus, dass K endlich ist, und behaupten, dass das Komplement von A eine positive Menge ist. Andernfalls könnten wir eine Teilmenge B von A^C finden, deren Maß negativ ist, und davon wieder eine negative Teilmenge D mit $\mu(D) \leq \mu(B) < 0$. Dann wäre aber $A \cup D$ eine negative Menge mit $\mu(A \cup D) < \mu(A) = K$ im Widerspruch zur Definition von K. A^C ist also positiv, und wir haben eine Hahn-Zerlegung mit $P = A^C$ und $N = A$ gefunden.

Nach der Existenz wollen wir uns jetzt kurz mit der Eindeutigkeit der Hahn-Zerlegung befassen. Es kann durchaus mehrere solche Zerlegungen geben, aber sie können sich nicht besonders stark unterscheiden: Wenn man nämlich etwa die Differenz $P_1 \setminus P_2$ von den positiven Mengen der beiden Zerlegungen betrachtet, dann kann man diese natürlich auch als Durchschnitt $P_1 \cap N_2$ schreiben. Damit ist sie sowohl eine Teilmenge der positiven Menge P_1 als auch der negativen Menge N_2, muss also eine Nullmenge sein. Zwei verschiedene Hahn-Zerlegungen unterscheiden sich also nur um Nullmengen.

Ganz eindeutig ist die andere Zerlegung:

Satz 6.2

μ sei ein signiertes Maß auf $(\Omega, \mathfrak{S})$. Dann gibt es zu μ genau eine Jordan-Zerlegung $\mu = \mu_+ - \mu_-$.

Zum Beweis der Existenz brauchen wir nur eine HAHN-Zerlegung (P, N) zu betrachten und

$$\mu_+(A) = \mu(P \cap A),$$

$$\mu_-(A) = -\mu(N \cap A)$$

zu setzen. Dadurch erhalten wir offensichtlich zwei Maßfunktionen, für die $\mu = \mu_+ - \mu_-$ gilt. Wegen $\mu_+(N) = \mu_-(P) = 0$ sind die beiden Funktionen auch singulär zueinander.

Wenn nun (μ_+, μ_-) eine Jordan-Zerlegung von μ ist, dann gibt es zwei disjunkte Mengen B und C aus $\mathfrak{S}$, sodass $\mu_+(B) = \mu_-(C) = 0$ und $B \cup C = \Omega$. Für ein beliebiges A gilt wegen $\mu_+(A \cap B) = \mu_-(A \cap C) = 0$

$$\mu_+(A) = \mu_+(A \cap B) + \mu_+(A \cap C) = \mu_+(A \cap C) = \mu_+(A \cap C) - \mu_-(A \cap C) = \mu(A \cap C),$$

und ebenso

$$\mu_-(A) = -\mu(A \cap B).$$

Daraus erhält man , dass B eine negative Menge und C eine positive Menge ist, also ist (C, B) eine Hahn-Zerlegung. Wie wir oben festgestellt haben, kann sich diese Zerlegung nur um Nullmengen von der Zerlegung (P, N) unterscheiden, und es ist

$$\mu_+(A) = \mu_+(A \cap C) = \mu_+(A \cap P) + \mu_+(A \cap (C \setminus P)) - \mu_+(A \cap (P \setminus C)) = \mu_+(A \cap P).$$

Es stimmt also jede Jordan-Zerlegung mit der speziellen Zerlegung überein, die wir zum Beweis der Existenz konstruiert haben, und damit ist auch die Eindeutigkeit gezeigt.

Für die einzelnen Komponenten der Jordan-Zerlegung haben wir spezielle Bezeichnungen:

Definition 6.6

μ_+ heißt die positive Variation von μ, μ_- die negative Variation und $|\mu| = \mu_+ + \mu_-$ die Totalvariation von μ.

Zum Abschluss wollen wir noch zeigen, dass der Sonderfall, mit dem wir unsere Untersuchungen begonnen haben, nämlich eine signierte Maßfunktion, die sich als unbestimmtes Integral einer messbaren Funktion bezüglich eines Maßes darstellen lässt, schon die ganze Geschichte ist, also jede signierte Maßfunktion so dargestellt werden kann:

Satz 6.3

Jedes signierte Maß lässt sich in der Form

$$\mu_A = \int_A f d|\mu|$$

mit $|f| = 1$ schreiben.

Das sollte (mit $f = P - N$) offensichtlich sein.

Damit können wir auch Integrale bezüglich einer signierten Maßfunktion definieren:

Definition 6.7

Das Integral einer messbaren Funktion g bezüglich des signierten Maßes μ ist

$$\int g d\mu = \int g d\mu_+ - \int g d\mu_- = \int f g d|\mu|,$$

wobei f die Funktion aus der Integraldarstellung des letzten Satzes ist.

Ein komplexwertiges Maß μ lässt sich natürlich in der Form

$$\mu = \mu_1 + i\mu_2$$

schreiben, wobei μ_1 und μ_2 signierte Maße sind (unsere grundsätzliche Annahme, dass komplexe Mengenfunktionen endlich sein sollen, impliziert natürlich, dass auch diese signierten Maße endlich sind). Jedes dieser Maße lässt sich in Positiv- und Negativteil zerlegen, was zur Darstellung

$$\mu = \mu_{1+} - \mu_{1-} + i\mu_{2+} - i\mu_{2-}$$

mit lauter (endlichen) Maßfunktionen auf der rechten Seite führt, bzw. als Integral schreiben. Diese letzte Darstellung ist noch nicht ideal, schöner wäre

$$\mu(A) = \int_A f d\mu_0$$

mit einem Maß μ_0 und einer komplexen integrierbaren Funktion f, vielleicht sogar mit $|f| = 1$. Das geht auch, braucht aber die Methoden aus dem nächsten Kapitel. Einstweilen können wir das Integral definieren:

Definition 6.8

Das Integral der komplexen Funktion $f = f_1 + if_2$ bezüglich des komplexen Maßes $\mu = \mu_1 + i\mu_2$ ist

$$\int f d\mu = \int f_1 d\mu_1 - \int f_2 d\mu_2 + i(\int f_1 d\mu_2 + \int f_2 d\mu_1).$$

Die Menge $\mathcal{M} = \mathcal{M}(\Omega, \mathfrak{S})$ aller endlichen signierten Maße auf dem Messraum $(\Omega, \mathfrak{S})$ ist offensichtlich bezüglich der Multiplikation mit Skalaren und der Addition abgeschlossen. Durch

$$\|\mu\| = |\mu|(\Omega)$$

ist auf $\mathcal{M}$ eine Norm definiert, die Variationsnorm. Wie so oft gilt auch hier

Satz 6.4

$\mathcal{M}$ ist bezüglich der Variationsnorm vollständig, also ein Banachraum.

Zum Nachweis der Vollständigkeit betrachten wir eine Cauchyfolge μ_n. Neben der Variationsnorm können wir auf $\mathcal{M}$ auch die Supremumsnorm

$$\|\mu\|_{\sup} = \sup\{|\mu(A)| : A \in \mathfrak{S}\}$$

betrachten. Wegen $\|\mu\|_{\sup} \leq \|\mu\| \leq 2\|\mu\|_{\sup}$ sind beide Normen äquivalent. μ_n ist also auch eine Cauchyfolge bezüglich der Supremumsnorm. Wie für Funktionen auf den reellen Zahlen folgt daraus einerseits die Existenz einer eindeutig bestimmten Grenzfunktion μ und andererseits, dass μ die Stetigkeitseigenschaften von μ_n erbt, etwa die Stetigkeit von oben bei $\emptyset$. Da μ auch additiv ist, ist μ ein signiertes Maß. Damit ist die Vollständigkeit von $\mathcal{M}$ gezeigt.

6.1 Ringe

Wir haben uns mit gutem Grund auf eine Sigmaalgebra als Definitionsbereich unserer signierten Maße festgelegt. Für Sigmaringe sieht die Sache noch relativ gut aus: der Beweis des Hilfssatzes, dass jede messbare Menge eine positive und eine negative Teilmwenge mit größerem bzw. kleinerem Maß enthält, gilt natürlich auch hier noch, und vom Zerlegungssatz von Hahn bleibt übrig, dass es eine positive (bzw. negative) Menge M mit endlichem Maß gibt, sodass für jede messbare Mange A die Differenz $A \setminus M$ negativ (positiv) ist. Der Zerlegungssatz von Lebesgue bleibt gültig, allerdings muss die Definition der Singularität angepasst werden (etwa: es gibt eine Menge A mit $\mu_+(A) = 0$, sodass für jede messbare Menge B $\mu_-(B \setminus A) = 0$).

Ringe sind etwas schwieriger. Es stellt sich nämlich heraus, dass nicht jede sigmaadditive Funktion auf einem Ring zu einem signierten Maß fortgesetzt werden kann. Eine notwendige Bedingung ist schnell gefunden: aus dem Zerlegungssatz von Hahn folgt für jedes $A \in \mathfrak{S}$

$$\mu(N) \leq \mu(A) \leq \mu(P),$$

und eine der beiden Schranken muss endlich sein. Deswegen ist μ entweder nach oben oder nach unten beschränkt, und diese Beschränktheit bleibt natürlich für die Einschränkung auf einen Ring erhalten. Dass diese Bedingung auch hinreichend ist, ergibt sich aus den Beziehungen

$$\mu_+(A) = \sup\{\mu(B) : \mathbb{B} \in \mathfrak{R}, B \subseteq A\}, \mu_-(A) = -\inf\{\mu(B) : \mathbb{B} \in \mathfrak{R}, B \subseteq A\},$$

die sich aus dem Approximationssatz ergeben. Deshalb gilt:

Satz 6.5

Die sigmaadditive Funktion μ auf dem Ring $\mathfrak{R}$ lässt sich genau dann zu einem signierten Maß auf der von $\mathfrak{R}$ erzeugten Sigmaalgebra fortsetzen, wenn μ nach oben oder unten beschränkt ist. Wenn μ totalsigmaendlich ist, dann ist diese Fortsetzung eindeutig bestimmt.

6.2 Signierte Maße auf $\mathbb{R}$

Wir wenden uns jetzt dem Spezialfall $(\mathbb{R}, \mathfrak{B})$ zu. Der Einfachheit halber betrachten wir endliche Maße. Ein endliches signiertes Maß μ können wir nach dem Zerlegungssatz von Jordan als Differenz zweier endlicher Maße μ_+ und μ_- darstellen. Zu diesen Maßen gibt es Verteilungsfunktionen F_+ und F_- und wir haben

$$F(x) = \mu((-\infty, x]) = F_+(x) - F_-(x),$$

also eine Differenz von zwei monotonen Funktionen.

Wir definieren

Definition 6.9

Die Totalvariation einer Funktion f auf dem Intervall $[a, b]$ ist

$$V_{[a,b]}(f) = \sup\{\sum_{i=1}^{n} |f(t_i) - f_(t_{i-1})|, n \in \mathbb{N}, a = t_0 < t_1, \ldots < t_n = b\}.$$

Ersetzt man in dieser Definition den Betrag durch den Positiv- bzw. Negativteil, dann ergibt sich die positive Variation $V^+_{[a,b]}(f)$ bzw. die negative Variation $V^-_{[a,b]}(f)$ von f.

Die Funktion f heißt von beschränkter Variation auf $[a, b]$, wenn $V_{[a,b]}(f) < \infty$.

Offensichtlich gilt für $a < b < c$

$$V_{[a,b]} + V_{[b,c]} = V_{[a,c]}.$$

Insbesondere ist $g(x) = V_{[a,x]}$ monoton nichtfallend. Man kann auch leicht feststellen, dass $g - f$ monoton nichtfallend ist. Deshalb stellen wir fest

Satz 6.6

Die Funktion $f : [a, b] \to \mathbb{R}$ ist genau dann von beschränkter Variation, wenn sie als Differenz von zwei monoton nichtfallenden Funktionen dargestellt werden kann.

Eine Richtung wurde bereits gezeigt, für die andere stellen wir fest, dass für nichtfallende Funktionen f und g

$$V_{[a,b]}(f - g) \leq f(b) - f(a) + g(b) - g(a).$$

Wir stellen also fest, dass die "Verteilungsfunktion" F von μ von beschränkter Variation ist. Sie ist auch rechtsstetig. Aus der Definition der Variationen ergibt sich

$$V_{[a,b]}F \leq |\mu|(]a, b]), V^+_{[a,b]}F \leq \mu_+(]a, b]), V^-_{[a,b]}F \leq \mu_-(]a, b]).$$

Auf der anderen Seite gibt es für jedes $\epsilon > 0$ eine Menge $A \subseteq]a, b]$ aus dem Ring, der vom Semiring der linkshalboffenen Intervalle erzeugt wird, sodass

$$|\mu|(A \triangle (P \cap]a, b])) < \epsilon$$

gilt. Daraus folgt

$$\mu_-(A) < \epsilon$$

und

$$\mu_+(A) \geq \mu_+(]a, b]) - \epsilon.$$

Insgesamt ist

$$\mu_(A) \geq \mu_+(]a,b]) - 2\epsilon.$$

Wir können A als Vereinigung von der Intervalle $]a_i, b_i], i = 1, \ldots, n$ darstellen (mit $b_i \leq a_{i+1}$). Damit haben wir

$$V^+_{[a,b]}F \geq \sum_{i=1}^{n}(F(b_i) - F(a_i))_+ \geq \sum_{i=1}^{n}(F(b_i) - F(a_i)) = \mu(A) \geq \mu_+(]a,b]) - 2\epsilon.$$

Analoges gilt für V^- und V. Es sind also die positive, die negative bzw. die Totalvariation von F mit variabler oberer Grenze die Verteilungsfunktionen der positiven, negativen bzw. Totalvariation von μ.

6.3 Wiederholungsfragen

1. Definieren Sie signierte Maßfunktionen.
2. Was ist eine positive Menge, negative Menge, Nullmenge?
3. Was ist eine Hahn-Zerlegung?
4. Was ist eine Jordan-Zerlegung?
5. Was sagt der Zerlegungssatz von Hahn?
6. Was sagt der Zerlegungssatz von Jordan?
7. Kann jedes signierte Maß als Integral bezüglich eines gewöhnlichen Maßes dargestellt werden?
8. Definieren Sie positive Variation, negative Variation und Totalvariation eines signierten Maßes.
9. Wie ist die Variationsnorm definiert? Welche Struktur hat der Raum der endlichen signierten Maße mit dieser Norm?
10. Wann lässt sich eine sigmaadditive Funktion auf einem Ring zu einem signierten Maß auf der erzeugten Sigmaalgebra fortsetzen?
11. Definieren Sie positive Variation, negative Variation und Totalvariation einer reellen Funktion.
12. Definieren Sie Funktion von beschränkter Variation.

Kapitel 7

Der Satz von Radon-Nikodym

Wir wollen uns nun der Frage zuwenden, unter welchen Bedingungen eine Maßfunktion ν sich als unbestimmtes Integral bezüglich einer anderen Maßfunktion μ darstellen lässt, also wann eine messbare Funktion f gibt, sodass

$$\nu(A) = \int_A f d\mu$$

gilt. Eine einfache notwendige Bedingung erhalten wir, wenn wir für A eine μ-Nullmenge einsetzen. Dann ist das Integral gleich 0 und damit auch $\nu(A)$. Diese Beziehung ist so wichtig, dass wir dafür einen eigenen Namen haben:

Definition 7.1

μ und ν seien zwei Maßfunktionen auf demselben Messraum $(\Omega, \mathfrak{S})$. Falls für jedes $A \in \mathfrak{S}$ mit $\mu(A) = 0$ auch $\nu(A) = 0$ ist, dann heißt ν absolutstetig bezüglich μ ($\nu \ll \mu$).

Unsere notwendige Bedingung ist also, dass ν absolutstetig bezüglich μ ist. Unter gewissen Umständen ist diese Bedingung auch hinreichend:

Satz 7.1: Radon-Nikodym

$(\Omega, \mathfrak{S}, \mu)$ sei ein sigmaendlicher Maßraum, und ν ein weiteres Maß auf $\mathfrak{S}$. ν ist genau dann als unbestimmtes Integral

$$\nu(A) = \int_A f d\mu$$

darstellbar, wenn $\nu \ll \mu$. Die Funktion f ist μ-fast überall eindeutig bestimmt und nicht-negativ, und wird die Radon-Nikodym Ableitung (Radon-Nikodym Dichte, Radon-Nikodym Derivierte) von ν nach μ,

$$f = \frac{d\nu}{d\mu}$$

bezeichnet.

Für den Beweis nehmen wir der Einfachheit halber an, dass μ endlich ist. Der allgemeine Fall lässt sich daraus erhalten, indem man Ω in abzählbar viele disjunkte Mengen E_n mit endlichem Maß zerlegt. Für die Einschränkung von ν und μ auf E_n kann man die Radon-Nikodym Dichte f_n berechnen (weil ja die Einschränkung von μ auf E_n ein endliches Maß ist), und schließlich kann man $f(x) = f_n(x)$ für $x \in E_n$ setzen.

Wir werden unser Wissen über signierte Maßfunktionen verwenden, um die Existenz einer Radon-Nikodym Dichte zu beweisen. Dazu überlegen wir uns folgendes: wenn wir annehmen, dass ν als Integral darstellbar ist, dann erhalten wir für

$$\nu(A) - c\mu(A) = \int_A (f - c) d\mu$$

eine Hahn-Zerlegung, wenn wir $P = \{x : f(x) > c\}$ setzen. Wir gehen den umgekehrten Weg, und werden die Funktion f so wählen, dass sie auf der positiven Menge einer Hahn-Zerlegung von $\nu - c\mu$ Werte $\geq c$ annimmt. Dabei haben wir natürlich ein kleines technisches Problem, weil die Hahn-Zerlegung nur bis auf Nullmengen eindeutig bestimmt ist. Wir können dieses Problem lösen, indem wir nur abzählbar viele verschiedene Werte für c betrachten. Es soll also für jedes rationale $c \geq 0$ (P_c, N_c) eine Hahn-Zerlegung von $\nu - c\mu$ sein. Es soll $P_0 = \Omega$ sein, und wir können annehmen, dass für $c < d$ die Inklusion $N_c \subseteq N_d$ gilt. Es gilt nämlich

$$\mu(N_c \setminus N_d) = 0,$$

weil einerseits

$$\nu(N_c \setminus N_d) - c\mu(N_c \setminus N_d) \leq 0$$

und andererseits

$$\nu(N_c \setminus N_d) - d\mu(N_c \setminus N_d) \geq 0$$

also

$$d\mu(N_c \setminus N_d) \leq \nu(N_c \setminus N_d) \leq c\mu(N_c \setminus N_d),$$

was nur für $\mu(N_c \setminus N_d) = 0$ möglich ist.

Wenn wir nun

$$\tilde{N}_c = \bigcup_{0 \leq d \leq c, d \in \mathbb{Q}} N_d$$

setzen, dann ist

$$\mu(\tilde{N}_c \setminus N_c) \leq \sum_{0 \leq d \leq c, d \in \mathbb{Q}} \mu(N_d \setminus N_c) = 0,$$

und $(\tilde{P}_c, \tilde{N}_c)$ mit $\tilde{P}_c = \tilde{N}_c^C$ ist auch eine Hahn-Zerlegung von $\nu - c\mu$, die $\tilde{N}_c \subseteq \tilde{N}_d$ für $c < d$ erfüllt.

Wir setzen jetzt

$$f(x) = \sup\{c \in Q : x \in P_c\}.$$

Es bleibt zu zeigen, dass f tatsächlich eine Radon-Nikodym Dichte ist. Dazu wählen wir eine beliebige Menge $A \in \mathfrak{S}$ und setzen

$$\nu(A) = \nu(A \cap [f = \infty]) + \nu(A \cap [f < \infty]).$$

Falls $\mu(A \cap [f = \infty]) = 0$, dann ist wegen der Absolutstetigkeit auch

$$\nu(A \cap [f = \infty]) = 0$$

und auch

$$\int_{A \cap [f = \infty]} f d\mu = 0.$$

Im anderen Fall ist $A \cap [f = \infty]$ eine positive Menge für jedes signierte Maß $\nu - c\mu$, also ist für jedes $c > 0$

$$\nu(A \cap [f = \infty]) \geq c\mu(A \cap [f = \infty]),$$

und daher

$$\nu(A \cap [f = \infty]) = \infty = \int_{A \cap [f = \infty]} f d\mu.$$

Für den zweiten Teil verwenden wir

$$\nu(A \cap [f < \infty]) = \sum_n \nu(A \cap [\epsilon(n-1) \leq f < \epsilon n]).$$

Wenn $f(x) \geq (n-1)\epsilon$, dann ist jedenfalls $x \in P_c$ für jedes $c < (n-1)\epsilon$, und wenn $f(x) < n\epsilon$, dann ist $x \in N_{n\epsilon}$. Daraus folgt, dass

$$(n-1)\epsilon\mu(A \cap [(n-1)\epsilon \leq f < n\epsilon]) \leq$$

$$\nu(A \cap [(n-1)\epsilon \leq f < n\epsilon]) \leq n\epsilon\mu(A \cap [(n-1)\epsilon \leq f < n\epsilon]),$$

und auch

$$(n-1)\epsilon\mu(A\cap[(n-1)\epsilon\le f<n\epsilon])\le\int_{A\cap[(n-1)\epsilon\le f<n\epsilon]}fd\mu\le$$
$$n\epsilon\mu(A\cap[(n-1)\epsilon\le f<n\epsilon]),$$

und daher

$$|\nu(A\cap[(n-1)\epsilon\le f<n\epsilon])-\int_{A\cap[(n-1)\epsilon\le f<n\epsilon]}fd\mu|\le$$
$$\epsilon\mu(A\cap[(n-1)\epsilon\le f<n\epsilon]).$$

Wir summieren über n und erhalten wegen

$$\sum_n\epsilon\mu(A\cap\{x:(n-1)\epsilon\le f(x)<n\epsilon\})\le\epsilon\mu(\Omega)$$

$$\nu(A\cap[f<\infty])-\epsilon\mu(\Omega)\le\int_{A\cap\{x:f(x)<\infty\}}fd\mu\le\nu(A\cap[f<\infty])+\epsilon\mu(\Omega).$$

Weil ϵ beliebig klein gewählt werden kann, gilt

$$\nu(A\cap[f<\infty])=\int_{A\cap[f<\infty]}fd\mu,$$

und schließlich

$$\nu(A)=\nu(A\cap[f<\infty])+\nu(A\cap[f=\infty])=$$
$$\int_{A\cap[f<\infty]}fd\mu+\int_{A\cap[f=\infty]}fd\mu=\int_A fd\mu.$$

f ist also wirklich eine Radon-Nikodym Dichte für ν.

Die Eindeutigkeit von f folgt aus der Tatsache, dass aus $\int_A fd\mu=\int_A gd\mu$ für alle A die fast überall Gleichheit von f und g folgt.

Die Konzepte der absoluten Stetigkeit und der Singularität sind in gewissem Sinn Gegensätze, wie der folgende Satz zeigt:

Satz 7.2: Zerlegungssatz von Lebesgue

$(\Omega,\mathfrak{S},\mu)$ sei ein sigmaendlicher Maßraum, ν sei ein weiteres sigmaendliches Maß auf diesem Messraum. Dann lässt sich ν in eindeutiger Weise in zwei Funktionen zerlegen:

$$\nu=\nu_1+\nu_2,$$

wobei $\nu_1\ll\mu$ und $\nu_2\perp\mu$ gilt.

Zum Beweis betrachten wir das Maß $\zeta=\mu+\nu$. Auch dieses Maß ist sigmaendlich, und offensichtlich ist sowohl μ als auch ν absolutstetig bezüglich ζ. Daher können wir nach dem Satz von Radon-Nikodym

$$f=\frac{d\mu}{d\zeta}$$

bestimmen, und es gilt

$$\frac{d\nu}{d\zeta}=1-f.$$

Wir setzen $B=\{x:f(x)=0\}$ und

$$\nu_1(A)=\nu(A\setminus B)$$

sowie

$$\nu_2(A)=\nu(A\cap B).$$

wegen

$$\nu_2(B^C)=0=\mu(B)$$

ist $\nu_2 \perp \mu$, und

$$\nu_1(A) = \int_{A\cap B^C} (1-f)d\zeta = \int_A B^C(1-f)d\zeta = \int_A B^C(\frac{1}{f}-1)d\mu.$$

ν_1 ist also als Integral bezüglich μ darstellbar und daher absolutstetig bezüglich μ.

Falls ν endlich ist, dann gibt es ein nettes Kriterium für die absolute Stetigkeit

Satz 7.3: ϵ-δ-Kriterium

$(\Omega, \mathfrak{S}, \mu)$ sei ein beliebiger Maßraum, ν ein endliches Maß auf $\mathfrak{S}$. Dann ist $\nu \ll \mu$ genau dann, wenn es zu jedem $\epsilon > 0$ ein $\delta > 0$ gibt, sodass aus $\mu(A) < \delta$ die Beziehung $\nu(A) < \epsilon$ folgt.

Es sollte offensichtlich sein, dass die ϵ-δ Bedingung die absolute Stetigkeit impliziert, auch wenn ν nicht endlich ist. Für die andere Richtung nehmen wir an, dass ν endlich ist und die ϵ-δ Bedingung nicht gilt, und zeigen, dass ν nicht absolutstetig bezüglich μ ist. Es soll also ein $\epsilon > 0$ geben, sodass zu jedem $\delta > 0$ eine Menge $A_\delta \in \mathfrak{S}$ existiert mit $\mu(A_\delta) < \delta$ und $\nu(A_\delta) > \epsilon$. Wir setzen

$$A = \limsup_n A_{1/2^n}.$$

Aus dem Borel-Cantelli Lemma folgt

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

Auf der anderen Seite folgt aus der Endlichkeit von ν

$$\nu(A) = \nu(\limsup_n A_{1/2^n}) \geq \limsup_n \nu(A_{1/2^n}) \geq \epsilon,$$

und das war's.

7.1 In den reellen Zahlen

There is a house in New Orleans
they call the rising sun

traditional

Black girl, black girl, don't lie to me,
tell me where did you sleep last night?
In the pines, in the pines, where the sun
will never shine,
I would shiver the whole night through.

traditional

Der zentrale Satz dieses Kapitels ist der folgende (un autre théorème de Lebesgue)

Satz 7.4

Eine monoton nichtfallende Funktion F ist λ-fast überall differenzierbar, und die Ableitung ist die Radon-Nikodym Dichte des (λ-) absolutstetigen Anteils von μ_F.

Für andere Lebesgue-Stieltjes Maße kann man feststellen.

Satz 7.5

F und G seien zwei Verteilungsfunktionen Dann existiert

$$\lim_{h\to 0} \frac{G(x+h)-G(x-h)}{F(x+h)-F(x-h)}$$

mit den Konventionen $0/0 = 0$ und $(> 0)/0 = \infty$ μ_F-fast überall und ist eine Radon-Nikodym Dichte des μ_F-absolutstetigen Anteils von μ_G.

Es sollte klar sein, dass in beiden Sätzen die Funktion im Zähler durch eine Funktion von beschränkter Variation ersetzt werden kann, die als Verteilungsfunktion eines signierten Maßes verstanden wird. Geht das (beim letzten Satz) auch im Nenner?

Die Ableitung muss nicht symmetrisch gewählt werden, wenn G stetig ist, kann die Ableitung analog zur gewöhnlichen Ableitung definiert werden, andernfalls muss nur sichergestellt sein, dass der Nenner gegen $\mu_G(\{x\})$ konvergiert.

Für den Beweis dieses Satzes verwenden wir den schönen

Satz 7.6: Riesz' Satz von der aufgehenden Sonne

$f : [a, b] \to \mathbb{R}$ sei eine Funktion, die nur Unstetigkeiten erster Art besitzt, also für die in jedem Punkt die Grenzwerte $f(x - 0)$ und $f(x + 0)$ existiert. Wir setzen $\tilde{f}(x) = \max(f(x - 0), f(x), f(x + 0))$ und nennen den Punkt x von rechts (links) unsichtbar, wenn es ein $y > x$ $(y < x)$ gibt mit $f(y) > \tilde{f}(x)$. Dann besteht die Menge der von rechts (links) unsichtbaren Punkte aus höchsten abzählbar vielen disjunkten offenen Intervallen $]a_n, b_n[$, und es gilt $f(a_n + 0) \leq \tilde{f}(b_n)$ $(f(b_n - 0) \leq \tilde{f}(a_n))$.

Wenn $\tilde{f}(x) < f(y)$, dann gibt es eine ganze Umgebung U von x, sodass für $z \in U$ $f(z) < f(y)$ gilt. Es sind also alle Punkte in einer Umgebung von x unsichtbar. Die Menge der unsichtbaren Punkte ist also offen, und damit erhalten wir unmittelbar die Darstellung als abzählbare Vereinigung von disjunkten offenen Intervallen.

Wir betrachten jetzt das Teilintervall $]a_n, b_n[$. Wenn $f(a_n + 0) > \tilde{f}(b_n)$ ist, dann können wir $x \in]a_n, b_n[$ finden mit $f(x) > \tilde{f}(b_n)$. Wir setzen

$$y = \sup\{z \in]a_n, b_n[: f(z) \geq f(x)\}.$$

Das Supremum ist entweder ein Maximum und $f(y) \geq f(x)$, oder es ist der Grenzwert einer wachsenden Folge und $f(y - 0) \geq f(x)$. Wegen der letzten Ungleichung gilt $y < b_n$. y liegt also in $]a_n, b_n[$ und ist daher von rechts unsichtbar. Es gibt also ein $z > y$ mit $f(z) > f(y)$. Die Definition von y hat zur Folge, dass $z > b_n$ sein muss, dann wäre aber b_n auch von rechts unsichtbar, und wir haben einen Widerspruch.

In der Folge seien μ und ν Maße mit Verteilungsfunktionen F bzw. G und wir definieren

$$\frac{dG}{dF}(x) = \lim_{h \to 0} \frac{G(x + h) - G(x)}{F(x + h) - F(x)},$$

wenn dieser Grenzwert existiert, und wieder mit der Konvention, dass wir die rechte Seite gleich 0 setzen, wenn Zähler und Nenner 0 sind, und gleich ∞, wenn nur der Nenner 0 ist.

Satz 7.7

F sei stetig und $\nu(A) = 0$. Dann ist μ-fast überall auf A $\frac{dG}{dF} = 0$.

Wir müssen nur zeigen, dass

$$D_+(x) = \limsup_{h \downarrow 0} \frac{G(x + h) - G(x)}{F(x + h) - F(x)} = 0$$

und

$$D_-(x) = \limsup_{h \downarrow 0} \frac{G(x) - G(x - h)}{F(x) - F(x - h)} = 0$$

fast überall auf A gilt (weil die entsprechenden limites inferiores jedenfalls nichtnegativ sind).

Dazu zeigen wir, dass für jedes $c > 0$ die Mengen $A \cap [D_+ > c]$ und $A \cap [D_- > c]$ Maß 0 haben.

Wir beginnen mit dem Beweis für D_+, der für D_- lässt sich fast darauf zurückführen, indem wir

$$\tilde{\mu}(B) = \mu(\ominus B), \tilde{\nu}(B) = \nu(\ominus B), \tilde{F}(x) = -F(-x), \tilde{G}(x) = -G(-x)$$

setzen. Ein kleiner Schönheitsfehler dabei besteht darin, dass $\tilde{G}$ nicht rechtsstetig ist, wenn G Sprünge hat. Um die Symmetrie herzustellen, verzichten wir auf die Annahme, dass G rechtsstetig ist, und nehmen nur an, dass $G(.+0)$, die rechtsstetige Modifikation von G, eine Verteilungsfunktion von ν ist.

Wir können zu jedem $\epsilon > 0$ eine Überdeckung von A durch abzählbar viele disjunkte Intervalle $]a_n, b_n]$ mit $\sum_n \nu(]a_n, b_n]) < \epsilon$ finden. Für $x \in [D_+ > c] \cap]a_n, b_n]$ ist entweder $x = b_n$, oder es gibt ein y zwischen x und b_n mit

$$G(y) - G(x) > c(F(y) - F(x)),$$

also

$$G(y) - cF(y) > G(x) - cF(x).$$

Nun ist entweder x Element der (abzählbaren) Menge $\mathcal{D}_G$ der Unstetigkeitspunkte von G, oder $f = G - cF$ ist in x stetig, also $\hat{f}(x) = f(x)$ und daher x ein von rechts unsichtbarer Punkt von f in $]a_n, b_n]$. Die Menge dieser Punkte ist nach dem Satz von Riesz eine Vereinigung von (höchstens) abzählbar vielen disjunkten Intervallen $]a_{nk}, b_{nk}[$, und es gilt

$$f(b_{nk} + 0) = \hat{f}(b_{nk}) \geq f(a_{nk} + 0).$$

Das gibt

$$F(b_{nk}) - F(a_{nk}) \leq \frac{G(b_{nk} + 0) - G(a_{nk} + 0)}{c},$$

also

$$\mu(]a_{nk}, b_{nk}]) \leq \frac{1}{c}\nu(]a_{nk}, b_{nk}]).$$

$[D_+ > c] \cap A$ ist in

$$\bigcup_{n,k}]a_{nk}, b_{nk}] \cup \{b_n, n \in \mathbb{N}\} \cup \mathcal{D}_G$$

enthalten. Da die beiden abzählbaren Mengen μ-Maß 0 haben, gilt

$$\mu([D_+ > c] \cap A) \leq \sum_{n,k} \mu(]a_{nk}, b_{nk}]) \leq \frac{1}{c} \sum_{n,k} \nu(]a_{nk}, b_{nk}]) \leq \frac{1}{c} \sum_n \nu(]a_n, b_n]) \leq \frac{\epsilon}{c}.$$

Da $\epsilon > 0$ beliebig ist, gilt also

$$\mu([D_+ > c] \cap A) = 0.$$

Aus dem letzten Satz ergibt sich die unmittelbare Folgerung:

Satz 7.8

Wenn F stetig ist und μ_G zu μ_F singulär ist, dann ist μ_F-fast überall

$$\frac{dG}{dF} = 0.$$

Als nächstes kommt der Fall $\mu_G \ll \mu_F$. Wir beginnen mit dem wichtigen Sonderfall $G = F$:

Satz 7.9

F sei eine stetige Verteilungsfunktion. Dann ist μ_F-fast überall

$$\frac{dF}{dF} = 1.$$

Der Beweis ist der Leserin überlassen. Es mag auf den ersten Blick verwundern, dass hier überhaupt etwas zu beweisen ist, aber nach unserer Definition der Ableitung kann eine einseitige Ableitung 0 sein, wenn F auf einem Intervall konstant ist.

Wenn nun $\mu_G \ll \mu_F$, beide endlich sind,

$$f = \frac{d\mu_G}{d\mu_F}$$

und F immer noch stetig, dann können wir wie im Beweis des Satzes von Radon-Nikodym für ein $c > 0$ wieder die Jordanzerlegung von

$$\nu_c = \mu_G - c\mu_F = \nu_{c+} - \nu_{c-}$$

bilden. Wir nennen die Verteilungsfunktionen von Positiv- und Negativteil H_{c+} und H_{c-}. Es sei nun $h > 0$. Dann ist

$$\nu_{nh-}([f > (n+1)h]) = \nu_{nh+}([f < (n-1)h]) = 0.$$

Damit ist auf $[f > (n+1)h]$ fast überall

$$\frac{dH_{nh-}}{dF} = 0.$$

Weil auf jeden Fall

$$\liminf_{t\to 0} \frac{H_{c\pm}(x+t) - H_{c\pm}(x)}{F(x+t) - F(x)} \geq 0$$

ist, ergibt sich daraus

$$\liminf_{t\to 0} \frac{G(x+t) - G(x)}{F(x+t) - F(x)} \geq nh.$$

Genauso ergibt sich fast überall auf $[f < (n-1)h]$

$$\limsup_{t\to 0} \frac{G(x+t) - G(x)}{F(x+t) - F(x)} \leq nh.$$

Auf $[nh \leq f \leq (n+1)h]$ ergibt sich

$$f(x)-2h \leq (n-1)h \leq \liminf_{t\to 0} \frac{G(x+t) - G(x)}{F(x+t) - F(x)} \leq \limsup_{t\to 0} \frac{G(x+t) - G(x)}{F(x+t) - F(x)} \leq (n+2)h \leq f(x)+2h.$$

Da f fast überall endlich ist, gibt es für jedes x ein n mit $nh \leq x < (n_1)h$, und daher gilt fast überall auf $\mathbb{R}$

$$f(x) - 2h \leq \liminf_{t\to 0} \frac{G(x+t) - G(x)}{F(x+t) - F(x)} \leq \limsup_{t\to 0} \frac{G(x+t) - G(x)}{F(x+t) - F(x)} \leq f(x) + 2h.$$

Für $h \to 0$ ergibt sich daraus

$$\frac{dG}{dF} = f$$

μ_F-fast überall.

Es bleibt noch der Fall, dass F nicht stetig ist, also Sprünge hat. Dann können wir μ_F in einen stetigen und einen diskreten Teil zerlegen. Wir bezeichnen diese Maße mit μ_c und μ_d, ihre Verteilungsfunktionen mit F_c und F_d und die Menge der Unstetigkeitspunkte von F mit D. Wir wissen, dass $\mu_c(D) = \mu_d(D^C) = 0$ ist. Deshalb gilt auf D^C μ_c-fast überall, was dasselbe ist wie μ_F-fast überall,

$$\frac{dG}{dF} = \frac{\frac{dG}{dF_c}}{\frac{dF_c}{dF_c} + \frac{dF_d}{dF_c}} = \frac{\frac{dG}{dF_c}}{1+0} = \frac{d_G}{d_F} = \frac{d\mu_{G\ll F_c}}{d\mu_{F_c}} = \frac{d\mu_{G\ll F}}{d\mu_F}.$$

Auf D benötigen wir die symmetrische Version

$$\lim_{h\to 0} \frac{G(x+h) - G(x-h)}{F(x+h) - F(x-h)} = \frac{\mu_G(\{x\})}{\mu_F(\{x\})},$$

und auch das ist die Radon-Nikodym Dichte.

7.1.1 Kochrezept für reelle Verteilungsfunktionen

Die beliebte Übungsaufgabe "Bestimmen Sie die Lebesgue-Zerlegung von μ_G bezüglich μ_F" kann man also so angehen:

1. Bestimme
$$f(x) = \lim_{\epsilon \to 0} \frac{G(x+\epsilon) - G(x-\epsilon)}{F(x+\epsilon) - F(x-\epsilon)}.$$
Das gibt die Radon-Nikodym Dichte.

2. Integriere f bezüglich μ_F:
$$G_c(x) = \begin{cases} \int_{]0,x]} f d\mu_F & \text{wenn } x \geq 0, \\ -\int_{]x,0]} f d\mu_F & \text{wenn } x < 0. \end{cases}$$

3. $G_s = G - G_c$.

In dem häufigen Fall, dass F und G endlich viele Sprungstellen haben und zwischen den Sprungstellen stetig differenzierbar sind, reduziert sich das auf die folgenden Regeln:

Betrachte zuerst die Punkte x, die Sprungstellen einer der beiden Funktionen sind:

- $F(x) \neq F(x-0)$:
$$\frac{d\mu_{G_c}}{d\mu_F}(x) = \frac{G(x) - G(x-0)}{F(x) - F(x-0)},$$
der Sprung von G kommt in den absolutstetigen Anteil, also
$$G_c(x) - G_c(x-0) = G(x) - G(x-0).$$

- $F(x) = F(x-0)$: $\frac{d\mu_{G_c}}{d\mu_F}(x)$ kann beliebig gesetzt werden, etwa $\frac{d\mu_{G_c}}{d\mu_F}(x) = 0$; der Sprung von G kommt in den singulären Anteil, also
$$G_s(x) - G_s(x-0) = G(x) - G(x-0).$$

Zwischen den Sprungstellen betrachten wir die Ableitungen:

- $F'(x) > 0$:
$$\frac{d\mu_{G_c}}{d\mu_F}(x) = \frac{G'(x)}{F'(x)},$$
$$G_c'(x) = G'(x).$$

- $F'(x) = 0$: $\frac{d\mu_{G_c}}{d\mu_F}(x)$ kann beliebig gesetzt werden, etwa $\frac{d\mu_{G_c}}{d\mu_F}(x) = 0$;
$$G_s'(x) = G'(x).$$

7.2 Bedingte Erwartungen

Der Satz von Radon-Nikodym erlaubt uns einen neuen Blick auf die bedingten Wahrscheinlichkeiten. Das wird uns die Möglichkeit geben, unter gewissen Umständen bedingte Wahrscheinlichkeiten zu definieren, wenn das bedingende Ereignis Wahrscheinlichkeit 0 hat.

Wir betrachten eine Zufallsvariable Y auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{S}, \mathbb{P})$, die nur endlich viele Werte annimmt, etwa $1, \ldots, n$. Alternativ können wir n disjunkte Mengen $A_1, \ldots, A_n$ wählen. Zwischen beiden Betrachtungsweisen können wir nach Belieben hin- und herschalten, indem wir einerseits $A_i = [Y = i]$ setzen, und andererseits $Y = \sum_i iA_i$. Wir nehmen an, dass wir die Wahrscheinlichkeiten $\mathbb{P}(A_i)$ und für eine Menge $B \in \mathfrak{S}$ die bedingten Wahrscheinlichkeiten $\mathbb{P}(B|A_i)$ kennen. Etwas allgemeiner können wir für eine integrierbare Zufallsvariable X und eine

Borelmenge C die Wahrscheinlichkeiten $\mathbb{P}(X \in C|A_i)$, also die bedingte Verteilung bestimmen, etwa die bedingte Verteilungsfunktion

$$F_i(x) = \mathbb{P}(X \leq x|A_i),$$

Den Erwartungswert dieser bedingten Verteilung nennen wir den bedingten Erwartungswert

$$\mathbb{E}(X|A_i) = \int x d\mu_{F_i}(x).$$

Der Satz von der vollständigen Wahrscheinlichkeit funktioniert auch hier: wegen

$$F_X(x) = \sum_i F_i(x)\mathbb{P}(A_i)$$

ist

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X|A_i)\mathbb{P}(A_i)$$

Wenn wir die Ereignisse A_i durch eine Zufallsvariable Y definiert haben, schreiben wir für den bedingten Erwartungswert natürlich

$$\mathbb{E}(X|Y = i).$$

Es ist leicht zu sehen, dass

$$\mathbb{E}(X|Y = i) = \frac{\mathbb{E}(X[Y = i])}{\mathbb{P}(Y = i)}$$

gilt.

Weil die Notation $\mathbb{E}(X|Y = Y)$ etwas seltsam aussieht, setzen wir

$$g(i) = \mathbb{E}(X|Y = i).$$

In diese Funktion können wir ohne Schwierigkeiten Y einsetzen. Diese Zufallsvariable $Z = g(Y)$ ist offensichtlich auf jeder der Mengen $A_i = [Y = i]$ konstant. In der alternativen Sichtweise können wir sie als

$$Z = \sum_i \mathbb{E}(X|A_i)A_i()$$

schreiben. Auf sehr hohem Niveau können wir sagen, dass Z bezüglich der Sigmaalgebra $\mathfrak{A}$ messbar ist, die von Y bzw. $A_1, \ldots, A_n$ erzeugt wird.

Wir können Z integrieren:

$$\mathbb{E}(Z) = \sum_i \mathbb{E}(X|Y = i)\mathbb{P}(Y = i) = \mathbb{E}(X).$$

Etwas komplizierter

$$\mathbb{E}([Y = i]Z) = \mathbb{P}(Y = i)\mathbb{E}(X|Y = i) = \mathbb{E}([Y = i]X).$$

Wenn wir für $[Y = i]$ wieder A_i setzen

$$\mathbb{E}(A_i Z) = \mathbb{E}(A_i X).$$

Wir können mehrere dieser Gleichungen addieren und so statt A_i eine beliebige Vereinigung dieser Mengen einsetzen. Alle diese Mengen bilden die Sigmaalgebra $\mathfrak{A}$; die Zufallsvariable Z erlaubt es uns also, die Erwartung $\mathbb{E}(AX)$ und damit den bedingten Erwartungswert $\mathbb{E}(X|A)$ zu bestimmen:

$$\mathbb{E}(AX) = \mathbb{E}(AZ). \tag{7.1}$$

Wir nennen Z daher die bedingte Erwartung von X unter $\mathfrak{A}$ bzw. unter Y und schreiben

$$Z = \mathbb{E}(X|\mathfrak{A}) = \mathbb{E}(X|Y).$$

Es wäre schön, etwas ähnliches wie (7.1) für Zufallsvariable bzw. Sigmaalgebren zu haben, die nicht ganz so primitiv sind. Wir wünschen uns also

Definition 7.2

$(\Omega, \mathfrak{S}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $\mathfrak{A}$ eine Untersigmaalgebra von $\mathfrak{S}$ und $X \in \mathcal{L}_1(\mathbb{P})$. Wir nennen Z die bedingte Erwartung von X unter $\mathfrak{A}$, wenn Z $\mathfrak{A}$-messbar ist und für jedes $A \in \mathfrak{A}$

$$\mathbb{E}(AX) = \mathbb{E}(AZ)$$

gilt. Wir schreiben

$$Z = \mathbb{E}(X|\mathfrak{A})$$

ist $Y : (\Omega, \mathfrak{S}) \to (\Omega_2, \mathfrak{S}_2)$, dann nennen wir

$$\mathbb{E}(X|Y) = \mathbb{E}(X|Y^{-1}(\mathfrak{S}_2))$$

die bedingte Erwartung von X unter Y. Ist $g : (\Omega_2, \mathfrak{S}_2) \to (\mathbb{R}, \mathfrak{B})$ mit

$$\mathbb{E}(X|Y) = g \circ Y,$$

dann schreiben wir

$$\mathbb{E}(X|Y = y) = g(y).$$

Für ein Ereignis $A \in \mathfrak{S}$ definieren wir die bedingten Wahrscheinlichkeiten $\mathbb{P}(A|\mathfrak{A})$, $\mathbb{P}(A|Y)$ und $\mathbb{P}(A|Y = y)$ als die entsprechenden bedingten Erwartungen der Indikatorfunktion $A()$.

Die Funktion $\mathbb{E}(X|Y = y)$ heißt auch die Regressionsfunktion von X bezüglich Y.

Der Satz von Radon-Nikodym garantiert die Existenz der bedingten Erwartung:

$$\nu(A) = \mathbb{E}(AX), A \in \mathfrak{A}$$

definiert ein endliches signiertes Maß auf $\mathfrak{A}$, das absolutstetig bezüglich $\mathbb{P}$ (genauer: bezüglich der Einschränkung von $\mathbb{P}$ auf $\mathfrak{A}$) ist. Es gibt daher eine $\mathfrak{A}$-messbare Zufallsvariable Z mit

$$\mathbb{E}(AZ) = \nu(A) = \mathbb{E}(AX)$$

für alle $A \in \mathfrak{A}$. Z ist also eine bedingte Erwartung von X unter $\mathfrak{A}$. Da jede Radon-Nikodym Dichte das erfüllt, ist eine bedingte Erwartung nur bis auf fast sichere Übereinstimmung eindeutig bestimmt. Dieser Mangel an Eindeutigkeit ist der Preis für die größere Allgemeinheit dieses Begriffes. Satz 4.7 garantiert dann die Existenz von $\mathbb{E}(X|Y = y)$.

Die Eigenschaften der bedingten Erwartung sind in weiten Teilen analog zu denen des gewöhnlichen Erwartungswertes. Es ist allerdings zu beachten, dass die Gleichungen und Ungleichungen zwischen bedingten Erwartungen nur fast sicher gelten.

Satz 7.10: Eigenschaften der bedingten Erwartung

In den folgenden Aussagen sind X, X_n, Y, Z integrierbare Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{S}, \mathbb{P})$, $\mathfrak{A}, \mathfrak{B}$ Untersigmaalgebren von $\mathfrak{S}$.

1. Linearität:
$$\mathbb{E}(aX + b|\mathfrak{A}) = a\mathbb{E}(X|\mathfrak{A}) + b.$$
2. Additivität:
$$\mathbb{E}(X + Y|\mathfrak{A}) = \mathbb{E}(X|\mathfrak{A}) + \mathbb{E}(Y|\mathfrak{A})$$
3. Monotonie: aus $X \leq Y$ folgt
$$\mathbb{E}(X|\mathfrak{A}) \leq \mathbb{E}(Y|\mathfrak{A}).$$
4. Satz von der monotonen Konvergenz: gilt $X_n \uparrow X$, dann gilt
$$\lim_n \mathbb{E}(X_n|\mathfrak{A}) = \mathbb{E}(X|\mathfrak{A}).$$

5. Satz von der dominierten Konvergenz: gilt $X_n \to X$ und $|X_n| \leq Y$, dann ebenfalls

$$\lim_n \mathbb{E}(X_n|\mathfrak{A}) = \mathbb{E}(X|\mathfrak{A}).$$

6. Ungleichung von Jensen: ist g konvex und $g \circ X$ integrierbar, dann gilt

$$\mathbb{E}(g(X)|\mathfrak{A}) \geq g(\mathbb{E}(X|\mathfrak{A})).$$

7. Die Turmeigenschaft: ist $\mathfrak{A} \subseteq \mathfrak{B}$, dann

$$\mathbb{E}(X|\mathfrak{A}) = \mathbb{E}(\mathbb{E}(X|\mathfrak{B})|\mathfrak{A}).$$

8. Falls X von $\mathfrak{A}$ unabhängig ist, dann gilt

$$\mathbb{E}(X|\mathfrak{A}) = \mathbb{E}(X).$$

9. Falls X bezüglich $\mathfrak{A}$-messbar ist und XY integrierbar dann gilt

$$\mathbb{E}(XY|\mathfrak{A}) = X\mathbb{E}(Y|\mathfrak{A}).$$

10. Der Satz von der vollständigen Wahrscheinlichkeit:

$$\mathbb{E}(\mathbb{E}(X|\mathfrak{A})) = \mathbb{E}(X).$$

Die elementaren bedingten Wahrscheinlichkeiten haben noch eine zusätzliche Eigenschaft, nämlich, dass sie sigmadditiv sind. Für endliche Sigmaalgebren $\mathfrak{A} \subseteq \mathfrak{S}$ gilt also, dass für fast alle ω durch

$$\nu(A) = \mathbb{P}(A|\mathfrak{A})(\omega)$$

ein Maß definiert wird. Wir können uns diese Eigenschaft auch im allgemeinen Fall wünschen:

Definition 7.3

$(\Omega_i, \mathfrak{S}_i)$ $(i = 1, 2)$ seien zwei Messräume. Die Funktion $p : \Omega_1 \times \mathfrak{S}_2 \to [0, \infty[$ heißt *Markovkern* oder *Übergangsfunktion* von $(\Omega_1, \mathfrak{S}_1)$ nach $(\Omega_2, \mathfrak{S}_2)$, wenn sie folgende Bedingungen erfüllt.

1. Für jedes $\omega \in \Omega_1$ ist $p(\omega, .)$ ein (endliches) Maß.
2. Für jedes $A \in \mathfrak{S}_2$ ist $p(., A)$ $\mathfrak{S}_1$-messbar.

Anmerkung: Diese Definition ist etwas allgemeiner als das, was wir hier brauchen, weil wir hauptsächlich an dem Fall interessiert sind, dass $p(\omega, .)$ ein Wahrscheinlichkeitsmaß ist. Andererseits könnte man sie noch etwas allgemeiner machen, indem wir statt der Endlichkeit von $p(\omega, .)$ nur die Sigmaendlichkeit fordern. Das geht allerdings nicht in beliebiger Weise, wir müssen annehmen, dass diese Sigmaendlichkeit gleichmäßig funktioniert, also das es eine Zerlegung von Ω_2 in abzählbar viele Mengen B_n gibt, sodass für alle n und $\omega \in \Omega_1$ $p(\omega, B_n) < \infty$ gilt.

Definition 7.4

Eine bedingte Wahrscheinlichkeit

$$p(\omega, A) = \mathbb{E}(A|\mathfrak{A})(\omega)$$

heißt reguläre bedingte Wahrscheinlichkeit, wenn p ein Markovkern von $(\Omega, \mathfrak{A})$ nach $(\Omega, \mathfrak{S})$ ist.

Das ist nicht so selbstverständlich, wie es auf den ersten Blick erscheint. Es ist natürlich die Sigmaadditivität, die hier Schwierigkeiten macht. Für eine bestimmte Folge A_n von disjunkten Ereignissen ist zwar die Sigmaadditivität fast sicher erfüllt, aber da es in nicht diskreten Maßräumen

überabzählbar viele solcher Folgen gibt, können sich die einzelnen Ausnahmemengen zu einer Menge von positivem Maß zusammenfinden, und in der Tat gibt es Beispiele von Räumen mit bedingten Wahrscheinlichkeiten, die nicht als Markovkern geschrieben werden können. In unserem Lieblingsraum funktioniert es aber:

Satz 7.11

Wenn der zugrundeliegende Raum $\mathbb{R}$ (oder $\mathbb{R}^n$) mit den Borelmengen ist, dann existiert eine reguläre Version der bedingten Wahrscheinlichkeit.

Zum Beweis erinnern wir uns daran, dass wir ein endliches Maß auf $\mathbb{R}$ durch seine Verteilungsfunktion festlegen können. Wir setzen also für $x \in \mathbb{Q}$

$$Y(x) = \mathbb{P}((-\infty, x]|\mathfrak{A}).$$

Für zwei rationale Zahlen $x_1 < x_2$ gilt fast sicher $Y(x_1) \leq Y(x_2)$ und $Y(x_1) = Y(x_1 + 0)$. Da es nur abzählbar viele solche Paare gibt, ist die Vereinigung der abzählbar vielen Nullmengen, auf denen diese Beziehungen nicht gelten, selbst eine Nullmenge, und auf ihrem Komplement ist $Y(.)$ nichtfallend und rechtsstetig (auf den rationalen Zahlen). Wir setzen $Y(.)$ zu einer monotonen und rechtsstetigen Funktion auf ganz $\mathbb{R}$ fort (eigentlich ist das eine Funktion $Y(x, \omega)$ von zwei Variablen, die für fast jedes ω eine monotone Funktion von x ist). Das von $Y(x, \omega)$ erzeugte Maß $p(\omega, A)$ ist die gesuchte reguläre Version der bedingten Wahrscheinlichkeit. Wegen der Additivität des Erwartungswerts gilt

$$p(., A) = \mathbb{P}(A|\mathfrak{A})$$

wenn A eine endliche Vereinigung von disjunkten Intervallen mit rationalen Endpunkten ist. Aus dem Satz von der dominierten Konvergenz folgt, dass das System aller Mengen A, für die diese Beziehung gilt, monoton ist. Damit enthält es den Sigmaring, der von den Intervallen mit rationalen Endpunkten erzeugt wird, also die Borelmengen.

Wir schließen diesen Abschnitt mit einem Beispiel, in dem keine reguläre bedingte Wahrscheinlichkeit existiert:

Es sei $\Omega = [0, 1]$, $\mathfrak{S} = \mathfrak{A}_\sigma(\mathfrak{B} \cup \{A\})$, wobei $A \subset [0, 1]$ mit $\lambda^*(A) = \lambda^*(A^C) = 1$. Die Elemente von $\mathfrak{S}$ sind bekanntlich als $(B_1 \cap A) \cup (B_2 \cap A^C)$, $B_1, B_2 \in \mathfrak{B}$ darstellbar. Wir definieren durch

$$\mathbb{P}((B_1 \cap A) \cup (B_2 \cap A^C)) = \frac{1}{2}(\lambda(B_1) + \lambda(B_2))$$

ein Wahrscheinlichkeitsmaß auf $\mathfrak{S}$ (das einzige, was hier noch eines Beweises bedarf, ist die Wohldefiniertheit, aber wenn $A \cap B_1 = A \cap B_1'$, dann ist $B_1 \triangle B_1'$ eine Teilmenge von A^C und daher vom Maß 0.

Die bedingte Wahrscheinlichkeit $\mathbb{P}(.|\mathfrak{B})$ hat keine bedingte Version. Wir betrachten

$$F(x, \omega) = \mathbb{P}(A \cap [0, x]|\mathfrak{B})(\omega).$$

Für jedes x gilt für fast alle ω

$$F(x, \omega) = \frac{1}{2}[0, \infty](\omega). \tag{7.2}$$

Für jedes x bezeichnen wir die Nullmenge, auf der diese Gleichung nicht gilt, mit N_x und setzen

$$N = \cap_{x \in \mathbb{Q} \cap [0,1]} N_x.$$

N ist auch eine Nullmenge, und für $\omega \in N^C$ gilt (7.2) für alle $x \in \mathbb{Q} \cap [0, 1]$. Nun soll $P(. \cap A|\mathfrak{B})(\omega)$ ein Maß auf $\mathfrak{B}$ sein und wird also durch seine Verteilungsfunktion festgelegt. Es muss also für alle $\omega \in N^C$

$$(A \cap B|\mathfrak{B})(\omega) = A(\omega).$$

Wenn wir $\omega \in A^C \cap N^C$ wählen, dann ist einerseits nach den vorigen Feststellungen

$$\mathbb{P}(A \cap \{\omega\}|\mathfrak{B})(\omega) = 1/2,$$

andererseits

$$\mathbb{P}(A \cap \{\omega\}|\mathfrak{B})(\omega) = \mathbb{P}(A \cap \emptyset|\mathfrak{B})(\omega) = 0.$$

Wir kommen also zu einem Widerspruch, und es gibt tatsächlich keine reguläre Version dieser bedingten Wahrscheinlichkeit.

7.3 Ergänzungen

7.3.1 Hauptsatz-Variationen

Der klassische Hauptsatz der Differential- und Integralrechnung besagt, dass das Integral einer stetigen Funktion eine Stammfunktion ist und dass jede stetig differenzierbare Funktion das Integral ihrer Ableitung ist.

Wir haben einige Ergänzungen dazu gefunden. Der Satz von Lebesgue sagt uns, dass die Ableitung des Integrals einer Lebesgue-integrierbaren bzw. Riemann-integrierbaren Funktion f fast überall existiert und mit f übereinstimmt. Das ist eine Verallgemeinerung des ersten Teils des Hauptsatzes, allerdings in abgeschwächter Form. In der anderen Richtung können wir sagen, dass eine absolutstetige Funktion fast überall differenzierbar ist und als Integral dieser (fast überall definierten) Ableitung darstellbar. Was aber geschieht, wenn wir nur wissen, dass eine Funktion F differenzierbar ist? Können wir dann diese Funktion aus ihrer Ableitung zurückgewinnen? Es ist klar, dass wir uns hier keine Hoffnungen machen dürfen, wenn F nur fast überall differenzierbar ist, weil ja jede singuläre monotone Funktion fast überall Ableitung 0 hat. Es bleibt also die Frage, ob es ein Integral gibt, das man auf die Ableitung einer überall differenzierbaren Funktion anwenden kann und so F zurückbekommt. Das gibt es tatsächlich und heißt verallgemeinertes Riemannintegral (leider versagen Riemann- und Lebesgue-Integral hier, wie die Funktion $F(x) = x^2 \sin(1/|x|^3)$ im Interval $[-1, 1]$ zeigt). Für seine Definition führen wir zuerst den Begriff einer Eichfunktion ein:

Definition 7.5

Eine Eichfunktion auf dem Intervall $[a, b]$ ist eine Abbildung $\gamma : [a, b] \to]0, \infty[$.

Eine markierte Partitionierung von $[a, b]$ ist ein Paar

$$\mathcal{T} = ((t_0, \ldots, t_n), (x_1, \ldots, x_n))$$

Mit $t_{i-1} \leq x_i \leq t_i$ und $t_0 = a$ und $t_n = b$.

γ sei eine Eichfunktion. Die markierte Patitionierung $\mathcal{T}$ heißt γ-fein, wenn für $i = 1, \ldots, n$

$$t_i - t_{i-1} \leq \gamma(x_i)$$

gilt.

Für eine reelle Funktion $f : [a, b] \to \mathbb{R}$ und eine markierte Partition $\mathcal{T}$ ist die Riemannsumme

$$I(f, \mathcal{T}) = \sum_{i=1}^{n} f(x_i)(t_i - t_{i-1}).$$

Definition 7.6

Eine Funktion $f : [a, b] \to \mathbb{R}$ heißt verallgemeinert Riemann-integrierbar, wenn es eine Zahl I und zu jedem $\epsilon > 0$ eine Eichfunktion γ_ϵ gibt, sodass für jede γ-feine Partition $\mathcal{T}$

$$|I(f, \mathcal{T}) - I| < \epsilon$$

gilt. I nennen wir das verallgemeinerte Riemann-Integral oder Henstock-Kurzweil-Integral

$$\text{HK-}\int_a^b f(x)dx = I.$$

7.4 Wiederholungsfragen

1. Was sagt der Satz von Radon-Nikodym?
2. Wie sind absolute Stetigkeit und Singularität definiert?

3. Was besagt der Zerlegungssatz von Lebesgue?
4. Was kann man über die Differenzierbarkeit von monotonen Funktionen sagen?
5. Wie ist die bedingte Erwartung einer Zufallsvariable bezüglich eine Sigmaalgebra bzw. einer Zufallsvariablen definiert?
6. Wie ist die Regressionsfunktion definiert?
7. Welche Eigenschaften hat die bedingte Erwartung?
8. Was ist eine reguläre bedingte Erwartung? Was kann man über ihre Existenz sagen?

Kapitel 8

Integralungleichungen und L_p-Räume

Wir beginnen mit der sehr vielseitig verwendbaren Ungleichung von Jensen:

Satz 8.1

μ sei ein Wahrscheinlichkeitsmaß, f sei integrierbar, und ϕ sei eine konvexe Funktion. Dann gilt

$$\int \phi \circ f \, d\mu \geq \phi(\int f d\mu).$$

Zum Beweis stellen wir fest, dass die Bedingung für die Konvexität von ϕ äquivalent zu der Forderung ist, dass es zu jedem x ein C gibt, sodass für alle y

$$\phi(y) \geq \phi(x) + C(y - x).$$

Wenn wir hier y durch $f(x)$ und x durch $m = \int f d\mu$ ersetzen und integrieren, erhalten wir

$$\int \phi(f) d\mu \geq \phi(m) + C(\int f d\mu - m),$$

also

$$\int \phi(f) d\mu \geq \phi(\int f d\mu) + C(m - m),$$

und die Ungleichung ist bewiesen.

Wir nennen eine messbare Funktion f p-fach integrierbar (p reell und positiv), wenn

$$\int |f|^p d\mu < \infty.$$

$\mathcal{L}_p$ sei die Menge aller p-fach integrierbaren Funktionen. Weiters sei

$$\|f\|_p = (\int |f|^p d\mu)^{1/p}.$$

Es gilt

Satz 8.2: Ungleichung von Hölder

Wenn $f \in \mathcal{L}_p$ und $g \in \mathcal{L}_q$ mit $1/p + 1/q = 1$ ($1 \leq p \leq \infty$, $q = \infty$, wenn $p = 1$ und umgekehrt). Dann gilt

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q.$$

Der Fall $p = 1$ bzw. $p = \infty$ kann recht einfach gesondert behandelt werden, dieser Teil bleibt de* Lese* überlassen.

Es bleibt der Fall $1 < p < \infty$. Wir verwenden die Ungleichung von Jensen mit $\phi(x) = |x|^p$, als Wahrscheinlichkeitsmaß verwenden wir

$$\nu(A) = \int_A |\frac{g}{\|g\|_q}|^q d\mu$$

und statt f setzen wir

$$\frac{|f|}{|g|^{q-1}}.$$

Wir erhalten

$$(\int \frac{|fg|}{\|g\|_q^q} d\mu)^p \leq \frac{1}{\|g\|_q^q} \int |f|^p |g|^{p+q-pq} d\mu,$$

und wegen $p + q = pq$

$$(\int |fg| d\mu)^p \leq \|g\|_q^{pq-q} \|f\|_p^p = \|g\|_q^p \|f\|_p^p,$$

und Ziehen der p-ten Wurzel gibt das Gewünschte.

Satz 8.3: Ungleichung von Minkowski

f und g seien aus $\mathcal{L}_p$, $p \geq 1$. Dann ist auch $f + g$ in $\mathcal{L}_p$ und

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Wegen

$$|f + g|^p \leq |2 \max(|f|, |g|)|^p \leq 2^p (|f|^p + |g|^p)$$

ist $f + g$ p-fach integrierbar. Außerdem ist

$$\int |f + g|^p d\mu = \int |f + g||f + g|^{p-1} d\mu \leq \int |f||f + g|^{p-1} d\mu + \int |g||f + g|^{p-1} d\mu.$$

Nach der Hölderschen Ungleichung gilt (mit $1/p + 1/q = 1$)

$$\int |f||f + g|^{p-1} d\mu \leq \|f\|_p (\int |f + g|^{q(p-1)} d\mu)^{1/q} = \|f\|_p \|f + g\|_p^{p-1}.$$

Zusammengefasst:

$$\|f + g\|^p \leq (\|f\|_p + \|g\|_p) \|f + g\|_p^{p-1}.$$

Falls $\|f+g\|_p$ positiv ist, können wir hier kürzen, im anderen Fall ist die Ungleichung von Minkowski trivial erfüllt.

Die Minkowski-Ungleichung ist die Dreiecksungleichung für $\|f\|_p$. Die Nichtnegativität und die positive Linearität sind offensichtlich, also ist $\|f\|_p$ auf $\mathcal{L}_p$ eine Seminorm. Das einzige, was zu einer vollen Norm fehlt, ist dass aus $\|f\|_p = 0$ auch $f = 0$ folgen soll (wir können nur folgern, dass μ-fast überall $f = 0$ ist). Dies kann man erreichen indem man fast überall übereinstimmende Funktionen identifiziert. Es sei also $\sim$ die Äquivalenzrelation, die durch $f \sim g$ wenn $f = g$ fast überall definiert ist, und

$$L_p = (\mathcal{L}_p)/\sim .$$

Dann ist L_p mit der Norm $\|f\|_p$ ein normierter Vektorraum.

Für $0 < p < 1$ folgt aus der relativ trivialen Ungleichung

$$(x + y)^p \leq x^p + y^p (x, y \geq 0)$$

(entweder ist $x + y \geq 0$ oder

$$(\frac{x}{x+y})^p + (\frac{y}{x+y})^p \geq \frac{x}{x+y} + \frac{y}{x+y} = 1),$$

dass

$$d(f,g) = \|f-g\|_p^p$$

eine Pseudometrik auf $\mathcal{L}_p$ definiert, die auf L_p zu einer Metrik wird. $\|.\|_p$ erfüllt statt der Dreiecksungleichung die Ungleichung

$$\|f+g\|_p \leq 2^{1/p-1}(\|f\|_p + \|g\|_p,$$

was $\|.\|_p$ zu einer *Quasinorm* macht (eine Quasinorm erfüllt alle Axiome einer Norm bis auf die Dreiecksungleichung, die zu $\|x+y\| \leq C(\|x\|+\|y\|)$ mit einer Konstanten $C > 1$ abgeschwächt wird).

Es gilt sogar

Satz 8.4: Riesz-Fischer

L$_p$ ist vollständig bezüglich $\|.\|_p$ und speziell für $p \geq 1$ ein Banachraum.

Es ist zu zeigen, dass jede Cauchyfolge konvergiert. Es sei also f_n eine Cauchyfolge bezüglich $\|.\|_p$. Aus der Ungleichung von Markov folgt, dass f_n auch eine Cauchyfolge im Maß ist und daher gegen eine Funktion f im Maß konvergiert. Wir können eine Teilfolge f_{n_k} auswählen, die fast überall gegen eine Funktion f konvergiert. Das Lemma von Fatou impliziert

$$\int |f|^p d\mu = \int \liminf |f_{n_k}|^p d\mu \leq \liminf \int |f_{n_k}|^p < \infty.$$

f ist also p-fach integrierbar. Wir wählen k_0 so, dass für $m,n \geq n_{k_0}$ $\|f_n - f_m\|_p \leq \epsilon$ gilt. Dasselbe Argument wie vorher liefert

$$\|f_{n_{k_0}} - f\|_p \leq \epsilon$$

und für $n \geq n_{k_0}$

$$\|f_n - f\|_p \leq \|f_n - f_{n_{k_0}}\|_p + \|f_{n_{k_0}} - f\|_p \leq 2\epsilon.$$

Also konvergiert die Folge f_n im p-ten Mittel gegen f.

Für die Konvergenz in $\mathcal{L}_p$ lassen sich Kriterien angeben:

Satz 8.5

(f_n) konvergiert genau dann gegen f im p-ten Mittel, wenn (f_n) gegen f im Maß konvergiert, und wenn eine der beiden folgenden Bedingungen erfüllt ist (und damit die andere):

1. $(|f_n|^p)$ ist gleichmäßig integrierbar.
2. $\limsup_{n\to\infty} \|f_n\|_p \leq \|f\|_p < \infty$.

Aus der Konvergenz im Maß von (f_n) können wir leider nicht die Konvergenz im Maß von $|f_n|^p$ folgern, weil die Potenzfunktion nicht gleichmäßig stetig ist. Auch hier kommt uns die Teilfolgenkonvergenz zu Hilfe, die aus der Konvergenz im Maß folgt unter unter stetigen Transformationen erhalten bleibt (ASAMOF könnten wir in der Formulierung des Satzes gleich statt der Konvergenz im Maß die schwächere Teilfolgenkonvergenz verwenden). Aus dieser folgt die Teilfolgenkonvergenz von $|f_n|^p$ und die von $|f_n - f|^p$. Die zweite Bedingung impliziert die erste, weil sie wegen des Lemmas von Fatou zu $\|f_n\|_p \to \|f\|_p$ äquivalent ist, und das nach Satz 5.10 zusammen mit der Teilfolgenkonvergenz zur gleichmäßigen Integrierbarkeit von $|f_n|^p$. Weil mit $|f_n|^p$ auch $|f_n - f|^p$ gleichmäßig integrierbar ist, implizieren Bedingung 1. und die Teilfolgenkonvergenz die Konvergenz der Integrale und damit die Konvergenz von f_n gegen f im p-ten Mittel. Das zeigt eine Richtung. Für die andere Richtung haben wir früher schon festgestellt, dass die Konvergenz im p-ten Mittel die Konvergenz im Maß impliziert, und $\|f_n\|_p \to \|f\|_p$ folgt ebenfalls daraus.

Ein wichtiger Begriff ist der *Dualraum* eines Banachraums:

Satz 8.6

Der Dualraum $\mathbf{B}^*$ des Banachraums $\mathbf{B}$ ist der Raum aller stetigen linearen Funktionale auf $\mathbf{B}$ (also aller stetigen linearen Abbildungen ℓ von $\mathbf{B}$ nach $\mathbb{R}$) mit der Norm

$$\|\ell\|^* = \sup\{|\ell(f)| : f \in \mathbf{B}, \|f\| \leq 1\}.$$

Für die L_p-Räume können die Dualräume schön beschrieben werden:

Satz 8.7: Darstellungssatz von Riesz

Für $1 < p < \infty$ gibt es für jedes lineare Funktional ℓ auf L_p ein Element $g \in L_q$ $(1/p+1/q = 1)$ mit $\ell(f) = \int gf$ und $\|\ell\|_p^* = \|g\|_q$. . Wenn μ sigmaendlich ist, dann gilt das auch für $p = 1$ (mit $q = \infty$). Es ist also L_p^* isometrisch zu L_q.

Es folgt direkt aus der Ungleichung von Hölder, dass für jedes $g \in L_q$ durch $\ell(f) = \int fg$ ein beschränktes lineares Funktional definiert wird, und dass $\|\ell\| = \|g\|_q$ gilt. Es kann also jedenfalls L_q isometrisch in L_p^* eingebettet werden.

Die Umkehrung, also dass jedes stetige lineare Funktional als Integral dargestellt werden kann, zeigen wir zunächst für endliche Maße. Daraus ergibt sich mit dem üblichen Zerlegungsargument der Satz für sigmaendliche Maße. Dass man für $0 < p < \infty$ auch auf die Sigmaendlichkeit verzichten kann, liegt daran, dass für $f \in L_p$ die Menge $[f \neq 0]$ sigmaendlich ist; wir werden auf die Details verzichten.

Es sei also μ endlich und ℓ ein stetiges lineares Funktional auf L_p. Für $A \in \mathfrak{S}$ setzen wir (wegen der Endlichkeit von μ ist $A() \in L_p$)

$$\nu(A) = \ell(A()).$$

Wir zeigen, dass ν ein signiertes Maß ist, das bezüglich μ absolutstetig ist. Der zweite Teil sollte klar sein, weil jeder Indikator einer Nullmenge zu 0 äquivalent ist und daher durch ℓ auf 0 abgebildet wird. Es bleibt die Sigmaadditivität von ν zu zeigen.

Die Additivität von ν ist offensichtlich. Für die Sigmaadditivität genügt also die Stetigkeit bei $\emptyset$ zu zeigen. Es sei also $A_n \in \mathfrak{S}$ mit $A_n \downarrow \emptyset$. Dann geht $\mu(A_n)$ gegen 0 und

$$|\nu(A_n)| \leq \|\ell\|\|A_n()\|_p = \|\ell\|\mu(A_n)^{1/p} \to 0.$$

Es ist also ν tatsächlich ein signiertes Maß und $\nu \ll \mu$. Daher gibt es die Radon-Nikodym Ableitung $g = \frac{d\nu}{d\mu}$. Deshalb gilt

$$\nu(f) = \int f g d\mu$$

für Indikatorfunktionen. Für Treppenfunktionen ergibt sich die Gültigkeit dieser Darstellung aus der Linearität von ℓ und dem Integral, und schließlich kann eine allgemeine Funktion in L_p durch Treppenfunktionen approximiert werden.

8.1 Wiederholungsfragen

- Ungleichung von Jensen?
- Ungleichung von Hölder?
- Ungleichung von Minkowski?
- Wie sind die p-Norm und die Räume $\mathcal{L}_p$ und L_p definiert?
- Was kann man über die Struktur von L_p sagen?
- Geben Sie Kriterien für die L_p-Konvergenz an.
- Wie ist der Dualraum eines Banachraums definiert?
- Wie kann man den Dualraum von L_p charakterisieren?

Kapitel 9

Produkträume

Multiplication
is the name of the game.

Bobby Darin

9.1 Das Produkt von zwei sigmaendlichen Maßräumen

In diesem Abschnitt vollziehen wir — natürlich in verallgemeinerter Form — den Schritt nach, der in der Elementargeometrie von der Längen- zur Flächenmessung führt. Dazu erinnern wir uns daran, dass die ersten Figuren, für die die Flächenmessung möglich war, die Rechtecke waren, und dass wir ihre Fläche damals als das Produkt der Seitenlängen erhielten. Dieses Modell wird uns auch hier leiten, nur müssen die "Seiten" unserer "Rechtecke" nicht unbedingt Intervalle sein, und statt der Längen können auf den Seiten natürlich beliebige Maße (in Grenzen) auftreten.

Wir zeigen zunächst

Satz 9.1

$(\Omega_1, \mathfrak{S}_1, \mu_1)$ und $(\Omega_2, \mathfrak{S}_2, \mu_2)$ seien zwei Maßräume (nicht notwendig sigmaendlich). Auf dem Semiring

$$\mathfrak{S}_1 \otimes \mathfrak{S}_2 = \{A_1 \times A_2 : A_1 \in \mathfrak{S}_1, A_2 \in \mathfrak{S}_2\}$$

ist durch

$$\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

ein Maß definiert. Wenn μ_1 und μ_2 sigmaendlich sind, dann auch $\mu_1 \times \mu_2$.

Es ist leicht zu sehen, dass $\mu_1 \times \mu_2$ auf $\mathfrak{S}_1 \otimes \mathfrak{S}_2$ additiv ist. Für den Nachweis der Sigmaadditivität verwenden wir einen Trick, der schon den Weg zu dem weist, was später kommt: es sei also

$$A = B \times C$$

HIER Wir definieren:

Definition 9.1

$(\Omega_1, \mathfrak{S}_1, \mu_1)$ und $(\Omega_2, \mathfrak{S}_2)$ seien zwei sigmaendliche Maßräume. Die Sigmaalgebra

$$\mathfrak{S}_1 \times \mathfrak{S}_2 = \mathfrak{A}_\sigma(\{A \times B : A \in \mathfrak{S}_1, B \in \mathfrak{S}_2\})$$

über der Menge $\Omega_1 \times \Omega_2$ heißt das *Produkt* der beiden Sigmaalgebren $\mathfrak{S}_1$ und $\mathfrak{S}_2$. Auf dieser Produktalgebra ist durch $\mu_1 \times \mu_2(A \times B) = \mu_1(A)\mu_2(B)$ ein sigmaendliches Maß definiert, das Produkt der beiden Maße μ_1 und μ_2.

Wir wissen bereits, dass die Menge

$$\mathfrak{T} = \{A \times B : A \in \mathfrak{S}_1, B \in \mathfrak{S}_2\}$$

ein Semiring ist, daher ist das Produktmaß durch die obige Definition auf der gesamten Produktalgebra eindeutig festgelegt (da es auf dem Semiring $\mathfrak{T}$ definiert und dort sigmaendlich ist). Was noch offen ist, ist die Verifizierung der Tatsache, dass $\mu_1 \times \mu_2$ auf T ein Maß ist. Das könnte man etwa tun, indem man auf dem erzeugten Ring die Stetigkeit von oben bei der leeren Menge zeigt. Wir gehen allerdings hier einen anderen Weg (indem wir uns an die Flächenberechnung in der Integralrechnung erinnern).

Dazu definieren wir

Definition 9.2

Für eine Menge $A \subseteq \Omega_1 \times \Omega_2$ bzw. eine Funktion $f : \Omega_1 \times \Omega_2 \to X$ nennen wir

$$A(\omega_1, .) = \{\omega_2 : (\omega_1, \omega_2) \in A\},$$

$$A(., \omega_2) = \{\omega_1 : (\omega_1, \omega_2) \in A\},$$

$$f(\omega_1, .) : \Omega_2 \to x, \omega_2 \mapsto f(\omega_1, \omega_2)$$

und

$$f(., \omega_2) : \Omega_2 \to x, \omega_2 \mapsto f(\omega_1, \omega_2)$$

die Schnitte von A bzw. f (in Richtung 2 bzw. 1 und bei ω_1 bzw. ω_2).

.

Die Schreibweise für die Funktionen ist relativ geläufig, wir verwenden sie in derselben Form für Mengen (weil wir sie ja auch als ihre eigenen Indikatorfunktionen lesen können).

Satz 9.2

Für jedes $A \in \mathfrak{S}_1 \times \mathfrak{S}_2$ gilt

$$\mu_1 \times \mu_2(A) = \int \mu_2(A(\omega_1, .))d\mu_1(\omega_1) = \int \mu_1(A(., \omega_2))d\mu_2(\omega_2).$$

In diesem Satz sind einige implizite Behauptungen enthalten, nämlich

1. $A(\omega_1, .) \in \mathfrak{S}_2$,
2. $\omega_1 \mapsto \mu_2(A(\omega_1, .))$ ist $\mathfrak{S}_1$-messbar, und schließlich
3. durch
$$\mu_1 \times \mu_2(A) = \int \mu_2(A(\omega_1, .))d\mu_1(\omega_1)$$
ist ein Maß auf $\mathfrak{S}_1 \times \mathfrak{S}_2$ definiert (und die analogen Behauptungen für $A(., \omega_2)$).

Wir führen den Beweis für den Fall, dass die Maße μ_1 und μ_2 endlich sind (der allgemeine Fall lässt sich darauf zurückführen, indem man $\Omega_1 \times \Omega_2$ in abzählbar viele Rechtecke mit endlichem Maß zerlegt und jedes dieser Rechtecke für sich betrachtet).

Zuerst beweisen wir die Messbarkeit der Schnitte. Dazu setzen wir

$$\mathfrak{A} = \{A \in \mathfrak{S}_1 \times \mathfrak{S}_2 : A(\omega_1, .) \in \mathfrak{S}_2 \text{ für alle } \omega_1 \text{ und } \omega_1 \mapsto \mu_2(A(\omega_1, .)) \text{ ist messbar}\}.$$

Die Menge $\mathfrak{A}$ enthält offensichtlich alle Rechtecke und alle disjunkten Vereinigungen von Rechtecken (also den von $\mathfrak{T}$ erzeugten Ring) und ist ein monotones System (beim Beweis, dass $\mathfrak{A}$ bezüglich der Grenzwerte von fallenden Mengenfolgen abgeschlossen ist, brauchen wir die Endlichkeit des Maßes).

Nach dem monotone class theorem enthält $\mathfrak{A}$ also den von $\mathfrak{T}$ erzeugten Sigmaring, und der ist genau die Produktsigmaalgebra. Also sind für alle Mengen in der Produktsigmaalgebra die Schnitte

messbar, und die Maße der Schnitte sind messbare Funktionen. Damit ist einmal gezeigt, dass die Integrale im obigen Satz definiert sind.

Von hier aus ist aber nicht mehr viel zu tun: das Integral

$$\int \mu_2(A(\omega_1,.))d\mu_1(\omega_1)$$

ist (als Funktion von A) offensichtlich additiv, wegen des Satzes von der monotonen Konvergenz auch sigmaadditiv, und falls $A = B \times C$ ein Rechteck ist, dann ist

$$\int \mu_2(A(\omega_1,.))d\mu_1(\omega_1) = \int_B \mu_2(C)d\mu_1(\omega_1) = \mu_2(C)\mu_1(A),$$

und daher stimmt das Integral mit dem Produktmaß überein (bzw. ist gezeigt, dass durch unsere Definition des Produktmaßes tatsächlich ein Maß festgelegt wird).

Nun wollen wir uns der Berechnung des Integrals bezüglich des Produktmaßes zuwenden. Dazu kann man natürlich wieder die Definition des Integrals nachvollziehen, aber wir hätten gern einen einfacheren Weg. Schön wäre es, wenn wir, wie es auch in der Analysis funktioniert, das Produktintegral als iteriertes Integral berechnen könnten, also wenn

$$\int f(\omega)\mu_1 \times \mu_2(d\omega) = \int (\int f(.,\omega_2)d\mu_1)\mu_2(d\omega_2)$$

gilt. Bedingungen dafür gibt der

Satz 9.3: FUBINI

$(\Omega_1, \mathfrak{S}_1, \mu_1)$ und $(\Omega_2, \mathfrak{S}_2, \mu_2)$ seien zwei sigmaendliche Maßräume. Dann gilt

1. Falls $f \geq 0$ bezüglich $\mathfrak{S}_1 \times \mathfrak{S}_2$ messbar ist, dann ist
$$\int f(\omega)\mu_1 \times \mu_2(d\omega) = \int (\int f(.,\omega_2)d\mu_1)\mu_2(d\omega_2).$$
2. Falls f bezüglich $\mu_1 \times \mu_2$ integrierbar ist, dann
   (a) ist $f(.,\omega_2)$ für fast alle $\omega_2 \in \Omega_2$ bezüglich μ_1 integrierbar,
   (b) ist die Funktion $h(\omega_2) = \int f(.,\omega_2)d\mu_1$ integrierbar,
   (c) ist
   $$\int f(\omega)\mu_1 \times \mu_2(d\omega) = \int (\int f(.,\omega_2)d\mu_1)\mu_2(d\omega_2).$$
3. f bezüglich $\mu_1 \times \mu_2$ integrierbar genau dann, wenn
$$\int (\int |f(\omega_1,\omega_2)|d\mu_1(\omega_1))d\mu_2(\omega_2) < \infty.$$

Für den Beweis des ersten Teils brauchen wir nur die Definition des Integrals nachzuvollziehen: für eine Indikatorfunktion ist die Behauptung genau der Inhalt des vorigen Satzes, daraus erhalten wir wegen der Linearität des Integrals, dass die Behauptung auch für Treppenfunktionen gilt, und im allgemeinen Fall approximieren wir f von unten durch eine Folge von nichtnegativen Treppenfunktionen und verwenden den Satz von der monotonen Konvergenz.

Der zweite Teil folgt durch Zerlegen von f in Positiv- und Negativteil. Auf diese ist Teil 1 anwendbar, Probleme können nur auftreten, wenn die inneren Integrale unendlich werden, aber das passiert (wieder als Folgerung aus Teil 1) nur auf einer Menge vom Maß null.

Der dritte Teil ist einfach die Anwendung des ersten Teils auf das Integral $\int |f| d(\mu_1 \times \mu_2)$.

Als Beispiel für die Anwendung dieses Satzes berechnen wir das (uneigentliche Riemann-) Integral

$$I = \int_0^\infty \frac{\sin(x)}{x} dx.$$

Dieses Integral kann natürlich nicht als Lebesgue-Integral verstanden werden, weil es nur bedingt konvergiert. Es hält uns aber nichts davon ab, den impliziten Grenzübergang des Riemann-Integrals im Lebesgue-Integral nachzubilden:

$$I = \lim_{M\to\infty} I_M$$

mit

$$I_M = \int_{[0,M]} \frac{\sin(x)}{x}\lambda(dx).$$

Den Faktor $1/x$ im Integranden schreiben wir als Integral:

$$\frac{1}{x} = \int_{[0,\infty)} e^{-tx}\lambda(dt)$$

und damit

$$I_M = \int_{[0,M]}\int_{[0,\infty)} e^{-tx}\sin(x)\lambda(dt)\lambda(dx).$$

Mit der Abschätzung $|\sin(x)| \leq x$ erkennt man leicht, dass das Integral des Absolutbetrags endlich bleibt (nämlich $\leq M$), also kann der Satz von Fubini angewendet und das iterierte Integral sowohl als Integral über das Produktmaß als auch als iteriertes Integral mit vertauschter Integrationsreihenfolge geschrieben werden. Letzteres ist, was wir wollen, und wir erhalten

$$I_M = \int_{[0,\infty)} \frac{1 - e^{-tM}(\cos(M) + t\sin(M))}{1+t^2}\lambda(dt).$$

Mit den Ungleichungen $|\cos(M)| \leq 1$ und $|\sin(M| \leq M$ kann man einen Teil abschätzen:

$$|\int_{[0,\infty)} e^{-tM}\frac{\cos(M) + t\sin(M)}{1+t^2}\lambda(dt)| \leq \frac{2}{M}.$$

Das geht gegen 0, wenn M über alle Grenzen wächst, und daher ist

$$I = \frac{\pi}{2}.$$

Als zweite Anwendung werden wir die Notwendigkeit im Dreireihensatz zeigen. Dazu beweisen wir zuerst eine untere Abschätzung in der Ungleichung von Kolmogorov:

Satz 9.4: Ungleichung von Kolmogorov 2

$X_1, \ldots, X_n$ seien unabhängige Zufallsvariable mit $|X_i| \leq c < \infty$, $\mathbb{E}(X_i) = 0$ und $\mathbb{V}(X_i) = \sigma_i^2$. Weiters sei $S_k = \sum_{i=1}^k X_i$ und $\lambda > 0$. Dann gilt

$$\mathbb{P}(\max_{k\leq n}|S_k| \geq \lambda) \geq 1 - \frac{(\lambda + c)^2}{\mathbb{V}(S_n)}.$$

Zum Beweis sei

$$A_i = [|S_j| < \lambda, j = 1, \ldots, i],$$
$$B_i = [|S_j| < \lambda, j = 1, \ldots, i-1, |S_i| \geq \lambda.$$

Damit ist

$$p = \mathbb{P}(\max_{k\leq n}|S_k| \geq \lambda) = \sum_{i=1}^n \mathbb{P}(B_i) = 1 - \mathbb{P}(A_n).$$

Wir stellen fest, dass

$$A_{i-1}S_i = A_iS_i + B_iS_i$$

und

$$A_{i-1}S_i = A_{i-1}S_{i-1} + A_{i-1}X_i.$$

Wir setzen also die rechten Seiten gleich, dann quadrieren wir und nehmen Erwartungswerte. Auf beiden Seiten fallen dabei die gemischten Glieder weg, einerseits weil $A_iB_i = 0$, auf der anderen, weil X_i von den anderen Faktoren unabhängig ist.

Wir summieren von 1 bis n, eliminieren identische Summanden auf beiden Seiten und erhalten

$$\sum_{i=1}^{n} \mathbb{P}(A_{i-1})\sigma_i^2 = \mathbb{E}(A_n S_n^2) + \sum_{i=1}^{n} \mathbb{E}(B_i S_i^2).$$

Auf der linken Seite gilt

$$\mathbb{P}(A_i) \geq \mathbb{P}(A_n) = 1 - p,$$

auf der rechten Seite haben wir

$$A_n S_n^2 \leq A_n \lambda^2 \leq A_n(\lambda + c)^2$$

und

$$B_i S_i^2 \leq B_i(\lambda + c)^2$$

(weil ja, wenn B_i zutrifft, $|S_i| \leq |S_{i-1}| + |X_i| \leq \lambda + c$). Das ergibt

$$(1 - p)\sum_{i=1}^{n} \sigma_i^2 \leq (\lambda + c)^2,$$

und nach trivialen Umformungen die gewünschte Ungleichung.

Zum Beweis der Notwendigkeit der Dreireihenbedingung zeigen wir zuerst, dass die erste Reihe konvergieren muss. Wäre nämlich

$$\sum_n \mathbb{P}(|X_i| > c) = \infty,$$

dann wäre nach dem zweiten Lemma von Borel-Cantelli mit Wahrscheinlichkeit 1 unendlich oft $|X_n| > c$, und damit kann die Reihe $\sum X_n$ nicht konvergieren. Wenn nun also die erste Reihe konvergiert, dann stimmen mit Wahrscheinlichkeit 1 die Zufallsvariablen X_n und

$$Y_n = [|X_n| < c]X_n$$

für fast alle n überein, und $\sum X_n$ konvergiert genau dann, wenn $\sum Y_n$ konvergiert. Wir müssen also die Aussage des Satzes nur noch für beschränkte Folgen (X_n) zeigen. Wir haben es immer noch mit zwei Reihen zu tun (der der Erwartungswerte und der der Varianzen), und es wäre schön, eine der beiden eliminieren zu können. Dazu hilft uns die Theorie der Produkträume. Statt des ursprünglichen Wahrscheinlichketsraums $(\Omega, \mathfrak{S}, \mathbb{P})$ betrachten wir $\Omega \times \Omega, \mathfrak{S} \times \mathfrak{S}, \mathbb{P} \times \mathbb{P})$ und darauf die Folgen $(X_n(\omega_1))$ und $(X_n(\omega_2))$. Das sind zwei unabhängige Kopien der ursprünglichen Folge, und daher ist

$$Z_n = X_n(\omega_1) - X_n(\omega_2)$$

eine Folge von unabhängigen Zufallsvariablen mit $|Z_n| \leq 2c$, $\mathbb{E}(Z_n) = 0$ und $\mathbb{V}(Z_n) = 2\mathbb{V}(X_n)$. wäre nun

$$\sum \mathbb{V}(Z_n) = \infty,$$

dann wäre für jedes $\lambda > 0$.

$$\mathbb{P}(\sup_{n\in\mathbb{N}} |\sum_{i=1}^{n} Z_i| \geq \lambda) = \lim_{k\to\infty} \mathbb{P}(\max_{n\leq k} |\sum_{i=1}^{n} Z_i| \geq \lambda) \geq \lim_{k\to\infty} (1 - \frac{(\lambda + 2c)^2}{\sum_{i=1}^{k} \mathbb{V}(Z_i)}) = 1.$$

Die Summe $\sum Z_n$ kann also gegen keinen endlichen Wert konvergieren.

Damit ist gezeigt, dass die dritte Reihe aus dem Dreireihensatz konvergieren muss, wenn die Reihe $\sum X_n$ konvergiert. Das impliziert aber die Konvergenz der Reihe

$$\sum (Y_n - \mathbb{E}(Y_n)),$$

und weil ja auch

$$\sum Y_n$$

konvergieren soll, konvergiert auch die zweite Reihe

$$\sum \mathbb{E}(Y_n).$$

9.2 Markovkerne

Etwas allgemeiner wird die Theorie der Produkträume, wenn das Maß auf dem zweiten Faktorraum nicht fix gewählt ist, sondern von ω_1 abhängen kann. Das führt uns zum Begriff eines Markovkerns, den wir schon im Zusammenhang mit bedingten Wahrscheinlichkeiten eingeführt haben.

Wir erinnern an Definition 7.3:

$(\Omega_i, \mathfrak{S}_i)$ $(i = 1, 2)$ seien zwei Messräume. Die Funktion $p : \Omega_1 \times \mathfrak{S}_2 \to [0, \infty[$ heißt *Markovkern* oder *Übergangsfunktion* von $(\Omega_1, \mathfrak{S}_1)$ nach $(\Omega_2, \mathfrak{S}_2)$, wenn sie folgende Bedingungen erfüllt.

1. Für jedes $\omega \in \Omega_1$ ist $p(\omega, .)$ ein (endliches) Maß.
2. Für jedes $A \in \mathfrak{S}_2$ ist $p(., A)$ $i\mathfrak{S}_1$-messbar.

Anmerkung: Für uns ist in der Folge der Fall am interessantesten, dass die Maße $p(\omega_1, .)$ Wahrscheinlichkeitsmaße sind. Die Endlichkeit könnte abgeschwächt werden, aber einfach nur zu verlangen, dass $p(\omega, .)$ ist zu wenig, etwa, wenn μp sigmaendlich sein soll. Dazu muss eine Zerlegung von Ω_2 in abzählbar viele Mengen B_n existieren, sodass jedes für alle n und ω_1 $p(\omega_1, B_n) < \infty$ gilt. Diese Eigenschaft wird auch als die "gleichmäßige Sigmaendlichkeit" bezeichnet.

Fast wörtlich genauso wie die Existenz des Produktmaßes zeigt man

Satz 9.5

$(\Omega_i, \mathfrak{S}_i)$ $(i = 1, 2)$ seien zwei Messräume, μ ein Maß auf $\mathfrak{S}_1$ und p ein Markovkern von $(\Omega_1, \mathfrak{S}_1)$ nach $(\Omega_2, \mathfrak{S}_2)$. Dann ist durch

$$\mu p(A) = \int p(\omega_1, A(\omega_1, .))\mu(d\omega_1)$$

ein Maß auf $(\Omega_1 \times \Omega_2, \mathfrak{S}_1 \times \mathfrak{S}_2)$ definiert.

Ebenso fast wörtlich ergibt sich

Satz 9.6

Satz von Fubini für Markovkerne f sei $\mathfrak{S}_1 \times \mathfrak{S}_2$-messbar.

1. Falls $f \geq 0$, dann ist
$$\int f(x, y)d\mu p(x, y) = \int (\int f(x, y)p(x, dy))\mu(dx).$$
2. Falls $f \in \mathcal{L}_1(\mu p)$, dann ist
$$\int f(x, y)d\mu p(x, y) = \int (\int f(x, y)p(x, dy))\mu(dx).$$
3. Falls
$$\int (\int |f|(x, y)p(x, dy))\mu(dx),$$
dann $f \in \mathcal{L}_1(\mu p)$.

Wir können nun die Frage stellen, ob ein Maß auf dem Produktraum in dieser Form darstellbar ist. Der Einfachheit halber werden wir nur endliche Maße betrachten, speziell Wahrscheinlichkeitsmaße. Wenn wir auch annehmen, dass die Maße $p(\omega, .)$ Wahrscheinlichkeitsmaße sind, dann ist

$$\mu(A) = \mu p(A \times \Omega_2).$$

Wenn wir die Projektionen

$$\pi_i : \Omega_1 \times \Omega_2 \to \Omega_i, \pi_i(\omega) = \omega_i$$

einführen und von einem Wahrscheinlichkeitsmaß $\mathbb{P}$ auf dem Produktraum ausgehen, dann ist das gesuchte μ genau die induzierte Verteilung von π_1:

$$\mu(A) = \mathbb{P}(A \times \Omega_2) = \mathbb{P}\pi_1^{-1}(A).$$

Dieses Maß heißt die Randverteilung von π_1:

Definition 9.3

$(\Omega_i, \mathfrak{S}_i)$ seien Messräume, $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\bigtimes_{i=1}^n \Omega_i, \bigtimes_{i=1}^n \mathfrak{S}_i)$. Dann heißen die Maße $\mathbb{P}\pi_i^{-1}$ die (eindimensionalen) Randverteilungen von $\mathbb{P}$.

Allgemeiner kann man für nichtleeres $J \subset \{1, \ldots, n\}$ die Projektion $\pi_J : \omega \to (\omega_i, i \in J)$ definieren und damit die (mehrdimensionalen, wenn $|J| > 1$) Randverteilungen $P\pi_J^{-1}$.

Wenn π_1 als Zufallsvariable angesehen wird, dann ist leicht zu sehen, dass

$$p(\omega_1, A) = \mathbb{P}(\Omega_1 \times A | \pi_1)$$

sein muss. Die Frage, ob p ein Markovkern sein kann, führt uns also zurück zur Frage nach der Existenz von regulären bedingten Wahrscheinlichkeiten. Wir haben am Ende von Kapitel 7 festgestellt, dass es diese nicht immer gibt, aber dass sie existieren, wenn die beiden Faktoren jeweils $\mathbb{R}$ (oder $\mathbb{R}^n$) mit den Boremengen sind.

9.3 Unendliche Produkträume

Mit den Methoden aus dem vorigen Abschnitt kann man natürlich auch (indem man den Produktraum wieder mit einem weiteren Faktor multipliziert) beliebige endliche Produkte behandeln. Interessanter wird es, wenn wir unendliche Produkte betrachten. Es sei also zunächst $((\Omega_i, \mathfrak{S}_i), i \in I)$ eine Familie von Messräumen. Als erstes müssen wir über dem Produkt

$$\bigtimes_{i \in I} \Omega_i = \{f : I \to \bigcup_{i \in I} \Omega_i : f(i) \in \Omega_i\}$$

eine geeignete Sigmaalgebra angeben. Das mindeste, das wir von einer solchen Sigmaalgebra verlangen müssen, ist, dass die Projektionen auf die einzelnen Koordinaten messbar sind. Wir definieren also:

Definition 9.4

$((\Omega_i, \mathfrak{S}_i), i \in I)$ sei eine Familie von Messräumen. Das Produkt $\bigtimes_{i \in I} \mathfrak{S}_i$ ist die kleinste Sigmaalgebra, bezüglich der alle Projektionen π_i ($\pi_i : \bigtimes_{i \in I} \Omega_i \to \Omega_i,\ \pi_i(f) = f(i)$) messbar sind.

Diese Sigmaalgebra wird von den Mengen $\pi_i^{-1}(A)$ mit $i \in I$ und $A \in \mathfrak{S}_i$ erzeugt. Eine solche Menge lässt sich in der Form $\bigtimes_{j \in I} B_j$ darstellen, wobei $B_i = A$ und $B_j = \Omega_j$ für $j \neq i$ gilt.

Ein etwas interessanteres Erzeugendensystem ist die Menge aller endlichen Durchschnitte von solchen Mengen, die einen Semiring bildet. Diese Mengen haben die Form $\bigtimes_{i \in I} B_i$ mit $B_i \in \mathfrak{S}_i$ und $B_i = \Omega_i$ für alle $i \in I$ bis auf endlich viele, und sollen "Pfeiler" heißen.

Man kann für jede Teilmenge J von i das Produkt $\bigtimes_{i \in J} \mathfrak{S}_i$ in $\bigtimes_{i \in I} \mathfrak{S}_i$ einbetten, indem man die Menge $A \in \bigtimes_{i \in J} \mathfrak{S}_i$ mit $A \times \bigtimes_{i \in I \setminus J} \Omega_i$ identifiziert (wobei der letzte Ausdruck als die Menge aller $f \in \bigtimes_{i \in I} \Omega_i$ zu verstehen ist, deren Einschränkung auf J in A liegt; diese Menge heißt Zylinder mit Basis A. Mit dieser Einschränkung ist das System $\mathfrak{Z}$ der Zylindermengen gleich der Vereinigung aller endlichen Produkte:

$$\mathfrak{Z} = \bigcup_{J \subseteq I, |J| < \infty} \bigtimes_{i \in J} \mathfrak{S}_i.$$

$\mathfrak{Z}$ ist also als Vereinigung einer gerichteten Menge von Ringen ein Ring, und die Pfeilermengen bilden einen Semiring, der $\mathfrak{Z}$ erzeugt.

Um ein Maß auf dieser Produktsigmaalgebra zu definieren, ist es notwendig, seinen Wert für alle Zylindermengen anzugeben. Wir werden uns auf endliche Maße beschränken — genauer gesagt, auf Wahrscheinlichkeitsmaße — und weil die Zylindermengen einen Ring bilden, ist damit (nach dem Fortsetzungssatz) das Maß eindeutig bestimmt.

Es ist also (wieder modulo Fortsetzungssatz) auf jedem endlichen Produkt $\mathfrak{S}_{i_1} \times \mathfrak{S}_{i_2} \times \ldots \times \mathfrak{S}_{i_n}$ ein Maß $\mu_{i1,i2,\ldots,i_n}$ anzugeben. Diese Maße werden gemeinhin (besonders in der Wahrscheinlichkeitstheorie) die endlichdimensionalen Randverteilungen genannt. Es erhebt sich die Frage, welche Bedingungen an eine solche Familie von endlichdimensionalen Verteilungen zu stellen sind, damit sie ein Maß auf dem unendlichen Produkt festlegen. Dazu müssen sie auf dem Ring

$$\mathfrak{R} = \bigcup_{J \subseteq I, |J| < \infty} \bigtimes_{i \in J} \mathfrak{S}_i,$$

der $\bigtimes_{i \in I} \mathfrak{S}_i$ erzeugt, ein Maß bilden, also sigmaadditiv sein, und μ muss durch die Forderung

$$\mu(A) = \mu_J(A) \text{ falls } A \in \bigtimes_{i \in J} \mathfrak{S}_i,\ |J| < \infty$$

wohlbestimmt sein. Die letzte Forderung bedeutet, dass für ein $A \in \mathfrak{S}_{i_1} \times \ldots \times \mathfrak{S}_{i_n}$, das man ja auch (wenn wir die endlichen Produkte mit ihrer Einbettung im unendlichen Produkt identifizieren, was wir in diesem Abschnitt ohne weitere Warnungen tun werden) als $A \times \Omega_{i_{n+1}}$ auffassen kann, gelten muss, dass

$$\mu_{i_1,\ldots,i_n,i_{n+1}}(A \times \Omega_{i_{n+1}}) = \mu_{i_1,\ldots,i_n}(A).$$

Eine Familie von endlichdimensionalen Maßen, die diese Forderung erfüllt, heißt konsistent. Es wäre schön zu berichten, dass jede konsistente Familie auf einem beliebigen Produkt von Messräumen ein Maß bestimmt, aber leider ist dem nicht so, man muss zumindest ein bisschen an zusätzlicher Struktur auf den einzelnen Faktorräumen verlangen oder aber Voraussetzungen über die Maße selbst treffen. Unser erster Satz fällt in die zweite Kategorie und stellt die direkte Verallgemeinerung des endlichen Produkts von Maßräumen dar:

Satz 9.7: Existenzsatz von Andersen-Jessen

$((\Omega_i, \mathfrak{S}_i, \mu_i), i \in I)$ sei eine Familie von Maßräumen. Durch $\mu_{i_1,\ldots,i_n} = \mu_{i_1} \times \ldots \times \mu_{i_n}$ wird auf $\bigtimes_{i \in I} \mathfrak{S}_i$ ein Maß bestimmt, das wir das Produkt der Maße μ_i nennen und mit $\bigtimes_{i \in I} \mu_i$.

Was wir zeigen müssen ist, dass das so definierte μ auf dem Ring

$$\mathfrak{R} = \bigcup_{J \subseteq I, |J| < \infty} \bigtimes_{i \in J} \mathfrak{S}_i$$

ein Maß ist (die Konsistenz dieser Familie ist offensichtlich). Die Additivität und Nichtnegativität von μ sind klar, und für die Sigmaadditivität müssen wir nur zeigen, dass μ stetig von oben bei $\emptyset$ ist, d.h., dass für eine monoton nichtwachsende Folge $A_n \in \mathfrak{R}$ mit $\bigcap A_n = \emptyset$ gilt, dass $\mu(A_n)$ gegen 0 konvergiert.

Wir führen den Beweis indirekt und zeigen, dass für eine monoton nichtwachsende Mengenfolge $A_n \in \mathfrak{R}$ mit $m = \lim \mu(A_n) > 0$ der Durchschnitt nicht leer ist.

Da $A_n \in \mathfrak{R}$, gibt es eine endliche Teilmenge $J_n \in I$, sodass $A_n \in \bigtimes_{i \in J_n} \mathfrak{S}_i$. $J = \bigcup J_n$ ist eine abzählbare Menge und $A_n \in \bigtimes_{i \in J} \mathfrak{S}_i$. Wir können also ohne Beschränkung der Allgemeinheit annehmen, dass I abzählbar ist, etwa $I = \mathbb{N}$. Ebenfalls ohne Beschränkung der Allgemeinheit können wir annehmen, dass $J_n = \{1, \ldots, n\}$ ist.

Wir bezeichnen mit ν_k das Produkt der Maße μ_i mit $i > k$ (mit dieser Bezeichnung ist $\mu = \nu_0$). Von diesen Funktionen wissen wir zwar noch nicht, dass Sie Maße sind, aber ihre Einschränkungen auf eine endliche Produktalgebra sind Maße.

Wir bezeichnen jetzt

$$A_n(x) = \{f \in A_n : f(1) = x\}.$$

Mit dieser Notation gilt

$$\mu(A_n) = \int \nu_1(A_n(x)) d\mu_1(x).$$

Jetzt sei B_n die Menge aller $x \in \Omega_1$ mit $\nu_1(A_n(x)) > m/2$). Im obigen Integral gilt:

$$\mu_1(B_n) \geq \int_{B_n} \nu_1(A_n(x)) d\mu_1(x) = \mu(A_n) - \int_{B_n^C} \nu_1(A_n(x)) \geq m - \frac{m}{2}\mu(B_n^C) \geq m/2.$$

B_n ist eine nichtwachsende Folge von Mengen. Daher ist

$$\mu_1(\bigcap B_n) \lim \mu_1(B_n) \geq m/2.$$

Wir können also ein $x = x_1$ finden, sodass $\nu_1(A_n(x)) > m/2$ ist. Das heißt insbesondere, dass jedes A_n ein Element f besitzt, für das $f(1) = x_1$ ist.

Wenn wir jetzt in diesem Argument A_n durch $A_n(x_1)$ ersetzen, erhalten wir ein x_2, sodass jedes $A_n(x_1)$ ein Element enthält, das mit x_2 beginnt. Das bedeutet aber, dass jedes A_n ein Element enthält, das mit (x_1, x_2) beginnt. Wenn wir dieses Verfahren ins Unendliche fortsetzen, erhalten wir eine Folge $x_1, x_2, \ldots$, sodass jedes A_n ein Element enthält, das mit $(x_1, dots, x_k)$ beginnt. Die Folge $(x_1, x_2, \ldots)$ ist aber in jedem A_n enthalten (denn aus $x \in A_n$ folgt, dass jedes y mit $y_i = x_i$ für alle $i \leq n$ auch in A_n liegt). Wir haben also eine Folge gefunden, die in allen A_n liegt, daher ist ihr Durchschnitt nicht leer, und der Satz ist bewiesen.

Für eine allgemeine Familie von endlichdimensionalen Randverteilungen können wir kein vollkommen allgemeines Ergebnis zeigen, aber es gibt den

Satz 9.8: Existenzsatz von Kolmogorov

Wenn $\Omega_i = \mathbb{R}$ und $\mathfrak{S}_i = \mathfrak{B}$, dann wird durch jede konsistente Familie von Maßen ein Maß auf dem Produkt definiert.

Der Beweis des vorigen Satzes kann fast wörtlich übernommen werden, wenn für ν_1 etc. eine reguläre Version von $\mathbb{P}(.|x_1)$ etc. eingesetzt wird.

Dieser Satz gilt natürlich auch für allgemeinere Räume (die Verallgemeinerung $\Omega_i = \mathbb{R}^k$ ist offensichtlich), aus unserer Version erhält man direkt die Gültigkeit für alle Borel'schen Räume, das sind solche, die sich isomorph (im Sinne der Maßtheorie, also mit einer bijektiven Abbildung, die samt ihrer inversen messbar ist) auf eine Borelsche Teilmenge von $\mathbb{R}$ abbilden lässt.

9.4 Wiederholungsfragen

1. Wie ist das Produkt von zwei Maßräumen definiert?
2. Was besagt der Satz von Fubini?
3. Was ist ein Markovkern?
4. Wie kann man mit einem Markovkern ein Maß auf dem Produktraum definieren?
5. Wie ist das Produkt einer Familie von Sigmaalgebren definiert?
6. Geben Sie einen Semiring bzw. Ring, der die Produktsigmaalgebra erzeugt.
7. Wie kann man ein (Wahrscheinichkeits-) Maß auf der Produktsigmaalgebra angeben (bzw. was ist eine konsistente Familie von Wahrscheinlichkeitsmaßen)?
8. Was besagt der Existenzsatz von Andersen-Jessen?
9. Was besagt der Existenzsatz von Kolmogorov?

Kapitel 10

Transformationssätze

In diesem Kapitel beschäftigen wir uns mit der Frage, wie die Verteilung einer Funktion einer Zufallsvariable bestimmt werden kann. Der Ausgangspunkt unserer Überlegungen ist der Transformationssatz aus Teil 1:

Satz 10.1

Transformationssatz für Integrale

$$\int g \circ X d\mathbb{P} = \int g d\mathbb{P}_X.$$

In der Folge werden wir uns mit einigen Spezialfällen auseinandersetzen. Zuerst nehmen wir an, dass X eindimensional ist mit der Verteilungsfunktion F, die wir zusätzlich als strikt monoton und stetig (also bijektiv von $\mathbb{R}$ nach $(0,1)$) annehmen. Dann existiert die Umkehrfunktion F^{-1}, und für $y \in (0,1)$

$$\mathbb{P}(F(X) \leq y) = \mathbb{P}(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y,$$

es ist also $F(X)$ Gleichverteilt auf $[0,1]$. Ist umgekehrt U gleichverteilt auf $[0,1]$, dann gilt (für $x \in \mathbb{R}$)

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

Diese Ergebnisse gelten auch in allgemeinerer Form, wenn wir eine geeignete Definition von F^{-1} haben:

Definition 10.1

F sei eine Verteilungsfunktion. Dann heißt

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}$$

die verallgemeinerte Inverse von F.

Satz 10.2

1. U sei gleichverteilt auf $[0,1]$, F eine Verteilungsfunktion (im engeren Sinn). Dann hat

   $$X = F^{-1}(U)$$

   Die Verteilungsfunktion F.

2. X sei nach der stetigen Verteilungsfunktion F verteilt. Dann ist $F(X)$ gleichverteilt auf $[0,1]$.

Aus der Definition der verallgemeinerten Inversen und der Rechtsstetigkeit einer Verteilungsfunktion folgt die Beziehung

$$x \geq F^{-1}(y) \Leftrightarrow F(x) \geq y.$$

Damit gilt

$$\mathbb{P}(F^{-1}(U) \geq x) = \mathbb{P}(U \geq F(x)) = 1 - F(x-0),$$

Daher ist $\mathbb{P}(F^{-1}(U) < x) = F(x-0)$. Die Verteilungsfunktion von $F^{-1}(U)$ stimmt also in ihren Stetigkeitspunkten mit F überein, und weil beide Verteilungsfunktionen rechtsstetig sind, stimmen Sie überall überein. Das beweist den ersten Teil des Satzes.

Für den zweiten Teil stellen wir fest, dass

$$\mathbb{P}(F(X) \geq y) = \mathbb{P}(X \geq F^{-1}(y)) = 1 - F(F^{-1}(y) - 0).$$

Weil F stetig ist, ist $F(x-0) = F(x)$ und $F(F^{-1}(y)) = y$ für alle x bzw. y. Insgesamt ergibt sich

$$\mathbb{P}(F(X) < y) = y,$$

und $F(X)$ ist tatsächlich gleichverteilt.

Etwas spannender ist der Fall, dass P absolutstetig und g differenzierbar ist:

Satz 10.3: Transformationssatz für Dichten

Ω sei eine offene Teilmenge von $\mathbb{R}^d$, μ ein absolutstetiges endliches Maß auf $(\Omega, \Omega \cap \mathfrak{B}_d)$ mit Dichte $f = \frac{d\mathbb{P}}{d\lambda_d}$, und $g : \Omega \to \mathbb{R}^d$ injektiv und stetig differenzierbar mit Funktionaldeterminante

$$\frac{\partial g}{\partial x} = \det(\mathcal{J}_g) \neq 0$$

auf Ω. Dabei ist

$$\mathcal{J}_g = (\frac{\partial g_i}{\partial x_j})_{d \times d}$$

die Jacobimatrix von g. Dann ist $\mathbb{P}g^{-1} \ll \lambda_d$ und

$$\frac{d\mathbb{P}g^{-1}}{d\lambda} = \begin{cases} f(g^{-1}(y))|\frac{\partial g^{-1}}{\partial y}(y)| = f(g^{-1}(y))\frac{1}{|\frac{\partial g}{\partial x}(g^{-1}(y))|} & \text{wenn } y \in g(\Omega) \\ 0 & \text{sonst.} \end{cases}$$

Wir benötigen das folgende Resultat aus der Analysis:

Satz 10.4

Wenn g in einer Umgebung von x stetig differenzierbar ist mit $\frac{\partial g}{\partial x}(x) \neq 0$, dann gibt es gibt es zu jeder Umgebung U von x, sodass die Einschränkung von g auf U injektiv und offen ist. Weiters ist die Inverse g^{-1} dieser Einschränkung stetig und in $g(x)$ differenzierbar mit $\mathcal{J}_{g^{-1}}(g(x)) = J_g^{-1}(x)$.

Für uns folgt daraus, dass die Funktion g auf Ω offen ist (d.h., die Bilder von offenen Mengen sind offen), insbesondere ist $g(\Omega)$ offen, und dass die Umkehrung g^{-1} auf $g(\Omega)$ stetig differenzierbar ist.

Wir zeigen zuerst

Satz 10.5

Unter denselben Voraussetzungen für g gilt für $A \in \mathfrak{B}$, $A \subseteq \Omega$

$$\lambda(g(A)) = \int_A |\frac{\partial g}{\partial x}| d\lambda.$$

Etwas ungenau lässt sich die Idee des Beweises für diesen Hilfssatz so beschreiben, dass wir A in kleine Würfelchen zerlegen; auf jedem dieser Würfelchen C (mit Mittelpunkt m) unterscheidet sich $g(x)$ nur wenig von $g(m) + \mathcal{J}_g(x)(x-m)$. Daher unterscheidet sich $\lambda(g(C))$ nur wenig von $|\frac{\partial g}{\partial x}(m)|\lambda(C)$, was sich wiederum nicht sehr von $\int_C |\frac{\partial g}{\partial x}|d\lambda$ unterscheidet. Diese Idee wollen wir streng machen.

Im Sinne unseres üblichen Tauschs von zwei Ungleichungen gegen eine Gleichung zeigen wir erst nur die Ungleichung

$$\lambda(g(A)) \leq \int_A |\frac{\partial g}{\partial x}|d\lambda.$$

Das Maß

$$\mu(A) = \lambda(g(A \cap \Omega))$$

ist sigmaendlich und daher von unten regulär. Da g stetig ist und λ regulär von oben (auf $g(\Omega)$) ist auch μ von oben regulär. Wegen der Sigmaendlichkeit genügt es, die Aussage des Satzes für den Fall zu beweisen, dass $\mu(A)$ endlich ist. Wir fixieren $\epsilon > 0$ und wählen eine kompakte Menge $K \subseteq A \cap \Omega$ mit $\mu(K) \geq \mu(A) - \epsilon$ und ein δ mit $0 < \delta < 1/2$, über das wir später noch verfügen werden. Zu jedem $x \in K$ können wir ein $h \in]0, \delta[$ finden, sodass $C(x) = \bigtimes_{i=1}^d]x_i - h, x_i + h] \subseteq \Omega$ und für $y \in C(x)$ die Ungleichungen

$$\|\mathcal{J}_g(y) - \mathcal{J}_g(x)\| \leq \delta \|\mathcal{J}_g(x)\|,$$

$$\|\mathcal{J}_g^{-1}(y) - \mathcal{J}_g^{-1}(x)\| \leq \delta \|\mathcal{J}_g^{-1}(x)\|$$

und

$$(1-\delta)|\frac{\partial g}{\partial x}(x)| \leq |\frac{\partial g}{\partial x}(y)| \leq (1+\delta)|\frac{\partial g}{\partial x}(x)|$$

gelten. Da die Vereinigung der offenen Mengen

$$C^*(x) = \bigtimes_{i=1}^d]x_i - h(x_i), x_i + h(x_i)[$$

die Menge K überdeckt, gibt es endlich viele x, sagen wir $x_1, \ldots, x_n$, sodass

$$K \subseteq \bigcup_{i=1}^n C(x_i)^*$$

gilt; diese Vereinigung können wir wiederum als disjunkte Vereinigung von endlich vielen Intervallen (Hyperquadern) $I_1, \ldots, I_N$ darstellen.

Wir betrachten nun eines dieser Intervalle

$$I = \bigtimes_{i=1}^d]y_i - h_i, y_i + h_i].$$

Wir können annehmen, dass für $1 \leq i, j \leq d$ stets $h_i \leq 2h_j$ gilt (notfalls können wir eine zu lange Seite halbieren) und dass I ganz in einem der Intervalle $C(x_i)$ enthalten ist. Dann gilt für $z \in I$

$$\|\mathcal{J}_g(z) - J_g(y)\| \leq \|\mathcal{J}_g(z) - \mathcal{J}_g(x_i)\| + \|\mathcal{J}_g(y) - \mathcal{J}_g(x_i)\| \leq 2\delta \|\mathcal{J}_g(x_i)\|,$$

und weil

$$\|\mathcal{J}_g(y)\| \geq \|\mathcal{J}_g(x_1)\| - \delta\|\mathcal{J}_g(x_i)\| \geq \frac{1}{2}\|\mathcal{J}_g(x_i)\|,$$

$$\|\mathcal{J}_g(z) - J_g(y)\| \leq 4\delta \|\mathcal{J}_g(y)\|.$$

Analog ergibt sich

$$(1-4\delta)|\frac{\partial g}{\partial x}(y)| \leq |\frac{\partial g}{\partial x}(z)| \leq (1+4\delta)|\frac{\partial g}{\partial x}(y)|.$$

Wegen

$$g(z) - g(y) = \int_0^1 \mathcal{J}_g(y + t(z-y))(z-y)dt$$

erhalten wir

$$|g(z) - g(y) - \mathcal{J}_g(y)(z-y)| \leq 4\delta \|\mathcal{J}_g(y)\||z-y|.$$

Daraus ergibt sich wiederum

$$|\mathcal{J}_g^{-1}(y)(g(z)-g(y)) - (z-y)| \leq 4\delta \|\mathcal{J}_g^{-1}(y)\| \|mathcal J_g(y)\||z-y|.$$

Die rechte Seite lässt sich

$$\delta M |z-y|$$

mit

$$M = 16 \max\{\|\mathcal{J}_g^{-1}(x)\| \|\mathcal{J}_g(x)\| : x \in K\}$$

abschätzen. Für die Koordinaten von $\mathcal{J}_g^{-1}(y)(g(z)-g(y))$ bedeutet das

$$|J_g^{-1}(y)(g(z)-g(y))_i| \leq |z_i - y_i| + \delta M |z-y| \leq h_i(1 + 2\delta M\sqrt{d}).$$

Deshalb ist $g(I)$ in

$$\mathcal{J}_g(y)(\bigtimes_{i=1}^{d}[y_i - h_i(1+\delta M'), y_i + h_i(1+\delta M')]$$

(mit $M' = 2M\sqrt{d}$) enthalten. Also

$$\mu(I) = \lambda(g(I)) \leq (1+\delta M')^d |\frac{\partial g}{\partial x}(y)\lambda(I) \leq \frac{(1+\delta M')^d}{1-4\delta} \int_I |\frac{\partial g}{\partial x}|(z) d\lambda(z).$$

Wenn wir das für alle Intervalle I summieren, erhalten wir (mit $D = \bigcup_{i=1}^{N} C(x_i)$)

$$\mu(K) \leq \mu(D) \leq \frac{(1+\delta M')^d}{1-4\delta} \int_D |\frac{\partial g}{\partial x}| d\lambda.$$

Jetzt ist D in

$$U(K,\delta) = \{y : \exists x \in K : |y-x| < \delta\sqrt{d}\}$$

enthalten, und

$$\bigcap \delta > 0 U(K,\delta) = K.$$

Deshalb geht für $\delta \to 0$ $\lambda(U(K,\delta))$ gegen $\lambda(K)$. Außerdem ist die stetige Funktion $\frac{\partial g}{\partial x}$ auf der kompakten Menge K beschränkt, und daher auch auf D. Daraus erhalten wir für hinreichend kleines δ

$$\int_D |\frac{\partial g}{\partial x}| d\lambda \leq \int_K |\frac{\partial g}{\partial x}| d\lambda + \epsilon$$

und

$$\mu(A) \leq \mu(K) + \epsilon \leq (1+\epsilon)(\int_K |\frac{\partial g}{\partial x}| d\lambda + \epsilon \leq\leq (1+\epsilon)(\int_A |\frac{\partial g}{\partial x}| d\lambda + \epsilon.$$

Für $\epsilon \to 0$ ergibt sich unsere Behauptung.

Aus der Ungleichung

$$\mu(g(A)) \leq \int_A |\frac{\partial g}{\partial x}| d\lambda$$

ergibt sich für messbares $f \geq 0$

$$\int_{g(\Omega)} f \circ g^{-1} d\lambda \leq \int_\Omega f |\frac{\partial g}{\partial x}| d\lambda.$$

Wir ersetzen nun g durch g^{-1}, Ω durch $g(\Omega)$ und f durch $f \circ g^{-1} |\frac{\partial g}{\partial x}| \circ g^{-1}$. Das ergibt

$$\int_\Omega f |\frac{\partial g}{\partial x}| d\lambda \leq \int_{g(\Omega)} f \circ g^{-1} d\lambda.$$

Damit haben wir die umgekehrte Ungleichung und insgesamt

$$\int_\Omega f |\frac{\partial g}{\partial x}| d\lambda = \int_{g(\Omega)} f \circ g^{-1} d\lambda.$$

Als Anwendung dieses Satzes können wir die Verteilung von Summe, Produkt und Quotienten von zwei unabhängigen Zufallsvariablen bestimmen:

Satz 10.6

X_1 und X_2 seien unabhängig mit Dichte f_1 bzw. f_2. Dann sind auch die Verteilungen von $X_1 + X_2$, X_1X_2 und X_1/X_2 absolutstetig und

$$f_{X_1+X_2}(z) = \int f_1(x)f_2(z-x)d\lambda(x),$$

$$f_{X_1X_2}(z) = \int f_1(x)f_2(z/x)\frac{1}{|x|}d\lambda(x),$$

$$f_{X_1/X_2}(z) = \int f_1(zx)f_2(x)|x|d\lambda(x).$$

Für die Anwendung des Transformationssatzes benötigen wir neben der Funktion $g_1(x) = x_1 + x_2$ (bzw. x_1x_2, x_1/x_2) noch eine eine zweite Funktion g_2, die g injektiv macht (zumindest fast überall). In allen drei Fällen funktioniert $g_2 = x_2$. Damit ist $g^{-1}(y) = (y_1 - y_2, y_2)$ mit der Funktionaldeterminante 1 und

$$f_{X_1+X_2,X_2}(y) = f_1(y_1 - y_2)f_2(y_2).$$

Durch Integrieren über y_2 erhalten wir die gesuchte Randdichte von $X_1 + X_2$. Die analogen Rechnungen für Summe und Quotient sind der Leserin überlassen (es empfiehlt sich $\Omega = \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ zu setzen).

10.1 Wiederholungsfragen

1. Wie ist die Verallgemeinerte Inverse einer Verteilungsfunktion definiert?
2. Wie kann man Zufallszahlen mit einer beliebigen Verteilungsfunktion erzeugen?
3. X sei nach der stetigen Verteilungsfunktion F verteilt. Was kann man über $F(X)$ sagen?
4. Was besagt der Transformationssatz für Dichten?

Kapitel 11

Null-Eins Gesetze

Die beiden Lemmata von Borel-Cantelli implizieren, dass für eine Folge (A_n) von unabhängigen Ereignissen $\limsup_n A_n$ nur Wahrscheinlichkeit 0 oder 1 haben kann. Diese Feststellung ist die Folge eines allgemeineren Prinzips. Dazu definieren wir

Definition 11.1

(A_n) sei eine Folge von Ereignissen,

$$\mathfrak{S}_n = \mathfrak{A}_\sigma(\{A_k, k \geq n\}).$$

$$\mathfrak{S}_\infty = \bigcap_{n\in\mathbb{N}} \mathfrak{S}_n$$

heißt die terminale Sigmaalgebra bzw. die Sigmaalgebra der terminalen Ereignisse.

Satz 11.1: Null-Eins Gesetz von Kolmogorov

Wenn (A_n) unabhängige Ereignisse sind, dann gilt für alle $A \in \mathfrak{S}_\infty$ $\mathbb{P}(A) \in \{0, 1\}$.

Wir setzen

$$\mathfrak{S}'_n = \mathfrak{A}_\sigma(\{A_1, \ldots, A_n\}).$$

Der Ring

$$\mathfrak{R} = \bigcup_n \mathfrak{S}'_n$$

erzeugt $\mathfrak{S}_1$, daher gibt es nach dem Approximationssatz zu $A \in \mathfrak{S}_\infty$ ein $B \in \mathfrak{R}$ mit

$$\mathbb{P}(A \triangle B) < \epsilon.$$

Es gibt ein n, sodass $B \in \mathfrak{S}'_n$, und weil $A \in S_{n+1}$, sind A und B unabhängig.

Damit haben wir

$$|\mathbb{P}(A) - \mathbb{P}(A \cap B)| \leq \mathbb{P}(A \triangle A \cap B) = \mathbb{P}(A \cap (A \triangle B)) \leq \mathbb{P}(A \triangle B) \leq \epsilon$$

und

$$|\mathbb{P}(A) - \mathbb{P}(B)| \leq \mathbb{P}(A \triangle B) \leq \epsilon.$$

Wegen $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ ergibt sich insgesamt

$$|\mathbb{P}(A) - \mathbb{P}(A)^2| \leq 2\epsilon,$$

und weil wir $\epsilon > 0$ beliebig wählen können, ist

$$\mathbb{P}(A) = \mathbb{P}(A)^2$$

und damit $\mathbb{P}(A) \in \{0, 1\}$.

Für eine Folge von unabhängigen Zufallsvariablen $(X_n, n \in \mathbb{N})$ können wir die terminale Sigmaalgebra $\mathfrak{S}_\infty$ analog definieren:

Definition 11.2

Die von den Zufallsvariablen $X_i, i \in I$ erzeugte Sigmaalgebra

$$\mathfrak{A}_\sigma(X_i, i \in I)$$

ist die kleinste Sigmaalgebra, bezüglich der alle X_i messbar sind. Die terminale Sigmaalgebra der Folge $(X_n, n \in \mathbb{N})$ ist

$$\mathfrak{S}_\infty = \bigcap_{n\in\mathbb{N}} \mathfrak{S}_n$$

mit

$$\mathfrak{S}_n = \mathfrak{A}_\sigma(X_i, i \geq n).$$

Nach demselben Muster wie für Ereignisse beweist man den

Satz 11.2: 0-1 Gesetz von Kolmogorov II

Wenn $(X_n, n \in \mathbb{N})$ eine Folge von unabhängigen Zufallsvariablen ist, dann gilt für $A \in \mathfrak{S}_\infty$

$$\mathbb{P}(A) \in \{0, 1\}.$$

Ein ähnliches Gesetz gilt, wenn die unabhängigen Zufallsvariablen im vorigen Beispiel durch Partialsummen von unabhängigen, identisch verteilten Zufallsvariablen ersetzt werden. In etwas allgemeinerer Form ist das die Aussage des nächsten Satzes:

Satz 11.3: 0-1 Gesetz von Hewitt-Savage

$(X_n, n \in N)$ sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen,

$$\mathfrak{G} = \{A \in \mathbb{R}^{\mathbb{N}} : \pi A = A \text{ für alle Permutationen } \pi\},$$

das heißt, für jedes n, $x \in A$ und jede Permutation π von $\{1, \ldots, n\}$ soll

$$(x_{\pi(1)}, \ldots, x_{\pi(n)}, x_{n+1}, \ldots) \in A.$$

Dann gilt für jedes A in

$$\mathfrak{H} = X^{-1}(\mathfrak{G}) = \{\{\omega : (X_1(\omega), \ldots) \in A\} : A \in \mathfrak{G}\}$$

$$\mathbb{P}(A) \in \{0, 1\}.$$

Aus dem Approximationssatz folgt, dass wir zu jedem $\epsilon > 0$ ein n und ein $B \in \mathfrak{B}_n$ finden können, sodass

$$\mathbb{P}(A \triangle C) \leq \epsilon$$

mit

$$C = [(X_1, \ldots X_n) \in B]$$

gilt. wegen der Symmetrie von A gilt dann auch

$$\mathbb{P}(A \triangle C') \leq \epsilon$$

mit

$$C' = [(X_{n+1}, \ldots X_{2n}) \in B].$$

Daraus ergibt sich einerseits

$$|\mathbb{P}(A) - \mathbb{P}(C \cap C')| \leq \mathbb{P}(A \triangle (C \cap C')) \leq 2\epsilon$$

und andererseits

$$|\mathbb{P}(A)^2 - \mathbb{P}(C \cap C')| = |\mathbb{P}(A)^2 - \mathbb{P}(C)\mathbb{P}(c')| \leq 2\epsilon.$$

Insgesamt erhalten wir

$$|\mathbb{P}(A) - \mathbb{P}(A)^2| \leq 4\epsilon,$$

und weil $\epsilon > 0$ beliebig gewählt werden kann, ergibt sich schließlich

$$\mathbb{P}(A) = \mathbb{P}(A)^2$$

und damit die Behauptung.

11.1 Wiederholungsfragen

1. Was besagt das Null-Eins-Gesetz von Kolmogorov?
2. Was besagt das Null-Eins-Gesetz von Hewitt-Savage?

Kapitel 12

Charakteristische Funktionen

12.1 Definition und Eigenschaften

X sei eine reelle Zufallsvariable. Die charakteristische Funktion von X ist gegeben durch

$$\phi_X(t) = \mathbb{E}(e^{iXt}) = \mathbb{E}(\cos(Xt)) + i\mathbb{E}(\sin(Xt)).$$

Für die charakteristische Funktion gelten folgende Eigenschaften:

Satz 12.1

ϕ_X sei die charakteristische Funktion einer Zufallsvariable X mit der Verteilungsfunktion F_X. Dann gilt

1. $|\phi_X(t)| \leq 1$,
2. $\phi_X(-t) = \bar{\phi}_X(t)$,
3. ϕ ist gleichmäßig stetig.
4. Falls $\mathbb{E}|X|^k < \infty$, dann ist ϕ_X k-mal differenzierbar und
$$\phi_X^{(k)}(0) = i^k \mathbb{E}(X^k).$$
5. Falls ϕ in 0 $2k$-mal differenzierbar ist, so ist $\mathbb{E}(X^{2k})$ endlich.
6. Falls X und Y unabhängig sind, dann ist
$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$
7. Falls $Y = aX + b$, dann ist
$$\phi_Y(t) = e^{ibt}\phi_X(at).$$
8. Falls a und b Stetigkeitspunkte von F_X sind, so ist
$$F_X(b) - F_X(a) = \lim_{A\to\infty} \int_{-A}^{A} \frac{e^{-iat} - e^{-ibt}}{2\pi it} \phi_X(t)dt.$$
(Da die Stetigkeitspunkte von F_X dicht liegen und F_X rechtsstetig ist, folgt daraus, dass die charakteristische Funktion die Verteilung von X eindeutig bestimmt)

9. ϕ_X ist positiv definit, d.h., für alle reellen $t_1, \ldots, t_n$ und alle komplexen $z_1, \ldots, z_n$ gilt

$$\sum_{i=1}^{n}\sum_{j=1}^{n} z_i \bar{z}_j \phi_X(t_i - t_j) \geq 0.$$

   (insbesondere ist diese Summe reell)

10. (Ohne Beweis) Eine Funktion ϕ ist genau dann eine charakteristische Funktion, wenn sie positiv definit und stetig in 0 ist und in 0 den Wert 1 annimmt.

Die ersten beiden Eigenschaften folgen direkt aus der Definition von ϕ_X.

Zum Beweis von Eigenschaft (3) seien s und t zwei Zahlen mit $|s-t| < \epsilon$. Dann ist

$$|\phi_X(t) - \phi_X(s)| \leq \int |e^{ixt} - e^{ixs}| dF(x) = \int 2|\sin(\frac{x(s-t)}{2})| dF_X(x)$$

$$\leq \int \min(|x(s-t)|, 2) dF_X(x) \leq \int \min(|\epsilon x|, 2) dFX(x).$$

Das letzte Integral konvergiert für $\epsilon \to 0$ gegen 0, somit ist ϕ_X gleichmäßig stetig.

Für die nächste Eigenschaft leiten wir in der Gleichung

$$\phi_X(t) = \int e^{ixt} dF_X(x)$$

unter dem Integralzeichen k mal ab. Der neue Integrand

$$i^k x^k e^{ixt}$$

hat die konvergente Majorante $|x|^k$. Deshalb können wir Ableitung und Integral vertauschen und erhalten

$$\phi_X^{(k)}(t) = \int i^k x^k e^{ixt} dF_X(x).$$

Für $t = 0$ ergibt sich insbesondere

$$\phi_X^{(k)}(0) = \int i^k x^k dF_X(x).$$

Wir zeigen Eigenschaft (5) für $k = 1$. Der allgemeine Fall lässt sich durch Induktion beweisen, indem man die neue Verteilungsfunktion

$$\tilde{F}(x) = \frac{1}{\mathbb{E}(X^{2k-2})} \int_0^x x^{2k-2} dF_X(x)$$

betrachtet. Es gilt

$$\phi''(0) = \lim_{t\to 0} \frac{2 - \phi(t) - \phi(-t)}{t^2} = \lim_{t\to 0} \int \frac{2(1-\cos(xt))}{t^2} dF_X(x).$$

Der Integrand ist nichtnegativ und konvergiert für $t \to 0$ gegen x^2. Daher gilt

$$\phi''(0) \geq \lim_{t\to 0} \int_{-A}^{A} \frac{2(1-\cos(xt))}{t^2} dF_X(x) = \int_{-A}^{A} x^2 dF_X(x).$$

Für $A \to \infty$ erhalten wir daraus

$$\phi''(0) \geq \int x^2 dF_X(x).$$

Es ist also $\mathbb{E}(X^2)$ endlich.

Die nächsten beiden Eigenschaften sind durch einfaches Einsetzen nachzuprüfen.

Für die Umkehrformel (8) haben wir

$$\int_{-A}^{A} \frac{e^{-iat} - e^{-ibt}}{2\pi it} \phi_X(t) dt = \int_{-A}^{A} \frac{e^{-iat} - e^{-ibt}}{2\pi it} \int e^{ixt} dF_X(x) dt$$

$$= \int dF_X(x) \frac{1}{\pi} \int_0^A \frac{\sin(x-a)t - \sin(x-b)t}{t} dt = \int (g(A(x-a)) - g(A(x-b))) dF_X(x),$$

wobei wir

$$g(x) = \frac{1}{\pi} \int_0^1 \frac{\sin tx}{t} dt = \frac{1}{\pi} \int_0^x \frac{\sin t}{t} dt$$

gesetzt haben. Die Funktion g ist beschränkt und konvergiert für $t \to \infty$ gegen $1/2$ und für $t \to -\infty$ gegen $-1/2$.

Der Integrand $g(A(x-a)) - g(A(x-b))$ konvergiert also für $A \to \infty$ gegen

$$h(x) = \begin{cases} 0 & \text{falls } x < a, \\ 1/2 & \text{falls } x = a, \\ 1 & \text{falls } a < x < b, \\ 1/2 & \text{falls } x = b, \\ 0 & \text{falls } x > b \end{cases}$$

Das obige Integral konvergiert also gegen

$$\int h(x) dF_X(x) = \frac{F_X(b) + F_X(b-0) - F_X(a) - F_X(a-0)}{2}.$$

Falls a und b Stetigkeitspunkte von F sind, ist der letzte Ausdruck $= F_X(b) - F_X(a)$.

Für den Beweis der positiven Definitheit setzen wir

$$Y = \sum_{j=1}^{n} z_j e^{iXt_j}.$$

Dann ist

$$\sum_{j=1}^{n} \sum_{k=1}^{n} z_j \bar{z}_k e^{iX(t_j - t_k)} = Y\bar{Y} \geq 0$$

und

$$\sum_{j=1}^{n} \sum_{k=1}^{n} z_j \bar{z}_k \phi(t_j - t_k) = \mathbb{E}(Y\bar{Y}) \geq 0.$$

Es ist etwas störend, dass für ungerades k die k-fache Differenzierbarkeit der charakteristischen Funktion in 0 zwar notwendig, aber nicht hinreichend für die Existenz des k-ten Moments ist. Wir fragen uns also, ob wir für beide Aussagen äquivalente Bedingungen in der jeweils "dualen" Welt finden können: das geht tatsächlich:

Satz 12.2

1. $\phi'_X(0) = ia$ genau dann, wenn

$$\lim_{M \to \infty} M\mathbb{P}(|X| > M) = 0$$

und

$$\lim_{M \to \infty} \mathbb{E}(X[|X| < M]) = a.$$

2. $\mathbb{E}(X) < \infty$ genau dann, wenn

$$\int_0^\infty \frac{1 - \Re\phi_X(t)}{t^2} dt < \infty.$$

Der Beweis bleibt dem Leser überlassen, es soll nur erwähnt werden, dass das Integral in der Bedingung für die Existenz des Erwartungswerts gleich $\frac{\pi}{2}\mathbb{E}(|X|)$ ist, und dass darin der Realteil durch den Betrag ersetzt werden kann.

12.2 Beispiele

Wie wollen nun einige charakteristische Funktionen für bekannte Verteilungen ausrechnen:

12.2.1 Diskrete Verteilungen

1. Die diskrete Gleichverteilung:

$$\mathbb{P}(X = k) = \frac{1}{b-a+1} \ (x = a, a+1, \ldots, b).$$

Wir erhalten

$$\phi_X(t) = \frac{1}{b-a+1}\sum_{k=a}^{b} e^{ikt} = \frac{e^{i(b+1)t} - e^{iat}}{(e^{it}-1)(b-a+1)} = e^{i(a+b)t/2}\frac{\sin((b-a+1)t/2)}{(b-a+1)\sin(t/2)}.$$

2. Die Binomialverteilung:

$$\phi_X(t) = \sum \binom{n}{k} p^k q^{n-k} e^{ikt} = (q + pe^{it})^n.$$

3. Die Poissonverteilung:

$$\phi_X(t) = \sum_k \frac{\lambda^k}{k!} e^{-\lambda} e^{ikt} = e^{\lambda(e^{it}-1)}.$$

12.2.2 Stetige Verteilungen

1. Die stetige Gleichverteilung

$$\phi_X(t) = \frac{1}{b-a}\int_a^b e^{ixt}dx = \frac{e^{ibt} - e^{iat}}{it(b-a)}.$$

2. Die Exponentialverteilung:

$$\phi_X(t) = \int_0^\infty \lambda e^{ixt} e^{-\lambda x} dx = \frac{\lambda}{\lambda - it}.$$

3. Die Gammaverteilung:

Diese charakteristische Funktion kann man mit Hilfe von komplexen Konturintegralen ausrechnen. Wir verwenden einen anderen Zugang; wenn nämlich

$$\psi(t) = \mathbb{E}(e^{Xt})$$

für ein $t > 0$ existiert, ist $\psi(z)$ für $0 < \mathbb{R}e(z) < t$ analytisch. Wir bekommen also durch analytische Fortsetzung

$$\phi_X(t) = \psi(it).$$

Im konkreten Fall gibt das

$$\phi_X(t) = (\frac{\lambda}{\lambda - it})^\alpha.$$

4. Die Standard-Normalverteilung:

   Mit demselben Trick wie vorhin ist

$$\psi(t) = \int_{-\infty}^{\infty} e^x t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} dx = e^{t^2/2}$$

   und daher

$$\phi_X(t) = e^{-t^2/2}.$$

5. Die Normalverteilung:

   Falls X normalverteilt ist mit Mittel μ und Varianz σ^2, dann kann man X schreiben als

$$X = \mu + \sigma Y,$$

   wobei Y standard-normalverteilt ist. Aus dem vorigen Punkt und Satz 1 folgt also

$$\phi_X(t) = e^{i\mu t - \sigma^2 t^2/2}.$$

12.2.3 Der Darstellungssatz von Riesz (Aldi-Version)

In seiner allgemeinen Form befasst sich dieser Satz mit kompakten bzw. lokalkompakten Räumen; wir beweisen hier die ursprüngliche und einfachste Version:

Definition 12.1

Ein lineares Funktional auf einem Raum $\mathfrak{X}$ von reellen Funktionen heißt positiv, wenn für alle $f \in \mathfrak{X}$ mit $f \geq 0$ $l(f) \geq 0$ gilt.

Satz 12.3: Darstellungssatz von Riesz I

Zu jedem positiven linearen Funktional l auf $C[0,1]$ gibt es ein Maß μ, sodass

$$l(f) = \int f d\mu.$$

Zuerst stellen wir fest, dass ein solches positives Funktional beschränkt ist (mit $\|l\| = l(1)$) und dass durch $l(f+ig) = l(f) + il(g)$ ein lineares Funktional auf dem Raum der komplexwertigen stetigen Funktionen definiert wird. Dann setzen wir $\phi(t) = l(e^{it\cdot})$. Für $z_1, \ldots, z_n \in \mathbb{C}, t_1, \ldots, t_n \in R$ gilt wegen

$$f(x) = \sum_{i,j} z_i \bar{z}_j e^{(t_i - t_j)x} = |\sum_i z_i e^{t_i x})|^2 \geq 0$$

$$\sum_{i,j} z_i \bar{z}_j \phi(t_i - t_j) = l(f) \geq 0,$$

also ist ϕ positiv definit. Außerdem konvergiert für $t \to 0$ e^{itx} für $x \in [0,1]$ gleichmäßig gegen 1, und daher wegen der Beschränktheit von l auch $\phi(t)$ gegen $\phi(0)$. ϕ ist also positiv definit und bei 0 stetig, also die charakteristische Funktion eines endlichen Maßes. Dieses Maß erfüllt also

$$l(f) = \int f d\mu$$

für alle Funktionen der Form $f(x) = e^{itx}$. Wegen der Linearität von l gilt das dann auch für alle Linearkombinationen von trigonometrischen Funktionen, und weil jede stetige Funktion auf $[0,1]$ gleichmäßig durch solche Funktionen approximiert werden kann (etwa nach dem Satz von Weierstrass), gilt die Gleichung für alle $f \in C[0,1]$.

Dieser Satz führt uns zur Charakterisierung des Dualraums von $C[0,1]$. Dazu beweisen wir zuerst, dass jedes stetige lineare Funktional auf $C[0,1]$ als Differenz zweier positiver Funktionale dargestellt werden kann. Wir setzen für $f \geq 0$

$$l_+(f) = \sup\{l(g) : 0 \leq g \leq f, g \in C[0,1]\}.$$

Man überzeugt sich leicht, dass l_+ positiv, additiv und positiv homogen auf $C_+[0,1]$ ist: für die Additivität kann man verwenden, dass aus $g_1 \leq f_1$ und $g_2 \leq f_2$ $g_1 + g_2 \leq f_1 + f_2$ folgt, und andererseits aus $g \leq f_1 + f_2$ $hf_i/(f_1 + f_2) \leq f_i$ (und dass die letzten Funktionen stetig sind). Mit $l_+(f-g) = l_+(f) - l_+(g)$ setzt man l_+ zu einem positiven linearen Funktional auf ganz $C[0,1]$ fort. $l_- = l_+ - l$ ist dann ebenfalls linear und positiv. Diese Darstellung gibt uns

Satz 12.4: Darstellungssatz von Riesz II

Der Dualraum zu $(C[0,1], \|.\|_\infty)$ ist isometrisch zum Raum der endlichen signierten Maße auf $([0,1], \mathbb{B})$ mit der Variationsnorm.

Das einzige an dieser Aussage, dass noch zu beweisen ist, ist die Isometrie, also, dass die Dualraumnorm des Funktionals $\ell : f \mapsto \int f d\mu$ gleich $\|\mu\|_V = |\mu|([0,1])|$. Einerseits ist für jede stetige Funktion f mit $\|f\|_{\sup} \leq 1$

$$\int f d\mu \leq \int |f| d|\mu| \leq |\mu|([0,1]),$$

andererseits gilt

$$\|mu\|([0,1]) = \int g d\mu$$

mit $g = P - N$, wobei (P, N) eine Hahn-Zerlegung von μ ist. g lässt sich in $\mathcal{L}_1(|\mu|)$ durch stetige Funktionen approximieren, also gibt es zu jedem $\epsilon > 0$ eine stetige Funktion f mit $\|f\|_{\sup} \leq 1$ und $\int f d\mu \geq \|\mu\|_V - \epsilon$, und damit ist auch die umgekehrte Ungleichung gezeigt.

12.3 Konvergenz von Verteilungen

Für eine Folge F_n von Verteilungsfunktionen definieren wir die Konvergenz wie folgt:

Definition 12.2

F_n konvergiert schwach gegen F (in Zeichen $F_n \Rightarrow F$) falls für jede stetige beschränkte Funktion f

$$\lim_{n\to\infty} \int f dF_n = \int f dF.$$

Definition 12.3

Eine Folge von X_n von Zufallsvariablen konvergiert in Verteilung gegen eine Zufallsvariable X, falls die Folge der Verteilungsfunktionen von X_n gegen die Verteilungsfunktion von X konvergiert.

Die schwache Konvergenz lässt sich folgendermaßen charakterisieren:

Satz 12.5

Die folgenden Behauptungen sind äquivalent:

1. $F_n \Rightarrow F$
2. $F_n(x) \to F(x)$ für alle Stetigkeitspunkte von F (wir werden im weiteren $\mathcal{C}(F)$ für die Menge der Stetigkeitspunkte einer Funktion schreiben).

3. $d(F_n, F) \to 0$, wobei

$$d(F, G) = \inf\{\epsilon \geq 0 : F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon \text{ für alle } x\}$$

die sogenannte Lévy-Metrik ist.

Wir zeigen zuerst (2) $\Rightarrow$ (1). Dazu sei f eine beschränkte stetige Funktion. Es sei

$$M = \sup |f(x)|.$$

Da die Menge der Unstetigkeiten einer monotonen Funktion höchstens abzählbar ist, können wir $a, b \in \mathcal{C}(F)$ wählen, sodass

$$F(a) < \epsilon$$

und

$$F(b) > 1 - \epsilon$$

Wegen der gleichmäßigen Stetigkeit von f auf $[a, b]$ gibt es Punkte

$$x_0 = a < x_1 < \ldots < x_N = b$$

aus $\mathcal{C}(F)$, sodass für $x_{i-1} \leq x \leq x_i$

$$|f(x) - f(x_i)| < \epsilon.$$

Wir setzen jetzt

$$\tilde{f}(x) = \begin{cases} f(x_i) & \text{falls } x_{i-1} \leq x \leq x_i, \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt

$$|\int f dF - \int \tilde{f} dF| < \int |f - \tilde{f}| dF < (2M+1)\epsilon.$$

Falls n groß genug ist, gilt für alle $i = 0, \ldots, n$

$$|F(x_i) - F_n(x_i)| < \frac{\epsilon}{NM}.$$

Deswegen erhalten wir

$$|\int f dF_n - \int \tilde{f} dF_n| < (2M+2)\epsilon.$$

Schließlich ist

$$|\int \tilde{f} dF_n - \int \tilde{f} dF| = |\sum_{i=1}^{N} f(x_i)(F_n(x_i) - F_n(x_{i-1}) - F(x_i) + F(x_{i-1}))| \leq 2\epsilon.$$

Insgesamt ergibt sich

$$|\int f dF_n - \int f dF| \leq (4M+5)\epsilon.$$

Da wir ϵ beliebig klein wählen können, konvergiert also $\int f dF_n$ nach $\int f dF$.

Als nächstes zeigen wir dass aus (3) (2) folgt. (Eigentlich ist noch zu zeigen, dass d wirklich eine Metrik ist; das sei dem Leser überlassen).

Da $d(F_n, F)$ gegen 0 geht, gibt es für jedes ϵ ein n_0, sodass für $n > n_0$ und für alle x

$$F(x-\epsilon) - \epsilon < F_n(x) < F(x+\epsilon) + \epsilon.$$

Falls $x \in \mathcal{C}(F)$, kann man ϵ so wählen, dass sich beide Außenterme in dieser Ungleichung beliebig wenig von $F(x)$ unterscheiden. $F_n(x)$ muss also für $x \in \mathcal{C}(F)$ gegen $F(x)$ konvergieren.

Als letztes zeigen wir nun (1) $\Rightarrow$ (3). Wir nehmen an, dass $d(F_n, F) \not\to 0$ und zeigen, dass dann auch nicht $F_n \Rightarrow F$ gelten kann.

Wir können ein $\epsilon > 0$ und eine Teilfolge von F_n finden, sodass entlang dieser Teilfolge $d(F_n, F) > \epsilon$ gilt. O.B.d.A. sei F_n selbst diese Teilfolge. Es gibt dann für jedes n ein x_n, sodass

$$F(x_n + \epsilon) + \epsilon < F_n(x_n).$$

oder

$$F(x_n - \epsilon) - \epsilon > F_n(x_n)$$

Eine dieser beiden Ungleichungen ist für unendlich viele n erfüllt, nehmen wir an, es sei die zweite. Indem wir evtl. wieder eine Teilfolge auswählen, können wir annehmen, dass diese Ungleichung für alle n gilt. Wegen der Beziehung $F(x_n) > \epsilon$ ist die Folge x_n nach unten beschränkt. Indem wir wieder eine Teilfolge auswählen, können wir annehmen, dass x_n gegen einen Grenzwert x konvergiert (dieser kann ∞ sein). Wir können jetzt

$$y < y' < x$$

so wählen, dass

$$F(y) > F(x - 0) - \epsilon/2.$$

Falls n groß genug ist, gilt $x_n > y$ und $x_n - \epsilon < x$.

Wir wählen jetzt

$$f(u) = \begin{cases} 1 & \text{falls } u < y, \\ \frac{y'-u}{y'-y} & \text{fall } y \leq u \leq y', \\ 0 & \text{falls } u > y' \end{cases}.$$

f ist stetig und beschränkt, und es gilt für n groß genug:

$$\int f dF_n \leq F_n(y') < F_n(x_n) < F(x_n - \epsilon) - \epsilon \leq F(x - 0) = \epsilon.$$

Andererseits ist

$$\int f dF \geq F(y) \geq F(x - 0) - \epsilon/2.$$

$\int f dF_n$ konvergiert also nicht gegen $\int f dF$; somit ist $F_n \not\Rightarrow F$.

Nun kommen wir zum Zusammenhang zwischen der Konvergenz von Verteilungen und der von charakteristischen Funktionen:

Satz 12.6

F_n sei eine Folge von Verteilungsfunktionen und ϕ_n seien die zugehörigen charakteristischen Funktionen. Dann gilt

1. Falls $F_n \Rightarrow F$, dann konvergiert ϕ_n gegen die charakteristische Funktion ϕ von F gleichmäßig auf kompakten Intervallen.
2. Falls ϕ_n gegen eine Funktion ϕ konvergiert, die bei 0 stetig ist, dann ist ϕ charakteristische Funktion zu einer Verteilungsfunktion F und $F_n \Rightarrow F$.

Teil (1) ist relativ schnell erledigt. Aus dem vorhergehenden Satz folgt nämlich dass ϕ_n gegen ϕ punktweise konvergiert. Es bleibt noch zu zeigen, dass diese Konvergenz im Bereich $|t| \leq M$ gleichmäßig erfolgt.

Dazu stellen wir wie im Beweis von Satz 1 fest, dass für $|s - t| < \delta$

$$|\phi(t) - \phi(s)| < \int \min(|\delta x|, 2) dF(x)$$

gilt. Dieses Integral kann kleiner als ein vorgegebenes ϵ gemacht werden, indem man ϵ klein genug wählt. Wir wählen jetzt $N > 2M/\delta$ und setzen

$$t_i = -M + i\delta.$$

Es gilt natürlich auch

$$|\phi_n(t) - \phi_n(s)| < \int \min(|\delta x|, 2) dF_n(x)$$

Da der Integrand eine stetige beschränkte Funktion ist, ist für n groß genug und für $|t - s| < \delta$

$$|\phi_n(t) - \phi_n(s)| < 2\epsilon$$

und für $i = 0, \ldots, N$

$$|\phi_n(t_i) - \phi(t_i)| < \epsilon.$$

Somit gilt für $t_{i-1} \leq t \leq t_i$, $i = 1, \ldots, N$

$$|\phi(t) - \phi_n(t)| \leq |\phi(t) - \phi(t_i)| + |\phi(t_i) - \phi_n(t_i)| + |\phi_n(t_i) - \phi_n(t)| \leq 4\epsilon.$$

Das zeigt die gleichmäßige Konvergenz.

Für den zweiten Teil unseres Satzes brauchen wir den folgenden Hilfssatz:

Satz 12.7

(Satz von HELLY) Aus jeder Folge F_n von Verteilungsfunktionen kann man eine Teilfolge aussuchen, die gegen eine monotone rechtsstetige Funktion G in deren Stetigkeitspunkten konvergiert.

Zum Beweis sei r_k eine Abzählung der rationalen Zahlen. Wir können aus F_n eine Teilfolge F_{1n} auswählen, die in r_1 gegen den Grenzwert $g_1$konvergiert. Aus F_{1n} wählen wir eine Teilfolge F_{2n} aus, die in r_2 gegen den Grenzwert g_2 konvergiert; allgemein sei F_{kn} eine Teilfolge von $F_{k-1,n}$, die in r_k gegen den Grenzwert g_k konvergiert.

Schließlich betrachten wir die Folge F_{nn}. Für $n > k$ ist das eine Teilfolge von F_{kn} und konvergiert daher in r_k gegen g_k.

Jetzt sei

$$G(x) = \inf\{g_k : r_k > x\}.$$

G ist rechtsstetig und monoton. Falls x ein Stetigkeitspunkt von G ist, so gibt es k und l so, dass

$$r_k < x < r_l$$

und

$$G(x) - \epsilon < g_k \leq g_l \leq G(x) + \epsilon.$$

Daher gilt

$$G(x) - \epsilon < g_k = \lim F_{nn}(r_k) \leq \liminf F_{nn}(x)$$

und

$$\limsup F_{nn}(x) \leq \lim F_{nn}(r_l) \leq G(x) + \epsilon.$$

Da wir ϵ beliebig wählen können, gilt also $F_{nn}(x) \to G(x)$.

Nun kehren wir zum Beweis unseres Konvergenzsatzes zurück. Wir setzen in der Umkehrformel aus Satz 1 $b = x$ und $a = -x$ und erhalten

$$F_n(x) - F_n(-x) = \frac{1}{\pi} \lim_{A\to\infty} \int_{-A}^{A} \frac{\sin xt}{t} \phi_n(t) dt.$$

Wenn man das über x integriert und berücksichtigt, dass $F_n(x) - F_n(-x)$ nichtfallend ist, erhält man

$$F_n(a) - F_n(-a) \geq \frac{1}{a} \int_0^a (F_n(x) - F_n(-x)) dx = \lim_{A\to\infty} \int_{-A}^{A} \frac{1 - \cos at}{\pi a t^2} \phi_n(t) dt.$$

$$= \int \frac{1 - \cos t}{\pi t^2} \phi_n(\frac{t}{a}) dt.$$

Wenn man im letzten Integral ϕ_n durch ϕ ersetzt, konvergiert $\phi(\frac{t}{a})$ für $a \to \infty$ wegen der Stetigkeit von ϕ in 0 gegen 1. Das ganze Integral konvergiert also für $a \to \infty$ gegen

$$\int \frac{1 - \cos t}{\pi t^2} dt = 1.$$

Wir können also ein $a > 0$ so wählen, dass Dieses Integral größer ist als $1 - \epsilon$. Da ϕ_n punktweise gegen ϕ konvergiert, ist für genügend großes n

$$F_n(a) - F_n(-a) > 1 - 2\epsilon.$$

Dann gilt aber auch (nach einem Grenzübergang)

$$G(a) - G(-a) > 1 - 2\epsilon.$$

Da ϵ beliebig gewählt werden kann, schließen wir daraus, dass G eine Verteilungsfunktion ist. Nach dem ersten Teil unseres Satzes muss dann auch ϕ die charakteristische Funktion sein, die zu G gehört.

Es bleibt noch zu zeigen, dass die ganze Folge F_n gegen G konvergiert. Wäre das nicht der Fall, gäbe es eine Teilfolge, entlang der $d(F_n, G)$ gegen einen von 0 verschiedenen Wert konvergiert. Aus dieser Teilfolge können wir aber wieder wie oben eine (Unter-) Teilfolge auswählen, die gegen G konvergiert, für die also $d(F_n, G)$ gegen 0 gehen müsste, was ein offensichtlicher Widerspruch ist.

12.4 Der zentrale Grenzwertsatz

Wir haben jetzt alle Hilfsmittel gesammelt, um den zweiten großen Satz der klassischen Wahrscheinlichkeitstheorie zu beweisen:

Satz 12.8: zentraler Grenzwertsatz

$X_1, \ldots$ sei eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit Mittel μ und Varianz σ^2. Dann gilt:

$$\lim_{n\to\infty} \mathbb{P}(\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq x) = \Phi(x),$$

wobei Φ die Verteilungsfunktion der Standard-Normalverteilung ist.

Als erstes stellen wir fest, dass wir den Beweis nur für $\mu = 0$ und $\sigma^2 = 1$ zu führen brauchen; im allgemeinen Fall können wir nämlich

$$Y = \frac{X - \mu}{\sigma}$$

setzen. Dann gilt

$$\mathbb{E}(Y) = 0,$$
$$\mathbb{V}(Y) = 1$$

und

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} = \frac{Y_1 + \ldots + Y_n}{\sqrt{n}}.$$

Falls nun X Erwartungswert 0 und Varianz 1 hat, gilt für die charakteristische Funktion

$$\phi_X(t) = 1 - \frac{t^2}{2} + o(t^2)$$

wenn $t \to 0$. Für die charakteristische Funktion ϕ_n von $(X_1 + \ldots + X_n)/\sqrt{n}$ ergibt sich daher

$$\phi_n(t) = (\phi_X(\frac{t}{\sqrt{n}}))^n = (1 - \frac{t^2}{2n} + o(\frac{1}{n}))^n.$$

Für $n \to \infty$ erhalten wir daraus

$$\lim_{n\to\infty} \phi_n(t) = e^{-t^2/2}.$$

Die Folge ϕ_n konvergiert also punktweise gegen die charakteristische Funktion der Standard-Normalverteilung. Nach unserem Konvergenzsatz konvergiert also die Verteilungsfunktion von $(X_1 + \ldots + X_n)/\sqrt{n}$ gegen Φ in allen Stetigkeitspunkten von Φ; weil Φ aber überall stetig ist, ist damit der Satz bewiesen.

Wenn wir darauf verzichten, dass die einzelnen Zufallsvariablen identisch verteilt sind, können wir immer noch fragen, ob (bzw. wann) die Folge

$$(S_n - \mathbb{E}(S_n))/\sqrt{\mathbb{V}(S_n)}$$

in Verteilung gegen die Standardnormalverteilung konvergiert. Dabei kann es passieren, dass die Summe von einem (oder wenigen) Summanden dominiert wird (etwa für $X_n = n!Y_n$ mit (Y_n) unabhängig und identisch verteilt). In diesem Fall wird das asymptotische Verhalten der Summe vom Verhalten der dominierenden Summanden bestimmt (in unserem Beispiel ist S_n genau dann asymptotisch normalverteilt, wenn Y_n normalverteilt ist). Diesen Fall werden wir ausschließen, indem wir verlangen, dass die einzelnen Summanden im Vergleich zur ganzen Summe klein sind; die genaue Formulierung sparen wir uns noch ein wenig auf, vorher machen wir unsere Summen noch ein wenig allgemeiner:

Definition 12.4

Wir nennen $(X_{nj}, 1 \leq j \leq n)$ eine Dreiecks folge, wenn für jedes n $(X_{n1}, \ldots, X_{nn})$ unabhängig sind.

Wir verlangen also Unabhängigkeit innerhalb jeder "Zeile". Über die Abhängigkeiten zwischen den einzelnen Zeilen treffen wir keine Annahmen. Die ursprüngliche Frage der Partialsummen einer einfachen Folge passt in dieses Modell, wenn wir $X_{nj} = X_j$ setzen.

Definition 12.5

Die Dreiecksfolge $(X_{nj}$ heißt gleichmäßig vernachlässigbar, wenn

$$\lim_{n\to\infty} \frac{\max_j \mathbb{V}(X_{nj})}{\sum_{k\leq n} \mathbb{V}(X_{nk}} = 0.$$

Satz 12.9: Lindeberg-Feller

(X_{nj}) sei eine Dreiecksfolge mit $\mathbb{E}(X_{nj}) = 0$, $\sigma_{nj}^2 = \mathbb{V}(X_{nj}) < \infty$. Wir setzen

$$S_n = \sum_{j=1}^{n} X_{nj}$$

und

$$\sigma_n^2 = \mathbb{V}(S_n) = \sum_{j=1}^{n} \sigma_{nj}^2.$$

Dann gilt:

1. Die Lindeberg-Feller Bedingung

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{j=1}^{n} \mathbb{E}(X_{nj}^2[|X_{nj}| > \epsilon\sigma_n]) \to 0 \text{ für alle } \epsilon > 0$$

   ist hinreichend für die Konvergenz der Verteilung von S_n/σ_n gegen die Standardnormalverteilung.

2. Ist X_{nj} zusätzlich gleichmäßig vernachlässigbar, dann ist diese Bedingung auch notwendig.

Wir können ohne Beschränkung der Allgemeinheit $\sigma_n = 1$ setzen. Wir haben die Ungleichungen

$$\mathbb{P}(X_{nj} \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathbb{E}(X_{nj}^2[X_{nj} > \epsilon])$$

und

$$|\mathbb{E}(X_{nj}[|X_{nj}| \geq \epsilon])| = |\mathbb{E}(X_{nj}[|X_{nj}| > \epsilon])| \leq \frac{1}{\epsilon}\mathbb{E}(X_{nj}^2[X_{nj} > \epsilon]).$$

Wenn wir $Y_{nj} = X_{nj}[|X_{nj}| \leq \epsilon]$ setzen, ergibt das, dass

$$\mathbb{P}(\exists i : Y_{nj} \neq X_{nj}) \to 0)$$

und

$$\sum_j \mathbb{E}(Y_{nj}) \to 0.$$

Die erste Gleichung impliziert, dass $\sum_j Y_{nj}$ und S_n dieselbe Grenzverteilung haben.

Für die charakteristische Funktion von Y_{nj} gilt

$$\phi_{nj}(t) = 1 + it\mathbb{E}(Y_{nj}) - \frac{t^2}{2}\mathbb{E}(Y_{nj}^2) + O(\epsilon\mathbb{E}(Y_{nj}^2).$$

Mit der Taylorentwicklung $\log(1+x) \approx x - x^2/2$ und wegen $\sum_j \mathbb{E}(Y_{nj}) \to 0$, $\sum_j \mathbb{E}(Y_{nj})^2 \to 0$, $\sum_j \mathbb{E}(Y_{nj}^2) \to 1$ ergibt sich

$$\lim_n \prod_j \phi_{nj}(t) = e^{-t^2/2+O(\epsilon)},$$

und weil ϵ beliebig ist, konvergiert S_n tatsächlich gegen eine Normalverteilung. Damit ist der erste Teil bewiesen.

Für den zweiten Teil bezeichnen wir die charakteristische Funktion von X_{nj} mit ϕ_{nj} und schätzen ab:

$$\phi_{nj}(t) \geq \Re\phi_{nj}(t) = \mathbb{E}(\cos(X_{nj}t)).$$

Mit den Ungleichungen

$$\cos(x) \geq 1 - (1 - \delta(epsilon))x^2/2(|x| > \epsilon).$$

mit $\delta(\epsilon) > 0$ für $\epsilon > 0$ erhalten wir

$$\mathbb{E}(\cos(X_{nj}) \geq 1 - \frac{1}{2}\mathbb{E}(X_{nj}^2) + \frac{\delta(\epsilon)}{2}\mathbb{E}(X_{nj}^2[|X_{nj}| \geq \epsilon]).$$

Das ergibt

$$-\frac{1}{2} = \lim_n \sum_j \log|\phi_{nj}(1)) \geq -\frac{1}{2} + \frac{\delta(\epsilon)}{2} \limsup_n \sum_j \mathbb{E}(X_{nj}^2[|X_{nj}| \geq \epsilon]).$$

Damit diese Ungleichung stimmt, muss der letzte Term 0 sein, und so ergibt sich die Gültigkeit der Lindebergbedingung.

12.4.1 Lokale Grenzwertsätze

Eine der ersten Versionen des zentralen Grenzwertsatzes ist das Gesetz von deMoivre-Laplace, das im wesentlichen besagt, dass für die Binomialwahrscheinlichkeiten gilt:

$$\mathbb{P}(X = np + x\sqrt{np(1-p)} \approx \frac{1}{\sqrt{2\pi np(1-p)}}e^{-x^2/2}.$$

Diese Art von Grenzwertsätzen wird uns jetzt beschäftigen.

Am einfachsten ist der Fall einer diskreten (genauer: gitterförmigen) Verteilung wie der Binomialverteilung:

Satz 12.10

Die Verteilung von X sei gitterförmig, d.h., X nimmt nur Werte der Form $a + kb$ an, und b sei der größte Wert, für den diese Darstellung möglich ist. Weiters sei $\mathbb{E}(X) = \mu$, $\mathbb{V}(X) = \sigma^2$. Dann gilt für x von der Form $na + kb$ gleichmäßig für $|x - n\mu| \leq C\sqrt{n}$

$$\mathbb{P}(S_n = x) = \frac{b}{\sqrt{2\pi n\sigma^2}} \exp(-\frac{(x-n\mu)^2}{2n\sigma^2})(1 + o(1)).$$

Wir nehmen ohne Beschränkung der Allgemeinheit $a = 0$ und $b = 1$ an. Die charakteristische Funktion einer solchen diskreten Verteilung hat die Form

$$\sum_{k \in Z} \mathbb{P}(X = k) e^{ikt}.$$

Multiplikation mit e^{-ixt} und ein bissel Fubini gibt

$$\mathbb{P}(X = x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ixt} \phi(t) dt.$$

Für S_n ergibt das

$$\mathbb{P}(S_n = x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ixt} \phi(t)^n dt.$$

Die Dichte der Normalverteilung ergibt sich aus der Umkehrformel

$$\frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(x-n\mu)^2}{n\sigma^2}} = \int_{-\infty}^{\infty} e^{-ixt} e^{in\mu t - n\sigma^2 t^2/2} dt.$$

Wir erhalten daraus

$$|\mathbb{P}(S_n = x) - \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(x-n\mu)^2}{n\sigma^2}}| \leq 2(\int_0^{\pi/2} |\phi(t)^n - e^{in\mu t - n\sigma^2 t^2/2}| dt + \int_{\pi}^{i} nfty |e^{in\mu t - n\sigma^2 t^2/2}| dt) =$$

Für das zweite Integral erhalten wir (indem wir noch einen Faktor $nt\sigma^2$ hineinstellen, der für hinreichend großes n größer als 1 ist), obere Abschätzung

$$e^{-n\pi^2\sigma^2/2}$$

. Für $\epsilon \leq |t| \leq \pi$ ist $|\phi(t)|$ von 1 weg beschränkt, und daher geht das erste Integral über diesen Bereich ebenfalls exponentiell schnell gegen 0.

Den Bereich $|t| < \epsilon$ teilen wir weiter in $0 \leq |t| \leq M/\sqrt{n}$ und $M/sqrtn < |t| < \epsilon$. Für $M/sqrtn < |t| < \epsilon$ haben wir die Abschätzung

$$|\phi(t)| \leq e^{-cx^2}$$

mit $c \geq 0$. Wird M groß genug gewählt, kann dieses Integral kleiner als

$$\frac{\epsilon}{\sqrt{n}}$$

gemacht werden. Im Bereich von 0 bis $M/\sqrt{n}$ substituieren wir t durch $u/\sqrt{n}$. Weil $e^{-in\mu t}\phi(u/\sqrt{n})^n$ für $|u| < M$ gleichmäßig gegen $e^{-\sigma^2 t^2/2}$ strebt, wird auch dieses Integral für hinreichend großes n kleiner als $\epsilon/\sqrt{n}$. Damit ist der Satz bewiesen.

Praktisch dasselbe Argument kann benutzt werden, um die folgende stetige Version zu beweisen:

Satz 12.11

(X_n) sei ein Folge von unabhängigen Zufallsvariablen mit charakteristischer Funktion $\phi \in \mathcal{L}_1$, Erwartungswert 0 und Varianz 1. Dann ist die Verteilung von S_n absolutstetig, und für ihre Dichte f_n gilt

$$f_n(x) = \frac{1}{\sqrt{2\pi n}} e^{-x^2/n} + o(\frac{1}{\sqrt{n}}).$$

Zum Beweis sei nur gesagt, dass wir jetzt über ganz R integrieren müssen, deswegen fordern wir die Integrierbarkeit von ϕ, die praktischerweise auch gleich die Absolutstetigkeit der Verteilungen impliziert. Es würde auch $\phi \in \mathcal{L}_p$ reichen, dann ist die Existenz der Dichten allerdings nur für $n \geq p$ gesichert. Für die exponentielle Abschätzung der "Enden" benötigen wir noch

Satz 12.12: Lemma von Riesz

Wenn die Verteilung von X absolutstetig ist, dann gilt

$$\lim_{t\to\infty} \phi_X(t) = 0.$$

Das bedeutet, dass für integrierbares f

$$\int f(x)e^{itx}d\lambda(x) \to 0$$

gilt. Wir können für jedes $\epsilon > 0$ eine Treppenfunktion $g = \sum_{j=1}^{n} g_j]a_j, b_j]$ finden mit $\|f - t\| < \epsilon$. Wegen

$$|\int e^{itx}g(x)d\lambda(x)| \leq \frac{2\sum_j g_j}{t}$$

folgt die Behauptung.

12.5 Mehrdimensionale charakteristische Funktionen

Falls $X = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor ist, definieren wir:

Definition 12.6

Die charakteristische Funktion des Zufallsvektors X ist gegeben durch

$$\phi_X(t) = \phi_X(t_1, \ldots, t_n) = \mathbb{E}(e^{i(t_1X_1+\ldots+t_nX_n)}).$$

Die n-dimensionale charakteristische Funktion bestimmt wieder eindeutig die Verteilung von X, und auch die meisten anderen Resultate aus den vorigen Abschnitten haben Analoga für den n-dimensionalen Fall. Wir geben hier nur (ohne Beweise) zwei Beispiele an:

Satz 12.13

Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn für die gemeinsame charakteristische Funktion gilt:

$$\phi_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \phi_{X_1}(t_1) * \ldots * \phi_{X_n}(t_n).$$

(Um die Unabhängigkeit zu beweisen, genügt es, zu zeigen, dass sich die gemeinsame charakteristische Funktion in n Faktoren zerlegen lässt, von denen jeder nur von einem der t_k abhängt.)

Wir definieren jetzt die n-dimensionale Normalverteilung wie folgt:

Definition 12.7

Der Vektor X ist normalverteilt, wenn es unabhängig standard-normalverteilte Zufallsvariable $Y1, \ldots, Y_k$ und Zahlen α_{jl} $(j = 1, \ldots, n; l = 1, \ldots, k)$ und μ_j $(j = 1, \ldots, n)$ gibt, sodass

$$X_j = \mu_j + \sum_{l=1}^{k} \alpha_{jl}Y_l.$$

Dann gilt der folgende Satz

Satz 12.14

X ist genau dann normalverteilt, wenn für die charakteristische Funktion

$$\phi_X(t) = e^{i\mu t - t^T\Sigma t/2}$$

gilt. Dabei bezeichnet ^{T} die Transposition, Σ ist eine positiv (semi)definite Matrix, die Kovarianzmatrix von X (d.h., $\mathbf{Cov}(X_j, X_l) = \Sigma_{jl}$), und t ist als Spaltenvektor, μ als Zeilenvektor zu lesen.

Falls Σ positiv definit ist, hat X die n-dimensionale Dichte

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp(-\frac{(x^T - \mu)\Sigma^{-1}(x - \mu^T)}{2}).$$

12.6 Wiederholungsfragen

1. Wie ist die schwache Konvergenz von Wahrscheinlichkeitsmaßen definiert?
2. Wie ist die Konvergenz in Verteilung definiert?
3. Geben Sie äquivalente Bedingungen zur schwachen Konvergenz an.
4. Wie ist die charakteristische Funktion einer Verteilung definiert?
5. Eigenschaften der charakteristischen Funktion.
6. Zusammenhang von schwacher Konvergenz und Konvergenz der charakteristischen Funktionen.
7. Was besagt der Darstellungssatz von Riesz?
8. Was besagt der zentrale Grenzwertsatz?
9. Der Satz von Lindeberg-Feller.
10. Was ist eine Dreiecksfolge?
11. Wann nennen wir eine Dreiecksfolge gleichmäßig vernachlßsigbar?
12. Lokale Grenzwertsätze.

Kapitel 13

Momente und die Momentenerzeugende

13.1 Definitionen

Definition 13.1

X sei eine reelle Zufallsvariable.

$$M_n = \mathbb{E}(X^n)$$

heißt das n-te Moment von X,

$$M_n^* = \mathbb{E}(|X|^n)$$

das n-te absolute Moment,

$$m_n = \mathbb{E}((X - \mathbb{E}(X))^n)$$

das n-te zentrale Moment.

$$M_X(t) = \mathbb{E}(e^{Xt})$$

ist die momentenerzeugende Funktion (kurz "Momentenerzeugende") von X.

$$K_X(t) = \log(M_X(t))$$

ist die Kumulantenerzeugende von X. Für eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ sprechen wir auch von den Momenten bzw. der Momentenerzeugenden von $\mathbb{P}$ und meinen damit die Momente einer Zufallsvariable, die nach $\mathbb{P}$ verteilt ist.

Der Name "Momentenerzeugende" erklärt sich aus der formalen Entwicklung

$$M_X(t) = \sum_{n \geq 0} M_n \frac{t^n}{n!}.$$

Aus dieser Entwicklung erhält man eine Entwicklung für K_X:

$$K_X(t) = \sum_{n \geq 0} \kappa_n \frac{t^n}{n!}.$$

Die Koeffizienten κ_n in dieser Entwicklung heißen die Kumulanten von X. Diese lassen sich für $n \geq 2$ durch die zentralen Momente ausdrücken und haben die Eigenschaft, dass für unabhängige Zufallsvariable X und Y

$$\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y)$$

gilt. Die ersten Kumulanten sind

$$\kappa_2 = m_2, \kappa_3 = m_3, \kappa_4 = m_4 - 3m_2^2, \kappa_5 = m_5 - 10m_3m_2.$$

Wenn die Reihe für M_X positiven Konvergenzradius hat (also $\limsup |M_n|^{1/n}/n < \infty$), dann kann man diese Reihe auch tatsächlich zur Berechnung von M_X verwenden, und weil die charakteristische Funktion nichts anderes ist als die Momentenerzeugende mit imaginärem Argument, haben wir damit schon eine Lösung für das sogenannte Momentenproblem — die Frage, ob durch die Momente eine Verteilung eindeutig festgelegt wird.

Es gilt der Satz

Satz 13.1

Wenn die Momente von X die Beziehung

$$\limsup |M_n|^{1/n}/n < \infty$$

erfüllen, dann legen Sie die Verteilung von X eindeutig fest.

Der Beweis ist im vorigen Absatz schon skizziert worden.

Ein Gegenbeispiel ist durch die logarithmische Normalverteilung gegeben: X heißt logarithmisch Normalverteilt, wenn $Y = \log(X)$ normalverteilt ist. Wenn wir eine Standardnormalverteilung voraussetzen, dann hat x die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}x} e^{-\log(x)^2/2}.$$

Die Momente von X sind

$$M_n = \mathbb{E}(X^n) = \mathbb{E}(e^{nY}) = e^{n^2/2}.$$

Wir betrachten die Funktion

$$g(x) = f(x)\sin(2k\pi \log(x)).$$

Es gilt

$$\int x^n g(x)dx = \int \frac{e^{ny}\sin(2k\pi y)}{\sqrt{2\pi}} e^{-y^2/2}dy =$$

$$\Re(\int \frac{e^{(n+2k\pi i)y - y^2/2}dy}{\sqrt{2\pi}} = \Re(e^{(n+2k\pi i)^2/2} = 0.$$

Wir können also feststellen, dass $f + cg$ für $|c| \leq 1$ (dann ist die Summe nichtnegativ) Dichte einer Wahrscheinlichkeitsverteilung ist, die dieselben Momente wie X hat.

13.2 Große Abweichungen

In diesem Abschnitt geht es um die Konvergenzrate im schwachen Gesetz der großen Zahlen. Dieses besagt ja, dass für $x > \mathbb{E}(X)$ $\mathbb{P}(S_n \geq nx) \to 0$ gilt (wie üblich ist $S_n = \sum_{i=1}^n X_i$, für $x < \mathbb{E}(X)$ gilt natürlich $\mathbb{P}(S_n \leq nx) \to 0$).

Wir nehmen an, dass

$$T_1 = \sup\{t : M_X(t) < \infty\} > 0$$

gilt, also dass die Momentenerzeugende von X für gewisse $t > 0$ existiert (genauer gesagt für $0 \leq t < T_1$). Dann existiert der Erwartungswert von X und ist $< \infty$ (kann aber $-\infty$ sein). Das folgt aus der Ungleichung von Jensen. Aus der Ungleichung von Hölder folgt die Ungleichung

$$M_X(\alpha t_1 + (1-\alpha)t_2) \leq M_X(t_1)^\alpha M_X(t_2)^{1-\alpha} (0 \leq \alpha \leq 1),$$

was einerseits zur Folge hat, dass die Menge aller t, für die $M(t) < \infty$ gilt, ein Intervall ist, und andererseits, dass K_X konvex ist. Wenn X nicht degeneriert ist, dann ist K_X sogar strikt konvex.

Eine erste Abschätzung für $\mathbb{P}(S_n \geq nx)$ gibt die Ungleichung von Markov: für $t \geq 0$ gilt

$$\mathbb{P}(S_n \geq nx)s \leq e^{-nxt} M_{S_n}(t) = e^{n(K_X(t) - nt)}$$

(interessant sind natürlich nur die Werte von t, für die $M_X(t)$ endlich ist). Unter diesen Abschätzungen wählen wir natürlich die kleinste und definieren:

Definition 13.2

Die Funktion

$$\rho_X(x) = \sup\{xt - K_X(t) : t \in \mathbb{R}\}$$

heißt die Chernov-Funktion von X.

Damit können wir sagen, dass für $x \geq \mathbb{E}(X)$ gilt:

$$\mathbb{P}(S_n \geq nx) \leq e^{-n\rho(x)}.$$

Das einzige, was daran noch zu beweisen ist, ist dass der Verzicht auf die Einschränkung $t \geq 0$ nichts ändert. Dazu stellen wir fest, dass aus der Ungleichung von Hölder folgt, dass K_X konvex ist. Deshalb ist $m(t) = K_X'(t)$ monoton nichtfallend und $\sigma^2(t) = m'(t)$ nichtnegativ (sogar positiv und damit K_X streng monoton, wenn X nicht degeneriert ist). Wir können

$$xt - K_X(t)$$

maximieren, indem wir nach t ableiten und die Ableitung gleich 0 setzen. Das gibt die Gleichung

$$m(t) = x.$$

Weil $m(0) = \mathbb{E}(X)$ gilt, muss $t > 0$ sein, wenn $x > \mathbb{E}(X)$ gilt. Für $x < \mathbb{E}(X)$ gilt

$$\mathbb{P}(S_n \leq nx) \leq e^{-n\rho(x)}.$$

Damit haben wir schon die Hälfte des folgenden Satzes bewiesen:

Satz 13.2

$M_X(t)$ sei für $0 \leq t < T_1$ endlich. Dann ist für $\mathbb{E}(X) < x < \lim_{t\uparrow T_1} m(t)$

$$\lim_{n\to\infty} \frac{1}{n} \log(\mathbb{P}(S_n \geq nx)) = -\rho(x).$$

Für den Beweis verwenden wir

Definition 13.3: Konjugierte Verteilungen

X sei eine Zufallsvariable mit Verteilung $\mathbb{P}_X$. Wir setzen

$$I = \{t : M_X(t) < \infty\}.$$

Wir nennen die Verteilungen $\mathbb{P}_{X;t}, t \in I$ mit

$$\mathbb{P}_{X;t}(A) = \frac{1}{M_X(t)} \int_A e^{tx} \mathbb{P}_X(dx)$$

die zu $\mathbb{P}_X$ konjugierten Verteilungen oder die Exponentialfamilie von $\mathbb{P}_X$.

$\mathbb{P}_{X;t}$ hat Erwartungswert

$$m(t) = K_X'(t)$$

und Varianz

$$\sigma^2(t) = m'(t),$$

wie man leicht nachrechnet.

Sind $Y_1, \ldots, Y_n$ unabhängige Zufallsvariable mit Verteilung $\mathbb{P}_{X;t}$ und $X_1, \ldots, X_n$ ebenfalls unabhängig mit Verteilung $\mathbb{P}_X$, dann hat die gemeinsame Verteilung von $Y_1, \ldots, Y_n$ (das Produktmaß $P_{X;t}^n$) bezüglich der gemeinsamen Verteilung von $X_1, \ldots, X_n$ die Dichte

$$\frac{1}{M_X(t)^n} e^{(x_1+\ldots+x_n)t}.$$

Insbesondere ist auch die Verteilung von $T_n = Y_1 + \ldots + Y_n$ absolutstetig bezüglich der von $S_n = X_1 + \ldots + X_n$ mit Dichte

$$\frac{d\mathbb{P}_{T_n}}{d\mathbb{P}_{S_n}}(s) = \frac{e^{ts}}{M_{S_n}(t)},$$

also ist

$$\mathbb{P}_{T_n} = \mathbb{P}_{S_n;t}.$$

Daraus erhalten wir

$$\mathbb{P}(S_n \geq nx) = M_X(t)^n \int_{[nx,\infty[} e^{-ut}\mathbb{P}_{T_n}(dx).$$

Wir wählen t so, dass $m(t) = x$ gilt. Y_n hat Erwartungswert $m(t) = x$ und Varianz $\sigma^2(t)$. Der zentrale Grenzwertsatz ergibt, dass T_n näherungsweise normalverteilt ist mit Mittel nx und Varianz $n\sigma^2(t)$. Für großes n ist damit

$$\mathbb{P}(nx \leq T_n \leq nx + 2\sqrt{n\sigma^2(t)}) \geq 0.4$$

und damit

$$\mathbb{P}(S_n \geq nx) \geq 0.4M(t)^n e^{-nt-2\sqrt{n\sigma^2(t)}} = 0.4e^{-n\rho(x)-2\sqrt{n\sigma^2(t)}}.$$

Der Logarithmus der rechten Seite, geteilt durch n, konvergiert gegen $-\rho$, und damit ist Satz 13.2 bewiesen.

Wenn man eine genauere Näherung für die Verteilungsfunktion hat (Stichwort Berry-Esseen bzw. Edgeworth), dann kann man diese Wahrscheinlichkeiten noch genauer bestimmen. Wir können hier für gitterförmige Verteilungen die Methode der konjugierten Verteilungen mit dem lokalen Grenzwertsatz verbinden, um die exakte Asymptotik für $\mathbb{P}(S_n = nx)$ zu bestimmen. Das ist hinreichend einfach, dass wir es an die Übungen delegieren können.

Die weniger genaue logarithmische Version gilt sehr viel allgemeiner.

Satz 13.3

(X_n) sei eine Folge von Zufallsvariablen, (a_n) eine positive Zahlenfolge mit $a_n \to \infty$; wenn für $0 \leq t_0 < t_1$ der Grenzwert

$$K(t) = \lim_{n\to\infty} \frac{1}{a_n} K_{X_n}(t)$$

existiert und stetig differenzierbar ist, dann gilt für $K'(t_0) < x < K'(t_1)$

$$\lim_{n\to\infty} \frac{1}{a_n}\mathbb{P}(X_n > a_n x) = -\rho(x)$$

mit

$$\rho(x) = \sup_{t_0<t<t_1} xt - K(t).$$

Die obere Abschätzung ist einfach: die Markov-Ungleichung liefert wieder

$$\mathbb{P}(X_n > a_n(x)) \leq e^{K_{X_n}(t)-a_n xt},$$

also

$$\frac{1}{a_n}\log(\mathbb{P}(X_n > a_n x)) \leq \frac{1}{a_n} K_{X_n}(t) - xt,$$

und mit $n \to \infty$ ergibt sich für $t_0 < t < t_1$

$$\limsup_{n\to\infty} \frac{1}{a_n}\log(\mathbb{P}(X_n > a_n x)) \leq K(t) - xt.$$

Die beste obere Abschätzung ergibt sich als das Minimum der rechten Seite, und das ist genau $-\rho(x)$.

Für die untere Abschätzung stellen wir zuerst fest, dass K als Limes von konvexen Funktionen konvex ist. Unsere Voraussetzungen über x haben zur Folge, dass es ein $t \in]t_0, t_1[$ gibt mit $\rho(x) = xt - K(t)$.

Die Beziehung zwischen K und ρ ist bekannt als die

Definition 13.4: Legendre-Transformation

$f :]t_0, t_1[$ sei konvex. Bekanntlich existieren in allen Punkten die einseitigen Ableitungen D_-f und D_+f. Wir setzen

$$x_0 = \lim_{t \downarrow t_0} D_+ f(t)$$

$$x_1 = \lim_{t \uparrow t_1} D_+ f(t)$$

und für $x_0 < x < x_1$

$$Lf(x) = \sup_{t_0 < t < t_1} (xt - f(x)).$$

Wir nennen Lf die Legendre-Transformation von f.

Hilfssatz 13.1

Lf ist konvex. Für $x \in]x_0, x_1[$ gibt es (mindestens) ein $t \in]t_0, t_1[$ mit $Lf(x) = xt - f(t)$. Jedes solche t erfüllt

$$D_-f(t) \leq x \leq D_+f(t)$$

und

$$D_-Lf(x) \leq t \leq D_+Lf(x).$$

Wenn f stetig differenzierbar ist, dann ist Lf streng konvex, wenn f streng konvex ist, dann ist Lf stetig differenzierbar.

Um die Konvexität von Lf zu zeigen, wählen wir x', x'', und $\alpha \in [0, 1]$ und setzen $x = \alpha x' + (1 - \alpha)x''$. Für $M < Lf(x)$ können wir ein t wählen, sodass $xt - f(t) \geq M$ gilt. Nach der Definition von Lf gilt

$$Lf(x') \geq x't - f(t)$$

und

$$Lf(x'') \geq x''t - f(t)$$

und damit

$$\alpha Lf(x') + (1 - \alpha)Lf(x'') \geq (\alpha x' + (1 - \alpha)x'')t - f(t) \geq M.$$

Da $M < Lf(x)$ beliebig gewählt werden kann, erhalten wir

$$\alpha Lf(x') + (1 - \alpha)Lf(x'') \geq Lf(x) = L(\alpha x' + (1 - \alpha)x'').$$

Lf ist also konvex (und wir haben die Konvexität von f nicht benötigt). Wenn f konvex ist, dann wird das Supremum entweder im Inneren des Intervalls $]t_0, t_1[$ angenommen, oder als Limes erreicht, wenn t gegen einen der Randpunkte strebt. Der letzte Fall kann wegen unserer Wahl von x nicht auftreten, und wir können t aus den Forderungen $D_-(xt - f(t)) \geq 0$ und $D_+(xt - f(t)) \leq 0$ bestimmen. Das gibt uns die Ungleichungen für x (und wir sehen, dass jedes solche t das Minimum gibt). Weil

$$Lf(x) = xt - f(t)$$

und für jedes x'

$$f(x') \geq x't - f(t)$$

ist $x't - f(t)$ eine Stützgerade von f in x, und das gibt die Ungleichungen für t. Zusammen sagen diese Ungleichungen, dass die Ableitungen von f und Lf (verallgemeinerte) Inverse voneinander sind; wenn Df streng monoton ist, dann ist DLf stetig und umgekehrt wenn Df stetig ist, dann ist f streng monoton.

Um den zweiten Teil von Satz 13.3 zu beweisen, wählen wir $t(x)$ so, dass $\rho(x) = tx - K(t)$ ist, etwa $t(x) = D_+\rho(x)$. Wegen der strikten Monotonie von ρ, die aus der geforderten stetigen Differenzierbarkeit von K folgt, ist t streng monoton. Dann wählen wir $x \in I =]K'(t_0), K'(t_1)[$ und $\epsilon > 0$ so, dass $x - \epsilon$ und $x + \epsilon$ ebenfalls in I liegen.

Wegen der strikten Konvexität von ρ, die aus der stetigen Differenzierbarkeit von K folgt, ist

$$\rho(x-\epsilon) > \rho(x) - \epsilon t(x)$$

und

$$\rho(x+\epsilon) > \rho(x) + \epsilon t(x).$$

Wir wählen δ so, dass

$$\rho(x-\epsilon) > \rho(x) + \delta - \epsilon t(x)$$

und

$$\rho(x+\epsilon) > \rho(x) + \delta + \epsilon t(x).$$

Durch partielle Integration erhalten wir

$$M_n(t(x)) = e^{K_n(t(x))} = t(x)\int_{-\infty}^{\infty} e^{ut(x)}\mathbb{P}(X_n \geq u)du.$$

Wir betrachten als erstes das Integral

$$\int_{-\infty}^{a_n(x-\epsilon)} e^{ut(x)}\mathbb{P}(X_n \geq u)du.$$

Wir schätzen ab

$$\mathbb{P}(X_n \geq u) \leq e^{K_n(t(x-\epsilon))-ut(x-\epsilon)}.$$

Das ergibt

$$\int_{-\infty}^{a_n(x-\epsilon)} e^{ut(x)}\mathbb{P}(X_n \geq u)du \leq \frac{1}{t(x)-t(x-\epsilon)} e^{a_n(x-\epsilon)(t(x)-t(x-\epsilon))+K_n(t(x-\epsilon))}.$$

Für hinreichend großes n gilt $K_n(t(x)) > a_n(K(t(x))-\delta/4)$ und $K_n(t(x-\epsilon) < a_n(K(t(x-\epsilon))+\delta/4)$.

Daraus folgt

$$t(x)\int_{-\infty}^{a_n(x-\epsilon)} e^{ut(x)}\mathbb{P}(X_n \geq u)du \leq \frac{t(x)}{t(x)-t(x-\epsilon)} e^{K_n(t(x))+a_n(\delta/2+K(t(x-\epsilon))-K(t(x))+(x-\epsilon)(t(x)-t(x-\epsilon)))} =$$

$$\frac{t(x)}{t(x)-t(x-\epsilon)} e^{K_n(t(x))+a_n(\delta/2-\rho(x-\epsilon)+\rho(x)-\epsilon t(x)} \leq \frac{t(x)}{t(x)-t(x-\epsilon)} e^{K_n(t(x))-a_n\delta/2}.$$

Für hinreichend großes n ist das kleiner als etwa $e^{K_n(t(x))}/4$. Das Integral von $a_n(x+\epsilon)$ bis ∞ kann ebenso abgeschätzt werden. Dadurch ergibt sich

$$t(x)\int_{a_n(x-\epsilon)}^{a_n(x+\epsilon)} e^{ut(x)}\mathbb{P}(X_n \leq u)du \geq \frac{1}{2}e^{K_n(t(x))}.$$

Der Integrand lässt sich trivial abschätzen:

$$2t(x)a_n\epsilon e^{a_n(x+\epsilon)t(x)}\mathbb{P}(X_n \geq a_n(x-\epsilon)) \geq \frac{1}{2}e^{K_n(t(x))}.$$

Es gilt also

$$\liminf_{n\to\infty} \frac{1}{n} \log(P(X_n \geq a_n(x-\epsilon))) \geq K(t(x)) - (x+\epsilon)t(x) = -\rho(x) - \epsilon t(x).$$

Wenn wir jetzt x durch $x+\epsilon$ ersetzen und dann ϵ gegen 0 gehen lassen, erhalten wir

$$\liminf_{n\to\infty} \frac{1}{n} \log(P(X_n \geq a_n(x-\epsilon))) \geq -\rho(x).$$

Damit ist Satz 13.2 bewiesen.

13.3 Das Gesetz vom iterierten Logarithmus

Dieser Satz bildet zusammen mit dem Gesetz der großen Zahlen und dem zentralen Grenzwertsatz die drei großen Sätze der klassischen Wahrscheinlichkeitstheorie.

Satz 13.4: Gesetz vom iterierten Logarithmus

(X_n) sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Erwartungswert 0 und Varianz 1. Dann gilt

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log(n)}} = 1.$$

Wie es die Tradition verlangt (und damit ein Brocken im Prüfungsformat daraus wird), führen wir den Beweis zuerst für den Sonderfall, dass X_n standardnormalverteilt ist. In diesem Fall ist auch S_n normalverteilt. Wir benötigen die folgenden Ungleichungen für die Normalverteilung:

Hilfssatz 13.2

Für $x > 0$ gilt

$$\frac{1}{\sqrt{2\pi}x}(1-\frac{1}{x^2})e^{-x^2/2} \leq 1-\Phi(x) \leq \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}.$$

Zum Beweis integrieren wir partiell:

$$\int_x^\infty \frac{1}{u}ue^{-u^2/2}du = \frac{1}{x}e^{-x^2/2} - \int_x^\infty \frac{1}{u^2}e^{-u^2/2}du =$$

$$(\frac{1}{x}-\frac{1}{x^3})e^{-x^2/2} + 3\int_x^\infty \frac{1}{u^4}e^{-u^2/2}du.$$

Insbesondere ist für jedes $\delta > 0$

$$1-\Phi(x) \geq e^{-(1+\delta)x^2/2},$$

wenn x groß genug ist. Die zweite Ungleichung ist vom Typ der Ungleichung von Kolmogorov:

Hilfssatz 13.3

$(X_i, i = 1,\ldots,n)$ seien unabhängige Zufallsvariable mit $\mathbb{E}(X_i) \geq 0$. Dann gilt

$$\mathbb{P}(\max_{0\leq k\leq n} S_k \geq x) \leq \inf_{t\geq 0} M_{S_n}(t)e^{-xt}.$$

Wenn X_i standardnormalverteilt ist, dann gilt

$$\mathbb{P}(\max_{0\leq k\leq n} S_k \geq x) \leq e^{-\frac{x^2}{2n}}.$$

Der Beweis verwendet dieselbe Idee wie die Ungleichung von Kolmogorov: Wir setzen

$$A_i = [S_i \geq x, S_j < x, j < i].$$

Damit ist

$$[\max_{i\leq n} S_i \geq x] = \sum_{i\leq n} A_i,$$

und

$$\mathbb{E}(e^{tS_n}) \geq \sum_{i\leq n}\mathbb{E}(A_i e^{tS_n}) = \sum_{i\leq n}\mathbb{E}(A_i e^{tS_i}e^{t(S_n-S_i)}) =$$

$$\sum_{i\leq n}\mathbb{E}(A_i e^{tS_i})E(e^{t(S_n-S_i)}) \geq \sum_{i\leq n}\mathbb{E}(A_i e^{tx}) = e^{tx}\mathbb{P}(\max_{i\leq n} S_n \geq x).$$

Wir können jetzt den Satz vom iterierten Logarithmus beweisen, was auf die beiden folgenden Punkte hinausläuft:

1. Mit Wahrscheinlichkeit 1 ist für jedes $\epsilon > 0$ für fast alle n

$$S_n \leq (1+\epsilon)\sqrt{2n\log\log(n)}.$$

2. Mit Wahrscheinlichkeit 1 ist für jedes $\epsilon > 0$ für unendlich viele n

$$S_n \geq (1-\epsilon)\sqrt{2n\log\log(n)}.$$

Für den ersten Teil wählen wir $\theta > 1$ und setzen $n_k = \theta^k$ und

$$A_k = [\max_{i<n_k} S_i \geq \sqrt{(1+\epsilon)2n_{k-1}\log\log(n_{k-1})}].$$

Aus unserem zweiten Hilfssatz folgt

$$\mathbb{P}(A_k) \leq \exp(-(1+\epsilon)\log\log(n_k)) = (k\log(\theta))^{-(1+\epsilon)/\theta}.$$

Daher ist

$$\sum_k \mathbb{P}(A_k) < \infty,$$

wenn $\theta < 1+\epsilon$, und nach dem Borel-Cantelli Lemma treten mit Wahrscheinlichkeit nur endlich viele der Ereignisse A_k auf. Für $n_{k-1} \leq n \leq n_k$ ist für genügend großes k

$$S_n \leq \max_{i<n_k} S_i \leq \sqrt{(1+\epsilon)2n_{k-1}\log\log(n_{k-1})}\sqrt{(1+\epsilon)2n\log\log(n)}.$$

Für den zweiten Teil wählen wir wieder $n_k = \theta^k$, nur wollen wir jetzt θ nicht mehr nahe bei 1 sondern sehr groß wählen. Wir definieren die Ereignisse

$$A_k = [S_{n_k} - S_{n_{k-1}} \geq (1-\epsilon)\sqrt{2(n_k - n_{k-1})\log\log(n_k)}.$$

Für beliebiges $\delta > 0$ und wenn k hinreichend groß ist, erhalten wir aus dem ersten Hilfssatz

$$\mathbb{P}(A_k) \geq k^{-(1+\delta)(1-\epsilon)^2},$$

und wenn δ hinreichend klein ist, divergiert diese Reihe. Das zweite Lemma von Borel-Cantelli ergibt, dass mit Wahrscheinlichkeit 1 für unendlich viele k

$$S_{n_k} - S_{n_{k-1}} \geq (1-\epsilon)\sqrt{2(n_k - n_{k-1})\log\log(n_k)} \geq (1-\epsilon)\sqrt{2((1-1/\theta)n_k\log\log(n_k)}.$$

Aus dem ersten Teil des Satzes folgt, dass

$$S_{n_{k-1}} \geq -(1+\epsilon)\sqrt{2n_{k-1}\log\log(n_{k-1})} \geq -\frac{1+\epsilon}{\sqrt{\theta}}\sqrt{2n_k\log\log(n_k)}.$$

Insgesamt ergibt sich

$$\frac{S_{n_k}}{\sqrt{2n_k\log\log(n_k)}} \geq (1-\epsilon)\sqrt{1-1/\theta} - \frac{1+\epsilon}{\sqrt{\theta}},$$

und hier kann die rechte Seite beliebig nahe an 1 gemacht werden. Damit ist der Satz bewiesen.

Für den allgemeinen Satz betrachten wir zuerst den Fall, dass die Momentenerzeugende $M_X(t)$ für alle t existiert. Wir können den Beweis von oben wörtlich verwenden, wenn wir nur entsprechende Abschätzungen für die Wahrscheinlichkeiten

$$\mathbb{P}(S_n \geq x\sqrt{n\log\log n})$$

haben, genauer, dass

$$\lim_n \to \infty \frac{1}{\log\log n}\log(\mathbb{P}(S_n \geq x\sqrt{n\log\log n}))\lim_n \to \infty \frac{1}{\log\log n}\log(\mathbb{P}(\max_{k\leq n} S_k \geq x\sqrt{n\log\log n})) = -\frac{x^2}{2}$$

gilt, zumindest in einer Umgebung von $x = \sqrt{2}$. Dazu verhilft uns Satz 13.3. Offensichtlich müssen wir $a_n = \log\log n$ setzen, und für X_n müssen wir dann

$$\sqrt{\frac{\log\log n}{n}} S_n$$

verenden. Die Momentenerzeugende davon ist

$$M_n(t) = M_X(t\sqrt{\frac{\log\log n}{n}})^n.$$

Aus der Reihenentwicklung von M_X erhalten wir

$$M_X(t) = 1 + \frac{t^2}{2} + O(t^3).$$

Für fixes t gilt das

$$M_X(t\sqrt{\frac{\log\log n}{n}} = 1 + \frac{t^2 \log\log n}{2n} + o(\frac{\log\log n}{n}.$$

Daraus folgt

$$K_X(t\sqrt{\frac{\log\log n}{n}} = \frac{t^2 \log\log n}{2n} + o(\frac{\log\log n}{n})$$

und

$$K_n(t) = nK_X(t\sqrt{\frac{\log\log n}{n}} = \frac{t^2 \log\log n}{2} + o(\log\log n),$$

also

$$\frac{1}{\log\log n} K_n(t) \to t^2/2,$$

und

$$\rho(x) = \min_t(tx - t^2/2) = x^2/2.$$

Das ist alles was wir brauchen.

Wenn wir die Existenz der Momentenerzeugenden nicht von vornherein annehmen, dann können versuchen, das Problem auf eines zurückzuführen, in dem die Momentenerzeugende existiert. Die erste Idee ist unser bewährter Trick des Abschneidens, der funktioniert aber nicht ohne Weiteres. Besser funktioniert, die Variablen

$$Y_n = S_n\sqrt{\frac{\log\log n}{n}} - \frac{\log\log n}{n}\sum_{i=1}^{n}(X_i^2 - 1)$$

zu betrachten. Diese sind beschränkt, und mit etwas Arbeit erhält man wieder, dass für $t \geq 0$

$$\frac{1}{\log\log n} K_{Y_n}(t) \to \frac{t^2}{2}$$

gilt.

Unser Beweis von oben zeigt uns dann, dass

$$\limsup(\frac{S_n}{\sqrt{2n\log\log n}} - \frac{\sum_{i=1}^{n}(X_i^2 - 1)}{n}) = 1.$$

Nach dem Gesetz der großen Zahlen geht der zweite Term gegen 0, und wir sind fertig.

13.4 Wiederholungsfragen

1. Definitionen: Momente, Momentenerzeugende, Kumulantenerzeugende.
2. Was versteht man unter "großen Abweichungen"?
3. Wie verhalten sich die Wahrscheinlichkeiten von großen Abweichungen?
4. Das Gesetz vom iterierten Logarithmus?

Kapitel 14

Martingale

Definition 14.1

Eine Filtration ist eine nichtfallende Folge $(\mathfrak{F}_n)$ von Untersigmaalgebren von $\mathfrak{S}$.

Eine Folge von Zufallsvariablen $X_0, X_1, \ldots$ heißt adaptiert (an die Filtrtation $(\mathfrak{F}_n)$), wenn für jedes n X_n $\mathfrak{F}_n$-messbar ist.

Die Filtration $\mathfrak{F}_n = \mathfrak{A}_\sigma(X_0, \ldots, X_n)$ heißt die natürliche Filtration der Folge (X_n).

Wir können $\mathfrak{F}_n$ als die Menge aller Ereignisse interpretieren, über deren Eintreten wir zum Zeitpunkt n Bescheid wissen. Die Adaptiertheit der Folge (X_n) and die Filtration $(\mathfrak{F}_n)$ ist gleichbedeutend damit, dass (F_n) feiner ist als die natürliche Filtration von (X_n).

Definition 14.2

Die Folge $(X_n, n \geq 0)$ heißt

1. $(\mathfrak{F}_n)$-Martingal, wenn sie adaptiert ist, alle X_n integrierbar sind und für $m \leq n$

$$\mathbb{E}(X_n|\mathfrak{F}_m) = X_m,$$

2. $(\mathfrak{F}_n)$-Submartingal, wenn sie adaptiert ist, alle X_n integrierbar sind und für $m \leq n$

$$\mathbb{E}(X_n|\mathfrak{F}_m) \geq X_m,$$

3. $(\mathfrak{F}_n)$-Supermartingal, wenn sie adaptiert ist, alle X_n integrierbar sind und für $m \leq n$

$$\mathbb{E}(X_n|\mathfrak{F}_m) \leq X_m.$$

Sub- Und Supermartingale werden als Semimartingale bezeichnet; wenn klar ist, welche Filtration gemeint ist, kann diese in der Bezeichnung weggelassen werden (im Zweifelsfall ist das die natürliche Filtration; wir werden allerdings in der Folge annehmen, dass wir eine fixe Filtration gewählt haben.

Es gibt einige Operationen, die aus (Semi-) Martingalen wieder Martingale machen:

Satz 14.1

(X_n) und (Y_n) seien Martingale bezüglich derselben Filtration. Dann ist auch $(aX_n + bY_n)$, $a, b \in \mathbb{R}$ ein Martingal.

(X_n) und (Y_n) seien Submartingale bezüglich derselben Filtration. Dann ist auch $(aX_n + bY_n)$, $a, b \geq 0$ ein Submartingal.

Satz 14.2

(X_n) sei ein Martingal und g eine konvexe Funktion und $g(X_n)$ integrierbar. Dann ist $(g(X_n))$ ein Submartingal.

(X_n) sei ein Submartingal und g eine konvexe nichtfallende Funktion und $g(X_n)$ integrierbar. Dann ist $(g(X_n))$ ein Submartingal.

Beispiele für Martingale:

1. (X_n) sei eine Folge von unabhängigen Zufallsvariablen mit Erwartungswert 0. Dann ist $S_n = sum_{i=1}^n X_i$ ein Martingal (bezüglich der natürlichen Filtration).
2. Es sei zusätzlich $X_n \in \mathcal{L}_2$. Dann ist
$$S_n^2 - \mathbb{E}(S_n^2)$$
ein Martingal (bezüglich der natürlichen Filtration von (X_n)i).
3. (F_n) sei eine Filtration, X integrierbar. Dann ist
$$X_n = \mathbb{E}(X|\mathfrak{F}_n)$$
ein Martingal, das Doob'sche Martingal.

Wenn von nun an davon die Rede ist, dass eine Folge (X_n) ein Martingal oder Semimartingal ist, ohne dass die entsprechende Filtration spezifiziert wird, dann soll das bedeuten, dass eine bestimmte Filtration festgelegt wurde

Beispiele für Semimartingale:

1. Klarerweise ist jedes Martingal auch Sub- und Supermartingal, und jede Folge, die gleichzeitig Super- und Submartingal ist, ein Martingal.
2. Eine monoton nichtfallende Folge von Zufallsvariablen ist ein Submartingal, und damit auch die Summe $(A_n + M_n)$, wobei A_n nichtfallend und (M_n)

Definition 14.3

Eine Stoppzeit τ ist eine ganzzahlige nichtnegative Zufallsvariable, für die gilt

$$[\tau \leq n] \in \mathfrak{F}_n \text{ für alle } n.$$

Anschaulich gesprochen beschreibt eine Stoppzeit ein Ereignis, über dessen Eintreten wir Bescheid wissen, wenn wir den Verlauf des Prozesses bis zu diesem Zeitpunkt kennen — es darf also nicht von der Zukunft abhängen. Typische Beispiele sind Übergangszeiten

$$\tau_A = \inf\{n : X_n \in A\}.$$

$\tau_A + c$ ist für $c > 0$ auch eine Stoppzeit, aber nicht für $c < 0$.

Satz 14.3: optional stopping

. (X_n) sei ein Martingal (Submartingal) und T eine Stoppzeit. Dann ist auch

$$Y_n = X(\min(\tau, n))$$

ein Martingal (Submartingal) und es gilt

$$\mathbb{E}(Y_n) = \mathbb{E}(X_n)$$

(wenn (X_n) ein Martingal ist) bzw.

$$\mathbb{E}(Y_n) \leq \mathbb{E}(X_n)$$

(wenn (X_n) Submartingal ist). (Wir verwenden die Notation $X(n)$ für X_n, weil diese für komplexe Argumente besser lesbar ist).

Satz 14.4: Maximalungleichung von Doob

(X_n) sei ein nichtnegatives Submartingal. Dann gilt

$$\mathbb{P}(\max_{i\leq n} X_i \geq \lambda) \leq \frac{1}{\lambda}\mathbb{E}(X_n).$$

Zum Beweis verwenden wir das optional stopping Theorem mit der Stoppzeit

$$\tau = \tau_{[\lambda,\infty)}.$$

Klarerweise gilt

$$[\max_{i\leq n} X_i \geq \lambda] = [\tau \leq n].$$

Dann erhalten wir

$$\mathbb{E}(X_n) \geq \mathbb{E}(X(\min(\tau, n)) \geq \mathbb{E}([\tau \leq n]X(\min(\tau, n)) \geq \mathbb{E}([\tau \leq n]\lambda) = \lambda\mathbb{P}(\tau \leq n).$$

Satz 14.5: optional selection

(X_n) sei ein Submartingal und τ_k und σ_k $(k \in \mathbb{N})$ Stoppzeiten mit $\sigma_k \leq \tau_k \leq \sigma_{k+1}$. Dann ist

$$Y(n) = X(0) + \sum_k (X(\min(n, \tau_k) - X(\min(n, sigma_k)))$$

ein Submartingal mit

$$\mathbb{E}(Y_n) \leq (X_n)$$

Wenn (X_n) ein Martingal ist, dann ist auch (Y_n) ein Martingal, und in der letzten Ungleichung gilt Gleichheit.

Satz 14.6: Upcrossing Ungleichung

(X_n) sei ein Submartingal mit $\sup_n \mathbb{E}((X_n)_+) < \infty$. N_{ab} sei die Anzahl der Durchquerungen des Intervalls $[a, b]$ durch X_n, also

$$N_{ab} = \sup\{N : \exists s_1 < t_1 < s_2 < \ldots < s_N < t_n : X(s_n) \leq a, X(t_n) \geq b\}.$$

Dann gilt

$$\mathbb{E}(N_{ab}) \leq \frac{\sup_n \mathbb{E}((X_n - a)_+)}{b - a}.$$

Satz 14.7: Martingal-Konvergenzsatz

(X_n) sei ein Submartingal mit $\sup_n \mathbb{E}((X_n)_+) < \infty$. Dann existiert

$$X = \lim_{n\to\infty} X_n$$

mit Wahrscheinlichkeit 1.

Wir stellen uns jetzt die Frage, wann ein Martingal als Doobsches Martingal geschrieben werden kann.

Definition 14.4

Ein Martingal (X_n) heißt abschließbar, wenn es eine integrierbare Zufallsvariable X_∞ gibt, sodass

$$X_n = \mathbb{E}(X_\infty|\mathfrak{F}_n).$$

Satz 14.8

Ein Doobsches Martingal ist gleichmäßig integrierbar.

Ein abschließbares Martingal muss also gleichmäßig integrierbar sein. Ist umgekehrt ein Martingal gleichmäßig integrierbar, dann erfüllt es die Voraussetzungen des Konvergenzsatzes, es gilt also

$$X_\infty = \lim_{n\to\infty} X_n.$$

Wegen der gleichmäßigen Integrierbarkeit ist der Grenzübergang mit dem Erwartungswert vertauschbar, und daher gilt

$$\mathbb{E}(X_\infty|\mathfrak{F}_k) = \lim_{n\to\infty} \mathbb{E}(X_n|\mathfrak{F}_k) = X_k.$$

Wir haben also

Satz 14.9

Ein Martingal ist genau dann abschließbar, wenn es gleichmäßig integrierbar ist.

Definition 14.5

Eine Folge X_n heißt vorhersagbar bezüglich der Filtration $(\mathfrak{F}_n)$, wenn für alle n X_n $\mathfrak{F}_{n-1}$-messbar ist (mit $\mathfrak{F}_{-1} = \{\emptyset, \Omega\}$).

Satz 14.10: Doob-Meyer

Jedes Submartingal (X_n) lässt sich in der Form

$$X_n = A_n + M_n$$

schreiben, wobei (A_n) monoton nichtfallend und (M_n) ein Martingal ist. Wird zusätzlich gefordert, dass (A_n) vorhersagbar und $A_0 = 0$ ist, dann ist die Zerlegung eindeutig bestimmt.

14.1 Wiederholungsfragen

1. Definitionen: Filtration, adaptiert, Martingal, Supermartingal, Submartingal, Stoppzeit.
2. Welche Operationen machen aus Martingalen/Submartingalen wieder Submartingale.
3. Optional stopping.
4. Optional selection.
5. Maximalungleichung.
6. Martingal-Konvergenzsatz.
7. Vorhersagbarkeit.
8. Doob-Meyer Zerlegung.

Anhang A

Tables

The distribution function of the normal distribution:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$$

	0	1	2	3	4	5	6	7	8	9
0.0	.500	.504	.508	.512	.516	.520	.524	.528	.532	.536
0.1	.540	.544	.548	.552	.556	.560	.564	.567	.571	.575
0.2	.579	.583	.587	.591	.595	.599	.603	.606	.610	.614
0.3	.618	.622	.626	.629	.633	.637	.641	.644	.648	.652
0.4	.655	.659	.663	.666	.670	.674	.677	.681	.684	.688
0.5	.691	.695	.698	.702	.705	.709	.712	.716	.719	.722
0.6	.726	.729	.732	.736	.739	.742	.745	.749	.752	.755
0.7	.758	.761	.764	.767	.770	.773	.776	.779	.782	.785
0.8	.788	.791	.794	.797	.800	.802	.805	.808	.811	.813
0.9	.816	.819	.821	.824	.826	.829	.831	.834	.836	.839
1.0	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	.873	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.893	.894	.896	.898	.900	.901
1.3	.903	.905	.907	.908	.910	.911	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932
1.5	.933	.934	.936	.937	.938	.939	.941	.942	.943	.944
1.6	.945	.946	.947	.948	.949	.951	.952	.953	.954	.954
1.7	.955	.956	.957	.958	.959	.960	.961	.962	.962	.963
1.8	.964	.965	.966	.966	.967	.968	.969	.969	.970	.971
1.9	.971	.972	.973	.973	.974	.974	.975	.976	.976	.977
2.0	.977	.978	.978	.979	.979	.980	.980	.981	.981	.982
2.1	.982	.983	.983	.983	.984	.984	.985	.985	.985	.986
2.2	.986	.986	.987	.987	.987	.988	.988	.988	.989	.989
2.3	.989	.990	.990	.990	.990	.991	.991	.991	.991	.992
2.4	.992	.992	.992	.992	.993	.993	.993	.993	.993	.994
2.5	.994	.994	.994	.994	.994	.995	.995	.995	.995	.995
2.6	.995	.995	.996	.996	.996	.996	.996	.996	.996	.996
2.7	.997	.997	.997	.997	.997	.997	.997	.997	.997	.997
2.8	.997	.998	.998	.998	.998	.998	.998	.998	.998	.998
2.9	.998	.998	.998	.998	.998	.998	.998	.999	.999	.999

Quantiles z_p of the normal distribution:

p	z_p	p	z_p	p	z_p
.51	.025	.71	.553	.91	1.341
.52	.050	.72	.583	.92	1.405
.53	.075	.73	.613	.93	1.476
.54	.100	.74	.643	.94	1.555
.55	.126	.75	.674	.95	1.645
.56	.151	.76	.706	.96	1.751
.57	.176	.77	.739	.97	1.881
.58	.202	.78	.772	.975	1.960
.59	.228	.79	.806	.98	2.054
.60	.253	.80	.842	.99	2.326
.61	.279	.81	.878	.991	2.366
.62	.305	.82	.915	.992	2.409
.63	.332	.83	.954	.993	2.457
.64	.358	.84	.994	.994	2.512
.65	.385	.85	1.036	.995	2.576
.66	.412	.86	1.080	.996	2.652
.67	.440	.87	1.126	.997	2.748
.68	.468	.88	1.175	.998	2.878
.69	.496	.89	1.227	.999	3.090
.70	.524	.90	1.282	.9999	3.719

Index

Absolute Stetigkeit, 106
adaptiert, 165
Additivität, 13
Äquivalenzsatz von Lévy, 89
äußeres Maß, 34, 35
    arithmetisches, 45
    induziertes, 34
Algebra, 16
Andersen-Jessen
    Existenzsatz, 131
Approximationssatz, 40
    für messbare Funktionen, 67

bedingte Erwartung, 115
bedingte Wahrscheinlichkeit
    reguläre, 116
bedingter Erwartungswert, 114
Beschränktheit in Wahrscheinlichkeit, 88
Blutgruppe, 32

Cauchy-Folge
    im Maß, 71
charakteristische Funktion, 91

Dreireihensatz, 90, 127
Dualraum, 122
Dynkin-System, 14
    erzeugtes, 14

Eichfunktion, 118
Elementarereignis, 68
Ereignis, 68
Erwartungswert
    bedingter, 114
essentielles Supremum, 69
Existenzsatz
    von Kolmogorov, 132
Exponentialfamilie, 158

fast sicher, 69
fast überall, 69
fast überall messbar, 69
Filtration, 165
    natürliche, 165
Fortsetzungssatz, 25, 39
Fubini, 126
Funktion
    charakteristische, 91
Funktion von beschränkter Variation, 104

Gesetz der großen Zahlen
    schwaches, 87
    starkes, 92, 93
gleichmäßig integrierbar, 80

Hahn, 25, 46
Hahn-Zerlegung, 100

induziertes Maß, 68
Inhalt, 14
Integral, 76
    komplexes, 84
integrierbar, 76
    gleichmäßig, 80

Jordan-Zerlegung, 100

Kolmogorov, 127
    Existenzsatz, 132
konjugierte Verteilungen, 158
Konsistenz, 131
Konvergenz
    fast gleichmäßig, 70
    fast sicher, 69
    fast überall, 69
    fast überall gleichmäßig, 69
    im Maß, 70
    in Verteilung, 89
    in Wahrscheinlichkeit, 70
Konvergenz lokal im Maß, 74
Kovarianz, 86

Lévy-Distanz, 72
Lévy-Metrik, 72
Legendre-Transformation, 160
Lemma von Fatou, 79
Lemma von Riesz, 154

Markovkern, 116, 129
Martingal, 165
Maß, 14
    äußeres, 34, 35
    induziertes, 68
Maßfunktion, 14
Maßraum, 41
maßtreu, 68

Menge
    negative, 100
    positive, 100
Mengenfunktion, 12
    additive, 13
    sigmaadditive, 13
Mengensystem, 12
messbar
    fast überall, 69
messbare Menge, 37
Messraum, 41

negative Variation, 102, 104
Nullmenge, 100

Pfeiler, 130
positive Variation, 102, 104
Produkt
    von zwei Sigmaalgebren, 19
Produktmaß, 124
    unendliches, 131
Produktsigmaalgebra, 19, 124, 130
    endlich, 124
    unendlich, 130
Projektion, 129

quasiintegrierbar, 76

Radon-Nikodym Dichte, 106
Randverteilung, 130
    endlichdimensionale, 131
Randverteilungen
    konsistente, 131
Regressionsfunktion, 115
reguläre bedingte Wahrscheinlichkeit, 116
Restriktion, 21
Ring, 16

Saks, 25, 46
Satz
    Beppo-Levi, 77
    Radon-Nikodym, 106
    Vitali-Hahn-Saks, 25, 46
    von der dominierten Konvergenz, 79
    von der monotonen Konvergenz, 77, 79
Satz von Bayes, 32
Satz von der vollständigen Wahrscheinlichkeit, 31
Schnitt, 125
Semimartingal, 165
Sigmaadditivität, 13
Sigmaalgebra, 15
    erzeugte, 15
Sigmaring, 15
    erzeugter, 15
signierte Maßfunktion, 99
singulär, 100

Spur, 21
Standardabweichung, 85
Streuung, 85
Submartingal, 165
Supermartingal, 165

Totalvariation, 102, 104
Treppenfunktion, 66
    kanonische Darstellung, 66

Übergangsfunktion, 116, 129
Unabhängigkeit, 30, 68
    von Ereignissen, 30
    von Zufallsvariablen, 68
Ungleichung
    Kolmogorov, 127
    Lévy-Ottaviani, 87
Ungleichung von Kolmogorov, 88
Unkorreliertheit, 86
Urbild, 20

Varianz, 85
Variationsnorm, 103
Verteilung, 58, 68
    deterministische, 59
    diskrete, 58
    konjugierte, 158
    stetige, 58
Vervollständigung, 40
Vitali, 25, 46
vollständige Sigmaalgebra, 40

Wahrscheinlichkeitsfunktion, 59
Wahrscheinlichkeitsraum, 41
Wohlordnung, 5

Zerlegungssatz von Hahn, 100
Zerlegungssatz von Jordan, 101
Zerlegungssatz von Lebesgue, 108
Zufallsvariable, 68
Zylinder, 130