Ergänzungen (Vorlesung Wise 2021-22)

Michael Weiss

CHAPTER 1

Notizen zur 1. Vorlesungswoche

Der vorläufige Plan ist, dass ich etwas über Masstheorie erzähle, mit geschichtlichen Einlagen soweit möglich, und dann ein paar hübsche Anwendungen oder Beispiele schildere. Masstheorie kommt auch in den "Grundlagenerweiterungen" unseres Bachelorstudiums vor, nämlich in Analysis III und in Stochastik. Überschneidungen mit diesen Kursen lassen sich deswegen nicht ganz vermeiden, und ich finde, das ist in Ordnung so. Andererseits muss ich (will ich) versuchen, Themen und/oder Schwerpunkte zu finden, die nicht schon in Analysis III oder Stochastik gründlich behandelt werden.

1.1. Zusammenhang zwischen Mass und Integral

Masstheorie hat natürlich etwas mit Integrationstheorie zu tun. Es ist nicht ganz dasselbe, aber fast. Es schadet nichts, wenn wir den Zusammenhang genauer beleuchten.

• Angenommen, wir sind Spezialisten der Masstheorie. Das heisst, wir sind kompetent in der Bestimmung, Berechnung oder Festsetzung von *Volumina*, etwa vom Volumen von gewissen Teilmengen von $\mathbb{R}^n$. (Wenn $n = 1$ ist, kann man auch Länge sagen; wenn $n = 2$, auch Fläche; aber wenn $n \geq 3$, dann heisst es n-dimensionales Volumen.) Jetzt kommt jemand daher und möchte ein Integral bestimmt haben. Sagen wir, es handelt sich um das Integral einer Funktion

$$f\colon K \to \mathbb{R}$$

wobei K ein Produkt von abgeschlossenen Intervallen im $\mathbb{R}^{n-1}$ ist, also etwa $K = \prod_{j=1}^{n-1}[a_j, b_j]$. Der Fall ist etwas einfacher, wenn $f(x) \geq 0$ ist für alle $x \in K$. Dann betrachten wir

$$L(f) := \{(x,t) \in \mathbb{R}^{n-1} \times \mathbb{R} \mid x \in K,\ 0 \leq t \leq f(x)\}.$$

Das ist eine Teilmenge von $\mathbb{R}^n$, weil wir $\mathbb{R}^{n-1} \times \mathbb{R}$ mit $\mathbb{R}^n$ gleichsetzen. Wenn wir Glück haben, ist das n-dimensionale Volumen $\mu(L(f))$ von $L(f)$ definiert, bestimmbar usw., und wir sagen dann: *das Integral von f ist $\mu(L(f))$*.

Wenn wir nicht annehmen dürfen, dass $f(x) \geq 0$ für alle $x \in K$, dann können wir immer noch schreiben

$$f = f^+ - f^-$$

wobei $f^+(x) = f(x)$ falls $f(x) \geq 0$, und $f^+(x) = 0$ sonst; ebenso $f^-(x) = -f(x)$ falls $f(x) \leq 0$ und $f^-(x) = 0$ sonst. Dann sind f^+ und f^- Funktionen mit nichtnegativen Werten, und wir können sagen: *das Integral von f is $\mu(L(f^+)) - \mu(L(f^-))$* (falls diese beiden Zahlen definiert sind).

• Angenommen, wir sind Spezialisten der Integrationstheorie. Jetzt kommt jemand daher und möchte ein Volumen bestimmt haben. Sagen wir, es handelt sich um das $(n-1)$-dimensionale Volumen einer beschränkten Teilmenge V von $\mathbb{R}^{n-1}$. Wir finden einen Quader $K = \prod_{j=1}^{n-1}[a_j, b_j]$ wie oben, der V enthält. Dann definieren wir eine Funktion

$$\chi_V\colon K \to \mathbb{R}$$

durch $\chi_V(x) = 1$ falls $x \in V$, dagegen $\chi_V(x) = 0$ falls $x \notin V$ (wobei $x \in K$ sowieso angenommen wird). Man sagt (wie Sie wahrscheinlich wissen), dass χ_V die *charakteristische Funktion* von V ist. Oder auch: *die Indikatorfunktion* von V. Und dann bestimmen wir: *das* $(n-1)$*-dimensionale Volumen von* V *ist das Integral von* χ_V. Dabei ist es egal, wie gross wir den Quader K wählen, solange $V \subset K$ gilt. Es sollte immer dasselbe herauskommen.

Aus dieser Predigt kann man schon ein paar Hinweise mitnehmen. Erstens, in der Masstheorie muss man sich manchmal damit abfinden, dass gewisse Mengen *nicht messbar* sind. Ihr Mass/Volumen lässt sich nicht angeben, wobei sie trotzdem Teilmengen von Mengen sein können, deren Mass/Volumen sich angeben lässt. Sogar im $\mathbb{R}^1$ gibt es beschränkte Teilmengen, deren Länge sich nicht bestimmen lässt. Zweitens, das Mass einer (messbaren) Menge ist nicht unbedingt von der Natur vorgeschrieben; es ist denkbar, dass wir es im Rahmen eines grösseren Unternehmens selber irgendwie definieren. Das ist nicht weiter überraschend.

1.2. Riemann, Cantor, Lebesgue

Warum diese drei? Riemann ist mit seinem Riemann-Integral natürlich ein Begründer der Integrationstheorie. Lebesgue ist aber auch ein wichtiger Begründer einer (neueren) Mass-und Integrationstheorie. Eigentlich wurde er gerade deswegen ein wichtiger Begründer einer neueren Mass-und Integrationstheorie, weil er an der Riemannschen Integrationstheorie etwas auszusetzen hatte. Es gab Unzulänglichkeiten. Aber was waren das für Unzulänglichkeiten? Und wie konnte es dazu kommen, dass Riemann, der doch so schlau war, eine Integrationstheorie entwickelte, die dann von Lebesgue verworfen werden musste? Ein interessantes Drama, in dem auch Cantor eine wichtige Rolle spielt.

Bei Wikipedia lese ich: Bernhard Riemann 1826 -1866. Er wurde nur knapp 40 Jahre alt, weil er an Tuberkulose litt. Weiter lese ich: Cantor 1845-1918, Lebesgue 1975-1941. (Geschichte ist, wenn man auf solche Lebensdaten achtet. Ich muss leider zugeben, dass ich als Schüler nicht gut in Geschichte war. Hat mich gelangweilt.) Eine erste Schlussfolgerung: Riemann konnte von Cantor und seiner Mengenlehre nichts wissen; Lebesgue dagegen sehr wohl. Ich glaube, Lebesgue war auch ein Bewunderer von Cantor und ein Fanatiker für die Mengenlehre. Das ist wichtig. (Dazu vielleicht später mehr.)

In der Masstheorie von Lebesgue spielt eine Forderung der Additivität eine wichtige Rolle. Ich kann sie jetzt nicht in voller Allgemeinheit formulieren, weil wir noch nicht so weit sind. Wenn es um das "gewöhnliche" Mass im $\mathbb{R}^n$ geht, also das n-dimensionale Volumen, hier μ genannt, dann lautet die Forderung so:

- Angenommen, wir haben eine (unendliche) Folge von Teilmengen $A_0, A_1, A_2, ...$ von $\mathbb{R}^n$. Wenn jedes A_i messbar ist (also ein bestimmbares n-dimensionales Volumen hat), dann ist auch die Vereinigung

  $$\bigcup_{i=0}^{\infty} A_i := A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4 \cup \cdots$$

  messbar, und für die Volumina gilt

  $$\mu(\bigcup_{i=0}^{\infty} A_i) \;\leq\; \sum_{i=0}^{\infty} \mu(A_i).$$

  Wenn ausserdem die A_i paarweise disjunkt sind (*besonders wichtiger Fall*), kann man hier ein Gleichheitszeichen setzen.

(In der Vorlesung habe ich die Vermutung geäussert, dass dieser Gedanke von Lebesgue stammt, aber bei Wikipedia steht, dass das nicht stimmt: Emile Borel ist dafür verantwortlich. Also, dieser Borel war auch wichtig für die Masstheorie.) Diese Forderung oder Eigenschaft heisst σ-Additivität. (Warum σ ... ich weiss es leider nicht.) Zu beachten: die rechte Seite der (Un-)gleichung muss nicht konvergieren. Aber in jedem Fall sind die Summanden sämtlich nicht-negativ, und wenn die Reihe divergiert, können wir mit ziemlich gutem Gewissen sagen: die Summe ist $+\infty$. Das ist in der Masstheorie auch so üblich. Interessanter ist es vielleicht, wenn die Summe konvergiert.

Was hätte Riemann dazu gesagt? Denkbar wäre folgende Reaktion.

- "Das geht nicht. Es führt zu einem Widerspruch. Um das zu zeigen, schauen wir uns das Intervall $A = [0,1]$ an. Es hat die Länge 1, also
$$\mu(A) = 1.$$
Andererseits ist A natürlich die (disjunkte) Vereinigung aller einelementigen Teilmengen von A. Diese einelementigen Teilmengen können wir A_0, A_1, A_2, A_3 usw. nennen, indem wir sie irgendwie durchnummerieren. Nach der oben beschriebenen Forderung müsste dann gelten
$$\mu(A) = \sum_{i=0}^{\infty} \mu(A_i).$$
Aber es ist klar, dass $\mu(A_i) = 0$ ist für alle i. Denn eine einelementige Teilmenge von $[0,1]$ kann man ja auch als Intervall der Länge 0 betrachten. Damit erhalten wir $1 = \mu(A) = \sum_{i=0}^{\infty} \mu(A_i) = \sum_{i=0}^{\infty} 0 = 0$. Ein klarer Fall von Widerspruch."

Allerdings vermute ich, dass Riemann nicht so argumentiert hätte. Irgendetwas wäre ihm dabei unangenehm gewesen; er hätte etwas gemerkt. Merken Sie etwas? (Ja, Sie haben etwas gemerkt; es wurde in der Vorlesung ausprobiert.)

1.3. Wozu brauchen wir σ-Additivität?

Die Masstheorie ist sehr wichtig in der Wahrscheinlichkeitstheorie. Seit Kolmogorov beruht die Wahrscheinlichkeitstheorie auf der Masstheorie. Das Rezept von Kolmogorov ist ungefähr so.

- Um ein *Experiment* im Sinn der Wahrscheinlichkeitstheorie zu beschreiben, versuchen wir zuerst, die Menge aller denkbaren *Ergebnisse* des Experiments zu erfassen. Nennen wir diese Menge X. (Diese Aufgabe muss keine eindeutige oder beste Lösung haben.)
- Die *Ereignisse*, von denen man bei so einem Experiment spricht, sollten eigentlich nichts anderes sein, als gewisse Teilmengen der Ergebnismenge X. (Später können wir das noch präzisieren.)
- Die Wahrscheinlichkeiten, die uns interessieren, bilden eine Funktion μ, die jedem Ereignis Y (also einer von den "gewissen" Teilmengen von X) eine reelle Zahl $\mu(Y)$ im Interval $[0,1]$ zuordnen.
- Diese Funktion μ hat allerhand gute Eigenschaften, also die, die man von einer Volumenfunktion erwarten würde. Kurz, μ ist ein Mass, und es hat noch die besondere Eigenschaft $\mu(X) = 1$. Die Wahrscheinlichkeit eines Ereignisses $Y \subset X$ ist demnach einfach das "Volumen" von Y im Sinne des Masses μ.
- Weil wir schon gewarnt worden sind, rechnen wir damit, dass nicht alle Teilmengen von X messbar im Sinn von μ sind. Das heisst, $\mu(Y)$ ist nicht unbedingt für jede Teilmenge Y von X definiert. Auf jeden Fall soll $\mu(X)$ und $\mu(\emptyset)$ definiert sein, und es soll gelten $\mu(X) = 1$, $\mu(\emptyset) = 0$. Wenn $\mu(Y)$ definiert ist, dann sagen wir: Y ist

μ-messbar. Dann können wir auch mit gutem Gewissen sagen: Y ist ein Ereignis, und seine Wahrscheinlichkeit ist eben $\mu(Y)$.

Nun kann es ja sein, dass wir in unserem Experiment, von dem wir glauben, dass es durch X und μ gut beschrieben wird, unendlich viele denkbare Ereignisse Y_0, Y_1, Y_2, Y_3 usw. formulieren können, die einander paarweise ausschliessen. Das bedeutet, dass die $Y_0, Y_1, Y_2, \ldots$ als Teilmengen von X *paarweise disjunkt* sind. Ausserdem bedeutet es auch, dass jedes Y_j μ-messbar ist. Dann kann es sein, dass wir uns für den Fall *es passiert* Y_0 *oder* Y_1 *oder* Y_2 *oder ... usw.* interessieren. Das ist dann, als Teilmenge von X, einfach die Vereinigung

$$\bigcup_{j=0}^{\infty} Y_j \,.$$

Natürlich wünschen wir uns jetzt, dass das auch eine Wahrscheinlichkeit hat, also μ-messbar ist. Ausserdem sollte die Wahrscheinlichkeit von $\bigcup_{j=0}^{\infty} Y_j$ genau die Summe der Wahrscheinlichkeiten der Y_j sein, also

$$\mu(\bigcup_{j=0}^{\infty} Y_j) = \sum_{j=0}^{\infty} \mu(Y_j).$$

Hier sehen wir also die Forderung der σ-Additivität.

Beispiel/Übungsaufgabe. Eine Münze mit zwei gleichberechtigten Seiten (die wir Kopf und Zahl nennen können) soll so oft geworfen werden, bis "Kopf" kommt. Wie gross ist die Wahrscheinlichkeit, dass die Anzahl der Würfe, die dafür benötigt wird (letzter Wurf eingeschlossen) durch 3 teilbar ist?

Und noch eine Frage, die allerdings ein bisschen zu früh kommt. Was ist die *durchschnittliche* Anzahl der Würfe, die man erwarten sollte, bis "Kopf" kommt (letzter Wurf eingeschlossen) ?

CHAPTER 2

Notizen 2. und 3. Vorlesungswoche

2.1. Offene Mengen in der Masstheorie

Topologische Räume (und ihre offenen Mengen) spielen in der Masstheorie eine wichtige Rolle, obwohl sie streng genommen in einem ganz abstrakten Aufbau der Masstheorie nicht gebraucht werden. Es gibt Situationen der folgenden Art.

- Wir haben einen topologischen Raum X mit Topologie $\mathcal{O}$ (das $\mathcal{O}$ ist also eine Auswahl von Teilmengen von X, die wir, ziemlich willkürlich, *offene* Teilmengen von X nennen).
- Wir brauchen oder suchen eine Volumenfunktion μ für (gewisse) Teilmengen V von X. Mit anderen Worten, μ soll ein Mass sein. Demnach sollen also "gewisse" Teilmengen von X als μ-messbar ausgezeichnet werden; welche das sind, muss man (kann man) manchmal schon entscheiden, bevor man festlegt, welche μ-Volumina sie haben sollen.
- Wir bestehen darauf, dass die offenen Teilmengen von X messbar sind.
- Dann ergibt sich allerdings aus den Forderungen der Masstheorie (die ich noch nicht sauber formuliert habe), dass noch viele andere Teilmengen von X messbar sein müssen. So zum Beipiel die abgeschlossenen Mengen, weil sie die Komplemente von offenen Mengen sind; dann auch Durschschnitte von offenen und abgeschlossenen, und viele, viele mehr.

So eine Situation hat man zum Beispiel, wenn man das (gewöhnliche) Lebesgue-Mass λ in $\mathbb{R}^n$ konstruiert, also $X = \mathbb{R}^n$. Es fängt an mit dem Gedanken oder Entschluss, dass das n-dimensionale Volumen $\lambda(V)$ einer offenen Teilmenge V leicht zu definieren sein sollte. (Dabei ist $+\infty$ erlaubt. Wenn V beschänkt ist, sollte nicht $+\infty$ herauskommen.)

Deswegen schlage ich vor, dass wir uns ein paar Eigenschaften von offenen Mengen (in speziellen topologischen Räumen) angucken, die für die Masstheorie interessant sein könnten. Der Fall $\mathbb{R}^1$ (als topologischer Raum mit der üblichen Topologie) ist schon sehr interessant.

2.2. Offene Mengen in $\mathbb{R}^1$ und ihre Länge

PROPOSITION 2.2.1. *Jede offene Menge U in $\mathbb{R}^1$ ist eine disjunkte Vereinigung von höchstens abzählbar vielen offenen Intervallen (die durch U eindeutig bestimmt werden).*

Beweis (ist wahrscheinlich bekannt): Wir führen eine Relation J auf der offenen Menge U ein wie folgt: für $x, y \in U$ soll xJy gelten genau dann, wenn das Intervall $[x, y]$ oder $[y, x]$ in U enthalten ist. Behauptung: J ist eine Äquivalenzrelation. Reflexiv: klarer Fall, xJx gilt für alle $x \in U$. Symmetrisch: nach Definition. Transitiv: ja, denn die Vereinigung von zwei kompakten Intervallen ist wieder ein kompaktes Intervall, wenn die beiden mindestens einen Punkt gemeinsam haben. Also: J ist Äquivalenzrelation. Weiter: jede Äquivalenzklasse ist offen, denn wenn $x \in U$, dann $\exists \varepsilon > 0$ so dass $[x - \varepsilon, x + \varepsilon] \subset U$, und dann sind alle Elemente von $[x - \varepsilon, x + \varepsilon]$ in der Äquivalenzklasse von x. Aus der Definition der Relation J ergibt sich ausserdem, dass jede Äquivalenzklasse ein Intervall ist, und wir wissen jetzt: ein offenes Intervall. Nun ist U also die disjunkte Vereinigung der Äquivalenzklassen von J (das ist immer so bei einer Äquivalenzrelation), und jede dieser Äquivalenzklassen ist ein offenes Intervall. – Es bleibt nur noch zu zeigen, dass die Anzahl dieser (paarweise disjunkten) Äquivalenzklassen höchstens abzählbar unendlich ist. Das folgt daraus, dass wir uns in jeder Äquivalenzklasse eine rationale Zahl aussuchen können. Damit gibt es ein injektive Abbildung von der Menge dieser Äquivalenzklassen in die Menge $\mathbb{Q}$, die bekanntlich abzählbar ist. Fertig. □

Bemerkung: Diese offenen Intervalle können auch als die Zusammenhangskomponenten von U beschrieben werden. Das kann man so sehen. Wenn $J =]a,b[\subset U$ eines dieser (nichtleeren) offenen Intervalle ist, dann gehört a nicht zu U, denn sonst müsste a in einem anderen der offenen Intervalle enthalten sein, nennen wir es J'. Dann würde J' aber auch $a+\varepsilon$ enthalten für genügend kleines $\varepsilon > 0$, so dass $J \cap J' \neq \emptyset$, damit Widerspruch. Ebenso ist $b \notin J$. Daraus folgt, dass die Funktion $f\colon U \to \mathbb{R}$ definiert durch $f(x) = 1$ falls $x \in J$ und $f(x) = 0$ falls $x \notin 0$ eine stetige Funktion ist. Daraus folgt, dass J eine Vereinigung von Zusammenhangskomponenten von U ist. Es ist aber zusammenhängend, also ist es *eine* Zusammenhangskomponente von U.

Wegen Proposition 2.2.1 liegt es nahe, die Länge $\lambda(U)$ einer offenen Menge U in $\mathbb{R}^1$ so zu definieren. Wir wissen, dass U disjunkte Vereinigung von höchstens abzählbar vielen offenen Intervallen $]a_i, b_i[$ ist, und wir setzen dann

$$\lambda(U) = \sum_i (b_i - a_i).$$

Ein paar Bemerkungen dazu. Das i läuft über eine endliche oder abzählbar unendliche Indexmenge (die hier leider keinen Namen hat). Alle Summanden sind nicht-negativ, so dass wir uns um die Reihenfolge der Summanden keine Sorge machen müssen. Es kann sein, dass die Summe divergiert; dann verstehen wir das als $\lambda(U) = +\infty$. Es kann ausserdem sein, dass eines der Intervalle unbeschränkt ist; dann wäre ein a_i gleich $-\infty$, oder ein b_i gleich $+\infty$, oder beides (im letzteren Fall ist $U = \mathbb{R}$). In diesen Fällen ist natürlich wieder $\lambda(U) = +\infty$.

Der Buchstabe λ erinnert einerseits an Länge, andererseits an das Lebesgue-Mass, und das passt beides ganz gut, denn es handelt sich hier um die Länge von U, so wie Lebesgue sie gesehen hat.

Die Proposition 2.2.1 könnte den Eindruck erwecken, dass offene Teilmengen von $\mathbb{R}^1$ harmlos und leicht zu verstehen sind, aber eigentlich ist das Gegenteil der Fall, wie das folgende Beispiel zeigen soll.

Lemma 2.2.2. *Zu jedem $\varepsilon > 0$ existiert eine beschränkte offene Teilmenge U von $\mathbb{R}$, die alle rationalen Zahlen im Intervall $[0,1]$ enthält und die $\lambda(U) \leq \varepsilon$ erfüllt.*

Beweis. Wir wählen eine Bijektion φ von $\mathbb{N} = \{0,1,2,\dots\}$ nach $\mathbb{Q} \cap [0,1]$. Die Menge U soll als disjunkte Vereinigung von offenen Intervallen $J_0, J_1, J_2, J_3, \dots$ konstruiert werden, von denen einige leer sein dürfen. Das Intervall J_n soll irrationale Endpunkte haben (wenn es nicht leer ist), und seine Länge soll höchstens $\varepsilon/2^{n+1}$ sein. Unter Beachtung dieser Bedingungen, die leicht zu erfüllen sind, verfahren wir so: wenn $J_0, J_1, ..., J_n$ schon konstruiert sind und wenn zufällig $\varphi(n+1)$ schon in einem von diesen Intervallen enthalten ist, dann sei $J_{n+1} = \emptyset$. Andernfalls wählen wir J_{n+1} so, dass es eben $\varphi(n+1)$ enthält, und achten darauf, dass es leeren Durchschnitt hat mit den vorher schon gewählten $J_1, J_2, \dots, J_n$. Da diese irrationale Endpunkte haben, ist das kein Problem, denn $\varphi(n+1)$ ist nicht einer von diesen Endpunkten. — Die Gesamtlänge $\lambda(U)$ ist die Summe der Längen von den J_n, und da die Länge von J_n nach Konstruktion immer $< \varepsilon/2^{n+1}$ ist, erhalten wir $\lambda(U) \leq \sum_{n=0}^{\infty} \varepsilon/2^{n+1} = \varepsilon$. □

Wenn nun U eine offene Teilmenge von $\mathbb{R}^1$ ist wie in Lemma 2.2.2, etwa mit $\varepsilon = 1/2$, dann ist die Indikatorfunktion χ_U nicht Riemann-Integrierbar, wie man leicht sieht. Ich nehme dazu die Darboux-Version von Riemann-Integral, mit Treppenfunktionen und Obersummen und Untersummen. Ausserdem können wir mit einem kompakten Intervall K arbeiten, das U enthält, zum Beispiel $K = [-1, +2]$. Wenn wir eine Treppenfunktion (step function) $f\colon K \to \mathbb{R}$ vor uns haben, die $\geq \chi_U$ ist, dann muss $f(x) \geq 0$ gelten für alle $x \in K$ und $f(x) \geq 1$ für alle $x \in [0,1]$, denn es gilt $\chi_U(x) = 1$ für alle $x \in \mathbb{Q} \cap [0,1]$. Dann ist das Integral von f mindestens 1. Wenn wir eine Treppenfunktion (step function) $g\colon K \to \mathbb{R}$ vor uns haben, die $\leq \chi_U$ ist, dann muss $g(x) \leq 1$ gelten für alle $x \in U$ und $g(x) \leq 0$ für alle $x \in K \smallsetminus U$, und dann folgt leicht, dass das Integral von g höchstens so gross ist wie $\lambda(U)$. Und zwischen diesen beiden Abschätzungen klafft eine Lücke von $1/2$.

Das ist (für mich) die schlimmste Unzulänglichkeit vom Riemann-Integral: die Indikatorfunktion χ_U einer beschränkten offenen Menge U im $\mathbb{R}^1$ ist nicht in jedem Fall Riemann-integrierbar. Damit ist gewissermassen das Riemann-*Mass* von U nicht bestimmbar (siehe Abschnitt 1.1), obwohl es doch eigentlich ein klarer Fall sein sollte, nämlich $\lambda(U)$ so wie oben erklärt. (Allerdings kann zugegeben werden, dass die Riemann-Untersumme zu χ_U den wahren Wert liefert. Nur leider versagt die Riemann-Obersumme.) Es ist nicht ganz leicht zu sagen, wer das zuerst gesehen hat und wann; Vito Volterra (1860-1940; um 1881) ist der Hauptverdächtige.

2.3. Kontrastierende Auffassungen in Masstheorie und Topologie

Das Lemma 2.2.2 zeigt, dass eine Teilmenge von einem sehr vernünftigen topologischen Raum (wie zum Beispiel $[0,1]$ mit der üblichen Topologie) offen und dicht[1] sein kann, und trotzdem beliebig klein im Sinn der (gewöhnlichen aber neueren) Masstheorie. (Sei U wie im Lemma; dann ist $V := U \cap [0,1]$ offen und dicht in $[0,1]$, und $\lambda(V) \leq \lambda(U) \leq \varepsilon$.) Andererseits gelten in der Topologie (abseits von der Masstheorie) Teilmengen eines topologischen Raumes X, die offen und dicht sein, als ziemlich "füllend". Die Rechtfertigung dafür ist der berühmte Satz von Baire, der sehr leicht zu beweisen ist.

Satz 2.3.1. (Satz von Baire) *Sei X ein topologischer Raum; angenommen X ist vollständig metrisierbar oder lokalkompakt (einschl. Hausdorff-Eigenschaft). Der Durchschnitt einer abzählbar unendlichen Auswahl von offenen dichten Teilmengen von X ist immer noch dicht in X.*

Beweis: 1. Fall: wir nehmen an, dass X vollständig metrisierbar ist, d.h., es existiert eine Metrik d auf X, die vollständig ist (Cauchy-Folgen sind konvergent) und die die gegebene Topologie induziert. Wir nehmen an, dass offene und dichte Teilmengen U_j für $j = 0, 1, 2, 3, \ldots$ ausgewählt sind. OBdA ist die Folge absteigend, also $U_j \supset U_{j+1}$; das vereinfacht die Bezeichnungen. Sei $z \in X$ beliebig und sei $\varepsilon > 0$ gegeben. Wir müssen zeigen, dass ein $y \in \bigcap_{j=0}^{\infty} U_j$ existiert mit $d(y,z) < \varepsilon$. Weil U_0 offen und dicht ist, finden wir $y_0 \in X$ und $\delta_0 > 0$, so dass der offene Ball $B(y_0, \delta_0)$ in $B(z, \varepsilon) \cap U_0$ enthalten ist. Weil U_1 offen und dicht in X ist, finden wir $y_1 \in X$ und $\delta_1 > 0$, so dass $B(y_1, \delta_1)$ in $B(y_0, \delta_0) \cap U_1$ enthalten ist. Weil U_2 offen und dicht in X ist, finden wir $y_2 \in X$ und $\delta_2 > 0$, so dass $B(y_2, \delta_2)$ in $B(y_1, \delta_1) \cap U_2$ enthalten ist. Usw. Ausserdem können wir sicherstellen, dass $\delta_0 \geq 2\delta_1 \geq 4\delta_2 \geq 8\delta_3 ...$ usw. Dann ist die Folge $y_0, y_1, y_2, ...$ Cauchy-konvergent und konvergiert damit gegen ein $y_\infty = y$. Dieses y ist nach Konstruktion in $B(z, \varepsilon)$ enthalten und auch in jedem der U_j, denn $y \in B(y_j, \delta_j) \subset U_j$.

2. Fall: wir nehmen an, dass X lokalkompakt ist (und die Hausdorff-Trennungseigenschaft hat). *Lokalkompakt* bedeutet, dass für jedes $z \in X$ und jede Umgebung V von z in X eine kompakte Umgebung K von z in X existiert, so dass $K \subset U$. Wir nehmen an, dass offene und dichte Teilmengen U_j für $j = 0, 1, 2, 3, \ldots$ ausgewählt sind. OBdA ist die Folge absteigend. Sei jetzt $z \in X$ beliebig und sei eine Umgebung V von $z \in X$ gegeben. Wir müssen zeigen, dass

[1] Dicht bedeutet: der Abschluss der Teilmenge ist der ganze topologische Raum.

$\bigcap_{j=0}^{\infty} U_j$ nichtleeren Schnitt mit V hat. Weil U_0 offen und dicht ist, finden wir $y_0 \in X$ und eine kompakte Umgebung K_0 von y_0 in X mit $K_0 \subset V \cap U_0$. Weil U_1 offen und dicht in X ist, finden wir $y_1 \in U_1$ und eine kompakte Umgebung K_1 von y_1 in X mit $K_1 \subset K_0 \cap U_1$. Weil U_2 offen und dicht in X ist, finden wir $y_2 \in X$ und eine kompakte Umgebung K_2 von y_2 in X mit $K_2 \subset K_1 \cap U_2$. Usw. Es ist bekannt (oder Übungsaufgabe), dass der Durchschnitt einer absteigenden Folge von kompakten nichtleeren Mengen wieder nichtleer ist. Also ist die Menge $\bigcap_{j=0}^{\infty} K_j$ nicht leer. Sie ist aber auch in V und jedem U_j enthalten. □

Übungsaufgaben. 1. Sei X ein vollständiger metrischer Raum, der keine isolierten Punkte hat. Zeigen Sie: Wenn $X \neq \emptyset$, dann ist X schon überabzählbar.

2. Definition: eine Teilmenge Y von $\mathbb{R}$ ist eine Lebesgue-*Nullmenge*, wenn zu jedem $\varepsilon > 0$ eine offene Menge $V \subset \mathbb{R}$ existiert, die Y enthält und $\lambda(V) < \varepsilon$ erfüllt. (Dazu mehr im nächsten Kapitel.) Zeigen Sie, dass eine Lebesgue-Nullmenge $Y \subset \mathbb{R}$ existiert, die Durchschnitt von (nur) abzählbar vielen offenen dichten Teilmengen von $\mathbb{R}$ ist. (Dann ist Y dicht in $\mathbb{R}$ wegen Satz von Baire.) So ein Y kann nicht abzählbar sein.

3. Eine von mir geschätzte Aufgabe aus der Topologie (die eigentlich eher zum nächsten Kapitel gehört). Sei $S = \{0,1\}$ und $T = \{0,1,2\}$, beide aufgefasst als topologische Räume mit der diskreten Topologie. Wir bilden die unendlichen Produkte

$$S^{\mathbb{N}} = S \times S \times S \times S \times \cdots ,$$
$$T^{\mathbb{N}} = T \times T \times T \times T \times \cdots$$

mit der Produkttopologie. (Wenn Sie mit der Produkttopologie nicht vertraut sind, können Sie die beiden als metrische Räume auffassen: zB für Elemente $(a_i)_{i\in\mathbb{N}}$ und $(b_i)_{i\in\mathbb{N}}$ von $S^{\mathbb{N}}$ ist der Abstand $\sum_{i=0}^{\infty} 2^{-i}|b_i - a_i|$. Für T können Sie genau dieselbe Formel benutzen. Das gibt in beiden Fällen die richtige Topologie.) — Zeigen Sie, dass $S^{\mathbb{N}}$ homöomorph ist zu $T^{\mathbb{N}}$.

CHAPTER 3

Notizen 3. und 4. Vorlesungswoche

3.1. Das Lebesgue-Mass im $\mathbb{R}^n$

Wir fangen an mit einem kompakten (aber eher grossen) Würfel $K = \prod_{j=1}^{n}[-c,c]$ und konstruieren das Lebesgue-Mass in K. Für $L \subset K$ endliche Vereinigung von Quadern wissen wir, was $\lambda(L)$ sein soll. (Unter einem Quader soll hier immer ein Produkt $\prod_{j=1}^{n}[a_j,b_j]$ von *abgeschlossenen* Intervallen verstanden werden.) Spezialfall: $\lambda(K) = 2^n c^n$.

Lemma 3.1.1. *Wenn $L, M \subset K$ endliche Vereinigungen von Quadern sind, dann $\lambda(L \cup M) = \lambda(L) + \lambda(M) - \lambda(L \cap M)$.* □

Definition 3.1.2. Für eine offene Teilmenge U von K (die nicht offen in $\mathbb{R}^n$ sein muss) definieren wir $\lambda(U) = \sup\{\lambda(L) \mid L$ endl. Verein. von Quadern$\}$.

Bemerkung: Diese Definition von $\lambda(U)$ sollte man ausprobieren mit $n = 1$, also U offene Teilmenge von $[-c,c]$. Dann ist U Vereinigung von höchstens abzählbar vielen Intervallen, die offen sind in $[-c,c]$. Vergleichen mit Lemma 2.2.2.

Lemma 3.1.3. *Für offene Mengen $U, V \subset K$ gilt $\lambda(U \cup V) = \lambda(U) + \lambda(V) - \lambda(U \cap V)$.*

Beweis. Zu gegebenem $\varepsilon > 0$ können wir finden $L_1 \subset U$, endliche Vereinigung von Quadern, und $L_2 \subset V$, endliche Vereinigung von Quadern, so dass $\lambda(U) - \lambda(L_1) < \varepsilon$, $\lambda(V) - \lambda(L_2) < \varepsilon$, $\lambda(U \cap V) - \lambda(L_1 \cap L_2) < \varepsilon$, $\lambda(U \cup V) - \lambda(L_1 \cup L_2) < \varepsilon$. Dann ist

$$\lambda(U \cup V) \geq \lambda(L_1 \cup L_2) = \lambda(L_1) + \lambda(L_2) - \lambda(L_1 \cap L_2) \geq \lambda(U) + \lambda(V) - \lambda(U \cap V) - 2\varepsilon$$

und es folgt $\lambda(U \cup V) \geq \lambda(U) + \lambda(V) - \lambda(U \cap V)$. Ebenso

$$\lambda(U \cup V) \leq \lambda(L_1 \cup L_2) + \varepsilon = \lambda(L_1) + \lambda(L_2) - \lambda(L_1 \cap L_2) + \varepsilon \leq \lambda(U) + \lambda(V) - \lambda(U \cap V) + 2\varepsilon$$

und es folgt $\lambda(U \cup V) \leq \lambda(U) + \lambda(V) - \lambda(U \cap V)$. □

Lemma 3.1.4. *Sei L endliche Vereinigung von Quadern, $L \subset K$. Dann ist $K \smallsetminus L$ offen in K und es gilt $\lambda(K \smallsetminus L) = \lambda(K) \smallsetminus \lambda(L)$.*

Beweis. Man soll sich überlegen, dass wir zu jedem $\varepsilon > 0$ ein offenes $U \subset K$ finden können mit $L \subset U$ und $0 \leq \lambda(U) - \lambda(L) \leq \varepsilon$. (Dieses U kann zum Beispiel in einer endlichen Vereinigung von Quadern enthalten sein, die etwas grösser ist als L.) Dann ist $U \smallsetminus L$ wieder offen in K und aus den Definitionen folgt leicht, dass $0 \leq \lambda(U \smallsetminus L) \leq \varepsilon$. Dann ist $K = (K \smallsetminus L) \cup U$ und daher

$$\lambda(K) = \lambda(K \smallsetminus L) + \lambda(U) - \lambda((K \smallsetminus L) \cap U) = \lambda(K \smallsetminus L) + \lambda(U) - \lambda(U \smallsetminus L)$$

wegen Lemma 3.1.3. Und wir sehen, dass $|\lambda(K) - \lambda(K \smallsetminus L) - \lambda(L)| < \varepsilon$. □

Definition 3.1.5. Für C abgeschlossen in K definieren wir: $\lambda(C) = \lambda(K) \smallsetminus \lambda(K \smallsetminus C)$, wobei die rechte Seite schon definiert ist. (Diese Definition ist im Einklang mit der schon vorher gegebenen für Quader und endliche Vereinigungen von Quadern; das ist eigentlich schon gezeigt worden.)

Definition 3.1.6. Das *äussere Mass* einer Teilmenge T von K ist

$$\lambda^*(T) := \inf\{\lambda(U) \mid U \supset T,\ U \text{ offen in } K\}$$

und das *innere Mass* ist $\lambda_*(T) := \sup\{\lambda(C) \mid C \subset T,\ C \text{ abgeschlossen in } K\}$. Daraus sollte ersichtlich sein, dass immer $\lambda_*(T) \leq \lambda^*(T)$ gilt. Wir sagen, dass T *messbar* ist, wenn $\lambda_*(T) = \lambda^*(T)$, und in diesem Fall definieren wir $\lambda(T) := \lambda^*(T) = \lambda_*(T)$.

Lemma 3.1.7. *Wenn $T_0, T_1 \subset K$ messbar sind, dann ist auch $T_0 \cup T_1$ messbar.*

Beweis. Sei $\varepsilon > 0$ gegeben. Wir wählen $U_0 \subset K$ offene Umgebung von T_0 und $U_1 \subset K$ offene Umgebung von T_1, so dass $\lambda(U_0) - \lambda(T_0) < \varepsilon$ und $\lambda(U_1) - \lambda(T_1) < \varepsilon$. Ausserdem wählen wir $A_0 \subset T_0$ and $A_1 \subset T_1$ derart, dass A_0, A_1 abgeschlossen in K und $\lambda(T_0) - \lambda(A_0) < \varepsilon$ sowie $\lambda(T_1) - \lambda(A_1) < \varepsilon$. Dann ist

$$\lambda(U_0) + \lambda(U_1) - \lambda(U_0 \cap U_1) = \lambda(U_0 \cup U_1) \geq \lambda^*(T_0 \cup T_1)$$

und ebenso

$$\lambda(A_0) + \lambda(A_1) - \lambda(A_0 \cap A_1) = \lambda(A_0 \cup A_1) \leq \lambda_*(T_0 \cup T_1).$$

Demnach ist

$$\begin{aligned}
\lambda^*(T_0 \cup T_1) - \lambda_*(T_0 \cup T_1) &\leq \lambda(U_0) + \lambda(U_1) - \lambda(U_0 \cap U_1) - (\lambda(A_0) + \lambda(A_1) - \lambda(A_0 \cap A_1)) \\
&= (\lambda(U_0) - \lambda(A_0)) + (\lambda(U_1) - \lambda(A_1)) - (\lambda(U_0 \cap U_1) - \lambda(A_0 \cap A_1)) \\
&\leq (\lambda(U_0) - \lambda(A_0)) + (\lambda(U_1) - \lambda(A_1)) \leq 4\varepsilon.
\end{aligned}$$

Weil das für beliebige $\varepsilon > 0$ gilt, stellen wir fest, dass $\lambda^*(T_0 \cup T_1) = \lambda_*(T_0 \cup T_1)$. □

Korollar 3.1.8. *Wenn $T_0, T_1 \subset K$ messbar sind, dann sind auch $T_0 \cap T_1$ und $T_0 \setminus T_1$ messbar.*

Beweis. Wenn $T_0, T_1 \subset K$ messbar sind, dann auch ihre Komplemente, dann auch die Vereinigung ihrer Komplemente wegen Lemma 3.1.7. Dann ist auch das Komplement davon messbar; aber das ist ja $T_0 \cap T_1$. Ebenso ist $T_0 \setminus T_1$ der Durchschnitt von T_0 und dem Komplement von T_1, und damit messbar. □

Die verbleibende Arbeit besteht jetzt darin, zu zeigen, dass erstens die Menge $\mathcal{A}$ der messbaren Teilmengen von K eine σ-Algebra ist, und natürlich zweitens, dass $\lambda\colon \mathcal{A} \to [0, 2^n c^n]$ die Eigenschaften hat, die man (nach Borel-Lebesgue) von einem Mass verlangt. Das heisst (in einer minimalistischen Formulierung, die ich aus dem Buch von Brokate-Kersting[1] habe):

(i) $K \in \mathcal{A}$, für jedes $T \in \mathcal{A}$ ist auch $K \setminus T$ Element von $\mathcal{A}$, und wenn $T_0, T_1, T_2, \ldots$ Elemente von $\mathcal{A}$ sind, dann ist auch $T := \bigcup_{j=0}^{\infty} T_j$ Element von $\mathcal{A}$.

(ii) $\lambda(\emptyset) = 0$, und wenn $T_0, T_1, T_2, \dots \in \mathcal{A}$ paarweise disjunkt sind, dann ist $\lambda(\bigcup_{j=0}^{\infty} T_j) = \sum_{j=0}^{\infty} \lambda(T_j)$.

Die beiden ersten Aussagen in (i) sind schon abgehakt. Für die dritte können wir annehmen, dass die T_j paarweise disjunkt sind, denn sonst ersetzen wir T_j durch $T_j \setminus (T_0 \cup T_1 \cup \dots \cup T_{j-1})$ für $j > 0$ (und wir wissen schon, dass $T_j \setminus (T_0 \cup T_1 \cup \dots \cup T_{j-1})$ messbar ist). Wenn die T_j paarweise disjunkt sind, dann ist es einigermassen klar, dass $\lambda_*(T) \geq \sum_{j=0}^{\infty} \lambda_*(T_j) = \sum_{j=0}^{\infty} \lambda(T_j)$ ist. (Denn wenn wir für $j = 1, 2, ..., r$ abgeschlossene Teilmengen A_j von K wählen mit $A_j \subset T_j$ und $\lambda(T_j) - \lambda(A_j) < \varepsilon$, dann sind die A_j paarweise disjunkt und $A_0 \cup A_1 \cup \dots \cup A_r$ ist abgeschlossene Teilmenge von K und ist enthalten in T, so dass $\lambda_*(T) \geq \sum_{j=0}^{r} \lambda(A_j) = (\sum_{j=0}^{r} \lambda(T_j)) - (r+1)\varepsilon$. Und da $\epsilon > 0$ beliebig klein war, bedeutet das $\lambda_*(T) \geq \sum_{j=0}^{r} \lambda(T_j)$, und da r beliebig gross war,

[1] M. Brokate und G. Kersting, *Mass und Integral*, Reihe "Mathematik kompakt", Birkhäuser, Springer Basel AG 2011,2019

folgt die Behauptung.) Andererseits können wir uns bei gegebenem $\varepsilon > 0$ für jedes j ein V_j offen in K mit $T_j \subset V_j$ wählen derart, dass $\lambda(V_j) - \lambda(T_j) < \varepsilon/2^{j+1}$, und wir setzen $V = \bigcup_{j=0}^{\infty} V_j$. Dann ist V offene Umgebung von T und daher

$$\lambda^*(T) \le \lambda(V) \le \sum_{j=0}^{\infty} \lambda(V_j) < \varepsilon + \sum_{j=0}^{\infty} \lambda(T_j).$$

Da das für jedes $\varepsilon > 0$ gilt, haben wir $\lambda^*(T) \le \sum_{j=0}^{\infty} \lambda(T_j)$, und damit $\lambda_*(T) = \lambda^*(T)$ wie gewünscht. Ausserdem haben wir damit auch den nicht-trivialen Teil von (ii) bewiesen. Damit ist die Konstruktion (Lebesgue-Mass für den Würfel K) beendet.

Definition 3.1.9. Eine Teilmenge T von $\mathbb{R}^n$ soll messbar heissen, wenn $S \cap K_c$ messbar ist als Teilmenge von K_c für jeden Würfel $K_c := \prod_{i=1}^{n} [-c, c]$; in diesem Fall ist das Lebesgue-Mass von T definiert als

$$\lambda(T) := \lim_{c \to \infty} \lambda(T \cap K_c).$$

Dieser Grenzwert kann auch ∞ sein.

Dann ist es leicht zu sehen, dass die messbaren Teilmengen von $\mathbb{R}^n$ eine σ-Algebra $\mathcal{A}$ bilden, und dass λ ein Mass auf $\mathcal{A}$ ist. Wir haben ausserdem erreicht: jede offene Teilmenge von $\mathbb{R}^n$ gehört zu $\mathcal{A}$.

3.2. Ähnliches Beispiel von Konstruktion eines Masses

Für $i = 0, 1, 2, 3, \ldots$ sei S_i eine endliche Menge augerüstet mit einem *Wahrscheinlichkeitsmass* μ_i. Das soll hier so verstanden werden, dass jede Teilmenge von S_i messbar ist. Also ist μ_i einfach eine Funktion von der Potenzmenge $\mathcal{P}(S_i)$ nach $[0, 1]$, die $\mu_i(S_i) = 1$ und $\mu_i(\emptyset) = 0$ erfüllt und $\mu_i(A \cup B) = \mu_i(A) + \mu_i(B)$ für disjunkte Teilmengen A und B von S_i. Es sollte klar sein, dass μ_i bestimmt ist durch seine Werte auf den einelementigen Teilmengen von S_i. Die Summe dieser Werte muss 1 sein.

Wir wollen uns überlegen, wie man ein Wahrscheinlichkeitsmass ω auf der Produktmenge

$$X := S_0 \times S_1 \times S_2 \times S_3 \times \cdots$$

definieren kann, das die Bezeichnung $\mu_0 \times \mu_1 \times \mu_2 \times \cdots$ verdient. (Wozu kann das gut sein? Wir können damit die Wahrscheinlichkeiten von Ereignissen in einem grossen Experiment beschreiben, das so abläuft: erst wählen wir zufällig ein Element von S_0, mit Wahrscheinlichkeiten wie durch μ_0 gegeben; dann wählen wir zufällig ein Eelement von S_1, mit Wahrscheinlichkeiten wie durch μ_1 gegeben; dann wählen wir zufällig ein Element von S_2, mit Wahrscheinlichkeiten wie durch μ_2 gegeben; usw.; nicht aufhören.) Wir wollen nicht darauf bestehen, dass jede Teilmenge von X messbar ist; aber wir wollen darauf bestehen, dass die *offenen* Teilmengen von X messbar sind. Dazu rufen wir uns nochmal in Erinnerung, was die offenen Teilmengen von X sind oder sein sollen. (Hier wird X aufgefasst als topologischer Raum, nämlich das Produkt der S_i, wobei jedes S_i mit der diskreten Topologie augerüstet ist; d.h. alle Teilmengen von S_i gelten als offen. Es folgt nicht, dass in X auch alle Teilmengen offen sind.)

Definition 3.2.1. Unter einem *Quader* in X soll eine Teilmenge von X von der Form

$$Q = T_0 \times T_1 \times T_2 \times \cdots$$

verstanden werden, wobei immer $T_i \subset S_i$ und fast immer $T_i = S_i$ gilt (also nur endlich viele Ausnahmen erlaubt).

Ein Durchschnitt von endlich vielen Quadern in X ist wieder ein Quader in X. — Eine Teilmenge von X ist offen genau dann, wenn sie eine Vereinigung von solchen Quadern ist. Man kann das als Satz oder als Definition auffassen. (Es ist eigentlich ein Satz, weil die Produkttopologie schon irgendwo vorher definiert worden ist. Wenn man es als Definition auffasst, dann ist der Preis dafür, dass man zeigen muss: diese Auswahl von Teilmengen von X hat die Eigenschaften, die von einer *Topologie* auf X gefordert werden. Dann wird man auch verstehen, warum es nützlich ist, zu bemerken, dass ein endlicher Durchschnitt von Quadern wieder ein Quader ist.)

Übrigens: ein Quader $Q \subset X$ ist demnach offen, aber auch abgeschlossen, denn das Komplement ist eine endliche Vereinigung von Quadern. Wir sehen, dass X sehr viele Teilmengen hat, die gleichzeitig offen und abgeschlossen sind; das heisst, X ist hochgradig *nicht* zusammenhängend, ausser in dem Fall, wo alle S_i nur jeweils ein Element haben[2].

Jetzt können wir ungefähr so verfahren, wie wir es bei der Konstruktion des Lebesgue-Masses gemacht haben. Für einen Quader

$$Q = T_0 \times T_1 \times T_2 \times \cdots$$

definieren wir:

$$\omega(Q) := \prod_{i=0}^{\infty} \mu_i(T_i)$$

wobei zu bedenken ist, dass $\mu_i(T_i) = 1$ für fast alle i; das angeblich unendliche Produkt ist also in Wirklichkeit ein endliches Produkt. Falls $L \subset X$ eine endliche Vereinigung von Quadern ist, wissen wir auch, wie wir $\omega(L)$ definieren sollen. Zum Beispiel wenn P, Q, R Quader sind, dann versuchen wir es mit

$$\omega(P \cup Q \cup R) = \omega(P) + \omega(Q) + \omega(R) - \omega(P \cap Q) - \omega(Q \cap R) - \omega(P \cap R) + \omega(P \cap Q \cap R).$$

(Es sollte überprüft werden, ob das wohldefiniert ist, aber das will ich hier lieber aussparen.)

Lemma 3.2.2. *Wenn $L \subset X$ und $M \subset X$ endliche Vereinigungen von Quadern sind, dann gilt* $\lambda(L \cup M) = \lambda(L) + \lambda(M) - \lambda(L \cap M)$. □

Für eine offene Teilmenge V von X definieren wir

$$\omega(V) := \sup_{L \subset V} \ \omega(L)$$

wobei L für eine endliche Vereinigung von Quadern steht (die in V enthalten ist).

Lemma 3.2.3. *Für offene Mengen $U, V \subset K$ gilt* $\lambda(U \cup V) = \lambda(U) + \lambda(V) - \lambda(U \cap V)$.

Beweis wie im Fall von Lebesgue-Mass λ. □

Lemma 3.2.4. *Sei L endliche Vereinigung von Quadern in X. Dann ist $X \smallsetminus L$ offen in X und es gilt* $\omega(X \smallsetminus L) = \omega(X) \smallsetminus \omega(L) = 1 - \omega(L)$.

Beweis. Folgt aus Lemma 3.2.2, denn $X \smallsetminus L$ ist wieder endliche Vereinigung von Quadern. □

Definition 3.2.5. Für C abgeschlossen in X definieren wir: $\omega(C) = 1 - \omega(X \smallsetminus C)$, wobei die rechte Seite schon definiert ist, denn $X \smallsetminus C$ ist offen. (Wenn C endliche Vereinigung von Quadern ist, dann ist das in Einklang mit der schon gegebenen Definition von $\omega(C)$ wegen Lemma 3.2.4.)

Alles weitere kann dann gemacht werden wie im Fall von Lebesgue-Mass λ.

[2]Man kann es auch so sagen: X ist total unzusammenhängend.

DEFINITION 3.2.6. Das *äussere Mass* einer Teilmenge T von X ist

$$\omega^*(T) := \inf\{\omega(U) \mid U \supset T,\ U \text{ offen in } X\}$$

und das *innere Mass* ist $\omega_*(T) := \sup\{\omega(C) \mid C \subset T,\ C \text{ abgeschlossen in } K\}$. Daraus sollte ersichtlich sein, dass immer $\omega_*(T) \le \omega^*(T)$ gilt. Wir sagen, dass T *messbar* ist, wenn $\omega_*(T) = \omega^*(T)$, und in diesem Fall definieren wir $\omega(T) := \omega^*(T) = \omega_*(T)$.

Wie im Fall vom Lebesgue-Mass λ in einem Würfel $K \subset \mathbb{R}^n$ kann man dann zeigen, dass die messbaren Teilmengen von X eine σ-Algebra $\mathcal{A}$ bilden, und dass ω ein Wahrscheinlichkeitsmass auf $\mathcal{A}$ ist. Das bedeutet, in minimalistischer Formulierung:

(i) $X \in \mathcal{A}$, für jedes $T \in \mathcal{A}$ ist auch $X \smallsetminus T$ Element von $\mathcal{A}$, und wenn $T_0, T_1, T_2, \dots$ Elemente von $\mathcal{A}$ sind, dann ist auch $T := \bigcup_{j=0}^{\infty} T_j$ Element von $\mathcal{A}$.

(ii) $\omega(\emptyset) = 0$, $\omega(X) = 1$, und wenn $T_0, T_1, T_2, \dots \in \mathcal{A}$ paarweise disjunkt sind, dann ist $\omega(\bigcup_{j=0}^{\infty} T_j) = \sum_{j=0}^{\infty} \omega(T_j)$.

Wir haben ausserdem erreicht: jede offene Teilmenge von X gehört zu $\mathcal{A}$.

Übungsaufgaben. 1. Sei $X := S_0 \times S_1 \times S_2 \times S_3 \times \cdots$ der topologische Raum von Abschnitt 3.2. Finden Sie eine offene Teilmenge von X, die nicht abgeschlossen ist.

2. Sei Y eine Menge und sei $\mathcal{A}$ eine σ-Algebra von Teilmengen von Y. Für eine Folge $T_0, T_1, T_2, \dots$ von Teilmengen von Y, die Elemente von $\mathcal{A}$ sind, sei

$$W := \{y \in Y \mid \text{es gibt unendlich viele } j \text{ mit } y \in T_j\}.$$

Zeigen Sie: $W \in \mathcal{A}$.

3. Sei $K = [-c, c]^n$ ein Würfel in $\mathbb{R}^n$ wie in Abschnitt 3.1. Sei $\mathcal{A}$ die σ-Algebra der messbaren Teilmengen von K wie in demselben Abschnitt. Sei $\mathcal{B}$ die kleinste σ-Algebra von Teilmengen von K, die alle offenen Teilmengen von K enthält (als Elemente). Dann ist $\mathcal{B} \subset \mathcal{A}$, denn wir haben gezeigt, dass alle offenen Teilmengen von K Elemente von $\mathcal{A}$ sind. Zeigen Sie: zu jedem $T \in \mathcal{A}$ existiert $T_1 \in \mathcal{B}$ derart, dass $\lambda(T \smallsetminus T_1) = 0$. (Hier wird benutzt, dass $T \smallsetminus T_1 \in \mathcal{A}$.) *Etwas schärfer*: Zu jedem $T \in \mathcal{A}$ existiert $T_1 \in \mathcal{B}$ derart, dass $T \smallsetminus T_1$ enthalten ist in einem $T_2 \in \mathcal{B}$, das $\lambda(T_2) = 0$ erfüllt. (In Worten: jede messbare Teilmenge T von K, im Sinne von Abschnitt 3.1, enthält eine Borel-Menge T_1, so dass $T \smallsetminus T_1$ eine Nullmenge ist, und sogar in einer Borel-Nullmenge enthalten ist.) Diese Aussage ist sehr bekannt und Sie können den Beweis vielleicht auch bei Wikipedia irgendwo finden.

CHAPTER 4

Notizen 5. und 6. Vorlesungswoche

4.1. Nachtrag

Wegen einer Frage, die in der Vorlesung gestellt wurde, ist mir der Gedanke gekommen, dass die Definitionen in Abschnitt 3.1 des Lebesgue-Masses $\lambda(U)$ bzw. $\lambda(C)$ für offene Teilmengen U bzw. abgeschlossene Teilmengen C eines Würfels K vielleicht eine bessere Rechtfertigung brauchen. Denn diese Definitionen sind ja sehr unterschiedlich.

Die entscheidende Beobachtung ist: jede offene Teilmenge U von K ist eine Vereinigung von abzählbar unendlich vielen Quadern. (Leere Quader sind dabei erlaubt.) Das ist nicht schwer zu sehen, und es wird weiter unten präzisiert. Wenn nun U die Vereinigung von Quadern Q_j ist wobei $j = 0, 1, 2, \ldots$, dann liegt es nahe, zu glauben, dass das Mass $\lambda(U)$, wenn wir es überhaupt definieren können, gleich dem Supremum oder Grenzwert der aufsteigenden Folge

$$\lambda(L_0), \lambda(L_1), \lambda(L_2), \ldots$$

ist mit $L_j = Q_0 \cup Q_1 \cup \cdots \cup Q_j$. Damit wir eine Definition haben, die unabhängig ist von der Wahl einer Überdeckung von U mit abzählbar vielen Quadern (und ihrer Anordnung), versuchen wir es gleich mit

$$\lambda(U) := \sup\{\lambda(L)\}$$

wobei L irgendeine endliche Vereinigung von Quadern ist, die in U enthalten ist.

Andererseits: es ist *nicht* richtig, dass jede abgeschlossene Teilmenge von K auch eine Vereinigung von abzählbar unendlich vielen Quadern ist. Deswegen sollten wir nicht versucht sein, die Definition, die bei offenen Teilmengen von K gut funktionierte[1], auch bei abgeschlossenen Teilmengen von K zu benutzen.

Hier ist ein Beispiel. Es ist vorhersehbar. Sei $n = 1$ und sei $K = [-1, 1]$. Das ist jetzt unser 1-dimensionaler Würfel. Nach dem, was in Lemma 2.2.2 geschrieben steht, können wir eine offene Teilmenge U von $[-1, 1]$ herstellen, die alle rationalen Zahlen in $[-1, 1]$ enthält und die trotzdem $\lambda(U) < 10^{-6}$ erfüllt. Für dieses Lemma hatten wir allerdings $\lambda(U)$ so definiert: U ist eine disjunkte Vereinigung von höchstens abzählbar unendlich vielen Intervallen, die offen in $[-1, 1]$ sind[2], und $\lambda(U)$ sollte die Summe ihrer Längen sein. Diese alte Definition von $\lambda(U)$ ist mit der neueren von Abschnitt 3.1 im Einklang; dazu keine weiteren Argumente, aber ich bilde mir trotzdem ein, dass der Beweis nicht schwer ist.[3] Nun betrachten wir $C = [-1, 1] \smallsetminus U$, eine abgeschlossene Teilmenge von K. Beachten, dass C keine rationalen Zahlen enthält. Sei $Q = [a, b]$ irgendein abgeschlossenes Teilintervall von $[-1, 1]$; dazu können wir auch 1-dimensionaler Quader sagen. Wenn $b > a$ ist, dann enthält $[a, b]$ rationale Zahlen, also ist $[a, b]$ nicht in C enthalten. Das heisst, C enthält überhaupt keine "Quader" ausser solchen, die nur aus

[1] Was einigermassen überprüft wurde, zB mit Lemma 3.1.3.

[2] So ein Intervall kann ein offenes Intervall in $\mathbb{R}$ sein, es kann aber auch die Form $[-1, b[$ oder $]b, 1]$ haben mit $-1 < b < 1$, oder es kann ganz $[-1, 1]$ sein.

[3] Ein endliche Vereinigung von Intervallen in $[-1, 1]$ kann immer als endliche Vereinigung von paarweise disjunkten Intervallen geschrieben werden. Das macht den Vergleich leichter.

einem einzigen Punkt bestehen. Das heisst, wenn wir es versuchen mit einer Definition wie $\lambda(C) = \sup\ \{\lambda(L)\}$ wobei L die endlichen Vereinigungen von Quadern durchläuft, die in C enthalten sind, dann würde das $\lambda(C) = 0$ nach sich ziehen. Und das würde zur Folge haben, dass $\lambda(C)$ viel kleiner ist als $\lambda(K) - \lambda(U)$, denn $\lambda(K) - \lambda(U)$ ist mindestens $2 - 10^{-6}$. Und das wäre schlecht.

Ich wollte jetzt noch präzisieren, wie man besonders *geschickt* eine offene Teilmenge U von K als abzählbare Vereinigung von Quadern schreiben kann. Der Einfachheit halber sei $K = [-1,1]^n$. Sei $\mathcal{M}$ die Menge aller Würfel der Form $\prod_{j=1}^n [a_j, b_j]$, die in K enthalten sind und ausserdem erfüllen

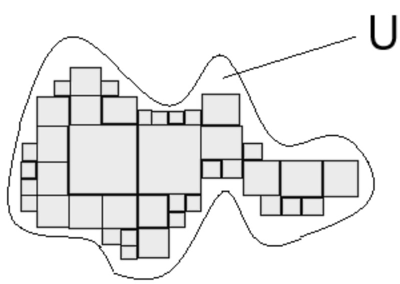
$$2^r b_j \in \mathbb{Z}, \quad 2^r a_j \in \mathbb{Z}, \qquad 2^r b_j - 2^r a_j = 1$$

für irgendein $r \in \{1, 2, 3, 4, \dots\}$. Dann ist U die Vereinigung aller $Q \in \mathcal{M}$, die in U enthalten sind und die maximal sind in dem Sinne, dass es kein grösseres $Q' \in \mathcal{M}$ gibt mit $Q' \subset U$. Und das ist eine abzählbare Vereinigung. Die Würfel, die an dieser Vereinigung teilnehmen, sind zwar nicht paarweise disjunkt, aber fast: das heisst, wenn wir ihre Randflächen entfernen, werden sie paarweise disjunkt. Das heisst, dass man sogar erwarten (oder ggf. definieren) darf:

$$\lambda(U) = \sum \lambda(Q_i)$$

wobei die maximalen Würfel sind, die zu $\mathcal{M}$ gehören und gerade noch in U enthalten sind.

Versuch einer künstlerischen Darstellung:



4.2. Beispiel einer nicht-messbaren Menge

Es handelt sich um das Standard-Beispiel (von Giuseppe Vitali, 1905) einer nicht-messbaren Teilmenge von $X = [0, 1[$, wobei X mit dem Lebesgue-Mass λ und mit der zugehörigen σ-Algebra $\mathcal{A}_X$ (der Lebesgue-messbaren Teilmengen) ausgerüstet ist, alles Spezialfall von Abschnitt 3.1. (Ich erlaube mir allerdings, das halb-offene Intervall $[0, 1[$ anstelle von $[0, 1]$ zu benutzen; Sie werden schon sehen, warum.) Bei dieser Konstruktion wird das Auswahlaxiom benutzt. (Es scheint sogar ein Satz der axiomatischen Mengenlehre zu sein, dass es ohne Auswahlaxiom nicht geht.)

Wir brauchen dazu eine Teilmenge S von X, die folgende Eigenschaft hat:

(i) für jedes $z \in X$ existiert genau ein $x \in S$ mit $x - z \in \mathbb{Q}$.

Um so ein S zu konstruieren, definieren wir zuerst eine Äquivalenzrelation $\mathfrak{R}$ auf der Menge X, nämlich durch $u\mathfrak{R}v \Leftrightarrow u - v \in \mathbb{Q}$. (Es ist leicht zu sehen, dass es sich um eine Äquivalenzrelation handelt: reflexiv, symmetrisch, transitiv ...) Dann wählen wir aus jeder Äquivalenzklasse genau ein Element aus. Das ergibt S. Hier wird das Auswahlaxiom benutzt, und es ist nicht leicht daran vorbeizukommen. (Es hat zB noch nie jemand eine “Formel” angeben können, die das Gewünschte leistet.)

Für eine rationale Zahl $a \in X$ sei $V_a \colon X \to X$ die bijektive Abbildung gegeben durch $V_a(x) = x + a$ falls $x + a < 1$, und $V_a(x) = x + a - 1$ sonst[4]. Dann gilt für eine beliebige Teilmenge $T \subset X$:

$$\lambda^*(V_a(T)) = \lambda^*(T) \text{ und ebenso } \lambda_*(V_a(T)) = \lambda_*(T).$$

Um das zu sehen, dürfen wir oBdA annehmen, dass $1 - a \notin T$. Dann ist T die disjunkte Vereinigung von $T \cap [0, 1-a[$ und $T \cap]1-a, 1[$, und diese beiden werden durch V_a nur “verschoben” (um a bzw. $a-1$).

Daraus folgt, dass für messbares $T \subset X$, also $T \in \mathcal{A}_X$, immer $V_a(T)$ auch messbar ist, und $\lambda(T) = \lambda(V_a(T))$. Ausserdem folgt aus (i):

(ii) die Mengen $V_a(S)$ für $a \in X \cap \mathbb{Q}$ sind paarweise disjunkt, und ihre Vereinigung ist X.

Wenn nun S messbar wäre, dann wären alle $V_a(S)$ ebenfalls messbar, mit $\lambda(V_a(S)) = \lambda(S)$, und dann müsste wegen σ-Additivität gelten:

$$1 = \lambda(X) = \sum_{a \in X \cap \mathbb{Q}} \lambda(V_a(S)) = \sum_{a \in X \cap \mathbb{Q}} \lambda(S).$$

Die Summe/Reihe rechts kann aber nur existieren/konvergieren, wenn $\lambda(S) = 0$, und in dem Fall hat sie den Wert 0. Widerspruch.

4.3. Dynamische Systeme (Vorschau, Definitionen, Beispiele)

Ein dynamisches System besteht im aller-einfachsten Fall aus einer Menge X und einer Abbildung $f \colon X \to X$. Wir interessieren uns für Iterationen von f und für statistische Eigenschaften dieser Iterationen.

Etwas interessanter wird es, wenn X mit zusätzlicher Struktur ausgestattet ist und f diese Struktur erhält oder teilweise erhält. Für uns kommt natürlich besonders in Frage: eine σ-Algebra $\mathcal{A}$ von Teilmengen von X und ein Mass μ auf $\mathcal{A}$.

Definition 4.3.1. Ein Paar $(X, \mathcal{A})$ bestehend aus Menge X und σ-Algebra $\mathcal{A} \subset \mathcal{P}(X)$ heisst *messbarer Raum*; ein Tripel $(X, \mathcal{A}, \mu)$ bestehend aus Menge X und σ-Algebra $\mathcal{A}$ und Mass $\mu \colon \mathcal{A} \to [0, \infty]$ heisst *Massraum*; wenn ausserdem $\mu(X) = 1$ ist, dann darf man *Wahrscheinlichkeitsraum* sagen.

Definition 4.3.2. Sei X eine Menge ausgestattet mit σ-Algebra $\mathcal{A} \subset \mathcal{P}(X)$, und sei Y Menge ausgestattet mit σ-Algebra $\mathcal{B} \subset \mathcal{P}(Y)$. Eine Abbildung $f \colon X \to Y$ heisst *messbar* (bezügl. $\mathcal{A}$ und $\mathcal{B}$), wenn für jedes $T \subset Y$, $T \in \mathcal{B}$ gilt: $f^{-1}(T) \in \mathcal{A}$.

Definition 4.3.3. Mit Bezeichnungen wie in der vorigen Definition: sei μ_0 ein Mass auf $\mathcal{A}$ und sei μ_1 ein Mass auf $\mathcal{B}$. Sei $f \colon X \to Y$ messbar bezüglich $\mathcal{A}$ und $\mathcal{B}$. Man sagt, dass f *masserhaltend* ist, wenn $\mu_0(f^{-1}(T)) = \mu_1(T)$ für alle $T \in \mathcal{B}$ gilt.

Wir kehren jetzt wieder zurück zur Situation $f \colon X \to X$. Zusätzlich soll angenommen werden, dass X mit σ-Algebra $\mathcal{A} \subset \mathcal{P}(X)$ ausgestattet ist, also $(X, \mathcal{A})$ ist Massraum, und dass $f \colon X \to X$ messbar ist (bezüglich $\mathcal{A}$).

Frage 4.3.4. Gibt es ein Mass μ auf dem messbaren Raum $(X, \mathcal{A})$, so dass f masserhaltend bezüglich $\mathcal{A}$ ist?

---

[4]Man sagt auch: $V_a(x) \equiv x + a$ modulo 1. Oder ähnlich.

Definition 4.3.5. Sei $(X, \mathcal{A}, \mu)$ Wahrscheinlichkeitsraum und sei $f\colon X \to X$ messbar und masserhaltend. Eine messbare Teilmenge $T \subset X$ (also $T \in \mathcal{A}$) heisst *invariant* bei f, wenn $f^{-1}(T) = T$, und *fast* invariant bei f, wenn die Mengen $f^{-1}(T) \smallsetminus T$ und $T \smallsetminus f^{-1}(T)$ das μ-Mass 0 haben.

Lemma 4.3.6. *Mit Bezeichnungen wie in Definition 4.3.5: zu jedem messbaren $T \subset X$, das fast invariant ist bei f, existiert messbares $S \subset X$, das invariant ist bei f und ausserdem erfüllt: $\mu(S \smallsetminus T) = 0$, $\mu(T \smallsetminus S) = 0$.*

Beweis: Sei $T_n = \{x \in X \mid f^n(x) \in T\} = f^{-n}(T)$, wobei der Exponent in f^n für die n-fache Iteration steht. (Spezialfall: $f^0 = \mathrm{id}$, $T_0 = T$.) Sei S_k die Vereinigung aller T_n für $n \geq k$. Wegen f-fast-Invarianz von T ist $\mu(T_n \smallsetminus T) = 0 = \mu(T \smallsetminus T_n)$. Daraus folgt mit σ-Additivität, dass $\mu(S_0 \smallsetminus T) = 0$. Die S_k bilden eine absteigende Folge, wobei $f^{-1}(S_k) = S_{k+1}$, so dass $\mu(S_{k+1}) = \mu(S_k)$. Daraus ergibt sich $\mu(S_k \smallsetminus S_{k+1}) = 0$. Sei $S = \bigcap_{k \geq 0} S_k$. (Also ist S die Menge aller $x \in X$, für die unendlich viele $n \geq 0$ existieren mit $f^n(x) \in T$.) Aus dem Obigen folgt $\mu(S_0 \smallsetminus S) = 0$ mit σ-Additivität, denn $S_0 \smallsetminus S = (S_0 \smallsetminus S_1) \cup (S_1 \smallsetminus S_2) \cup (S_2 \smallsetminus S_3) \cup \ \dots$, und wir hatten auch schon $\mu(S_0 \smallsetminus T) = 0$, wobei $S \subset S_0 \supset T$. Dann ist $\mu(S \smallsetminus T) = 0$ und $\mu(T \smallsetminus S) = 0$. Andererseits ist klar, dass $f^{-1}(S) = S$. $\square$

Definition 4.3.7. Sei $(X, \mathcal{A}, \mu)$ Wahrscheinlichkeitsraum. Ein messbares und masserhaltendes $f\colon X \to X$ heisst *ergodisch*, falls jede messbare Teilmenge T von X, die (fast) invariant ist bei f, entweder μ-Mass 0 oder μ-Mass 1 hat. (Wegen Lemma 4.3.6 ist es egal, ob wir hier *fast invariant* oder *invariant* schreiben.)

Jetzt endlich ein paar Beispiele.

Beispiel 4.3.8. Sei $X = [0, 1[$, ausgestattet mit dem üblichen Lebesgue-Mass λ und der dazugehörigen σ-Algebra der Lebesgue-messbaren Teilmengen. Dann ist $(X, \mathcal{A}, \lambda)$ ein Wahrscheinlichkeitsraum. (Für das Weglassen der 1, also $X = [0, 1[$ statt $X = [0, 1]$, gibt es Gründe, die so ähnlich sind wie die in Abschnitt 4.2.) Sei nun α eine reelle Zahl, $0 < \alpha < 1$. Sei $g_\alpha\colon X \to X$ die Abbildung gegeben durch $g_\alpha(x) = x + \alpha$ falls $x + \alpha < 1$, und $g_\alpha(x) = x + \alpha - 1$ sonst. Es ist leicht zu sehen, dass g_α messbar und masserhaltend ist, und auch bijektiv. (Die Inverse dazu ist auch messbar, und automatisch masserhaltend.)

Wenn $\alpha \in \mathbb{Q}$, dann ist g_α nicht ergodisch. Sei etwa $\alpha = p/q$ mit positiven ganzen Zahlen p, q, wobei $p < q$. Sei T die Vereinigung der endlich vielen Intervalle $[r/q, r/q + 1/2q[$ mit $r = 0, 1, 2, \dots, q-1$. Dann ist $\lambda(T) = 1/2$ und $g_\alpha^{-1}(T) = T$. Also ist g_α nicht ergodisch. Andererseits, wenn $\alpha \notin \mathbb{Q}$, dann ist g_α ergodisch, und das ist nicht leicht zu zeigen; wir werden es aber hoffentlich bald schaffen.

Beispiel 4.3.9. Sei $(X, \mathcal{A}, \lambda)$ der Wahrscheinlichkeitsraum von Beispiel 4.3.8. Sei $f\colon X \to X$ definiert durch $f(x) = 2x$ falls $2x < 1$, und $f(x) = 2x - 1$ sonst. Es ist nicht schwer zu zeigen, dass f messbar ist (und ich lasse das aus). Sehr überraschend: f ist auch masserhaltend (obwohl weit davon entfernt, invertierbar zu sein). Zum Beweis sei $T \subset X$ und $T \in \mathcal{A}$. Wir sollen zeigen, dass $\lambda(f^{-1}(T)) = \lambda(T)$ gilt. Man kann schreiben $f^{-1}(T) = C \cup D$, disjunkte Vereinigung, wobei (etwas informell) $C = \frac{1}{2}T$ und $D = \frac{1}{2}T + \frac{1}{2}$. Diese beiden, C und D, haben jeweils das Mass $\lambda(T)/2$, wie man leicht sieht, so dass das Mass von $f^{-1}(T)$ tatsächlich gleich $\mu(T)$ ist. — Dieses f ist wieder ergodisch, und wir werden das hoffentlich auch bald zeigen können.

Beispiel 4.3.10. Sei $X = [0, 1]$ ausgerüstet mit der Lebesgue-σ-Algebra $\mathcal{A}$ und dem üblichen Lebesgue-Mass λ. Sei $f\colon X \to X$ gegeben durch $f(x) = x^{-1} - \lfloor x^{-1} \rfloor$ für $x > 0$ und $f(0) = 0$. Dann ist f messbar, zum Beispiel weil die Einschränkung von f auf jedes der abzählbar vielen Intervalle $]1/(k+1), 1/k]$ messbar ist (für positive ganze Zahlen k). Es ist leicht zu sehen, dass f *nicht* masserhaltend ist für das Mass λ. Wir werden aber ein anderes Mass μ auf X und $\mathcal{A}$ finden oder konstruieren, für das f masserhaltend ist. (Das ist nicht sehr schwer. Dieses Mass μ ist äquivalent zu λ in dem Sinn, dass eine messbare Teilmenge T von X genau dann $\mu(T)=0$ erfüllt, wenn sie $\lambda(T)=0$ erfüllt.) Dann soll auch gezeigt werden, dass f ergodisch ist bezüglich μ.

Bemerkung 4.3.11. Die Abbildung f in Beispiel 4.3.10 hat etwas mit Kettenbrüchen zu tun, und auch mit dem Euklidischen Algorithmus. Wenn wir eine *irrationale* Zahl $x_0 \in [0,1]$ in einen Kettenbruch

$$\frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}}}$$

mit positiven ganzen Zahlen $a_1, a_2, a_3, a_4, \dots$ entwickeln wollen, dann schreiben wir erst

$$x_0 = \frac{1}{x_0^{-1}} = \frac{1}{a_1 + x_1}$$

wobei $a_1 = \lfloor x_0^{-1} \rfloor$ und $x_1 = x_0 - \lfloor x_0^{-1} \rfloor = f(x_0)$ ist. Dann schreiben wir

$$x_1 = \frac{1}{a_2 + x_2}$$

wobei $a_2 = \lfloor x_1^{-1} \rfloor$ und $x_2 = f(x_1)$ ist. Undsoweiter. Mit anderen Worten, die reellen Zahlen x_j usw. erhalten wir als $f^j(x_0)$, und die positiven ganzen Zahlen a_j können wir beschreiben durch $a_j = \lfloor f^{j-1}(x_0) \rfloor$, wobei der Exponent j in beiden Fällen für die j-fache Iteration steht.

Wenn wir eine *rationale* Zahl $x_0 = p/q$ auf dieselbe Weise in einen Kettenbruch entwickeln wollen, dann bricht der Prozess irgendwann ab. Genauer: wenn wir $x_0 = p_0/q_0$ schreiben mit ganzen teilerfremden Zahlen p_0, q_0 wobei $0 < p_0 < q_0$, dann wird $x_1 = r_1/p_0$ wobei r_1 der ganzzahlige Rest von q_0 nach Teilen durch p_0 ist. Ebenso wird $x_2 = r_2/r_1$ wobei r_2 der ganzzahlige Rest von p_0 nach Teilen durch r_1 ist. Wir führen also den Euklidischen Algorithmus durch für das Zahlenpaar (p_0, q_0). Nach endlich vielen Schritten wird der Rest r_j gleich Null und der Prozess endet. Beispiel:

$$10/23 = \frac{1}{2 + \frac{1}{3 + \frac{1}{3}}}.$$

In so einem Fall kürzt man gerne ab $10/23 = [2,3,3]$, indem man nur die positiven ganzen Zahlen auflistet, die bei der Kettenbruchentwicklung entstehen. Ebenso kann man im Fall einer irrationalen Zahl $x_0 \in [0,1]$ schreiben:

$$x_0 = [a_1, a_2, a_3, a_4, \dots]$$

wobie die a_j die positiven ganzen Zahlen sind, die bei der Kettenbruchentwicklung entstehen. Die Gleichung $x_0 = [a_1, a_2, a_3, a_4, \dots]$ darf dann auch so verstanden werden:

$$x_0 = \lim_{r \to \infty} [a_1, a_2, \dots, a_r]$$

wobei $[a_1, a_2, \dots, a_r]$ für jedes $r \geq 1$ eine *rationale* Zahl ist. Hier sollte die Existenz des Grenzwerts eigentlich bewiesen werden; das ist aber auch nicht schwer.

4.4. Der Wiederkehrsatz von Poincaré

Proposition 4.4.1. *Sei $(X, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum, $f\colon X \to X$ messbar und masserhaltend. Sei $E \subset X$ messbar mit $\mu(E) > 0$. Für fast jedes $x \in E$ gibt es unendlich viele $n \geq 0$ derart, dass $f^n(x) \in E$.*

Beweis. Die Menge $\{x \in X \mid f^n(x) \in E \text{ für } \infty \text{ viele } n \geq 0\}$ kann so beschrieben werden:

$$\bigcap_{n\geq 0}\left(\bigcup_{k\geq n} f^{-k}(E)\right).$$

Deswegen interessieren wir uns für

$$S_n = \bigcup_{k\geq n} f^{-k}(E)$$

wobei $n \geq 0$. Dann ist $S_n \supset S_{n+1}$. Andererseits ist $f^{-1}(S_{n+1}) = S_n$. Weil f masserhaltend ist, folgt $\mu(S_{n+1}) = \mu(S_n)$, das heisst, $S_n \smallsetminus S_{n+1}$ ist eine μ-Nullmenge. Daraus folgt mit σ-Additivität, dass

$$S_0 \smallsetminus \bigcap_{n\geq 0} S_n = (S_0 \smallsetminus S_1) \cup (S_1 \smallsetminus S_2) \cup (S_2 \smallsetminus S_3) \cup \cdots$$

immer noch eine μ-Nullmenge ist. Dann ist erst recht $E \smallsetminus \bigcap_{n\geq 0} S_n$ eine μ-Nullmenge, denn $E \subset S_0$. Und das war die Behauptung. □

FRAGE 4.4.2. Geht das auch, wenn wir erlauben, dass $\mu(X) = \infty$ ist (also $(X, \mathcal{A}, \mu)$ kein Wahrscheinlichkeitsraum, sondern nur Massraum) ? Wenn nicht, wo geht es schief?

Ein paar Übungsaufgaben. Lange Geschichte: da wir schon den Begriff *messbare Abbildung* haben, liegt es nahe, zu fragen, ob stetige Abbildungen unter halbwegs vernünftigen Voraussetzungen auch messbar sind.

1. Sei $f\colon X \to Y$ ein stetige Abbildung zwischen topologischen Räumen. Sei $\mathcal{B}_X \subset \mathcal{P}(X)$ die Borel-Algebra von X, also die kleinste σ-Algebra, die alle offenen Mengen von X als Elemente enthält. Ebenso sei $\mathcal{B}_Y$ die Borel-Algebra von Y. Dann ist f messbar bezüglich $\mathcal{B}_X$ und $\mathcal{B}_Y$. (Da ist kaum etwas zu beweisen.)

2. Sei $S \subset [0,1[$ die nicht-messbare Menge von Vitali, Abschnitt 4.2. Zeigen Sie, dass $\lambda_*(S) = 0$. Leiten Sie daraus her: wenn $A \subset [0,1]$ messbar ist (im Sinn von Abschnitt 3.1) und $\lambda(A) > 0$, dann ist mindestens eine der Mengen $A \cap V_a(S)$ nicht messbar (wobei $a \in \mathbb{Q} \cap [0,1[$, Bezeichnung V_a wie Abschnitt 4.2).

3. Sei $A \subset [0,1]$ abgeschlossen mit $\lambda(A) > 0$; angenommen $W = [0,1] \smallsetminus A$ ist dicht in $[0,1]$, wie im Beispiel von Abschnitt 2.2. (Genauer, man nehme $A = [0,1] \smallsetminus U$ wobei U wie in Lemma 2.2.2.) Dann ist $\lambda(W) > 0$, folglich $\lambda(A) < 1$. Wir definieren eine Abbildung[5] $g\colon [0,1] \longrightarrow [0,1]$ durch

$$g(t) = \frac{\lambda(W \cap [0,t])}{\lambda(W)}.$$

Zeigen Sie: g ist ein Homöomorphismus von $[0,1]$ nach $[0,1]$. Es folgt, dass $g(W)$ offen ist in $[0,1]$. Zeigen Sie auch, dass $\lambda(g(W)) = 1$. Es folgt, dass $\lambda(g(A)) = 0$.

4. Leiten Sie aus den Aufgaben 2 und 3 her: der Homöomorphismus $g\colon [0,1] \to [0,1]$ ist nicht messbar, wenn $[0,1]$ mit der σ-Algebra $\mathcal{A}$ aller Lebesgue-messbaren Mengen ausgestattet wird. (Wegen Aufgabe 1 ist g allerdings messbar, wenn $[0,1]$ mit der Borel-σ-Algebra $\mathcal{B}_{[0,1]}$ ausgestattet wird. Ausserdem ist g als stetige Funktion Riemann-integrierbar. Zum Integral von g fällt mir leider nicht viel ein.)

5. (Anderes Thema.) Sei $(X, \mathcal{A}, \lambda)$ der Wahrscheinlichkeitsraum von Beispiel 4.3.9 und $f\colon X \to X$ wie dort. Sei $S = \{0,1\}$ ausgerüstet mit der σ-Algebra $\mathcal{B} = \mathcal{P}(S)$ und dem

[5] Eine hübsche Idee, die ich vorübergehend für meine eigene Idee gehalten habe. Aber sie kommt wohl aus dem Beweis des Lebesgue Dichtesatzes von C.A.Faure, Abschnitt 5.2.

Wahrscheinlichkeitsmass μ, bei dem $\{0\}$ und $\{1\}$ beide das Mass $\frac{1}{2}$ haben. Dann ist $(S, \mathcal{B}, \mu)$ ebenfalls Wahrscheinlichkeitsraum. Hier interessieren wir uns aber mehr für das unendliche Produkt

$$Y := \prod_{i=1}^{\infty} S \quad = \quad S \times S \times S \times S \times \cdots$$

ausgerüstet mit der Produkt-σ-Algebra und dem Produktmass $\mu \times \mu \times \cdots$, die ich jetzt mit $\mathcal{B}_\infty$ bzw. mit μ_∞ bezeichne. Einzelheiten wie in Abschnitt 3.2. Dann ist also $(Y, \mathcal{B}_\infty, \mu_\infty)$ wieder ein Wahrscheinlichkeitsraum. — Eine Abbildung $\Phi\colon X \to Y$ kann definiert werden durch

$$x \mapsto (f(x), f(f(x)), f(f(f(x))), f(f(f(f(x)))), \dots) \in Y.$$

Man zeige, dass Φ messbar (bezüglich $\mathcal{A}$ und $\mathcal{B}_\infty$), masserhaltend (bezüglich λ und μ_∞) und fast invertierbar ist. Das heisst, es existiert messbares und masserhaltendes $\Psi\colon Y \to X$ derart, dass $\Psi \circ \Phi = \mathrm{id}_X$ fast überall und $\Phi \circ \Psi = \mathrm{id}_Y$ fast überall.

CHAPTER 5

Notizen zur 7. Vorlesungswoche

5.1. Dichtesatz von Lebesgue (als Hilfsmittel)

SATZ 5.1.1. (Lebesgues Dichtesatz für $\mathbb{R}$.) *Sei E eine beliebige Teilmenge von $\mathbb{R}$. Dann gilt für fast alle*[1] $a \in E$, *dass*

$$\lim_{\varepsilon_1,\varepsilon_2\to 0} \frac{\lambda^*(E\cap[a-\varepsilon_1,a+\varepsilon_2])}{\varepsilon_1+\varepsilon_2} = 1.$$

Dabei ist λ^* wie üblich das äussere Lebesgue-Mass. Beachten: wenn E messbar ist, dann ist auch $E\cap[a-\varepsilon_1,a+\varepsilon_2]$ messbar, und man kann überall λ statt λ^* schreiben. Dieser Fall würde uns auch genügen, aber es ist nicht klar, ob der Beweis dadurch einfacher wird. Ich schiebe den Beweis erstmal auf, denn er ist nicht einfach.

KOROLLAR 5.1.2. *Sei $(X,\mathcal{A},\lambda)$ der Wahrscheinlichkeitsraum und $g = g_\alpha$ die masserhaltende und bijektive Abbildung von Beispiel 4.3.8, und sei $\alpha\notin\mathbb{Q}$. Dann ist g ergodisch.*

Beweis. Wir überlegen uns zuerst: *zu jedem positiven $m\in\mathbb{Z}$ existiert $n\in\mathbb{Z}$ mit $|n|\le m$ derart, dass* $0<g^n(0)<1/m$ (der Exponent bedeutet Iteration). Das ist eine typische und bekannte Anwendung von Dirichlets Schubfachprinzip. Dazu teilen wir das Intervall $X=[0,1[$ in m Teilintervalle $[0,1/m[$, $[1/m,2/m[$, ... gleicher Länge $1/m$ auf (die Schubfächer). Es muss dann zwei verschiedene $j,k\in\{0,1,\dots,m\}$ geben mit $g^j(0)$ und $g^k(0)$ im selben Schubfach. OBdA ist $g^j(0)\le g^k(0)$. Dafür können wir auch schreiben $j\alpha-\lfloor j\alpha\rfloor < k\alpha-\lfloor k\alpha\rfloor$; Gleichheit ausgeschlossen, weil sonst $\alpha\in\mathbb{Q}$ wäre. Sei $n=k-j$; dann liegt $g^n(0)=n\alpha-\lfloor n\alpha\rfloor$ im Intervall $[0,1/m]$ wie gewünscht (und es kann kein Endpunkt davon sein).

Daraus folgt weiter: *zu jedem $t\in X$ und $\varepsilon>0$ existiert $r\in\mathbb{Z}$ derart, dass $|g^r(0)-t|<\varepsilon$.* Denn oBdA ist $\varepsilon=1/m$ mit $m\in\mathbb{Z}$ wie oben, und wir können dann $n\in\mathbb{Z}$ so finden, dass $s:=g^n(0)$ die Ungleichungen $0<s<1/m=\varepsilon$ erfüllt. Unter den Zahlen $s,2s,3s,\dots,\lfloor s^{-1}\rfloor s$ muss es eine geben, sagen wir qs mit $q\in\mathbb{Z}$, $q\neq 0$, deren Abstand von t minimal ist, und er ist dann höchstens s, also $<\varepsilon$. Dann können wir auch schreiben $qs=g^{qn}(0)$.

So weit, so gut; nun sei $T\subset X$ messbar und es möge gelten $g^{-1}(T)=T$; gleichwertig dazu $g(T)=T$, weil g invertierbar ist. Für eine positive ganze Zahl n und $p\in\{0,\dots,n-1\}$ sei $J_{p,n}=[\frac{p}{n},\frac{p}{n}+\frac{1}{n}[$. Wir zeigen jetzt:

$$\lambda(T\cap J_{p,n})=\lambda(T\cap J_{0,n}).$$

Dazu sei irgendein $\varepsilon>0$ vorgegeben. OBdA ist $p>0$. Wie vorher gezeigt, finden wir dann $r\in\mathbb{Z}$ derart, dass $|g^r(0)-\frac{p}{n}|<\varepsilon$. Andererseits ist

$$\lambda(T\cap J_{0,n})=\lambda(g^r(T)\cap g^r(J_{0,n}))=\lambda(T\cap g^r(J_{0,n}))$$

und weil $g^r(J_{0,n})\smallsetminus J_{p,n}$ sowie $J_{p,n}\smallsetminus g^r(J_{0,n})$ beide λ-Mass $<\varepsilon$ haben, schliessen wir daraus

$$|\lambda(T\cap J_{0,n})-\lambda(T\cap J_{p,n})|<2\varepsilon$$

[1] Bedeutet: diejenigen $a\in E$, für die es nicht gilt, bilden eine Menge vom Lebesgue-Mass Null; gleichwertig dazu, eine Menge vom äusseren Lebesgue-Mass 0.

und damit Gleichheit, $\lambda(T \cap J_{0,n}) = \lambda(T \cap J_{p,n})$, denn ε war beliebig klein. Daraus folgt natürlich, dass

$$\lambda(T \cap J_{p,n}) = \lambda(T)/n.$$

Und daraus folgt weiter, dass für alle $a \in T$ gilt:

$$\liminf_{\varepsilon_1,\varepsilon_2\to 0} \frac{\lambda(T \cap [a-\varepsilon_1, a+\varepsilon_2])}{\varepsilon_1+\varepsilon_2} \leq \lambda(T),$$

denn wir können ε_1 und ε_2 so wählen, dass $[a-\varepsilon_1, a+\varepsilon_2[$ eins der Intervalle $J_{n,p}$ wird, mit beliebig grossem n (und dann wird $\varepsilon_1+\varepsilon_2 = 1/n$). Wenn nun $0 < \lambda(T) < 1$ wäre, dann würde das dem Satz 5.1.1 widersprechen. Also ist $\lambda(T) = 0$ oder $\lambda(T) = 1$. □

Korollar 5.1.3. *Sei $(X, \mathcal{A}, \lambda)$ der Wahrscheinlichkeitsraum und f die masserhaltende und bijektive Abbildung von Beispiel 4.3.9. Dieses f ist ergodisch.*

Beweis. Der Beweis ähnelt dem des vorigen Korollars, ist aber kürzer. Für beliebiges $T \subset X$ können wir (etwas informell) sagen, dass $f^{-n}(T) \subset X$ die disjunkte Vereinigung der 2^n Mengen

$$2^{-n}T + k \cdot 2^{-n}$$

ist, wobei $k = 0, 1, \ldots, 2^n - 1$. Das ist so gemeint: $aT + b := \{ax + b \mid x \in T\} \subset \mathbb{R}$. Wenn T messbar ist, dann sind diese Mengen $2^{-n}T + k \cdot 2^{-n}$ auch messbar und alle haben das gleiche Mass, $2^{-n}\lambda(T)$. Wenn ausserdem $f^{-1}(T) = T$ gilt, dann folgt $f^{-n}(T) = T$, und es folgt auch, dass

$$T \cap [k \cdot 2^{-n}, (k+1) \cdot 2^{-n}[$$

das Mass $2^{-n}\lambda(T)$ hat. Daraus folgt wieder, dass für fast alle $a \in T$ gilt:

$$\liminf_{\varepsilon_1,\varepsilon_2\to 0} \frac{\lambda(T \cap [a-\varepsilon_1, a+\varepsilon_2])}{\varepsilon_1+\varepsilon_2} \leq \lambda(T),$$

denn wenn $a \notin \mathbb{Q}$, dann können wir ε_1 und ε_2 so wählen, dass $[a-\varepsilon_1, a+\varepsilon_2[$ eins der Intervalle $[k \cdot 2^{-n}, (k+1) \cdot 2^{-n}[$ wird, mit beliebig grossem n (und dann wird $\varepsilon_1+\varepsilon_2 = 2^{-n}$). Wenn nun $0 < \lambda(T) < 1$ wäre, dann würde das dem Satz 5.1.1 widersprechen. Also ist $\lambda(T) = 0$ oder $\lambda(T) = 1$. □

5.2. Beweis vom Dichtesatz

(Nach C.A. Faure, *A short proof of Lebesgue's density theorem*, Amer. Math. Monthly 109 (2002), 194–196.)

Lemma 5.2.1. *Sei $g\colon [a,b] \to \mathbb{R}$ stetig und sei $U \subset \mathbb{R}$ offen, ausserdem $U \subset [a,b]$. Dann ist $U_g := \{x \in U \mid \exists y > x \text{ mit }]x,y[\subset U \text{ und } g(x) > g(y)\}$ auch offen in $\mathbb{R}$. Wenn $]c,d[$ Zusammenhangskomponente von U_g, dann $g(c) \geq g(d)$.*

Beweis. U_g offen: für $x_0 \in U_g$ existiert $y > x_0$ mit $]x_0, y[\subset U$ und $g(x_0) > g(y)$. Wenn $\delta > 0$ klein genug ist, dann ist $]x_0 - \delta, x_0 + \delta[\subset U$, denn U offen in $\mathbb{R}$, ausserdem $g(x) > g(y)$ für alle $x \in]x_0 - \delta, x_0 + \delta[$, denn g stetig; ausserdem $x < y$ für alle $x \in]x_0 - \delta, x_0 + \delta[$. Damit $x \in U_g$ für alle $x \in]x_0 - \delta, x_0 + \delta[$. — Behauptung $g(c) \geq g(d)$: Angenommen falsch, dann $\exists x_1 \in]c,d[$ mit $g(x_1) < g(d)$. OBdA ist $g(x_1) \leq g(x)$ für alle $x \in [x_1, d]$, denn sonst ersetzen wir x_1 durch eine Stelle in $[x_1, d]$, an der g ein Min annimmt. Nach Voraussetzung $\exists y > x_1$ mit $]x_1, y[\subset U$ und $g(x_1) > g(y)$. Letztere Ungleichung hat zur Folge $y \notin [x_1, d]$, also $y > d$. Daher $d \in U_g$, Widerspruch zu Annahme. □

Beweis vom Dichtesatz. Sei n positive ganze Zahl und $I_n = [-n, n]$. Sei $s \in \mathbb{R}$ fest gewählt, $0 < s < 1$. Es genügt, zu zeigen, dass die Menge

$$A_n :=]-n, n[\cap \left\{ x \in E \;\middle|\; \liminf_{\varepsilon \to 0} \frac{\lambda^*(E \cap [x, x+\varepsilon])}{\varepsilon} < s \right\}$$

eine λ-Nullmenge ist, also $\lambda^*(A_n) = 0$. Wir definieren $g \colon I_n \to \mathbb{R}$ durch

$$g(x) := \lambda^*(E \cap [-n, x]) - sx.$$

Dann ist g stetig. Wähle offenes $U \subset \mathbb{R}$ mit $A_n \subset U \subset I_n$. Die Definition von A_n ergibt $A_n \subset U_g$ wobei U_g wie in Lemma 5.2.1. Sei $]c, d[$ eine Zusammenhangskomponente von U_g. Dann sagt das Lemma $g(c) \geq g(d)$ und daher $g(d) - g(c) \leq 0$ und daher

$$\lambda^*(E \cap]c, d]\,) = \lambda^*(E \cap]c, d[\,) \leq s(d - c)$$

wegen Definition von g. Wenn wir jetzt über die höchstens abzählbar vielen Zusammenhangskomponenten von U_g summieren, erhalten wir

$$\lambda^*(E \cap U_g) \leq s\lambda(U_g).$$

Ausserdem ist $A_n \subset E \cap U_g \subset U$, also $\lambda^*(A_n) \leq s\lambda(U_g) \leq s\lambda(U)$. Andererseits wurde bei Wahl von U nur vorausgesetzt $A_n \subset U \subset I_n$ und U offen in $\mathbb{R}$. Das heisst nach Definition von λ^*, dass wir bei vorgegebenem $\varepsilon > 0$ durch geeignete Wahl von U erreichen können $\lambda(U) - \lambda^*(A_n) < \varepsilon$. Die Ungleichung $\lambda^*(A_n) \leq s\lambda(U)$ kann damit nur in Einklang sein, wenn $\lambda^*(A_n) = 0$. $\square$

CHAPTER 6

Notizen zur 8. und 9. Vorlesungswoche

Eins der nächsten Ziele sollte sein: Beweis der Behauptung, dass $f\colon X \to X$ in Beispiel 4.3.10 ergodisch ist. Dabei war $X = [0,1]$. Weil aber f nicht masserhaltend ist für das Lebesgue-Mass λ auf $[0,1]$, wurde Bezug genommen auf ein anderes Mass μ, das aus λ durch Multiplizieren mit einer passenden *Gewichtsfunktion* θ entsteht. Die Definition ist ungefähr

$$\mu(T) := \int_T \theta(x)\ dx$$

wobei allerdings die rechte Seite noch nicht definiert ist (weil T ja nicht unbedingt ein Intervall sein muss) und die Bezeichnungen überhaut fragwürdig sind. (Man schreibt wohl auch $\int_T \theta\ d\lambda$.) Deswegen müssen wir uns erst noch mit *Integration* im Sinne von Lebesgue beschäftigen. Darüber ist in Abschnitt 1.1 schon etwas gesagt worden. Die dort angegebene Methode gibt leider nicht die allerbeste und allgemeinste Grundlage, obwohl sie für die Zwecke der ersten Abschnitte schon ganz gut war.

6.1. Integration nach Lebesgue

Sei $(X, \mathcal{E}, \omega)$ ein Massraum. Sei $f\colon X \to \mathbb{R}$ eine Funktion, die messbar ist bezüglich $\mathcal{E}$ und $\mathcal{B}$, wobei $\mathcal{B} \subset \mathcal{P}(\mathbb{R})$ die Borel-σ-Algebra ist (kleinste σ-Unteralgebra von $\mathcal{P}(\mathbb{R})$, die alle offenen Teilmengen von $\mathbb{R}$ enthält). Diese Bedingung an f ist äquivalent zur Folgenden: für jedes abgeschlossene Intervall $J \subset \mathbb{R}$ ist $f^{-1}(J) \in \mathcal{E}$. (Das liegt daran, dass $\mathcal{B}$ auch die kleinste σ-Unteralgebra von $\mathcal{P}(\mathbb{R})$ ist, die alle abgeschlossenen Intervalle J enthält.) Wir wollen versuchen, so etwas wie $\int f\ d\omega \in \mathbb{R}$ zu definieren, das Integral von f bezüglich ω. Um es einfach zu halten, nehmen wir vorläufig an:

- $\omega(X) < \infty$;
- f ist beschränkt, genauer, $\exists c > 0$ so dass $-c < f(x) < c$ für alle $x \in X$.

Die Definition des Integrals ist dann sehr einfach. Wir halten die Schranke c fest. Für jede Partition

$$-c = a_0 < a_1 < \cdots < a_{r-1} < a_r = c$$

von $[-c, c]$ können wir

$$\mathcal{I}_f(a_0, a_1, \dots, a_r) := \sum_{j=0}^{r-1} a_j \cdot \omega(f^{-1}([a_j, a_{j+1}[\,))$$

bilden. Wichtig: weil f messbar ist bezüglich $\mathcal{E}$ und $\mathcal{B}$, ist $\omega(f^{-1}([a_j, a_{j+1}[\,))$ definiert. Die rechte Seite kann man auffassen als das ω-Integral einer Art von *Treppenfunktion*, die auf den Mengen $f^{-1}([a_j, a_{j+1}[\,)$ jeweils konstant ist mit Wert a_j. Und diese treppen-artige Funktion ist eine gute Annäherung an f, wenn die Abstände $a_j - a_{j-1}$ allesamt klein sind. Deswegen ist es naheliegend, die Bezeichnung $\|(a_0, a_1, \dots, a_r)\|$ für das Maximum dieser Abstände $a_j - a_{j-1}$ einzuführen, $j = 1, 2, \dots, r$.

DEFINITION 6.1.1. (Eigentlich mehr ein Satz.)

$$\int f\ d\omega := \lim_{\substack{(a_0,a_1,\ldots,a_{r-1},a_r) \\ \|(a_0,a_1,\ldots,a_{r-1},a_r)\|\to 0}} \mathcal{I}_f(a_0,a_1,\ldots,a_r).$$

Die Existenz des Limes ergibt sich aus der folgenden Behauptung: wenn $(a_0,a_1,\ldots,a_r)$ und $(b_0,b_1,\ldots,b_s)$ Partitionen von $[-c,c]$ sind mit $\|(a_0,a_1,\ldots,a_r)\|\le\varepsilon$ und $\|(b_0,b_1,\ldots,b_s)\|\le\varepsilon$, dann ist

$$|\mathcal{I}_f(a_0,a_1,\ldots,a_r)-\mathcal{I}_f(b_0,b_1,\ldots,b_r)|\le\varepsilon\cdot\omega(X).$$

Um das zu zeigen, nehmen wir zuerst an, dass $(b_0,\ldots,b_r)$ feiner ist als $(a_0,a_1,\ldots,a_r)$, das heisst, alle a_i kommen unter den b_j vor. Für $j=0,1,2,\ldots,s$ sei $u(j)$ das grösste der $i\in\{0,1,\ldots,r\}$ mit $a_i\le b_j$. Dann ist

$$\mathcal{I}_f(a_0,a_1,\ldots,a_r)=\textstyle\sum_{i=0}^{r-1}a_i\cdot\omega(f^{-1}([a_i,a_{i+1}[\,))=\sum_{j=0}^{s-1}a_{u(j)}\cdot\omega(f^{-1}([b_j,b_{j+1}[\,))$$

(weil $[a_i,a_{i+1}[$ die disjunkte Vereinigung von gewissen $[b_j,b_{j+1}[$ ist) und daher

$$\begin{array}{rcccl} 0 & \le & \mathcal{I}_f(b_0,b_1,\ldots,b_r)-\mathcal{I}_f(a_0,a_1,\ldots,a_r) & = & \sum_{j=0}^{s-1}(b_j-a_{u(j)})\cdot\omega(f^{-1}([b_j,b_{j+1}[\,)) \\ & \le & \sum_{j=0}^{s-1}\varepsilon\cdot\omega(f^{-1}([b_j,b_{j+1}[)) & = & \varepsilon\cdot\omega(X). \end{array}$$

Jetzt der allgemeine Fall: wenn $(a_0,a_1,\ldots,a_r)$ und $(b_0,b_1,\ldots,b_s)$ zwei Partitionen von $[-c,c]$ sind mit $\|(a_0,a_1,\ldots,a_r)\|\le\varepsilon$ und $\|(b_0,b_1,\ldots,b_s)\|\le\varepsilon$, dann finden wir eine dritte Partition $(c_0,c_1,\ldots,c_t)$, die feiner ist als beide. Dann ist nach dem Vorhergehenden

$$0\le\mathcal{I}_f(c_0,c_1,\ldots,c_t)-\mathcal{I}_f(a_0,a_1,\ldots,a_r)\le\varepsilon\cdot\omega(X)$$

und ebenso $0\le\mathcal{I}_f(c_0,c_1,\ldots,c_t)-\mathcal{I}_f(a_0,a_1,\ldots,a_r)\le\varepsilon\cdot\omega(X)$, und daraus folgt

$$|\mathcal{I}_f(a_0,a_1,\ldots,a_r)-\mathcal{I}_f(b_0,b_1,\ldots,b_s)|\le\varepsilon\cdot\omega(X).$$

BEMERKUNG 6.1.2. (i) Die Definition 6.1.1 erinnert etwas an die Definition des Riemann-Integrals. Ein wichtiger Unterschied liegt darin, dass beim Riemann-Integral die “Quelle”, in unserer Situation wäre das X, in kleine Stücke geteilt wird, während hier das Ziel $\mathbb{R}$ oder $[-c,c]$ in kleine Stücke geteilt wird. Manche Spezialisten vertreten die Ansicht, dass das überhaupt der wesentliche Unterschied zwischen Riemann-Integration und Lebesgue-Integration ist. Ich tue das nicht; aber es ist schon richtig, dass Definition 6.1.1 uns auf den guten Weg führt, der darin besteht, dass wir die Integrationstheorie auf der Masstheorie aufbauen und nicht umgekehrt.

(ii) Es ist wichtig, dass wir die Schranke c so wählen, dass $-c<f(x)<c$ für alle $x\in X$ gilt (und nicht nur etwa $-c\le f(x)\le c$). Ebenso ist es wichtig, dass wir in der Definition von $\mathcal{I}_f(a_0,a_1,\ldots,a_r)$ die halboffenen Intervalle $[a_i,a_{i+1}[$ benutzen und nicht etwa die abgeschlossenen $[a_i,a_{i+1}]$. Man versteht, dass es wichtig ist, wenn man für f eine konstante Funktion wählt. Andererseits ist es klar, dass der Wert von $\int f\ d\omega$ wie in Definition 6.1.1 nicht von der Wahl der Schranke c abhängt, solange $-c<f(x)<c$ für alle $x\in X$ gilt.

(iii) Wenn man darauf Wert legt, dass es viele messbare Funktionen von $(X,\mathcal{E},\omega)$ nach $\mathbb{R}$ gibt, dann sollte man $\mathcal{E}$ möglichst gross wählen, soweit das Mass ω sich dazu hergibt. Andererseits hilft es, wenn man für das Ziel $\mathbb{R}$ eine eher kleine σ-Algebra wählt; deswegen fiel die Entscheidung für die Borel-Algebra $\mathcal{B}$ und nicht für $\mathcal{A}$, die σ-Algebra *aller* Lebesgue-messbaren Teilmengen von $\mathbb{R}$ (die grösser ist als $\mathcal{B}$). Wenn man stattdessen $\mathcal{A}$ wählt (für das Ziel $\mathbb{R}$), dann wird man sich darüber ärgern, dass es stetige Funktionen von beispielsweise $[0,1]$ nach $\mathbb{R}$ gibt, die nicht messbar sind bezüglich $\mathcal{E}$ und $\mathcal{A}$, wobei $\mathcal{E}$ jetzt die Menge aller Lebesgue-messbaren Teilmengen von $[0,1]$ ist. (Davon handeln einige der Übungsaufgaben am Ende von Kapitel 4.)

(iv) Anstelle der Definition 6.1.1 kann man auch schreiben $\int f\, d\omega := \sup\{\mathcal{I}_f(a_0, a_1, \ldots, a_r)\}$, wobei das Supremum über alle Partitionen von $[-c, c]$ zu nehmen ist. Dann ist es eine echte Definition. Trotzdem sollte man sich ein paar Dinge klar machen, die in dieser kurzen Formulierung nicht genügend betont werden. Die Menge der Partitionen $(a_0, a_1, \ldots, a_r)$ von $[-c, c]$ ist mit einer (partiellen) Ordnungsrelation ausgerüstet, weil wir definieren können: $(a_0, a_1, \ldots, a_r) \leq (b_0, b_1, \ldots, b_s)$ genau dann, wenn $\{a_0, ..., a_r\} \subset \{b_0, ..., b_s\}$. Die Funktion $\mathcal{I}_f$ ist *monoton*, das heisst $(a_0, a_1, \ldots, a_r) \leq (b_0, b_1, \ldots, b_s)$ hat zur Folge $\mathcal{I}_f(a_0, a_1, \ldots, a_r) \leq \mathcal{I}_f(b_0, b_1, \ldots, b_s)$. Das haben wir schon nachgerechnet. Das bedeutet: $\mathcal{I}_f(a_0, a_1, \ldots, a_r)$ kann den wahren Wert $\int f\, d\omega$ nur unterschätzen, nicht überschätzen, und die Abschätzung kann nur besser werden, nicht schlechter, wenn wir eine Partition durch eine feinere ersetzen.

(v) In Brokate-Kersting (Kapitel 4) wird noch radikaler vorgegangen. Unter einer *Elementarfunktion* $h\colon X \to \mathbb{R}$ wird eine messbare Funktion verstanden, die nur endlich viele verschiedene Werte annimmt, also $h(X)$ endliche Teilmenge von $\mathbb{R}$. Die radikale Definition geht etwa so:

$$\int f\, d\omega := \sup_{0 \leq h \leq f} \left(\sum_{t>0} t \cdot \omega(h^{-1}(t)) \right).$$

Hier wird *nicht* vorausgesetzt, dass f beschränkt ist (sondern nur messbar); es wird *nicht* vorausgesetzt, dass $\omega(X) < \infty$; aber es wird vorausgesetzt, dass $f \geq 0$. Das Supremum wird genommen über alle Elementarfunktionen h, die $0 \leq h \leq f$ erfüllen. Der Ausdruck $\sum_{t \in \mathbb{R}} t \cdot \omega(h^{-1}(t))$ ist genau das, was wir uns unter dem Integral einer Elementarfunktion h vorstellen, falls $h \geq 0$; beachten, dass die Summe eigentlich eine endliche Summe ist, da $h^{-1}(t)$ nach Voraussetzung nur für endlich viele $t \in \mathbb{R}$ nicht leer ist. Es kann vorkommen, dass $\omega(h^{-1}(t)) = \infty$ für gewisse $t > 0$; dann ist unweigerlich $\int f\, d\omega = \infty$. — In den Fällen, wo $\omega(X) < \infty$ und f von unten beschränkt ist, aber nicht unbedingt von oben, kann man auch gefahrlos definieren

$$\int f\, d\omega := \sup_{h \leq f} \left(\sum_{t>0} t \cdot \omega(h^{-1}(t)) \right)$$

wobei nicht $h \geq 0$ verlangt wird. Diese Definition ist im Einklang mit Definition 6.1.1, falls f beidseitig beschränkt ist. Der einzige Gedanke zum Beweis davon ist, dass $\mathcal{I}_f(a_0, a_1, \cdots, a_r) = \sum_{t>0} t \cdot \omega(h^{-1}(t))$ gilt, wenn h die *grösste* unter allen Elementarfunktionen ist, die $\leq f$ sind und deren Bild in $\{a_0, a_1, \ldots, a_r\}$ enthalten ist. Dieses h muss auf der Menge $f^{-1}([a_j, a_{j+1}[\,)$ konstant sein mit dem Wert a_j.

(v) Wir können auch die Abkürzung

$$\mathcal{I}_f^+(a_0, a_1, \ldots, a_r) := \sum_{j=1}^{r} a_j \cdot \omega(f^{-1}(\,]a_{j-1}, a_j]))$$

einführen und dann neu definieren

$$\int f\, d\omega := \inf_{(a_0, \ldots, a_r)} \mathcal{I}_f^+(a_0, a_1, \ldots, a_r) = \lim_{\substack{(a_0, a_1, \ldots, a_{r-1}, a_r) \\ \|(a_0, a_1, \ldots, a_{r-1}, a_r)\| \to 0}} \mathcal{I}_f^+(a_0, a_1, \ldots, a_r).$$

Das ist in Übereinstimmung mit Definition 6.1.1, weil

$$|\mathcal{I}_f^+(a_0, a_1, \ldots, a_r) - \mathcal{I}_f(a_0, a_1, \ldots, a_r)| \leq 2 \cdot \|(a_0, a_1, \ldots, a_r)\| \cdot \omega(X)$$

wie man leicht nachrechnet. (An dieser Gymnastik kommt man kaum vorbei, wenn man zeigen will, dass $\int -f\, d\omega = -\int f\, d\omega$.)

(vi) Aus alledem folgt, dass wir mit der Definition von $\int f\, d\omega$ flexibel sein müssen, je nach Gegebenheiten und Wünschen. Wenn wir zum Beispiel wollen, dass das Integral gute

Linearitätseigenschaften hat, dann kann es klug sein, nur beschränkte f zuzulassen und ausserdem $\omega(X) < \infty$ anzunehmen. Wenn wir andererseits $\omega(X) = \infty$ erlauben wollen und auch unbeschränkte f, dann müssen wir wohl so etwas wie $f \geq 0$ annehmen. Für allgemeinere messbare f kann man es mit Tricks wie $f = f^+ - f^-$ versuchen. Dann ist es mit der Linearität vom Integral nicht so gut bestellt.

6.2. Satz von der monotonen Konvergenz (Beppo Levi)

In diesem Abschnitt soll (zuerst) angenommen werden, dass $(X, \mathcal{E}, \omega)$ irgendein Massraum ist, und wir benutzen die Integraldefinition wie in Bemerkung 6.1.2, also

$$\int f \; d\omega = \sup_{0 \leq h \leq f} \sum_{z \in \mathbb{R}} z \cdot \omega(h^{-1}(z))$$

wobei $f \colon X \to [0, \infty]$ messbar sein soll, $f \geq 0$, und das h für Elementarfunktionen steht (die jeweils nur endlich viele *reelle* Werte annehmen dürfen, also ∞ ausgeschlossen). Der Zielraum $[0, \infty]$ ist mit der Borel-σ-Algebra $\mathcal{B}$ ausgestattet[1] (kleinste σ-Algebra, die alle offenen Teilmengen von $[0, \infty]$ enthält).

Satz 6.2.1. *Sei* $(f_j \colon X \to [0, \infty])_{j=0,1,2,\ldots}$ *eine Folge von messbaren Funktionen bezüglich* $\mathcal{E}$ *und* $\mathcal{B}$, *wobei* $f_j \leq f_{j+1}$ *für alle* j *und* $0 \leq f_0$. *Dann ist die Funktion* $f = \sup_j f_j = \lim_{j \to \infty} f_j$ *wieder messbar, und es gilt*

$$\int f \; d\omega = \sup_j \int f_j \; d\omega = \lim_{j \to \infty} \int f_j \; d\omega < \infty.$$

Beweis. (Nach Brokate-Kersting.) Messbarkeit von f: wir wählen $a, b \in [0, \infty]$ und schreiben $K = [a, \infty]$ und $L = [0, b]$. Dann ist

$$f^{-1}(K) = \bigcap_{n=1}^{\infty} \bigcup_{j=0}^{\infty} f_j^{-1}([a - n^{-1}, \infty]), \qquad f^{-1}(L) = \bigcap_{j \geq 0} f_j^{-1}(L)$$

und daher $f^{-1}(K), f^{-1}(L) \in \mathcal{E}$. Messbarkeit von f bewiesen.

Zur Abkürzung schreibe ich jetzt $\mathcal{J}(h) := \sum_{z \in \mathbb{R}} z \cdot \omega(h^{-1}(z))$ falls $h \colon X \to [0, \infty[$ Elementarfunktion. Es genügt, zu zeigen: für jede Elementarfunktion h mit $0 \leq h \leq f$ ist

$$\mathcal{J}(h) \leq \sup_j \int f_j \; d\omega.$$

OBdA gilt $h(x) < f(x)$ für alle $x \in X$ mit $f(x) > 0$. Denn sonst können wir h ersetzen durch $(1 - \varepsilon)h$ für kleines $\varepsilon > 0$; dann wird $\mathcal{J}(h)$ ersetzt durch $(1 - \varepsilon)\mathcal{J}(h)$, was für unsere Abschätzungsaufgabe kein Schaden ist, weil wir dann noch $\varepsilon \to 0$ machen können. Sei h_j die grösste Elementarfunktion, die $h_j \leq \min\{h, f_j\}$ erfüllt und nur Werte annimmt, die auch schon von h angenommen werden; das heisst

$$h_j(x) = \max\{t \in \mathrm{im}(h) \mid t \leq \min\{f_j(x), h(x)\}\}$$

für $x \in X$. Wir können schreiben $\mathrm{im}(h) = \{a_0, \ldots, a_r\}$ mit $a_0 < a_1 < \cdots < a_r$. Sei nun $A_t := \{x \in X \mid h(x) \leq a_t\}$ für jedes $t \in \{0, 1, \ldots, r\}$, und ebenso $A_{j,t} := \{x \in X \mid h_j(x) \leq a_t\}$. Dann ist

$$\bigcup_j A_{j,t} = A_t$$

[1]Dabei soll $[0, \infty]$ so topologisiert werden, dass die Abbildung $x \mapsto x/(x+1)$ nach $[0, 1]$ ein Homöomorphismus ist. Oder anders und besser gesagt: $[0, \infty]$ ist die Einpunkt-Kompaktifizierung von $[0, \infty[$.

für jedes $t \in \{0,1,\dots,r\}$ nach Konstruktion (wegen $f = \sup_j f_j$ und $h(x) < f(x)$ wenn $f(x)$ positiv). Das ist eine aufsteigende Vereinigung, also $\lim_j \omega(A_{j,t}) = \omega(A_t)$ für jedes t in $\{0,1,\dots,r\}$. Daraus ergibt sich

$$\begin{aligned}\sup_j \int f_j \, d\omega \;\geq\; \sup_j \mathcal{J}(h_j) &= \sup_j \big(a_0 \cdot \omega(A_{j,0}) + \sum_{t=1}^{r}(a_t - a_{t-1})\omega(A_{j,t})\big) \\ &= a_0 \cdot \omega(A_0) + \sum_{t=1}^{r}(a_t - a_{t-1})\omega(A_t) \;=\; \mathcal{J}(h) \quad \square\end{aligned}$$

Definition 6.2.2. (Anwendung.) Sei $(X,\mathcal{E},\omega)$ ein Massraum und sei $g\colon X \to [0,\infty]$ messbar. Für messbares $T \subset X$ sei

$$\mu(T) = \int_T g \, d\omega := \int g \cdot \chi_T \, d\omega$$

wobei χ_T die Indikatorfunktion von T ist, also $\chi_T(x) = 1$ falls $x \in T$ und $\chi_T(x) = 0$ sonst. (Die Funktion $g \cdot \chi_T$ ist messbar, denn $(g \cdot \chi_T)^{-1}(V) = g^{-1}(V) \cap T$ für messbares $V \subset [0,\infty]$.) Aus dem Satz von B Levi ergibt sich sofort, dass die Funktion $\mu\colon \mathcal{E} \to [0,\infty]$ eine σ-additive Funktion ist; das heisst, μ ist wieder ein Mass auf dem messbaren Raum $(X,\mathcal{E})$. (Hier wird die minimalistische Definition von *Mass* benutzt, die wir schon in Abschnitt 3.1 gesehen haben, kurz nach Korollar 3.1.8.)

Man sagt, dass das Mass μ die *Dichte* g bezüglich ω hat. Eine wichtige Beobachtung dazu: jedes messbare $T \subset X$ mit der Eigenschaft $\omega(T) = 0$ erfüllt auch $\mu(T) = 0$. Dazu sagt man auch: *μ ist absolut stetig in Bezug auf ω.*

6.3. Zurück zu dynamischen Systemen

Sei $f\colon X \to X$ die Abbildung von Beispiel 4.3.10, also $X = [0,1]$ und $f(x) = x^{-1} - \lfloor x^{-1} \rfloor$ falls $x > 0$. Es wurde schon gezeigt, dass f messbar ist für die Lebesgue-σ-Algebra $\mathcal{A}$ auf X. Sei $g\colon X \to \mathbb{R}$ definiert durch

$$g(x) = (\ln(2))^{-1}\frac{1}{1+x}.$$

Der Faktor $(\ln(2))^{-1}$ soll hier nur sicherstellen, dass $\int_X g \, d\lambda = 1$. Sei γ das Mass auf X mit der Dichtefunktion g bezüglich Lebesgue-Mass λ, so dass

$$\gamma(T) := \int_T g(x) \, dx := \int \chi_T \cdot g \, d\lambda$$

für messbares $T \subset X$. Dann ist γ wieder ein Wahrscheinlichkeitsmass auf X, denn $\gamma(X) = \int g \, d\lambda = 1$. (Es wird γ genannt wegen Gauss. Die folgende Proposition ist eine Beobachtung von Gauss, und er wusste mehr darüber ...)

Proposition 6.3.1. *$f\colon X \to X$ ist masserhaltend für das Mass γ.*

Beweis. Wir sollen zeigen $\gamma(f^{-1}(T)) = \gamma(T)$ für messbares $T \subset X$, also $T \in \mathcal{A}$. Es ist einigermassen klar (aus der Definition von f), dass das richtig ist, wenn $\gamma(T) = 0$. Da zu jedem $T \in \mathcal{A}$ ein $S \in \mathcal{A}$ existiert mit $S \subset T$ und $\gamma(T \smallsetminus S) = 0$ und S Vereinigung einer aufsteigenden Folge von abgeschlossenen Teilmengen von X, haben wir auf den Fall reduziert, wo T selber Vereinigung einer aufsteigenden Folge von abgeschlossenen Mengen $T_0 \subset T_1 \subset T_2 \subset \dots$ ist. Dann ist auch $f^{-1}(T)$ die aufsteigende Vereinigung der Mengen $f^{-1}(T_i)$. Damit haben wir reduziert auf den Fall wo T abgeschlossen ist; oder, durch Uebergang zu den Komplementen, auf den Fall, wo T offen ist. Wenn T offen ist, dann ist es abzählbare disjunkte Vereinigung von Intervallen. Das heisst, wir haben auf den Fall reduziert, wo $T \subset X$ ein Intervall ist, und hier ist es ziemlich egal, ob offen, abgeschlossen oder halboffen. Sei also $T = [a, b] \subset [0, 1]$, wobei $0 \le a \le b \le 1$. Dann ist $f^{-1}(T)$ die disjunkte Vereinigung der Mengen

$$C_n :=](n+b)^{-1}, (n+a)^{-1}]$$

für $n = 1, 2, 3, \dots$, und noch $\{0, 1\}$, falls $0 \in T$. (Den Beitrag $\{0, 1\}$ können wir aber vernachlässigen, weil γ-Nullmenge.) Es ergibt sich

$$\gamma(C_n) = (\ln(2))^{-1} \int_{(n+b)^{-1}}^{(n+a)^{-1}} \frac{1}{1+x}\, dx = (\ln(2))^{-1}\big(\ln(1 + (n+a)^{-1}) - \ln(1 + (n+b)^{-1}))\big)$$
$$= (\ln(2))^{-1}\big(\ln(\frac{n+1+a}{n+a}) - \ln(\frac{n+1+b}{n+b}))\big) = (\ln(2))^{-1}\big(\ln(\frac{n+b}{n+a}) - \ln(\frac{n+1+b}{n+1+a}))\big).$$

Jetzt kann man leicht aufsummieren. Es kommt heraus, dass

$$\gamma(f^{-1}(T)) = \sum_{n \ge 1} \gamma(C_n) = (\ln(2))^{-1}\big(\ln(\frac{1+b}{1+a})\big) = \gamma(T). \quad \square$$

CHAPTER 7

Notizen zur 10., 11. und 12. Vorlesungswoche

7.1. Transformationsformel

Sei $f\colon U \to V$ eine stetig differenzierbare Abbildung zwischen offenen Mengen im $\mathbb{R}^n$, die eine stetig differenzierbare Inverse besitzt. Aus der Analysis II oder III ist wahrscheinlich bekannt:

$$\int_U \varphi(f(x)) \cdot |\det(Df(x)|\ dx_1 dx_2 ... dx_n = \int_V \varphi(y)\ dy_1 dy_2 \dots dy_n$$

wobei φ etwa eine stetige Funktion mit kompaktem Träger sein kann (damit das Riemann-Integral rechts definiert ist), und Df die totale Ableitung von f bezeichnet. Im Fall $n = 1$ (und wenn U, V Intervalle sind) ist das die Substitutionsregel. Mit dem Begriff *Dichtefunktion* aus Abschnitt ... können wir das umformulieren. Mit $\mathcal{A}_U$ bzw $\mathcal{A}_V$ soll die σ-Algebra der Lebesgue-messbaren Teilmengen von U bzw V bezeichnet werden, und mit $\lambda_U\colon \mathcal{A}_U \to [0,\infty]$ bzw $\lambda_V\colon \mathcal{A}_V \to [0,\infty]$ die entsprechenden Lebesgue-Masse.

Proposition 7.1.1. *Durch $\tau(T) := \lambda_V(f(T))$ für $T \in \mathcal{A}_U$ wird ein Mass τ auf $\mathcal{A}_U$ definiert. Dieses τ lässt sich auch beschreiben als das Mass mit der Dichtefunktion $|\det(Df)|$ bezüglich λ_U.*

Anmerkung: wer die (obige) Transformationsformel schon in grösserer Allgemeinheit kennt, zB für messbares φ, könnte/sollte den folgenden Beweis überflüssig finden. Denn man kann dann für φ in der Transformationsformel die Indikatorfunktion von $f(T)$ wählen, $T \in \mathcal{A}_U$. Dann ist $\varphi \circ f$ die Indikatorfunktion von T. Die rechte Seite der Transformatiomsformel ergibt also $\lambda_V(f(T)) = \tau(T)$, und die linke Seite ist genau $\tau_1(T)$, wobei τ_1 das Mass mit der Dichtefunktion $|\det(Df)|$ bezüglich λ_U bezeichnet.

Beweis. Sei $\mathcal{B}_U \subset \mathcal{A}_U$ die Borel-σ-Algebra. Die Abbildung $T \mapsto f(T)$ (wobei $T \in \mathcal{B}_U$) bestimmt einen Isomorphismus $\mathcal{B}_U \to \mathcal{B}_V$, und deswegen bestimmt $\tau(T) := \lambda_V(f(T))$ (für $T \in \mathcal{B}_U$) ein Mass auf $\mathcal{B}_U$. Aus unserer Definition des Integrals folgt sofort, dass

$$\int g\ d\lambda_V = \int g \circ f\ d\tau$$

für jede messbare (bezügl $\mathcal{B}_V$) Funktion $g \geq 0$ auf V. Andererseits wissen wir aus der Transformationsformel, dass $\int g\ d\lambda_V = \int (g \circ f) \cdot |\det(Df)|\ d\lambda_U$ gilt, falls $g\colon V \to \mathbb{R}$ stetig mit kompaktem Träger, so dass

$$\int g \circ f\ d\tau = \int (g \circ f) \cdot |\det(Df)|\ d\lambda_U$$

für solche speziellen g. Wir denken uns jetzt für $g\colon V \to \mathbb{R}$ die Indikatorfunktion von $f(U_0)$, wobei $U_0 \subset U$ offen ist; dann ist $f(U_0)$ offen in V, und wir können wir eine aufsteigende Folge $(g_j\colon V \to [0,1])_{j=1,2,\dots,n}$ von stetigen Funktionen mit kompaktem Träger wählen, so dass

$g = \sup_j g_j$. Es ergibt sich dann wieder

$$\int g \circ f \ d\tau = \sup_j \int g_j \circ f \ d\tau = \sup_j \int (g_j \circ f) \cdot |\det(Df)| \ d\lambda_U = \int (g \circ f) \cdot |\det(Df)| \ d\lambda_U$$

(zweifache Anwendung von B Levi), was sich vereinfacht zu

$$\tau(U_0) = \int_{U_0} |\det(Df)| \ d\lambda_U.$$

Da die offenen Teilmengen von U ganz $\mathcal{B}_U$ erzeugen, folgt, dass τ auf ganz $\mathcal{B}_U$ mit dem Mass übereinstimmt, das durch die Dichtefunktion $|\det(Df)|$ bezüglich λ_U gegeben ist.

Ausserdem: für $T \in \mathcal{A}_U$ gibt es $T_0, T_1 \in \mathcal{B}_U$ so dass $T_0 \subset T \subset T_1$ und $\lambda_U(T_1 \smallsetminus T_0) = 0$. Mit dem Vorhergehenden ergibt sich dann, dass $\lambda_U(f(T_1 \smallsetminus T_0)) = 0$, denn diese Zahl lässt sich als Integral über $T_1 \smallsetminus T_0$ ausdrücken. Dann hat auch $f(T) \smallsetminus f(T_0)$ das Lebesgue-Mass 0, und daher $f(T) \in \mathcal{A}_V$. Demnach können wir ohne Skrupel definieren: $\tau(T) := \lambda_V(f(T))$. Es gilt wieder

$$\tau(T) = \int_T |\det(Df)| \ d\lambda_U$$

weil beide Seiten mit $\tau(T_0) = \int_{T_0} |\det(Df)| \ d\lambda_U$ übereinstimmen. □

7.2. Wieder zurück zu dynamischen Systemen

Bezeichnungen wie in Abschnitt 6.3. Es soll endlich gezeigt werden:

SATZ 7.2.1. *Das dynamische System* $f\colon [0,1] \to [0,1]$*, wie dort definiert, ist ergodisch.*

Dabei ist das Intervall $[0,1]$ ausgestattet mit der σ-Algebra $\mathcal{A}$ der Lebesgue-messbaren Mengen und dem Mass γ von Gauss, so dass f masserhaltend ist für γ.

Ich stütze mich auf eine Bachelorarbeit, die ich im letzten Semester betreut habe (Autorin Nina Durner). Die Argumente stammen zwar nicht in erster Linie von N.D., sondern von einem gewissen O. Karpenkov, aber N.D. hat sie im Internet gefunden und diese Wahl getroffen, dazu Einzelheiten beigesteuert, die bei Karpenkov nicht ausgeführt waren, usw. Die Strategie ist wieder ein Zurückführen auf den Dichtesatz von Lebesgue. Allerdings muss erst eine ganze Menge Vorarbeit geleistet werden.

DEFINITION 7.2.2. Für positive ganze Zahlen $a_1, a_2, \ldots, a_n$ ist $\Delta_{(a_1,a_2,\ldots,a_n)} \subset [0,1]$ die Menge aller $x \in [0,1]$, deren Kettenbruchentwicklung die ersten n Teilnenner $a_1, a_2, ..., a_n$ hat. Die Mengen $\Delta_{(a_1,a_2,\ldots,a_n)}$ heissen *Zylindermengen.*

Was ist hier wichtig:

(i) Für positive ganze Zahl a und $x \in [0,1]$ sei $\psi_a(x) = (a+x)^{-1}$. Dann ist $\Delta_{(a_1,a_2,\ldots,a_n)}$ die Menge der Zahlen

$$\psi_{a_1}(\psi_{a_2}(\cdots(\psi_{a_n}(z)))...)$$

wobei $z \in [0,1[$. Statt dieser Formel kann man auch einen endlichen Kettenbruch mit den *Teilnennern* $a_1, a_2, ..., a_{n-1}, a_n + z$ hinschreiben (bitte selber machen, denn es ist drucktechnisch eine Herausforderung für mich) oder kürzer schreiben

$$[a_1, a_2, \ldots, a_n + z]$$

was ungefähr mit den Bezeichnungen von Bemerkung 4.3.11 übereinstimmen sollte. Weil ψ_a ein Diffeomorphismus aufs Bild ist (und monoton fallend), folgt aus dieser Beschreibung, dass $\Delta_{(a_1,a_2,\ldots,a_n)}$ ein halboffenes Intervall ist, begrenzt von den rationalen Zahlen $[a_1, a_2, \ldots, a_n]$ und $[a_1, a_2, \ldots, a_n + 1]$. Die Erstgenannte gehört zu

$\Delta_{(a_1,a_2,\ldots,a_n)}$, die Zweitgenannte nicht. Erstgenannte ist grösser als Zweitgenannte wenn n ungerade, dagegen kleiner wenn n gerade.

(ii) Das Maximum der Längen der Intervalle $\Delta_{(a_1,a_2,\ldots,a_n)}$ geht gegen 0 für $n \to \infty$. Dazu Genaueres später.

(iii) Die Einschränkung von f^n auf $\Delta_{(a_1,a_2,\ldots,a_n)}$ ist eine umkehrbar-glatte Abbildung von $\Delta_{(a_1,a_2,\ldots,a_n)}$ auf $[0,1[$. Die Umkehrung haben wir schon in (i) gesehen: es ist

$$\psi_{a_1} \circ \psi_{a_2} \circ \cdots \circ \psi_{a_n}$$

eingeschränkt auf $[0,1[$. Stattdessen kann man für diese Umkehrung auch schreiben $z \mapsto [a_1, a_2, \ldots, a_n + z]$.

Noch mehr zur Strategie. Sei $S \subset [0,1]$ eine invariante Menge für f, also $f^{-1}(S) = S$. Angenommen sei ausserdem $\lambda(S) > 0$ (oder $\gamma(S) > 0$, was dazu äquivalent ist). Dann müssen wir zeigen, dass $\lambda(S) = 1$ (oder $\gamma(S) = 1$, was dazu äquivalent ist). Denn wir wollen ja zeigen, dass f ergodisch ist. Wegen Dichtesatz von Lebesgue genügt es, wenn wir für jedes $\varepsilon > 0$ das Intervall $[0,1]$ durch viele Intervalle J_β mit Endpunkten in $\mathbb{Q}$ und von der Länge $\leq \varepsilon$ überdecken können derart, dass für die Zahlen $\lambda(S \cap J_\beta)/\lambda(J_\beta)$ eine positive untere Schranke angegeben werden kann, die unabhängig von ε und vom Index β ist. Dafür empfehlen sich die Intervalle $\Delta_{(a_1,a_2,\ldots,a_n)}$, also die *Zylindermengen*! — Ausserdem können wir folgende Argumentation benutzen. Zur Abkürzung sei $\Delta = \Delta_{(a_1,a_2,\ldots,a_n)}$ Zylindermenge mit fest gewählten positive ganzen Zahlen $a_1, a_2, ..., a_n$. Sei g die in (iv) beschriebene Abbildung $z \mapsto [a_1, a_2, \ldots, a_n + z]$ von $[0,1[$ nach $\Delta_{(a_1,a_2,\ldots,a_n)}$, ein Diffeomorphismus. Dann gilt wegen Transformationsformel, Proposition 7.1.1:

$$\lambda(S \cap \Delta) = \int_{S\cap\Delta} 1 \; d\lambda = \int_{g^{-1}(S)} |g'(x)| \; dx \, .$$

Hier ist aber $g^{-1}(S) = g^{-1}(f^{-n}(S)) = (f^n g)^{-1}(S) = S \setminus \{1\}$, denn $f^n g = \mathrm{id}$ auf $[0,1[$. Also erhalten wir

$$\lambda(S \cap \Delta) = \int_S |g'(x)| \; dx \;\; \geq \;\; \lambda(S) \cdot \min_{x\in[0,1]} \{ \, |g'(x)| \, \}$$

und damit

$$\frac{\lambda(S \cap \Delta)}{\lambda(\Delta)} \geq \frac{\min_{x\in[0,1]}\{ \, |g'(x)\,| \}}{\lambda(\Delta)}.$$

Also genügt es, für die rechte Seite dieser letzten Ungleichung eine positive untere Schranke zu finden, unabhängig von $a_1, ..., a_n$.

Aufgrund dieser Sachlage müssen wir mit den Zylindermengen ein paar Rechnungen anstellen. Gegeben sei dafür irgeneine Folge von positiven ganzen Zahlen $a_1, a_2, \ldots, a_n, \ldots$; man darf sich dabei vorstellen, dass es die Folge der Teilnenner in der Kettenbruchentwiclung einer irrationalen Zahl in $[0,1]$ ist (aber das ist nicht wichtig). Wir interessieren uns für die rationalen Zahlen

$$[a_1, a_2, \ldots, a_n] = \frac{p_n}{q_n}$$

wobei p_n, q_n teilerfremd, positiv, und $n \geq 1$ beliebig. Dann gilt:

(iv) $q_n p_{n-1} - p_n q_{n-1} = (-1)^n$ falls $n > 1$.

(v) $p_n = p_{n-2} + a_n p_{n-1}$ und $q_n = q_{n-2} + a_n q_{n-1}$ falls $n > 2$.

(vi) Für den Ausdruck $[a_1, a_2, \ldots, a_n + z]$ in (iii) kann man schreiben

$$\frac{p_n + p_{n-1} z}{q_n + q_{n-1} z}.$$

Beweis von (iv),(v) *und* (vi). Angenommen (vi) gilt mit $n = k$; wir setzen $z = a_{k+1}^{-1}$ und erhalten

$$\frac{p_{k+1}}{q_{k+1}} = [a_1, a_2, \ldots, a_k, a_{k+1}] = [a_1, a_2, \ldots, a_k + z] \stackrel{\text{(vi)}}{=} \frac{p_k + p_{k-1}z}{q_k + q_{k-1}z} = \frac{a_{k+1}p_k + p_{k-1}}{a_{k+1}q_k + q_{k-1}}.$$

Um zu sehen, dass Zähler und Nenner im letzten Ausdruck teilerfremd sind, multipliziere diese mit q_k bzw p_k und bilde Differenz. Dann benutze (iv) mit $n = k$. Wir sehen also: (vi) und (iv) für $n = k$ implizieren zusammen (v) für $n = k + 1$. Ausserdem: (v) für $n = k + 1$ zusammen mit (iv) für $n = k$ impliziert (iv) für $n = k + 1$:

$$q_{k+1}p_k - p_{k+1}q_k \stackrel{\text{(v)}}{=} (q_{k-1} + a_{k+1}q_k)p_k - (p_{k-1} + a_{k+1}p_k)q_k = q_{k-1}p_k - p_{k-1}q_k \stackrel{\text{(iv)}}{=} (-1)^{k+1}.$$

Ausserdem: (vi) für $n = k$ zusammen mit (v) für $n = k + 1$ implizieren (vi) für $n = k + 1$:

$$\frac{p_{k+1} + p_k z}{q_{k+1} + q_k z} \stackrel{\text{(v)}}{=} \frac{p_{k-1} + a_{k+1}p_k + p_k z}{q_{k-1} + a_{k+1}q_k + q_k z} = \frac{p_k + p_{k-1}(\frac{1}{a_{k+1}+z})}{q_k + q_{k-1}(\frac{1}{a_{k+1}+z})}$$

$$\stackrel{\text{(vi)}}{=} [a_1, a_2, \ldots, a_k + \frac{1}{a_{k+1} + z}] = [a_1, a_2, \ldots, a_k, a_{k+1} + z].$$

Damit sind alle nötigen Induktionsschritte gemacht. Es fehlt noch der Induktionsanfang, der darin bestehen sollte, dass wir (iv) und (vi) für $n = 2$ beweisen. Ich lasse das aber aus. □

Beweis von (ii). Wegen (vi) und mit den Bezeichnungen von (vi) ist die Länge das Intervalls $\Delta_{(a_1,a_2,\ldots,a_n)}$ gleich dem Betrag von

$$\frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}} = \frac{p_n(q_n + q_{n-1}) - q_n(p_n + p_{n-1})}{q_n(q_n + q_{n-1})} = \frac{p_n q_{n-1} - q_n p_{n-1}}{q_n(q_n + q_{n-1})}$$

und damit haben wir sie berechnet zu $(q_n(q_n + q_{n-1}))^{-1}$, unter Benutzung von (iv). Also genügt es zu zeigen, dass $q_n \to \infty$ für $n \to \infty$; aber das ist klar wegen (v). □

Das folgende Lemma stellt die Verbindung mit dem Dichtesatz von Lebesgue her. Beachten, dass die Formulierung das Lebesgue-Mass λ benutzt, nicht γ. Das mag unsportlich sein, aber die Rechnungen werden gerade deswegen einfacher.

LEMMA 7.2.3. *Sei $\Delta_{(a_1,a_2,\ldots,a_n)}$ eine Zylindermenge und sei $S \subset [0, 1]$ messbar und invariant unter f, also $f^{-1}(S) = S$. Dann ist $\lambda(S \cap \Delta_{(a_1,a_2,\ldots,a_n)}) = \int_S (q_n + q_{n-1}x)^{-2}\ dx$ und daher*

$$\lambda(S \cap \Delta_{(a_1,a_2,\ldots,a_n)}) \geq \tfrac{1}{2}\lambda(S)\lambda(\Delta_{(a_1,a_2,\ldots,a_n)}).$$

Beweis. Zur Abkürzung sei $\Delta := \Delta_{(a_1,a_2,\ldots,a_n)}$ in diesem Beweis. Für $x \in [0, 1[$ sei

$$g(x) = [a_1, a_2, \ldots, a_n + x]\,.$$

Wie schon im Absatz *Noch mehr zur Strategie* erklärt, ist $\lambda(S \cap \Delta) = \int_S |g'(x)|\ dx$. Ausserdem haben wir die Formel $g(x) = (p_n + p_{n-1}x)/(q_n + q_{n-1}x)$ aus (vi) und damit nach der Quotientenregel

$$g'(x) = \frac{p_{n-1}(q_n + q_{n-1}x) - q_{n-1}(p_n + p_{n-1}x)}{(q_n + q_{n-1}x)^2} = \frac{(-1)^n}{(q_n + q_{n-1}x)^2}$$

unter Benutzung von (iv). Deswegen wird

$$\lambda(S \cap \Delta) = \int_S |g'(x)|\ dx = \int_S (q_n + q_{n-1}x)^{-2}\ dx \geq \int_S (q_n + q_{n-1})^{-2}\ dx = \lambda(S)(q_n + q_{n-1})^{-2}$$

und das ist mindestens so gross wie $\lambda(S)(2q_n)^{-1}(q_n + q_{n-1})^{-1} = \frac{1}{2}\lambda(S)\lambda(\Delta_{(a_1,a_2,\ldots,a_n)})$. Die letzte Gleichung kommt aus dem Beweis von (ii). □

Beweis von Satz 7.2.1. Sei $S \subset [0,1]$ messbar und invariant unter f, also $f^{-1}(S) = S$. Wir können annehmen $\lambda(S) > 0$ (gleichbedeutend mit $\gamma(S) > 0$) und wir müssen dann zeigen $\lambda(S) = 1$ (gleichbedeutend mit $\gamma(S) = 1$). Sei $z \in [0,1] \setminus \mathbb{Q}$. In Kettenbruchschreibweise, Bemerkung 4.3.11, ist $z = [a_1, a_2, a_3, ...]$ für eine (unendliche) Folge von positiven ganzen Zahlen $a_1, a_2, a_3, \dots$. Für jedes $n \geq 1$ liegt z im Inneren vom halboffenen Intervall $\Delta_{(a_1,\dots,a_n)}$. Die Länge dieser Intervalle geht gegen 0 für $n \to \infty$. Aus Lemma 7.2.3 und der Annahme $\lambda(S) > 0$ folgt also, dass

$$\lim_{\varepsilon_1,\varepsilon_2 \to 0} \frac{\lambda(S \cap [z-\varepsilon_1, z+\varepsilon_2])}{\varepsilon_1 + \varepsilon_2}$$

entweder nicht existiert oder $\geq \lambda(S)/2$ ist (für fast alle $z \in [0,1]$, nämlich die irrationalen). Andererseits hat für fast alle z aus $E := [0,1] \setminus S$ dieser lim den Wert 0, denn

$$\lim_{\varepsilon_1,\varepsilon_2 \to 0} \frac{\lambda(E \cap [z-\varepsilon_1, z+\varepsilon_2])}{\varepsilon_1 + \varepsilon_2} = 1$$

für fast alle $z \in E$ (Lebesgues Dichtesatz 5.1.1). Daraus folgt, dass $\lambda(E) = 0$, also $\lambda(S) = 1$. $\square$

CHAPTER 8

Notizen zur 13. Vorlesungswoche

8.1. Der Satz von Radon-Nikodym

LEMMA 8.1.1. (Nach H.Hahn.) *Sei X eine Menge ausgerüstet mit σ-Algebra $\mathcal{A}$, und ρ, τ zwei endliche Masse auf $\mathcal{A}$. Wenn $\rho(X) \geq \tau(X) > 0$ gilt, dann existiert messbares $T \subset X$ derart, dass $\rho(T) > 0$ und $\rho(U) \geq \tau(U)$ für alle messbaren $U \subset T$.*

Beweis. Sei $S_0 := X$ und

$$t_1 = \inf\{\rho(U) - \tau(U) \mid U \text{ messbar, } U \subset S_0\}.$$

Wenn $t_1 \geq 0$, dann sind wir fertig und setzen $T := S_0$. Sonst wählen wir ein messbares $U_1 \subset S_0$ mit $\rho(U_1) - \tau(U_1) < t_1/2$. Setze $S_1 = S_0 \smallsetminus U_1$. Sei

$$t_2 = \inf\{\rho(U) - \tau(U) \mid U \text{ messbar, } U \subset S_1\}.$$

Wenn $t_2 \geq 0$, dann sind wir fertig und setzen $T := S_1$. Sonst wählen wir ein messbares $U_2 \subset S_1$ mit $\rho(U_2) - \tau(U_2) < 3t_2/4$. Setze $S_2 = S_1 \smallsetminus U_2$. Sei

$$t_3 = \inf\{\rho(U) - \tau(U) \mid U \text{ messbar, } U \subset S_2\}.$$

Wenn $t_3 \geq 0$, sind wir fertig und setzen $T =: S_2$. Sonst wählen wir ein messbares $U_3 \subset S_2$ mit $\rho(U_3) - \tau(U_3) < 7t_3/8$. Undsoweiter. Wenn der Prozess abbricht, sind wir fertig. Sonst bemerken wir (für später), dass $t_1 \leq t_2 \leq t_3 \leq ... \leq 0$ und wir setzen

$$T := \bigcap_{j=0}^{\infty} S_j = S_0 \smallsetminus \bigcup_{j=1}^{\infty} U_j$$

(wobei die U_j paarweise disjunkt sind). Behauptung: T hat die gewünschten Eigenschaften. Denn erstens ist $\rho(T) = \rho(S_0) - \sum_{i=1}^{\infty} \rho(U_i)$ und $\tau(T) = \tau(S_0) - \sum_{i=1}^{\infty} \tau(U_i)$, so dass

$$\rho(T) - \tau(T) = \rho(S_0) - \tau(S_0) - \sum_{i=1}^{\infty}(\rho(U_i) - \tau(U_i)) > \rho(S_0) - \tau(S_0) \geq 0,$$

und damit $\rho(T) > 0$. Zweitens gibt es kein messbares $V \subset T$ mit $\rho(V) < \tau(V)$, denn sonst wäre $\rho(V) - \tau(V) < t_j/2^j$ für genügend grosses j, und damit

$$\rho(V \cup U_j) - \tau(V \cup U_j) = \rho(V) - \tau(V) + \rho(U_j) - \tau(U_j) < t_j/2^j + (2^j - 1)t_j/2^j = t_j$$

im Widerspruch zur Wahl von t_j. $\square$

SATZ 8.1.2. (Radon-Nikodym). *Gegeben sei ein Massraum $(X, \mathcal{A})$ und zwei endliche Masse ρ, μ auf $\mathcal{A}$. Angenommen, für jedes $S \in \mathcal{A}$ mit $\mu(S) = 0$ gilt auch $\rho(S) = 0$. Dann gibt es eine messbare Funktion $f \colon X \to [0, \infty]$ derart, dass $f \geq 0$ und*

$$\rho(T) = \int_T f \; d\mu$$

für alle $T \in \mathcal{A}$.

(Die Voraussetzung wird auch so augedrückt: ρ ist absolut stetig in Bezug auf μ. Die Schlussfolgerung wird auch so ausgedrückt, dass ρ ein Mass ist, das durch eine Dichtefunktion bezüglich μ beschrieben kann.)

Beweis. (Hält sich eng an Brokate-Kersting.) Wir beschäftigen uns zuerst mit messbaren Funktionen $h\colon X \to [0,\infty]$, die $\int_T h \; d\mu \leq \rho(T)$ erfüllen für alle messbaren T. So ein h soll *zulässig* heissen. Für zulässiges h ist

$$\int h \; d\mu = \int_X h \; d\mu \leq \rho(X) < \infty$$

nach Definition. Sei also

$$\beta = \sup\{\int h \; d\mu \mid h \text{ zulässig}\}$$

so dass $0 \leq \beta \leq \rho(X)$. Wir machen jetzt folgende Schritte:

(i) Die Menge der zulässigen Funktionen ist abgeschlossen unter der Operation max, das heisst, das (punktweise) Maximum von zwei zulässigen Funktionen ist wieder zulässig.
(ii) Es gibt ein zulässiges h mit $\int h \; d\mu = \beta$.
(iii) Es muss gelten $\beta = \rho(X)$.

Beweis von (i): gegeben zulässige h_1 und h_2. Sei $h_3 = \max(h_1, h_2)$; das ist wieder eine messbare Funktion. Sei $B = \{x \in X \mid h_1(x) \leq h_2(x)\}$. Dann ist B messbar, und $h_3 = h_2$ auf B, wohingegen $h_3 = h_1$ auf $X \smallsetminus B$. Für $T \in \mathcal{A}$ folgt dann

$$\int_T h_3 \; d\mu = \int_{T\cap B} h_3 \; d\mu + \int_{T\smallsetminus B} h_3 \; d\mu = \int_{T\cap B} h_2 \; d\mu + \int_{T\smallsetminus B} h_1 \; d\mu \leq \rho(T\cap B) + \rho(T \smallsetminus B) = \rho(T)$$

was zu beweisen war. Beweis von (ii): wir wählen eine Folge $h_1, h_2, h_3, \ldots$ von zulässigen Funktionen derart, dass

$$\int h_j \; d\mu > \beta - 2^{-j}.$$

OBdA ist die Folge aufsteigend, denn sonst können wir h_j ersetzen durch das punktweise Maximum von $h_1, h_2, \ldots, h_j$. Wenn die Folge in diesem Sinn aufsteigend ist, dann ist das Supremum h_∞ der h_j wieder eine messbare Funktion. Aus dem Satz über die monotone Konvergenz folgt, dass

$$\int_T h_\infty \; d\mu = \sup_j \int_T h_j \; d\mu,$$

und die rechte Seite ist $\leq \rho(T)$, so dass h_∞ wieder zulässig ist. Für $T = X$ erhalten wir sogar $\int_X h_\infty \; d\mu = \int h_\infty \; d\mu = \beta$, so dass $h := h_\infty$ eine Lösung ist. — Beweis von (iii): Sei zulässiges h gewählt wie in (ii), also $\int h \; d\mu = \beta$. Angenommen, $\beta < \rho(X)$. Dann können wir auch ein $\varepsilon > 0$ finden mit $\beta + \varepsilon \cdot \rho(X) \leq \rho(X)$. Sei nun τ das Mass auf X, das durch

$$\tau(T) = \int_T h + \varepsilon \; d\mu$$

definiert ist. Nach Konstruktion ist $\rho(X) \geq \tau(X)$. Wegen Lemma ... existiert ein messbares $T \subset X$ mit $\rho(U) \geq \tau(U)$ für alle messbaren $U \subset T$, und $\rho(T) > 0$. Nach Voraussetzung muss dann auch $\mu(T) > 0$ gelten. Es folgt dann, dass die Funktion

$$h + \varepsilon \cdot \chi_T$$

wieder zulässig ist (χ_T ist wie üblich die Indikatorfunktion von T). Ihr μ-Integral ist aber

$$\beta + \varepsilon \cdot \mu(T),$$

und das ist grösser als β, weil $\mu(T) > 0$. Widerspruch, womit (iii) bewiesen ist.

Zusammenfassend: es gibt ein messbares $f \geq 0$, $f\colon X \to \mathbb{R}$ (vorher auch h oder h_∞ genannt), so dass $\int_T f \; d\mu \leq \rho(T)$ für alle messbaren T gilt und $\int_X f \; d\mu = \rho(X)$. Daraus folgt aber leicht, dass $\int_T f\ d\mu = \rho(T)$ gelten muss für alle messbaren T, denn wenn das falsch wäre für ein gewisses $T = T_0$, dann hätte man

$$\int_X f\ d\mu = \int_{T_0} f\ d\mu + \int_{X-T_0} f\ d\mu < \rho(T_0) + \rho(X \smallsetminus T_0) = \rho(X)$$

im Widerspruch zu $\int_X h\ d\mu = \rho(X)$. □

Bemerkung. Das f im Satz von Radon-Nikodym ist fast eindeutig; das heisst, wenn f_1 und f_2 beide die im Satz genannten Eigenschaften haben, dann ist die Menge

$$\{x \in X \mid f_1(x) \neq f_2(x)\}$$

eine μ-Nullmenge (und damit auch eine ρ-Nullmenge). *Beweis.* Für $n = 1, 2, 3, \ldots$ sei $T_n = \{x \in X \mid f_1(x) \geq f_2(x) + 1/n\}$. Dann gilt einerseits

$$\int_{T_n} f_1\ d\mu = \rho(T_n) = \int_{T_n} f_2\ d\mu$$

und andererseits $\int_{T_n} f_1\, d\mu - \int_{T_n} f_2\, d\mu \geq \int_{T_n} 1/n\, d\mu = \mu(T_n)/n$. Also ist $\mu(T_n) = 0$. Daraus folgt $\mu(\bigcup_{n=1}^{\infty} T_n) = 0$, und dass heisst, die Menge $\{x \in X \mid f_1(x) > f_2(x)\}$ ist eine μ-Nullmenge. □

Bemerkung. Wir haben zugelassen, dass f im Satz von Radon-Nikodym den Wert ∞ annimmt, weil es im Beweis praktisch war. Aber: die messbare Menge $T := f^{-1}(\infty)$ ist eine μ-Nullmenge, denn sonst wäre $\rho(T) = \int_T f\ d\mu \geq \int_T c\ d\mu = c \cdot \mu(T)$ für jede positive reelle Zahl c, was nicht stimmt. Denn es wurde $\rho(T) < \infty$ vorausgesetzt.

8.2. Bedingte Erwartungen

Sei $(X, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $f\colon X \to \mathbb{R}$ messbar bezüglich $\mathcal{A}$ (und Borel-Algebra von $\mathbb{R}$). Zur Vereinfachung nehmen wir ausserdem an, dass $\int |f|\ d\mu < \infty$. Die Zahl $\int_X f\ d\mu$ nennt man dann den *erwarteten Wert* von f, und schreibt dafür $E(f)$. Etwas allgemeiner: sei $Y \subset X$ eine messbare Teilmenge, $\mu(Y) > 0$. Dann heisst die Zahl

$$\frac{\int_Y f\ d\mu}{\mu(Y)} =: E(f|Y)$$

die *bedingte Erwartung* von f, gegeben Y. Ich hoffe, dass diese Wortwahl selbsterklärend ist.

Aus dem Satz von Radon-Nikodym ergibt sich eine Verallgemeinerung von diesen Begriffen. Sei nämlich $\mathcal{C}$ eine σ-Unteralgebra von $\mathcal{A}$. Von f (wie oben) nehmen wir weiterhin an, dass $\int |f|\ d\mu < \infty$, aber jetzt auch $f \geq 0$. Dann können wir ein Mass ρ auf dem Massraum $(X, \mathcal{C})$ definieren durch

$$\rho(T) = \int_T f\ d\mu$$

für $T \in \mathcal{C}$. Wenn $\mu(T) = 0$, dann ist auch $\rho(T) = 0$; also ist ρ absolut stetig in Bezug auf $\mu|_{\mathcal{C}}$. Nach Radon-Nikodym existiert deswegen eine Funktion $g\colon X \to [0, \infty]$, die messbar ist bezüglich $\mathcal{C}$, derart dass

$$\rho(T) = \int_T g\ d(\mu|_{\mathcal{C}})$$

für jedes $T \in \mathcal{C}$. Ausserdem folgt aus den Definitionen, dass $\int_T g\ d(\mu|_{\mathcal{C}}) = \int_T g\ d\mu$. Etwas kürzer gesagt: das bestimmende Gleichungssystem für g ist

$$\forall T \in \mathcal{C}: \int_T f\ d\mu = \int_T g\ d\mu\,.$$

(Beachten, dass g messbar sein soll bezüglich $\mathcal{C}$, während das von f nicht verlangt wurde.) Man schreibt dann

$$g = E(f|\mathcal{C})$$

und nennt g die bedingte Erwartung von f, gegeben $\mathcal{C}$. Etwas allgemeiner: wenn $\int |f|\ d\mu < \infty$ ist, dann können wir

$$E(f|\mathcal{C}) := E(f_+|\mathcal{C}) - E(f_-|\mathcal{C})$$

definieren, ohne die Voraussetzung $f \geq 0$ zu machen. Es wird dann klar, dass die bedingte Erwartung eine Linearitätseigenschaft hat, also $E(af_1 + bf_2\,|\mathcal{C}) = aE(f_1|\mathcal{C}) + bE(f_2|\mathcal{C})$.

Beispiel 8.2.1. Sei $(X, \mathcal{A}, \mu)$ Wahrscheinlichkeitsraum wie vorher und sei $Y \subset X$, wobei $Y \in \mathcal{A}$ und $\mu(Y) > 0$. Sei $\mathcal{C}$ die σ-Unteralgebra $\mathcal{C}$ von $\mathcal{A}$, die nur aus den vier Mengen $\emptyset, Y, X \smallsetminus Y$ und X besteht. Dann muss die Funktion $g = E(f|\mathcal{C})$ jeweils konstant sein auf Y und $X \smallsetminus Y$, damit sie $\mathcal{C}$-messbar ist. Da $\int_Y f\ d\mu = \int_Y g\ d\mu$ sein soll, ist der konstante Wert auf Y gleich $\int_Y f\ d\mu$ geteilt durch $\mu(Y)$, also gleich der "klassischen" bedingten Erwartung $E(f|Y)$. Ebenso ist der konstante Wert von g auf $X \smallsetminus Y$ gleich der bedingten Erwartung $E(f|X \smallsetminus Y)$.

Lemma 8.2.2. *Sei f messbar. Die bedingte Erwartung von $|f|$, gegeben $\mathcal{C}$, ist $|E(f|\mathcal{C})|$.*

Beweis. Man schreibt $f = f_+ - f_-$ und $|f| = f_+ + f_-$ und benutzt die Linearitätseigenschaft der bedingten Erwartung. □

Korollar 8.2.3. *Wenn $\int |f|\ d\mu < \infty$, dann auch $\int |E(f|\mathcal{C})|\ d\mu < \infty$.*

Beweis. Folgt aus dem Lemma und der Regel $\int g\ d\mu = \int E(g|\mathcal{C})\ d\mu$, mit $g = |f|$. □

CHAPTER 9

Notizen zur 14. Vorlesungswoche

9.1. Satz von Birkhoff-Chintschin

SATZ 9.1.1. *Sei $(X, \mathcal{A}, \mu)$ Wahrscheinlichkeitsraum und sei $w\colon X \to X$ eine messbare und masserhaltende Abbildung. Sei $\mathcal{A}^w \subset \mathcal{A}$ die Unter-σ-Algebra bestehend aus den $Y \in \mathcal{A}$, die $w^{-1}(Y) = Y$ erfüllen. Für jedes $\mathcal{A}$-messbare $f\colon X \to \mathbb{R}$, das die Bedingung $\int |f|\ d\mu < \infty$ erfüllt, konvergiert das "zeitliche Mittel"*

$$\frac{1}{n}\sum_{j=0}^{n-1} f \circ w^j$$

für $n \to \infty$ ausserhalb einer μ-Nullmenge punktweise gegen die bedingte Erwartung $E(f|\mathcal{A}^w)$.

BEISPIEL 9.1.2. Wenn wir im Satz noch zusätzlich annehmen, dass w ergodisch ist, dann besteht $\mathcal{A}^w \subset \mathcal{A}$ nur aus Mengen, die das μ-Mass 0 oder 1 haben. Dann kann $E(f|\mathcal{A}^w)$ dargestellt werden durch eine konstante Funktion mit dem konstanten Wert $\int f\ d\mu$. Also gilt für fast alle $x \in X$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ w^j(x) \ = \ \int f\ d\mu\,.$$

Das wird auch oft so ausgedrückt (nach L Boltzmann): *im ergodischen Fall stimmt das zeitliche Mittel von f mit dem räumlichen Mittel überein* (für jedes messbare f mit $\int |f|\ d\mu < \infty$). Dabei soll man sich klar machen, dass das räumliche Mittel $\int f\ d\mu$ eine reelle Zahl oder Konstante ist, wohingegen das zeitliche Mittel eine Funktion von $x \in X$ ist, nämlich eben

$$Z_f(x) := \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ w^j(x).$$

Dann versteht man auch, dass die Boltzmann-Bedingung gleichwertig ist zur Bedingung *w ist ergodisch*. Die eine Richtung davon haven wir gerade schon nachgewiesen. Für die andere Richtung: sei $K \in \mathcal{A}$ eine invariante Menge, $w^{-1}(K) = K$. Sei $f = \chi_K$ die Indikatorfunktion von K. Dann ist auch $Z_f = \chi_K$. Andererseits soll Z_f ja fast konstant sein, konstanter Wert gleich dem räumlichen Mittel von f. Wenn χ_K fast konstant ist, dann ist $\mu(K) \in \{0, 1\}$.

Den Beweis des Satzes habe ich bis auf kleine Änderungen von Einsiedler-Schmidt (S.52) übernommen; diese beiden haben ihn allerdings übernommen aus einem sehr langen Buch von A Katok und B Hasselblatt[1]. Dieser Beweis ist hat sich wahrscheinlich sehr lange entwickelt (1930-1995). In den älteren Formulierungen wird $E(f|\mathcal{A}^w)$ nicht erwähnt; die Aussage ist dann nur, dass

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f \circ w^j(x)$$

[1] *Introduction to the modern theory of dynamical systems*, Cambridge University Press, 1995.

für fast alle $x \in X$ existiert. Im Beweis ist es ein grosser Vorteil, den Kandidaten $E(f|\mathcal{A}^w)$ für die Grenz(wert)funktion zu haben.

Zur Vorbereitung gibt es ein paar Lemmas. Von diesen ist Lemma 9.1.5 auch für sich genommen schon sehr interessant. Es ist auch der Hauptbeitrag zum Beweis von Satz 9.1.1.

LEMMA 9.1.3. *Für jede reellwertige messbare Funktion $h \geq 0$ auf X gilt $\int h\ d\mu = \int h \circ w\ d\mu$.*

Beweis. Das kann man auf den Fall einer Indikatorfunktion $h = \chi_T$ zurückführen. Dann ist $h \circ w$ die Indikatorfunktion von $w^{-1}(T)$, und der Fall ist klar, weil w masserhaltend ist. □

LEMMA 9.1.4. *Für jede $\mathcal{A}^w$-messbare reellwertige Funktion h gilt $h = h \circ w$.*

Beweis. Das lässt sich leicht auf den Fall zurückführen, wo h die Indikatorfunktion von einem $T \in \mathcal{A}^w$ ist, und dann ist die Sache klar, denn $w^{-1}(T) = T$. □

LEMMA 9.1.5. *Sei $(X, \mathcal{A}, \mu)$ Wahrscheinlichkeitsraum und $w\colon X \to X$ masserhaltend. Sei $h\colon X \to \mathbb{R}$ messbar, mit $\int |h|\ d\mu < \infty$. Sei*

$$B = \{x \in X \mid \sup_n \textstyle\sum_{j=0}^{n-1} h(w^j(x)) = \infty\}.$$

Dann ist $\int_B h\ d\mu \geq 0$.

Beweis. Zur Abkürzung sei $S_k h := \sum_{j=0}^{k-1} h \circ w^j$. Im Fall $k = 0$ soll das gelesen werden als $S_0 h \equiv 0$. Ausserdem sei, für positives n,

$$M_n h := \max_{0 \leq k \leq n} S_k h.$$

Man überlegt sich $S_{k+1} h = h + (S_k h \circ w)$, und auf ähnliche Weise $M_{n+1} h = \max\{0, h + M_n h \circ w\}$. Aus dieser letzten Gleichung folgt

$$(*) \qquad M_{n+1} h - M_n h \circ w = \max\{-M_n h \circ w,\ h\}.$$

Die rechte Seite von $(*)$ stellt eine fallende Folge von (messbaren) Funktionen dar, die durch h von unten beschränkt ist, denn die Folge der $M_n h$ ist wachsend. Es folgt dann leicht aus dem Satz von Beppo Levi, dass die Funktion

$$D_\infty h := \lim_{n \to \infty} M_{n+1} h - M_n h \circ w$$

messbar/integrierbar ist und dass

$$\infty \ > \ \int D_\infty h\ d\mu = \lim_{n \to \infty} \int M_{n+1} h - M_n h \circ w\ d\mu \geq \int h\ d\mu \ > \ -\infty$$

ist. Die Funktion $D_\infty h$ hat ausserdem folgende nützliche Eigenschaft: für jedes $C \in \mathcal{A}^w$ gilt

$$(**) \qquad \int_C D_\infty h\ d\mu \geq 0.$$

Denn erstens ist

$$\int_C D_\infty h\ d\mu = \lim_n \int_C M_{n+1} h - M_n h \circ w\ d\mu = \lim_n \int_C M_{n+1} h\ d\mu - \int_C M_n h \circ w\ d\mu.$$

Zweitens kann im letzten Integral auf der rechten Seite die Funktion $M_n h \circ w$ durch $M_n h$ ersetzt werden wegen Lemma 9.1.3, angewandt auf $M_n h \cdot \chi_C$. Dann erhalten wir

$$\cdots = \lim_n \int_C M_{n+1} h\ d\mu - \int_C M_n h\ d\mu = \lim_n \int_C M_{n+1} h - M_n h\ d\mu \geq 0,$$

also $(**)$. Jetzt können wir $C := B$ nehmen, denn B (wie oben definiert) gehört zu $\mathcal{A}^w$, wie man leicht sieht. Also ist $\int_B D_\infty h\ d\mu \geq 0$. Andererseits ist auch $D_\infty h \equiv h$ auf B. Das folgt aus $(*)$, denn es ist klar, dass $\lim_{n\to\infty}(M_n h)(x) = \infty$ für alle $x \in B$, wegen Definition von B. Also erhalten wir $\int_B h\ d\mu \geq 0$. □

Beweis von Satz 9.1.1. Wir wählen ein $\varepsilon > 0$ und setzen $g := f - E(f|\mathcal{A}^w) - \varepsilon$. Dann ist g wieder messbar und integrierbar, wegen Lemma 8.2.2. Wir wollen zuerst zeigen , dass

$$\limsup_{n\to\infty} \frac{1}{n} S_n g$$

fast sicher ≤ 0 ist. Für diejenigen $x \in X$, die $\limsup_{n\to\infty} \frac{1}{n} S_n g(x) > 0$ erfüllen, muss auch gelten $\limsup_{n\to\infty} S_n g(x) = \infty$. Deswegen genügt es, zu zeigen, dass die Menge

$$B = \{x \in X \mid \limsup_{n\to\infty} S_n g(x) = \infty\}$$

eine μ-Nullmenge ist.

Aus Lemma 9.1.5 (mit g anstelle von h) folgt, dass $\int_B g \; d\mu \geq 0$. Ausserdem folgt aus $B \in \mathcal{A}^w$, dass $\int_B f \; d\mu = \int_B E(f|\mathcal{A}^w) \; d\mu$, wegen Definition der bedingten Erwartung, so dass $\int_B g \; d\mu = -\varepsilon \cdot \mu(B)$ wegen Definition von g. Also muss tatsächlich gelten $\mu(B) = 0$.

Wenn wir jetzt g_0 für $f - E(f|\mathcal{A}^w)$ schreiben, dann ist $g = g_0 - \varepsilon$ und wir dürfen zusammenfassen: bei fest gewähltem ε gilt

$$\limsup_{n\to\infty} \frac{1}{n}(S_n g_0)(x) \leq \varepsilon$$

für fast alle $x \in X$. Daraus folgt sofort $\limsup_{n\to\infty} \frac{1}{n}(S_n g_0)(x) \leq 0$ für fast alle $x \in X$. Wenn wir jetzt auch noch im ganzen Argument f durch $-f$ ersetzen, was erlaubt ist, dann erhalten wir ebenso $\liminf_{n\to\infty} \frac{1}{n}(S_n g_0)(x) \geq 0$ für fast alle $x \in X$. Damit ist bewiesen:

$$\lim_{n\to\infty} \frac{1}{n}(S_n g_0)(x) = 0$$

für fast alle $x \in X$. Andererseits können wir schreiben

$$\frac{1}{n} S_n g_0 = \frac{1}{n}(S_n f - S_n E(f|\mathcal{A}^w)) = \frac{1}{n} S_n f - E(f|\mathcal{A}^w)$$

wegen Lemma 9.1.4, angewandt mit $h = E(f|\mathcal{A}^w)$. Damit ist bewiesen

$$\lim_{n\to\infty} \frac{1}{n}(S_n f)(x) = E(f|\mathcal{A}^w)(x)$$

für fast alle $x \in X$, was zu zeigen war. □

Beispiel 9.1.6. Wir nehmen an, dass w ergodisch ist, und wählen für f eine Indikatorfunktion χ_K, wobei $K \in \mathcal{A}$. Dann ergibt der Satz 9.1.1, dass für fast alle $x \in X$ das zeitliche Mittel

$$Z_K(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_K(w^j(x))$$

existiert und mit $\mu(K)$ übereinstimmt, denn $E(\chi_K|\mathcal{A}^w)$ ist die konstante Funktion mit dem einzigen Wert $\int \chi_K \; d\mu = \mu(K)$. (Siehe auch Beispiel 9.1.2.) Das bedeutet grob gesagt, dass wir das Mass μ aus dem statistischen Verhalten von w rekonstruieren können. Weniger grob kann man das so sagen. Wenn μ_1 und μ_2 zwei Wahrscheinlichkeitsmasse auf $\mathcal{A}$ sind, und $\mu_2 \ll \mu_1$ (jede μ_1-Nullmenge ist eine μ_2-Nullmenge), und $w\colon X \to X$ ergodisch ist für beide, dann ist $\mu_1 = \mu_2$. — Denn das räumliche Mittel $\int \chi_K \; d\mu_1 = \mu_1(K)$ ist gleich dem zeitlichen Mittel $Z_K(x)$ (für μ_1-fast alle $x \in X$, und daher auch für μ_2-fast alle $x \in X$), und das ist gleich dem räumlichen Mittel $\int \chi_K \; d\mu_2 = \mu_2(K)$.

Beispiel 9.1.7. Chintschin war besessen von Kettenbrüchen und im Zusammenhang damit vom dynamischen System von Beispiel 4.3.10. (Hier muss ich die Bezeichnungen ändern, also $w\colon [0,1] \to [0,1]$ oder $w\colon X \to X$ statt f für das dynamische System, und f für irgendwelche messbaren Funktionen auf $[0,1] = X$.) Dieses dynamische System *ist* ergodisch, wie wir schon bewiesen haben. Als Beispiel für ein f wie in Satz 9.1.1 nennt Chintschin

$$f(x) := \ln\lfloor x^{-1} \rfloor$$

für $x > 0$ (der Wert an der Stelle 0 kann uns egal sein). Wenn wir $x \notin \mathbb{Q}$ annehmen (keine grosse Einschränkung, da $\mathbb{Q} \cap [0,1]$ eine γ-Nullmenge ist) und x als Kettenbruch mit den Teilnennern $a_1, a_2, \dots$ schreiben, dann ist $f(x) = a_1$, $f(w(x)) = a_2$, $f(w(w(x))) = a_3$ usw., und daher

$$\frac{1}{n}(S_n f)(x) = \frac{\ln(a_1 a_2 a_3 \cdots a_{n-1} a_n)}{n} = \ln\big((a_1 a_2 a_3 \cdots a_{n-1} a_n)^{1/n}\big).$$

Das ist der Logarithmus vom geometrischen Mittel von $a_1, a_2, \dots, a_n$. Die Aussage des Satzes ist also in diesem Fall: für $x \in [0,1] \smallsetminus \mathbb{Q}$ mit Kettenbruchdarstellung $[a_1, a_2, a_3, ...]$ ist fast sicher

$$\lim_{n\to\infty} (a_1 a_2 a_3 \cdots a_{n-1} a_n)^{1/n} = \exp\big(\int f \, d\gamma\big)$$

wobei die rechte Seite unabhängig von x ist. (Sie heisst natürlich: die Chintschin-Konstante.) Um sie genauer zu bestimmen, müssen wir etwas rechnen. Ich glaube, man versteht die Rechnung besser, wenn man hier nicht eine genaue Formel für f benutzt, sondern nur $f(x) = u(\lfloor x^{-1} \rfloor)$ schreibt für eine (verhandelbare) Funktion u auf der Menge der positiven ganzen Zahlen. Dann ergibt sich

$$\int f \, d\gamma = \sum_{n=1}^{\infty} \frac{1}{\ln 2} \int_{1/(n+1)}^{1/n} \frac{u(n)}{1+x} \, dx = \sum_{n=1}^{\infty} \frac{u(n)}{\ln(2)} \cdot \ln\big(1 + \frac{1}{n(n+2)}\big).$$

Wenn $u(n) = \ln(n)$, dann konvergiert diese Reihe zum Glück (wird dominiert durch die Reihe $\sum_n \log_2(n) \cdot n^{-2}$). Die Chintschin-Konstante ist also exp von der Summe. Dafür kann man auch schreiben

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{n(n+2)}\right)^{\log_2(n)}.$$

Bemerkung 9.1.8. Die Kombination von Kettenbruchtheorie und Theorie der dynamischen Systeme geht auf Gauss und Kuzmin zurück. Die Bachelorarbeit, die ich erwähnt habe, hiess deshalb auch *Der Satz von Gauss-Kuzmin aus der Sicht der Ergodentheorie.* Aber was war eigentlich der Satz von Gauss-Kuzmin, und haben wir ihn jetzt bewiesen, oder wurde er in der Bachelorarbeit vollständig bewiesen?

Bei mathoverflow[2] lese ich, dass Gauss vorgeschlagen oder notiert hatte

$$\lim_{n\to\infty} \lambda(w^n \le s) = \log_2(1+s).$$

Dabei ist λ das Lebesguemass auf $X = [0,1]$, und $(w^n \le s)$ ist natürlich eine Abkürzung für die Teilmenge $\{x \in X \mid w^n(x) \le s\}$ von X, bei festem $s \in [0,1]$. Ausserdem ist $\log_2(1+s)$ dasselbe wie $\ln(1+s)/\ln(2)$, was uns vielleicht vertrauter vorkommt. Möglicherweise hatte Gauss behauptet, einen Beweis zu haben, hatte aber keinen aufgeschrieben, so dass R Kuzmin im Jahr 1928 die Gelegenheit ergreifen konnte, genau das zu tun.

Diese Gleichung von Gauss oder Kuzmin kann auch so geschrieben werden:

$$\lim_{n\to\infty} \lambda(w^n \le s) = \gamma([0,s]).$$

[2]Ein Internet-Diskussionsforum. Vorsicht, kann abhängig machen.

Haben wir das bewiesen? Ich fürchte, wir haben weniger bewiesen.

Ich schlage vor, dass wir Satz 9.1.1 anwenden mit $X = [0,1]$, $\mu := \gamma$, $w(x) = x^{-1} - \lfloor x^{-1} \rfloor$ für $x > 0$, und ausserdem $f = \chi_{[0,s]}$ nehmen, Indikatorfunktion von $[0,s]$. Wenn wir dann nämlich die Gleichung "Zeitmittel = räumliches Mittel" (für dieses f) mit dem gewöhnlichen (!) Lebesgue-Mass λ integrieren, und wenn wir es wagen, die Grenzwertbildung mit dem Integral zu vertauschen[3], dann erhalten wir

$$\int Z_f \, d\lambda = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \lambda(w^j \leq s).$$

Andererseits soll ja Z_f nach dem Satz konstant sein (ausserhalb von einer γ-Nullmenge, die dann auch eine λ-Nullmenge ist), und diese Konstante ist $\gamma([0,s])$. Wenn wir diese Konstante integrieren (auf $[0,1]$, mit λ), erhalten wir nur wieder dieselbe Konstante. Also ergibt sich

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \lambda(w^j \leq s) = \gamma([0,s]).$$

Das ist schwächer als die Aussage von Gauss-Kuzmin[4].

Natürlich wäre es schön, wenn man die Gauss-Kuzmin-Aussage mit allgemeinen Sätzen der Ergodentheorie beweisen könnte. Das geht wohl auch, und ich habe dazu ein paar Ideen und Definitionen gesammelt. (Dazu ist aber in dieser Vorlesung keine Zeit mehr.)

- Das dynamische System von Beispiel 4.3.10 und Beispiel 9.1.7 ist *exakt*. Siehe Definition 9.1.10.
- Aus *exakt* folgt *mischend*. Siehe Definition 9.1.12. Das sieht eher schwierig aus. Ich habe es selber noch nicht verstanden. Bei Einsiedler-Schmidt steht es nicht.
- Die Gauss-Kuzmin-Gleichung ist eine Anwendung der Eigenschaft *mischend*.

DEFINITION 9.1.9. Sei X Menge mit σ-Algebra $\mathcal{A}$, also $(X, \mathcal{A})$ messbarer Raum. Eine Abbildung $w\colon X \to X$ soll *beidseitig messbar* heissen, wenn für jedes $S \in \mathcal{A}$ gilt, dass $w(S) \in \mathcal{A}$ und $w^{-1}(S) \in \mathcal{A}$. (Zum Vergleich: für *messbar* verlangen wir nur, dass $w^{-1}(S) \in \mathcal{A}$ für alle $S \in \mathcal{A}$.)

DEFINITION 9.1.10. Sei $(X, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und sei $w\colon X \to X$ beidseitig messbar und masserhaltend.[5] Eine messbare Teilmenge $Y \subset X$ soll *w-gesättigt* heissen[6], wenn für jedes $k \geq 0$ die Mengen Y und $w^{-k}(w^k(Y)) := \{x \in X \mid w^k(x) \in w^k(Y)\}$ fast

---

[3] Satz von der dominierten Konvergenz benutzen. Brokate-Kersting Seite 43, Satz 5.4.

[4] Es ist leicht, eine Folge von reellen Zahlen $(z_j)_{j=0,1,2,\ldots}$ zu konstruieren derart, dass

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} z_j$$

existiert, während $\lim_{j\to\infty} z_j$ nicht existiert. Zum Beispiel: $z_j = 1$ wenn j eine Potenz von 2 ist, $z_j = 0$ sonst. Wenn allerdings $\lim_{j\to\infty} z_j$ existiert, dann existiert auch der andere Limes und die beiden stimmen überein:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} z_j = \lim_{j\to\infty} z_j\,.$$

[5] *Masserhaltend* hat hier immer noch die alte Bedeutung: $\mu(w^{-1}(S)) = \mu(S)$ für alle $S \in \mathcal{A}$.

[6] Meine Wortwahl.

übereinstimmen[7]. Das dynamische System w ist *exakt*, wenn $\mu(Y) \in \{0,1\}$ gilt für jede w-gesättigte (messbare) Teilmenge Y von X.

Bemerkung. Diese Definition ist im Falle von invertierbarem w nutzlos. Denn dann ist die Bedingung *w-gesättigt* für alle messbaren Teilmengen Y von X erfüllt. Das heisst, ein solches w kann nur exakt sein, wenn μ nur die Werte 0 und 1 annimmt.

Bemerkung. Wenn $w\colon X \to X$ exakt ist (wie in der Definition), dann ist w auch ergodisch. Denn jede w-invariante Teilmenge Y von X ist auch w-gesättigt:

$$w^{-k}(w^k(Y)) = w^{-k}(w^k(w^{-k}(Y))) \subset w^{-k}(Y) = Y$$

und $Y \subset w^{-k}(w^k(Y))$ is trivial.

Bemerkung. Es gibt auch eine Version der Definition von *exakt*, die ohne die Annahme der beidseitigen Messbarkeit auskommt (d.h. nur mit der üblichen Annahme von Messbarkeit). Diese Version ist aber komplizierter, und ich weiss nicht, ob dann noch ein guter Zusammenhang mit *mischend* besteht.

SATZ 9.1.11. *Das dynamische System von Gauss-Kuzmin und Beispiel 9.1.7, Beispiel 4.3.10* (Bezeichnungen wie in Beispiel 9.1.7) *ist exakt.*

Beweis. Der Beweis von *beidseitig messbar* ist Routine, und ich lasse das aus. Den anderen Teil von *exakt* haben wir eigentlich schon bewiesen (als wir bewiesen haben, dass w ergodisch ist), nur haben wir es nicht bemerkt. Im Abschnitt *Noch mehr zur Strategie* wurde geschrieben: *Hier ist aber $g^{-1}(S) = g^{-1}(f^{-n}(S)) = \dots$.* Das muss ersetzt werden durch *Hier ist aber $g^{-1}(S)$ fast identisch mit $g^{-1}(f^{-n}(f^n(S)))$ ....* In Lemma 7.2.3 kann dann die Annahme, dass S invariant ist unter f, ersetzt werden durch die schwächere Annahme, dass S nur f-gesättigt ist. Dann erhält man zwar nicht $\lambda(S \cap \Delta) = \int_S |g'(x)|\, dx$ wie vorher, aber man erhält immer noch

$$\lambda(S \cap \Delta) = \int_{f^n(S)} |g'(x)|\, dx\,.$$

Damit kann man weiter verfahren wie im Beweis von diesem Lemma. Beachten, dass $\mu(f^n(S)) = \mu(f^{-n}(f^n(S))$ weil f masserhaltend ist, und ausserdem $\mu(f^{-n}(f^n(S)) = \mu(S)$ weil $f^{-n}(f^n(S))$ fast mit S übereinstimmt. Also $\mu(f^n(S)) = \mu(S)$. □

DEFINITION 9.1.12. Sei $(X, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und sei $w\colon X \to X$ messbar und masserhaltend. Das dynamische System w heisst (stark) *mischend*, wenn für beliebige messbare Teilmengen U und V von X gilt, dass

$$\lim_{k\to\infty} \mu(w^{-k}(A) \cap B) = \mu(A)\mu(B).$$

Hier ist $w^{-k}(A)$ nur eine Abkürzung für $\{x \in X \mid w^k(x) \in A\}$.

Bemerkung. Aus *mischend* folgt sofort *ergodisch*. Denn angenommen, w ist mischend, und ein messbares $S \subset X$ erfüllt die Invarianzbedingung $w^{-1}(S) = S$. In der Definition von *mischend* können wir $A = S = B$ nehmen und erhalten $\mu(S) = \mu(S \cap S) = \mu(S)\mu(S)$, also $\mu(S) \in \{0,1\}$. Damit ist gezeigt, dass w ergodisch ist. □

LEMMA 9.1.13. *Sei $(X, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und sei $w\colon X \to X$ messbar, masserhaltend und mischend. Dann gilt für beschränkte messbare Funktionen $u, v\colon X \to \mathbb{R}$, dass*

$$\lim_{n\to\infty} \int_B u \cdot (v \circ w^n)\; d\mu = \int u\; d\mu \cdot \int v\; d\mu.$$

[7] Zwei messbare Teilmengen A, B von X stimmen fast überein, wenn $\mu(A) = \mu(A \cap B) = \mu(B)$ gilt.

Beweis-Skizze. Man kann das leicht auf den Fall zurückführen, wo u und v Indikatorfunktionen sind, also $u = \chi_A$ und $v = \chi_B$ für messbare $A, B \subset X$. Dann wird die zu beweisende Behauptung

$$\lim_{n\to\infty} \mu(w^{-n}(A) \cap B) = \mu(B)\mu(A)$$

und das gilt nach Definition. □

Beweis der Gauss-Kuzmin-Aussage mit Benutzung von exakt $\Rightarrow$ mischend*.*

Unser dynamisches System (wie in Beispiel 9.1.7 und Beispiel 4.3.10, aber mit Bezeichnungen wie in Beispiel 9.1.7) ist exakt nach Satz 9.1.11, also mischend. Für die linke Seite der Gauss-Kuzmin-Gleichung,

$$\lim_{n\to\infty} \lambda(w^n \leq s),$$

können wir auch schreiben

$$\lim_{n\to\infty} \int g \cdot (\chi_{[0,s]} \circ w^n)\ d\gamma$$

wobei $g(x) = \ln(2) \cdot (1 + x)$. (Denn λ kann durch γ mit der Dichtefunktion g ausgedrückt werden; normalerweise machen wir es umgekehrt unter Benutzung von $1/g$.) Wegen Lemma 9.1.13 vereinfacht sich das zu

$$\int g\ d\gamma \cdot \int \chi_{[0,s]}\ d\gamma = 1 \cdot \gamma([0, s]).$$

Und das ist die rechte Seite der Gauss-Kuzmin-Gleichung. □