(Elemente der) Maß- und Wahrscheinlichkeitstheorie

Ein paar Notizen

"Diese Überzeugung von der Lösbarkeit eines jeden mathematischen Problems ist uns ein kräftiger Ansporn während der Arbeit; wir haben in uns den steten Zuruf: Da ist das Problem, suche die Lösung. Du kannst sie durch reines Denken finden; denn in der Mathematik gibt es kein Ignorabimus[1]"[Hilbert, 1900]

[1] Lat. Ignoramus et ignorabimus: "Wir wissen es nicht und wir werden es niemals wissen" ist ein Ausspruch von Emil Heinrich Du Bois-Reymond, einem deutschen Physiologen, der damit seine Skepsis gegenüber den Erklärungsansprüchen der Naturwissenschaften zum Ausdruck brachte. Dem entgegnete Hilbert 1930 in Königsberg auf der Versammlung Deutscher Naturforscher und Ärzte mit den Worten "Wir müssen wissen, wir werden wissen". Diesen Ausspruch sollte man allerdings sicherlich nicht allzu naiv wortwörtlich als Forderung nach einer positiv- oder negativ-affirmativen Antwort auf jedwedes mathematische Problem auffassen, vielmehr sah Hilbert auch Unmöglichkeitsbeweise als akzeptable Lösung eines Problems an. Im Schlußwort der *Grundlagen der Geometrie* betont Hilbert den *"Grundsatz, eine jede sich darbietende Frage in der Weise zu erörtern, daß wir zugleich prüften, ob ihre Beantwortung auf einem vorgeschriebenen Wege mit gewissen eingeschränkten Hilfsmitteln möglich ist. Dieser Grundsatz scheint mir eine allgemeine und naturgemäße Vorschrift zu enthalten; in der Tat wird, wenn wir bei unseren mathematischen Betrachtungen einem Probleme begegnen oder einen Satz vermuten, unser Erkenntnistrieb erst dann befriedigt, wenn uns entweder die völlige Lösung jenes Problems und der strenge Beweis dieses Satzes gelingt oder wenn der Grund für die Unmöglichkeit des Gelingens und damit zugleich die Notwendigkeit des Mißlingens von uns klar erkannt worden ist. So spielt denn in der neueren Mathematik die Frage nach der Unmöglichkeit gewisser Lösungen oder Aufgaben eine hervorragende Rolle, und das Bestreben, eine Frage solcher Art zu beantworten, war oftmals der Anlaß zur Entdeckung neuer und fruchtbarer Forschungsgebiete. Wir erinnern nur an Abels Beweis für die Unmöglichkeit der Auflösung der Gleichungen fünften Grades durch Wurzelziehen, ferner an die Erkenntnis der Unbeweisbarkeit des Parallelenaxioms und an Hermites und Lindemanns Sätze von der Unmöglichkeit, die Zahlen e und π auf algebraischem Wege zu konstruieren."[Hilbert, 1899, S.124f]*

Die Axiomatisierung der Physik, das sechste von Hilberts 23 Problemen[2], die Hilbert 1900 auf dem parallel zur Pariser Weltausstellung abgehaltenen zweiten internationalen Mathematikerkongress vorstellte, gilt heute immer noch als größtenteils ungelöst. Jedoch ein Aspekt von Hilberts sechstem Problem, die erfolgreiche Axiomatisierung der Wahrscheinlichkeitstheorie, die wesentlich durch Kolmogorov vorangetrieben wurde, kann als ein beeindruckender Beitrag zur Befriedigung des Hilbertschen Anspruchs angesehen werden. Im Vorwort der *Grundbegriffe der Wahrscheinlichkeitsrechnung* schreibt Kolmogorov:

"Zweck des vorliegenden Heftes ist eine axiomatische Begründung der Wahrscheinlichkeitsrechnung. Der leitende Gedanke des Verfassers war dabei, die Grundbegriffe der Wahrscheinlichkeitsrechnung, welche noch unlängst für ganz eigenartig galten, natürlicherweise in die Reihe der allgemeinen Begriffsbildungen der modernen Mathematik einzuordnen. Vor der Entstehung der Lebesgueschen Maß- und Integrationstheorie war diese Aufgabe ziemlich hoffnungslos. Nach den Lebesgueschen Untersuchungen lag die Analogie zwischen dem Maße einer Menge und der Wahrscheinlichkeit eines Ereignisses sowie zwischen dem Integral einer Funktion und der mathematischen Erwartung einer zufälligen Größe auf der Hand. Diese Analogie ließ sich auch weiter fortführen: so sind z. B. mehrere Eigenschaften der unabhängigen zufälligen Größen den entsprechenden Eigenschaften der orthogonalen Funktionen völlig analog. Um, ausgehend von dieser Analogie, die Wahrscheinlichkeitsrechnung zu begründen, hätte man noch die Maß-und Integrationstheorie von den geometrischen Elementen, welche bei Lebesgue noch hervortreten, zu befreien. Diese Befreiung wurde von Fréchet vollzogen. Der diesen allgemeinen Gesichtspunkten entsprechende Aufbau der Wahrscheinlichkeitsrechnung war in den betreffenden mathematischen Kreisen seit einiger Zeit geläufig; es fehlte jedoch eine vollständige und von überflüssigen Komplikationen freie Darstellung des ganzen Systems..."[Kolmogorov, 1933]

Wollen wir den oben anklingenden Optimismus in Bezug auf die fruchtbare Anwendung maßtheoretischer Ergebnisse zur Aufklärung des Wahrscheinlichkeitsbegriffs im Sinne der axiomatischen Methode zum Anlass

[2] Siehe z.B. `https://en.wikipedia.org/wiki/Hilbert%27s_problems`.

nehmen, auf den folgenden Seiten ein bisschen Maß-, Integrations- und Wahrscheinlichkeitstheorie zu betreiben.

Inhaltsverzeichnis

1 Maßtheorie 5
1 Einstimmung und Vorbereitungen 5
2 Mengensysteme und Maße 16
2.1 Hüllensysteme und Hüllenoperatoren 19
2.2 Konstruktion von Maßen 25
3 Messbare Abbildungen und Bildmaße 37

2 Integrationstheorie 43
4 Integral für nichtnegative einfache Funktionen 46
5 Integral für nichtnegative messbare Funktionen 51
6 Integral für integrierbare Funktionen 57
7 Abschließende Bemerkungen zur Integrationstheorie 64

3 Konvergenzbegriffe für Zufallsvariablen 69
8 Die L^p-Räume, der Hilbertraum L^2 und Konvergenz im p-ten Mittel 69
9 Weitere Konvergenzbegriffe 71

4 Bedingte Erwartung und bedingte Verteilung 85

5 Produktmaß und stochastische Unabhängigkeit 93

6 Grenzwertsätze 103
10 Klassische Grenzwertsätze 103

11 Gleichmäßige Konvergenz, Glivenko-Cantelli, Vapnik-Chervonenkis: Der “Hauptsatz der Statistik” und ein bisschen mehr 107
11.1 Glivenko-Cantelli-Klassen 108
11.2 Ein bisschen Vapnik-Chervonenkis-Theorie 125
Literaturverzeichnis 133

Kapitel 1
Maßtheorie

1 Einstimmung und Vorbereitungen

Wir beginnen damit, ein paar (vorwiegend negative) Resultate der Maßtheorie zusammenzutragen. (Wir orientieren uns hier stark an [Elstrodt, 1996, S.4-6], allgemein sind die Abschnitte zur Maß- und Integrationstheorie vorwiegend an Elstrodt [1996] angelehnt.)

Das Maßproblem. Gesucht ist eine auf der Potenzmenge des $\mathbb{R}^d$ erklärte "Maßfunktion"

$$\mu : 2^{\mathbb{R}^d} \longrightarrow [0,\infty]$$

mit folgenden Eigenschaften:

i) σ-Additivität: Für jede Folge $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkter Teilmengen von $\mathbb{R}^d$ gilt
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

ii) Bewegungsinvarianz: Für jede Bewegung $\beta : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ und für alle $A \subseteq \mathbb{R}^d$ gilt
$$\mu(\beta(A)) = \mu(A).$$

iii) Normiertheit: $\mu([0,1]^p) = 1$.

Satz von Vitali (1905). Das Maßproblem ist unlösbar.

Satz von Banach und Tarski (1924). Es sei $p \geq 3$ und $A, B \subseteq \mathbb{R}^d$ seien beschränkte Mengen mit nicht-leerem Inneren. Dann gibt es Mengen

$C_1, \dots, C_n \subseteq \mathbb{R}^d$ und Bewegungen $\beta_1, \dots, \beta_n$, so daß A die disjunkte Vereinigung der Mengen $C_1, \dots, C_n$ ist und B die disjunkte Vereinigung der Mengen $\beta_1(C_1), \dots, \beta_n(C_n)$.

Satz von Banach und Tarski über das Maßproblem (1924). Es sei $d \geq 1$ und $A, B \subseteq \mathbb{R}^d$ seien beliebige (möglicherweise auch unbeschränkte) Mengen mit nicht-leerem Inneren. Dann gibt es abzählbar viele Mengen $(C_k)_{k \in \mathbb{N}} \in (\mathbb{R}^d)^{\mathbb{N}}$ und Bewegungen $\beta_k : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, so daß A die disjunkte Vereinigung der $(C_k)_{k \in \mathbb{N}}$ ist und B die disjunkte Vereinigung der $(\beta_k(C_k))_{k \in \mathbb{N}}$.

Von den oben aufgelisteten Resultaten werden wir nun stellvertretend den Satz von Vitali für $d = 1$ beweisen. Dazu bedarf es zunächst einiger Vorbereitungen:

Auswahlaxiom. Das Auswahlaxiom ist ein Axiom der axiomatischen Mengenlehre nach Ernst Zermelo und Abraham Adolf Fraenkel. Es wurde erstmals 1904 von Zermelo formuliert und besagt:

"Für jede Menge A von nichtleeren Mengen gibt es eine Auswahlfunktion F, die aus jeder Menge $X \in A$ ein $x \in X$ auswählt:

$$\forall A \text{ Menge von nichtleeren Mengen } \exists F : A \longrightarrow \bigcup A : X \mapsto F(X) \in X.''$$

Diese eventuell intuitiv unmittelbar einleuchtende Aussage ist in dieser Allgemeinheit jedoch nicht in naheliegender Weise irgendwie mathematisch beweisbar. Um die Aussage des Auswahlaxioms universell verwenden zu können, muss man sich schon "dazu entschließen", sie eben als Axiom in seiner vollen Pracht einfach zu akzeptieren. Man kann das Auswahlaxiom selbstverständlich auch einfach nicht akzeptieren[3], dann kann man allerdings bestimmte eventuell zeigenswerte Dinge nicht beweisen: Das

[3] Natürlich besteht auch noch die Möglichkeit, das Auswahlaxiom nur in einer abgeschwächten Form zu akzeptieren. Ein Beispiel für eine Abschwächung des Auswahlaxioms wäre das Axiom der abhängigen Auswahl, siehe `https://en.wikipedia.org/wiki/Axiom_of_dependent_choice`. In dem Fall wäre die Konstruktion einer nicht-messbaren Menge nicht mehr möglich, siehe [Solovay, 1970].

Auswahlaxiom wird oft, wie auch hier im Beweis des Satzes von Vitali, verwendet, um abstrakte Existenzbeweise zu führen, die aber nur die Existenz von gewissen Objekten zeigen, nicht aber über deren konkrete Beschaffenheit Auskunft geben. Mit Hilfe des Auswahlaxioms werden wir eine "nicht vernünftig bemessbare" Menge, die sogenannte Vitali-Menge "konstruieren", ohne deren Gestalt zu begreifen.

Das Auswahlaxiom ist äquivalent zu einer Reihe anderer prominenter mathematischer Prinzipien bzw. Aussagen wie beispielsweise:

i) Dem Zornschen Lemma, das besagt, dass für jede nichtleere geordnete Menge, in der jede Kette (d.h. jede total geordnete Teilmenge) eine obere Schranke besitzt, jedes Element unter einem maximalen Element liegt.
ii) Dem Wohlordnungsprinzip, das besagt, dass man jede Menge wohlordnen kann, d.h., dass man auf jeder Menge eine totale Ordnung einführen kann, so dass jede beliebige nicht-leere Teilmenge ein kleinstes Element besitzt.
iii) Der Aussage, dass jeder Vektorraum eine Basis besitzt.
iv) Der Aussage, dass jede unendliche Menge A gleichmächtig zum kartesischen Produkt $A \times A$ ist.[4]

Im Folgenden werden wir das Auswahlaxiom einfach als geltend annehmen. (Wenn wir das Auswahlaxiom nicht akzeptierten, dann würden wir zwar das Vitalische Unmöglichkeitsergebnis vermeiden, wir würden dann aber trotzdem nicht wissen, ob es eine "Maßfunktion" gibt, geschweige denn, dass wir eine "Maßfunktion" konstruieren könnten.)

Für den Beweis des Satzes von Vitali benötigen wir jetzt noch ein weiteres wichtiges mathematisches Konzept, das uns auch noch an anderen Stellen behilflich sein wird:

---

[4] Die Tatsache, dass aus dieser Aussage das Auswahlaxiom folgt, ist genau die Aussage von Tarski's theorem of choice (1924, siehe `http://pldml.icm.edu.pl/pldml/element/bwmeta1.element.bwnjournal-article-fmv5i1p18bwm`). Die andere Richtung der Äquivalenz war damals schon bekannt. Tarski berichtete Jan Mycielski (Mycielski [2006]) dazu, dass er sein Theorem im Journal Comptes Rendus de l'Académie des Sciences Paris veröffentlichen wollte, jedoch Féchet und Lebesgue beide ablehnten, Fréchet mit der Begründung, dass eine Implikation zwischen zwei wohlbekannten Propositionen kein neues Ergebnis sei, Lebesgue mit der Erwiderung, dass eine Implikation zwischen zwei falschen Aussagen nicht interessiere.

Definition 1 (Äquivalenzrelation). Sei X eine beliebige Menge. Eine binäre Relation $R \subseteq X \times X$ heißt **Äquivalenzrelation**, falls sie folgende Eigenschaften besitzt:

$$\begin{aligned} \text{Reflexivität: } & \forall x \in X : xRx \\ \text{Transitivität: } & \forall x,y,z \in X : xRy \;\&\; yRz \Longrightarrow xRz \\ \text{Symmetrie: } & \forall x,y \in X : xRy \Longrightarrow yRx \end{aligned}$$

Bemerkung 1.1 *Die Notation xRy meint, dass das Paar (x,y) in R ist (sogenannte Infixnotation). Äquivalenzrelationen werden wir im Folgenden typischerweise mit $\sim$ bezeichnen. Für $x \in X$ bezeichnen wir mit $[x]_\sim := \{y \in X \mid x \sim y\}$ die Menge aller Elemente, die äquivalent zu x sind (die sogenannte Äquivalenzklasse von x). Der Raum aller Äquivalenzklassen wird als Faktorraum oder auch als Quotientenraum bezeichnet und mit $X_{/\sim}$ notiert. Eine Äquivalenzrelation auf einer Grundmenge X zerlegt X in die Menge aller Äquivalenzklassen:*

$$X = \dot{\bigcup} \{[x]_\sim \mid x \in X\}.$$

Beweis (Zum eigentlichen Beweis des Satzes von Vitali für $p = 1$). Wir werden eine Menge V konstruieren, die eine abzählbare, disjunkte Vereinigung $V = \dot{\bigcup}_{k \in \mathbb{N}} V_k$ von Mengen V_k, die Translationen voneinander sind, ist. Wir werden dann zeigen, dass

$$[0,1] \subseteq V \subseteq [-1,2]$$

gilt. Daraus folgt, dass für jede “Maßfunktion” μ gilt:

$$1 = \mu([0,1]) \overset{(*)}{\leq} \mu(V) \overset{(**)}{\leq} \mu([-1,2]) = 3,$$

denn μ ist notwendigerweise isoton. (Isoton bedeutet: $A \subseteq B \Longrightarrow \mu(A) \leq \mu(B)$.) Dies führt zu folgendem Widerspruch:

$$\mu(V) = \mu\left(\dot{\bigcup_{k \in \mathbb{N}}} V_k\right) = \sum_{k \in \mathbb{N}} \mu(V_k) = \sum_{k \in \mathbb{N}} c$$

für ein $c \geq 0$. Falls $c = 0$, dann folgt $\mu(V) = 0$, was im Widerspruch zu $(*)$ ist. Falls $c > 0$, dann folgt $\mu(V) = \infty$, was im Widerspruch zu $(**)$ ist.

Nun zur eigentlichen Konstruktion von

$$V = \dot{\bigcup_{k\in\mathbb{N}}} V_k :$$

Definiere $\sim\subseteq [0,1]\times[0,1]$ durch

$$\sim := \{(x,y)\in[0,1]\times[0,1] \mid x-y\in\mathbb{Q}\}.$$

Dies ist in der Tat eine Äquivalenzrelation, siehe Hausübung 0, Aufgabe 2. Betrachte nun

$$S = [0,1]_{/\sim}.$$

Gemäß Auswahlaxiom ist es möglich, aus jeder Äquivalenzklasse von S genau ein Element auszuwählen. Die Menge aller so ausgewählter Elemente sei mit T bezeichnet. Betrachte nun eine Abzählung $(x_k)_{k\in\mathbb{N}}$ der rationalen Zahlen zwischen -1 und 1, d.h., $[-1,1]\cap\mathbb{Q} = \dot{\bigcup\limits_{k\in\mathbb{N}}}\{x_k\}$.
Sei nun

$$V_k := T\oplus x_k \;(=\{x+x_k \mid x\in T\}).$$

Dann gilt: Für $k\neq l$ ist $V_k\cap V_l=\emptyset$ nach Konstruktion, denn für $x\in V_k\cap V_l$ wäre $x=z+x_k=\tilde{z}+x_l$ für geeignete $z,\tilde{z}\in T$. Damit wäre aber $z-\tilde{z} = x-x_k-(x-x_l) = x_l-x_k\in\mathbb{Q}$ und damit $z\sim\tilde{z}$, also $z=\tilde{z}$, denn aus jeder Äquivalenzklasse von S wurde nur genau ein Vertreter ausgewählt. Damit wäre aber $x_k = x-z = x-\tilde{z} = x_l$, was im Widerspruch zur Annahme $x_k\neq x_l$ bzw. $k\neq l$ ist.

Außerdem gilt für $k,l\in\mathbb{N}$ beliebig:

$$\begin{aligned} V_k &= T\oplus x_k \\ &= T\oplus(x_l+x_k-x_l) \\ &= (T\oplus x_l)\oplus(x_k-x_l) \\ &= V_l\oplus(x_k-x_l), \end{aligned}$$

d.h., V_k ist eine Translation von V_l.
Es bleibt noch $[0,1]\subseteq V\subseteq[-1,2]$ zu zeigen:

Zu $[0,1] \subseteq V$: Sei $x \in [0,1]$ beliebig. Dann existiert ein $z \in T$ mit $x \in [z]_\sim$ bzw. $x - z \in \mathbb{Q}$. Daraus folgt $x = \underbrace{z}_{\in T} + \underbrace{x - z}_{\in \mathbb{Q}} \in T \oplus \underbrace{(x-z)}_{\in \mathbb{Q} \cap [-1,1]}$, also ist x aus V.
Zu $V \subseteq [-1,2]$: Sei $x \in V$. Dann ist $x = z + x_k$ mit $z \in T \subseteq [0,1]$ und $x_k \in \mathbb{Q} \cap [-1,1]$. Damit ist $x \in [-1,2]$.

Nach diesen negativen Resultaten folgt abschließend noch ein positives, konstruktives Ergebnis für offene Teilmengen von $\mathbb{R}^1$, das uns Motivation genug sein wird, das Maßproblem in abgeschwächter Form trotzdem anzugehen:

Satz 1.2 (Eindeutige Darstellung offener Teilmengen von $\mathbb{R}^1$) *Jede offene Menge $O \subseteq \mathbb{R}^1$ ist eindeutig (bis auf Permutation der Indizierung) als höchstens abählbare, disjunkte Vereinigung (nicht-leerer) offener Intervalle darstellbar:*

$$O = \dot{\bigcup_{k \in M}} (u_k, o_k)$$

mit M höchstens abzählbar.

Korollar 1.3 ("Bemessbarkeit" offener Teilmengen von $\mathbb{R}^1$) *Jeder offenen Menge $O \subseteq \mathbb{R}^1$ kann eindeutig die Maßzahl*

$$\mu(O) = \mu\left(\dot{\bigcup_{k \in M}} (u_k, o_k)\right) = \sum_{k \in M} (o_k - u_k)$$

zugeordnet werden.

Bemerkung: Die Abbildung

$$\mu : \{O \subseteq \mathbb{R} \mid O \text{ offen } \} \longrightarrow [0,\infty] : O \mapsto \mu(O)$$

ist sogar σ-additiv, vergleiche später.

Vorbemerkungen zum Beweis:

Zur Erinnerung: $(\mathbb{R}, \leq)$ ist eine ordnungsvollständige (total) geordnete Menge. Ordnungsvollständig bedeutet: Jede beschränkte Teilmenge $M \subseteq \mathbb{R}$ besitzt ein Infimum (größte untere Schranke) und ein Supremum (kleinste obere Schranke). Außerdem ist $\mathbb{R}$, aufgefasst als metrischer Raum (, d.h. mit $d(x,y) := |x-y|$), separabel. Separabel bedeutet, dass es eine höchstens abzählbare Menge (nämlich z.B. $\mathbb{Q}$) gibt, deren Abschluss ganz $\mathbb{R}$ ist (i.e.: $\mathbb{Q} \cup \partial\mathbb{Q} = \mathbb{R}$). Anders ausgedrückt: In jeder noch so kleinen Umgebung um einen beliebigen Punkt $x \in \mathbb{R}$ existiert auch ein rationaler Punkt $q \in \mathbb{Q}$. Nun zum eigentlichen Beweis:

Sei $O \subseteq \mathbb{R}$ offen. Für jedes $x \in O$ existiert dann ein (bezüglich Mengeninklusion) maximales offenes Intervall, das x enthält und noch vollständig in O liegt, nämlich das Intervall

$$(u_x, o_x)$$

mit

$$u_x = \inf\{u \in \mathbb{R} \mid (u,x] \subseteq O\}$$
$$o_x = \sup\{o \in \mathbb{R} \mid [x,o) \subseteq O\},$$

wobei die Fälle $u_x = -\infty$ bzw. $o_x = \infty$ hier mit eingeschlossen sind. Außerdem sind obige Mengen nicht-leer, da O offen ist und deshalb gilt $u_x < o_x$, d.h., die Intervalle (u_x, o_x) sind nicht-leer.
Um $(u_x, o_x) \subseteq O$ einzusehen, betrachte z.B. $z \in [x, o_x)$. Da $o_x = \sup\{o \in \mathbb{R} \mid [x,o) \subseteq O\}$, existiert ein $\tilde{z}$ mit $z < \tilde{z} < o_x$ und $[x, \tilde{z}) \subseteq O$, woraus $z \in O$ folgt. Analoges gilt für $z \in (u_x, x]$. Außerdem gilt $o_x \notin O$ und $u_x \notin O$, denn wäre z.B. $o_x \in O$, so gäbe es eine Umgebung $B_\varepsilon(o_x)$ mit $B_\varepsilon(o_x) \subseteq O$, was im Widerspruch zur Konstruktion von o_x als Supremum ist.
Wir haben bis jetzt: $O = \bigcup\limits_{x \in O} (u_x, o_x)$ und werden nun zeigen, dass wir diese Vereinigung disjunkt machen können:

Für $x, x' \in O$ gilt entweder $(u_x, o_x) = (u_{x'}, o_{x'})$ oder $(u_x, o_x) \cap (u_{x'}, o_{x'}) = \emptyset$, denn wäre $(u_x, o_x) \neq (u_{x'}, o_{x'})$ und $(u_x, o_x) \cap (u_{x'}, o_{x'}) \neq \emptyset$, dann wäre $(u_x, o_x) \cup (u_{x'}, o_{x'})$ ein offenes Intervall, dass echt größer als (u_x, o_x) oder als $(u_{x'}, o_{x'})$ wäre und x bzw x' enthielte, was aber im Widerspruch zur Maximalität von (u_x, o_x) bzw. $(u_{x'}, o_{x'})$ wäre. Definiere nun die Äquivalenzrelation

$$\sim = \{(x,x') \in O \times O \mid (u_x, o_x) = (u_{x'}, o_{x'})\}.$$

Dann gilt:

$$O = \bigcup_{x \in O} (u_x, o_x) = \dot{\bigcup_{[x]_\sim \in O_{/\sim}}} (u_x, o_x).$$

Jetzt wäre zunächst noch zu zeigen, dass obige disjunkte Vereinigung eine höchstens abzählbare Vereinigung ist. (Damit ist gemeint, dass es höchstens abzählbar viele Vereinigungspartner gibt, nicht, dass die Menge O selbst höchstens abzählbar ist.) Wäre $O_{/\sim}$ überabzählbar, dann wäre auch die Menge

$$S := \dot{\bigcup_{[x]_\sim \in O_{/\sim}}} \underbrace{(u_x, o_x) \cap \mathbb{Q}}_{\substack{\neq \emptyset, \\ \text{denn } \mathbb{R} \text{ ist separabel}}} \subseteq \mathbb{Q}$$

überabzählbar, was im Widerspruch zur Abzählbarkeit von $\mathbb{Q}$ steht.
Abschließend zeigen wir noch, dass die oben entwickelte Darstellung von O als höchstens abzählbare Vereinigung von offenen Intervallen eindeutig (bis auf Permutation der Abzählung) ist:

Sei dazu

$$O = \dot{\bigcup_{k \in M}} I_k = \dot{\bigcup_{l \in N}} J_l$$

mit (nicht-leeren) offenen Intervallen I_k und J_l und M, N höchstens abzählbar. Betrachte ein beliebiges Intervall I_k und ein beliebiges $x \in I_k \subseteq O$. Dann muss ein J_l existieren, das x enthält. Dieses Intervall muss aber bereits schon mit I_k übereinstimmen, denn wäre $I_k \neq J_l$, dann würde eines der beiden Intervalle einen Randpunkt des anderen Intervalls enthalten. Dies würde bedeuten, dass dieser Randpunkt in der einen Vereinigung von Intervallen enthalten ist, in der anderen aber nicht, was ein Widerspruch zur Tatsache ist, dass beide Vereinigungen die gleiche Menge O darstellen. (Man beachte, dass jeder Randpunkt eines Intervalls nicht in diesem Intervall enthalten ist, aber auch nicht in den anderen offenen Intervallen der betrachteten disjunkten Vereinigung sein kann, sonst wäre die Vereinigung nämlich nicht disjunkt.) Insgesamt können wir also für jedes I_k ein J_l finden, dass mit I_k identisch ist. Aus Symmetriegründen können wir ebenso auch für jedes J_l ein I_k finden, dass zu J_l identisch ist. Damit gilt, dass $(I_k)_{k\in M}$ und $(J_k)_{k\in N}$ bis auf Permutation der Indizierung identische Familien von paarweise disjunkten offenen Intervallen sind.

Mit obigem Ergebnis können wir jetzt schon offene Mengen reeller Zahlen bemessen. Es ist jetzt naheliegend zu fragen, ob wir damit nicht auch schon eine größere Klasse von Mengen in "sinnvoller" und "eindeutiger" Weise bemessen können. Beispielsweise würde es für eine abgeschlossene Menge $C \subseteq [0,1]$ auf der Hand liegen, $\mu(C)$ als $1-\mu([0,1]\backslash C)$ zu definieren, denn $[0,1]\backslash C$ ist eine offene Menge, der wir ein Maß zuordnen können und es ist naheliegend anzunehmen, dass sich die Maßzahlen der beiden disjunkten Teile C und $[0,1]\backslash C$ zur Maßzahl ihrer disjunkten Vereinigung $[0,1]$, also zu 1 addieren. Ganz allgemein stellt sich also die Frage, ob wir eine gegebene Mengenfunktion $\mu : \mathcal{S} \longrightarrow [0,\infty]$ in sinnvoller und eindeutiger Weise auf einen größeren Definitionsbereich erweitern können:

Fortsetzungs- und Eindeutigkeitsproblem. Gegeben sei eine Grundmenge Ω, zwei Mengensysteme $\mathcal{S},\mathcal{F} \subseteq 2^\Omega$ mit $\mathcal{S} \subseteq \mathcal{F}$ und eine Mengenfunktion

$$\mu : \mathcal{S} \longrightarrow [0,\infty]$$

mit bestimmten Eigenschaften $E()$.
Gesucht ist dann eine Fortsetzung

$$\nu : \mathcal{F} \longrightarrow [0,\infty]$$

mit $\nu_{|\mathcal{S}} = \mu_{|\mathcal{S}}$, die ebenfalls die Eigenschaften $E()$ besitzt.

Außerdem stellt sich die Frage, ob diese eindeutig ist.

Beispiel 1.4 (Fortsetzungskonstruktionen) *Für $\mu : \mathcal{S} \longrightarrow [0,\infty]$ kann man sich beispielsweise*

$$\tilde{\mu} : 2^\Omega \longrightarrow [0,\infty] : A \mapsto \inf\{\mu(B) \mid B \in \mathcal{S}, B \supseteq A\}$$

bzw.

$$\underset{\sim}{\mu} : 2^\Omega \longrightarrow [0,\infty] : A \mapsto \sup\{\mu(B) \mid B \in \mathcal{S}, B \subseteq A\}$$

oder auch

$$\mu^* : 2^\Omega \longrightarrow [0,\infty] : A \mapsto \inf\left\{\sum_{n\in\mathbb{N}} \mu(B_n) \mid (B_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}, \bigcup_{n\in\mathbb{N}} B_n \supseteq A\right\}$$

als naheliegende Fortsetzung vorstellen. (Hier ist $\inf\emptyset := \infty$ *und* $\sup\emptyset = 0$ *gesetzt.)*

Konkret würde sich hier beispielsweise für $\mathcal{S} = \mathfrak{I}^1 := \{]a,b] \mid a,b \in \mathbb{R}^1, a < b\}$ *und* $A := \mathbb{Q} \cap [0,1]$ *ergeben:*

$$\begin{aligned}
&\mu^*(A) = 0 \textit{ (warum?)},\\
&\underset{\sim}{\mu}(A) = 0 \textit{ und } \tilde{\mu}(A) = 1 \textit{ (warum?)},
\end{aligned}$$

bzw. auch

$$\inf\left\{\sum_{n=1}^{N} \mu(B_n) \mid N \in \mathbb{N}, B_1, \ldots, B_N \in \mathfrak{I}^1, \bigcup_{n=1}^{N} B_n \supseteq A\right\} = 1 \; (warum?).$$

Dies zeigt bereits, dass z.B. $\tilde{\mu}$ *nicht mehr* σ*-additiv ist, denn*

$$\tilde{\mu}(A) = 1 \neq \sum_{q\in A} \underbrace{\tilde{\mu}(\{q\})}_{=0}.$$

Die Frage ist nun allgemein, welche Eigenschaften von μ sich auf die jeweilige Fortsetzung übertragen und unter welchen Voraussetzungen an μ es sich wirklich um eine Fortsetzung handelt. Weiterhin kann man fragen, für welche A die Gleichheit $\underset{\sim}{\mu}(A) = \tilde{\mu}(A)$ oder auch die Gleichheit $\underset{\sim}{\mu}(A) = \mu^*(A)$ gilt. Im Folgenden ist für uns wegen obiger Überlegungen insbesondere vor allem μ^* von Interesse. Im Fall $\mu(\emptyset) = 0$ kann man zeigen (vgl. Übung 3, Aufgabe 1):

$AM1: \mu^*(\emptyset) = 0$ (Leere Menge ist Nullmenge)

$AM2: A \subseteq B \Longrightarrow \mu^*(A) \leq \mu^*(B)$ (Isotonie)

$AM3: \forall (A_n)_{n\in\mathbb{N}} \in (2^\Omega)^\mathbb{N}: \quad \mu^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu^*(A_n)$ (σ-Subadditivität)

Eine Mengenabbildung mit ganz 2^Ω als Definitionsbereich, die die Bedingungen $AM1$ bis $AM3$ erfüllt, werden wir im Folgenden als **äußeres Maß** bezeichnen.

Wenn μ bereits σ-subadditiv war[5], dann stimmen μ und μ^* auf $\mathcal{S}$ überein und wir haben es tatsächlich mit einer Fortsetzung zu tun (vgl. Hausübung 3, Aufgabe 1)

Wenn $\Omega = \mathbb{R}^d$ und wenn $\mathcal{S}$ abgeschlossen unter Bewegungen[6] ist, dann ist μ^* bewegungsinvariant falls μ bereits bewegungsinvariant[7] ist (vgl. auch Übung 3, Aufgabe 2).

Natürlich kann in diesem Fall die Mengenfunktion μ^*, sofern sie normiert ist, nicht σ-additiv sein. Es ist aber möglich, μ^* auf ein (möglichst großes) Mengensystem $\mathcal{F} \subseteq 2^\Omega$ einzuschränken, auf dem μ^* dann wieder σ-additiv ist. Falls μ nicht zu "schräg" gewählt wurde, sollte $\mathcal{F}$ genügend reichhaltig wählbar sein. Das Mengensystem $\mathcal{F}$ sollte darüber hinaus so beschaffen sein, dass wir möglichst bedenkenlos "rechnen" können, d.h., es wäre wünschenswert, wenn gewisse Mengenoperationen nicht aus dem Mengensystem hinausführen, und man so bei allen denkbaren nötigen

[5] Da μ auf irgendeinem beliebigen Mengensystem $\mathcal{S}$ definiert ist, ist nicht ganz klar, wie man hier die σ-Subadditivität im Fall einer Folge $(A_n)_{n\in\mathbb{N}}$ von Mengen aus $\mathcal{S}$, deren Vereinigung aber nicht in $\mathcal{S}$ ist, verstehen soll. Die folgende Konkretisierung des Begriffs der σ-Subadditivität ist hier zielführend (und im Fall $\mathcal{S} = 2^\Omega$ konsistent mit obiger Definition, sofern man eine isotone Mengenfunktion betrachtet): Eine Mengenfunktion $\mu: \mathcal{S} \longrightarrow [0,\infty]$ heißt σ-subadditiv, wenn für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Mengen aus $\mathcal{S}$ und jede Menge $A \in \mathcal{S}$ mit $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$ die Ungleichung $\mu(A) \leq \sum_{n\in\mathbb{N}} \mu(A_n)$ gilt.

[6] Ein Mengensystem $\mathcal{S} \subseteq 2^{(\mathbb{R}^d)}$ heißt abgeschlossen unter Bewegungen, falls für jede Menge $A \in \mathcal{S}$ und jede Bewegung β auch die bewegte Menge $\beta(A)$ in $\mathcal{S}$ ist.

[7] Eine Mengenfunktion $\mu: \mathcal{S} \longrightarrow [0,\infty]$ auf einem beliebigen Mengensystem $\mathcal{S}$ heißt bewegungsinvariant, falls $\mathcal{S}$ abgeschlossen unter Bewegungen ist und wenn für jede Menge $A \in \mathcal{S}$ und jede Bewegung β gilt: $\mu(\beta(A)) = \mu(A)$.

Umformungen immer auf die σ-Additivität des eingeschränkten äußeren Maßes zurückgreifen kann.

2 Mengensysteme und Maße

Definition 2 (Algebra, vollständige Algebra, σ-Algebra). Gegeben sei eine Grundmenge Ω und ein Mengensystem $\mathcal{A} \subseteq 2^{\Omega}$. Dann heißt $\mathcal{A}$ eine

Algebra, falls

$$A1: \ \emptyset, \Omega \in \mathcal{A} \qquad \text{(Nichttrivialität)}$$
$$A2: \ A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A} \qquad \text{(Abgeschlossenheit unter Komplementbildung)}$$
$$A3: \ A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A} \qquad \text{(Abgeschlossenheit unter endlichen Vereinigungen)}$$

vollständige Algebra, falls

$$V1: \ \emptyset, \Omega \in \mathcal{A} \qquad \text{(Nichttrivialität)}$$
$$V2: \ A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A} \qquad \text{(Abgeschlossenheit unter Komplementbildung)}$$
$$V3: \ I \text{ beliebig}, (A_i)_{i \in I} \in \mathcal{A}^I \Longrightarrow \bigcup_{i \in I} A_i \in \mathcal{A} \qquad \text{(Abgeschlossenheit unter beliebigen Vereinigungen)}$$

σ-Algebra, falls

$$S1: \ \emptyset, \Omega \in \mathcal{A} \qquad \text{(Nichttrivialität)}$$
$$S2: \ A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A} \qquad \text{(Abgeschlossenheit unter Komplementbildung)}$$
$$S3: \ I \text{ abzählbar}, (A_i)_{i \in I} \in \mathcal{A}^I \Longrightarrow \bigcup_{i \in I} A_i \in \mathcal{A} \qquad \text{(Abgeschlossenheit unter abzählbaren Vereinigungen)}$$

Bemerkung 2.1 *Wegen der De Morganschen Gesetze folgt bereits, dass eine Algebra / vollständige Algebra / σ-Algebra auch abgeschlossen ist unter endlichen / beliebigen / abzählbaren Schnitten, vergleiche Tutorium 2, Aufgabe 1.*

Bemerkung 2.2 *Eine vollständige Alegbra, die alle einelementigen Mengen enthält, enthält bereits alle Teilmengen von Ω, denn jedes beliebige $A \subseteq \Omega$ lässt sich darstellen als $A = \bigcup_{\omega \in A} \{\omega\}$. In Anbetracht der Unlösbarkeit des Maßproblems wäre der Wunsch nach einer vollständigen Algebra also eher utopisch. Auf der anderen Seite würde die bescheidene Forderung lediglich einer Algebra als unterliegendem Mengensystem zu einer nicht allzu weit tragenden Theorie führen. Deshalb werden wir im Folgenden Mengenfunktionen auf σ-Algebren betrachten.[8] Kommen wir nun dazu, zu formulieren, was wir von einer Maßfunktion auf einer σ-Algebra gerne hätten:*

Definition 3 (Maß, Wahrscheinlichkeitsmaß). Sei Ω ein Grundraum und $\mathcal{S} \subseteq 2^\Omega$ ein Mengensystem. Ist $\mathcal{S}$ eine σ-Algebra, so wird das Paar $(\Omega, \mathcal{S})$ als Messraum bezeichnet. Mengen $A \in \mathcal{S}$ werden als messbar (bezüglich $\mathcal{S}$) bezeichnet und Mengen $A \in 2^\Omega \setminus \mathcal{S}$ werden als nicht messbar (bezüglich $\mathcal{S}$) bezeichnet.

Eine Mengenfunktion $\mu : \mathcal{S} \longrightarrow M$ mit $M \subseteq \overline{\mathbb{R}}$ heißt Maß, falls $\mathcal{S}$ eine σ-Algebra ist und

$$M1: \ \forall A \in \mathcal{S} : \mu(A) \geq 0 \qquad \text{(Nichtnegativität)}$$

$$M2: \ \mu(\emptyset) = 0 \qquad \text{(Leere Menge ist Nullmenge)}$$

$$M3: \ \text{Für } (A_k)_{k \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}} \text{ paarweise disjunkt gilt:}$$

$$\mu\left(\dot{\bigcup_{k \in \mathbb{N}}} A_k\right) = \sum_{k \in \mathbb{N}} \mu(A_k) \qquad (\sigma\text{-Additivität})$$

In diesem Fall wird das Tripel $(\Omega, \mathcal{S}, \mu)$ als Maßraum bezeichnet. Ein Maß μ mit $\mu(\Omega) = 1$ heißt normiert bzw. Wahrscheinlichkeitsmaß. Ist $\mu(\Omega) < \infty$, so heißt μ endlich. Existiert eine Folge $(A_n)_{n \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ mit $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$, die Ω überdeckt (i.e.: $\Omega = \bigcup_{n \in \mathbb{N}} A_n$), so wird μ als σ-endlich bezeichnet. Mengen A mit $\mu(A) = 0$ werden als μ-Nullmengen bezeichnet. Eine Mengenfunktion μ auf einem beliebigen Mengensystem $\mathcal{S}$ wird als σ-additiv bezeichnet, falls für beliebige paarweise disjunkte Folgen $(A_n)_{n \in \mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$, deren Vereinigung ebenfalls wieder in $\mathcal{S}$ liegt, gilt:

[8] Die Einsicht in die Fruchtbarkeit der σ-Abgeschlossenheit bzw. der σ-Additivität für die Maßtheorie haben wir in großen Teilen Borel zu verdanken, vergleiche zum Beispiel [Elstrodt, 1996, S.3f,S.39].

$$\mu\left(\dot{\bigcup_{n\in\mathbb{N}}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Eine Mengenfunktion μ auf einem beliebigen Mengensystem $\mathcal{S}$ wird als σ-subadditiv bezeichnet, falls für beliebige paarweise disjunkte Folgen $(A_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ und jede Menge $A \in \mathcal{S}$ mit $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$ die Ungleichung

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

gilt. Ferner heißt im Fall $\Omega = \mathbb{R}^d$ eine Mengenfunktion auf einem beliebigen Mengensystem $\mathcal{S} \subseteq 2^{(\mathbb{R}^d)}$ translationsinvariant, falls für alle $c \in \mathbb{R}^d$ und alle $A \in \mathcal{S}$ gilt:

$$A \oplus c := \{a + c \mid a \in A\} \in \mathcal{S} \quad \& \quad \mu(A) = \mu(A \oplus c).$$

Satz 2.3 (Elementare Eigenschaften von Maßen) *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dann gilt:*

i) Endliche Additivität: Für n paarweise disjunkte messbare Mengen $A_1, \dots, A_n$ gilt:

$$\mu\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mu(A_k).$$

ii) Jede messbare Menge A zerlegt jede andere messbare Menge B additiv:

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

iii) μ ist subtraktiv, d.h., für $A, B \in \mathcal{A}$ mit $A \subseteq B$ und $\mu(A) < \infty$ gilt:

$$\mu(B \backslash A) = \mu(B) - \mu(A).$$

iv) μ ist isoton, d.h., für messbare Mengen A, B mit $A \subseteq B$ gilt:

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

v) μ ist modular, d.h., für beliebige $A, B \in \mathcal{A}$ gilt:

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

vi) μ *ist* σ*-subadditiv.*

vii) μ *ist stetig von unten: Für jede aufsteigende Folge* $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ *gilt:*

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right).$$

(Aufsteigend bedeutet: $A_k \subseteq A_l$ *für* $k < l$*.)*

viii) μ *ist stetig von oben: Für jede absteigende Folge* $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ *mit* $\mu(A_n) < \infty$ *für ein* $n \in \mathbb{N}$ *gilt:*

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n\in\mathbb{N}} A_n\right).$$

(Absteigend bedeutet: $A_k \supseteq A_l$ *für* $k < l$*.)*

ix) Ist μ σ*-endlich, so ist jede Familie paarweise disjunkter messbarer Mengen mit positivem Maß höchstens abzählbar.*

Beweis. Siehe Hausübung 2, Aufgabe 1 und Übung 2, Aufgabe 3.

2.1 Hüllensysteme und Hüllenoperatoren

Nachdem wir nun ungefähr wissen, was unser Ziel ist (nämlich ein Maß auf einer möglichst großen σ-Algebra), beschäftigen wir uns jetzt zunächst mit der Frage, wie wir von einem beliebigen Mengensystem $\mathcal{S}$ auf kanonische Weise zu einer σ-Algebra gelangen, die das gegebene Mengensystem $\mathcal{S}$ umfasst. Erst später werden wir uns damit beschäftigen, wie wir ein gegebenes Maß μ, das wir auf $\mathcal{S}$ bereits kennen, auf diese σ-Algebra (oder auf eine noch größere σ-Algebra) fortsetzen können. Allgemein empfiehlt es sich, im Hinterkopf zu behalten, dass wir immer mit zwei Aspekten zu kämpfen haben werden:

i) Der Frage, welche Mengen als messbar deklariert werden. Dies betrifft also das Mengensystem, auf dem eine Mengenfunktion definiert ist.
ii) Der Frage, welche Maßzahlen den als messbar deklarierten Mengen zugeordnet werden sollen.

Wir werden uns jetzt also zunächst mit Frage $i)$ beschäftigen. Danach werden wir Frage $ii)$ angehen, wir werden dabei jedoch bei einer σ-Algebra landen, die typischerweise echt größer ist, als die σ-Algebra, die sich allein aufgrund von Überlegungen zu Frage $i)$ ergibt. Den folgenden Abschnitt hätten wir auch knapper halten können, jedoch stellt er Hilfsmittel bereit, die durchaus zum "strukturellen" Verständnis von Mengensystemen, wie beispielsweise von σ-Algebren dienlich sind.

Definition 4 (Hüllensystem und Hüllenoperator). Sei V eine beliebige Menge und $\mathcal{H} \subseteq 2^V$. Dann heißt $\mathcal{H}$ ein Hüllensystem, falls es V selbst enthält und abgeschlossen ist unter beliebigen Schnitten. Die Mengen aus $\mathcal{H}$ werden als Hüllen bezeichnet. Einem Hüllensystem ist auf kanonische Weise der Operator

$$h_{\mathcal{H}} : 2^V \longrightarrow 2^V : A \mapsto \bigcap \{B \in \mathcal{H} \mid B \supseteq A\}$$

zugeordnet. Für $A \in 2^V$ bezeichnen wir $h_{\mathcal{H}}(A)$ als die Hülle von A.

Satz 2.4 (Hüllenoperatoreigenschaften) *Der Operator $h_{\mathcal{H}}$ hat die folgenden Eigenschaften:*

$$\begin{aligned} \textit{Extensivität: } & \forall A \in 2^V : A \subseteq h_{\mathcal{H}}(A) \\ \textit{Isotonie: } & \forall A, B \in 2^V : A \subseteq B \Longrightarrow h_{\mathcal{H}}(A) \subseteq h_{\mathcal{H}}(B) \\ \textit{Idempotenz: } & \forall A \in 2^V : h_{\mathcal{H}}(h_{\mathcal{H}}(A)) = h_{\mathcal{H}}(A). \end{aligned}$$

Außerdem gilt

$$A \in \mathcal{H} \iff h_{\mathcal{H}}(A) = A.$$

("Eine Menge ist eine Hülle genau dann, wenn sie mit ihrer Hülle übereinstimmt.")

Beweis: Vergleiche Übung 3, Aufgabe 1.

Bemerkung 2.5 *Jeder Operator $h : 2^V \longrightarrow 2^V$, der extensiv, isoton und idempotent ist, wird als (mengentheoretischer) Hüllenoperator bezeichnet. Außerdem wird jeder Hüllenoperator h von dem Hüllensystem $\mathcal{H}_h := \{h(A) \mid A \in 2^V\}$ erzeugt, d.h., es gilt:*

$$h_{\mathcal{H}_h} = h.$$

Weiterhin gilt für ein beliebiges Hüllensystem $\mathcal{H}$:

$$\mathcal{H}_{h_{\mathcal{H}}} = \mathcal{H}.$$

Bemerkung 2.6 *Die Eigenschaft einer Abbildung $h : 2^V \longrightarrow 2^V$, Hüllenoperator zu sein, lässt sich einfach charakterisieren über die Eigenschaft*

$$\forall A, B \in 2^V : A \subseteq h(B) \iff h(A) \subseteq h(B)$$

("Eine Menge ist Teilmenge einer Hülle genau dann, wenn es ihre Hülle ist.")
Den Beweis haben Sie in Übung 1, Aufgabe 2 geführt.

Bemerkung 2.7 *Hüllensysteme bzw. Hüllenoperatoren trifft man innerhalb vieler mathematischer Gebiete an, es folgen ein paar Beispiele:*

i) Ist X ein metrischer oder ein topologischer Raum, dann ist die Menge aller abgeschlossenen Teilmengen von M ein Hüllensystem. Der zugehörige Hüllenoperator

$$\overline{} : 2^M \longrightarrow 2^M : A \mapsto \overline{A} := \bigcap \{B \subseteq M, B \text{ abgeschlossen } \mid B \supseteq A\}$$

ordnet jeder Menge A deren Abschluss (, d.h., die kleinste abgeschlossene Menge, die Obermenge von A ist,) zu. In diesem Fall kann der Abschluss auch als $\overline{A} = A \cup \partial A$ angegeben werden, was zumindest im Fall eines metrischen Raumes konkreter ist.

ii) Die Menge aller konvexen Teilmengen des $\mathbb{R}^d$ ist ein Hüllensystem. Die Abbildungsvorschrift des zugehörigen Hüllenoperators

$$conv : 2^{\left(\mathbb{R}^d\right)} \longrightarrow 2^{\left(\mathbb{R}^d\right)} : A \mapsto \bigcap \{B \subseteq \mathbb{R}^d, B \text{ konvex } \mid B \supseteq A\}$$

kann hier ebenfalls etwas konkreter als

$$conv(A) = \left\{ \sum_{i=1}^{n} \lambda_i x_i \mid n \in \mathbb{N}, x_1, \ldots, x_n \in A, \lambda_1, \ldots, \lambda_n \geq 0, \sum_{i=1}^{n} \lambda_i = 1 \right\}$$

bzw. wegen des Satzes von Carathéodory über konvexe Hüllen[9] noch einfacher als

$$conv(A) = \left\{ \sum_{i=1}^{d+1} \lambda_i x_i \mid x_1, \ldots, x_{d+1} \in A, \lambda_1, \ldots, \lambda_{d+1} \geq 0, \sum_{i=1}^{d+1} \lambda_i = 1 \right\}$$

angegeben werden.

---

[9] Siehe zum Beispiel `https://en.wikipedia.org/wiki/Carath%C3%A9odory%27s_theorem_(convex_hull)`

iii) Für gegebene Grundmenge M und $2^{M\times M}$ der Menge aller Relationen auf M ist die Menge aller transitiven Relationen ein Hüllensystem und der zugehörige Hüllenoperator

$$trans : 2^{M\times M} \longrightarrow 2^{M\times M} : R \mapsto \bigcap \left\{ S \in 2^{M\times M} \text{ transitiv } \mid S \supseteq R \right\}$$

ordnet jeder Relation ihrer transitive Hülle zu. Auch hier kann (unabhängig von der Mächtigkeit von M) die Hülle $trans(R)$ konkreter dargestellt werden als $trans(R) = \bigcup\limits_{n\in\mathbb{N}} R^n$, wobei hier R^n das n-fache Relationenprodukt von R mit sich selbst bezeichnet. (Konkret ist hier $R^1 := R$, $R * R = \{(x,y) \in M \mid \exists z \in M : xRz \;\&\; zRy\}$ und $R^{n+1} := (R^n) * R$.)

Betrachtet man nicht Selbstabbildungen auf einer Potenzmenge 2^V, sondern Selbstabbildungen auf einer geordneten Menge[10] $(V, \leq)$ und ersetzt man in der Formulierung der 3 Eigenschaften Extensivität, Isotonie und Idempotenz die Teilmengenrelation $\subseteq$ jeweils durch die Ordnungsrelation $\leq$, so gelangt man in Abstraktion eines mengentheoretischen Hüllenoperators zu einem ordnungstheoretischen Hüllenoperator. Ein sehr einfaches Beispiel für einen ordnungstheoretischen Hüllenoperator auf der total geordneten Menge $(\mathbb{R}, \leq)$ ist der Rundungsoperator

$$\lceil\ \rceil : \longrightarrow \mathbb{R} : a \mapsto \lceil a \rceil := \inf\{b \in \mathbb{Z} \mid b \geq a\}.$$

(Man beachte die formale Analogie zur mengentheoretischen Hüllenoperatorkonstruktion.)

Satz 2.8 (System aller σ-Algebren ist Hüllensystem) *Sei Ω eine beliebige Grundmenge und $V = 2^\Omega$. Dann ist das System*

$$\mathfrak{M} := \{\mathcal{A} \subseteq 2^\Omega \mid \mathcal{A} \text{ ist } \sigma\text{-Algebra über } \Omega\}$$

aller σ-Algebren über Ω ein Hüllensystem. Der zugehörige Hüllenoperator

$$h_{\mathfrak{M}} : 2^{(2^\Omega)} \longrightarrow \mathfrak{M} : \mathcal{S} \mapsto \bigcap\{\mathcal{A} \in \mathfrak{M} \mid \mathcal{A} \supseteq \mathcal{S}\}$$

[10] Eine geordnete Menge $(V, \leq)$ ist ein Paar bestehend aus einer Grundmenge V und einer reflexiven, transitiven und antisymmetrischen Relation $\leq$ auf V. Antisymmetrisch bedeutet dabei, dass für beliebige $a, b \in V$ die Implikation $a \leq b \;\&\; b \leq a \Longrightarrow a = b$ gilt. Man beachte, dass eine geordnete Menge nicht total geordnet sein muss, d.h., es kann durchaus Paare (a, b) geben, für die weder $a \leq b$, noch $b \leq a$ gilt.

wird als σ-Operator bezeichnet und mit σ notiert. Ebenso ist das System aller Algebren und auch das System aller vollständigen Algebren über Ω ein Hüllensystem.

Beweis: Vergleiche Hausübung 1, Aufgabe 1.

Bemerkung 2.9 *Der σ-Operator ordnet jedem Mengensystem $\mathcal{S}$ über Ω die (eindeutig bestimmte und immer existierende) kleinste σ-Algebra zu, die $\mathcal{S}$ umfasst. Ausführlicher ausgedrückt bedeutet dies dreierlei:*

i) $\sigma(\mathcal{S})$ *ist eine σ-Algebra, die $\mathcal{S}$ umfasst im Sinne von $\sigma(\mathcal{S}) \supseteq \mathcal{S}$.*
ii) $\sigma(\mathcal{S})$ *ist minimal bezüglich Eigenschaft i), d.h., für jede weitere σ-Algebra $\mathcal{F}$, die ebenfalls $\mathcal{S}$ umfasst, gilt bereits $\mathcal{F} \supseteq \sigma(\mathcal{S})$.*
iii) $\sigma(\mathcal{S})$ *ist eindeutig bestimmt in dem Sinne, dass jede weitere σ-Algebra $\mathcal{F}$, die ebenfalls die Eigenschaften i) und ii) erfüllt, bereits mit $\sigma(\mathcal{S})$ übereinstimmt, denn für solch eine σ-Algebra $\mathcal{F}$ würde mit ii) die Relation $\sigma(\mathcal{S}) \subseteq \mathcal{F}$ gelten, und wenn $\mathcal{F}$ ebenfalls die Minimalitätseigenschaft ii) erfüllt, folgt $\mathcal{F} \subseteq \sigma(\mathcal{S})$, woraus sich die Gleicheit von $\sigma(\mathcal{S})$ und $\mathcal{F}$ ergibt.*

Bemerkung 2.10 *Für die von einem Mengensystem $\mathcal{S}$ erzeugte σ-Algebra $\sigma(\mathcal{S})$ scheint es für den Fall eines überabzählbaren Grundraumes Ω im Allgemeinen keine so einfache „konstruktive“ Beschreibung zu geben, wie es z.B. für die obigen Beispiele von Hüllenoperatoren der Fall ist, vergleiche z.B. [Elstrodt, 1996, S. 16f].*

Die von einem Mengensystem erzeugte vollständige Algebra ist jedoch leichter zu beschreiben:

Satz 2.11 (Darstellung der erzeugten vollständigen Algebra) *Sei Ω ein beliebiger Grundraum und $\mathcal{S} \subseteq 2^{\Omega}$ ein beliebiges Mengensystem auf Ω. Definiere zunächst die Ununterscheidbarkeitsrelation*

$$\sim_{\mathcal{S}} := \{(\omega, \omega') \in \Omega \times \Omega \mid \forall A \in \mathcal{S} : \mathbb{1}_A(\omega) = \mathbb{1}_A(\omega')\}.$$

Diese Relation ist eine Äquivalenzrelation. Die von $\mathcal{S}$ erzeugte vollständige Algebra $\Sigma_{\mathcal{S}}$ lässt sich nun darstellen als

$$\Sigma_{\mathcal{S}} = \{A \subseteq \Omega \mid A \text{ ist Vereinigung von Äquivalenzklassen von } \Omega_{/\sim_{\mathcal{S}}}\},$$

wobei hier die leere Vereinigung als $\emptyset$ definiert sei.

Dies zeigt insbesondere auch, dass jede vollständige Algebra ordnungstheoretisch ismorph zu einem Potenzmengenverband ist.

Ist weiterhin Ω oder wenigstens $\Omega_{/\sim_S}$ abzählbar, so stimmt die von $\mathcal{S}$ erzeugte σ-Algebra mit der von $\mathcal{S}$ erzeugten vollständigen Algebra überein, denn in dem Fall sind alle Äquivalenzklassen von $\Omega_{/\sim_S}$ aus $\sigma(\mathcal{S})$ und alle Vereinigungen von Äquivalenzklassen sind als abzählbare Vereinigungen darstellbar. Dies bedeutet, dass in diesem Fall auch die erzeugte σ-Algebra einfacher zu beschreiben ist.

Beweis: Vergleiche Übung 2, Aufgabe 2.

Bemerkung 2.12 *Aus obiger Darstellung folgt sofort, dass für Ω endlich oder auch für lediglich $\Omega_{/\sim_S}$ endlich die Mächtigkeit einer σ-Algebra immer eine Potenz von 2 sein muss. Für den Fall, dass $\Omega_{/\sim_S}$ abzählbar unendlich ist, ist $\sigma(\mathcal{S})$ bereits überabzählbar. (Das folgt aus dem Satz von Cantor: Jede Menge ist weniger mächtig als ihre Potenzmenge.) Allgemein gilt für Ω und $\mathcal{S} \subseteq 2^{\Omega}$ beliebig, dass $\sigma(\mathcal{S})$ nur endlich oder überabzählbar sein kann.*

Definition 5 (Borelsche σ-Algebra). Sei $X = \mathbb{R}^d$ (oder auch ein metrischer oder ein topologischer Raum). Bezeichne $\mathcal{O}(X)$ die Menge aller offenen Mengen von X. Die Borelsche σ-Algebra $\mathfrak{B}(X)$ ist die von $\mathcal{O}(X)$ erzeugte σ-Algebra:

$$\mathfrak{B}(X) := \sigma(\mathcal{O}(X)).$$

Satz 2.13 (Erzeuger der Borelschen σ-Algebra) *Sei $X = \mathbb{R}^d$. Die Borelsche σ-Algebra wird beispielsweise ebenso von den folgenden Mengensystemen erzeugt:*

$$
\begin{aligned}
\mathcal{A} &:= \{]a,b[\mid a,b \in \mathbb{R}^d, a < b\} \\
\mathcal{B} &:= \{[a,b] \mid a,b \in \mathbb{R}^d, a < b\} \\
\mathcal{C} &:= \{[a,b[\mid a,b \in \mathbb{R}^d, a < b\} \\
\mathcal{D} &:= \{]a,b] \mid a,b \in \mathbb{R}^d, a < b\} \\
\mathcal{E} &:= \{B_\varepsilon(x) \mid x \in \mathbb{R}^d, \varepsilon > 0\} \\
\mathcal{F} &:= \{]-\infty, c] \mid c \in \mathbb{R}^d\} \\
\mathcal{G} &:= \{[c, \infty[\mid c \in \mathbb{R}^d\}.
\end{aligned}
$$

Zum Beweis: Zum Beispiel für $\mathcal{A}$ und $\mathbb{R}^1$ gilt:
$\mathcal{A} \subseteq \mathcal{O}(\mathbb{R}) \Longrightarrow \sigma(A) \subseteq \sigma(\mathcal{O}(\mathbb{R}))$ *und*
$\mathcal{O}(\mathbb{R}) \subseteq \sigma(\mathcal{A})$ *(warum?)* $\Longrightarrow \sigma(\mathcal{O}(\mathbb{R})) \subseteq \sigma(\sigma(\mathcal{A})) = \sigma(\mathcal{A})$.
Also ist $\sigma(\mathcal{O}(\mathbb{R})) = \sigma(\mathcal{A})$.

Bemerkung 2.14 *An dieser Stelle ist noch nicht ganz klar, dass $\mathfrak{B}(\mathbb{R}^d) \subsetneq 2^{\mathbb{R}^d}$ gilt.*

Bemerkung 2.15 *Allgemein gilt für einen beliebigen,* ***separablen*** *metrischen Raum X:*

$$\sigma(\mathcal{O}(X)) = \sigma(\{B_\varepsilon(x) \mid x \in X, \varepsilon > 0\}).$$

2.2 Konstruktion von Maßen

Kommen wir nun zur eigentlichen Konstruktion von Maßen. Die Idee ist hier, das äußere Maß μ^*, das auf 2^Ω definiert war, wieder einzuschränken auf eine σ-Algebra, auf der μ^* dann hoffentlich wieder alle Eigenschaften eines Maßes besitzt. Eine Wahl wäre z.B. für $\mathbb{R}^d$ die Borelsche σ-Algebra. Da die Borelsche σ-Algebra die kleinste σ-Algebra ist, die alle Intervalle enthält, drängt sich die Frage auf, ob man denn nicht auch noch mit einer größeren σ-Algebra arbeiten kann. Außerdem wäre noch offen zu zeigen, dass (unter gewissen Voraussetzungen an μ) die Einschränkung von μ^* auf $\mathfrak{B}(\mathbb{R}^d)$ zum Beispiel σ-additiv ist, was gar nicht so leicht zu bewerkstelligen scheint, man erinnere sich daran, dass die Borelsche σ-Algebra $\mathfrak{B}(\mathbb{R}^d)$ nur sehr implizit gegeben ist. In der Tat werden wir zunächst eine

σ-Algebra betrachten, die größer als $\mathfrak{B}(\mathbb{R}^d)$ ist, und auf der die Maßeigenschaften von μ^* verhältnismäßig einfach zu zeigen sind. Erst nachträglich werden wir die Einschränkung von μ^* auf $\mathfrak{B}(\mathbb{R}^d)$ betrachten, auf der μ^* dann natürlich ebenfalls die Eigenschaften eines Maßes besitzt. Eine naheliegende Idee wäre jetzt, zunächst nach Mengen zu suchen, die "möglichst einfach zu bemessen" sind. Beispielsweise Intervalle sind einfach zu bemessen. Für kompliziertere Mengen entsteht das äußere Maß μ^* durch eine "Approximation von außen". Um zu analysieren, ob die Approximation von außen "gut genug" ist, wäre es naheliegend, diese Approximation mit einer Approximation von innen zu vergleichen. Für den Fall von beschränkten Teilmengen von $\mathbb{R}$ hat z.B. Lebesgue ein inneres Maß als

$$\mu_*(A) := \sup\{\mu^*(K) \mid K \subseteq A, K \text{ kompakt}\}$$

definiert und Mengen A als messbar bezeichnet, wenn das innere und das äußere Maß übereinstimmen. Das System aller so als messbar definierten Teilmengen eines gegebenen beschränkten Intervalls ist dann eine σ-Algebra und μ^* bzw. μ_* ist ein Maß auf dieser σ-Algebra. Um anschließend ein Maß auf ganz $\mathbb{R}$ zu definieren müsste man dann noch ein bisschen weiter tun. Außerdem kommt in der Konstruktion des inneren Maßes das topologische Konzept der Kompaktheit ins Spiel, was auf abstraktere Situationen, wo man lediglich eine Mengenfunktion auf einem Mengensystem hat, nicht unmittelbar übertragbar ist. Wir werden deshalb im Folgenden nicht den Lebesgueschen Zugang, sondern den Zugang nach Carathéodory, der ganz ohne ein inneres Maß auskommt, verfolgen:

"Borel und Lebesgue haben ... jeder Punktmenge A ein äußeres Maß m^*A und ein inneres Maß m_* A zugeordnet .... Die Punktmenge A wurde meßbar genannt, wenn $m^*A = m_*$ A ist ... Nun habe ich im Juli 1914 den Satz bewiesen: Ist A nach Borel-Lebesgue meßbar, so ist für jede Punktmenge X, ob meßbar oder nicht,

$$m^*X = m^*A \cap X + m^*X \setminus A. \tag{2}$$

Nimmt man (2) *als Definition für die Meßbarkeit, so geht in der Borel-Lebesgueschen Theorie keine meßbare Menge verloren... Die neue Definition hat große Vorteile: ... Die Beweise*[11] *der Hauptsätze der Theorie sind unvergleichlich einfacher und kürzerer als vorher."* [Carathéodory, 1956, S. 276]

[11] Zugegebenermaßen wird unser Vorgehen diesbezüglich nicht ganz konsequent sein, da wir nicht alle Teile z.B. des Fortsetzungssatzes von Carathéodory beweisen werden.

Wir beginnen mit ein paar Vorüberlegungen:

$$\mu : \mathcal{S} \subsetneq 2^\Omega \to [0,\infty] \qquad \longrightarrow \qquad \mu^* : 2^\Omega \to [0,\infty] \qquad \longrightarrow \qquad \mu^*_{|\mathcal{F}} : \mathcal{S} \subsetneq \mathcal{F} \subsetneq 2^\Omega \to [0,\infty]$$

Äußere Maßkonstruktion
Definitionsbereich wird größer,
Eigenschaften von μ,
wie z.B. die (σ)-Additivität
können verloren gehen
Außerdem kann

$\mu \neq \mu^*_{|\mathcal{S}}$ passieren.

Einschränkung auf
möglichst großes
Mengensystem $\mathcal{F}$,
wo alles wieder schön ist.
Am besten: σ-Algebra,
auf der μ^* σ-additiv ist.

$$\mu^*(A) = \inf\left\{ \sum_{n\in\mathbb{N}} \mu(B_n) \mid B_n \in \mathcal{S}^{\mathbb{N}},\ \bigcup_{n\in\mathbb{N}} B_n \supseteq A \right\}$$

Beobachtung: $$B = (B \cap A) \dot\cup (B \cap A^c)$$

Dies sollte ein vernünftige Maß ν "mitmachen" i.S.v:

$$\nu(B) = \nu(B \cap A) + \nu(B \cap A^c) \qquad (Z)$$

Voraussetzung (*):
$\mathcal{S}$ abgeschlossen unter Schnitt und Mengendifferenz
und μ endlich additiv

$$\Longrightarrow \forall A, B \in \mathcal{S} : \mu(B) = \mu(B \cap A) + \mu(B \cap A^c)$$

Was ist mit äußerem Maß μ^*?

(vorausgesetzt $\emptyset \in \mathcal{S}$ und $\mu(\emptyset) = 0$)

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

gilt für beliebige Mengen $A, B \in 2^\Omega$ nach Konstruktion von μ^*, sofern $\mu^*(\emptyset) = 0$,
dies folgt aus der σ-Subadditivität von μ^* (vgl. Übung)

-Wann gilt "$\geq$"?

-Falls $A, B \in \mathcal{S}$ und $\mu^*_{|\mathcal{S}} = \mu$, dann gilt "$\geq$".
-Dies ist der Fall, wenn μ σ-subadditiv ist (Voraussetzung (**),vgl. Hausübung)

-Für welche Mengen gilt "$\geq$" noch?

- Wenn A eine μ^*-Nullmenge ist und $B \in 2^\Omega$ beliebig. Dies folgt aus der Isotonie von μ^*.
- μ^*-Nullmengen sind also kein Problem!
- Wenn A, B beliebig?... schwer zu analysieren
- Wenn nur $A \in \mathcal{S}$ und B beliebig?

$$B \quad = \quad (B \cap A) \ \dot{\cup} \ (B \cap A^c)$$

Für $\mu^*(B)$ betrachte $\sum\limits_{n\in\mathbb{N}} \mu(B_n)$
mit $(B_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ und $\bigcup\limits_{n\in\mathbb{N}} B_n \supseteq B$:

Da $A, B_n \in \mathcal{S}$ gilt:
$B_n = \underbrace{(B_n \cap A)}_{=:C_n \in \mathcal{S}} \dot{\cup} \underbrace{B_n \cap A^c)}_{=:D_n \in \mathcal{S}}$.

$\Longrightarrow$ Es existieren $(C_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ und $(D_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ mit $B_n = C_n \dot{\cup} D_n$ für $n \in \mathbb{N}$.

Ist $\mathcal{S}$ abgeschlossen unter endlichen Vereinigungen (Voraussetzung (***)), so gilt:

$\mu(B_n) = \mu(C_n) + \mu(D_n)$.

Ferner gilt:

$\bigcup\limits_{n\in\mathbb{N}} C_n \supseteq B \cap A$ und $\bigcup\limits_{n\in\mathbb{N}} D_n \supseteq B \cap A^c$.

Dies bedeutet: Für jede Überdeckung $\bigcup\limits_{n\in\mathbb{N}} B_n$ von B mit $B_n \in \mathcal{S}$ für $n \in \mathbb{N}$
existiert eine Überdeckung $\bigcup\limits_{n\in\mathbb{N}} C_n$ von $B \cap A$ und eine Überdeckung $\bigcup\limits_{n\in\mathbb{N}} D_n$ von $B \cap A^c$ jeweils mit Mengen aus $\mathcal{S}$

$$\text{mit: } \sum_{n\in\mathbb{N}} \mu(B_n) = \sum_{n\in\mathbb{N}} \mu(C_n) + \sum_{n\in\mathbb{N}} \mu(D_n)$$

Daraus folgt: $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$

Damit gilt unter den gemachten Voraussetzungen für alle $A \in \mathcal{S}$ und erstaunlicherweise für beliebige $B \in 2^\Omega$ die Gleichung (Z)

Obiges Ergebnis könnte man jetzt so interpretieren, dass bezüglich additiver Zerlegung unter den gemachten Voraussetzungen die äußere Maßkonstruktion "nichts kaputt macht". Weiterhin kann man sich vorstellen, dass über die vorher in $\mathcal{S}$ enthaltenen Mengen hinaus weitere Mengen existieren, die bezüglich additiver Zerlegung unproblematisch sind, und die man deshalb ebenfalls als messbar deklarieren könnte, was zu folgender Definition führt:

Definition 6 (System aller additiv zerlegenden Mengen). Sei $\mathcal{S}$ ein Mengensystem, das abgeschlossen unter endlichen Schnitten und Mengendifferenzen ist und sei $\mu : \mathcal{S} \longrightarrow [0,\infty]$. Wir sagen, dass eine Menge $A \in \mathcal{S}$ eine Menge $B \in \mathcal{S}$ additiv zerlegt, falls

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

Mit

$$\mathcal{A}_\mu := \{A \in \mathcal{S} \mid \forall B \in \mathcal{S} : \mu(B) = \mu(B \cap A) + \mu(B \cap A^c)\}$$

bezeichnen wir die Menge aller $A \in \mathcal{S}$, die alle $B \in \mathcal{S}$ additiv zerlegen.

Bemerkung 2.16 *Für isotones und σ-subadditives μ mit $\mu(\emptyset) = 0$ enthält $\mathcal{A}_\mu$ bereits alle μ-Nullmengen.*

Satz 2.17 (System aller additiv zerlegenden Mengen ist σ-Algebra) *Ist μ^* ein äußeres Maß, so ist die Menge $\mathcal{A}_{\mu^*}$ eine σ-Algebra. Diese bezeichnen wir als Carathéodory-σ-Algebra (zu μ^*). Außerdem ist in diesem Fall $\mu^*_{|\mathcal{A}_{\mu^*}}$ ein Maß.*

Zum Beweis: Vergleiche Übung 4, Aufgabe 1.

Bemerkung 2.18 *Ausgehend von irgendeinem μ mit $\mu(\emptyset) = 0$ bekommen wir also immer ein Maß $\mu^*_{|\mathcal{A}_{\mu^*}}$. Es ist jedoch nicht unmittelbar klar, was dieses Maß mit dem urspünglichen μ zu tun hat. Insbesondere stellen sich die Fragen:*

- $\mathcal{S} \subseteq \mathcal{A}^*_\mu$ *? (Darüber wissen wir schon ein bisschen etwas, siehe Vorüberlegungen)*
- $\mu^*_{|\mathcal{S}} = \mu$ *? (Das gilt z.B., wenn μ σ-subadditiv ist, vgl. Hausübung 3, Aufgabe 1)*

Wenn wir uns jetzt die gemachten Voraussetzungen in unseren Vorüberlegungen vergegenwärtigen, dann stellen wir fest, dass wir z.B. in $\mathbb{R}^1$, wo wir μ zunächst auf Intervallen (oder auch auf disjunkten abzählbaren Vereinigungen von offenen Intervallen) definieren könnten, insbesondere die Abgeschlossenheit unter Mengendifferenzen nicht gegeben ist, d.h., wir sind hier noch nicht am Ziel. Um zu formulieren, unter welchen Voraussetzungen wir μ zunächst so fortsetzen können, dass unsere Voraussetzungen der Vorüberlegung erfüllt sind, folgen zunächst einige Definitionen.

Definition 7 (Ring, Halbring, Inhalt, Prämaß). Ein Mengensystem $\mathcal{S} \subseteq 2^{\Omega}$ heißt Ring (im Sinne der Maßtheorie), falls:

$$\begin{aligned} R1 &: \emptyset \in \mathcal{S} \\ R2 &: A,B \in \mathcal{S} \Longrightarrow A \cup B \in \mathcal{S} \\ R3 &: A,B \in \mathcal{S} \Longrightarrow A \backslash B \in \mathcal{S} \end{aligned}$$

Ein Mengensystem $\mathcal{S} \subseteq 2^{\Omega}$ heißt Halbring oder Semiring, falls:

$HR1: \emptyset \in \mathcal{S}$

$HR2: A,B \in \mathcal{S} \Longrightarrow A \cap B \in \mathcal{S}$

$HR3: A,B \in \mathcal{S} \Longrightarrow$ es existieren $C_1, \ldots, C_n \in \mathcal{S}$ disjunkt, mit $A \backslash B = \dot{\bigcup\limits_{k \in \{1,\ldots,n\}}} C_k$

Eine nichtnegative Mengenfunktion $\mu : \mathcal{S} \longrightarrow [0,\infty]$ auf einem Halbring heißt Prämaß, falls sie σ-additiv ist und $\mu(\emptyset) = 0$ gilt. Ist $\mu(\emptyset) = 0$ und ist μ nur endlich additiv, d.h., gilt für beliebige paarweise disjunkte $A_1, \ldots, A_n$ mit $\bigcup\limits_{k=1}^{n} A_k \in \mathcal{S}$ immer die Gleicheit $\mu\left(\bigcup\limits_{k=1}^{n} A_k\right) = \sum\limits_{k=1}^{n} \mu(A_k)$, dann heißt μ ein Inhalt.

Bemerkung 2.19 *Für einen Ring $\mathcal{S}$ gilt immer schon $A,B \in \mathcal{S} \Longrightarrow A \cap B \in \mathcal{S}$, denn $A \cap B = A \backslash (A \backslash B)$.*

Beispiel 2.20 *Für $\Omega = \mathbb{R}^d$ ist $\mathfrak{I}^d = \{(a,b] \mid a,b \in \mathbb{R}^d, a \leq b\}$ ein Halbring und*

$$\lambda^d : \mathfrak{I}^d \longrightarrow [0,\infty] : (a,b] \mapsto \prod_{i=1}^{d} (b_i - a_i)$$

ein Prämaß. Dieses Prämaß wird als das Lebesguesche Prämaß auf $\mathfrak{I}^d$ bezeichnet.

Satz 2.21 (Fortsetzung eines Inhalts auf einen Ring)
Sei $\mu : \mathcal{S} \longrightarrow [0,\infty]$ ein Inhalt auf einem Halbring $\mathcal{S}$. Dann gibt es genau eine Fortsetzung $v : \mathfrak{R}(\mathcal{S}) \longrightarrow [0,\infty]$ zu einem Inhalt auf den von $\mathcal{S}$ erzeugten Ring

$$\begin{aligned}\mathfrak{R}(\mathcal{S}) &= \bigcap \{R \text{ Ring auf } \Omega \mid R \supseteq \mathcal{S}\} \\ &= \left\{ \bigcup_{k=1}^{n} A_k \mid n \in \mathbb{N}, A_1, \ldots, A_n \in \mathcal{S} \text{ paarweise disjunkt} \right\}.\end{aligned}$$

Insbesondere gilt für $A = \dot{\bigcup\limits_{k \in \{1,\ldots,n\}}} A_k$ *mit paarweise disjunkten* $A_1, \ldots, A_n \in \mathcal{S}$ *dann*

$$\nu(A) = \sum_{k=1}^{n} \mu(A_k).$$

Satz 2.22 (Charakterisierung endlicher Inhalte und Prämaße auf $\mathfrak{I}^1$)

Gegeben seien $\Omega = \mathbb{R}$ *und der Halbring* $\mathfrak{I}^1 = \{(a,b] \mid a,b \in \mathbb{R}, a \leq b\}$.

i) Ist $F : \mathbb{R} \longrightarrow \mathbb{R}$ *eine isotone Funktion, d.h., gilt* $x \leq y \Longrightarrow F(x) \leq F(y)$, *so ist*

$$\mu_F : \mathfrak{I}^1 \longrightarrow \mathbb{R} : (a,b] \mapsto F(b) - F(a)$$

ein endlicher Inhalt. Dieser Inhalt wird auch als der zu F *gehörige Stieltjessche Inhalt bezeichnet.*

ii) Für zwei isotone Funktionen $F, G : \mathbb{R} \longrightarrow \mathbb{R}$ *gilt genau dann* $\mu_F = \mu_G$, *wenn* $F - G$ *eine konstante Funktion ist.*

iii) Ist $\mu : \mathfrak{I}^1 \longrightarrow \mathbb{R}$ *ein endlicher Inhalt, dann ist z.B. die Funktion* $F : \mathbb{R} \longrightarrow \mathbb{R}$ *definiert durch*

$$F(x) := \begin{cases} \mu((0,x]) & \text{falls } x \geq 0 \\ -\mu((x,0]) & \text{falls } x < 0 \end{cases}$$

isoton und es gilt $\mu = \mu_F$.

iv) Ist $F : \mathbb{R} \longrightarrow \mathbb{R}$ *isoton, dann ist der zugehörige Stieltjessche Inhalt* μ_F *genau dann ein Prämaß, wenn* F *rechtsseitig stetig ist.*

Bemerkung 2.23 *Ähnliche Überlegungn sind auch für* $\mathbb{R}^d$ *mit* $d > 1$ *möglich, allerdings muss dann die Eigenschaft der Isotonie der Funktion* F *zur stärkeren Eigenschaft der „vollständigen Isotonie“ von* F *verschärft werden, vergleiche z.B. [Schmidt, 2009, S. 293ff] (da wird die Eigenschaft als „rechtecksmonoton“ bezeichent) oder [Elstrodt, 1996, S. 45ff] (da wird die Eigenschaft einfach als „(monoton) wachsend“ bezeichnet).*

Damit sind wir jetzt endlich in der Lage, den wesentlichen Satz dieses Abschnitts zur Konstruktion von Maßen zu formulieren:

Satz 2.24 (Fortsetzungssatz von Carathéodory) *Es sei $\mu : \mathcal{S} \longrightarrow [0,\infty]$ ein Inhalt auf einem Halbring $\mathcal{S} \subseteq 2^{\Omega}$. Dann gilt für das assoziierte Maß*

$$\mu^* : 2^{\Omega} \longrightarrow [0,\infty] : A \mapsto \inf\left\{ \sum_{n\in\mathbb{N}} \mu(B_n) \mid (B_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}, \bigcup_{n\in\mathbb{N}} B_n \supseteq A \right\}:$$

i) μ^* ist ein äußeres Maß.
ii) Alle Mengen aus $\mathcal{S}$ sind $\mathcal{A}_{\mu^*}$-messbar.
iii) Ist μ ein Prämaß, so gilt ${\mu^*}_{|\mathcal{S}} = \mu$. Insbesondere ist dann ${\mu^*}_{|\mathcal{A}_{\mu^*}}$ eine Fortsetzung von μ zu einem Maß auf einer σ-Algebra, die $\mathcal{S}$ (und damit auch $\sigma(\mathcal{S})$) umfasst.
iv) Ist μ kein Prämaß, so gibt es ein $A \in \mathcal{S}$ mit $\mu^*(A) < \mu(A)$.
v) Ist μ ein σ-endliches[12] Prämaß, dann ist $\mu^*_{|A_{\mu^*}}$ die einzige Fortsetzung von μ zu einem Maß auf $\mathcal{A}_{\mu^*}$, d.h., die Fortsetzung ist in diesem Fall eindeutig.
vi) Das Maß $\eta := {\mu^*}_{|\mathcal{A}_{\mu^*}}$ ist ein vollständiges Maß auf $\mathcal{A}_{\mu^*}$, d.h., jede Teilmenge einer η-Nullmenge ist η-messbar (i.e.: $A \in \mathcal{A}_{\mu^*}, \eta(A) = 0, B \subseteq A \Longrightarrow B \in \mathcal{A}_{\mu^*}$).

Beispiel 2.25 (Lebesgue-Maß und Borel-Maß auf $\mathbb{R}^d$) *Wendet man den Fortsetzungssatz von Carathéodory auf das Lebesguesche Prämaß λ^d an, so kommt man zum äußeren Maß $(\lambda^d)^* : 2^{\Omega} \longrightarrow [0,\infty]$ gegeben durch*

$$(\lambda^d)^*(A) = \inf\left\{ \sum_{n\in\mathbb{N}} \lambda^d(I_n) \mid \forall n \in \mathbb{N} : I_n \in \mathfrak{I}^d, \bigcup_{n\in\mathbb{N}} I_n \supseteq A \right\}.$$

Das System aller $\mathcal{A}_{(\lambda^d)^*}$-messbaren Mengen ist eine σ-Algebra, die wir als die Lebesguesche σ-Algebra bezeichnen und mit $\mathfrak{L}^d$ notieren. Die Lebesguesche σ-Algebra entält alle Intervalle aus $\mathfrak{I}^d$ und ist damit eine Obermenge der Borelschen σ-Algebra $\mathfrak{B}(\mathbb{R}^d)$. Achtung: Im Fall von $\mathbb{R}^d$ ist die Lebesguesche σ-Algebra echt größer als die Borelsche σ-Algebra. Die

[12] Definition analog zur σ-Endlichkeit eines Maßes.

Einschränkung von $(\lambda^d)^*$ auf die Lebesguesche-σ-Algebra bezeichnen wir als das Lebesgue-Maß und notieren es ebenfalls (mit einem bisschen abuse of notation) als λ^d. Die Einschränkung von $(\lambda^d)^*$ auf die Borelsche σ-Algebra $\mathfrak{B}(\mathbb{R}^d)$ bezeichnen wir als das Borel-Maß bzw. auch als das Lebesgue-Borel-Maß. Da sowohl das Borel-Maß, als auch das Lebesgue-Maß nach Konstruktion normiert und translationsinvariant sind (vergleiche Übung 3, Aufgabe 3, beachte aber, dass wir die Translationsabgeschlossenheit von $\mathfrak{B}(\mathbb{R}^d)$ bzw. $\mathfrak{L}^d$ noch nicht gezeigt haben), gilt insgesamt

$$\mathfrak{B}(\mathbb{R}^d) \subsetneq \mathfrak{L}^d \subsetneq 2^{\mathbb{R}^d}.$$

Abschließend beantworten wir noch die Frage, wie sich $\mathfrak{B}(\mathbb{R}^d)$ und $\mathfrak{L}^d$ bzw. das Borel-Maß und das Lebesgue-Maß zueinander verhalten: Das Lebesgue-Maß entsteht aus dem Borel-Maß durch Vervollständigung:

Definition 8 (Vollständiger Maßraum). Ein Maßraum $(\Omega, \mathcal{A}, \mu)$ heißt vollständig, wenn jede Teilmenge einer μ-Nullmenge $A \in \mathcal{A}$ ebenfalls zu $\mathcal{A}$ gehört (und somit selbst eine μ-Nullmenge ist). In diesem Fall sagen wir auch, dass das Maß μ vollständig ist.

Satz 2.26 (Vervollständigung eines Maßes) *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $\mathfrak{N} := \{B \in 2^\Omega \mid \exists A \in \mathcal{A} : A \supseteq B \ \& \ \mu(A) = 0\}$ die Menge aller Teilmengen von μ-Nullmengen.*
Definiere

$$\begin{aligned} \mathcal{A}' &:= \{A \cup N \mid A \in \mathcal{A}, N \in \mathfrak{N}\} \\ \mu' &: \mathcal{A}' \longrightarrow [0,\infty] : A \cup N \mapsto \mu(A) \text{ für } A \in \mathcal{A} \text{ und } N \in \mathfrak{N}. \end{aligned}$$

Dann gilt:

i) Das Mengensystem $\mathcal{A}'$ ist eine σ-Algebra, μ' ist wohldefiniert und $(\Omega, \mathcal{A}', \mu')$ ist ein vollständiger Maßraum. Außerdem ist $\tilde{\mu}$ die einzige Fortsetzung von μ zu einem Inhalt auf $\mathcal{A}'$.
ii) Jede vollständige Fortsetzung von μ ist eine Fortsetzung von μ'.

Satz 2.27 (Carathéodory-Fortsetzung ist Vervollständigung) *Sei μ ein σ-endliches Prämaß auf einem Halbring $\mathcal{S}$ und μ^* das zugehörige äußere Maß. Dann ist $\mu^*_{|\mathcal{A}_{\mu^*}}$ die Vervollständigung von $\mu^*_{|\sigma(\mathcal{S})}$.*

Korollar 2.28 *Das Lebesgue-Maß ist die Vervollständigung des Lebesgue-Borel-Maßes.*

Bemerkung 2.29 *Da es Lebesgue-messbare Mengen $A \subseteq \mathbb{R}^d$ gibt, die gleichmächtig zu $\mathbb{R}$ sind, aber trotzdem Lebesgue-Maß Null besitzen, folgt $\mathfrak{L}^d \cong 2^{\mathbb{R}}$, denn mit A sind auch alle Teilmengen $B \in 2^A$ Lebesgue-messbar. Außerdem gilt $\mathfrak{B}(\mathbb{R}^d) \cong \mathbb{R}$.*

Satz 2.30 (Charakterisierung der Lebesgue-Messbarkeit) *Eine Menge $A \subseteq \mathbb{R}^d$ ist genau dann Lebesgue-messbar, wenn für jedes $\varepsilon > 0$ eine offene Menge O und eine abgeschlossene Menge C existiert mit $C \subseteq A \subseteq O$ und $\lambda^d(O \backslash C) < \varepsilon$.*

Bemerkung 2.31 *In obigem Satz ist wichtig, dass O offen ist und dass C abgeschlossen ist. Insbesondere kann der Rand einer Menge echt positives Lebesgue-Maß besitzen (vgl. Übung 4, Aufgabe 3). Lässt man offen und abgeschlossen weg und fordert man von O und C nur die Lebesgue-Messbarkeit, dann wird der Satz trivial. Fordert man demgegenüber beispielsweise, dass O nicht offen, sondern abgeschlossen ist, dann wird der Satz falsch, wie Sie in Übung 4, Aufgabe 2 sehen werden.*

Korollar 2.32 *Ist $F : \mathbb{R} \longrightarrow \mathbb{R}$ isoton und rechtsseitig stetig, so gibt es ganu eine Fortsetzung des Lebesgue-Stieltjesschen Prämaßes μ_F zu einem Maß auf der σ-Algebra $\mathcal{A}_{\mu_F^*}$. Diese wird als das Lebesgue-Stieltjessche Maß λ_F zu F bezeichnet und ist die Vervollständigung von $\lambda_{F|\mathfrak{B}(\mathbb{R})}$.*

Von der Fortsetzung eines σ-endlichen Prämaßes von einem Halbring auf die Carathédory-σ-Algebra wissen wir, dass diese eindeutig bestimmt ist. Im Abschnitt zur Integrationstheorie werden wir jedoch hauptsächlich mit der wesentlich kleineren Borelschen σ-Algebra arbeiten, da diese "handlicher" ist (vergleiche die Bemerkung zu Korollar 3.3) und für die Integration reellwertiger Funktionen grundsätzlich völlig ausreicht. Deshalb bleibt noch die Frage der Eindeutigkeit der Fortsetzung auf die Borelsche σ-Algebra zu beantworten:

Satz 2.33 (Maßeindeutigkeitssatz) *Sei $\mathcal{E} \subseteq 2^\Omega$ und μ sowie ν zwei Maße, die auf der von $\mathcal{E}$ erzeugten σ-Algebra definiert sind. Ist $\mathcal{E}$ abgeschlossen unter endlichen Schnitten und gibt es eine Folge $(E_n)_{n\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ mit $\mu(E_n) = \nu(E_n) < \infty$ für alle $n \in \mathbb{N}$ und $\Omega = \bigcup\limits_{n\in\mathbb{N}} E_n$, dann gilt:*

Stimmen μ und ν auf $\mathcal{E}$ überein, so sind μ und ν bereits identisch.

Korollar 2.34 *Ist $F : \mathbb{R} \longrightarrow \mathbb{R}$ eine isotone, rechtsseitig stetige Funktion mit $\lim_{x\to\infty} F(x) = 1$ und $\lim_{x\to-\infty} F(x) = 0$, so ist das zugehörige Prmaß μ_F eindeutig zu einem Wahrscheinlichkeitsmaß auf $\mathfrak{B}(\mathbb{R})$ bzw. auch auf $\mathfrak{L}^1$ fortsetzbar. In diesem Fall nennen wir F eine Verteilungsfunktion. Außerdem wird jedes Wahrscheinlichkeitsmaß μ auf $\mathfrak{B}(\mathbb{R})$ bzw. auch auf $\mathfrak{L}^1$ eindeutig[13] durch die zugehörige Verteilungsfunktion*

$$F_\mu : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto \mu(]-\infty, x])$$

induziert.

Bemerkung 2.35 *Eine Verteilungsfunktion $F : \mathbb{R} \longrightarrow \mathbb{R}$ ist zerlegbar in 3 Teile:*

$$\begin{aligned} F &= F_{diskret} + F_{stetig} \\ &= F_{diskret} + F_{absolutstetig} + F_{stetigsingulär}. \end{aligned}$$

Dabei ist $F_{diskret}$ isoton und konstant bis auf höchstens abzählbar viele Sprungstellen, konkret den Stellen $x \in \mathbb{R}$ mit $\lim_{x' \nearrow x} F(x') < F(x)$. (Die Höhe der Sprungstellen von $F_{diskret}$ beträgt dann $F(x) - \lim_{x' \nearrow x} F(x')$.)

F_{stetig} ist isoton und stetig und kann weiter in die zwei Teile $F_{absolutstetig}$ und $F_{stetigsingulär}$ zerlegt werden. Dabei besitzt $F_{absolutstetig}$ eine Integraldarstellung als

$$F_{absolutstetig}(x) = \int_{-\infty}^{x} f(\omega) d\lambda(\omega),$$

wobei f eine reellwertige, nichtnegative Lebesgue-integrierbare Funktion und das rechte Integral das Lebesgue-Integral (bezüglich des Lebesguemaßes) ist (siehe später). Der Teil $F_{stetigsingulär}$ ist stetig und isoton, aber im Allgemeinen schwierig zu beschreiben (vgl. auch Übung 5, Aufgabe 2).

Außerdem sind alle 3 Teile singulär zueinander, d.h., es gibt 3 μ_F-messbare Mengen $S_{diskret}, S_{absolutstetig}$ und $S_{stetigsingulär}$, die Ω in 3 Teile zerlegen, so dass für die den 3 Teilen zugeordneten Maße gilt:

[13] Mit eindeutig ist hier gemeint, dass es keine weitere Verteilungsfunktion gibt, die ebenfalls das Maß μ induziert.

$$\mu_{F_{diskret}}(\Omega \setminus S_{diskret}) = 0$$
$$\mu_{F_{absolutstetig}}\left(\Omega \setminus S_{absolutstetig}\right) = 0$$
$$\mu_{F_{stetigsingulär}}\left(\Omega \setminus S_{stetigsingulär}\right) = 0.$$

Abschließende Frage: Tangiert das Vitalische Unmöglichkeitsresultat auch spezifisch die Wahrscheinlichkeitstheorie und nicht nur die Maßtheorie?

Antwort: Ja, siehe [Georgii, 2015, S.9 f]:

Sei $\Omega = \{0,1\}^{\mathbb{N}}$ der Ergebnisraum des unendlich oft durchgeführten Münzwurfs. Dann gibt es keine Abbildung $P : 2^{\Omega} \longrightarrow [0,\infty]$ mit den Eigenschaften:

$$\begin{array}{rl} \text{Normierung:} & P(\Omega) = 1 \\ \sigma\text{-Additivität:} & \forall (A_n)_{n\in\mathbb{N}} \in \left(2^{\Omega}\right)^{\mathbb{N}} \text{ paarweise disjunkt: } P\left(\dot{\bigcup_{n\in\mathbb{N}}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n) \\ \text{Invarianz:} & \forall A \subseteq \Omega \forall n \in \mathbb{N}: \; P(T_n(A)) = P(A). \end{array}$$

Hier ist $T_n(\omega_1, \omega_2, \ldots, \omega_n, \ldots) = (\omega_1, \omega_2, \ldots, \omega_{n-1}, 1-\omega_n, \omega_{n+1}, \ldots)$ die Abbildung von Ω auf Ω, die das Ergebnis des n-ten Wurfes umdreht, und $T_n(A) := \{T_n(\omega) \mid \omega \in A\}$. Die Invarianz unter den Transformationen T_n drückt hier die Fairness der Münze und die Unabhängigkeit der Würfe aus.

Insgesamt bedeutet dies, dass es bereits bei der Modellierung des unendlich oft wiederholten unabhängigen Wurfes einer fairen Münze nicht mehr möglich ist, allen beliebigen Teilmengen des Ergebnisraumes in sinnvoller Weise eine Wahrscheinlichkeit zuzuordnen. (Natürlich nur, sofern man auf die σ-Additivität und das Auswahlaxiom besteht.)
Den Beweis des Satzes finden Sie in [Georgii, 2015, S.9 f]. Der Beweis verläuft in etwa "isomorph" zum oben geführten Beweis des Satzes von Vitali, man stelle sich eine Folge $\omega = (\omega_1, \omega_2, \ldots)$ als Binärdarstellung einer reellen Zahl $x \in [0,1]$ im Beweis des Satzes von Vitali vor. Die darin eingeführte Äquivalenzrelation $\sim = \{(x,y) \in [0,1] \mid x-y \in \mathbb{Q}\}$ würde sich dann übersetzen in $\sim = \{(\omega, \omega') \in \Omega \times \Omega \mid \exists n_0 \in \mathbb{N} : n \geq n_0 \Longrightarrow \omega_n = \omega'_n\}$ und man könnte wieder mit Hilfe des Auswahlaxioms einen Widerspruch konstruieren...

3 Messbare Abbildungen und Bildmaße

Definition 9 (Messbare Abbildung, Zufallsvariable). Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ zwei Messräume. Eine Abbildung $f : \Omega \longrightarrow \Omega'$ heißt $\mathcal{A}-\mathcal{A}'$-messbar, wenn jede $\mathcal{A}'$-messbare Menge B ein $\mathcal{A}$-messbares Urbild $f^{-1}(B)$ besitzt:

$$\forall B \in \mathcal{A}' : f^{-1}(B) \in \mathcal{A}.$$

Sind die σ-Algebren $\mathcal{A}$ und $\mathcal{A}'$ aus dem Kontext heraus klar, so nennen wir f kurz messbar. Eine messbare Abbildung wird im Kontext der Stochastik auch als Zufallsvariable bezeichnet und typischerweise mit X, Y oder Z notiert.

Definition 10 (Bildmaß). Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $(\Omega', \mathcal{A}')$ ein Messraum, sowie $X : \Omega \longrightarrow \Omega'$ eine $\mathcal{A}-\mathcal{A}'$-messbare Abbildung. Dann ist auf der σ-Algebra $\mathcal{A}'$ das sogenannte Bildmaß von μ unter X gegeben durch

$$\mu_X : \mathcal{A}' \longrightarrow [0, \infty] : B \mapsto \mu(X^{-1}(B)).$$

Ist $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum (, d.h., μ ist ein Wahrscheinlichkeitsmaß), dann nennen wir im Fall $\Omega' = \mathbb{R}$ die zum Bildmaß μ_X gehörige Verteilungsfunktion auch die Verteilungsfunktion von X und notieren sie mit F_X.

Bemerkung 3.1 *Das Bildmaß μ_X ist tatsächlich ein Maß auf dem Maßraum $(\Omega', \mathcal{A}')$. Der Nachweis folgt einfach durch Übertragung der Maßeigenschaften von μ auf μ_X unter Ausnutzung der Operationstreue des Urbilds (vergleiche Übung 1, Aufgabe 2). Die Messbarkeit von X ist erforderlich für die Wohldefiniertheit von μ_X.*

Satz 3.2 (Charakterisierung der Messbarkeit) *Sei $X : \Omega \longrightarrow \Omega'$ eine Abbildung und $\mathcal{A}'$ eine σ-Algebra auf Ω', die von einem Mengensystem $\mathcal{E} \subseteq 2^{\Omega'}$ erzeugt wird (, d.h. $\sigma(\mathcal{E}) = \mathcal{A}'$). Sei*

$$X^{-1}[\mathcal{E}] := \{X^{-1}(B) \mid B \in \mathcal{E}\}.$$

Dann gilt:

i) *Jede σ-Algebra $\mathcal{A}$ auf Ω, für die X eine $\mathcal{A} - \mathcal{A}'$-messbare Abbildung ist, muss mindestens $X^{-1}[\mathcal{E}]$ und damit auch $\sigma(X^{-1}[\mathcal{E}])$ umfassen.*

ii) *Die kleinste σ-Algebra auf Ω, für die X messbar ist, ist genau $\sigma(X^{-1}[\mathcal{E}])$, denn X ist $\sigma(X^{-1}[\mathcal{E}]) - \mathcal{A}'$-messbar genau dann, wenn*

$$\sigma(\mathcal{E}) \subseteq \underbrace{\{B \subseteq \Omega' \mid X^{-1}(B) \in \sigma(X^{-1}[\mathcal{E}])\}}_{\text{Menge aller } B \text{ mit } \sigma(X^{-1}[\mathcal{E}])\text{-messbarem Urbild}},$$

was äquivalent zur Tautologie

$$\mathcal{E} \subseteq \{B \subseteq \Omega' \mid X^{-1}(B) \in \underbrace{\sigma(X^{-1}[\mathcal{E}])}_{\supseteq X^{-1}[\mathcal{E}]}\}$$

ist, denn $\{B \subseteq \Omega' \mid X^{-1}(B) \in \sigma(X^{-1}[\mathcal{E}])\}$ ist eine σ-Algebra (vergleiche Übung 2, Aufgabe 3).

iii) *Dies bedeutet für eine σ-Algebra $\mathcal{A}$ auf Ω, dass X eine $\mathcal{A} - \mathcal{A}'$-messbare Abbildung ist genau dann, wenn $\mathcal{A} \supseteq \sigma(X^{-1}[\mathcal{E}])$ bzw. $\mathcal{A} \supseteq X^{-1}[\mathcal{E}]$ gilt, also <u>wenn alle Urbilder des Erzeugers $\mathcal{E}$ messbar bezüglich $\mathcal{A}$ sind.</u>*

iv) *Darüber hinaus gilt*

$$\sigma(X^{-1}[\mathcal{E}]) = X^{-1}[\sigma(\mathcal{E})].$$

(Begründung: $\subseteq$ gilt , weil $X^{-1}[\sigma(\mathcal{E})]$ eine σ-Algebra ist (vgl. Übung 2, Aufgabe 3) und $\supseteq$ folgt aus ii), denn links steht die kleinste σ-Algebra auf Ω, bezüglich derer X messbar ist, also insbesondere eine σ-Algebra auf Ω, bezüglich derer X messbar ist.)

Korollar 3.3 *Jede stetige Abbildung* $X : \mathbb{R}^d \longrightarrow \mathbb{R}^q$ *ist eine* $\mathfrak{B}(R^d) - \mathfrak{B}(\mathbb{R}^q)$*-messbare Abbildung, denn* $\mathfrak{B}(\mathbb{R}^q)$ *wird von den offenen Mengen (in* $\mathbb{R}^q$*) erzeugt und jede offene Menge (in* $\mathbb{R}^q$*) hat unter einer stetigen Abbildung ein (in* $\mathbb{R}^d$*) offenes und somit* $\mathfrak{B}(\mathbb{R}^d)$*-messbares Urbild (vgl. Übung 0, Aufgabe 4).*

Achtung: Hier geht es nur um Messbarkeit bezüglich der Borelschen σ*-Algebren. Demgegenüber ist zum Beispiel die stetige Abbildung*

$$X : \mathbb{R} \longrightarrow \mathbb{R}^2 : x \mapsto \begin{pmatrix} x \\ 0 \end{pmatrix}$$

nicht $\mathfrak{L}^1 - \mathfrak{L}^2$*-messbar: Beispielsweise ist* $V \times \{0\}$ *mit* V *der Vitali-Menge eine Menge mit zweidimensionalem Lebesgue-Maß Null, denn* $V \times \{0\}$ *ist eine Teilmenge der Menge* $[-1,2] \times \{0\}$*, die offensichtlich Lebesgue-Maß Null besitzt. Insbesondere ist also* $V \times \{0\}$ *messbar bezüglich* $\mathfrak{L}^2$*. Das Urbild von* $V \times \{0\}$ *ist aber die Vitali-Menge selbst, also nicht* $\mathfrak{L}^1$*-messbar.*

Satz 3.4 (Komposition messbarer Abbildungen ist messbar) *Seien* $(\Omega, \mathcal{A})$*,* $(\Omega', \mathcal{A}')$ *und* $(\Omega'', \mathcal{A}'')$ *Messräume und* $X : \Omega \longrightarrow \Omega'$ *sowie* $Y : \Omega' \longrightarrow \Omega''$ *zwei Abbildungen. Ist* X *eine* $\mathcal{A} - \mathcal{A}'$ *messbare Abbildung und ist* Y *eine* $\mathcal{A}' - \mathcal{A}''$*-messbare Abbildung, so ist die Komposition*

$$Y \circ X : \Omega \longrightarrow \Omega'' : \omega \mapsto (Y(X(\omega)))$$

eine $A - \mathcal{A}''$*-messbare Abbildung.*

Beweis (Durchmeditieren). Für $B \in \mathcal{A}''$ ist $Y^{-1}(B) \in \mathcal{A}'$, da Y messbar und weiterhin ist $X^{-1}(Y^{-1}(B)) \in \mathcal{A}$, weil X messbar. Da $(Y \circ X)^{-1}(B) = \{\omega \in \Omega \mid Y(X(\omega)) \in B\} = \{\omega \in \Omega \mid X(\omega) \in Y^{-1}(B)\} = X^{-1}(Y^{-1}(B))$ gilt, haben wir damit die $\mathcal{A} - \mathcal{A}''$-Messbarkeit von $Y \circ X$ gezeigt.

Satz 3.5 (Charakterisierung der $\mathfrak{B}(\mathbb{R}^d)$-Messbarkeit) *Sei* $(\Omega, \mathcal{A})$ *ein Messraum und*

$$X : \Omega \longrightarrow \mathbb{R}^d : \omega \mapsto \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_p(\omega) \end{pmatrix}.$$

Dann ist X *eine* $\mathcal{A} - \mathfrak{B}(\mathbb{R}^d)$*-messbare Abbildung genau dann, wenn jede der Abbildungen* $X_1, \ldots, X_p$ *eine* $\mathcal{A} - \mathfrak{B}(\mathbb{R})$*-messbare Abbildung ist.*

Beweis. "$\Longrightarrow$": Betrachte für $i = 1, \ldots, p$ die Projektion

$$\pi_i : \mathbb{R}^d \longrightarrow \mathbb{R} : \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_p \end{pmatrix} \mapsto \omega_i$$

auf die i-te Komponente. Diese ist stetig und deshalb $\mathfrak{B}(\mathbb{R}^d) - \mathfrak{B}(\mathbb{R})$-messbar. Damit ist $X_i = \pi_i \circ X$ als Komposition der $\mathcal{A} - \mathfrak{B}(\mathbb{R}^d)$-messbaren Abbildung X und der $\mathfrak{B}(\mathbb{R}^d) - \mathfrak{B}(\mathbb{R})$-messbaren Abbildung π_i eine $\mathcal{A} - \mathfrak{B}(\mathbb{R})$-messbare Abbildung.

"$\Longleftarrow$": Das Mengensystem $\mathcal{F} = \{(-\infty, c] \mid c \in \mathbb{R}^d\}$ ist ein Erzeuger von $\mathfrak{B}(\mathbb{R}^d)$. Für beliebiges $B = (-\infty, c] \in \mathcal{F}$ gilt:

$$\begin{aligned} X^{-1}(B) &= \{\omega \in \Omega \mid X(\omega) \in (-\infty, c]\} \\ &= \{\omega \in \Omega \mid X_1(\omega) \in (-\infty, c_1], \ldots, X_p(\omega) \in (-\infty, c_p]\} \\ &= \bigcap_{i=1}^{p} \underbrace{X_i^{-1}((-\infty, c_i])}_{\in \mathcal{A}, \text{ da } X_i \text{ messbar}} \in \mathcal{A}. \end{aligned}$$

Damit ist X eine $\mathcal{A} - \mathfrak{B}(\mathbb{R}^d)$-messbare Funktion.

Satz 3.6 (Borel-Messbarkeit von zusammengesetzten Funktionen) *Sei $(\Omega, \mathcal{A})$ ein Messraum und seien $X, Y : \Omega \longrightarrow \mathbb{R}$ zwei $\mathcal{A} - \mathfrak{B}(\mathbb{R})$-messbare Funktionen und sei $c \in \mathbb{R}$. Dann sind auch*

$$X + Y, \qquad X - Y, \qquad X \cdot Y \qquad \textit{und} \qquad c \cdot X$$

$\mathcal{A} - \mathfrak{B}(\mathbb{R})$-messbar. (Alle Operationen sind hier punktweise zu verstehen, d.h., z.B. $X + Y$ ist definiert als $X + Y : \Omega \longrightarrow \mathbb{R} : \omega \mapsto X(\omega) + Y(\omega)$.) Ist $Y > 0$ (,d.h. $X(\omega) > 0$ für alle $\omega \in \Omega$) , so ist auch $\frac{X}{Y}$ eine $\mathcal{A} - \mathfrak{B}(\mathbb{R})$-messbare Funktion.

Beweis. Nur für $X + Y$, die anderen Aussagen können analog bewiesen werden. Definiere

$$Z : \Omega \longrightarrow \mathbb{R}^2 : \omega \mapsto \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix}.$$

Diese Abbildung ist $\mathcal{A}-\mathfrak{B}(\mathbb{R}^2)$-messbar, da jede Komponente X und Y jeweils $\mathcal{A}-\mathfrak{B}(\mathbb{R})$-messbar war. Definiere weiter

$$W : \mathbb{R}^2 \longrightarrow \mathbb{R} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto x+y.$$

Dann ist W stetig und somit $\mathfrak{B}(\mathbb{R}^2) - \mathfrak{B}(\mathbb{R})$-messbar. Daraus folgt, dass $X+Y = W \circ Z$ als Komposition zweier messbarer Abbildungen wieder $\mathcal{A}-\mathfrak{B}(\mathbb{R})$-messbar ist.

Kapitel 2
Integrationstheorie

Kommen wir nun zur Lebesgueschen Integrationstheorie. Ziel ist es unter anderem, für eine möglichst beliebige Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ einen Erwartungswert $\mathbb{E}_P(X)$ zu definieren. Um keine falschen Intuitionen zu wecken, werden wir zunächst nicht $\mathbb{E}_P(X)$, sondern $I_\mu(X)$ schreiben, wobei I für Integral steht und der Index μ bedeutet, dass wir bezüglich des Maßes μ, das kein Wahrscheinlichkeitsmaß zu sein braucht, integrieren. Im Vergleich zur Riemannschen Integrationstheorie gibt es hier wenigstens zwei wesentliche Unterschiede:

i) Bei der Riemann-Integration wird der (reelle) Definitionsbereich einer Funktion gedanklich in kleinere Intervalle zerlegt, um das Integral einer Funktion als “Fläche unter dem Funktionsgraphen” in einfachere “Teilflächen” zu zerlegen.
Bei der Lebesgueschen Integrationstheorie hingegen wird der Wertebereich einer rellwertigen Funktion gedanklich in Intervalle zerlegt, um die “Fläche unter dem Funktionsgraphen” in einfachere “Teilflächen” zu zerlegen.
ii) Da beim Lebesgueschen Zugang der Wertebereich gedanklich zerlegt wird, ist es möglich, auch Funktionen mit einem sehr abstrakten Definitionsbereich D, auf dem man nur ein Maß μ hat, zu integrieren, denn es ist lediglich nötig, für Intervalle I im Wertebereich von f (also reellwertige Intervalle, dessen Bemessung unproblematisch ist,) die “Größe” des Urbilds $f^{-1}(I)$ zu bemessen. (Man beachte auch die geschmeidigeren Struktureigenschaften des Urbilds $f^{-1}(I)$, das in der Lebesgueschen Integrationstheorie zum tragen kommt, gegenüber denen des Bildes $f(I)$, das in der Riemannschen Integrationstheorie von

wesentlicher Bedeutung ist.) Die Möglichkeit, auf einem abstrakten Maßraum $(\Omega, \mathcal{A}, \mu)$ integrieren zu können, ist ganz wesentlich für die Überzeugungskraft der Lebesgueschen Integrationstheorie im Kontext der Wahrscheinlichkeitstheorie.

Lebesgue äußert sich zum Verhältnis des Lebesgue-Integrals zum Riemann-Integral in einem Vortrag 1926 so (Übersetzung nach [Elstrodt, 1996, S.85]):

“Man kann auch sagen, daß man sich bei dem Vorgehen von Riemann verhält wie ein Kaufmann ohne System, der Geldstücke und Banknoten zählt in der zufälligen Reihenfolge, wie er sie in die Hand bekommt; während wir vorgehen wie ein umsichtiger Kaufmann, der sagt:

Ich habe $m(E_1)$ Münzen zu einer Krone, macht $1 \cdot m(E_1)$,
ich habe $m(E_2)$ Münzen zu zwei Kronen, macht $2 \cdot m(E_2)$,
ich habe $m(E_3)$ Münzen zu fünf Kronen, macht $5 \cdot m(E_3)$,
usw., ich habe also insgesamt $S = 1 \cdot m(E_1) + 2 \cdot m(E_2) + 5 \cdot m(E_3) + \ldots$

Die beiden Verfahren führen sicher den Kaufmann zum gleichen Resultat, weil er - wie reich er auch sei - nur eine endliche Anzahl von Banknoten zu zahlen hat; aber für uns, die wir unendlich viele Indivisiblen zu addieren haben, ist der Unterschied zwischen den beiden Vorgehensweisen wesentlich.”

Im Folgenden sei nun immer gegeben ein Maßraum $(\Omega, \mathcal{A}, \mu)$. Gesucht ist für eine hinreichend große Menge $D \subseteq \mathbb{R}^{\Omega}$ von reellwertigen Funktionen ein Integrationsfunktional $I_\mu : D \longrightarrow \overline{\mathbb{R}}$, das in Anlehnunug an gewisse Vorstellungen, die man von der “Fläche unter einem Funktionsgraphen” haben kann, möglichst die folgenden Eigenschaften besitzt:

1. I_μ setzt μ fort im Sinne von:

$$\forall A \in \mathcal{A} : \; \mathbb{1}_A \in D \quad \& \quad I_\mu(\mathbb{1}_A) = \mu(A).$$

2. I_μ ist homogen, d.h.:

$$\forall X \in D, \forall c \in \mathbb{R} : \; c \cdot X \in D \quad \& \quad I_\mu(c \cdot X) = c \cdot I_\mu(X).$$

3. I_μ ist additiv, d.h.:

$$\forall X, Y \in D: \ X + Y \in D \quad \& \quad I_\mu(X+Y) = I_\mu(X) + I_\mu(Y).$$

4. I_μ ist isoton, d.h.:

$$\forall X, Y \in D: \ X \geq Y \Longrightarrow I_\mu(X) \geq I_\mu(Y).$$

5. I_μ ist "nichtnegativ definit" auf den nichtnegativen Funktionen, d.h., für beliebige $X \geq 0$ gilt:

$$I_\mu(X) = 0 \iff \mu(\{\omega \in \Omega \mid X(\omega) \neq 0\}) = 0.$$

6. I_μ ist (in einem noch zu definierenden Sinne) stetig.

Bemerkung 3.7 *Anders als hier wird die Integrationstheorie üblicherweise etwas allgemeiner für Klassen von sogenannten numerischen Funktionen entwickelt. Numerische Funktionen sind Funktionen $X : \Omega \longrightarrow \overline{\mathbb{R}}$, die auch die Werte $-\infty$ und $+\infty$ annehmen können. Wir beschränken uns hier der Einfachheit halber auf reellwertige Funktionen. (Die Beweise für numerische Funktionen sind oft nicht wirklich schwieriger als für reellwertige Funktionen, nur manchmal etwas länger z.B. wegen eventuell nötiger Fallunterscheidungen.)*

Wir werden I_μ zunächst für eine kleine Klasse D von Funktionen definieren und diese Klasse dann schrittweise vergrößern:

i) Wir definieren I_μ für Indikatorfunktionen von messbaren Mengen. Aus Forderung 1 folgt dann schon, dass wir für $A \in \mathcal{A}$ den Wert von $I_\mu(\mathbb{1}_A)$ auf $\mu(A)$ setzen müssen.

ii) Wir definieren I_μ für sogenannte einfache nichtnegative Funktionen.

iii) Wir definieren I_μ für beliebige nichtnegative messbare Funktionen.

iv) Wir definieren I_μ für beliebige messbare Funktionen, für die die Konstruktion noch gut geht.

4 Integral für nichtnegative einfache Funktionen

Definition 1 (Einfache Funktion). Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Funktion $X : \Omega \longrightarrow \mathbb{R}$ heißt einfach, falls sie $\mathcal{A}-\mathfrak{B}(\mathbb{R})$-messbar ist und ein endliches Bild hat, d.h., die Menge

$$Im(X) := \{X(\omega) \mid \omega \in \Omega\}$$

ist endlich. Die Menge aller einfachen Funktionen notieren wir mit $\mathcal{T}$ und die Menge aller nichtegativen einfachen Funktionen notieren wir mit $\mathcal{T}_{\geq 0}$.

Bemerkung 4.1 *Jede einfache Funktion X ist darstellbar als*

$$X = \sum_{z \in Im(X)} z \cdot \mathbb{1}_{X^{-1}(\{z\})},$$

wobei alle $X^{-1}(\{z\})$ messbare Mengen (bezüglich $\mathcal{A}$) sind. Diese Darstellung von X bezeichnen wir als die Standarddarstellung von X.

Beispiel 4.2 *Sei $A \in \mathcal{A}$. Dann ist*

$$X := \mathbb{1}_A$$

eine einfache Funktion, denn $Im(X) = \{0,1\}$ und X ist $\mathcal{A}-\mathfrak{B}(\mathbb{R})$-messbar, denn für beliebiges $B \subseteq \mathbb{R}$ gilt

$$X^{-1}(B) \in \{\emptyset, \Omega, A, A^c\}.$$

Beispiel 4.3 *Sei $A \in \mathcal{A}^n$ und $\alpha \in \mathbb{R}^n$. Dann ist*

$$X := \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}$$

eine einfache Funktion, denn $|Im(X)| \leq |\{\sum_{k\in I} \alpha_k \mid I \subseteq \{1,\ldots,n\}\}| \leq 2^n$ und X ist als Linearkombination messbarer Funktionen wieder messbar.

Achtung: Es kann vorkommen, dass für verschiedene $\alpha \in \mathbb{R}^n$, $\beta \in \mathbb{R}^m$ und verschiedene $A \in \mathcal{A}^n$, $B \in \mathcal{A}^m$ die zugehörigen einfachen Funktionen $\sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}$ und $\sum_{k=1}^{m} \beta_k \cdot \mathbb{1}_{B_k}$ trotzdem übereinstimmen. Dies ist zu beachten, wenn man mit Hilfe einer konkreten Darstellung einer einfachen Funktion X irgend etwas definieren will. Bei der Definition des Integrals für nichtnegative einfache Funktionen gibt es nun zwei Möglichkeiten, vorzugehen, die aber auf das gleiche Ergebnis führen:

1) Definiere $I_\mu(X)$ für eine nichtnegative einfache Funktion über die Standarddarstellung. Eine naheliegende Wahl wäre:

$$I_\mu(X) = I_\mu\left(\sum_{z\in Im(X)} z \cdot \mathbb{1}_{X^{-1}(\{z\})}\right) := \sum_{z\in Im(X)} z \cdot \mu(X^{-1}(\{z\}))$$

mit der Konvention $0 \cdot \infty := 0$. Dies wäre wohldefiniert und würde zum Ziel führen. Eigenschaften, wie beispielsweise die Additivität von I_μ wären dann natürlich noch zu zeigen.

2) Definiere $I_\mu(X)$ über eine beliebige Darstellung von X und zeige, dass die Definition nicht von der konkreten Darstellung von X abhängt. Vorüberlegung dazu:

Wenn I_μ linear (, d.h. homogen und additiv) sein soll, dann muss für $X = \sum_{k=1}^{n} \alpha_k \mathbb{1}_{A_k}$ notwendigerweise gelten:

$$\begin{aligned} I_\mu(X) &= I_\mu\left(\sum_{k=1}^{n} \alpha_k \mathbb{1}_{A_k}\right) \\ &= \sum_{k=1}^{n} I_\mu(\alpha_k \mathbb{1}_{A_k}) \\ &= \sum_{k=1}^{n} \alpha_k I_\mu(\mathbb{1}_{A_k}) \\ &= \sum_{k=1}^{n} \alpha_k \mu(A_k), \end{aligned}$$

was auch zur Definition des Integrals für nichtnegative einfache Funktionen verwendet werden kann. Wir müssen uns jetzt nur noch darum kümmern, das diese Definition nicht von der konkreten Darstellung von X abhängt. Dies ist aber sichergestellt: Für zwei Darstellungen von X als

$$X = \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k} = \sum_{k=1}^{m} \beta_k \cdot \mathbb{1}_{B_k}$$

mit $A_1, \ldots, A_n, B_1, \ldots, B_m \in \mathcal{A}$, betrachte die Ununterscheidbarkeitsrelation $\sim$ des Mengensystems $\mathcal{E} := \{A_1, \ldots, A_n, B_1, \ldots, B_m\}$. Jede Äquivalenzklasse von $\Omega_{/\sim}$ entsteht als Schnitt von Mengen aus $\mathcal{E}$ bzw. Komplementen von Mengen aus $\mathcal{E}$ (vgl. Übung.). Da $\mathcal{E}$ endlich ist und da alle $A \in \mathcal{E}$ messbar sind, ist jede Äquivalenzklasse messbar. (Achtung: Ω kann hier auch überabzählbar sein.) Außerdem gibt es nur endlich viele Äquivalenzklassen, konkret gilt $|\Omega_{/\sim}| \leq 2^{2\cdot|\mathcal{E}|}$. Damit folgt:

$$
\begin{aligned}
\sum_{k=1}^{n} \alpha_k \cdot \mu(A_k) &= \sum_{k=1}^{n} \alpha_k \cdot \mu\left(\dot{\bigcup}\{[\omega]_\sim \mid [\omega]_\sim \subseteq A_k\}\right) \\
&= \sum_{k=1}^{n} \alpha_k \sum_{\substack{[\omega]_\sim \in \Omega_{/\sim}: \\ [\omega]_\sim \subseteq A_k}} \mu([\omega]_\sim) \\
&= \sum_{[\omega]_\sim \in \Omega_{/\sim}} \mu([\omega]_\sim) \sum_{\substack{k \in \{1,\dots,n\}: \\ A_k \supseteq [\omega]_\sim}} \alpha_k \\
&= \sum_{[\omega] \in \Omega_{/\sim}} \mu([\omega]_\sim) \cdot X(\omega) \\
&= \dots \text{ (Symmetrie)} \\
&= \sum_{k=1}^{m} \beta_k \cdot \mu(B_k).
\end{aligned}
$$

Aus der Darstellungsunabhängigkeit folgt dann sofort auch, dass I_μ additiv und positiv homogen[14] ist: Für $\alpha \in \mathbb{R}^n;\beta \in \mathbb{R}^m$, $A \in \mathcal{A}^n, B \in \mathcal{A}^m$ sowie $X = \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}$ und $Y = \sum_{k=1}^{m} \beta_k \cdot \mathbb{1}_{B_k}$ gilt:

$$
\begin{aligned}
I_\mu(X+Y) &= I_\mu\left(\sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k} + \sum_{k=1}^{m} \beta_k \cdot \mathbb{1}_{B_k}\right) \\
&= \sum_{k=1}^{n} \alpha_k \cdot \mu(A_k) + \sum_{k=1}^{m} \beta_k \cdot \mu(B_k) \\
&= I_\mu(X) + I_\mu(Y)
\end{aligned}
$$

und für $\lambda \in [0, \infty[$ folgt:

[14] Positiv homogen bedeutet homogen für nichtnegative Skalare λ. Da wir I_μ zunächst nur für nichtnegative einfache Funktionen definieren, müssen wir sicherstellen, dass für $X \in D$ auch $\lambda \cdot X$ wieder in D, also eine nichtnegative einfache Funktion ist, was mit der Restriktion auf nichtnegative λ an dieser Stelle sichergestellt ist.

$$
\begin{aligned}
I_\mu(\lambda \cdot X) &= I_\mu\left(\lambda \cdot \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}\right) \\
&= I_\mu\left(\sum_{k=1}^{n} \lambda \cdot \alpha_k \cdot \mathbb{1}_{A_k}\right) \\
&= \sum_{k=1}^{n} \lambda \cdot \alpha_k \cdot \mu(A_k) \\
&= \lambda \cdot \sum_{k=1}^{n} \alpha_k \cdot \mu(A_k) \\
&= \lambda \cdot I_\mu(X).
\end{aligned}
$$

Außerdem sieht man sofort, dass die Definitionen nach 1) und nach 2) zum gleichen Ergebnis führen. Nun zur positiven Definitheit: Für $X \in \mathcal{T}_{\geq 0}$ gilt nach Definition des Integrals:

$$
\begin{aligned}
I_\mu(X) = 0 &\iff \sum_{z \in Im(X)} z \cdot \mu(X^{-1}(\{z\})) = 0 \\
&\iff \forall z \in Im(X) \backslash \{0\} : \mu(X^{-1}(\{z\})) = 0 \\
&\iff \mu(\Omega \backslash X^{-1}(\{0\})) = 0 \\
&\iff \mu(\{\omega \in \Omega \mid X(\omega) \neq 0\}) = 0.
\end{aligned}
$$

Zur Stetigkeit sei $X = \sum\limits_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}$, $Y = \sum\limits_{k=1}^{m} \beta_k \cdot \mathbb{1}_{B_k}$ und $\sim$ die zu $\mathcal{E} = \{A_1, \ldots, A_n, B_1, \ldots, B_m\}$ gehörige Ununterscheidbarkeitsrelation. Dann folgt:

$$
\begin{aligned}
I_\mu(X) - I_\mu(Y) &= \sum_{[\omega] \in \Omega_{/\sim}} \mu([\omega]_\sim) \cdot X(\omega) - \sum_{[\omega] \in \Omega_{/\sim}} \mu([\omega]_\sim) \cdot Y(\omega) \\
&= \sum_{[\omega] \in \Omega_{/\sim}} \mu([\omega]_\sim) \cdot (X(\omega) - Y(\omega)) \qquad (*) \\
&\leq \mu(\Omega) \cdot \sup_{\omega \in \Omega} |X(\omega) - Y(\omega)| \\
&= \mu(\Omega) \cdot ||X - Y||_\infty,
\end{aligned}
$$

d.h., falls $\mu(\Omega)$ endlich ist, dann ist I_μ Lipschitz-stetig bezüglich der Supremumsnorm mit Lipschitz-Konstante $L = \mu(\Omega)$.

Bemerkung 4.4 *Hier hätte man anstelle der Supremumsnorm auch die wesentliche Supremumsnorm*
$||X-Y||_\infty^{ess} := \inf_{M:\mu(M)=0} \left\{ \sup_{\omega\in\Omega\setminus M} |X(\omega)-Y(\omega)| \right\}$ *verwenden können. (Vergleiche Beispiel 6.3 zu* μ*-fast-überall bestehenden Eigenschaften.)*

Abschließend folgt noch die Isotonie von I_μ aus *: Für $X \geq Y$ ist nämlich der Ausdruck * als Summe nichtnegativer Zahlen nichtnegativ und somit ist $I_\mu(X) - I_\mu(Y) \geq 0$, d.h., es gilt $I_\mu(X) \geq I_\mu(Y)$.

Bemerkung 4.5 *Es gibt Autoren, die das Integral im ersten Schritt bereits für messbare Funktionen mit abzählbarem Bild einführen. In dem Fall wäre Vorgehen 1) naheliegender und z.B. die Additivität würde man anders zeigen.*

5 Integral für nichtnegative messbare Funktionen

Definition 2 (Menge aller messbaren Funktionen). Sei $(\Omega, \mathcal{A})$ ein Messraum. Die Menge aller messbaren Funktionen $X : \Omega \longrightarrow \mathbb{R}$ sei mit $\mathcal{M}$ bezeichnet. Weiter bezeichne $\mathcal{M}_{\geq 0}$ die Menge aller nichtnegativen messbaren Funktionen.

Definition 3 (Konvergenz von unten für Funktionen). Sei Ω ein Grundraum, $(u_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger Funktionen und u eine weitere reellwertige Funktion. Wir sagen, dass die Folge u_n von Funktionen von unten gegen die Funktion $u : \Omega \longrightarrow \mathbb{R}$ konvergiert, wenn für alle $\omega \in \Omega$ gilt:

$$n \geq m \implies u_n(\omega) \geq u_m(\omega) \text{ und}$$
$$\lim_{n\to\infty} u_n(\omega) = u(\omega)$$

Schreibweise: $u_n \nearrow u$.

Satz 5.1 (Approximation durch einfache Funktionen) *Für jedes* $X \in \mathcal{M}_{\geq 0}$ *gibt es eine Folge* $u_n : \Omega \longrightarrow \mathbb{R}$ *von einfachen Funktionen, die von unten gegen X konvergiert, nämlich zum Beispiel die Folge*

$$u_n : \Omega \longrightarrow \mathbb{R} : \omega \mapsto \sup\left\{ a \in \frac{1}{n}\cdot\mathbb{Z} \mid a \leq X(\omega), a \leq n \right\}.$$

Satz 5.2 (Zur Wohldefiniertheit des Integrals für $X \in \mathcal{M}_{\geq 0}$) *Für jede wachsende Folge $(u_n)_{n\in\mathbb{N}} \in {\mathcal{T}_{\geq 0}}^{\mathbb{N}}$ und für jedes $v \in \mathcal{T}_{\geq 0}$ mit*

$$v \leq \lim_{n\to\infty} u_n$$

gilt

$$I_\mu(v) \leq \lim_{n\to\infty} I_\mu(u_n).$$

Beweis. Für $\beta > 1$ und $n \in \mathbb{N}$ betrachte

$$B_n := \{\omega \in \Omega \mid \beta \cdot u_n(\omega) \geq v(\omega)\} \in \mathcal{A}.$$

Dann ist B_n eine aufsteigende Folge mit $\bigcup\limits_{n\in\mathbb{N}} B_n = \Omega$ (kurz $B_n \nearrow \Omega$) und

$$\underbrace{\mathbb{1}_{B_n} \cdot v}_{\in \mathcal{T}_{\geq 0}} \leq \underbrace{\beta \cdot u_n}_{\in \mathcal{T}_{\geq 0}}.$$

Für v gibt es eine Darstellung als $v = \sum\limits_{k=1}^{m} \alpha_k \cdot \mathbb{1}_{A_k}$ mit $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ und $\mathcal{A}_1, \ldots, \mathcal{A}_m \in \mathcal{A}$ und damit ist $\mathbb{1}_{B_n} \cdot v$ darstellbar als $\mathbb{1}_{B_n} \cdot v = \sum\limits_{k=1}^{m} \alpha_k \cdot \mathbb{1}_{A_k \cap B_n}$. Es folgt nun

$$\begin{aligned}
I_\mu(v) &= \sum_{k=1}^{m} \alpha_k \cdot \mu(A_k) \\
&\overset{*}{=} \sum_{k=1}^{m} \alpha_k \cdot \lim_{n\to\infty} \mu(A_k \cap B_n) \\
&= \lim_{n\to\infty} \sum_{k=1}^{m} \alpha_k \cdot \mu(A_k \cap B_n) \\
&= \lim_{n\to\infty} I_\mu(v \cdot \mathbb{1}_{B_n}) \\
&\overset{**}{\leq} \beta \lim_{n\to\infty} I_\mu(u_n).
\end{aligned}$$

---

* $A_k \cap B_n \nearrow A_k$ und μ ist stetig von unten.

** $v \cdot \mathbb{1}_{B_n} \leq \beta \cdot u_n$.

Mit einem Grenzübergang $\beta \searrow 1$ folgt daraus die Behauptung.

Korollar 5.3 *Für zwei monoton wachsende Folgen* $(u_n)_{n\in\mathbb{N}} \in \mathcal{T}_{\geq 0}{}^{\mathbb{N}}$ *und* $(v_n)_{n\in\mathbb{N}} \in \mathcal{T}_{\geq 0}{}^{\mathbb{N}}$ *mit*

$$\lim_{n\to\infty} v_n = \lim_{n\to\infty} u_n \in \mathbb{R}^\Omega$$

gilt

$$\lim_{n\to\infty} I_\mu(v_n) = \lim_{n\to\infty} I_\mu(u_n).$$

Definition 4 (Integral für nichtnegative messbare Funktionen).
Sei $X \in \mathcal{M}_{\geq 0}$. Dann existiert eine Folge $(u_n)_{n\in\mathbb{N}}$ von einfachen Funktionen mit $u_n \nearrow X$ und das Integral von X bezüglich μ ist definiert als

$$I_\mu(X) := \lim_{n\to\infty} I_\mu(u_n).$$

Dies ist nach obigen Überlegungen wohldefiniert.

Bemerkung 5.4 *Für eine messbare nichtnegative Funktion*

$$X = \sum_{k=1}^{\infty} \underbrace{\alpha_k}_{\in\mathbb{R}_{\geq 0}} \cdot \mathbb{1}_{\underbrace{A_k}_{\in\mathcal{A}}} = \sum_{k=1}^{\infty} \underbrace{\beta_k}_{\in\mathbb{R}_{\geq 0}} \cdot \mathbb{1}_{\underbrace{B_k}_{\in\mathcal{A}}}$$

mit zwei verschiedenen Darstellungen mit jeweils nichtnegativen Koeffizienten gilt $\sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k} \nearrow X$ *bzw.* $\sum_{k=1}^{n} \beta_k \cdot \mathbb{1}_{B_k} \nearrow X$ *für* $n \to \infty$ *und somit ist*

$$\begin{aligned} I_\mu(X) &= I_\mu\left(\sum_{k=1}^{\infty} \alpha_k \cdot \mathbb{1}_{A_k}\right) \\ &= \lim_{n\to\infty} I_\mu\left(\sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}\right) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k \cdot \mu(A_k) \end{aligned}$$

und analog ist auch $I_\mu(X) = \sum_{k=1}^{\infty} \beta_k \cdot \mu(B_k)$. *Dies bedeutet, dass auch für messbare Funktionen mit abzählbarem Bild und nichtnegativen Koeffizienten* α_k *die Setzung*

$$I_\mu\left(\sum_{k=1}^{\infty}\alpha_k\cdot\mathbb{1}_{A_k}\right):=\sum_{k=1}^{\infty}\alpha_k\cdot\mu(A_k)$$

darstellungsunabhängig ist. Dies wissen wir aber erst jetzt, nachdem wir den "Umweg" über die messbaren Funktionen mit endlichem Bild gegangen sind.

Bemerkung 5.5 *Für $X\in\mathcal{M}_{\geq 0}$ gilt auch*

$$I_\mu(X)=\sup\{I_\mu(u)\mid u\in\mathcal{T}_{\geq 0},u\leq X\}.$$

Außerdem folgen die Eigenschaften $1.-6.$ für nichtnegative messbare Funktionen relativ unmittelbar durch eine Grenzwertbetrachtung. Beispielsweise die Additivität folgt so: Für $X,Y\in\mathcal{M}_{\geq 0}$ gibt es $(u_n)_{n\in\mathbb{N}}\in{\mathcal{T}_{\geq 0}}^{\mathbb{N}}$ und $(v_n)_{n\in\mathbb{N}}\in{\mathcal{T}_{\geq 0}}^{\mathbb{N}}$ monoton wachsend mit $u_n\nearrow X$ und $v_n\nearrow Y$. Damit gilt $u_n+v_n\nearrow X+Y$ und es folgt

$$\begin{aligned}I_\mu(X+Y)&=\lim_{n\to\infty}I_\mu(u_n+v_n)\\&=\lim_{n\to\infty}\big(I_\mu(u_n)+I_\mu(v_n)\big)\\&=\lim_{n\to\infty}I_\mu(u_n)+\lim_{n\to\infty}I_\mu(v_n)\\&=I_\mu(X)+I_\mu(Y).\end{aligned}$$

Satz 5.6 (Beppo Levi 1906: Von der monotonen Konvergenz)
Sei $(u_n)_{n\in\mathbb{N}}\in{\mathcal{M}_{\geq 0}}^{\mathbb{N}}$ monoton wachsend und sei

$$u:=\lim_{n\to\infty}u_n\in\mathbb{R}^{\Omega},$$

wobei der Grenzwert punktweise zu verstehen ist. Dann ist $u\in\mathcal{M}_{\geq 0}$ und es gilt

$$I_\mu(u)=\lim_{n\to\infty}I_\mu(u_n).$$

Zum Beweis.
i) Der Grenzwert u ist als punktweiser Grenzwert nichtnegativer Funktionen wieder nichtnegativ. Die Tatsache, dass u messbar ist, haben Sie in Übung 6, Aufgabe 2 gezeigt.
ii) Aus $u_n\leq u$ folgt $I_\mu(u_n)\leq I_\mu(u)$ und somit auch $\lim\limits_{n\to\infty}I_\mu(u_n)\leq I_\mu(u)$.

iii) Der Beweis von $I_\mu(u) \leq \lim_{n\to\infty} I_\mu(u_n)$ ist sehr ähnlich zum Beweis zur Wohldefiniertheit des Integrals für $X \in \mathcal{M}_{\geq 0}$ und kann zum Beispiel in [Elstrodt, 1996, S. 125] nachgelesen werden.

Bemerkung 5.7 *Die Voraussetzung, dass die Folge $(u_n)_{n\in\mathbb{N}}$ monoton wachsend ist, ist nicht entbehrlich, vergleiche Übung 8, Aufgabe 1. Unter der Einschränkung, dass $(u_n)_{n\in\mathbb{N}}$ monoton wachsend ist, ist obige monotone Konvergenz eine sehr viel stärkere Stetigkeitseigenschaft des Integrationfunktionals I_μ, als es die Lipschitz-Stetigkeit bezüglich der (wesentlichen) Supremumsnorm ist. Im Folgenden werden wir, wenn wir von Eigenschaft 6 des Integrationsfunktionals sprechen, aber weiterhin die Lipschitz-Stetigkeit bezüglich der (wesentlichen) Supremumsnorm meinen.*

Korollar 5.8 *Für jede Folge $(u_n)_{n\in\mathbb{N}}$ nichtnegativer messbarer Funktionen mit $\sum_{n=1}^{\infty} u_n \in \mathbb{R}^\Omega$ gilt:*

$$I_\mu\left(\sum_{n=1}^{\infty} u_n\right) = \sum_{n=1}^{\infty} I_\mu(u_n),$$

d.h., auf $\mathcal{M}_{\geq 0}$ ist I_μ ein σ-additives Funktional.

Dies ermöglicht die folgende Definition:

Definition 5 (Maß mit Dichte). Für $f \in \mathcal{M}_{\geq 0}$ definiere

$$f \odot \mu : \mathcal{A} \longrightarrow [0,\infty] : A \mapsto I_\mu\left(f \cdot \mathbb{1}_A\right).$$

Dies ist in der Tat ein Maß auf dem Messraum $(\Omega, \mathcal{A})$, das wir als Maß mit Dichte f in Bezug auf μ bezeichnen.

Zum Beweis der σ-Additivität sei $A = \bigcup_{n\in\mathbb{N}} A_n$ mit $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}^\mathbb{N}$ paarweise disjunkt. Dann gilt $f \cdot \mathbb{1}_A = \sum_{n\in\mathbb{N}} f \cdot \mathbb{1}_{A_n}$ und es folgt sofort

$$
\begin{aligned}
(f \odot \mu)(A) &= I_\mu(f \cdot \mathbb{1}_A) \\
&= I_\mu\left(\sum_{n\in\mathbb{N}} f \cdot \mathbb{1}_{A_n}\right) \\
&= \sum_{n\in\mathbb{N}} I_\mu\left(f \cdot \mathbb{1}_{A_n}\right) \\
&= \sum_{n\in\mathbb{N}} (f \odot \mu)(A_n).
\end{aligned}
$$

Bemerkung 5.9 *Für $f \in \mathcal{M}_{\geq 0}$ ist $f \circ \mu$ "stetig" bezüglich μ in folgendem Sinne: Ist $A \in \mathcal{A}$ eine μ-Nullmenge, so ist A auch eine $f \circ \mu$-Nullmenge, denn die Menge $\{\omega \in \Omega \mid f(\omega) \cdot \mathbb{1}_A > 0\} = \{\omega \in \Omega \mid f(\omega) > 0\} \cap A$ ist eine μ-mesbare Teilmenge von A und somit eine μ-Nullmenge, falls A eine μ-Nullmenge ist.*

Allgmein schreiben wir für zwei Maße ν und μ (auf dem gleichen Messraum)

$$\nu \ll \mu,$$

falls jede μ-Nullmenge auch eine ν-Nullmenge ist. In dem Fall sagen wir auch, dass ν absolut stetig bezüglich μ ist.

Gibt es demgegenüber eine Menge $A \in \mathcal{A}$ mit

$$\nu(A) = 0 \;\&\; \mu(A^c) = 0,$$

dann sagen wir, dass ν singulär bezüglich μ ist, in Zeichen:

$$\nu \perp \mu.$$

Bemerkung 5.10 *Für $f, g \in \mathcal{M}_{\geq 0}$ gilt*

$$g \odot (f \odot \mu) = (g \cdot f) \odot \mu.$$

und

$$I_{f\odot\mu}(g) = I_\mu(g \cdot f).$$

6 Integral für integrierbare Funktionen

Definition 6 (Integral für integrierbare Funktionen). Sei $X \in \mathcal{M}$. Definiere den Positivteil X^+ als

$$X^+ : \Omega \longrightarrow \mathbb{R} : \omega \mapsto \max\{0, X(\omega)\}$$

und den Negativteil X^- als

$$X^- : \Omega \longrightarrow \mathbb{R} : \omega \mapsto \max\{0, -X(\omega)\}.$$

Dann sind X^+ und X^- nichtnegativ und messbar und es gilt

$$X = X^+ - X^-.$$

Ist eines der Integrale $I_\mu(X^+)$ oder $I_\mu(X^-)$ endlich, dann definieren wir

$$I_\mu(X) := I_\mu(X^+) - I_\mu(X^-)$$

mit den Konventionen $\infty - c = \infty$ und $c - \infty = -\infty$ für $c \in \mathbb{R}$. In diesem Fall sagen wir, dass das Integral von X existiert. Sind beide Integrale endlich, so nennen wir X (Lebesgue-)integrierbar (bezüglich μ). Die Menge aller integrierbaren Funktionen (auf dem Maßraum $(\Omega, \mathcal{A}, \mu)$) bezeichnen wir mit $\mathfrak{L}^1(\Omega, \mathcal{A}, \mu)$ bzw. kurz mit $\mathfrak{L}^1(\mu)$ oder $\mathfrak{L}^1$.

Bemerkung 6.1 *Die Eigenschaften* 1. – 6. *des Integrals für messbare nichtnegative Funktionen übertragen sich entsprechend auf die Klasse aller integrierbaren Funktionen.*

Satz 6.2 (Satz von der majorisierten Konvergenz) *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $(X_n)_{n\in\mathbb{N}} \in \mathcal{M}^{\mathbb{N}}$ eine Folge messbarer Funktionen, die punktweise gegen eine reellwertige Funktion X konvergieren. Sei weiter Y eine nichtnegative messbare Funktion mit $I_\mu(Y) < \infty$ und $|X_n| \leq Y$ für alle $n \in \mathbb{N}$. Dann sind alle X_n sowie X integrierbar und es gilt*

$$\lim_{n\to\infty} I_\mu(X_n) = I_\mu(X).$$

Definition 7 (μ-fast überall bestehende Eigenschaften). Sei $E : \Omega \longrightarrow \{0,1\}$ eine "Eigenschaft" mit der Interpretation $E(\omega) = 1 \iff$ Eigenschaft E gilt "im Fall ω". Wir sagen dann, dass die Eigenschaft E μ-fast überall gilt, falls es eine μ-Nullmenge $N \subseteq \Omega$ gibt, so dass

$$E_{|\Omega\setminus N} = 1$$

gilt, d.h., dass für alle $\omega \in \Omega\setminus N$ die Eigenschaft E gilt.

Beispiel 6.3 *Seien X und Y zwei integrierbare Funktionen. Sind X und Y μ-fast überall identisch (kurz $X = Y$ μ-f.ü. bzw. formal: mit*

$$E : \Omega \longrightarrow \{0,1\} : \omega \mapsto \begin{cases} 1 & \textit{falls } X(\omega) = Y(\omega) \\ 0 & \textit{sonst} \end{cases}$$

gelte die Eigenschaft E μ-fast überall.) Dann gilt

$$I_\mu(X) = I_\mu(Y).$$

Grund: $X = Y$ μ-f.ü. bedeutet: $\exists N \in \mathcal{A} : \mu(N) = 0$ und $X_{|\Omega\setminus N} = Y_{|\Omega\setminus N}$.

Damit ist

$$X = \underbrace{X \cdot \mathbb{1}_{\Omega\setminus N}}_{=Y\cdot \mathbb{1}_{\Omega\setminus N}} + X \cdot \mathbb{1}_N$$

und somit ist

$$\begin{aligned}
I_\mu(X) &= I_\mu\left(X \cdot \mathbb{1}_{\Omega\setminus N} + X \cdot \mathbb{1}_N\right) \\
&= I_\mu(X \cdot \mathbb{1}_{\Omega\setminus N}) + I_\mu(\underbrace{X \cdot \mathbb{1}_N}_{\neq 0 \text{ nur auf Nullmenge } N}) \\
&= I_\mu(X \cdot \mathbb{1}_{\Omega\setminus N}) \\
&= I_\mu(Y \cdot \mathbb{1}_{\Omega\setminus N}) \\
&= I_\mu(Y \cdot \mathbb{1}_{\Omega\setminus N}) + \underbrace{I_\mu(Y \cdot \mathbb{1}_N)}_{=0} \\
&= I_\mu(Y \cdot \mathbb{1}_{\Omega\setminus N} + Y \cdot \mathbb{1}_N) \\
&= I_\mu(Y).
\end{aligned}$$

Satz 6.4 (Verhalten des Integrals unter Komposition) *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\Omega', \mathcal{A}')$ ein Messraum und $X : \Omega \longrightarrow \Omega'$ sowie $Y : \Omega' \longrightarrow \mathbb{R}$. Ist X $\mathcal{A} - \mathcal{A}'$-messbar und ist Y $\mathcal{A}' - \mathfrak{B}(\mathbb{R})$-messbar, dann gilt:*

$$I_\mu(Y \circ X) \textit{ existiert} \iff I_{\mu_X}(Y) \textit{ existiert}$$

und in diesem Fall gilt

$$I_\mu(Y \circ X) = I_{\mu_X}(Y).$$

Zum Beweis: Nur für einfache Funktion Y: Sei $Y = \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{A_k}$ mit $A_1, \dots, A_n \in \mathcal{A}'$ und $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. Dann ist $Y \circ X = \sum_{k=1}^{n} \alpha_k \cdot \mathbb{1}_{X^{-1}(A_k)}$ und $I_\mu(Y \circ X) = \sum_{k=1}^{n} \alpha_k \mu(X^{-1}(A_k))$. Auf der anderen Seite ist $I_{\mu_X}(Y) = \sum_{k=1}^{n} \alpha_k \mu_X(A_k) = \sum_{k=1}^{n} \alpha_k \mu(X^{-1}(A_k)) = I_\mu(Y \circ X)$.

Satz 6.5 (Transformationssatz für Dichten) *Seien*

$$\begin{aligned} X :& \Omega \longrightarrow G \\ Y :& G \longrightarrow \mathbb{R}^d \end{aligned}$$

mit $G \subseteq \mathbb{R}^d$ offen und Y injektiv mit Umkehrabbildung $Y^{-1} : Im(Y) \longrightarrow G$ und Y sowie Y^{-1} messbar und stetig differenzierbar.
Besitzt das Bildmaß von X die (p-dimensionale) Lebesgue-Dichte f_X (ist also $\mu_X = f_X \circ \lambda^p$), dann ist die Dichte des Bildmaßes der Zufallsvariable

$$Z := Y \circ X$$

(bezüglich des p-dimensionalen Lebesgue-Maßes) gegeben durch

$$f_Z(z) = \begin{cases} |\det(J_{Y^{-1}}(z))| \cdot f_X(Y^{-1}(z)) & \text{falls } z \in Im(Z) \\ 0 & \text{sonst} \end{cases}.$$

(Hier bezeichnet $J_{Y^{-1}}(z)$ die Jakobi-Matrix von Y^{-1} ausgewertet an der Stelle z.)

Satz 6.6 (Satz von Radon Nikodym, Variante) *Sei μ ein σ-endliches Maß und $\nu \ll \mu$ ein Maß (jeweils auf $(\Omega, \mathcal{A})$). Dann ist ν darstellbar als ein Maß mit Dichte bezüglich μ, d.h., es gibt eine nichtnegative messbare Funktion $f : \Omega \longrightarrow \mathbb{R}$, so dass*

$$\nu = f \odot \mu.$$

Außerdem ist die Funktion f μ-fast überall eindeutig bestimmt, d.h., für jedes weitere g mit obiger Eigenschaft gilt $f = g$ μ-fast überall.

Satz 6.7 (Lebesguescher Zerlegungssatz, Variante) *Sind μ und ν zwei σ-endliche Maße auf dem Messraum $(\Omega, \mathcal{A})$, so gibt es genau eine Zerlegung*

$$\nu = \rho + \delta$$

mit $\rho \ll \mu$ und $\delta \perp \mu$.

Beispiel 6.8 *Sei λ das Lebesgue-Maß und ν ein Wahrscheinlichkeitsmaß. Dann kann ν zerlegt werden in $\nu = \rho + \delta$ mit $\rho \ll \lambda$ und $\delta \perp \lambda$. Der Anteil ρ ist absolut stetig bezüglich λ, d.h. insbesondere gibt es eine Dichte f mit $\rho = f \odot \lambda$. Der Rest δ ist singulär zu λ und kann weiter zerlegt werden in:*

-Den diskreten Teil $\delta_{diskret}$: Mit $D := \{\omega \in \Omega \mid \delta(\{\omega\}) > 0\}$ (maximal abzählbar) ist

$$\delta_{diskret} : \mathcal{A} \longrightarrow [0, \infty] : A \mapsto \delta(A \cap D).$$

-Den (schwer zu behandelnden) Rest: $\delta_{Rest} = \delta - \delta_{diskret}$.

Satz 6.9 (Zusammenhang zum eigentlichen Riemann-Integral) *Seien $a, b \in \mathbb{R}$ und (mit einem bisschen abuse of notation) λ das Lebesguemaß eingeschränkt auf alle Lebesgue-messbaren Teilmengen des Intervalls $[a, b]$. Sei weiter $f : [a, b] \longrightarrow \mathbb{R}$ auf dem Intervall $[a, b]$ (eigentlich) Riemann-integrierbar. Dann ist f Lebesgue-integrierbar bezüglich λ und beide Integralwerte stimmen überein, d.h.:*

$$\int_a^b f(t)\, dt = I_\lambda(f).$$

Bemerkung 6.10 *Es gibt natürlich Lebesgue-integrierbare Funktionen, die nicht Riemann-integrierbar sind, beispielsweise die Dirichlet-Funktion*

$$D : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto \begin{cases} 1 & \textit{falls } x \textit{ rational} \\ 0 & \textit{falls } x \textit{ irrational} \end{cases}.$$

Etwas erstaunlicher ist die Tatsache, dass es uneigentlich Riemann-integrierbare Funktionen f gibt, die nicht Lebesgue-integrierbar sind, z.B.:

$$f : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R} : x \mapsto \begin{cases} \frac{1}{\lceil 1+x \rceil} & \textit{falls } \lfloor x \rfloor \textit{ gerade} \\ \frac{-1}{\lfloor 1+x \rfloor} & \textit{falls } \lfloor x \rfloor \textit{ ungerade} \end{cases}.$$

Dann gilt für das uneigentliche Riemann-Integral:

$$\int_0^\infty f(t)\,dt = \lim_{x\to\infty} \int_0^x f(t)\,dt = \ldots = 0.$$

Für das Lebesgue-Integral (bezüglich des Lebesgue-Maßes eingeschränkt auf $\mathbb{R}_{\geq 0}$) ist jedoch

$$I_\lambda(f^+) = I_\lambda(f^-) = \infty,$$

d.h., das Lebesgue-Integral existiert nicht. Frage: Warum nicht einfach $\infty - \infty =: 0$ in der Lebesgueschen Integrationstheorie definieren?
Antwort: Weil das "alles zerstört":
Beispielsweise mit $g : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R} : x \mapsto \frac{1}{(1+x)^2}$ hätten wir $\int_0^\infty \frac{1}{(1+t)^2}\,dt = \lim_{x\to\infty} \int_0^x \frac{1}{(1+t)^2}\,dt = \lim_{x\to\infty} \left[\frac{-1}{1+t}\right]_0^x = 1$, also auch $I_\lambda(g) = 1$, denn $\int_0^x g(t)\,dt = I_\lambda(g \cdot \mathbb{1}_{[0,x]})$ für $x \in \mathbb{R}$ und $\underbrace{g \cdot \mathbb{1}_{[0,x]}}_{\in \mathcal{M}_{\geq 0}} \nearrow \underbrace{g}_{\in \mathcal{M}_{\geq 0}}$ für $x \nearrow \infty$. Wenn wir $I_\lambda(f) = \infty - \infty = 0$ setzten, dann würde einerseits

$$I_\lambda(g+f) = I_\lambda(g) + I_\lambda(f) = 1 + 0 = 1$$

folgen, sofern wir an der Additivität von I_λ festhielten. Andererseits würde dies jedoch in Widerspruch zur Integraldefinition

$$I_\lambda(g+f) = I_\lambda((f+g)^+) - I_\lambda((f+g)^-) = \ldots = \infty - \infty = 0$$

stehen.
Frage: Ist die Tatsache, dass es uneigentlich Riemann-integrierbare Funktionen gibt, die nicht Lebesgue-integrierbar sind, ein "Problem" für die

Lebesguesche Integrationstheorie?

Ein Aspekt: Wenn es um "abstrakte Integration" geht (, d.h., auf Ω ist nur eine σ-Algebra $\mathcal{A}$ und ein Maß μ gegeben, und darüber hinaus nichts), dann ist die Riemannsche Integrationstheorie eh schon aus dem Rennen, und es macht mangels vorhandener Struktur auch keinen Sinn, mit einer Grenzwertüberlegung im Definitionsbereich Ω der betrachteten Funktion (auf dem ja eben nur $\mathcal{A}$ und μ anzutreffen sind) zu versuchen, den Fall "$\infty - \infty$" zu retten.

Definition 8 (Diskretes Maß). Ein Maß μ auf einem Grundraum $(\Omega, \mathcal{A})$ heißt diskret, falls es eine höchstens abzählbare Menge $D_\mu \subseteq \Omega$ gibt mit

D1: Für alle $x \in D_\mu$ ist $\{x\}$ messbar bezüglich $\mathcal{A}$.
D2: Es gilt $\mu(\Omega \backslash D_\mu) = 0$.

Satz 6.11 (Integration bezüglich eines diskreten Maßes) *Für ein diskretes Maß μ auf einem Messraum $(\Omega, \mathcal{A})$ berechnet sich das Integral einer integrierbaren Funktion zu*

$$I_\mu(X) = \sum_{\substack{\omega \in D_\mu: \\ X(\omega) > 0}} X(\omega) \cdot \mu(\{\omega\}) + \sum_{\substack{\omega \in D_\mu: \\ X(\omega) < 0}} X(\omega) \cdot \mu(\{\omega\}).$$

Beweis. Die Menge $\Omega \backslash D_\mu$ ist eine μ-Nullmenge, d.h., die Funktion X^+ und die Funktion $\tilde{X}^+ := X^+ \cdot \mathbb{1}_{D_\mu}$ haben das gleiche Integral, denn $X^+ = \tilde{X}^+ \mu$-f.ü. Ist $D_\mu = \dot{\bigcup\limits_{k \in \mathbb{N}}} \{\omega_k\}$ abzählbar, dann ist

$$\tilde{X}^+ = \sum_{\substack{k \in \mathbb{N}: \\ X(\omega_k) > 0}} X^+(\omega_k) \cdot \underbrace{\mathbb{1}_{\{\omega_k\}}}_{\text{messbar}}$$

und für

$$\tilde{X}_n^+ := \sum_{\substack{k \in \{1, \dots, n\} \\ X(\omega_k) > 0}} X^+(\omega_k) \cdot \mathbb{1}_{\{\omega_k\}} \in \mathcal{T}_{\geq 0}$$

gilt $\tilde{X}_n^+ \nearrow \tilde{X}^+$. Also ist

$$\begin{aligned}
I_\mu(X^+) &= I_\mu(\tilde{X}^+) \\
&= \lim_{n\to\infty} \sum_{\substack{k\in\{1,\ldots,n\}:\\ X(\omega_k)>0}} X^+(\omega_k)\cdot\mu(\{\omega_k\}) \\
&= \sum_{\substack{k\in\mathbb{N}:\\ X(\omega_k)>0}} X^+(\omega_k)\cdot\mu(\{\omega_k\}) \\
&= \sum_{\substack{\omega\in D_\mu:\\ X(\omega)>0}} X(\omega)\cdot\mu(\{\omega\}).
\end{aligned}$$

(Für D_μ endlich folgt obiges direkt ohne eine Approximation von unten.) Analog folgt auch $I_\mu(X^-) = \sum_{\substack{\omega\in D_\mu:\\ X(\omega)<0}} -X(\omega)\cdot\mu(\{\omega\})$ und insgesamt ist damit

$$\begin{aligned}
I_\mu(X) &= I_\mu(X^+) - I_\mu(X^-) \\
&= \sum_{\substack{\omega\in D_\mu:\\ X(\omega)>0}} X(\omega)\cdot\mu(\{\omega\}) - \sum_{\substack{\omega\in D_\mu:\\ X(\omega)<0}} -X(\omega)\cdot\mu(\{\omega\}) \\
&= \sum_{\substack{\omega\in D_\mu:\\ X(\omega)>0}} X(\omega)\cdot\mu(\{\omega\}) + \sum_{\substack{\omega\in D_\mu:\\ X(\omega)<0}} X(\omega)\cdot\mu(\{\omega\}).
\end{aligned}$$

Definition 9 (Erwartungswert, Varianz, Kovarianz). Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable, für die $I_P(X)$ existiert und endlich ist. Dann definieren wir den Erwartungswert von X einfach als

$$\mathbb{E}_P(X) := I_P(X)$$

bzw. kurz

$$\mathbb{E}(X) := \mathbb{E}_P(X).$$

Ist $I_p(X^2)$ endlich, so ist die Varianz von X definiert als

$$\begin{aligned} Var_P(X) &:= I_p\left((X - \mathbb{E}_P(X))^2\right) \\ &= I_P(X^2) - (\mathbb{E}_P(X))^2 \end{aligned}$$

bzw. kurz

$$Var(X) := Var_P(X).$$

Für eine Zufallsvariable $Z : \Omega \longrightarrow \mathbb{R}^d$, für die das Integral

$$I_P(Z_i \cdot Z_j)$$

für alle $i \in \{1, \ldots, d\}$ existiert und endlich ist, ist die Kovarianzmatrix $\Sigma_Z \in \mathbb{R}^{d \times d}$ definiert als

$$(\Sigma_Z)_{ij} := I_P\left((Z_i - \mathbb{E}(Z_i)) \cdot \left(Z_j - \mathbb{E}(Z_j)\right)\right).$$

Bemerkung 6.12 *Im Folgenden werden wir anstelle von $I_\mu(X)$ auch die folgenden Notationen verwenden:*

$\int X \, d\mu$ *oder*

$\int\limits_\Omega X \, d\mu$ *oder auch*

$\int\limits_\Omega X(\omega) \, d\mu(\omega)$.

Außerdemsetzen wir für $A \in \mathcal{A}$:

$$\int_A X \, d\mu := I_\mu(X \cdot \mathbb{1}_A).$$

7 Abschließende Bemerkungen zur Integrationstheorie

Es gäbe noch viele andere Möglichkeiten, das Funktional I_μ einzuführen:

Ein Vorgehen wäre beispielsweise über die Survivorfunktion

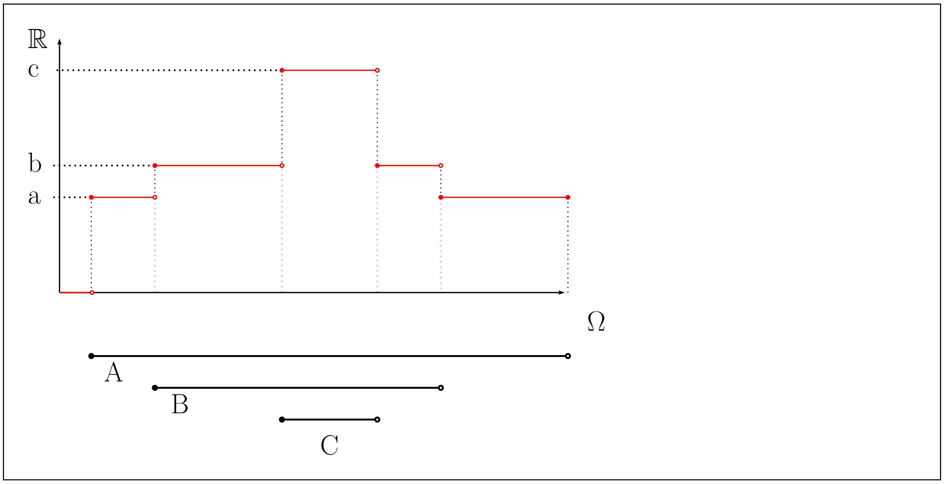
$$S : [0, \infty) \longrightarrow [0, \infty] : t \mapsto \mu(\{\omega \in \Omega \mid X(\omega) > t\})$$

für $X \in \mathcal{M}_{\geq 0}$. Dann kann I_μ einfach als

$$I_\mu(X) = \int_0^\infty S(t)\, dt$$

definiert werden[15].
Da die Survivorfunktion S eine antitone Funktion ist (, d.h., $x \leq y \Longrightarrow S(x) \geq S(y)$), ist (, sofern $S(t) < \infty$ sichergestellt ist,) S in jedem Fall (uneigentlich) Riemann-integrierbar. (Beachte: $S \geq 0$.) Natürlich kann $I_\mu(X)$ auch unendlich sein.

Illustration:

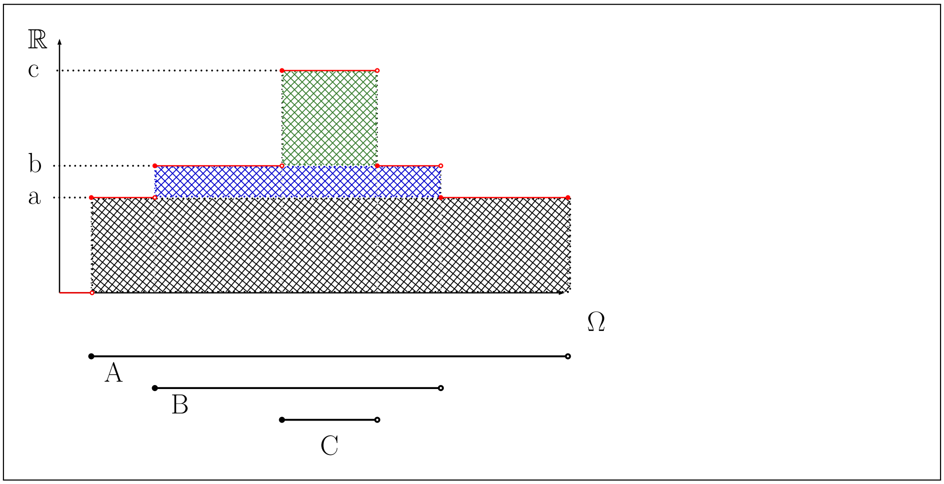


$$\begin{aligned} X &= a \cdot \mathbb{1}_A + (b-a) \cdot \mathbb{1}_B + (c-b) \cdot \mathbb{1}_C && i) \\ &= a \cdot \mathbb{1}_{X^{-1}(\{a\})} + b \cdot \mathbb{1}_{X^{-1}(\{b\})} + c \cdot \mathbb{1}_{X^{-1}(\{c\})} && ii) \end{aligned}$$

[15] Für beliebiges $X \in \mathcal{M}$ kann dann $I_\mu(X)$ definiert werden als $I_\mu(X) = \int_0^\infty S(t)\, dt - \int_{-\infty}^0 F(t)\, dt$, sofern dies wohldefiniert ist. Hier ist F die Verteilungsfunktion von X.

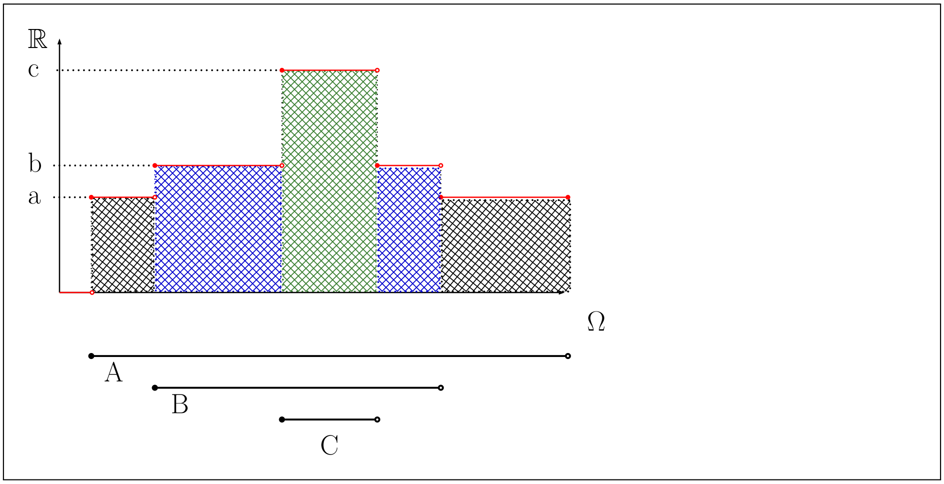
Dann ist nach der hier betrachteten Integrationstheorie gemäß i):

$$I_\mu(X) = a \cdot \mu(A) + (b-a) \cdot \mu(B) + (c-b) \cdot \mu(C)$$



Gleichzeitig gilt auch gemäß ii):

$$I_\mu(X) = a \cdot \mu(X^{-1}(\{a\})) + b \cdot \mu(X^{-1}(\{b\})) + c \cdot \mu(X^{-1}(\{c\})).$$



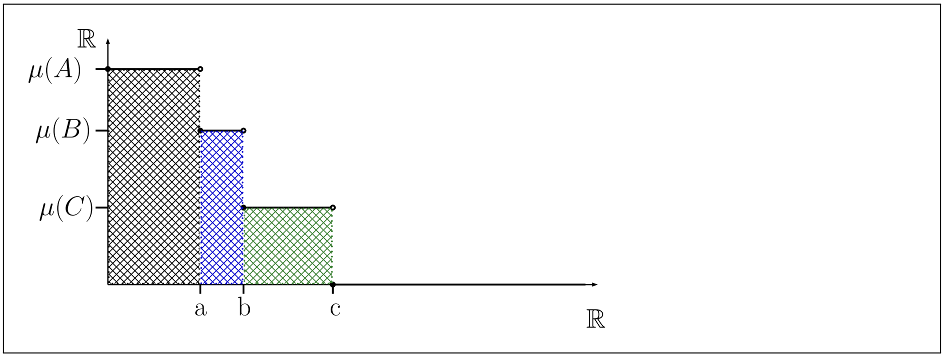
Wenn μ ein Maß (und damit insbesondere (σ-)additiv ist), dann sind $I_\mu(X)$ nach $i)$ und nach $ii)$ gemäß allen bisherigen Überlegungen identisch.

Alternativ mit der Definition des Integrals über die Survivorfunktion

$$S(t) = \mu(\{\omega \in \Omega \mid X(\omega) > t\})$$

erhält man:

$$I_\mu(X) \int_0^\infty S(t)\, dt = a \cdot \mu(A) + (b-a) \cdot \mu(B) + (b-c) \cdot \mu(C)$$



, was im Falle eines klassischen Maßes μ identisch zu I_μ von oben ist. (Der Nachweise dieser Tatsache, bzw. der Nachweis, dass das über die Survivorfunktion definierte Integral alle wünschneswerten Eigenschaften $1-6$ besitzt, wäre dann naürlich noch offen.)

Ist μ (σ-)additiv, dann spielt es keine Rolle, wie wir die "Flächenstücke" zusammensetzten, es ergibt sich immer der gleiche Wert für das Integral $I_\mu(X)$. Ist μ jedoch nicht-additiv, dann fallen die verschiedenen Zugänge zu I_μ nicht mehr zusammen. In diesem Fall ist jedoch z.B. der Ansatz über die Survivorfunktion trotzdem noch wohldefiniert. Man erhält einen Integralbegriff, der für bestimmte Fragestellungen immer noch fruchtbar ist, das Integral ist dann aber natürlich im Allgemeinen nicht mehr additiv. Das so über die Survivorfunktion definierte Integral wird als das Choquet-Integral bezeichnet. (Hintergrund ist die Arbeit [Choquet, 1954],vergleiche auch [Denneberg, 2013] zur Integrationstheorie nicht-additiver Mengenfunktionen.)

Ein noch anderes Vorgehen bestünde in der “gleichzeitigen” Abhandlung der äußeren Maßkonstruktion und der eigentlichen Integrationstheorie in einem einzigen Schritt:
Für $\mu : \mathcal{S} \longrightarrow [0,\infty]$, oder auch allgemeiner für $\mu : \mathcal{K} \longrightarrow \overline{\mathbb{R}}$ mit $\mathcal{K} \subseteq \mathbb{R}^{\Omega}$ definiere:

$$\overline{I_\mu}(X) := \inf\left\{ \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \cdot \mu(A_k) \mid \alpha_0 \in \mathbb{R}, \alpha_1, \alpha_2, \ldots \in \mathbb{R}_{\geq 0}, A_1, A_2, \ldots \in \mathcal{S}, \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \cdot \mathbb{1}_{A_k} \geq X \right\}$$

bzw.

$$\overline{I_\mu}(X) := \inf\left\{ \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \cdot \mu(A_k) \mid \alpha_0 \in \mathbb{R}, \alpha_1, \alpha_2, \ldots \in \mathbb{R}_{\geq 0}, X_1, X_2, \ldots \in \mathcal{K}, \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \cdot X_k \geq X \right\}$$

als “obere Erwartung” von X. Anstelle einer Approximation von X durch abzählbar unendlich viele Mengen/Funktionen wäre auch eine Approximation mit nur endlich vielen Mengen/Funktionen denkbar. Dieser Weg wird z.B. (in leicht anderer Darstellung) in Walley [1991] eingeschlagen, wobei die Betrachtung von nur endlich vielen Mengen/Funktionen dort mehr inhaltlich und weniger rein mathematisch motiviert ist.

Kapitel 3
Konvergenzbegriffe für Zufallsvariablen

8 Die L^p-Räume, der Hilbertraum L^2 und Konvergenz im p-ten Mittel

Gegeben sei ein Maßraum $(\Omega, \mathcal{A}, \mu)$ und $p \in [0, \infty)$. Betrachte die Menge

$$\mathfrak{L}^p(\Omega, \mathcal{A}, \mu) := \{X : \Omega \longrightarrow \mathbb{R} \text{ messbar } \mid \int_\Omega |X|^p \, d\mu < \infty\}.$$

Diese Menge, ausgestattet mit der punktweisen Addition $+$ von Zufallsvariablen und der punktweisen Skalarmultiplikation $\cdot$, bildet einen linearen Raum (Vektorraum) über $\mathbb{R}$. Die Abbildung

$$||\cdot||_{\mathfrak{L}^p} : \mathfrak{L}^p \longrightarrow [0, \infty) : X \mapsto \left[\int_\Omega |X|^p \, d\mu\right]^{\frac{1}{p}}$$

ist für $p \geq 1$ eine Halbnorm auf $\mathfrak{L}^p$. Dabei heißt eine Abbildung $||\cdot||$ von einem Vektorraum V (über $\mathbb{R}$) nach $[0, \infty)$ Halbnorm, falls für $X, Y \in V$ und $\lambda \in \mathbb{R}$ beliebig gilt:

i) $||\lambda \cdot X|| = |\lambda| \cdot ||X||$ (absolute Homogenität)
ii) $||X + Y|| \leq ||X|| + ||Y||$ (Dreiecksungleichung)

Zur Norm fehlt einer Halbnorm lediglich die Eigenschaft

$$||X|| = 0 \Longrightarrow X = 0.$$

Die Relation

$$\begin{aligned}\sim_\mu := &\{(X,Y) \in \mathbb{R}^\Omega \times \mathbb{R}^\Omega \mid X = Y \quad \mu\text{-f.ü}\} \\ &= \{(X,Y) \in \mathbb{R}^\Omega \times \mathbb{R}^\Omega \mid X - Y = 0 \quad \mu\text{-f.ü.}\}\end{aligned}$$

ist eine Kongruenzrelation bezüglich der Operationen $+$ und $\cdot$, d.h., $\sim_\mu$ ist eine Äquivalenzrelation, die mit $+$ und $\cdot$ verträglich ist im Sinne von:

$$\begin{aligned}\forall X, \tilde{X}, Y, \tilde{Y} \in \mathbb{R}^\Omega : \ & X \sim_\mu \tilde{X}, Y \sim_\mu \tilde{Y} \Longrightarrow X + Y \sim_\mu \tilde{X} + \tilde{Y} \\ \forall X, \tilde{X} \in \mathbb{R}^\Omega, \lambda \in \mathbb{R} : \ & X \sim_\mu \tilde{X} \Longrightarrow \lambda \cdot X \sim_\mu \lambda \cdot \tilde{X}\end{aligned}$$

(Zum Beweis dieser Tatsache siehe Hausübung 3, Aufgabe 2 und Übung 10, Aufgabe 1.) Dies macht die folgende Konstruktion wohldefiniert:

Betrachte den Quotientenraum

$$L^p := \mathfrak{L}^p(\Omega, \mathcal{A}, \mu)_{/\sim_\mu},$$

ausgestattet mit der "vertreterweisen" Addition

$$+ : L^p \times L^p \longrightarrow L^p : \big([X]_{\sim_\mu}, [Y]_{\sim_\mu}\big) \mapsto [X+Y]_{\sim_\mu}$$

und der "vertreterweisen" Skalarmultiplikation

$$\cdot : \mathbb{R} \times L^p \longrightarrow L^p : \big(\lambda, [X]_{\sim_\mu}\big) \mapsto [\lambda X]_{\sim_\mu}.$$

Dieser Raum ist wieder ein linearer Raum und die Abbildung

$$||\cdot||_{L^p} : L^p \longrightarrow [0, \infty[: [X]_{\sim_\mu} \mapsto \left[\int |X|^p \, d\mu\right]^{\frac{1}{p}}$$

ist wohldefiniert und (für $p \geq 1$) jetzt sogar eine Norm auf L^p. Versehen mit dieser Norm ist L^p ein vollständiger normierter Raum (Banachraum). Vollständig bedeutet, dass jede Cauchyfolge[16] konvergent ist.
Speziell für $p = 2$ wird $||\cdot||_{L^2}$ vom Skalarprodukt

[16] Eine Folge $(x_n)_{n \in \mathbb{N}}$ von Elementen eines metrischen Raumes (M,d) heißt Cauchyfolge, falls für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert mit $m, n \geq n_0 \Longrightarrow d(x_n, x_m) \leq \varepsilon$.

$$\langle \cdot,\cdot \rangle_{L^2} : L^2 \times L^2 \longrightarrow \mathbb{R} : \big([X]_{\sim\mu},[Y]_{\sim\mu}\big) \mapsto \int X \cdot Y \, d\mu$$

erzeugt im Sinne von

$$||X||_{L^2} = \sqrt{\langle X,X \rangle_{L^2}}.$$

Dies bedeutet, dass wir für $p = 2$ einen vollständigen Skalarproduktraum (Hilbertraum) haben.

Definition 1 (Konvergenz im p-ten Mittel). Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $p \in [1,\infty)$ und $([X_n]_{\sim\mu})_{n\in\mathbb{N}}$ eine Folge von Elementen aus L^p, sowie $[X]_{\sim\mu}$ ein weiteres Element aus L^p. Wir sagen, dass $[X_n]_{\sim\mu}$ gegen $[X]_{\sim\mu}$ im p-ten Mittel konvergiert, oder auch, dass X_n gegen X im p-ten Mittel konvergiert, wenn

$$\mathbb{E}(|X_n - X|^p) \longrightarrow 0$$

für $n \to \infty$ gilt.
Achtung: dies impliziert insbesondere $\mathbb{E}(|X_n - X|) < \infty$ für $n \in \mathbb{N}$. Außerdem wird vorausgesetzt, dass $X \in \mathfrak{L}^p$ und $X_n \in \mathfrak{L}^p$ für $n \in \mathbb{N}$ ist.
Schreibweise: $[X_n]_{\sim\mu} \xrightarrow{p} [X]_{\sim\mu}$ oder auch $X_n \xrightarrow{p} X$.

Bemerkung 8.1 *Obige Definition bezieht sich nur auf reellwertige Zufallsvariablen bzw. entsprechende Äquivalenzklassen. Für $\mathbb{R}^d$-wertige Zufallsvariablen würde man entsprechend von Konvergenz im p-ten Mittel sprechen, wenn Konvergenz bezüglich jeder Dimension $i = 1,\dots,p$ vorliegt.*

9 Weitere Konvergenzbegriffe

Konvergenz in Wahrscheinlichkeit

Definition 2 (Konvergenz in Wahrscheinlichkeit). Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen und X eine weitere $\mathbb{R}^d$-wertige Zufallsvariable (jeweils auf $(\Omega,\mathcal{A},P)$). Wir sagen, dass $(X_n)_{n\in\mathbb{N}}$ in Wahrscheinlichkeit (oder stochastisch) gegen X konvergiert, falls für jedes $\varepsilon > 0$

$$\lim_{n\to\infty} P(\{\omega \in \Omega \mid ||X_n(\omega) - X(\omega)|| > \varepsilon\}) = 0$$

gilt. Hier ist $||\cdot||$ irgendeine Norm in $\mathbb{R}^d$, z.B. die euklidische Norm. (Beachte, dass in $\mathbb{R}^d$ alle Normen äquivalent sind.) Schreibweise: $X_n \xrightarrow{\mathbb{P}} X$.

Fast sichere Konvergenz

Definition 3 (Sichere und fast sichere Konvergenz). Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen und X eine weitere $\mathbb{R}^d$-wertige Zufallsvariable (jeweils auf $(\Omega, \mathcal{A}, P)$). Wir sagen, dass die Folge $(X_n)_{n\in\mathbb{N}}$ sicher gegen X konvergiert, falls für alle $\omega \in \Omega$ gilt:

$$\lim_{n\longrightarrow\infty} X_n(\omega) = X(\omega).$$

Weiter sagen wir, dass die Folge $(X_n)_{n\in\mathbb{N}}$ P-fast sicher gegen X konvergiert, falls

$$P\Big(\{\omega \in \Omega \mid \lim_{n\to} X_n(\omega) = X(\omega)\}\Big) = 1$$

gilt. Schreibweise: $X_n \overset{P-f.s.}{\longrightarrow} X$.

Bemerkung 9.1 *$\lim_{n\to\infty} \underbrace{X_n(\omega)}_{\in\mathbb{R}^d} = \underbrace{X(\omega)}_{\in\mathbb{R}^d}$ ist hier punktweise (also für jede Dimension $i = 1, \ldots, p$) zu verstehen. Da $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ äquivalent zu $\liminf_{n\to\infty} X_n(\omega) = \limsup_{n\to\infty} X_n(\omega) = X(\omega)$ ist, und da für eine messbare Folge $(X_n)_{n\in\mathbb{N}}$ sowohl $\liminf_{n\to\infty} X_n$, als auch $\limsup_{n\to\infty} X_n$ messbar sind (vergleiche Übung 6, Aufgabe 2), ist die Menge*

$$\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$$

immer messbar.

Satz 9.2 (Fast sicher konvergent impliziert stochastisch konvergent) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge von $\mathbb{R}^d$-wertigen Zufallsvariablen und X eine weitere $\mathbb{R}^d$-wertige Zufallsvariable (jeweils auf $(\Omega, \mathcal{A}, P)$). Dann gilt:*

$$X_n \overset{P-f.s.}{\longrightarrow} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X\,.$$

Beweis. Gelte $X_n \overset{P-f.s.}{\longrightarrow} X$, also $P(\underbrace{\{\omega \in \Omega \mid X_n(\omega) \longrightarrow X(\omega)\}}_{=:B}) = 1$. Sei jetzt $\varepsilon > 0$ beliebig. Definiere

$$A_n := \{\omega \in \Omega \mid ||X_N(\omega) - X(\omega)|| < \varepsilon \text{ für alle } N \geq n\} \in \mathcal{A}.$$

und

$$A := \bigcup_{n \in \mathbb{N}} A_n = \{\omega \in \Omega \mid \exists n_0 \in \mathbb{N} : ||X_n(\omega) - X(\omega)|| < \varepsilon \text{ für } n \geq n_0\} \in \mathcal{A}.$$

Dann ist $B \subseteq A$ und somit $P(A) \geq P(B) = 1$, also auch $P(A) = 1$. Da $A_n \nearrow A$ folgt $P(A) = \lim_{n \to \infty} P(A_n) = 1$, denn P ist stetig von unten. Außerdem gilt

$$P(\underbrace{\{\omega \in \Omega \mid ||X_n(\omega) - X(\omega)|| < \varepsilon\}}_{\supseteq A_n}) \geq P(A_n),$$

und da $P(A_n)$ gegen 1, konvergiert, muss auch $P(\{\omega \in \Omega \mid ||X_n(\omega) - X(\omega)|| < \varepsilon\})$ gegen 1 konvergieren.

Bemerkung 9.3 *Aus $X_n \overset{\mathbb{P}}{\longrightarrow} X$ folgt nicht notwendigerweise $X_n \overset{P-f.s.}{\longrightarrow} X$.*
Beispiel (Wandernde Türme, siehe [Schmidt, 2009, S.333])
Sei der Wahrscheinlichkeitsraum $((0,1], \mathfrak{B}[(0,1]), \lambda_{|\mathfrak{B}((0,1])})$ gegeben. Für $m \in \mathbb{N}$ und $k \in \{1, 2^m\}$ sei $B_{m,k} := ((k-1) \cdot 2^{-m}, k \cdot 2^{-m}]$ und $X_{2^m+k-2} := \mathbb{1}_{B_{m,k}}$.
Dann gilt $X_n \overset{\mathbb{P}}{\longrightarrow} 0$, aber für jedes $\omega \in \Omega$ gilt $\liminf_{n \to \infty} X_n(\omega) = 0$ und $\limsup_{n \to \infty} X_n(\omega) = 1$.

Der nächste Satz wird oft in Verbindung mit der darauffolgenden Anwendung verwendet, um die fast sichere Konvergenz einer Folge von Zufallsvariablen nachzuweisen:

Satz 9.4 (Borel-Cantelli-Lemma) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(E_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ eine Folge von messbaren Ereignissen. Ist die Summe*

$$\sum_{n \in \mathbb{N}} P(E_n)$$

endlich, dann ist die Wahrscheinlichkeit, dass unendlich viele der Ereignisse eintreten gleich 0 :

$$\underbrace{P\left(\bigcap_{n=1}^{\infty}\bigcup_{k\geq n}^{\infty}E_k\right)}_{=\limsup\limits_{n\to\infty}E_n} = 0.$$

Ist dagegen obige Summe unendlich und sind die Ereignisse $(E_n)_{n\in\mathbb{N}}$ *paarweise stochastisch unabhängig, so gilt*

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{k\geq n}^{\infty}E_k\right) = 1.$$

Beweis (Nur erste Aussage). Wenn $\sum\limits_{n\in\mathbb{N}} P(E_n) < \infty$, dann musss notwendigerweise $\sum\limits_{k=n}^{\infty} P(E_k)$ für $n \to \infty$ gegen 0 konvergieren. Damit ist
$P\left(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty}E_k\right) \overset{\text{Stetigkeit von oben}}{=} \lim\limits_{n\to\infty} P\left(\bigcup\limits_{k=n}^{\infty}E_k\right) \overset{\sigma\text{-Subadditivität}}{\leq} \lim\limits_{n\to\infty}\sum\limits_{k=n}^{\infty} P(E_k) = 0.$

Anwendung.
Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge $\mathbb{R}^d$-wertiger Zufallsvariablen und X eine weitere $\mathbb{R}^d$-wertige Zufallsvariable. Gilt für beliebiges $\varepsilon > 0$

$$\sum_{n\in\mathbb{N}} P(||X_n - X|| > \varepsilon) < \infty,$$

so folgt $X_n \overset{P-f.s.}{\longrightarrow} X$.

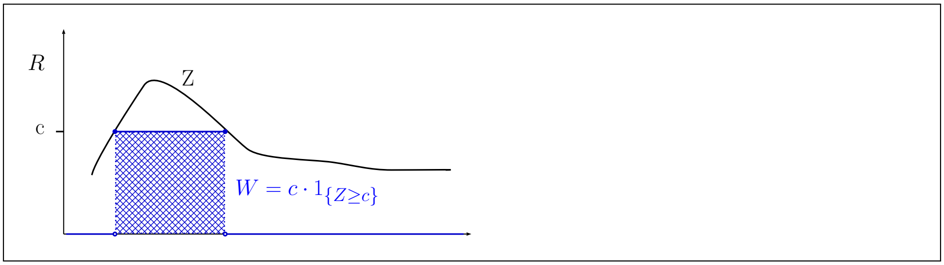
Beweis. Für $\varepsilon \in \mathbb{Q}_{>0}$ fix setze

$$E_n^\varepsilon = \{\omega \in \Omega \mid ||X_n(\omega) - X(\omega)|| > \varepsilon\}.$$

Dann ist $\sum\limits_{n\in\mathbb{N}} P(E_n^\varepsilon) < \infty$ und nach Borel-Cantelli ist somit
$P\left(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty}E_k^\varepsilon\right) = 0$. Nun ist $\{\omega\in\Omega \mid X_n(\omega) \not\to X(\omega)\} = \bigcup\limits_{\varepsilon\in\mathbb{Q}_{>0}}\left(\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty}E_k^\varepsilon\right)$,
also ist $P(\{\omega\in\Omega \mid X_n(\omega) \not\to X(\omega)\}) = 0$.

Einschub: Ungleichungen

Ungleichung von Markov
Gegeben sei ein Maßraum $(\Omega, \mathcal{A}, \mu)$ und wir betrachten eine nichtnegative Funktion $Z : \Omega \longrightarrow \mathbb{R}$. Das Ziel ist es, den Wert $\mu(Z \geq c)$ mit $c > 0$ abzuschätzen.
Idee:



Betrachte $W := c \cdot \mathbb{1}_{\{Z \geq c\}}$. Dann ist $W \leq Z$ und es folgt

$$I_\mu(W) \leq I_\mu(Z),$$

also $I_\mu\left(c \cdot \mathbb{1}_{\{Z \geq c\}}\right) \leq I_\mu(Z)$ bzw. $c \cdot \mu(\{Z \geq c\}) \leq I_\mu(Z)$, also

$$\mu(\{Z \geq c\}) \leq \frac{I_\mu(Z)}{c}.$$

Diese Ungleichung wird als Ungleichung von Markov bezeichnet.
Im Kontext der Wahrscheinlichkeitstheorie würde diese sich in

$$P(\{Z \geq c\}) \leq \frac{\mathbb{E}_P(Z)}{c}$$

übersetzen. Die Ungleichung von Markov hat mehrere Anwendungen, z.B.:

a) Für eine reellwertige Zufallsvariable X mit endlichem Erwartungwert μ und endlicher Varianz definiere $Z := (X - \mathbb{E}_P(X))^2$.
Dann ist

$$P(\{Z \geq c\}) \leq \frac{\mathbb{E}_P(Z)}{c} \iff P(\{(X - E_P(X))^2 \geq c\}) \leq \frac{Var(X)}{c}$$
$$\iff P(\{|X - \mu| \geq \underbrace{\sqrt{c}}_{=:\varepsilon}) \leq \frac{Var(X)}{\varepsilon^2}.$$

Die letzte Ungleichung oben wird als Ungleichung von Tschebyscheff bezeichnet.

b) **Satz.** Aus der Konvergenz im p-ten Mittel folgt die Konvergenz in Wahrscheinlichkeit.

Beweis. Gelte $X_n \xrightarrow{p} X$ und sei $\varepsilon > 0$ beliebig. Betrachte $Z := ||X_n - X||^p$. Dann folgt:

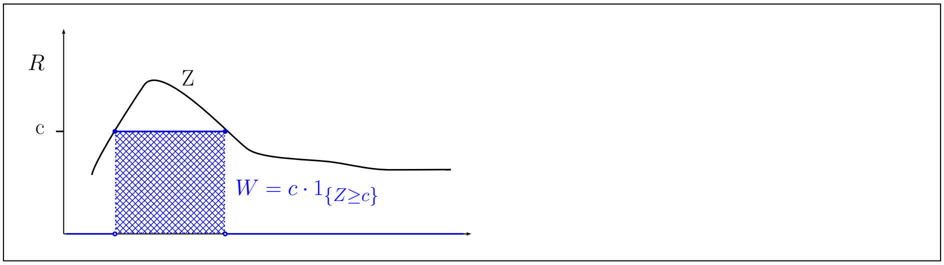
$$P(\{\omega \in \Omega \mid ||X_n(\omega) - X(\omega)|| \geq \varepsilon\}) = P(\{\omega \in \Omega \mid ||X_n(\omega) - X(\omega)||^p \geq \varepsilon^p\})$$
$$\leq \frac{1}{\varepsilon^p} \cdot \underbrace{\mathbb{E}(||X_n - X||^p)}_{\to 0,\text{ da } X_n \xrightarrow{p} X} \longrightarrow 0 \text{ für } n \to \infty$$

Bemerkung 9.5 *Im Allgemeinen gilt nicht $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{p} X$. Beispiel: Für*

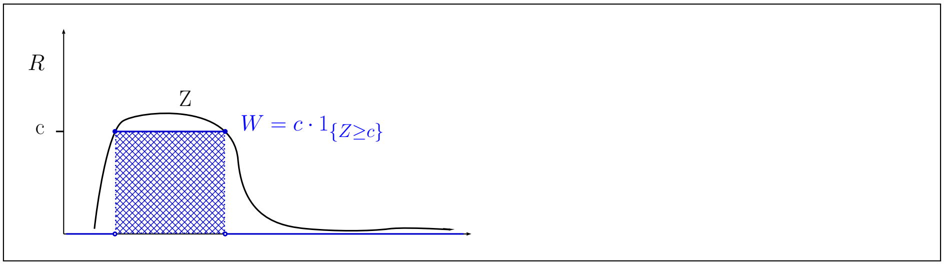
$$X_n = \begin{cases} n & \textit{mit Wahrscheinlichkeit } \frac{1}{n} \\ 0 & \textit{sonst} \end{cases}$$

gilt $X_n \xrightarrow{\mathbb{P}} X = 0$, denn $P(|X_n - 0| > \varepsilon) \leq \frac{1}{n}$. Aber z.B. für $p = 1$ ist $\mathbb{E}_P(|X_n - 0|^1) = n \cdot \frac{1}{n} = 1$, also $X_n \not\xrightarrow{1} X$.
Außerdem gilt im Allgemeinen auch nicht $X_n \xrightarrow{P-f.s.} X \implies X_n \xrightarrow{p} X$, denn zum Beispiel für den Wahrscheinlichkeitsraum $([0,1], \mathfrak{B}([0,1]), \lambda_{|\mathfrak{B}([0,1])})$ und $X_n = n \cdot \mathbb{1}_{[0,\frac{1}{n}]}$ gilt $X_n \xrightarrow{P-f.s.} 0$, aber nicht $X_n \xrightarrow{p} 0$.

c) Allgemein: Idee der Markov-Ungleichung:



Die Ungleichung wird schärfer, wenn wir eher etwa so etwas haben:



Idee: Betrachte nicht die interessierende Zufallsvariable X selbst, sondern eine geeignete (injektive und isotone[17]) Transformation $Z := T \circ X$.

Beispiel: Chernoff-Ungleichung
Für die Analyse einer interessierenden Zufallsvariable X betrachte die Markov-Ungleichung angewandt auf

$$Z := e^{\lambda X}$$

und wähle $\lambda > 0$ derart, dass die entstehende Ungleichung möglichst "scharf" wird.
Beispiel: $X_1, \ldots, X_n$ i.i.d. und $X = \sum_{i=1}^{n} X_i$. Dann ist

$$Z = e^{\lambda X} = e^{\lambda \sum_{i=1}^{n} X_i} = \prod_{i=1}^{n} e^{\lambda X_i}$$

und mit der Markov-Ungleichung angewandt auf Z erhalten wir:

[17] Die Injektivität und die Isotonie der Transformation T sind hilfreich für die äquivalente Umformung des Ereignisses $\{X \geq c\}$ in $\{T \circ X \geq T(c)\}$.

$$
\begin{aligned}
P(X \geq a) &= P(e^{\lambda X} \geq e^{\lambda a}) \\
&\leq \frac{\mathbb{E}(e^{\lambda X})}{e^{\lambda a}} \\
&= \frac{\mathbb{E}\left(\prod_{i=1}^{n} e^{\lambda X_i}\right)}{e^{\lambda a}} \\
&= \frac{\prod_{i=1}^{n} \mathbb{E}\left(e^{\lambda X_i}\right)}{e^{\lambda a}} \\
&= \frac{\left[\mathbb{E}\left(e^{\lambda X_1}\right)\right]^n}{e^{\lambda a}}
\end{aligned}
$$

Anschließend kann $\lambda > 0$ so gewählt werden, dass die rechte Seite obiger Ungleichung möglichst klein wird: Konkret für das Beispiel $X_1, \ldots, X_n$ binomialverteilt mit Erfolgswahrscheinlichkeit[18] $p \geq 0.5$ ergibt sich mit einer Abschätzung über die Exponentialreihe zunächst:

$$
\begin{aligned}
\mathbb{E}\left(e^{\lambda X_1}\right) &= p \cdot e^{\lambda} + (1-p) \cdot 1 \\
&= 1 + p(e^{\lambda} - 1) \\
&\leq \sum_{i=0}^{\infty} \frac{(p(e^{\lambda}-1))^i}{i!} \\
&= e^{p(e^{\lambda}-1)}
\end{aligned}
$$

und damit

$$
P(X \geq a) \leq \frac{\left[e^{p(e^{\lambda}-1)}\right]^n}{e^{\lambda a}} = \frac{\left[e^{(e^{\lambda}-1)}\right]^{np}}{e^{\lambda a}}.
$$

Dasjenige $\lambda > 0$, das die rechte Seite obiger Ungleichung minimiert, ergibt sich hier (nach einer kurzen, hier unterschlagenen Kurvendiskussion) zu

$$
\lambda = \ln\left(\frac{a}{np}\right).
$$

[18] Der Fall $p \leq 0.5$ kann analog abgehandelt werden.

Da sich $\frac{1}{n} \cdot X$ für große n um den Wert p konzentrieren wird, sind für das Ereignis $\{X \geq a\}$ Werte von a in der Umgebung von $a = np$ interessant. Betrachte also in einer multiplikativen Darstellung $a = (1+\delta)np$ mit $\delta > 0$ sehr klein. Dann erhalten wir $\lambda = \ln\left(\frac{(1+\delta)np}{np}\right) = \ln(1+\delta)$ und

$$\begin{aligned} P(X \geq (1+\delta)np) &\leq \frac{\left[e^{\delta}\right]^{np}}{(1+\delta)^{(1+\delta)np}} \\ &= \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{np} \\ &\leq e^{-Cn} \end{aligned}$$

mit einer Konstante $C > 0$, die nur von δ und $p \geq 0.5$, aber insbesondere nicht von n abhängt. Durch eine analoge Analyse des Falls $p \leq 0.5$ kann C auch so gewählt werden, dass die obige Ungleichung völlig unabhängig von $p \in [0,1]$ gilt.

Für den Mittelwert $\bar{X}^{(n)} := \frac{1}{n} \sum_{i=1}^{n} X_i$ ergibt sich daraus

$$P(\bar{X}^{(n)} \geq (1+\delta)p) \leq e^{-Cn}.$$

Mit einer analogen Analyse folgt auch

$$P(\bar{X}^{(n)} \leq (1-\delta)p) \leq e^{-\tilde{C}n}$$

mit $\tilde{C} > 0$ und schließlich

$$P(|\bar{X}^{(n)} - p| \geq \varepsilon) \leq 2e^{-Dn}$$

mit einer Konstanten $D > 0$, die nur von ε abhängt.

Die letzte Ungleichung besagt insbesondere, dass die Wahrscheinlichkeit, dass die relative Häufigkeit $\bar{X}^{(n)}$ von der unbekannten Wahrscheinlichkeit p um mehr als ε abweicht, für $n \to \infty$ mindestens exponentiell schnell gegen Null konvergiert. Dies bedeutet, dass nicht nur $\bar{X}^{(n)} \xrightarrow{\mathbb{P}} p$, sondern wegen Borel-Cantelli auch $\bar{X}^{(n)} \xrightarrow{P-f.s.} p$ gilt.

Bemerkung 9.6 *Die oben angesprochnen Ungleichungen werden auch als Chernoff-Schranken bezeichnet. Etwas allgemeiner ist die Hoeffding-Ungleichung:*

Für $X_1,\ldots,X_n$ stochastisch unabhängig (nicht notwendigerweise identisch verteilt) mit $P(X_1 \in [a_1,b_1], X_2 \in [a_2,b_2],\ldots,X_n \in [a_n,b_n]) = 1$ für geeignete $a_1,\ldots,a_n,b_1,\ldots,b_n \in \mathbb{R}$ gilt:

$$P\left(\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \geq \varepsilon\right) \leq \exp\left(\frac{-2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

bzw.

$$P\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \geq \varepsilon\right) \leq \exp\left(\frac{-2\varepsilon^2 n^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Außerdem gilt auch eine entsprechende Ungleichung für ein Samplen aus einer endlichen Population ohne Zurücklegen, siehe beispielsweise Serfling [1974].

Konvergenz in Verteilung, schwache Konvergenz

Motivation
i) Bei der schwachen Konvergenz interessiert man sich weniger für das Verhalten der gesamten Zufallsvariable X_n als Abbildung, vielmehr ist nur die Verteilung von X_n, also das Bildmaß P_{X_n} von Interesse.

ii) Da das Bildmaß $P_{X_n} : \mathcal{A} \longrightarrow [0,1]$ (,wobei $\mathcal{A}$ meist die Borelsche-σ-Algebra des zugehörigen Wertebereichs der Zufallsvariablen ist,) eine Abbildung ist, könnte eine naheliegende Definition einer schwachen Konvergenz etwa so aussehen:

a) $X_n \longrightarrow X :\Longleftrightarrow \forall A \in \mathcal{A} : P_{X_n}(A) \longrightarrow P_X(A)$.

b) Da das Bildmaß P_{X_n} auch eine Integration bezüglich P_{X_n} zulässt, wäre auch die Definition

$$X_n \longrightarrow X :\Longleftrightarrow \forall f \in \mathcal{F} : I_{P_{X_n}}(f) \longrightarrow I_{P_X}(f)$$

denkbar, wobei man sich über die Klasse $\mathcal{F}$ von Funktionen, für die die Integralbedingung gelten soll, noch Gedanken machen müsste. (Wenn die

Klasse von Indikatorfunktionen von messbaren Mengen in $\mathcal{F}$ enthalten wäre, dann wäre b) eine mindestens so starke Forderung wie a).)

iii) Aber: Bereits die Definition nach a) würde schon zu viel fordern. Betrachte dazu folgendes Beispiel: Sei $X_n \sim U[-\frac{1}{n}, \frac{1}{n}]$. Dann würde man intuitiv erwarten, dass X_n gegen $X = 0$ konvergiert. (In der Tat gilt hier sogar schon $X_n \overset{P-f.s.}{\longrightarrow} X$.) Es ist jedoch $P_{x_n}(\{0\}) = 0$, aber $P_X(\{0\}) = 1$, also gilt nicht $X_n \longrightarrow X$ im Sinne von ii) a). Dies bedeutet, dass man obige Definitionen abschwächen muss, um einen nützlichen Begriff der schwachen Konvergenz zu erhalten. Man kann beispielsweise Bedingung ii) a) nicht für alle $A \in \mathcal{A}$, sondern nur für bestimmte $A \in \mathcal{A}$ fordern. Konkret würde die Einschränkung auf Mengen $A \in \mathcal{A}$ mit $P_X(\partial A) = 0$ zielführend sein. (Achtung: Die Einschränkung bezieht sich nur auf P_X, nicht jedoch auf P_{X_n}.) Einen äquivalenten Begriff der schwachen Konvergenz erhält man, wenn man in ii) b) die Klasse $\mathcal{F}$ aller reellwertigen, stetigen und beschränkten Funktionen heranzieht:

Definition 4 (Konvergenz in Verteilung, schwache Konvergenz).
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge $\mathbb{R}^d$-wertiger Zufallsvariablen und X eine weitere $\mathbb{R}^d$-wertige Zufallsvariable. Wir sagen, dass X_n in Verteilung (oder schwach) gegen X konvergiert, falls für jede stetige und beschränkte (und damit $\mathfrak{B}(\mathbb{R}^d) - \mathfrak{B}(\mathbb{R})$-messbare und integrierbare) Funktion $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ gilt:

$$I_{P_{X_n}}(f) \longrightarrow I_{P_X}(f).$$

Schreibweise: $X_n \xrightarrow{\mathcal{D}} X$.

Bemerkung 9.7 *Da bezüglich der Bildmaße integriert wird, spielen nur die Verteilungen von X_n bzw. X eine Rolle, nicht jedoch die genauen Abbildungen $X_n : \Omega \longrightarrow \mathbb{R}^d$ bzw. $X : \Omega \longrightarrow \mathbb{R}^d$, insbesondere auch nicht, für welche ω genau welche Werte angenommen werden. Ein einfaches Beispiel dazu: $\Omega = \{a,b\}$; $P(\{a\}) = P(\{b\}) = 0.5$ und*

$$X_n : \quad \begin{array}{rl} a \mapsto & -1 \\ b \mapsto & 1 \end{array}$$

$$X : \quad \begin{array}{rl} a \mapsto & 1 \\ b \mapsto & -1 \end{array}.$$

Dann ist $P_{X_n} = P_X$ und somit $X_n \xrightarrow{\mathcal{D}} X$, aber $X_n \overset{\mathbb{P}}{\not\longrightarrow} X$ und $X_n \overset{P-f.s.}{\not\longrightarrow} X$ sowie $X_n \overset{p}{\not\longrightarrow} X$.

Die schwache Konvergenz $X_n \xrightarrow{\mathcal{D}} X$ kann man sich in etwa als "$X_n \longrightarrow X$ in Bezug auf jeden "stetigen, beschränkten" Verteilungsaspekt" vorstellen. Beispielsweise für reellwertige Zufallsvariablen mit $X_n \xrightarrow{D} X$ (die Existenz und Endlichkeit der involvierten Integrale jeweils vorausgesetzt) gilt unter der technischen Voraussetzung, dass ein Y existiert mit

$$\begin{aligned} |X_n| &\leq Y \quad P-f.s. \\ |X| &\leq Y \quad P-f.s. \\ \mathbb{E}(Y) &< \infty, \end{aligned}$$

das

$$\mathbb{E}(X_n) = I_{P_{X_n}}(id_{\mathbb{R}}) = \lim_{k\to\infty} I_{P_{X_n}}(id_{\mathbb{R}}^k) \overset{n\to\infty}{\longrightarrow} \lim_{k\to\infty} I_{P_X}(id_{\mathbb{R}}^k) = I_{P_X}(id_{\mathbb{R}}) = \mathbb{E}(X)$$

ist, wobei hier $id_{\mathbb{R}}^k$ die gestutzte identische Abbildung

$$id_{\mathbb{R}}^k : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto \max\{-k, \min\{x, k\}\}$$

ist, und diese Abbildung ist offensichtlich stetig und beschränkt. Mit einer gleichen Argumentation folgt, dass (unter einer entsprechend modifizierten technischen Voraussetzung) $Var(X_n)$ gegen $Var(X)$ konvergiert, gleiches gilt auch für höhere Momente von X_n bzw. X (immer vorausgesetzt, dass diese existieren und endlich sind und das die entsprechenden Potenzen von X_n bzw. X gleichmäßig von einem Y mit existierender endlicher Erwartung dominiert werden).

Der folgende Satz charakterisiert für reellwertige Zufallsvariablen die schwache Konvergenz über die Verteilungsfunktion:

Satz 9.8 (Charakterisierung der schwachen Konvergenz) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger Zufallsvariablen und X eine weitere reellwertige Zufallsvariable. Dann gilt:*

$$X_n \xrightarrow{\mathcal{D}} X \iff \text{Für alle Stetigkeitsstellen } c \text{ von } F_X \text{ gilt:}$$
$$\lim_{n\to\infty} \underbrace{F_{X_n}(c)}_{=P_{X_n}(]-\infty,c])} = \underbrace{F_X(c)}_{=P_X(]-\infty,c])}.$$

Beispiel 9.9 *Sei X_n gegeben durch die Verteilungsfunktion F_{X_n}:*

$$F_{X_n}(x) = \begin{cases} 0 & \text{falls } x < 0 \\ x & \text{falls } x \in [0, 1 - \frac{1}{n}[\\ 1 - \frac{1}{n} & \text{falls } x \in [1 - \frac{1}{n}, n[\\ 1 & \text{sonst} \end{cases}$$

Dann konvergiert X_n in Verteilung gegen eine auf dem Intervall $[0,1]$ stetig gleichverteilte Zufallsvariable X mit Verteilungsfunktion $F_X(x) := \begin{cases} 0 & \text{falls } x < 0 \\ x & \text{falls } x \in [0,1] \\ 1 & \text{sonst} \end{cases}$. Man beachte, dass hier $\mathbb{E}(X) = 1/2$ ist, aber da F_{X_n} einen Sprung der Höhe $1/n$ an der Stelle $x = n$ besitzt, gilt $\mathbb{E}(X_n) \geq 1/n \cdot n = 1$, d.h., dass der Erwartungswert von X_n hier nicht gegen den Erwartungswert von X konvergiert. Dies bedeutet, dass für die Konvergenz von Momenten die obige technische Voraussetzung im Zweifelsfall wirklich benötigt wird.

Satz 9.10 (Schwache Konvergenz und stetige Abbildungen) *Für eine stetige Abbildung $g : \mathbb{R}^d \longrightarrow \mathbb{R}^q$ gilt:*

$$X_n \xrightarrow{\mathcal{D}} X \quad \Longrightarrow \quad g \circ X_n \xrightarrow{\mathcal{D}} g \circ X.$$

Bemerkung 9.11 *Analoge Aussagen gelten für die fast sichere Konvergenz und die Konvergenz in Wahrscheinlichkeit.*

Abschließend wollen wir noch wesentliche Zusammenhänge zwischen den Konvergenzbegriffen, die wir größtenteils bereits gezeigt haben, in einem Satz zusammenfassen:

Satz 9.12 (Zusammenhang zwischen den Konvergenzbegriffen) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $(X_n)_{n\in\mathbb{N}}$ eine Folge $\mathbb{R}^d$-wertiger Zufallsvariablen und X eine weitere $\mathbb{R}^d$-wertige Zufallsvariable. Dann gilt:*

i) $X_n \xrightarrow{p} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X.$

ii) *Existiert ein Y mit $P(|X_n| \leq Y) = 1$ für alle $n \in \mathbb{N}$ und $\mathbb{E}(|Y|^p) < \infty$, so gilt auch* $X_n \xrightarrow{\mathbb{P}} X \quad \Longrightarrow \quad X_n \xrightarrow{p} X.$

iii) $X_n \xrightarrow{P-f.s} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X.$

iv) $X_n \xrightarrow{\mathbb{P}} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathcal{D}} X.$

v) Falls X eine konstante Zufallsvariable ist, so gilt auch

$$X_n \xrightarrow{\mathcal{D}} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X.$$

vi) (Skorochod-Darstellung) Gilt $X_n \xrightarrow{\mathcal{D}} X$*, so gibt es einen weiteren Wahrscheinlichkeitsraum* $(\Omega', \mathcal{A}', P')$ *und Zufallsvariablen* $X_n : \Omega' \longrightarrow \mathbb{R}^d, X : \Omega' \longrightarrow \mathbb{R}^d$ *mit* $X_n' \stackrel{d}{=} X$ *(meint* $P_{X_n} = P_X$*) für* $n \in \mathbb{N}$ *und* $X' \stackrel{d}{=} X$*, sowie*

$$X_n' \xrightarrow{P'-f.s.} X'.$$

Kapitel 4
Bedingte Erwartung und bedingte Verteilung

Die gewöhnliche bedingte Wahrscheinlichkeit und der gewöhnliche bedingte Erwartungswert

Definition 1. (Gewöhnliche bedingte Wahrscheinlichkeit) Sei jetzt immer $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Für $B \in \mathcal{A}$ mit $P(B) > 0$ definieren wir die gewöhnliche Wahrscheinlichkeit gegeben B als

$$P(\cdot \mid B) : \mathcal{A} \longrightarrow \mathbb{R} : A \mapsto P(A \mid B) := \frac{P(A \cap B)}{P(B)}.$$

Bemerkung 9.13 *Die Abbildung $P(\cdot \mid B)$ ist selbst wieder ein Wahrscheinlichkeitsmaß. Das zugehörige Erwartungswertfunktional sei mit $\mathbb{E}_{(\cdot|B)}$ bezeichnet. Für $X \in \mathfrak{L}^1(P)$ gilt dann:*

$$\mathbb{E}(X \mid B) := \mathbb{E}_{(\cdot|B)}(X) = \frac{\mathbb{E}(X \cdot \mathbb{1}_B)}{P(B)} = \frac{\mathbb{E}(X \cdot \mathbb{1}_B)}{\mathbb{E}(\mathbb{1}_B)}.$$

Das Ziel wird nun unter anderem sein, für Ereignisse $B \in \mathcal{A}$ mit $P(B) = 0$ ebenfalls einen bedingten Erwartungswert bzw. eine bedingte Verteilung zu definieren. Dazu ist es nötig, nicht nur ein einzelnes bedingendes Ereignis B, sondern gleich eine ganze Familie von bedingenden Ereignissen zu betrachten. Diese Familie wird im Folgenden immer eine Unter-σ-Algebra der unterliegenden σ-Algebra $\mathcal{A}$ sein, die wir meist mit Σ bezeichnen werden. Zunächst betrachten wir den einfachen Fall, dass der Grundraum Ω von Σ in eine abzählbare Partition von Äquivalenzklassen positiver Wahrscheinlichkeit zerlegt wird:

Definition 2. (Bedingte Erwartung für den Fall eines abzählbaren Faktorraums von nicht-null-Ereignissen)
Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine weitere σ-Algebra Σ mit $\Sigma \subseteq \mathcal{A}$ (, d.h., alle $A \in \Sigma$ sind messbar bezüglich $\mathcal{A}$) und mit $\Omega_{/\sim_\Sigma} = \{A_1, A_2, \ldots\}$ höchstens abzählbar und $P(A_i) > 0$ für alle $A_i \in \Omega_{/\sim_\Sigma}$.
Dann ist die bedingte Erwartung $\mathbb{E}(X \mid \Sigma) : \Omega \longrightarrow \mathbb{R}$ gegeben durch

$$\mathbb{E}(X \mid \Sigma) := \sum_{A_i \in \Omega_{/\sim_\Sigma}} \frac{\mathbb{E}(X \cdot \mathbb{1}_{A_i})}{P(A_i)} \cdot \mathbb{1}_{A_i}.$$

Achtung: $\mathbb{E}(X \mid \Sigma)$ ist kein Wert, sondern eine Zufallsvariable. Konkret ist $\mathbb{E}(X \mid \Sigma)$ eine Zufallsvariable, die auf den Äquivalenzklassen A_i konstant ist, für $\omega \in A_i$ gilt

$$(\mathbb{E}(X \mid \Sigma))(\omega) = \mathbb{E}(X \mid A_i).$$

Bemerkung 9.14 *Für die bedingte Erwartung $\mathbb{E}(X \mid \Sigma)$ gilt hier*

$$\mathbb{E}(\mathbb{E}(X \mid \Sigma) \cdot \mathbb{1}_{A_i}) = \mathbb{E}(X \cdot \mathbb{1}_{A_i}),$$

die bedingte Erwartung verhält sich auf den A_i (und damit auch auf allen $A \in \Sigma$) in Erwartung wie X. Diese Eigenschaft wird auch später im Fall, wo Σ nicht so einfach ist und wo Ereignisse $A \in \Sigma$ auch Wahrscheinlichkeit 0 haben können, als Definition für die bedingte Erwartung herhalten müssen. Bevor wir dazu kommen, wollen wir jedoch noch eine geometrische Vorstellung von der bedingten Erwartung als Projektion in einem Hilbertraum betrachten:

Die Bedingte Erwartung als Projektion in einem Hilbertraum

Zunächst: Der übliche Erwartungswert minimiert die Funktion

$$f : \mathbb{R} \longrightarrow \mathbb{R} : c \mapsto \mathbb{E}((X - c)^2).$$

Man kann sich den Erwartungswert also als denjenigen (eindeutig bestimmten) Wert c einer konstanten Zufallsvariable vorstellen, die den kleinsten (semi-)Abstand im Sinne der $\mathfrak{L}^2$- Halbnorm zu X besitzt. Anders ausgedrückt: Der Erwartungswert einer Zufallvariablen X ist beschrieben durch die (in diesem Fall eindeutig bestimmte) Projektion von X auf den Unterraum aller konstanten Zufallsvariablen. Diese Vorstellung legt die Verallgemeinerung auf Projektionen auf beliebige Unterräume nahe.

Einige Fakten zur Projektion auf Unterräume (Vergleiche auch die Vorlesungen zur linearen Algebra und zu linearen Modellen)
Gegeben sei ein (reeller) halbnormierter Raum $(M, ||\cdot||)$, dessen Halbnorm von einem semidefiniten Skalarprodukt $\langle\cdot,\cdot\rangle$ erzeugt wird, und für den der Faktorraum $(M||\cdot||)_{/\sim}$ mit

$$\sim := \{(X,Y) \in M^2 \mid ||X-Y|| = 0\}$$

ein Hilbertraum (also insbesondere vollständig) ist. Sei weiter $U \subseteq M$ ein abgeschlossener Unterraum von M. Dann gilt der folgende

Satz 9.15 *(Projektion auf abgeschlossene Unterräume) Für $x \in M$ und $x_P \in U$ ist äquivalent:*

i) $||x_P - x|| = \inf_{y \in U} ||y - x||$.
ii) $\langle x_P - x, y\rangle = 0$ *für beliebige* $y \in U$.

Betrachten wir jetzt wie oben den Fall einer σ-Algebra $\Sigma \subseteq \mathcal{A}$, für die der Faktorraum $\Omega_{/\sim_\Sigma} = \{A_1, A_2, \ldots\}$ höchstens abzählbar ist. Dann ist der Raum aller quadratintegrierbaren Zufallsvariablen, die zusätzlich Σ-messbar sind (, d.h. in diesem Fall genau, dass sie auf allen A_i konstant sind,) ein Unterraum von $\mathfrak{L}^2(\Omega, \mathcal{A}, P)$, den wir mit U_Σ bezeichnen wollen. Wie sieht nun die Projektion X_P einer Zufallsvariable $X \in \mathfrak{L}^2((\Omega, \mathcal{A}, P))$ auf U_Σ aus?

Für die Projektion X_P und für beliebige Σ-messbare $Y \in \mathfrak{L}^2(\Omega, \mathcal{A}, P))$ muss

$$\langle X_P - X, Y\rangle = \mathbb{E}((X - X_P) \cdot Y) = 0$$

gelten. Speziell für $Y := \mathbb{1}_A$ mit $A \in \Sigma$ müsste notwendigerweise

$$\mathbb{E}((X_P - X) \cdot \mathbb{1}_A) = \mathbb{E}(X_P \cdot \mathbb{1}_A - X \cdot \mathbb{1}_A) = 0,$$

bzw.

$$\mathbb{E}(X_P \cdot \mathbb{1}_A) = \mathbb{E}(X \cdot \mathbb{1}_A) \quad (*)$$

gelten, d.h., auf allen $A \in \Sigma$ muss sich X_P in Erwartung genau so wie X verhalten. Diese notwendige Bedingung ist auch hinreichend dafür, dass X_P eine Projektion von X auf U_Σ ist (im Sinne von $||X_P - X||_2 = \inf_{y \in U} ||Y - X||_2$). Für die einfache Situation eines abzählbaren Faktorraums $\Omega_{/\sim_\Sigma}$ mit Äquivalenzklassen positiver Wahrscheinlichkeit ist die Projektion X_P sogar bereits dadurch bestimmt, dass X_P auf allen A_i konstant den Wert

$$X_P(\omega) = \mathbb{E}(X \cdot \mathbb{1}_{A_i}) \text{ für } \omega \in A_i$$

besitzt, was genau der Definition der bedingten Ewartung $\mathbb{E}(X \mid \Sigma)$ in dieser einfachen Situation entspricht. Die Darstellung als Projektion bzw. die Bedingung $(*)$ ist nun auch auf beliebige σ-Algebren $\Sigma \subseteq \mathcal{A}$ übertragbar, was zur folgenden allgemeinen Definition des bedingten Erwartungswertes herangezogen werden kann:

Definition 3. (Bedingte Erwartung, allgemeiner Fall) Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\Sigma \subseteq \mathcal{A}$ und $X \in \mathfrak{L}^1(\Omega, \mathcal{A}, P)$. Dann heißt $Y : \Omega \longrightarrow \mathbb{R}$ eine bedingte Erwartung von X gegeben Σ, falls gilt:

i) Y ist Σ-messbar.
ii) Für alle $A \in \Sigma$ gilt $\mathbb{E}(Y \cdot \mathbb{1}_A) = \mathbb{E}(X \cdot \mathbb{1}_A)$.

Satz 9.16 (Existenz und Eindeutigkeit der bedingten Erwartung) *Es sei $X \in \mathfrak{L}^1(\Omega, \mathcal{A}, P)$ und $\Sigma \subseteq \mathcal{A}$ eine σ-Algebra. Dann existiert die bedingte Erwartung Y von X gegeben Σ und diese ist P-fast sicher eindeutig bestimmt, d.h., für jede weitere Zufallsvariable Z, die ebenfalls i) und ii) erfüllt, gilt $P(Y = Z) = 1$. Die bedingte Erwartung von X gegeben Σ werden wir im Folgenden mit $\mathbb{E}(X \mid \Sigma)$ bezeichnen. Da die bedingte Erwartung von X gegeben Σ nur P-fast sicher eindeutig bestimmt ist, sprechen wir auch von einer Version der bedingten Erwartung von X gegeben Σ.*

Bemerkung 9.17 *Da die bedingte Erwartung über Integrale definiert ist, ist wirklich jede beliebige Σ-messbare Zufallsvariable, die P-fast sicher mit der bedingten Erwartung von X gegeben Σ übereinstimmt, ebenfalls eine Version der bedingten Erwartung.*

Satz 9.18 (Eigenschaften bedingter Erwartungen) *Es gilt:*

i) *Ist X eine Σ-messbare Zufallsvariable, dann gilt $\mathbb{E}(X \mid \Sigma) = X$, $P - f.s.$*
ii) *Linearität: $\mathbb{E}(aX + bY \mid \Sigma) = a\mathbb{E}(X \mid \Sigma) + b\mathbb{E}(Y \mid \Sigma)$, $P - f.s.$*
iii) *Monotonie $X \leq Y$, $P - f.s. \Longrightarrow \mathbb{E}(X \mid \Sigma) \leq \mathbb{E}(Y \mid \Sigma)$, $P - f.s.$*
iv) $\mathbb{E}(\mathbb{E}(X \mid \Sigma)) = \mathbb{E}(X)$.
v) *Falls X Σ-messbar und $\mathbb{E}(|Y|) < \infty$ und $\mathbb{E}(|X \cdot Y|) < \infty$, so gilt $\mathbb{E}(XY \mid \Sigma) = X\mathbb{E}(Y \mid \Sigma)$, $P - f.s.$*
vi) *Monotone Konvergenz: Ist $X_n \geq 0$ und $X_n \nearrow X$ mit $\mathbb{E}(X) < \infty$, so gilt $\mathbb{E}(X_n \mid \Sigma) \nearrow \mathbb{E}(X \mid \Sigma)$, $P - f.s.$*
vii) *Turmeigenschaft: Für $\Sigma_1 \subseteq \Sigma_2 \subseteq \mathcal{A}$ σ-Algebren gilt*

$$\mathbb{E}(\mathbb{E}(X \mid \Sigma_1)\Sigma_2) = \mathbb{E}(\mathbb{E}(X \mid \Sigma_2)\Sigma_1) = \mathbb{E}(X \mid \Sigma_1),\ P - f.s.$$

viii) *Falls $X \in \mathfrak{L}^2$, dann gilt für $Y^* := \mathbb{E}(X \mid \Sigma)$ und $M = \{Y \in \mathfrak{L}^2,\ Y \text{ ist } \Sigma\text{-messbar}\}$ das*

$$\mathbb{E}((X - Y^*)^2) \leq \mathbb{E}((X - Y)^2)$$

für jedes $Y \in M$.

Wir verfolgen nun das Ziel, für Zufallsvariablen X, Y einen bedingetn Erwatungswert $\mathbb{E}(X \mid Y = y)$ bzw. eine bedingte Wahrscheinlichkeit $P(X \in B \mid Y = y)$ zu definieren.

Definition 4. Für $X, Y \in \mathfrak{L}^1$ definiere

$$\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \sigma(Y))$$

mit $\sigma(Y) := Y^{-1}[\mathfrak{B}(\mathbb{R})] = \{Y^{-1}(B) \mid B \in \mathfrak{B}(\mathbb{R})\}$.

Bemerkung 9.19 *Zur „Interpretation“: $\mathbb{E}(X \mid Y)$ kann man sich als beste Schätzung von X vorstellen, gegeben mann kennt Y. Es gibt zwei Extremfälle: X, Y stochastisch unabhängig, dann gilt $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$, $P - f.s.$ versus $X = f(Y)$ mit f Funktion, dann ist $\mathbb{X}(X \mid Y) = X = f(Y)$, $P - f.s.$*

Gegeben zwei Zufallsvariablen X, Y ermöglicht der folgende Satz die Definition eines bedingten Erwartungswertes $\mathbb{E}(X \mid Y = y)$ auf Basis der bedingten Erwartung $\mathbb{E}(X \mid \sigma(Y))$, die ja noch eine Zufallsvariable, und kein Wert ist:

Satz 9.20 *Ist X eine (reellwertige) $\sigma(Y)$-messbare Zufallsvariable, so existiert eine $\mathfrak{B}(\mathbb{R})$-messbare Funktion $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ mit $X = \varphi \circ Y$.*

Definition 5 (Bedingter Erwartungswert). Seien X,Y reelwertige Zufallsvariablen und sei φ derart, dass

$$\varphi \circ Y = \underbrace{\mathbb{E}(X \mid Y)}_{\text{ist } \sigma(Y)\text{-messbar}}$$

gilt. Dann definieren wir $\mathbb{E}(X \mid Y) := \varphi(y)$ und nennen diesen Wert den (bzw. einen) bedingten Erwartungswert von X gegeben $Y = y$.

Bemerkung 9.21 *Für reellwertige Zufallsvariablen X,Y und $\Sigma := \sigma(Y)$ sowie $\tilde{X}$ eine Version der bedingten Erwartung von X gegeben Σ gilt*

$$\begin{aligned}\mathbb{E}(X \cdot \mathbb{1}_{\{Y \in B\}}) &= \mathbb{E}(\tilde{X} \cdot \mathbb{1}_{\{Y \in B\}}) = \mathbb{E}((\varphi \circ Y) \cdot \mathbb{1}_{\{Y \in B\}}) \\ &= \int\limits_B \varphi(y) dP_Y = \int_B \mathbb{E}(X \mid Y = y) dP_Y\end{aligned}$$

für beliebiges $B \in \mathfrak{B}(\mathbb{R})$.

Definition 6 (Bedingte Verteilung). Für $B \in \mathcal{A}$ setze $P(B \mid Y = y) := \mathbb{E}(\mathbb{1}_B \mid Y = y)$. und für $\Sigma := \sigma(Y)$ setze $P(B \mid \Sigma) := \mathbb{E}(\mathbb{1}_B \mid \Sigma)$.

Bemerkung 9.22 *$P(B \mid Y = y)$ ist eine Zahl und $P(B \mid \Sigma)$ ist eine Zufallsvariable.*

Frage: Ist für alle $\omega \in \Omega$ die Funktion $(P(\cdot \mid \Sigma))(\omega)$ ein Wahrscheinlichkeitsmaß? Für eine Folge $(B_n)_{n \in \mathbb{N}}$ paarweise disjunkter Mengen gilt zwar

$$P\left(\bigcup_{i=1}^{\infty} B_i \mid \Sigma\right) = \sum_{i=1}^{\infty} P(B_1 \mid \Sigma)$$

bis auf eine Ausnahmemenge vom Maß 0. Wenn es aber überabzählbar viele solcher disjunkter Folgen von gewissen B_n's gibt, dann kann es vorkommen, dass die gesamte Menge aller Ausnahmen keine Nullmenge mehr ist. Natürlich hätten wir gern, dass wie gewohnt $(P(\cdot \mid \Sigma))(\omega)$ ein Wahrscheinlichkeitsmaß ist und etablieren deshalb folgende

Definition 7 (Reguläre Version der bedingten Wahrscheinlichkeit bzw. Verteilung). Eine Funktion $\tilde{P} : \Omega \times \Sigma \longrightarrow [0,1]$ heißt (eine) reguläre Version der bedingten Wahrscheinlichkeit gegeben Σ, falls gilt:

a) $\forall \omega \in \Omega : \tilde{P}(\omega, \cdot)$ ist ein Wahrscheinlichkeitsmaß
b) $\forall B \in \Sigma : \tilde{P}(\cdot, B)$ ist eine Version von $P(B \mid \Sigma)$.

Analog ist $\tilde{P} : \Omega \times \mathfrak{B}(\mathbb{R}) \longrightarrow [0,1]$ die (eine) reguläre Version der bedingten Verteilung von X gegeben Σ, falls

a) $\forall \omega \in \Omega : \tilde{P}(\omega, \cdot)$ ist ein Wahrscheinlichkeitsmaß
b) $\forall B \in \mathfrak{B}(\mathbb{R}) : \tilde{P}(\cdot, B)$ ist eine Version von $P(X \in B \mid \Sigma)$.

Satz 9.23 (Eistenz einer regulären Version der bedingten Verteilung)
Für reellwertige Zufallsvariablen existiert eine reguläre Version der bedingten Verteilung. Für diese gilt dann

$$\mathbb{E}(X \mid \Sigma)(\omega) = \int_{\Omega} X(\omega') d\tilde{P}(\omega, \cdot), \; P - f.s.$$

Kapitel 5
Produktmaß und stochastische Unabhängigkeit

Satz 9.24 (Produkt-σ-Algebra) *Seien $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2), \ldots, (\Omega_p, \mathcal{A}_p)$ Messräume. Sei für $i = 1, \ldots, p$ jeweils $\mathcal{E}_i$ ein Erzeuger der σ-Algebra A_i, der Ω_i enthalte. (Beispielsweise für $(\Omega_1, \mathcal{A}_1) = (\Omega_2, \mathcal{A}_2) = \ldots = (\Omega_p, \mathcal{A}_p) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ denke man an $\mathcal{E}_i = \{(-\infty, c] \mid c \in \overline{R}\}$.) Sei*

$$\Omega := \Omega_1 \times \Omega_2 \times \ldots \times \Omega_p$$

und sei

$$\pi_i : \Omega \longrightarrow \Omega_i : \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_p \end{pmatrix} \mapsto \omega_i$$

die Projektion auf die i-te Komponente. Dann erzeugen die folgenden Mengensysteme die gleiche σ-Algebra auf Ω:

$$\mathcal{D}_1 := \{A_1 \times A_2 \times \ldots \times A_p \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, \ldots, A_p \in \mathcal{A}_p\}$$
$$\mathcal{D}_2 := \{\pi_i^{-1}(A) \mid i \in \{1, \ldots, p\}, A \in \mathcal{A}_i\}$$
$$\mathcal{D}_3 := \{\pi_i^{-1}(A) \mid i \in \{1, \ldots, p\}, A \in \mathcal{E}_i\}$$
$$\mathcal{D}_4 := \{A_1 \times A_2 \times \ldots \times A_p \mid A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2, \ldots, A_p \in \mathcal{E}_p\}$$

Die von $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ bzw. $\mathcal{D}_4$ erzeugte σ-Algebra nennen wir die Produkt-σ-Algebra (von $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_p$) auf $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_p$ und notieren sie mit

$$\bigotimes_{i=1}^{p} \mathcal{A}_i.$$

Bemerkung 9.25 *Das kartesische Produkt* $\bigtimes_{i=1}^{p} \mathcal{A}_i$ *ist nicht mit der Produkt-σ-Algebra*

$$\bigotimes_{i=1}^{p} \mathcal{A}_i = \sigma\left(\{A_1 \times \ldots \times A_p \mid A_1 \in \mathcal{A}_1, \ldots, A_p \in \mathcal{A}_p\}\right)$$

(und auch nicht mit $\{A_1 \times \ldots \times A_p \mid A_1 \in \mathcal{A}_1, \ldots, A_p \in \mathcal{A}_p\} \subseteq 2^{\Omega} = 2^{\Omega_1 \times \ldots \times \Omega_p}$*) zu verwechseln.*
Aus der Darstellung der Produkt-σ-Algebra über $\mathcal{D}_2$ sieht man leicht, dass $\bigotimes_{i=1}^{p} \mathcal{A}_i$ *die kleinste σ-Algebra auf Ω ist, bezüglich derer alle Projektionen $\pi_1, \ldots, \pi_p$ messbar sind. Diese Charakterisierung würde man auch benutzen, um für den Fall einer unendlichen Familie von Messräumen die Produkt-σ-Algebra zu definieren. Achtung: Für den Fall einer überabzählbaren Familie von Messräumen fallen z.B. $\sigma(\mathcal{D}_1)$ und $\sigma(\mathcal{D}_2)$ im Allgemeinen nicht mehr zusammen.*

Bemerkung 9.26 *Achtung: Die Eigenschaft, dass alle Projektionen π_i messbar sind, bedeutet lediglich, dass die Urbilder von $\mathcal{A}_i$-messbaren Mengen A_i unter π_i ein bezüglich der Produkt-σ-Algebra messbares Urbild besitzen. Es kann erstaunlicherweise vorkommen, dass das Bild einer bezüglich der Produkt-σ-Algebra messbaren Menge unter der so unscheinbaren Projektionsabbildung π_i eine nicht $\mathcal{A}_i$-messbare Menge ist.*

Beweis. Zunächst: Für eine σ-Algebra $\mathcal{A}$ auf Ω gilt

$$\begin{aligned}
\mathcal{A} \supseteq \mathcal{D}_2 &\iff \text{alle Projektionen } \pi_i \text{ sind } \mathcal{A} - \mathcal{A}_i\text{-messbar} \\
&\iff \text{alle } \pi_i^{-1}(A) \text{ mit } A \in \mathcal{E}_i \text{ sind aus } \mathcal{A} \\
&\iff \mathcal{A} \supseteq \mathcal{D}_3.
\end{aligned}$$

Nun ist:

i) $\mathcal{D}_3 \subseteq \mathcal{D}_2 \Longrightarrow \sigma(\mathcal{D}_3) \subseteq \sigma(\mathcal{D}_\in)$

ii) $\sigma(\mathcal{D}_3) \supseteq \mathcal{D}_3 \Longrightarrow_{\text{(mit obiger Bemerkung)}} \sigma(\mathcal{D}_3) \supseteq \mathcal{D}_2 \Longrightarrow \sigma(\mathcal{D}_3) \supseteq \sigma(\mathcal{D}_2)$

Aus *i*) und *ii*) folgt somit

$$\sigma(\mathcal{D}_2) = \sigma(\mathcal{D}_3).$$

(Man hätte hier auch kürzer argumentieren können, dass nach Konstruktion sowohl $\mathcal{D}_2$, als auch $\mathcal{D}_3$ die kleinste $\sigma-Algebra$ ist, für die alle Projektionen messbar sind, und dass deshalb $\mathcal{D}_2$ und $\mathcal{D}_3$ identisch sind, vergleiche Satz 3.2 zur Charakterisierung der Messbarkeit.)

Außerdem gilt

$$\sigma(\mathcal{D}_2) = \sigma(\mathcal{D}_3) \subseteq \sigma(\mathcal{D}_1),$$

denn $\mathcal{D}_2 \subseteq \mathcal{D}_1$, denn $\pi_i^{-1}(A) = \Omega_1 \times \ldots \times \Omega_{i-1} \times A \times \Omega_{i+1} \times \ldots \times \Omega_p$. Weiterhin gilt $\mathcal{D}_1 \subseteq \sigma(\mathcal{D}_2)$, woraus $\sigma(\mathcal{D}_1) \subseteq \sigma(\mathcal{D}_2)$ und damit insgesamt

$$\sigma(\mathcal{D}_1) = \sigma(\mathcal{D}_2) = \sigma(\mathcal{D}_3)$$

folgt.

Abschließend ist $\mathcal{D}_4 \subseteq \sigma(\mathcal{D}_3)$, also $\sigma(\mathcal{D}_4) \subseteq \sigma(\mathcal{D}_3)$, sowie $\mathcal{D}_3 \subseteq \mathcal{D}_4$, also $\sigma(\mathcal{D}_3) \subseteq \sigma(\mathcal{D}_4)$, d.h., insgesamt gilt

$$\sigma(\mathcal{D}_1) = \sigma(\mathcal{D}_2) = \sigma(\mathcal{D}_3) = \sigma(\mathcal{D}_4).$$

Satz 9.27 (Messbarkeit bezüglich der Produkt-σ-Algebra) *Seien* $(\Omega, \mathcal{A})$ *und* $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2), \ldots, (\Omega_p, \mathcal{A}_p)$ *Messräume, sowie*

$$X : \Omega \longrightarrow \bigtimes_{i=1}^{p} \Omega_i : \omega \mapsto \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_p(\omega) \end{pmatrix}.$$

Dann gilt:

X ist $\mathcal{A} - \bigotimes_{i=1}^{p} \mathcal{A}_i$-messbar genau dann, wenn für alle $i \in \{1, \ldots, p\}$ die Abbildung $X_i : \Omega \longrightarrow \Omega_i : \omega \mapsto X_i(\omega)$ eine $\mathcal{A} - \mathcal{A}_i$-messbare Abbildung ist.

Zum Beweis. Der Beweis läuft völlig analog zum Beweis des Satzes zur Charakterisierung der $\mathfrak{B}(\mathbb{R}^p)$-Messbarkeit, man ersetze lediglich in der Hinrichtung “π_i stetig und deshalb $\mathfrak{B}(\mathbb{R}^p) - \mathfrak{B}(\mathbb{R})$-messbar” durch “$\pi_i$ ist $\bigotimes_{k=1}^{p} \mathcal{A}_k - \mathcal{A}_i$-messbar nach Definition der Produkt-σ-Algebra”. Für die Rückrichtung ersetze man den Erzeuger $\mathcal{F} = \{(-\infty, c] \mid c \in \mathbb{R}^p\}$ durch beispielsweise den Erzeuger $\mathcal{D}_4$.

Frage: Wir hatten $\mathfrak{B}(\mathbb{R}^p)$ betrachtet. Wie verhält sich $\mathfrak{B}(\mathbb{R}^p)$ zur Produkt-σ-Algebra $\bigotimes_{i=1}^{p} \mathfrak{B}(\mathbb{R})$?

Es ist

$$\bigotimes_{i=1}^{p} \mathfrak{B}(\mathbb{R}) = \sigma\left(\{(a_1,b_1) \times (a_2,b_2) \times \ldots \times (a_p,b_p) \mid a_1,\ldots,a_p,b_1,\ldots,b_p \in \overline{\mathbb{R}}\}\right)$$

die von den offenen Hyperrechtecken erzeugte σ-Algebra. Nun gilt für den $\mathbb{R}^p$, dass beispielsweise die euklidische Metrik d und die Supremumsmetrik

$$d_\infty(x,y) = max\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_p - y_p|\}$$

äquivalent sind, d.h., dass es zwei Konstanten $C, D > 0$ gibt mit

$$\forall x,y \in \mathbb{R}^p : d(x,y) \leq C \cdot d_\infty(x,y) \;\&\; d_\infty(x,y) \leq D \cdot d(x,y).$$

Dies bedeutet, dass beide Metriken das gleiche System von offenen Mengen erzeugen. Da $\mathbb{R}^p$ bezüglich beider Metriken separabel ist, und da in diesem Fall die Borelsche σ-Algebra ebenfalls bereits von den offenen ε-Umgebungen erzeugt wird, kann man sich leicht überlegen, dass

$$\mathfrak{B}(\mathbb{R}^p) = \bigotimes_{i=1}^{p} \mathfrak{B}(\mathbb{R})$$

gilt, denn die ε-Umgebungen bezüglich der Supremumsmetrik sind genau die p-dimensionalen offenen Hyperquadrate, die bereits die p-dimensionalen offenen Hyperrechtecke und damit somit sowohl die Produkt-σ-Algebra $\bigotimes_{i=1}^{p} \mathfrak{B}(\mathbb{R})$, als auch die Borelsche σ-Algebra auf $\mathbb{R}^p$ erzeugen.

Definition 1 (Produktmaß).
Seien $(\Omega_1, \mathcal{A}_1, \mu_1), (\Omega_2, \mathcal{A}_2, \mu_2), \ldots, (\Omega_p, \mathcal{A}_p, \mu_p)$ Maßräume und sei μ ein Maß auf dem Messraum

$$\left(\bigtimes_{i=1}^{p} \Omega_i, \bigotimes_{i=1}^{p} \mathcal{A}_i \right).$$

Gilt für beliebige $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, \ldots, A_p \in \mathcal{A}_p$ die Faktorisierungseigenschaft

$$\mu(A_1 \times A_2 \times \ldots \times A_p) = \mu_1(A_1) \cdot \mu_2(A_2) \cdot \ldots \cdot \mu_2(A_p),$$

so nennen wir μ das Produktmaß von $\mu_1, \mu_2, \ldots, \mu_p$ auf $\left(\bigtimes_{i=1}^{p} \Omega_i, \bigotimes_{i=1}^{p} \mathcal{A}_i \right)$ und notieren es als

$$\mu = \bigotimes_{i=1}^{p} \mu_i.$$

Satz 9.28 (Existenz und Eindeutigkeit des Produktmaßes) *Seien $(\Omega_1, \mathcal{A}_1, \mu_1), (\Omega_2, \mathcal{A}_2, \mu_2), \ldots, (\Omega_p, \mathcal{A}_p, \mu_p)$ jeweils σ-endliche Maßräume. Dann existiert ein Produktmaß*

$$\mu = \bigotimes_{i=1}^{p} \mu_i$$

auf dem Messraum

$$\left(\bigtimes_{i=1}^{p} \Omega_i, \bigotimes_{i=1}^{p} \mathcal{A}_i \right).$$

Dieses ist eindeutig bestimmt und ebenfalls σ-endlich.

Zum Beweis. Die Existenz kann beispielsweise über den Fortsetzungssatz von Carathéodory geschehen. Betrachte dazu $\mathcal{D}_1 = \{A_1 \times \ldots \times A_p \mid A_1 \in \mathcal{A}_1, \ldots, A_p \in \mathcal{A}_p\}$. Dies ist ein Halbring, vergleiche Hausübugng 11, Aufgabe 2. Definiert man nun μ auf $\mathcal{D}_1$ vermöge

$$\mu(A_1 \times \ldots \times A_p) := \mu_1(A_1) \cdot \ldots \cdot \mu_p(A_p),$$

dann ist μ ein Prämaß. Zumindest das μ ein Inhalt ist, haben Sie für $p = 2$ ebenfalls in Hausübung 11, Aufgabe 2 gezeigt. Die Idee ist hier, Ω_1 und Ω_2 so fein wie eben nötig zu zerlegen (mit Hilfe geeignet gewählter Ununterscheidbarkeitsrelationen auf Ω_1 und Ω_2), und für $H_1, \ldots, H_n \in \mathcal{D}_1$ mit $H = \dot{\bigcup}_{i \in \{1, \ldots n\}} H_i \in \mathcal{D}_1$ die Menge H in kleinere kartesische Produkte zu zerlegen und die Additivität von oben definiertem μ auf die Additivität von μ_1 und μ_2 mittels Umordnung von Summanden zurückzuführen.

Die Eindeutigkeit folgt direkt über den Maßeindeutigkeitssatz, denn $\mathcal{D}_1$ ist als Halbring schnittstabil und da $\mu_1, \ldots, \mu_p$ jeweils σ-endlich sind, gibt es Folgen $(E_n^1)_{n \in \mathbb{N}} \in \mathcal{A}_1^{\mathbb{N}}, \ldots, (E_n^p)_{n \in \mathbb{N}} \in \mathcal{A}_p^{\mathbb{N}}$ von Mengen mit $\mu_i(E_n^i) < \infty$ für $i \in \{1, \ldots, p\}$ und $n \in \mathbb{N}$, sowie $\bigcup_{n \in \mathbb{N}} E_n^i = \Omega_i$ für $i \in \{1, \ldots, p\}$. Mit

$$F_{n_1, n_2, \ldots, n_p} := E_{n_1}^1 \times E_{n_2}^2 \times \cdots \times E_{n_p}^p \in \mathcal{D}_1$$

für $n_1, \ldots, n_p \in \mathbb{N}$ ist

$$\mu(F_{n_1, n_2, \ldots, n_p}) = \mu_1(E_{n_1}^1) \cdot \mu_2(E_{n_2}^2) \cdot \cdots \cdot \mu_p(E_{n_p}^p)) < \infty$$

und es gilt

$$\Omega = \underbrace{\bigcup_{n_1, \ldots, n_p \in \mathbb{N}}}_{\text{abzählbare Vereinigung}} F_{n_1, n_2, \ldots, n_p},$$

d.h., die zweite Voraussetzung des Eindeutigkeitssatzes ist ebenfalls erfüllt.

Satz 9.29 (Fubini-Tonelli: Integration bezüglich eines Produktmaßes)

Seien $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ zwei σ-endliche Maße und $f : \Omega_1 \times \Omega_2 \longrightarrow \mathbb{R}$ eine $\mathcal{A}_1 \otimes \mathcal{A}_2$-$\mathfrak{B}(\mathbb{R})$-messbare Funktion. Ist f nichtnegativ oder ist f bezüglich $\mu_1 \otimes \mu_2$ integrierbar, so gibt es eine μ_2-Nullmenge $N \subseteq \Omega_2$, so dass für $\omega_2 \in \Omega_2 \backslash N$ die Funktion $f(\cdot, \omega_2) : \Omega_1 \longrightarrow \mathbb{R} : \omega_1 \mapsto f(\omega_1, \omega_2)$ integrierbar bezüglich μ_1 ist. Mit

$$f_{\Omega_1} : \Omega_2 \backslash N \longrightarrow \mathbb{R} : \omega_2 \mapsto I_{\mu_1}(f(\cdot, \omega_2))$$

gilt weiter

$$I_{\mu_1 \otimes \mu_2}(f) = I_{\mu_{2|\Omega_2 \backslash N}}(f_{\Omega_1})$$

bzw.

$$\int\limits_{\Omega_1\times\Omega_2} f(\omega_1,\omega_2)\, d(\mu_1\otimes\mu_2)(\omega_1,\omega_2) = \int\limits_{\Omega_2\setminus N} \left(\int\limits_{\Omega_1} f(\omega_1,\omega_2)\, d\mu_1(\omega_1) \right) d\mu_{2|\Omega_2\setminus N}(\omega_2).$$

Definition 2 (Stochastische Unabhängigkeit von Ereignissen).
Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum. Die Mengen $A_1,\ldots,A_n \in \mathcal{A}$ heißen (gemeinsam) stochastisch unabhängig (bezüglich P), falls

$$P\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} P(A_i)$$

für alle beliebigen Indexmengen $I \subseteq \{1,\ldots,n\}$ gilt.

Definition 3 (Stochastische Unabhängigkeit von Mengensystemen). Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum und $\mathcal{D} \in \left(2^{\mathcal{A}}\right)^n$, d.h., für jedes $i \in \{1,\ldots,n\}$ ist $\mathcal{D}_i \subseteq \mathcal{A}$ ein Mengensystem. Dann heißen $\mathcal{D}_1,\ldots,\mathcal{D}_n$ stochastisch unabhängig, falls für beliebige $A_1 \in \mathcal{D}_1, A_2 \in \mathcal{D}_2,\ldots,A_n \in \mathcal{D}_n$ die Mengen $A_1,\ldots,A_n$ stochastisch unabhängig sind.

Definition 4 (Stochastische Unabhängigkeit von Zufallsvariablen). Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum, $(\Omega_1,\mathcal{A}_1),\ldots,(\Omega_n,\mathcal{A}_n)$ Messräume und für $i \in \{1,\ldots,n\}$

$$X_i : \Omega \longrightarrow \mathcal{A}_i$$

eine $\mathcal{A}-\mathcal{A}_i$-messbare Abbildung. Dann heißen $X_1,\ldots,X_n$ stochastisch unabhängig (bezüglich P), falls die Mengensysteme

$$X_1^{-1}[\mathcal{A}_1], X_2^{-1}[\mathcal{A}_2],\ldots,X_n^{-1}[\mathcal{A}_n]$$

stochastisch unabhängig sind, also wenn für beliebige $B_1 \in \mathcal{A}_1, B_2 \in \mathcal{A}_2,\ldots,B_n \in \mathcal{A}_n$ gilt:

$$P\left(\bigcap_{i\in I} X_i^{-1}(B_i)\right) = \prod_{i\in I} P(X_i^{-1}(B_i))$$

für jede Indexmenge $I \subseteq \{1,\ldots,n\}$.

Satz 9.30 (Schnittstabiler Erzeuger und Unabhängigkeit) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und seien $(\Omega_1, \mathcal{A}_1), \ldots, (\Omega_n, \mathcal{A}_n)$ Messräume sowie*

$$X_i : \Omega \longrightarrow \Omega_i$$

für $i \in \{1, \ldots, n\}$ jeweils eine $\mathcal{A} - \mathcal{A}_i$-messbare Abbildung und $\mathcal{E}_i$ ein Erzeuger von $\mathcal{A}_i$, der jeweils schnittstabil ist. Dann gilt: Die Zufallsvariablen $X_1, \ldots, X_n$ sind stochastisch unabhängig (bezüglich P) genau dann, wenn

$$P\left(\bigcap_{i \in I} X_i^{-1}(E_i)\right) = \prod_{i \in I} P(X_i^{-1}(E_i))$$

für beliebige $E_i \in \mathcal{E}_i$ und beliebige Indexmengen $I \subseteq \{1, \ldots, n\}$.

Satz 9.31 (Unabhängigkeit von Zufallsvariablen und Produktbildmaß) *Sei $(\Omega, \mathcal{A}, P)$ en Wahrscheinlichkeitsraum und für $i \in \{1, \ldots, n\}$ jeweils $(\Omega_i, \mathcal{A}_i)$ ein Messraum, sowie $X_i : \Omega \longrightarrow \mathcal{A}_i$ eine $\mathcal{A} - \mathcal{A}_i$-messbare Abbildung mit Bildmaß P_{X_i}. Sei weiter*

$$X : \Omega \longrightarrow \bigtimes_{i=1}^{n} \Omega_i : \omega \mapsto \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$$

und sei P_X das Bildmaß von X (definiert auf der Produkt-σ-Algebra $\bigotimes_{i=1}^{n} \mathcal{A}_i$). Dann gilt:

$$X_1, \ldots, X_n \text{ stochastisch unabhängig} \iff P_X = \bigotimes_{i=1}^{n} P_{X_i}.$$

Satz 9.32 (Unabhängigkeit von Funktionen von Zufallsvariablen) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum /und sei für $i = 1, \ldots, n$ jeweils $(\Omega_i, \mathcal{A}_i)$ ein Messraum sowie $X_i : \Omega \longrightarrow \Omega_i$ eine $\mathcal{A} - \mathcal{A}_i$-messbare Abbildung. Sei weiter für $i = 1, \ldots, n$ jeweils $(\tilde{\Omega}_i, \tilde{\mathcal{A}}_i)$ ein weiterer Messraum und $g_i : \Omega_i \longrightarrow \tilde{\Omega}_i$ eine $\mathcal{A}_i - \tilde{\mathcal{A}}_i$-messbare Abbildung. Dann gilt: Sind*

$$X_1, X_2, \ldots, X_n$$

stochastisch unabhängig, so sind auch

$$g_1 \circ X_1, g_2 \circ X_2, \ldots, g_n \circ X_n$$

stochastisch unabhängig.

Beweis. Siehe Übung 9, Aufgabe 1.

Satz 9.33 (Erwartungswert für unabhängige Zufallsvariablen) *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und sei für $i = 1, \ldots, n$ jeweils $X_i : \Omega \longrightarrow \mathbb{R}$ eine $\mathcal{A}-\mathfrak{B}(\mathbb{R})$-messbare Funktion, die integrierbar bezüglich P sei. Sind $X_1, \ldots, X_n$ stochastisch unabhängig, dann ist auch $X_1 \cdot X_2 \cdot \cdots \cdot X_n$ integrierbar bezüglich P und es gilt*

$$\mathbb{E}_P\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}_P(X_i).$$

Zum Beweis. Vergleiche Übung 9, Aufgabe 2.

Korollar 9.34 *Für zwei Zufallsvariablen X, Y, die bezüglich P integrierbar und stochastisch unabhängig sind, gilt*

$$\begin{aligned} COV(X,Y) &= \mathbb{E}_P((X - \mathbb{E}_P(X)) \cdot (Y - \mathbb{E}_P(Y)) \\ &= \mathbb{E}_P(X \cdot Y) - \mathbb{E}_P(X) \cdot \mathbb{E}_P(Y) = 0. \end{aligned}$$

Kapitel 6
Grenzwertsätze

10 Klassische Grenzwertsätze

Wenden wir uns jetzt einigen klassischen Sätzen zu, die Auskunft darüber geben, unter welchen Voraussetzungen gewisse zufällige Größen, wie beispielsweise der Mittelwert einer Stichprobe, in welchem Sinne gegen welchen Grenzwert konvergieren.

Satz 10.1 (Schwaches Gesetz der großen Zahlen) *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge paarweise unkorrelierter reellwertiger Zufallsvariablen (auf dem gleichen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$) mit $Var(X_n) \leq c < \infty$ für alle $n \in \mathbb{N}$. Dann genügt $(X_n)_{n\in\mathbb{N}}$ einem schwachen Gesetz der großen Zahlen, d.h.,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_P\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \quad \xrightarrow{\mathbb{P}} 0.$$

Konkret gilt

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_P\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right| \geq \varepsilon\right) \leq \frac{c}{n\varepsilon^2}.$$

Beweis. Mit $Z := \frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mathbb{E}_P\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right)$ gilt $\mathbb{E}_P(Z) = 0$ und mit der Tschebyscheff-Ungleichung folgt

$$
\begin{aligned}
P(|Z-0| \geq \varepsilon) &\leq \frac{Var(Z)}{\varepsilon^2} \\
&= \frac{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_P\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right)}{\varepsilon^2} \\
&= \frac{Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} \\
&= \frac{\frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)}{\varepsilon^2} \\
&\leq \frac{\frac{1}{n^2}\cdot nc}{\varepsilon^2} = \frac{c}{n\varepsilon^2}.
\end{aligned}
$$

Bemerkung 10.2 *Es gilt auch für $(X_n)_{n\in\mathbb{N}}$ i.i.d. mit lediglich $\mathbb{E}_P(|X_1|) < \infty$, dass $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}_P(X_1)$. Der Beweis benutzt sogenannte "Stutzungstechniken", bei denen man mit den gestutzten Zufallsvariablen $\tilde{X}_n = X_n \cdot \mathbb{1}_{\{|X_n|\leq M\}}$ arbeitet. Details dazu kann man beispielsweise in [Durrett, 2010, S.52ff] nachlesen.*

Satz 10.3 (Starkes Gesetz der großen Zahlen) *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge paarweise unabhängiger und identisch verteilter reellwertiger Zufallsvariablen (auf dem gleichen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$) mit $\mathbb{E}(|X_1|) < \infty$. Dann genügt $(X_n)_{n\in\mathbb{N}}$ einem starken Gesetz der großen Zahlen, d.h., es gilt*

$$
\frac{1}{n}\sum_{i=1}^{n} \xrightarrow{P-f.s.} \mathbb{E}_P(X_1).
$$

Bemerkungen zum Beweis.

i) Unter der stärkeren Voraussetzung, dass $(X_n)_{n\in\mathbb{N}}$ i.i.d. ist und dass X_1 P-fast sicher beschränkt ist, kann man ein starkes Gesetz der großen Zahlen mit Hilfe der Hoeffding-Ungleichung und Borel-Cantelli zeigen.

ii) Unter der Annahme $(X_n)_{n\in\mathbb{N}}$ i.i.d. und $\mathbb{E}_P(|X_1|^4) < \infty$, die weniger stark als die Annahme unter i) ist, kann man ein starkes Gesetz der großen

Zahlen unter Anwendung der Markov-Ungleichung auf

$$\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_P(X_1)\right]^4$$

plus Borel-Cantelli zeigen.

iii) Um den Satz unter der oben aufgeführten schwachen Voraussetzung $(X_n)_{n\in\mathbb{N}}$ paarweise unabhängig und identisch verteilt mit $\mathbb{E}_P(|X_1|) < \infty$ zu zeigen, wären wieder Stutzungstechniken plus Borel-Cantelli plus ein bisschen mehr zu bemühen. Details dazu sind in [Durrett, 2010, S.63ff] zu finden.

Satz 10.4 (Zentraler Grenzwertsatz) *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger, identisch und unabhängig verteilter Zufallsvariablen auf dem Wahrscheinlichkeitraum $(\Omega, \mathcal{A}, P)$ mit endlichem Erwartungswert $\mu := \mathbb{E}_P(X_1)$ und endlicher Varianz $\sigma^2 := Var(X_1) \in (0,\infty)$. (Der Fall $Var(X_1) = 0$ bedeutete $X_1 = \mu$ $P.-f.s.$ und ist uninteressant.) Dann gilt:*

$$Z_n := \frac{\sum\limits_{i=1}^{n} X_i \; - \; n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} Z,$$

wobei Z eine standardnormalverteilte Zufallsvariable ist.

Zum Beweis. Der Beweis wird meist über die charakteristische Funktion

$$\varphi_X : \mathbb{R} \longrightarrow \mathbb{C} : t \mapsto \mathbb{E}_P\left(e^{i\cdot t\cdot X}\right)$$

geführt, vergleiche beispielsweise [Schmidt, 2009, S.383ff]. Ein eher elementarer Beweis ist in Lindeberg [1922] zu finden.

Satz 10.5 (Berry-Esseen) *Seien die Voraussetzungen des zentralen Grenzwertsatzes erfüllt und sei darüber hinaus*

$$\gamma := \mathbb{E}_P(|X-\mu|^3) < \infty.$$

Dann gilt für die Verteilungsfunktion F_{Z_n} der gemäß des zentralen Grenzwertsatzes asymptotisch standardnormalverteilten Zufallsvariable

$$Z_n := \frac{\sum_{i=1}^{n} X_i \; - \; n\mu}{\sigma\sqrt{n}},$$

dass

$$|F_{Z_n} - \Phi(x)| \leq \frac{C \cdot \gamma}{\sigma^3\sqrt{n}}$$

ist, wobei $C = 0.4748$ eine Konstante ist.

Bemerkung 10.6 *Der zentrale Grenzwertsatz kann genutzt werden, um ein **approximatives** Konfidenzintervall für den Erwartungswert einer Zufallsvariablen zu konstruieren. (Man beachte aber, dass σ^2 üblicherweise unbekannt ist, und deshalb aus den Daten geschätzt werden muss.) Über die Güte der Approximation durch die Normalverteilung gibt der Satz von Berry-Esseen Auskunft. Da aber insbesondere auch das dritte zentrale Moment γ, das in diesem Satz von wesentlicher Bedeutung ist, üblicherweise unbekannt ist, müsste man auch γ aus den Daten schätzen, und erhielte selbst bei bekannter Varianz nur eine **approximative** Einschätzung über die Güte der Normalverteilungsapproximation. Es stellt sich somit die recht naheliegende Frage, ob man ohne starke Annahmen überhaupt exakte Konfidenzintervalle für den Erwartungswert konstruieren kann. Es gilt leider folgender*

Fakt: Im vollständig nichtparamterischen Setting (, d.h., nur i.i.d.-Annahme, insbesondere keine Annahme über die Beschränktheit des Trägers der interessierenden Zufallsvariable,) gibt es keine "nichttrivialen" Konfidenzintervalle bzw. Hypothesentests, wie in Bahadur and Savage [1956] gezeigt wurde:

"It seems plausible that if the population distribution of a real random variable is entirely unknown, then a sample from the population can yield little or no information about the tails of the distribution, even if the sample is obtained according to a sequential procedure. This paper gives evidence supporting and clarifying this proposition. The paper treats in some detail problems of inference concerning the population mean μ. It is shown that there is neither an effective test of the hypothesis that $\mu = 0$, nor an effective confidence interval for μ, nor an effective point estimate of μ. These

conclusions concerning μ flow from the fact that μ is sensitive to the tails of the population distribution; parallel conclusions hold for other sensitive parameters, and they can be established by the same methods as are here used for μ. It is also shown that there exists no confidence band for the population distribution function such that the upper and lower limits of the band are themselves distribution functions; that is, no confidence band fits very well." [Bahadur and Savage, 1956, S.1115]

11 Gleichmäßige Konvergenz, Glivenko-Cantelli, Vapnik-Chervonenkis: Der "Hauptsatz der Statistik" und ein bisschen mehr

Im vorherigen Abschnitt haben wir uns für den Erwartungswert als Verteilungsparameter und die Konvergenz des Mittelwertes gegen den Erwartungswert interessiert. Der folgende Abschnitt ist jetzt nicht nur einzelnen Verteilungsparametern, sondern vielmehr der Verteilung einer Zufallvariable schlechthin gewidmet. Wir stellen uns nun die folgende Frage: Kann man, gegeben ein i.i.d.-Sample $(X_1,\ldots,X_n)$ einer Zufallsvariable mit Bildmaß P_{X_1}, das unbekannte P_{X_1} irgendwie "vernünftig" schätzen? Das starke Gesetz der großen Zahlen sagt uns bereits, dass für ein beliebiges (messbares) Ereigniss A die relative Häufigkeit des Eintretens des Ereignisses in einer i.i.d.-Stichprobe der Göße n für $n \to \infty$ P-fast sicher gegen die wahre Wahrscheinlichkeit konvergiert. Ist man aber nich nur an einem einzigen Wahrscheinlichkeitswert, sondern am Wahrscheinlichkeitsmaß als Ganzem interessiert (Warum könnte man sich wohl dafür interessieren?), dann wird man eher an einer Konvergenz bezüglich einer Metrik in einem geeigneten Raum von Wahrscheinlichkeitsmaßen (auf einem gemeinsam unterliegenden Messraum $(\Omega,\mathcal{A})$) interessiert sein. Eine naheliegende Metrik wäre hier vielleicht die Supremumsmetrik

$$d_\infty(P,P') = ||P-P'||_\infty = \sup_{A\in\mathcal{A}} |P(A)-P'(A)|,$$

die aber oft(, insbesondere in nicht-diskreten Situationen, vgl. später,) zu grobschlächtig ist. Eine Möglichkeit der Modifikation der Supremumsmetrik wäre, das Supremum nicht über alle messbaren Mengen $A \in \mathcal{A}$ zu betrachten, sondern die Klasse der betrachteten Ereignisse auf eine kleinere Klasse $\mathcal{S} \subseteq \mathcal{A}$ einzuschränken.

11.1 Glivenko-Cantelli-Klassen

Die Frage, welche Klassen $\mathcal{S}$ unter gewissen Voraussetzungen gewisse Konvergenzresultate zulassen, ist Gegenstand dieses Abschnitts. Im Folgenden sei jetzt immer ein i.i.d.-Sample $(X_1, \ldots, X_n)$ einer Zufallsvariable mit Bildmaß P_{X_1} vorausgesetzt. Achtung: Die Zufallsvariable muss nicht unbedingt reellwertig zu sein. Eine naheliegende Möglichkeit, für ein Ereignis A die Wahrscheinlichkeit $P_{X_1}(A)$ aus einer Stichprobe zu schätzen, bestünde darin, einfach zu zählen, wie viele Beobachtungen in A liegen. Als Schätzwert für $P_{X_1}(A)$ kann man dann einfach die relative Häufigkeit des Eintretens des Ereignisses A in der Stichprobe ansetzen. Dies führt auf das sogenannte empirische Wahrscheinlichkeitsmaß:

Definition 1 (Empirisches Wahrscheinlichkeitsmaß). Es sei $(\Omega, \mathcal{A})$ ein Messraum. Bezeichne $\mathscr{P}(\Omega, \mathcal{A})$ die Menge aller Wahrscheinlichkeitsmaße auf dem Messraum $(\Omega, \mathcal{A})$. Sei jetzt weiter P ein konkretes Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ und $(X_n)_{n\in\mathbb{N}}$ eine Folge von i.i.d. Zufallsvariablen (auf (Ω, A, P)) mit Werten in einer Menge V, die mit einer σ-Algebra $\mathcal{A}'$ ausgestattet sei. Wir interessieren uns für das (unbekannte) Bildmaß P_{X_1}. Definiere für $A \subseteq V$

$$\nu_{A,n} : \Omega \longrightarrow \mathbb{R} : \omega \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(X_i(\omega))$$

bzw. kurz:

$$\nu_{A,n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A \circ X_i$$

und

$$\nu'_{A,n} = \frac{1}{n} \sum_{i=n+1}^{2n} \mathbb{1}_A \circ X_i.$$

Dann ist

$$P_n : 2^V \times \Omega \longrightarrow \mathbb{R} : (A,\omega) \mapsto \nu_{A,n}(\omega)$$

das sogenannte empirische Maß zum Sample $(X_1,\ldots,X_n)$ und

$$P_n' : 2^V \times \Omega \longrightarrow \mathbb{R} : (A,\omega) \mapsto \nu_{A,n}'(\omega)$$

ist das empirische Maß zum Sample $(X_{n+1},\ldots,X_{2n})$. Das empirische-Maß P_n' wird später lediglich zu Beweiszwecken benutzt. Das Sample $(X_{n+1},\ldots,X_{2n})$, das ebenfalls nur eine Rolle im Rahmen eines Symmetrisierungsarguments innerhalb eines Beweises spielen wird, wird auch oft auch als "ghost-sample" bezeichnet.

Sei jetzt $\mathcal{S} \subseteq \mathcal{A}'$ eine Klasse von $\mathcal{A}'$-messbaren Teilmengen von V. Definiere dann zunächst für eine Abbildung $f : M \longrightarrow \mathbb{R}$ mit $M \supseteq \mathcal{S}$ die Norm $||\cdot||_{\mathcal{S}}$ via $||f||_{\mathcal{S}} := \sup\limits_{A\in\mathcal{S}} |f(A)|$.

Sei jetzt weiter

$$\begin{aligned} D_n : \Omega \longrightarrow \mathbb{R} : \omega \mapsto ||P_n(\cdot,\omega) - P_{X_1}||_{\mathcal{S}} &= \sup_{A\in\mathcal{S}} |P_n(A,\omega) - P_{X_1}(A)| \\ &= \sup_{A\in\mathcal{S}} |\nu_{A,n}(\omega) - P_{X_1}(A)| \end{aligned}$$

und

$$\begin{aligned} D_n' : \Omega \longrightarrow \mathbb{R} : \omega \mapsto ||P_n'(\cdot,\omega) - P_{X_1}||_{\mathcal{S}} &= \sup_{A\in\mathcal{S}} |P_n'(A,\omega) - P_{X_1}(A)| \\ &= \sup_{A\in\mathcal{S}} |\nu_{A,n}'(\omega) - P_{X_1}(A)| \end{aligned}$$

sowie

$$\begin{aligned} \tilde{D}_n : \Omega \longrightarrow \mathbb{R} : \omega \mapsto ||P_n(\cdot,\omega) - P_n'(\cdot,\omega)||_{\mathcal{S}} &= \sup_{A\in\mathcal{S}} |P_n(A,\omega) - P_n'(A,\omega)| \\ &= \sup_{A\in\mathcal{S}} |\nu_{A,n}(\omega) - \nu_{A,n}'|. \end{aligned}$$

Die Zufallsvariable D_n (und analog auch D'_n und $\tilde{D}_n$) werden wir im Folgenden auch (mit einem bisschen abuse of notation) mit $||P_n - P_{X_1}||_{\mathcal{S}}$ notieren. Außerdem werden wir im Folgenden immer etwas einschränkend annehmen, dass $\nu_{A,n}, \nu'_{A,n}, D_n, D'_n$ und $\tilde{D}_n$ jeweils $\mathcal{A} - \mathfrak{B}(\mathbb{R})$-messbar sind. (Ansonsten müsste man zum äußeren Maß P^* übergehen, was die Analyse schwieriger macht.) Nun kommen wir zum Eigentlichen:

Definition 2 (Glivenko-Cantelli-Klassen). Gegeben die Voraussetzungen von Definition 1 heißt eine Klasse $\mathcal{S} \subseteq \mathcal{A}'$

i) Glivenko-Cantelli-Klasse bezüglich P, falls eine der beiden äquivalenten[19] Bedingungen erfüllt ist

   a) $D_n \xrightarrow{P-f.s.} 0$
   b) $D_n \xrightarrow{\mathbb{P}} 0$;

ii) universelle Glivenko-Cantelli-Klasse, falls $\mathcal{S}$ bezüglich jedes Wahrscheinlichkeitsmaßes auf dem Messraum $(\Omega, \mathcal{A})$ eine Glivenko-Cantelli-Klasse ist;

iii) gleichmäßige Glivenko-Cantelli-Klasse, falls gilt:

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} : \forall n \geq n_0 \forall P \in \mathscr{P}(\Omega, \mathcal{A}) : P(D_n \geq \varepsilon) \leq \varepsilon$$

bzw.

$$\sup_{P \in \mathscr{P}(\Omega, \mathcal{A})} \mathbb{E}_P(D_n) \quad \longrightarrow \quad 0 \qquad \text{für } n \to \infty.$$

Bemerkung 11.1 $D_n \xrightarrow{P-f.s.} 0$ *bedeutet: Auf einer Menge mit Wahrscheinlichkeit* 1 *konvergiert* $P_{n|\mathcal{S}\times\Omega}$ *gleichmäßig in Bezug auf alle* $A \in \mathcal{S}$ *gegen* $P_{X_1|\mathcal{S}\times\Omega}$. *Ist* $\mathcal{S}$ *eine gleichmäßige Glivenko-Cantelli-Klasse, so ist die Konvergenz (in Wahrscheinlichkeit) gleichmäßig sowohl in Bezug auf alle* $A \in \mathcal{S}$, *als auch in Bezug auf alle* $P \in \mathscr{P}(\Omega, \mathcal{A})$.

Frage: Die Formulierung des Begriffs der gleichmäßigen Glivenko-Cantelli-Klasse sieht aus wie eine Verschärfung der Konvergenz in Wahrscheinlichkeit im Sinne einer Gleichmäßigkeit bezüglich des unterliegenden Wahrscheinlichkeitsmaßes $P \in \mathscr{P}(\Omega, \mathcal{A})$. Warum ergibt es konzeptionell keinen

[19] Für den Beweis dieser Äquivalenz vergleiche beispielsweise [Dudley et al., 1991, Theorem 1, S.488].

Sinn, nach einer analogen Verschärfung der P-fast sicheren Konvergenz, gleichmäßig bezüglich $P \in \mathscr{P}(\Omega, \mathcal{A})$, zu fragen?

Beispiel 11.2 *Sei V endlich und $\mathcal{S} = 2^V$, sowie P beliebig. Ist $\mathcal{S}$ dann eine Glivenko-Cantelli-Klasse bezüglich P?*

Da für festes $\omega \in \Omega$

$$\sup_{A \in \mathcal{S}} |\nu_{A,n}(\omega) - P_{X_1}(A)| \leq \varepsilon \iff \forall A \in \mathcal{S} : |\nu_{A,n}(\omega) - P_{X_1}(A)| \leq \varepsilon$$

gilt, und da $\mathcal{S} = 2^V$ hier endlich ist, gilt

$$\sup_{A \in \mathcal{S}} |\nu_{A,n}(\omega) - P_{X_1}(A)| \longrightarrow 0 \iff \forall A \in \mathcal{S} : |\nu_{A,n}(\omega) - P_{X_1}(A)| \longrightarrow 0 \ \textit{(warum?)}$$

Für festes $A \in \mathcal{S}$ gilt wegen des starken Gesetzes der großen Zahlen

$$|\nu_{A,n} - P_{X_1}(A)| \overset{P-f.s.}{\longrightarrow} 0,$$

wir haben also

$$\forall A \in \mathcal{S} : |\nu_{A,n} - P_{X_1}(A)| \overset{P-f.s.}{\longrightarrow} 0.$$

Da $\mathcal{S}$ endlich ist, haben wir auch, dass die Eigenschaft

$$\forall A \in \mathcal{S} : |\nu_{A,n} - P_{X_1}(A)| \longrightarrow 0$$

P-fast sicher gilt, vergleiche Übung 7, Aufgabe 2 zu μ-fast überall bestehenden Eigenschaften. Damit gilt nach obiger Überlegung, dass

$$D_n = \sup_{A \in \mathcal{S}} |\nu_{A,n} - P_{X_1}(A)|$$

fast sicher gegen 0 konvergiert. Da P beliebig war, ist $\mathcal{S}$ eine universelle Glivenko-Cantelli Klasse.

Ist $\mathcal{S}$ auch eine gleichmäßige Glivenko-Cantelli-Klasse? Es gilt:

$$
\begin{aligned}
P\left(\sup_{A\in\mathcal{S}}|\nu_{A,n}-P_{X_1}(A)|\geq\varepsilon\right) &= P\left(\exists A\in\mathcal{S}:|\nu_{A,n}-P_{X_1}(A)|\geq\varepsilon\right)\\
&= P\left(\bigcup_{A\in\mathcal{S}}\{|\nu_{A,n}-P_{X_1}(A)|\geq\varepsilon\}\right)\\
&\overset{\text{sog. union-bound}}{\leq} \sum_{A\in\mathcal{S}}P\left(\{|\nu_{A,n}-P_{X_1}(A)|\geq\varepsilon\}\right)\\
&\overset{\text{Chernoff}}{\leq} \sum_{A\in\mathcal{S}}C\cdot e^{-D\cdot n}\\
&= |\mathcal{S}|\cdot Ce^{-D\cdot n}
\end{aligned}
$$

mit Konstanten $C,D>0$ (unabhängig von P). Da $|\mathcal{S}|$ endlich ist, konvergiert die rechte Seite für $n\to\infty$ gleichmäßig bezüglich $P\in\mathscr{P}(\Omega,\mathcal{A})$ gegen 0, d.h., $\mathcal{S}$ ist sogar eine gleichmäßige Glivenko-Cantelli-Klasse.

Beispiel 11.3 *Sei $V=\mathbb{N}$ und $\mathcal{S}=2^V$ (überabzählbar). Hier lässt sich die Frage, ob $\mathcal{S}$ eine universelle Glivenko-Cantelli-Klasse ist, auf die Analyse in Beispiel 11.2 zurückführen. Betrachte dazu $A_m=\{m,m+1,m+2,\ldots\}$. Da*

$$1=P_{X_1}(V)=P_{X_1}\left(\bigcup_{m\in\mathbb{N}}\{m\}\right)=\sum_{m=1}^{\infty}P_{X_1}(\{m\})$$

muss notwendigerweise $P_{X_1}(A_m)=\sum_{k=m}^{\infty}P_{X_1}(\{k\})$ für $m\to\infty$ gegen 0 konvergieren. Gegeben ein $\varepsilon>0$ kann man dann m so groß wählen, dass $P_{X_1}(A_m)\leq\varepsilon/4$ ist, und $|\nu_{A_m,n}-P_{X_1}(A_m)|$ sehr wahrscheinlich kleinergleich $\varepsilon/4$ ist (schwaches Gesetz der großen Zahlen). Für alle Teilmengen $A\subseteq A_m$ ist dann $|\nu_{A,n}-P_{X_1}(A)|$ ebenfalls sehr wahrscheinlich kleinergleich $\varepsilon/2$. Eine Zerlegung von Mengen $A\subseteq\mathcal{S}$ in $A=A\cap A_m^c\ \dot\cup\ A\cap A_m$ und eine Analyse von $A\cap A_m^c$ gemäß Beispiel 11.2 ($A\cap A_m^c$ und damit auch $2^{A\cap A_m^c}$ ist endlich) sowie eine Analyse von $A\cap A_m$ gemäß obiger Epsilontik würde dann ergeben, dass $\mathcal{S}$ eine universelle Glivenko-Cantelli-Klasse ist. Allerdings ist $\mathcal{S}$ keine gleichmäßige Glivenko-Cantelli-Klasse mehr, man beachte, dass wir in obiger Analyse m so groß gewählt haben, dass $P_{X_1}(A_m)\leq\varepsilon/4$ ist, die Wahl von m ist deshalb abhängig von P_{X_1} bzw. P. Um wirklich einzusehen, dass $\mathcal{S}$ keine gleichmäßige Glivenko-Cantelli-Klasse ist, vergleiche Beispiel 11.6.

Beispiel 11.4
*i) Sei $V = \mathbb{R}$ und P_{X_1} absolut stetig bezüglich λ. Sei weiter $\mathcal{S} := \mathrm{Fin}(A) := \{A \subseteq V \mid A \text{ endlich }\}$ (überabzählbar). Dann ist $\mathcal{S}$ **keine** Glivenko-Cantelli-Klasse bezüglich P:*

Betrachte

$$\sup_{A\in\mathcal{S}} |\quad \underbrace{\nu_{A,n}}_{\substack{=1 \text{ für} \\ A=\{X_1(\omega),\ldots,X_n(\omega)\}}} \quad - \quad \underbrace{P_{X_1}(A)}_{=0} \quad | \geq 1.$$

Damit kann D_n nicht gegen 0 konvergieren und $\mathcal{S}$ ist keine Glivenko-Cantelli-Klasse.

ii) Sei $V = \mathbb{R}$, $\mathcal{S} = \mathrm{Fin}(V)$ (oder aber auch $\mathcal{S} = 2^V$). Sei aber P_{X_1} "deterministisch", d.h., es existiert ein $x_0 \in V$ mit $P_{X_1}(\{x_0\}) = 1$. Dann ist $P_{X_1}(A) = 1$ falls $x_0 \in A$ und $P_{X_1}(A) = 0$ sonst. Andererseits ist $\nu_{A,n} = 1$ falls $x_0 \in A$ und falls $X_1(\omega) = X_2(\omega) = \ldots = X_n(\omega) = x_0$, was im Fall $x_0 \in A$ mit Wahrscheinlichkeit 1 der Fall ist. Analog ist $\nu_{A,n}(A) = 0$ mit Wahrscheinlichkeit 1 im Fall $x_0 \notin A$.
Dies bedeutet, dass

$$P\left(\sup_{A\in\mathcal{S}} |\nu_{A,n} - P_{X_1}(A)| = 0\right) = 1$$

gilt, d.h., $\mathcal{S}$ ist eine gleichmäßige Glivenko-Cantelli-Klasse.

iii) Sei $V = \mathbb{R}$, $\mathcal{S} = \mathrm{Fin}(V)$ und $P_{X_1}(A) = 1$ für eine endliche Menge A. Dann kann man analog zu Beispiel 11.2 zeigen, dass $\mathcal{S}$ ebenfalls eine Glivenko-Cantelli-Klasse bezüglich des betrachteten Wahrscheinlichkeitsmaßes P ist.

Bemerkung 11.5 *Beispiele 11.4 ii) und iii) könnte man jetzt etwa so auffassen, dass man im Fall, dass das unbekannte Wahrscheinlichkeitsgesetz P kontingenterweise nicht "zu arg" ist, trotzdem P_{X_1} gleichmäßig auf einer großen Klasse $\mathcal{S}$ "lernen" kann. Man kann also auch "Glück" haben.*

Beispiel 11.6 *Sei $V = \mathbb{N}$ und $\mathcal{S} = 2^V$ (oder auch $\mathcal{S} = \mathrm{Fin}(V)$, vergleiche auch Beispiele 11.3 und 11.4). Dann ist $\mathcal{S}$ nach Beispiel 11.3 eine universelle Glivenko-Cantelli-Klasse. Sei jetzt $P^k_{X_1}$ die diskrete Gleichverteilung auf $\{1,\ldots,k\}$. Dann gilt für jedes noch so große $n \in \mathbb{N}$:*

$$\sup_{A\in\mathcal{S}} \Big| \underbrace{\nu_{A,n}}_{\substack{=1 \text{ für} \\ A=\{X_1(\omega),\dots,X_n(\omega)\}}} - \underbrace{P_{X_1}^{2n}(A)}_{\substack{\leq 1/2 \text{ für} \\ A=\{X_1(\omega),\dots,X_n(\omega)\}}} \Big| \geq 1/2,$$

d.h., $\mathcal{S}$ kann keine gleichmäßige Glivenko-Cantelli-Klasse sein.

Nachdem wir jetzt ein grobes erstes Gefühl für Glivenko-Cantelli-Klassen bekommen haben, stellt sich nun zunächst die Frage, ob es überhaupt nicht-triviale (,d.h., wenigstens unendliche) gleichmäßige Glivenko-Cantelli-Klassen gibt. Eine ganz einfache Glivenko-Cantelli-Klase ist Gegenstand des nächsten Satzes:

Satz 11.7 (Satz von Glivenko-Cantelli, "Hauptsatz der Statistik") *Sei $(X_n)_{n\in\mathbb{N}}$ eine i.i.d.-Folge von reellwertigen Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},P)$ mit P beliebig. Dann gilt für die Klasse*

$$\mathcal{S} := \{(-\infty,c] \mid c\in\mathbb{R}\}$$

dass

$$||P_n - P_{X_1}||_{\mathcal{S}} \quad \overset{P-f.s.}{\longrightarrow} \quad 0,$$

d.h., $\mathcal{S}$ ist eine universelle Glivenko-Cantelli-Klasse.

Bemerkung 11.8 *Anders ausgedrückt gilt, dass*

$$||P_n - P_{X_1}||_{\mathcal{S}} = \sup_{(-\infty,c]\in\mathcal{S}} |P_n((-\infty,c]) - P_{X_1}((-\infty,c])| = \sup_{c\in\mathbb{R}} |F_n(c) - F_{X_1}(c)|$$

fast sicher gegen 0 konvergiert, wobei F_n die (zufällige) empirische Verteilungsfunktion zum Sample $(X_1,\dots,X_n)$ ist. Darüber hinaus gilt sogar, dass $\mathcal{S}$ eine gleichmäßige Glivenko-Cantelli-Klasse ist. Konkreter gilt:

i) [Vapnik and Chervonenkis, 1968, 1971, Vapnik, 2006a]: Für $n \geq \frac{2}{\varepsilon^2}$ gilt

$$P(D_n \geq \varepsilon) \leq 6(2n+1)e^{-\frac{\varepsilon^2 n}{4}}.$$

ii) [Dvoretzky et al., 1956]:

$$P(D_n \geq \varepsilon) \leq C\cdot e^{-\varepsilon^2 n},$$

wobei $C > 0$ eine unspezifizierte Konstante ist.

iii) [Massart, 1990]:

$$P(D_n \geq \varepsilon) \leq 2 \cdot e^{-\varepsilon^2 n}$$

und die Konstante $C = 2$ vor der Exponentialfunktion ist scharf.

Frage: Warum wird der Satz von Glivenko-Cantelli manchmal auch als Hauptsatz der Statistik bezeichnet?

- Historisch: Mir ist nicht bekannt, wer den Namen *Hauptsatz* in diesem Zusammenhang geprägt hat. Mir ist lediglich beispielsweise das folgende Zitat bekannt:

  "Analysis of Bernoulli's law of large numbers has been the subject of intensive research since the 1930s. Also in the 1930s it was shown that for one particular set of events the uniform law of large numbers always holds. This fact is the Glivenco-Cantelli theorem. The corresponding bound on the rate of convergence forms Kolmogorov's bound. Classical statistics took advantage of these results (the Glivenco-Cantelli theorem and Kolmogorov's bound are regarded as the foundation of theoretical statistics)."[Vapnik, 2006b]

- Die Verteilungsfunktion F_{X_1} charakterisiert P_{X_1}.
- Jeder beliebige Verteilungsparameter θ, für den man sich interessieren könnte, kann als $\theta = \theta(F_{X_1})$ geschrieben werden.
- Plug-in-Prinzip/Substitutionsprinzip: Mit $\theta : \mathscr{F} \longrightarrow \mathbb{R} : F \mapsto \theta(F)$ mit $\mathscr{F}$ der Menge aller Verteilungsfunktionen besteht immer die naheliegende (aber nicht immer sehr kluge) Möglichkeit, θ zu schätzen, indem man einfach die empirische Verteilungsfunktion in die Abbildung θ einsetzt:

  $$\hat{\theta} := \theta(F_n).$$

  Bemerkung dazu: Da F_n der (nichtparametrische) Maximum-Likelihood-Schätzer für F ist, ist der Schätzer $\hat{\theta} := \theta(F_n)$ der Maximum-Likelihood-Schätzer für θ im komplett nichtparametrischen Setting.
- Was folgt nun aus dem Hauptsatz?
- Da F_n fast sicher gegen F_{X_1} bezüglich der Supremumsnorm konvergiert, würde sich die Konvergenz auf $\hat{\theta}(F_n)$ übertragen, sofern θ stetig ist (bezüglich der Supremumsnorm).

Schauen wir also einmal die Stetigkeit verschiedener Funktionale

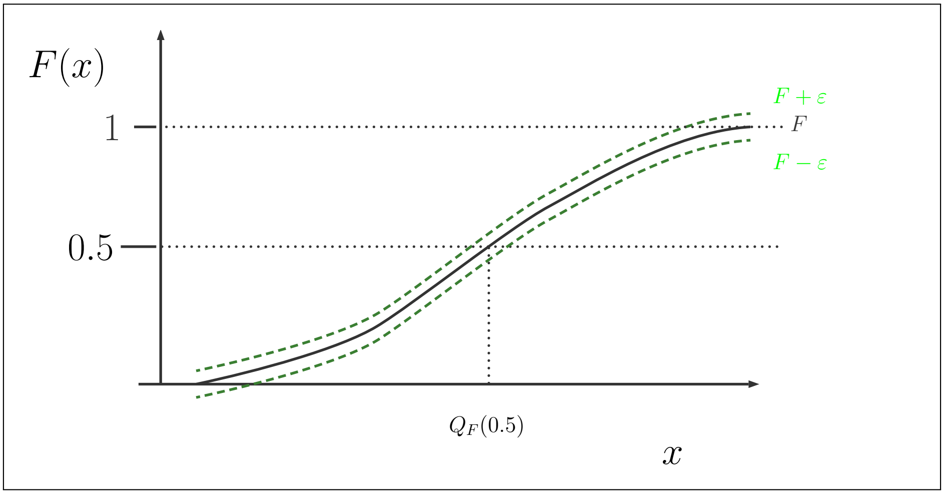
$$\theta : \mathscr{F} \longrightarrow \mathbb{R}$$

an:

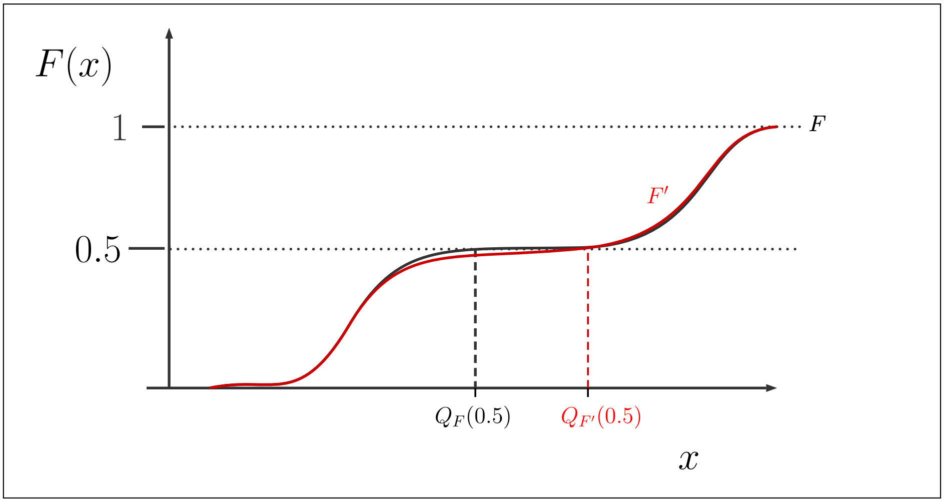
Beispiel 11.9 (Median)

$$\theta : \mathscr{F} \longrightarrow \mathbb{R} : F \mapsto median(F) = Q_F(0.5),$$

wobei Q_F die Quantil-Funktion zu F ist.



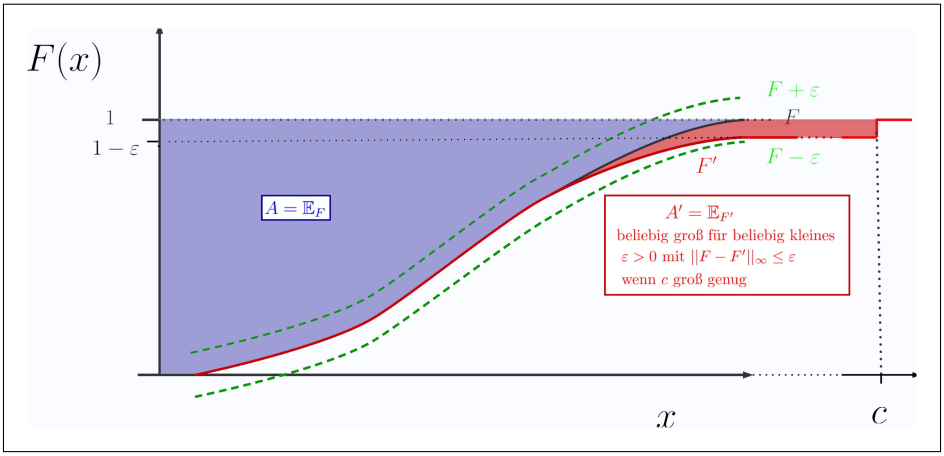
Das Median-Funktional ist an allen Stellen F, für die F in einer Umgebung um den Median von F bijektiv ist, stetig.



Fällt der Median von F auf ein Plateau von F, so ist das Median-Funktional an der Stelle F nicht stetig.

Beispiel 11.10 (Erwartungswert) *Für $X \geq 0$ ist*

$$\mathbb{E}_{F_X} := \mathbb{E}(X) = I_P(X) = \int_0^\infty 1 - F_X(t)\, dt = \int_0^1 Q_X(t)\, dt$$



Damit ist das Funktional

$$\theta : \mathscr{F} \longrightarrow \mathbb{R} : F \mapsto \mathbb{E}_F$$

nach obiger Skizze an keiner Stelle F stetig.

- Die konkreten Ungleichungen für die Größe $\sup_{c\in\mathbb{R}} |F_n(c) - F_{X_1}(c)|$ erlauben es auch, im Prinzip für jeden beliebigen Parameter θ ein exakt valides (im Sinne von $P(KI \ni \theta) \geq 1-\alpha$ unabhängig von θ), allerdings oft meist konservatives Konfidenzintervall zu konstruieren:

  i) Konstruiere zunächst ein (simultanes) Konfidenzband KB für F_{X_1}:

  $$KB := [L,U]$$

  mit $L : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto \max\{F_n(x) - \varepsilon, 0\}$ und $U : \mathbb{R} \longrightarrow \mathbb{R} : x \mapsto \min\{F_n(x) + \varepsilon, 1\}$ und ε derart, dass

  $$P\left(KB \ni F_{X_1}\right) \geq 1-\alpha$$

  gilt. Dies ist mit der Dvoretzky-Kiefer-Wolfowitz-Ungleichung erreichbar.

  ii) Konstruiere anschließend ein Konfidenzintervall für θ via

  $$KI = \left[\inf_{\substack{F \text{ Verteilungsfunktion,} \\ F\in KB}} \theta(F), \sup_{\substack{F \text{ Verteilungsfunktion,} \\ F\in KB}} \theta(F) \right].$$

- **Warnung:** Wir haben hier einen Umweg $(X_1,\ldots,X_n) \mapsto F_n \mapsto \hat{\theta}$ über die Verteilungsfunktion gemacht, d.h., die Konfidenzintervalle können unnötig groß sein.

- Für den Erwartungswert ist aufgrund der schlechten Stetigkeitseigenschaften des Funktionals θ das zugehörige Konfidenzintervall immer $[-\infty,\infty]$, was sich auch mit der Aussage von Bahadur und Savage deckt.

- Auch der Weg über die Dichte ist unter Umständen ein unkluger Umweg, insbesondere ist die Dichte nicht sehr schön abhängig von P_{X_1} bzw. F_{X_1}, es handelt sich hier um ein sogenanntes "schlecht gestelltes Problem"(vgl. z.B. [Vapnik, 2006b, S.477]), dem man mindestens mit irgendeiner Form von Regularisierung begegnen müsste (wenn man es denn mit der Dichte nicht gleich ganz ließe).

- Generell scheint (sofern umsetzbar) an dieser Stelle der folgende Imperativ angebracht:

"When solving a problem of interest, do not solve a more general problem as an intermediate step. Try to get the answer that you really need but not a more general one."[Vapnik, 2006b, S.477]

Zum Beweis des Hauptsatzes.

Es gibt verschiedene Möglichkeiten, den Hauptsatz zu beweisen. Eine Variante wäre "eher elementar": Für $\varepsilon > 0$ zerlege das Intervall $[0,1] \supseteq Im(F_{X_1})$ in Intervalle $[(k-1)\varepsilon, k\varepsilon]$, $k = 1, \ldots, m$ der Länge $\varepsilon = 1/m$. Betrachte dann die endlich vielen Stellen $x_k = Q_{X_1}((k-1)\varepsilon)$, $k = 2, \ldots, m-1$, wobei Q_{X_1} die Quantilsfunktion zu F_{X_1} ist. Dann kann man die P-fast sichere Konvergenz von $\sup_{x\in\mathbb{R}} |F_n(x) - F_{X_1}(x)|$ auf die P-fast sichere Konvergenz von

$$\sup_{k=2,\ldots,m-1} |F_n(x_k) - F_{X_1}(x_k)|$$

zurückführen, indem man für Zwischenstellen $x \in (x_k, x_{k+1})$ mit der Isotonie von F_{X_1} und F_n argumentiert, dass $F_n(x)$ von $F_{X_1}(x)$ um nicht mehr als 2ε abweichen kann, falls $|F_n(x_k) - F_{X_1}(x_k)| \leq \varepsilon$ und $|F_n(x_{k+1}) - F_{X_1}(x_{k+1})| \leq \varepsilon$ gilt.

Bei obigem Beweis wird wesentlich ausgenutzt, dass $Im(F_{X_1})$ kompakt ist, dass F_n bzw. F_{X_1} isoton ist, und dass der Wertebereich $\mathbb{R}$ der betrachteten Zufallsvariablen $X_1, \ldots, X_n$ total geordnet ist. Ein vollständiger Beweis kann z.B. in [Schmidt, 2009, S.355ff] nachgelesen werden.

Obige Beweisidee lässt sich auch für den multivariaten Fall von $\mathbb{R}^d$-wertigen Zufallsvariablen und die multivariate Verteilungsfunktion

$$F : \mathbb{R}^d \longrightarrow [0,1] : \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} \mapsto P(X_1 \leq c_1, \ldots, X_d \leq c_d)$$

adaptieren. (Beachte, dass $\mathbb{R}^d$ nicht mehr total geordnet ist und man nicht eine Zerlegung von $Im(F_{X_1})$, sondern eine Zerlegung von $\mathbb{R}^d$ betrachten muss.)

Neben dieser “eher elementaren” Beweisidee wollen wir jetzt noch einen etwas anderen Beweis kennenlernen, von dem erstens eine sehr viel konkretere Ungleichung abfällt, und der zweitens darüber hinaus eine erstaunliche Verallgemeinerung zulässt. Diese Verallgemeinerung wird sogar Anlass für ein weiteres Kapitel zur sogenannten Vapnik-Chervonenkis-Theorie sein.

Beweis nach Vapnik-Chervonenkis.

Der Beweis, den wir hier nur skizzieren wollen, um die wesentlichen Beweisideen zu verstehen, erfolgt in mehreren Schritten:

i) Symmetrisierung: Führe die Analyse von D_n auf die Analyse von $\tilde{D}_n$ zurück. (Dies ist der wesentliche Unterschied zum “eher elementaren” Beweis von oben.)

ii) Zurückführung der Supremumsstatistik

$$\tilde{D}_n = \sup_{A \in \mathcal{S}} |\nu_{A,n} - \nu'_{A,n}|$$

auf eine Supremumsstatistik

$$\sup_{A \in \mathcal{S}^{f,n}} |\nu_{A,n} - \nu'_{A,n}|$$

über einem endlichen Mengensystem $\mathcal{S}^{f,n}$. Das Mengensystem $\mathcal{S}^{f,n}$ ist dann aber ein zufälliges Mengensystem.

iii) Permutations- bzw. Symmetrieargument. Dies erlaubt eine bedingte Analyse, bei der das zufällige Mengensystem $\mathcal{S}^{f,n}$ als fix betrachtet werden kann.

iv) Union-bound plus eine Konzentrationsungleichung für relative Häufigkeiten unter Ziehen ohne Zurücklegen ([Serfling, 1974]).

Fangen wir also an:

i) Symmetrisierung: Führe die Analyse von D_n auf die Analyse von $\tilde{D}_n$ zurück: Man kann zeigen, dass für $n \geq 2/\varepsilon^2$ gilt:

$$P(D_n \geq \varepsilon) \leq 2P(\tilde{D}_n \geq \varepsilon/2).$$

Dies bedeutet, dass D_n in Wahrscheinlichkeit gegen 0 konvergiert, falls $\tilde{D}_n$ bereits gegen 0 konvergiert. (Wir werden sogar erhalten, dass die Konvergenz von $P(\tilde{D}_n \geq \varepsilon/2)$ und damit auch die Konvergenz von $P(D_n \geq \varepsilon)$ exponentiell in n ist, und wir somit wegen Borel-Cantelli auch P-fast sichere Konvergenz erhalten.) Um obige Ungleichung zu zeigen könnte man versuchen, die Dreiecksungleichung zu bemühen. Man erhielte unmittelbar

$$\tilde{D}_n = ||P_n - P'_n||_{\mathcal{S}} \leq ||P_n - P_{X_1}||_{\mathcal{S}} + ||P_{X_1} - P'_n||_{\mathcal{S}}.$$

Damit ist $P(\tilde{D}_n \geq \varepsilon) \leq P(D_n \geq \varepsilon/2) + P(D'_n \geq \varepsilon/2) = 2P(D_n \geq \varepsilon/2)$, was aber eine Ungleichung in die genau falsche Richtung ist. Wir müssen hier genauer in die Supremumsstatistik hineinschauen. Man braucht hier in der Tat eine richtige Idee:

Die Analyse von $\sup\limits_{A\in\mathcal{S}} |\nu_{A,n} - P_{X_1}(A)|$ ist vor allem deshalb schwierig, weil diejenige Stelle $A^* \in \mathcal{S}$, an der das Supremum "angenommen" wird, zufällig ist. Man kann also nicht so argumentieren, als ob A^* fest wäre, denn dann wäre man mit dem starken Gesetz der großen Zahlen fertig und alle Klassen wären universelle Glivenko-Cantelli-Klassen, was aber, wie wir bereits wissen, nicht der Fall ist. Aber: mit dem zusätzlichen "ghost-sample" kann man sich leicht das Folgende überlegen:

Sei A^* die Menge aus $\mathcal{S}$, auf der $\sup\limits_{A\in\mathcal{S}} |\nu_{A,n} - P_{X_1}(A)|$ "angenommen" wird. Das ghost-sample hat nun diese zufällige Menge A^* "nicht gesehen" ($(X_1, \ldots, X_n)$ ist stochastisch unabhängig von $(X_{n+1}, \ldots, X_{2n})$). Damit gilt wirklich, dass auch für die zufällige Menge A^* und für das ghost-sample der Wert von $|\nu_{A^*,n} - P_{X_1}(A^*)|$ gemäß des schwachen Gesetzes der großen Zahlen sehr wahrscheinlich sehr klein ist. Diese Einsicht macht nun die folgende Überlegung anwendbar:

Ist

$$D_n = ||P_n - P_{X_1}||_{\mathcal{S}} \approx |\nu_{A^*,n} - P_{X_1}(A^*)| \geq \varepsilon$$

und ist gleichzeitig $|\nu'_{A^*,n} - P_{X_1}(A^*)| \leq \frac{\varepsilon}{2}$, dann ist

$$\tilde{D}_n = ||P_n - P'_n||_{\mathcal{S}} \geq |\nu_{A^*,n} - \nu'_{A^*,n}| \geq \frac{\varepsilon}{2},$$

denn (Dreiecksungleichung)

$$|\nu_{A^*,n} - P_{X_1}(A^*)| \leq |\nu_{A^*,n} - \nu'_{A^*,n}| + |\nu'_{A^*,n} - P_{X_1}(A^*)|,$$

woraus

$$|\nu_{A^*,n} - \nu'_{A^*,n}| \geq \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}$$

folgt. Wir haben also

$$\{D_n \geq \varepsilon\} \cap \left\{|\nu'_{A^*,n} - P_{X_1}(A^*)| \leq \varepsilon/2\right\} \subseteq \left\{\tilde{D}_n \geq \varepsilon/2\right\}.$$

Daraus folgt nun

$$\begin{aligned} P(\{\tilde{D}_n \geq \varepsilon/2\}) &\geq P(\{D_n \geq \varepsilon \ \& \ |\nu'_{A^*,n} - P_{X_1}(A*)| \leq \varepsilon/2\}) \\ &= P(\{D_n \geq \varepsilon\}) \cdot \underbrace{P(|\nu'_{A^*,n} - P_{X_1}(A^*) \leq \varepsilon/2 \ \mid \ D_n \geq \varepsilon)}_{\substack{\geq 1/2 \ \ \text{für } n \geq 2/\varepsilon^2, \\ \text{Tschebyscheff}}}, \end{aligned}$$

also

$$P(D_n \geq \varepsilon) \leq 2P(\tilde{D}_n \geq \varepsilon/2).$$

Es reicht damit, $\tilde{D}_n$ zu analysieren.

ii) Zurückführung auf ein endliches Mengensystem:

$$\begin{aligned} \tilde{D}_n &= \sup_{x \in \mathbb{R}} |F_n(x) - F'_n(x)| \\ &= \sup_{x \in \underbrace{\{X_1(\omega), \ldots, X_{2n}(\omega)\}}_{\text{zufällig, aber endlich}}} |F_n(x) - F'_n(x)| \qquad \text{(warum?)} \end{aligned}$$

bzw. allgemein:

$$\sup_{A \in \mathcal{S}} |\nu_{A,n} - \nu'_{A,n}| = \sup_{A \in \mathcal{S}^{f,n}} |\nu_{A,n} - \nu'_{A,n}|$$

mit $\mathcal{S}^{f,n} = \{A \cap \{X_1(\omega), \ldots, X_{2n}(\omega)\} \mid A \in \mathcal{S}\}$.

- $\mathcal{S}^{f,n}$ ist zufällig, aber endlich. Konkret gilt $|\mathcal{S}^{f,n}| \leq 2^{2n}$.

- Achtung: im Allgemeinen wächst $|\mathcal{S}^{f,n}|$ mit n, schlimmstenfalls sogar exponentiell in n.

- Obiger Trick funktioniert nur wegen der vorherigen Symmetrisierung:

$$\begin{aligned} P_n(A) &= P_n(A \cap \{X_1(\omega), \ldots, X_n(\omega)\}) \text{ bzw.} \\ P_n'(A) &= P_n'(A \cap \{X_{n+1}(\omega), \ldots, X_{2n}(\omega)\}), \text{ aber i.A.:} \\ P_{X_1}(A) &\neq P_{X_1}(A \cap \{X_1(\omega), \ldots, X_{2n}(\omega)\}). \end{aligned}$$

iii) Permutations-/Symmetrieargument:

Wir lösen uns jetzt gedanklich von dem Fakt, dass $\mathcal{S}^{f,n}$ zufällig ist. Man kann sich den Entstehungsprozess von $(X_1, \ldots, X_{2n})$ auch so vorstellen: Ziehe Zufallsvariablen $(Y_1, \ldots, Y_{2n})$ i.i.d. gemäß P_{X_1}. Ziehe anschließend ohne Zurücklegen und stochastisch unabhängig von $(Y_1, \ldots, Y_{2n})$ Indizes $i_1, \ldots, i_{2n} \in \{1, \ldots, 2n\}$, wobei jede Permutation von $\{1, \ldots, 2n\}$ mit gleicher Wahrscheinlichkeit gezogen werden soll. Definiere dann

$$\begin{aligned} X_1 &= Y_{i_1} \\ &\vdots \\ X_{2n} &= Y_{i_{2n}}. \end{aligned}$$

Aus Symmetriegründen ist dann $(X_1, \ldots, X_{2n})$ in der Tat i.i.d. und gemäß P_{X_1} verteilt. Obige Vorstellung von der Entstehung von $(X_1, \ldots, X_{2n})$ erlaubt nun eine bedingte Analyse gegeben $(Y_1 = y_1, \ldots, Y_{2n} = y_{2n})$:

Betrachte
$$\tilde{P} := P(\cdot \mid Y_1 = y_1, \ldots, Y_{2n} = y_{2n}).$$

Dann kann

$$
\begin{aligned}
\mathcal{S}^{f,n} &= \{A \cap \{X_1(\omega), \ldots, X_{2n}(\omega)\} \mid A \in \mathcal{S}\} \\
&= \{A \cap \{Y_1(\omega), \ldots, Y_{2n}(\omega)\} \mid A \in \mathcal{S}\}
\end{aligned}
$$

in der bedingten Analyse als fest betrachtet werden.

iv) Union-bound plus Konzentrationsungleichung für Ziehen ohne Zurücklegen, [Serfling, 1974]:

$$
\begin{aligned}
\tilde{P}\left(\sup_{A\in\mathcal{S}} |\nu_{A,n} - \nu'_{A,n}| \geq \varepsilon/2\right) &= \tilde{P}\left(\sup_{A\in\mathcal{S}^{f,n}} |\nu_{A,n} - \nu'_{A,n}| \geq \varepsilon/2\right) \\
&= \tilde{P}\left(\bigcup_{A\in\mathcal{S}^{f,n}} \{|\nu_{A,n} - \nu'_{A,n}| \geq \varepsilon/2\}\right) \\
&\overset{\text{union-bound}}{\leq} \sum_{A\in\mathcal{S}^{f,n}} \tilde{P}\left(|\nu_{A,n} - \nu'_{A,n}| \geq \varepsilon/2\right) \\
&\overset{\text{Serfling}}{\leq} \sum_{A\in\mathcal{S}^{f,n}} Ce^{-Dn} \\
&= |\,\mathcal{S}^{f,n}\,| Ce^{-Dn}
\end{aligned}
$$

mit $C > 0, D > 0$ Konstanten, unabhängig von n (aber abhängig von ε).

In der Situation des Hauptsatzes haben wir

$$
|\,\mathcal{S}^{f,n}\,| \leq (2n+1) \qquad \text{(warum?)}
$$

und es folgt

$$
\tilde{P}\left(\sup_{A\in\mathcal{S}} |\nu_{A,n} - \nu'_{A,n}| \geq \varepsilon/2\right) \leq (2n+1)Ce^{-Dn}.
$$

Damit erhalten wir insgesamt:

$$
\begin{aligned}
P(D_n \geq \varepsilon) &\leq 2P(\tilde{D}_n \geq \varepsilon/2) \\
&\underset{\text{warum?}}{\leq} 2 \cdot \left((2n+1)Ce^{-Dn}\right),
\end{aligned}
$$

d.h., für jedes noch so kleine $\varepsilon > 0$ konvergiert $P(D_n \geq \varepsilon)$ exponentiell schnell (warum?) in n gegen 0. Wir haben also unter den Voraussetzungen des Hauptsatzes

$$D_n \xrightarrow{\mathbb{P}} 0,$$

und wegen Borel-Cantelli sogar

$$D_n \xrightarrow{P-f.s.} 0.$$

11.2 Ein bisschen Vapnik-Chervonenkis-Theorie

"In the early 1930s Popper suggested his idea of falsifiability. It was considered both as the demarcation line between metaphysics and natural science as well as a justification of Occam's Razor principle. Popper developed the falsifiability idea over his entire lifetime: his first publication appeared in German in 1934 his last addition was in an English edition in 1972. This idea is considered as one of the most important achievements in the philosophy of science of the 20th century. Fisher's philosophy of applied statistics and Popper's justification of the dependence of generalization ability on the number of entities formed the classical paradigm of philosophical realism for induction[20]. The continuation of the Kolmogorov-Glivenco-Cantelli line of theoretical statistics led to the development of the VC theory that reflected the philosophical instrumentalism paradigm."[Vapnik, 2006b, S.482]

In diesem Abschnitt wollen wir uns mit der Frage befassen, unter welchen Voraussetzungen die Mächtigkeit von $\mathcal{S}^{f,n}$ auch in allgemeineren Situationen als der Situation des Hauptsatzes nur polynomial in n wächst, und wir somit immer noch mit obigem Beweis der gleichmäßigen Konvergenz durchkommen. Diese Fragestellung ist Gegenstand der sogenannten

[20] Ich bin mir nicht ganz sicher, ob man im Zusammenhang mit Popper von "induction" reden sollte: *"And the success of science is not based upon rules of induction, but depends upon luck, ingenuity, and the purely deductive rules of critical argument.... Induction, i.e. inference based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor one of scientific procedure."*[Popper, 2014, S.53]

Vapnik-Chervonenkis Theorie, die wesentlich durch die Aufsätze Vapnik and Chervonenkis [1968] bzw. Vapnik and Chervonenkis [1971] initiiert wurde. (Eine historische Perspektive dazu gibt [Bottou, 2013].) Die Vapnik-Chervonenkis Theorie ist üblicherweise eher Gegenstand von theoretischen Veranstaltungen zum maschinellen Lernen, was hat sie also in einer Vorlesung zur Maß- und Wahrscheinlichkeitstheorie zu suchen? Die Antwort ist ganz einfach: Es geht hier ganz klar um theoretische Statistik / Kombinatorik zur gleichmäßigen Konvergenz des empirischen Maßes gegen das wahre unterliegende Maß. Der Fakt, dass die Vapnik-Chervonenkis-Theorie eher im Kontext des theoretischen maschinellen Lernens geläufig ist, hat unter anderem auch eine historisch-wissenschaftskulturelle Kontingenz zum Grund:

“... The philosophy described above was more or less clear by the end of the 1960s. By that time there was no doubt that in analyzing the pattern recognition problem we came up with a new direction in the theory of generalization. The only question that remained was how to describe this new direction. Is this a new branch of science or is it a further development in classical statistics? This question was the subject of discussions in the seminars at the Institute of Control Sciences of the Academy of Sciences of USSR (Moscow). The formal decision, however, was made when it came time to publish these results in the Reports of Academy of Sciences of USSR. The problem was in which section of Reports it should be published - in 'Control Sciences (Cybernetics)' or in 'Statistics'. It was published as a contribution in the 'Control Science' section. This is how one of the leading statisticians of the time, Boris Gnedenco, explained why it should not be published in the 'Statistics' section:

'It is true that this theory came from the same roots and uses the same formal tools as statistics. However, to belong to the statistical branch of science this is not enough. Much more important is to share the same belief in the models and to share the same philosophy. Whatever you are suggesting is not in the spirit of what I am doing or what A. Kolmogorov is doing. It is not what our students are doing nor will it be what the students of our students do. Therefore, you must have your own students, develop your own philosophy, and create your own community.'

More than 35 years have passed since this conversation. The more time passed, the more impressed I became with Gnedenco's judgment. The next three decades (1970s, 1980s, and 1990s) were crucial for developments in statistics. After the shocking discovery that the classical approach suffers from the curse of dimensionality, statisticians tried to find methods that could replace classical methods in solving real-life problems. During this time statistics was split into two very different parts: theoretical statistics that continued to develop the classical paradigm of generative models, and applied statistics that suggested a compromise between theoretical justification of the algorithms and heuristic approaches to solving real-life problems. They tried to justify such a position by inventing special names for such activities (exploratory data analysis), where in fact the superiority of common sense over theoretical justification was declared. However, they never tried to construct or justify new algorithms using V.C.-theory. Only after SVM technology became a dominant force in data mining methods did they start to use its technical ideas (but not its philosophy) to modify classical algorithms..." [Vapnik, 2006b, S.422f]

Definition 3 (Projektion, growth-function, shatterable set). Sei $\mathcal{S} \subseteq 2^V$ ein Mengensystem auf der Grundmenge V. Definiere für eine **beliebige** Menge $A \subseteq V$ (A muss nicht aus $\mathcal{S}$ sein) die Projektion von $\mathcal{S}$ auf A als:

$$\mathcal{S}_A := \{A \cap B \mid B \in \mathcal{S}\}.$$

Die sogenannte growth-function (Wachstumsfunktion) $m^{\mathcal{S}}$ ist definiert als

$$m^{\mathcal{S}} : \mathbb{N} \longrightarrow \mathbb{N} : r \mapsto \max_{\substack{A \subseteq V, \\ |A| = r}} |\mathcal{S}_A|.$$

Für eine **beliebige** Menge $A \subseteq V$ sagen wir, dass A shatterable (zerschmetterbar) bezüglich $\mathcal{S}$ ist, wenn

$$\mathcal{S}_A = 2^A$$

gilt, d.h., wenn alle Teilmengen von A als Schnitte von A mit Mengen aus $\mathcal{S}$ erzeugbar sind. Mit $sh_V(\mathcal{S})$ bezeichnen wir die Menge aller bezüglich $\mathcal{S}$ zerschmetterbaren Teilmengen von V.

Bemerkung 11.11 *Es gilt:*

$$\mathcal{S}_A \subseteq 2^A;\ |\mathcal{S}_A| \leq 2^{|A|}$$
$$m^{\mathcal{S}}(r) \leq 2^r;\ |\mathcal{S}^{f,n}| \leq m^{\mathcal{S}}(2n).$$

Ziel der nun folgenden Überlegungen ist es, die Mächtigkeit eines Mengensystems $\mathcal{S}$ mit rein kombinatorischen Mitteln nach oben abzuschätzen. Man kann sich leicht überlegen, dass große zerschmetterbare Mengen das Mengensystem $\mathcal{S}$ groß machen: Ist A shatterable, dann kann jede Teilmenge $B \subseteq A$ als Schnitt von A und einer Menge aus $\mathcal{S}$ dargestellt werden, woraus bereits $\mathcal{S} \geq |2^A|$ folgt. Wir sind aber an einer Art Umkehrung dieser Aussage interessiert und wollen nun der folgenden Spekulation nachgehen:

Vermutung. Wenn $\mathcal{S}$ sehr groß ist, dann muss es eine große zerschmetterbare Menge (bezüglich $\mathcal{S}$) geben, bzw. umgekehrt: Gibt es keine große zerschmetterbare Menge (bezüglich $\mathcal{S}$), so ist $\mathcal{S}$ nicht zu groß.

Wenn dem so wäre, dann könnten wir unter Umständen die growth-function $m^{\mathcal{S}}$ geeignet abschätzen. Wenn die growth-function nur polyniomial wächst, dann würden wir mit dem Beweis des vorigen Abschnitts zur gleichmäßigen Glivenko-Cantelli-Eigenschaft durchkommen. Erstaunlicherweise gilt in der Tat eine sehr verblüffende kombinatorische Relation zwischen der Mächtigkeit von $\mathcal{S}$ und der Mächtigkeit der Menge der zerschmetterbaren Mengen von $\mathcal{S}$, die wir zur Beschränkung der growth-function nutzen können:

Satz 11.12 (Mächtigkeit eines Mengensystems, [Pajor, 1985]) *Für eine endliche Menge $V = \{1, \dots, m\}$ und ein beliebiges Mengensystem $\mathcal{S} \subseteq 2^V$ gilt*

$$|\mathcal{S}| \leq |sh_V(\mathcal{S})|. \tag{*}$$

Beweis. Der Beweis läuft über vollständige Induktion über die Mächtigkeit von V.

Induktionsanfang: Für $m = 1$ gilt (*), denn die leere Menge ist immer zerschmetterbar und für den Fall $\mathcal{S} = \{\emptyset, \{1\}\}$ ist $sh_V(\mathcal{S}) = \{\emptyset, \{1\}\}$.

Induktionsvoraussetzung: Gelte (*) für $\tilde{V} = \{1, \ldots, m-1\}$ und $\tilde{S} = 2^{\tilde{V}}$ beliebig.

Induktionsschritt: Betrachte $V = \{1, \ldots, m\}$ und $S \subseteq 2^V$ sowie

$$\begin{aligned}
\mathcal{S}_0 &= \{A \mid A \in \mathcal{S}, m \notin A\} && \subseteq 2^{\tilde{V}} \\
\mathcal{S}_1 &= \{A \backslash \{m\} \mid A \in \mathcal{S}, m \in A\} && \subseteq 2^{\tilde{V}}.
\end{aligned}$$

Dann ist $|\,\mathcal{S}\,| = |\,\mathcal{S}_0\,| + |\,\mathcal{S}_1\,|$. Nach Induktionsvoraussetzung ist $|sh_{\tilde{V}}(\mathcal{S}_0)| \geq |\,\mathcal{S}_0\,|$ und $sh_{\tilde{V}}(\mathcal{S}_1)| \geq |\,\mathcal{S}_1\,|$, woraus

$$|sh_{\tilde{V}}(\mathcal{S}_0)| + |sh_{\tilde{V}}(\mathcal{S}_1)| \geq |\,\mathcal{S}_0\,| + |\,\mathcal{S}_1\,| = |\,\mathcal{S}\,|$$

folgt. Weiterhin ist

$$sh_{\tilde{V}}(\mathcal{S}_0) \cup sh_{\tilde{V}}(\mathcal{S}_1) \subseteq sh_V(\mathcal{S}).$$

Betrachte jetzt ein beliebiges A aus dem Schnitt $sh_{\tilde{V}}(\mathcal{S}_0) \cap sh_{\tilde{V}}(\mathcal{S}_1)$.
Dann ist $m \notin A$ und $A \cup \{m\}$ ist in $sh_V(\mathcal{S})$. Das bedeutet, dass für jedes A, das im Schnitt $sh_{\tilde{V}}(\mathcal{S}_0) \cap sh_{\tilde{V}}(\mathcal{S}_1)$ liegt, die Menge $A \cup \{m\}$ eine weitere zerschmetterbare Menge ist.
Damit folgt aus

$$sh_V(\mathcal{S}) \supsetneq sh_{\tilde{V}}(\mathcal{S}_0) \cup sh_{\tilde{V}}(\mathcal{S}_1)$$

das

$$|sh_V \mathcal{S}| \geq |sh_{\tilde{V}}(\mathcal{S}_0)| + |sh_{\tilde{V}}(\mathcal{S}_1)| \geq |\,\mathcal{S}\,|$$

gilt.

Obiger Satz impliziert nun das folgende Resultat, das von mehreren Autoren innerhalb verschiedener mathematischer Gebiete (Kombinatorik, Modelltheorie, Statistik) publiziert wurde. (Die erste Publikation dazu scheint [Vapnik and Chervonenkis, 1968] zu sein.)

Satz 11.13 ([Sauer, 1972, Shelah, 1972, Vapnik and Chervonenkis, 1968]) *Sei $\mathcal{S} \subseteq 2^{\{1,\dots,m\}}$ derart, dass es keine zerschmetterbare Menge der Kardinalität $k+1$ gibt. Dann ist*

$$|\mathcal{S}| \leq \sum_{i=0}^{k} \binom{m}{i} \leq m^k + 1.$$

Beweis (Nur erste Ungleichung). Da mit jeder zerschmetterbaren Menge A auch jede Teilmenge $B \subseteq A$ zerschmetterbar ist, kann es unter obiger Voraussetzung keine zerschmetterbaren Mengen der Kardinalität größer als k geben. Damit ist

$$sh_V \subseteq \bigcup_{i=0}^{k} \{A \subseteq \{1,\dots,m\} \mid |A| = i\}$$

und es folgt

$$|\mathcal{S}| \leq |sh_V(\mathcal{S})| \leq \sum_{i=0}^{k} \binom{m}{i}.$$

Obiges Resultat gibt Anlass zu folgender Definition:

Definition 4 (Vapnik-Chervonenkis-Dimension). Für ein Mengensystem $\mathcal{S} \subseteq 2^V$ mit V nicht notwendigerweise endlich ist die Vapnik-Chervonenkis-Dimension (kurz V.C.-Dimension) von $\mathcal{S}$ definiert als die größte mögliche Kardinalität einer zerschemtterbaren Teilmenge von V. Gibt es zu jedem $r \in \mathbb{N}$ eine zerschmetterbare Menge der Kardinaliät r, so ist die V.C.-Dimension von $\mathcal{S}$ definiert als ∞.

Satz 11.14 (Hinreichende Bedingung für gleichmäßige Konvergenz) *Ist die V.C.-Dimension von $\mathcal{S}$ endlich, so gilt*

$$D_n \overset{P-f.s.}{\longrightarrow} 0.$$

Darüber hinaus ist $\mathcal{S}$ eine gleichmäßige Glivenko-Cantelli-Klasse genau dann, wenn die V.C.-Dimension von $\mathcal{S}$ endlich ist.

Satz 11.15 (Notwendige und hinreichende Konvergenzbedingung) *Eine notwendige und hinreichende Bedingung dafür, dass*

$$D_n \stackrel{P-f.s.}{\longrightarrow} 0$$

gilt, ist die Bedingung

$$\lim_{n\to\infty} \frac{\mathbb{E}_P(\log_2 |\mathcal{S}_{\{X_1,\ldots,X_n\}}|)}{n} = 0.$$

Anwendungen.

i) $V = \mathbb{R}$, $\mathcal{S} = \{(-\infty, c] \mid c \in \mathbb{R}\} \rightsquigarrow$ klassischer Kolmogorov-Smirnov Test:

$$H_0 : F = F_0 \quad vs. \quad H_1 : F \neq F_0 :$$

$$D_n \xrightarrow{H_0} 0 \quad \rightsquigarrow \text{Wähle } D_n \text{ als sensible Teststatistik.}$$

ii) $V = \mathbb{R}^d$, $\mathcal{S} = \{(-\infty, c] \mid c \in \mathbb{R}^d\}$ oder
$V = \mathbb{R}^d$ und $\mathcal{S} = \{A \subseteq \mathbb{R}^d \mid A \text{ konvex }\}$ (Achtung: V.C.-Dimension unendlich.) $\rightsquigarrow$ Verallgemeinerung des klassischen Kolmogorov-Smirnov Tests.

iii) Binäre Klassifikation $Z = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ i.i.d. verteilt gemäß unbekanntem P_Z mit $X_i : \Omega \longrightarrow V$ und $Y_i : \Omega \longrightarrow \{0, 1\}$. Ziel: Vorhersage von Y für gegebene Ausprägung von X mit deterministischer Funktion $f : V \longrightarrow \{0, 1\}$.

Gegeben eine Verlustfunktion L wäre eine natürliche Wahl einer Klassifikationsfunktion dasjenige f aus einer gegebenen Klasse $\mathscr{F}$ von Funktionen, das

$$\mathbb{E}_P(L(f(X), Y))$$

minimiert. Aber: P unbekannt $\rightsquigarrow$ minimiere nicht wahres Risiko, sondern empirisches Risiko. Wichtig dafür ist, dass das empirische Risiko das wahre Risiko nicht zu stark unterschätzt.

$$\text{generalization error} \;\leq\; \text{training error} + O\left(\sqrt{\frac{d}{n}}\right)$$

mit d= V.C.-Dimension von $\mathcal{S} = \{f^{-1}(\{1\}) \mid f \in \mathscr{F}\}$.

Literaturverzeichnis

R. R. Bahadur and L. J. Savage. The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Mathematical Statistics*, 27(4):1115–1122, 1956.

L. Bottou. In hindsight: Doklady akademii nauk SSSR, 181(4), 1968. In B. Schölkopf, Z. Luo, and V. Vovk, editors, *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pages 3–5. Springer, 2013.

C. Carathéodory. *Gesammelte Mathematische Schriften.* C.H. Beck, 1956. Bd. IV.

G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1954. doi: 10.5802/aif.53. URL `http://www.numdam.org/item/AIF_1954__5__131_0`.

D. Denneberg. *Non-additive measure and integral*, volume 27. Springer Science & Business Media, 2013.

R. M. Dudley, E. Giné, and J. Zinn. Uniform and universal Glivenko-Cantelli classes. *Journal of Theoretical Probability*, 4(3):485–510, Jul 1991. ISSN 1572-9230. doi: 10.1007/BF01210321. URL `https://doi.org/10.1007/BF01210321`.

R. Durrett. *Probability: Theory and examples*. Cambridge university press, 2010.

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

J. Elstrodt. *Maß-und Integrationstheorie*. Springer, 1996.

H.-O. Georgii. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter GmbH & Co KG, 2015.

D. Hilbert. *Grundlagen der Geometrie*. Teubner, 1899.

D. Hilbert. Mathematische Probleme: Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900. Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse (pp. 253–297), 1900.

A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, 1933*. Springer, Berlin, 1933.

J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15(1):211–

225, Dec 1922. ISSN 1432-1823. doi: 10.1007/BF01494395. URL https://doi.org/10.1007/BF01494395.

P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990.

J. Mycielski. A system of axioms of set theory for the rationalists. *Notices of the AMS*, 53(2):206–217, 2006.

A. Pajor. *Sous-espaces LN/L des espaces de Banach*. Number 16. Hermann, 1985.

K. Popper. *Conjectures and refutations: The growth of scientific knowledge*. Routledge, 2014.

N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145 – 147, 1972. ISSN 0097-3165. doi: https://doi.org/10.1016/0097-3165(72)90019-2. URL http://www.sciencedirect.com/science/article/pii/0097316572900192.

K. D. Schmidt. *Maß und Wahrscheinlichkeit*. Springer-Verlag, 2009.

R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, pages 39–48, 1974.

S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1): 247–261, 1972.

R. M. Solovay. A model of set-theory in which every set of reals is Lebesgue measurable. *Annals of Mathematics*, pages 1–56, 1970.

V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer, 2006a.

V. N. Vapnik. *Estimation of dependences based on empirical data. Empirical inference science: Afterword of 2006 / Vladimir Vapnik*. Springer, New York, NY and [Heidelberg], 2006b. ISBN 0387308652.

V. N. Vapnik and A. Y. Chervonenkis. The uniform convergence of frequencies of the appearance of events to their probabilities. In *Doklady Akademii Nauk*, volume 181, pages 781–783. Russian Academy of Sciences, 1968.

V. N. Vapnik and A. Y. Chervonenkis. Theory of uniform convergence of frequencies of events to their probabilities and problems of search for an optimal solution from empirical data(average risk minimization based on empirical data, showing relationship of problem to uniform convergence of averages toward expectation value). *Automation and Remote Control*, 32:207–217, 1971.

P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, 1991.

Notation

$\mathbb{N} = \{1,2,3,\ldots\}$ .... Menge der natürlichen Zahlen
$\mathbb{Q}$ .... Menge der rationalen Zahlen
$\mathbb{R}$ .... Menge der reellen Zahlen
$[a,b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ ..... abgeschlossenes reellwertiges Intervall
$]a,b[= \{x \in \mathbb{R} \mid a < x < b\}$ .... offenes reellwertiges Intervall
$[a,b[= \{x \in \mathbb{R} \mid a \leq x < b\}$ .... halboffenes reellwertiges Intervall
$]a,b] = \{x \in \mathbb{R} \mid a < x \leq b\}$ .... halboffenes reellwertiges Intervall
$\overline{R} = \mathbb{R} \cup \{-\infty, +\infty\}$ .... Menge der erweiterten reellen Zahlen
$[0,\infty] = \{x \in \overline{\mathbb{R}} \mid x \geq 0\}$ ... Menge der nichtnegativen erweiterten reellen Zahlen
$[0,\infty[= \{x \in \mathbb{R} \mid x \geq 0\}$ .... Menge der nichtnegativen reellen Zahlen
$A, B, C,$ .... typischerweise Mengen
A^c .... Komplement einer Menge A
$\mathcal{A}, \mathcal{D}, \mathcal{E}, \mathcal{F}, \mathcal{S}$ .... typischerweise Mengensysteme
$\mathfrak{B}(X)$ .... Borelsche σ-Algebra auf einem metrischen Raum X
$\mathfrak{L}^p$ .... Lebesguesche σ-Algebra auf $\mathbb{R}^d$
μ, ν, η, m .... typischerweise Mengenfunktionen
$f : A \longrightarrow B : x \mapsto f(x)$ .... Notation für eine Abbildung f
$\mathrm{im}(f) := \{f(x) \mid x \in A\}$ .... Bild eine Abbildung $f : A \longrightarrow B$
$f : A_{|C} \longrightarrow B$ Einschränkung der Abbildung f im Definitionsbereich auf die Teilmenge $C \subseteq A$
A^B .... Menge aller Abbildungen von B nach A
2^A .... Potenzmenge von A
$\mathbb{1}_A : \Omega \longrightarrow \{0,1\} : \omega \mapsto \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{sonst} \end{cases}$ ..... Indikatorfunktion von A
$A^{\mathbb{N}}$ .... Menge aller Folgen von Elementen aus A
$(x_n)_{n\in\mathbb{N}}$ .... Folge von Elementen
$(x_n)_{n\in\mathbb{N}} \in V^{\mathbb{N}}$ .... Folge von Elementen aus V
$(x_n)_{n\in\mathbb{N}} \subseteq V$ .... alternative Notation für Folge von Elementen aus V

I .... typischerweise (beliebige nicht-leere) Indexmenge
$(x_i)_{i\in I} \in V^I$ .... Familie von Elementen aus V
$\mathcal{T}$ .... Menge aller einfachen Funktionen
$\mathcal{T}_{\geq 0}$ .... Menge aller nichtnegativen einfachen Funktionen
$\mathcal{M}$ .... Menge aller messbaren Funktionen
$\mathcal{M}_{\geq 0}$ .... Menge aller nichtegativen messbaren Funktionen

$id_M : M \longrightarrow M : x \mapsto x$ . . . . identische Abbildung auf der Menge M
$x \oplus A := \{x + a \mid a \in A\}$ Komplexnotation für die um x verschobene Menge A
Φ . . . . Verteilungsfunktion der Standardnormalverteilung

Sachverzeichnis

L^p-Raum, 69
μ-fast überall bestehende Eigenschaften, 57
σ-Additivität, 17
σ-Algebra, 16
σ-Operator, 23
σ-Subadditivität, 15, 18
σ-endliches Maß, 17
Äquivalenzrelation, 8
äußeres Maß, 15

absolut stetig, 56
Algebra, 16
Approximation durch einfache Funktionen, 51
Auswahlaxiom, 6

Bildmaß, 37
Borel-Cantelli-Lemma, 73
Borel-Maß, 33
Borelsche σ-Algebra, 24

Carathéodory-σ-Algebra, 29
Charakterisierung der Messbarkeit, 38
Chernoff-Ungleichung, 77
Choquet-Integral, 67

Darstellungsunabhängigkeit, 48
diskretes Maß, 62
Dvoretzky-Kiefer-Wolfowitz-Ungleichung, 114

einfache Funktion, 46
empirisches Wahrscheinlichkeitsmaß, 108
endliches Maß, 17
Erwartungswert, 63
Erwartungswert für unabhängige Zufallsvariablen, 101
Erzeuger der Borelschen σ-Algebra, 24
Existenz und Eindeutigkeit des Produktmaßes, 97

Faktorraum, 8
fast sichere Konvergenz, 72
Fortsetzungs- und Eindeutigkeitsproblem, 13
Fortsetzungssatz von Carathéodory, 32

gleichmäßige
Glivenko-Cantelli-Klasse, 110
Glivenko-Cantelli-Klasse, 110
growth-function, 127

Hülle, 20
Hüllenoperator, 20
Hüllensystem, 20
Halbnorm, 69
Halbring, 30
Hauptsatz der Statistik, 114
Hinreichende Bedingung für gleichmäßige Konvergenz, 130

Inhalt, 30
inneres Maß, 26
Integral für integrierbare Funktionen, 57
Integral für nichtnegative messbare Funktionen, 53
Integration bezüglich eines diskreten Maßes, 62
Integration bezüglich eines Produktmaßes, 98
isotone Funktion, 31
Isotonie eines Maßes, 18

Kongruenzrelation, 70
Konvergenz fast sicher, 72
Konvergenz im p-ten Mittel, 71
Konvergenz in Verteilung, 81
Konvergenz in Wahrscheinlichkeit, 71
Konvergenz von unten für Funktionen, 51
Kovarianz, 63

Lebesgue-Borel-Maß, 33
Lebesgue-Maß, 33
Lebesguesche σ-Algebra, 32
Lebesguescher Zerlegungssatz, 60
Lebesguesches Prm̈aß, 30

Markov-Ungleichung, 75
Maß, 17
Maß mit Dichte, 55
Maßeindeutigkeitssatz, 34
Maßproblem, 5
Maßraum, 17
messbare Abbildung, 37
messbare Funktion, 51
Messbarkeit bezüglich der Produkt-σ-Algebra, 95
Modularitätsgleichung, 18

Nullmenge, 17

obere Erwartung, 68

Prämaß, 30
Produkt-σ-Algebra, 93
Produktmaß, 97
Projektion, 127

Quotientenraum, 8

Riemann-Integration, 60
Ring, 30

Satz von Banach und Tarski, 5
Satz von Banach und Tarski über das Maßproblem, 6
Satz von Beppo Levi, 54
Satz von Berry-Esseen, 105
Satz von der monotonen Konvergenz, 54

Satz von Fubini-Tonelli, 98
Satz von Glivenko-Cantelli, 114
Satz von Radon Nikodym, 59
Satz von Vitali, 5
Sauer-Shelah-Lemma, 130
schnittstabiler Erzeuger und Unabhängigkeit, 100
schwache Konvergenz, 81
Schwaches Gesetz der großen Zahlen, 103
Semiring, 30
separabler metrischer Raum, 11
shatterable set, 127
sichere Konvergenz, 72
Standarddarstellung, 47
Starkes Gesetz der großen Zahlen, 104
Stetigkeit von oben, 19
Stetigkeit von unten, 19
stochastische Unabhängigkeit, 99
Supremumsnorm, 50
Survivorfunktion, 65

Transformationssatz für Dichten, 59
translationsinvariantes Maß, 18
Tschebyscheff-Ungleichung, 76

Unabhängigkeit von Funktionen von Zufallsvariablen, 100
Unabhängigkeit von Zufallsvariablen und Produktbildmaß, 100
Ungleichung von Markov, 75
Ungleichung von Tschebyscheff, 76
universelle Glivenko-Cantelli-Klasse, 110
Ununterscheidbarkeitsrelation, 23

Vapnik-Chervonenkis-Dimension, 130
Vapnik-Chervonenkis-Lemma, 130
Vapnik-Chervonenkis-Theorie, 125
Varianz, 63
Verteilungsfunktion, 35
Verteilungsfunktion einer Zufallsvariable, 37
Vervollständigung eines Maßes, 33
vollständige Algebra, 16
Vollständiger Maßraum, 33

Wachstumsfunktion, 127
Wahrscheinlichkeitsmaß, 17
wesentliche Supremumsnorm, 51
Wohldefiniertheit des Integrals für $X \in \mathcal{M}_{\geq 0}$, 52

Zentraler Grenzwertsatz, 105
zerschmetterbare Menge, 127
Zufallsvariable, 37
Zusammenhang zwischen den Konvergenzbegriffen, 83

Sachverzeichnis

L^p-Raum, 69
μ-fast überall bestehende Eigenschaften, 57
σ-Additivität, 17
σ-Algebra, 16
σ-Operator, 23
σ-Subadditivität, 15, 18
σ-endliches Maß, 17
Äquivalenzrelation, 8
äußeres Maß, 15

absolut stetig, 56
Algebra, 16
Approximation durch einfache Funktionen, 51
Auswahlaxiom, 6

Bildmaß, 37
Borel-Cantelli-Lemma, 73
Borel-Maß, 33
Borelsche σ-Algebra, 24

Carathéodory-σ-Algebra, 29
Charakterisierung der Messbarkeit, 38
Chernoff-Ungleichung, 77
Choquet-Integral, 67

Darstellungsunabhängigkeit, 48
diskretes Maß, 62
Dvoretzky-Kiefer-Wolfowitz-Ungleichung, 114

einfache Funktion, 46
empirisches Wahrscheinlichkeitsmaß, 108
endliches Maß, 17
Erwartungswert, 63
Erwartungswert für unabhängige Zufallsvariablen, 101
Erzeuger der Borelschen σ-Algebra, 24
Existenz und Eindeutigkeit des Produktmaßes, 97

Faktorraum, 8
fast sichere Konvergenz, 72
Fortsetzungs- und Eindeutigkeitsproblem, 13
Fortsetzungssatz von Carathéodory, 32

gleichmäßige
Glivenko-Cantelli-Klasse, 110
Glivenko-Cantelli-Klasse, 110
growth-function, 127

Hülle, 20
Hüllenoperator, 20
Hüllensystem, 20
Halbnorm, 69
Halbring, 30
Hauptsatz der Statistik, 114
Hinreichende Bedingung für gleichmäßige Konvergenz, 130

Inhalt, 30
inneres Maß, 26
Integral für integrierbare Funktionen, 57
Integral für nichtnegative messbare Funktionen, 53
Integration bezüglich eines diskreten Maßes, 62
Integration bezüglich eines Produktmaßes, 98
isotone Funktion, 31
Isotonie eines Maßes, 18

Kongruenzrelation, 70
Konvergenz fast sicher, 72
Konvergenz im p-ten Mittel, 71
Konvergenz in Verteilung, 81
Konvergenz in Wahrscheinlichkeit, 71
Konvergenz von unten für Funktionen, 51
Kovarianz, 63

Lebesgue-Borel-Maß, 33
Lebesgue-Maß, 33
Lebesguesche σ-Algebra, 32
Lebesguescher Zerlegungssatz, 60
Lebesguesches Prm̈aß, 30

Markov-Ungleichung, 75
Maß, 17
Maß mit Dichte, 55
Maßeindeutigkeitssatz, 34
Maßproblem, 5
Maßraum, 17
messbare Abbildung, 37
messbare Funktion, 51
Messbarkeit bezüglich der Produkt-σ-Algebra, 95
Modularitätsgleichung, 18

Nullmenge, 17

obere Erwartung, 68

Prämaß, 30
Produkt-σ-Algebra, 93
Produktmaß, 97
Projektion, 127

Quotientenraum, 8

Riemann-Integration, 60
Ring, 30

Satz von Banach und Tarski, 5
Satz von Banach und Tarski über das Maßproblem, 6
Satz von Beppo Levi, 54
Satz von Berry-Esseen, 105
Satz von der monotonen Konvergenz, 54

Satz von Fubini-Tonelli, 98
Satz von Glivenko-Cantelli, 114
Satz von Radon Nikodym, 59
Satz von Vitali, 5
Sauer-Shelah-Lemma, 130
schnittstabiler Erzeuger und
    Unabhängigkeit, 100
schwache Konvergenz, 81
Schwaches Gesetz der großen
    Zahlen, 103
Semiring, 30
separabler metrischer Raum, 11
shatterable set, 127
sichere Konvergenz, 72
Standarddarstellung, 47
Starkes Gesetz der großen
    Zahlen, 104
Stetigkeit von oben, 19
Stetigkeit von unten, 19
stochastische Unabhängigkeit, 99
Supremumsnorm, 50
Survivorfunktion, 65

Transformationssatz für Dichten,
    59
translationsinvariantes Maß, 18
Tschebyscheff-Ungleichung, 76

Unabhängigkeit von Funktionen
    von Zufallsvariablen, 100
Unabhängigkeit von
    Zufallsvariablen und
    Produktbildmaß, 100
Ungleichung von Markov, 75
Ungleichung von Tschebyscheff,
    76
universelle
    Glivenko-Cantelli-Klasse,
    110
Ununterscheidbarkeitsrelation,
    23

Vapnik-Chervonenkis-
    Dimension,
    130
Vapnik-Chervonenkis-Lemma,
    130
Vapnik-Chervonenkis-Theorie,
    125
Varianz, 63
Verteilungsfunktion, 35
Verteilungsfunktion einer
    Zufallsvariable, 37
Vervollständigung eines Maßes,
    33
vollständige Algebra, 16
Vollständiger Maßraum, 33

Wachstumsfunktion, 127
Wahrscheinlichkeitsmaß, 17
wesentliche Supremumsnorm, 51
Wohldefiniertheit des Integrals
    für $X \in \mathcal{M}_{\geq 0}$, 52

Zentraler Grenzwertsatz, 105
zerschmetterbare Menge, 127
Zufallsvariable, 37
Zusammenhang zwischen den
    Konvergenzbegriffen, 83