

Inhaltsverzeichnis

0	Maß- und Integrationstheorie	5
0.1	Messräume und Maße	5
	Algebren und σ -Algebren	5
	Stetigkeitssatz	7
0.2	Konkrete Konstruktionen der Maß-Fortsetzung	8
	σ -Stetigkeit des eindimensionalen Inhalts	8
	Eindeutigkeitsatz für Maße	8
	Fortsetzungssatz	8
0.3	Integrierbare Funktionen	9
	Treppenfunktionen	9
	Nullmengen	11
	Fast überall bestehende Eigenschaften	11
	Messbare Funktionen und Abbildungen	11
	Integrierbare Funktionen	14
	Satz von Beppo Levi, monotone Konvergenz	14
	Eigenschaften des Integrals	15
	Integrale stetiger Funktionen auf Intervallen	16
	Satz von Lebesgue, majorisierte Konvergenz	17
	Uneigentliche Integrale	18
0.4	Produktmaße	19
	Existenz von Produktmaßen	20
	Satz von Fubini	21
0.5	Der Transformationssatz	23
	Der Transformationssatz	23
1	Grundlagen der Wahrscheinlichkeitstheorie	25
1.1	Elementare kombinatorische Probleme	25
	Permutationen mit Wiederholungen	25
	Permutationen ohne Wiederholungen	25
	Anzahl der r -elementigen Teilmengen	26
	Anzahl ungeordneter Proben ohne Wiederholungen	26
1.2	Wahrscheinlichkeitsräume	28
	Definition eines Wahrscheinlichkeitsraumes	28
	Bedingte Wahrscheinlichkeit	36
	Totale Wahrscheinlichkeit	36
	Formel von Bayes	36
1.3	Zufallsvariable	38

	Definition einer Zufallsvariablen	38
	Erwartungswert einer Zufallsvariablen	38
	Erwartungswert der Binomialverteilung	39
	Erwartungswert der geometrischen Verteilung	40
	Erwartungswert der Poissonverteilung	40
	Verteilungsfunktion, Verteilungsdichte	40
	Berechnung des Erwartungswertes bei Dichten	41
	Erwartungswert der Gamma-Verteilung	42
	Median einer Zufallsvariablen	42
1.4	Unabhängige Zufallsvariablen	44
	Unabhängige Ereignisse	44
	Unabhängigkeit von Zufallsvariablen	46
	Unkorrelierte Zufallsvariable	47
	Varianz, Standard-Abweichung	48
	Kovarianz	48
	Tschebyscheff Ungleichung	49
	Varianz der Binomial-Verteilung	49
	Varianz der Normal-Verteilung	50
	Varianz der geometrischen Verteilung	51
	Varianz der Poisson-Verteilung	51
	Varianz der Gamma-Verteilung	52
	Faltung	53
	Faltung von Maßen mit Dichten	54
1.5	Grenzwertsätze, der zentrale Grenzwertsatz	56
	Unendliche Produkte von Wahrscheinlichkeitsräumen	56
	Lemma von Borel-Cantelli	58
	Starkes Gesetz der großen Zahl	59
	Konvergenzbegriffe von Folgen von Zufallsvariablen	60
	Der zentrale Grenzwertsatz	62
	Approximation der Binomialverteilung	66
	Approximation der Binomialverteilung mit Korrekturterm	66
1.6	Mit der Normalverteilung zusammenhängende Verteilungen	68
	Multivariate Normalverteilung	68
	Beta-Verteilung	70
	Chiquadrat-Verteilung und Fisher-Verteilung	71
	Student t -Verteilung	72
2	Grundbegriffe der Statistik	75
2.1	Lineare Regression	75
	Methode der kleinsten Quadrate	76
2.2	Parameterabschätzung der Statistik	79
	Statistisches Modell	79
	Parametrische Modelle	79
	Standard-Modell	79
	Schätzer	80
	Konsistente und erwartungstreue Schätzer	81
	Maximum-Likelihood-Schätzer	82

Log-Likelihood-Schätzer	83
Das Gaußsche Produktmodell	84
Bias eines Schätzers	85
Beste Schätzer	87
2.3 Aussagen zur Testtheorie, Nichtparametrische Tests	89
Neyman-Pearcy-Tests	90
Der Wilcoxonsche Vorzeichentest	90
Der Signierte Rang Test von Wilcoxon	93
Runs-Tests	95
2.4 Mit der Normalverteilung zusammenhängende Tests	97
Beste Tests und unverfälschte Tests	97
Macht eines Tests	99
Relative Entropie	99
Testen des Mittelwertes	103
2.5 Verteilungstest	106
Empirische Verteilungsfunktion	106
Test von Kolmogorov–Smirnov	106

0. Maß- und Integrationstheorie

In diesem Abschnitt werden als Voraussetzungen für den weiteren Verlauf einige Resultate der allgemeinen Integrationstheorie bereitgestellt. Dabei werden schwierigere oder längere Beweise nicht wiedergegeben; diese können an anderer Stelle nachgelesen werden. Es wird eine allgemeine Integrationstheorie dargestellt, die für das Verständnis der Aussagen der Wahrscheinlichkeitstheorie und für die Integration von Funktionen in n Veränderlichen, also zur Berechnung von Mehrfachintegralen benötigt wird. Die hier behandelte Klasse der Lebesgue-integrierbaren Funktionen umfasst im Fall einer reellen Veränderlichen die Klasse der stetigen Funktionen und auf einem kompakten Intervall $[a, b]$. In diesem Fall stimmt auch das hier definierte Integral mit dem bisher betrachteten überein. Betrachten wir andererseits den Bereich der natürlichen Zahlen \mathbb{N} , so führt das Integral bezüglich des Zählmaßes unmittelbar zu der Theorie der absolut konvergenten Reihen.

0.1 Messräume und Maße

Bevor wir mit der eigentlichen Theorie beginnen, diskutieren wir zunächst ein Beispiel für die Bestimmung eines Flächeninhaltes eines Dreiecks mit Hilfe der als bekannt vorausgesetzten Flächeninhalt von Rechtecken:

0.1.1 Beispiel: (Flächenbestimmung eines rechtwinkligen Dreiecks) Gegeben sei ein rechtwinkliges Dreieck mit den Katheten der Längen a, b . Wir erhalten durch jeweilige Halbierung der Strecken 2^{n-1} Rechtecke $R_{n,j}$ mit den Kantenlängen $a2^{-n}$ beziehungsweise $b2^{-n}$ für alle $j = 1, \dots, 2^{n-1}$. Damit ergibt sich als Fläche

$$\lambda^2(F) = \sum_{n=1}^{\infty} \sum_{j=1}^{2^{n-1}} ab(2^{-n})^2 = \sum_{n=1}^{\infty} 2^{n-1} ab2^{-2n} = \frac{ab}{2}.$$

Dieses Beispiel zeigt, dass es sinnvoll ist, zu fordern, dass wenn eine Fläche F sich als abzählbare Vereinigung paarweise disjunkter Flächen F_n mit bekannter Fläche $\lambda^2(F_n)$ darstellen lässt, wir auch $\lambda^2(F) = \sum_{n=1}^{\infty} \lambda^2(F_n)$ erhalten.

Diese Eigenschaft charakterisiert entscheidend die Maße in der allgemeine Theorie. Es ergibt sich allerdings eine Schwierigkeit: Man kann zeigen, dass es keine σ -additive Funktion $\lambda : \mathcal{P}(\mathbb{R}^2) \rightarrow [0, \infty]$ gibt mit $\lambda([a_1, b_1] \times [a_2, b_2]) = (b_1 - a_1)(b_2 - a_2)$ für alle Rechtecke $[a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$. Daher betrachtet man nicht die Potenzmenge $\mathcal{P}(\mathbb{R}^2)$ sondern kleinere Mengensysteme $\mathcal{A} \subset \mathcal{P}(\mathbb{R}^2)$. Diese Mengensysteme sind allerdings so groß, dass sie alle "vernünftigen" Mengen enthalten. Wir kommen also zur folgenden Definition; dabei fixieren wir ab jetzt stets eine nicht leere Menge Ω . Von besonderen Interesse sind natürlich die Fälle $\Omega = \mathbb{N}, \mathbb{R}, \mathbb{R}^q$.

0.1.2 Definition: i) Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ heißt eine σ -**Algebra**, wenn die folgenden Bedingungen erfüllt sind:

- 1) $\emptyset, \Omega \in \mathcal{A}$,
- 2) $A, B \in \mathcal{A}$ impliziert $A^c = \text{Kompl}(A)$, $A \cap B$, $A \cup B$, $B \setminus A \in \mathcal{A}$.
- 3) Für jede Folge $(A_n)_{n=1}^\infty \subset \mathcal{A}$ gilt $\bigcup_{n=1}^\infty A_n \in \mathcal{A}$.

ii) Eine Abbildung $\mu : \mathcal{A} \rightarrow [0, \infty]$ auf einer σ -Algebra \mathcal{A} mit $\mu(\emptyset) = 0$ heißt ein **Maß**, wenn für jede **disjunkte Folge** $(A_n)_{n=1}^\infty \subset \mathcal{A}$ – also einer Folge mit $A_n \cap A_m = \emptyset$ für alle $n \neq m$ – stets gilt:

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

iii) Für ein Mengensystem $\mathcal{E} \subset \mathcal{P}(\Omega)$ bezeichne $\mathcal{A}(\mathcal{E})$ die von \mathcal{E} erzeugte σ -Algebra, also die kleinste σ -Algebra, die \mathcal{E} enthält.

Anmerkung: Die Mengensysteme $\{\emptyset, \Omega\}$ und $\mathcal{P}(\Omega)$ sind offenbar σ -Algebren. Weiter genügt es in der Definition i), 2) zu verlangen, dass aus $A, B \in \mathcal{A}$ stets $A \cup B$, $A^c \in \mathcal{A}$ folgt: Man beachte, dass für alle $A, B \in \mathcal{P}(\Omega)$ stets $A \cap B = (A^c \cup B^c)^c$ gilt. Da $\mathcal{P}(\Omega)$ eine σ -Algebra ist, folgt für jedes Mengensystem $\mathcal{E} \subset \mathcal{P}(\Omega)$ unmittelbar

$$\mathcal{A}(\mathcal{E}) = \bigcap \left\{ \mathcal{C} \mid \mathcal{C} \text{ ist } \sigma\text{-Algebra mit } \mathcal{E} \subset \mathcal{C} \right\}.$$

Ist $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine σ -Algebra, so heißt (Ω, \mathcal{A}) ein **Messraum**, ist zusätzlich μ ein Maß auf \mathcal{A} , so heißt $(\Omega, \mathcal{A}, \mu)$ ein **Maßraum**. Aus beweistechnischen Gründen ist es oft erforderlich, allgemeinere Mengensysteme zu betrachten als σ -Algebren: Ein Mengensystem $\mathcal{A} \subset \mathcal{P}(\Omega)$ mit den Eigenschaften 1) und 2) aus Teil i) der Definition heißt auch eine **Algebra** und eine Mengenfunktion $\mu : \mathcal{A} \rightarrow [0, \infty]$ mit $\mu(A \cup B) = \mu(A) + \mu(B)$ für alle $A, B \in \mathcal{A}$ mit $A \cap B = \emptyset$ und mit $\mu(\emptyset) = 0$ heißt **endlich additiv**; gilt dann zusätzlich

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

für alle disjunkten Folgen $(A_n)_{n=1}^\infty \subset \mathcal{A}$ mit $A = \bigcup_{n=1}^\infty A_n \in \mathcal{A}$, so heißt μ σ -**additiv**. Demnach sind alle Maße auf einer σ -Algebra genau die σ -additiven Mengenfunktionen.

In metrischen Räumen und dabei speziell in den Fällen \mathbb{R} und \mathbb{R}^q ist eine spezielle σ -Algebra von Interesse, die von dem System \mathcal{O} aller offenen Teilmengen von Ω erzeugte σ -Algebra $\mathcal{B}(\Omega)$: die **Borel-Algebra** oder genauer die σ -Algebra der Borelschen Mengen. Da eine Teilmenge $A \subset \Omega$ genau dann offen ist, wenn das Komplement $A^c \subset \Omega$ abgeschlossen ist, wird die Borel-Algebra $\mathcal{B}(\Omega)$ gleichermaßen von dem System aller abgeschlossenen Teilmengen von Ω erzeugt. Im Fall $\Omega = \mathbb{R}^q$ sei $\mathcal{B}^q = \mathcal{B}(\mathbb{R}^q)$ und speziell $\mathcal{B} = \mathcal{B}^1 = \mathcal{B}(\mathbb{R})$. Weiter identifizieren wir hier \mathbb{C} mit \mathbb{R}^2 und erhalten damit $\mathcal{B}(\mathbb{C}) = \mathcal{B}^2$. In vielen Fällen wird die Borel-Algebra $\mathcal{B}(\Omega)$ auch von einem Teilsystem der offenen oder abgeschlossenen Mengen erzeugt:

Wir diskutieren hier den Fall \mathbb{R}^q : Für $a = (a_1, \dots, a_q)$, $b = (b_1, \dots, b_q)$ mit $a_j < b_j$ sei $]a, b[=]a_1, b_1[\times \dots \times]a_q, b_q[$ das von a und b erzeugte offene **Intervall** $\subset \mathbb{R}^q$. \mathcal{I}^q sei die Gesamtheit aller offenen Intervalle des \mathbb{R}^q .

0.1.3 Bemerkung: Für alle $q \in \mathbb{N}$ gilt $\mathcal{B}^q = \mathcal{A}(\mathcal{I}^q)$.

Beweis: Wegen $\mathcal{A}(\mathcal{I}^q) \subset \mathcal{B}^q$ müssen wir $U \in \mathcal{A}(\mathcal{I}^q)$ nachweisen für jede offene Teilmenge $U \subset \mathbb{R}^q$ mit $U^c \neq \emptyset$. Da \mathbb{Q} abzählbar ist, muss auch $U_0 = U \cap \mathbb{Q}^q$ abzählbar sein. Da U^c abgeschlossen ist gilt $r(u) := \min\{\|x - u\|_\infty \mid x \in U^c\} > 0$ für alle $u = (u_1, \dots, u_q)^T \in U$. Zu jedem $u \in U$ fixieren wir ein $u^0 \in U_0$ mit $2\|u - u^0\|_\infty < r(u)$ und setzen $r = r(u^0)$. Für alle $x \in U^c$ gilt $r(u) \leq \|u - x\|_\infty \leq \|u - u^0\|_\infty + \|u^0 - x\|_\infty$. Durch Infimumsbildung über alle $x \in U^c$ erhalten wir damit

$$r(u) \leq \|u - u^0\|_\infty + r(u^0) = \|u - u^0\|_\infty + r < \frac{1}{2} r(u) + r$$

also $2\|u - u^0\|_\infty < r(u) < 2r$. Damit folgt

$$u \in I(u^0) :=]u_1^0 - r, u_1^0 + r[\times \dots \times]u_q^0 - r, u_q^0 + r[\subset U.$$

Demnach gilt $U = \bigcup_{u^0 \in U_0} I(u^0)$, und da U_0 abzählbar ist, folgt $U \in \mathcal{A}(\mathcal{I}^q)$. □

0.1.4 Beispiel: Auf $\mathcal{A} = \mathcal{P}(\mathbb{N})$ definieren wir das **Zählmaß**

$$\mu(A) = \text{Anz}(A) \quad \text{für jede Teilmenge } A \subset \mathbb{N},$$

wobei die Anzahl der elemente $\text{Anz}(A)$ wie früher für jede Menge $A \subset \mathbb{N}$ definiert wird. Offenbar gilt $\mu(A \cup B) = \mu(A) + \mu(B)$ für alle disjunkten $A, B \in \mathcal{P}(\mathbb{N})$. Durch Fallunterscheidungen kann man auch leicht die σ -Additivität von $\mu = \text{Anz}$ nachweisen. $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ ist also ein Maßraum.

Wir notieren jetzt einige elementare Eigenschaften additiver und σ -additiver Mengenfunktionen. Die zweite Aussage wird auch als **Stetigkeitssatz** bezeichnet; diese Stetigkeitsaussagen sind dann in vielen Fällen leichter nachweisbar.

0.1.5 Satz: \mathcal{A} sei eine Algebra auf Ω und $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ sei eine additive Mengenfunktion.

i) μ ist monoton auf \mathcal{A} in dem Sinne, dass $\mu(A) \leq \mu(B)$ gilt für alle $A, B \in \mathcal{A}$ mit $A \subset B$.

ii) Es gilt $\mu(A \cup B) \leq \mu(A) + \mu(B)$ für alle $A, B \in \mathcal{A}$.

iii) Für $A, B \in \mathcal{A}$ mit $\mu(A) < \infty$ und $B \subset A$ gilt $\mu(A \setminus B) = \mu(A) - \mu(B)$.

Beweis: i) Für alle $A, B \in \mathcal{A}$ mit $A \subset B$ gilt $B = A \cup (B \setminus A)$, wegen $B \setminus A = B \cap A^c \in \mathcal{A}$ folgt $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$.

ii) Wegen $(A \cup B) = A \cup (B \setminus A)$ und i) folgt unmittelbar

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

iii) Es gilt $B \cap (A \setminus B) = \emptyset$ und daher $\mu(A) = \mu(B) + \mu(A \setminus B)$. □

0.1.6 Satz: \mathcal{A} sei eine Algebra auf Ω und $\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ sei eine additive Mengenfunktion. Die folgenden Aussagen sind äquivalent:

i) μ ist σ -additiv.

ii) Es gilt $\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$ für jede Folge $(A_n)_{n=1}^{\infty} \subset \mathcal{A}$ und $A \in \mathcal{A}$ mit $A \subset \bigcup_{n=1}^{\infty} A_n$.

iii) μ ist stetig von unten in dem Sinne, dass $\mu(A_n) \uparrow \mu(A)$ gilt bei $n \rightarrow \infty$ für jede Folge $(A_n)_{n=1}^\infty \subset \mathcal{A}$ mit $A_n \subset A_{n+1}$ für alle $n \in \mathbb{N}$ und $A = \bigcup_{n=1}^\infty A_n \in \mathcal{A}$.

Zusatz: Im Fall $\mu(\Omega) < \infty$ sind diese Aussage weiterhin äquivalent zu:

iv) μ ist stetig von oben in dem Sinne, dass $\mu(B_n) \downarrow 0$ gilt bei $n \rightarrow \infty$ für jede Folge $(B_n)_{n=1}^\infty \subset \mathcal{A}$ mit $B_n \supset B_{n+1}$ für alle $n \in \mathbb{N}$ und $\bigcap_{n=1}^\infty B_n = \emptyset$.

Auf den Beweis soll hier verzichtet werden.

0.2 Konkrete Konstruktionen der Maß-Fortsetzung

In diesem Abschnitt behandeln wir die mehr technischen Aspekte der Maß- und Integrationstheorie, die für den weiteren Verlauf der Theorie unverzichtbar sind. Allerdings werden später fast ausschließlich die hier hergeleiteten Resultate benutzt aber nicht die dazu erforderlichen Konstruktionsmethoden und Beweise; daher verzichten wir hier vollständig auf die Beweise und verweisen wieder auf die Literatur.

In diesem Abschnitt beschäftigen wir uns mit der Konstruktion der Fortsetzung einer σ -additiven Mengen-Funktion auf einer Algebra \mathcal{A}_0 zu einem Maß auf die erzeugte σ -Algebra $\mathcal{A}(\mathcal{A}_0)$ und formulieren auch eine Eindeutigkeitsaussage. Aus diesen Sätzen folgt dann speziell die Existenz des Lebesgue-Maßes λ_1 auf der Borel-Algebra $\mathcal{B}^1 \subset \mathcal{P}(\mathbb{R})$.

0.2.1 Satz: Es sei $\mathcal{F} = \mathcal{F}(\mathbb{R})$ die Gesamtheit aller Vereinigungen von jeweils endlich vielen Intervallen $\subset \mathbb{R}$. Dann ist \mathcal{F} eine Algebra, und die auf \mathcal{F} definierte Mengenfunktion λ , die jedem Intervall seine Länge als Wert zuordnet, ist σ -additiv.

Im weiteren Verlauf dieses Abschnittes befasst sich mit der eindeutigen Fortsetzbarkeit einer σ -additiven und σ -**endlichen Mengenfunktion** auf einer Algebra \mathcal{F} eindeutig zu einem eindeutig bestimmten Maß auf der von \mathcal{F} erzeugten σ -Algebra $\mathcal{A}(\mathcal{F})$. Dabei heißt eine Mengenfunktion $\mu : \mathcal{F} \rightarrow [0, \infty]$ σ -endlich, wenn eine Folge $(B_n)_{n=1}^\infty \subset \mathcal{F}$ existiert mit $\mu(B_n) < \infty$ für alle $n \in \mathbb{N}$ und mit $\Omega = \bigcup_{n=1}^\infty B_n$. Betrachten wir jetzt wieder die Situation von $\mathcal{F}(\mathbb{R})$ und der dort definierten σ -additiven Mengenfunktion λ , so lässt sich diese eindeutig zu einem Maß auf die von \mathcal{F} erzeugte Borel-Algebra \mathcal{B} fortsetzen. Wir bezeichnen diese Fortsetzung wieder mit λ und nennen es das eindimensionale **Lebesgue-Maß**. Für den allgemeinen Fall weisen wir jetzt die Eindeutigkeit und die Existenz getrennt nach.

0.2.2 Satz: \mathcal{A} sei eine σ -Algebra mit einem Erzeugendensystem \mathcal{F} im Sinne von $\mathcal{A} = \mathcal{A}(\mathcal{F})$. Es seien $\mu, \nu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ zwei Maße mit $\mu(A) = \nu(A) < \infty$ für alle $A \in \mathcal{F}$. Es gelte ferner $A \cap B \in \mathcal{F}$ für alle $A, B \in \mathcal{F}$, und es existiere eine Folge $(S_n)_{n=1}^\infty \subset \mathcal{F}$ mit $S_n \subset S_{n+1}$ für alle $n \in \mathbb{N}$ und mit $\bigcup_{n=1}^\infty S_n = \Omega$. Dann gilt $\mu(A) = \nu(A)$ für alle $A \in \mathcal{A}$.

Anmerkung: Dieser Satz besitzt eine Vielzahl von Anwendungen etwa in der Wahrscheinlichkeitstheorie: er gestattet sehr oft den Nachweis bestimmter Darstellungen von Maßen durch das Nachrechnen auf dazu besonders geeigneten Mengen.

0.2.3 Satz: \mathcal{A}_0 sei eine Algebra auf Ω , und $\mu : \mathcal{A}_0 \rightarrow \overline{\mathbb{R}}_+$ eine σ -additive und σ -endliche Mengenfunktion. Dann ist μ eindeutig fortsetzbar zu einem Maß auf die von \mathcal{A}_0 erzeugte σ -Algebra $\mathcal{A} = \mathcal{A}(\mathcal{A}_0)$.

Wir notieren jetzt die wichtigsten Spezialfälle der Theorie: das Lebesgue-Maß λ_q auf der Borel-Algebra \mathcal{B}^q . Die Aussage im Fall $q = 1$ folgt unmittelbar aus den vorstehenden Resultaten. der Fall $q > 1$ wird dann in Zusammenhang mit der Theorie der Produkträume behandelt: er folgt ohne weitere Konstruktion einfach aus der dort entwickelten allgemeinen Theorie.

0.2.4 Satz: Es existiert ein eindeutig bestimmtes Maß – das **Lebesgue-Maß** λ_q – auf \mathcal{B}^q mit $\lambda_q([a_1, b_1] \times \dots \times [a_q, b_q]) = (b_1 - a_1) \cdot \dots \cdot (b_q - a_q)$ für alle $a_1, \dots, a_q, b_1, \dots, b_q \in \mathbb{R}$ mit $a_j \leq b_j$ für alle $j = 1, \dots, q$.

0.3 Integrierbare Funktionen

In diesem Abschnitt führen wir die Klasse der integrierbaren Funktionen auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$ ein. Wir betrachten zunächst das Integral auf der Gesamtheit aller Stufenfunktionen; später wird dieses Integral dann auf eine Klasse nichtnegativer numerischer ($\overline{\mathbb{R}}$ -wertiger) Funktionen und dann auf komplexwertige Funktionen ausgedehnt. Für stetige Funktionen auf einem kompakten Intervall stimmt dann das so gewonnene Integral mit dem bisherigen Begriff überein. Zur Abkürzung verwenden wir im weiteren Verlauf auch häufig die Schreibweise \mathbb{K} für den Körper der reellen oder komplexen Zahlen. Wir fixieren jetzt stets einen Maßraum $(\Omega, \mathcal{A}, \mu)$. Weiter bilden wir für eine Teilmenge $A \in \mathcal{P}(\Omega)$ die **Indikatorfunktion** oder auch **charakteristische Funktion** $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$ durch

$$\mathbb{1}_A(t) = \begin{cases} 1 & \text{für } t \in A \\ 0 & \text{für } t \notin A \end{cases} .$$

Ohne Beweis notieren wir die einfachen Rechenregeln:

0.3.1 Bemerkung: Für alle $A, B \in \mathcal{P}(\Omega)$ gilt

$$\mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B, \quad \mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B}.$$

Ein System von Teilmengen $\{A_j \mid j \in I\} \subset \mathcal{A}$ nennen wir (paarweise) **disjunkt**, wenn $A_j \cap A_k = \emptyset$ gilt für alle $j, k \in I$ mit $j \neq k$. Eine Funktion $f : \Omega \rightarrow \mathbb{R}$ oder $f : \Omega \rightarrow \mathbb{C}$ heißt eine **Stufenfunktion** (in disjunkter Darstellung) oder eine **Treppenfunktion**, wenn endlich viele $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ und disjunkte $A_1, \dots, A_n \in \mathcal{A}$ existieren mit $f = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$.

$\mathcal{T}(\mathcal{A})$ sei die Gesamtheit aller Treppenfunktionen $f : \Omega \rightarrow \mathbb{C}$ und $\mathcal{T}_0(\mathcal{A})$ die Gesamtheit aller Treppenfunktionen $f = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$ mit $A_k \in \mathcal{A}$ und $\mu(A_k) < \infty$ für alle $k = 1, \dots, n$.

Ist $f \in \mathcal{T}_0(\mathcal{A})$ reellwertig, so schreiben wir $f \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})$, ist f zusätzlich nichtnegativ, also $f \geq 0$, so schreiben wir dann $f \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$. Für $f \in \mathcal{T}_0(\mathcal{A})$ definieren wir das **Integral** $\int f d\mu$ durch

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

Damit diese Definition sinnvoll ist, müssen wir nachweisen, dass der Wert dieses Integrals unabhängig von der speziellen Art der Darstellung von f definiert ist. Dazu zeigen wir die folgende Hilfsaussage, in der auch weitere Eigenschaften des Integrals auf $\mathcal{T}_0(\mathcal{A})$ hergeleitet werden.

0.3.2 Lemma: Für $f, g \in \mathcal{T}_0(\mathcal{A})$ und $\alpha \in \mathbb{K}$ gelten die folgenden Aussagen.

i) Im Fall $f = g$ folgt $\int f d\mu = \int g d\mu$; der Wert des Integrals ist also unabhängig von der speziellen Darstellung.

ii) $f, g \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})$ mit $f \leq g$ impliziert $\int f d\mu \leq \int g d\mu$.

iii) $f + \alpha g \in \mathcal{T}_0(\mathcal{A})$ mit $\int (f + \alpha g) d\mu = \int f d\mu + \alpha \int g d\mu$.

iv) $|\int f d\mu| \leq \int |f| d\mu$.

Beweis: Für den weiteren Verlauf des Beweises verwenden wir die folgenden Bezeichnungen: Es seien $\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_m \in \mathbb{K}$, $A_1, \dots, A_r \in \mathcal{A}$, und $B_1, \dots, B_m \in \mathcal{A}$ jeweils paarweise disjunkt mit

$$f = \sum_{j=1}^r \alpha_j \mathbb{1}_{A_j}, \quad g = \sum_{k=1}^m \beta_k \mathbb{1}_{B_k}.$$

Wir nehmen $A_1 \cup \dots \cup A_r = B_1 \cup \dots \cup B_m$ an; ansonsten ergänzen wir die Darstellung der Funktionen durch $\alpha_0 = 0$, $\beta_0 = 0$ und

$$A_0 = (B_1 \cup \dots \cup B_m) \setminus (A_1 \cup \dots \cup A_r) \quad \text{und} \quad B_0 = (A_1 \cup \dots \cup A_r) \setminus (B_1 \cup \dots \cup B_m).$$

Damit erhalten wir dann eine Darstellung mit $A_0 \cup \dots \cup A_r = B_0 \cup \dots \cup B_m$. Wegen der Disjunktheit beider Mengensysteme gilt für alle j, k weiter

$$\begin{aligned} A_j &= (B_1 \cup \dots \cup B_m) \cap A_j = (B_1 \cap A_j) \cup \dots \cup (B_m \cap A_j), \text{ also} \\ \mu(A_j) &= \mu(B_1 \cap A_j) + \dots + \mu(B_m \cap A_j) \quad \text{und entsprechend} \\ \mu(B_k) &= \mu(A_1 \cap B_k) + \dots + \mu(A_r \cap B_k). \end{aligned}$$

i) Im Fall $x \in B_k \cap A_j$ folgt $\alpha_j = f(x) = g(x) = \beta_k$. Wir erhalten also stets

$$\alpha_j \mu(A_j \cap B_k) = \beta_k \mu(A_j \cap B_k) \quad \text{für alle } j = 1, \dots, r, k = 1, \dots, m.$$

Damit ergibt sich weiter aus der Vorbemerkung

$$\begin{aligned} \int f d\mu &= \sum_{j=1}^r \alpha_j \mu(A_j) = \sum_{j=1}^r \sum_{k=1}^m \alpha_j \mu(A_j \cap B_k) \\ &= \sum_{k=1}^m \sum_{j=1}^r \beta_k \mu(A_j \cap B_k) = \sum_{k=1}^m \beta_k \mu(B_k) = \int g d\mu. \end{aligned}$$

ii) Wie in i) erhalten wir stets

$$\alpha_j \mu(A_j \cap B_k) \leq \beta_k \mu(A_j \cap B_k) \quad \text{für alle } j = 1, \dots, r, k = 1, \dots, m$$

und daher weiter wie eben

$$\int f d\mu = \sum_{j=1}^r \alpha_j \mu(A_j) = \sum_{j=1}^r \sum_{k=1}^m \alpha_j \mu(A_j \cap B_k) \leq \sum_{k=1}^m \sum_{j=1}^r \beta_k \mu(A_j \cap B_k) = \int g d\mu.$$

iii) Wegen der Vorbemerkung erhalten wir unmittelbar

$$\begin{aligned} f + \alpha g &= \sum_{j=1}^r \alpha_j \mathbb{1}_{A_j} + \alpha \sum_{k=1}^m \beta_k \mathbb{1}_{B_k} = \sum_{j=1}^r \sum_{k=1}^m \alpha_j \mathbb{1}_{A_j \cap B_k} + \alpha \sum_{k=1}^m \sum_{j=1}^r \beta_k \mathbb{1}_{A_j \cap B_k} \\ &= \sum_{j=1}^r \sum_{k=1}^m (\alpha_j + \alpha \cdot \beta_k) \mathbb{1}_{A_j \cap B_k} \in \mathcal{T}_0(\mathcal{A}) \quad \text{und weiter} \end{aligned}$$

$$\begin{aligned} \int (f + \alpha g) d\mu &= \sum_{j=1}^r \sum_{k=1}^m (\alpha_j + \alpha \cdot \beta_k) \mu(A_j \cap B_k) \\ &= \sum_{j=1}^r \alpha_j \sum_{k=1}^m \mu(A_j \cap B_k) + \alpha \sum_{k=1}^m \beta_k \sum_{j=1}^r \mu(A_j \cap B_k) \\ &= \sum_{j=1}^r \alpha_j \mu(A_j) + \alpha \sum_{k=1}^m \beta_k \mu(B_k) = \int f d\mu + \alpha \int g d\mu . \end{aligned}$$

iv) Wegen der Disjunktheit der $A_1, \dots, A_r \in \mathcal{A}$ gilt $|f| = \sum_{j=1}^r |\alpha_j| \mathbb{1}_{A_j}$, also folgt

$$\left| \int f d\mu \right| = \left| \sum_{j=1}^r \alpha_j \mu(A_j) \right| \leq \sum_{j=1}^r |\alpha_j| \mu(A_j) = \int |f| d\mu . \quad \square$$

Um sofort die in dieser Theorie üblichen und sehr nützlichen Bezeichnungen verwenden zu können führen wir folgende Notationen ein: Eine Teilmenge $A \subset \Omega$ heißt eine μ -**Nullmenge** oder kurz Nullmenge, wenn ein $A' \in \mathcal{A}$ existiert mit $A \subset A'$ und mit $\mu(A') = 0$. Offenbar ist wegen 0.1.6 die abzählbare Vereinigung von Nullmengen wieder eine Nullmenge. Weiter sagen wir, dass eine Eigenschaft (*) auf Ω μ -**fast überall** gilt, wenn diese Eigenschaft überall außerhalb einer geeigneten μ -Nullmenge A gilt. Sind also etwa $f, g : \Omega \rightarrow \mathbb{R}$ Funktionen, so bedeutet $f \leq g$ μ -fast-überall (oder auch $f(t) \leq g(t)$ für μ -fast alle $t \in \Omega$), dass eine μ -Nullmenge $A \in \mathcal{A}$ existiert mit $f(t) \leq g(t)$ für alle $t \notin A$. Sind weiter $f_n, f : \Omega \rightarrow \mathbb{R}$ für alle $n \in \mathbb{N}$ Funktionen, so bedeutet $f_n \rightarrow f$ punktweise μ -fast-überall bei $n \rightarrow \infty$, dass eine μ -Nullmenge $A \in \mathcal{A}$ existiert mit $f_n(t) \rightarrow f(t)$ bei $n \rightarrow \infty$ für alle $t \in A^c$.

0.3.3 Definition: (Ω, \mathcal{A}) und $(\Omega_1, \mathcal{A}_1)$ seien Messräume. Eine Abbildung $f : \Omega \rightarrow \Omega_1$ heißt **messbar**, wenn $f^{-1}(A_1) \in \mathcal{A}$ gilt für alle $A_1 \in \mathcal{A}_1$.

Diese Definition hat formal Ähnlichkeit mit der Charakterisierung stetiger Funktionen mit Hilfe offener Mengen. Um die Messbarkeit besser beschreiben zu können benötigen wir folgende Bezeichnungen: Für eine Funktion $f : \Omega \rightarrow \overline{\mathbb{R}}$ und $\alpha \in \mathbb{R}$ bilden wir die Mengen

$$\begin{aligned} [f \leq \alpha] &= \{t \in \Omega \mid f(t) \leq \alpha\}, & [f < \alpha] &= \{t \in \Omega \mid f(t) < \alpha\}, \\ [f \geq \alpha] &= \{t \in \Omega \mid f(t) \geq \alpha\}, & [f > \alpha] &= \{t \in \Omega \mid f(t) > \alpha\}, \\ [f = \alpha] &= \{t \in \Omega \mid f(t) = \alpha\} \end{aligned}$$

Offenbar gelten die folgenden Eigenschaften:

$$[f < \alpha] = \bigcup_{n=1}^{\infty} [f \leq \alpha - \frac{1}{n}], \quad [f > \alpha] = ([f \leq \alpha])^c, \quad [f \geq \alpha] = ([f < \alpha])^c.$$

Besonders einfach lässt sich die Messbarkeit einer Funktion oder Abbildung überprüfen oder charakterisieren, wenn ein Erzeugendensystem vorliegt:

0.3.4 Bemerkung: (Ω, \mathcal{A}) und $(\Omega_i, \mathcal{A}_i)$ $i = 1, 2$ seien Messräume und $\mathcal{E} \subset \mathcal{P}(\Omega_1)$ mit $\mathcal{A}(\mathcal{E}) = \mathcal{A}_1$.

- i) $f : \Omega \rightarrow \Omega_1$ ist genau dann messbar, wenn $f^{-1}(A_1) \in \mathcal{A}$ gilt für alle $A_1 \in \mathcal{E}$.
- ii) $f : \Omega \rightarrow (\overline{\mathbb{R}}, \mathcal{B})$ ist genau dann messbar, wenn $[f \leq \alpha] \in \mathcal{A}$ gilt für alle $\alpha \in \mathbb{R}$.
- iii) $f : \Omega \rightarrow \Omega_1$ und $g : \Omega_1 \rightarrow \Omega_2$ seien messbar, dann ist $g \circ f : \Omega \rightarrow \Omega_2$ messbar.

Beweis: i) Ist f messbar, so gilt offenbar $f^{-1}(A_1) \in \mathcal{A}$ für alle $A_1 \in \mathcal{E} \subset \mathcal{A}_1$. Um die Umkehrung einzusehen setzen wir

$$\mathcal{A}_0 = \{B \in \mathcal{P}(\Omega_1) \mid f^{-1}(B) \in \mathcal{A}\}.$$

Sind jetzt $B, B_j \in \mathcal{P}(\Omega_1)$ für $j \in I$ so gilt

$$f^{-1}\left(\bigcup_{j \in I} B_j\right) = \bigcup_{j \in I} f^{-1}(B_j), \quad f^{-1}\left(\bigcap_{j \in I} B_j\right) = \bigcap_{j \in I} f^{-1}(B_j), \quad f^{-1}(B^c) = (f^{-1}(B))^c.$$

Daher ist \mathcal{A}_0 eine σ -Algebra auf Ω_1 mit $\mathcal{E} \subset \mathcal{A}_0$, also mit $\mathcal{A}_1 = \mathcal{A}(\mathcal{E}) \subset \mathcal{A}_0$.

ii) Für $\mathcal{E} = \{] - \infty, \alpha] \mid \alpha \in \mathbb{R}\}$ gilt $\mathcal{A}(\mathcal{E}) = \mathcal{B}^1$. Die Aussage folgt also aus i)

iii) Die Aussage gilt wegen $(g \circ f)^{-1}(A_2) = f^{-1}(g^{-1}(A_2)) \in \mathcal{A}$ für alle $A_2 \in \mathcal{A}_2$. \square

0.3.5 Beispiel: i) Es sei $f \in \mathcal{T}(\mathcal{A}, \mathbb{R}) : f = \sum_{j=1}^r \alpha_j \mathbb{1}_{A_j}$ mit paarweise disjunkten $A_1, \dots, A_r \in \mathcal{A}$ und $\alpha_1, \dots, \alpha_r \in \mathbb{R}$. Dann gilt $[f \leq \alpha] = \bigcup \{A_j \mid \alpha_j \leq \alpha\} \in \mathcal{A}$ für alle $\alpha \in \mathbb{R}$. Folglich ist f messbar.

ii) Ist (X, τ) ein metrischer Raum, so ist jede stetige Funktion $f : X \rightarrow \mathbb{K}$ messbar bezüglich der Borel-Algebren; da das Urbild jeder offenen Menge wieder offen ist folgt diese Aussage direkt aus 0.3.4 i).

In der folgenden Aussage zeigen wir, dass die Gesamtheit aller messbaren Funktionen sehr groß ist. So kann im Fall $(\mathbb{R}, \mathcal{B}^1)$ keine nichtmessbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ formelmäßig im üblichen Sinn angegeben werden, rein theoretisch kann die Existenz einer derartigen Funktion nachgewiesen werden. Derartige Aussagen, die für uns hier nicht wichtig sind, befinden sich in Büchern über Maßtheorie.

0.3.6 Satz: i) Es seien die Funktionen $f_n : \Omega \rightarrow \overline{\mathbb{R}}$ messbar für alle $n \in \mathbb{N}$. Dann sind auch $\sup\{f_n \mid n \in \mathbb{N}\}$, $\inf\{f_n \mid n \in \mathbb{N}\}$, $\limsup_{n \rightarrow \infty} f_n$, $\liminf_{n \rightarrow \infty} f_n$ messbare Funktionen.

ii) Eine Funktion $f : \Omega \rightarrow \overline{\mathbb{R}}$ ist genau dann messbar, wenn eine Folge $(f_n)_{n=1}^{\infty} \subset \mathcal{T}(\mathcal{A}, \mathbb{R})$ existiert mit $f_n(t) \rightarrow f(t)$ bei $n \rightarrow \infty$ im weiteren Sinne für alle $t \in \Omega$. Die Folge kann so gewählt werden, dass $|f_n(t)| \uparrow |f(t)|$ bei $n \rightarrow \infty$ für alle $t \in \Omega$ gilt.

iii) Sind die Funktionen $f, g : \Omega \rightarrow \mathbb{R}$ messbar, so sind auch $f + g$, $f \cdot g$, $\max\{f, g\}$, $\min\{f, g\}$ messbar.

iv) Ist $f : \Omega \rightarrow \mathbb{C}$ messbar, so sind auch die Funktionen $\operatorname{Re} f$, $\operatorname{Im} f$, $|f|$ messbar.

v) $f : \Omega \rightarrow \mathbb{C}$ ist genau dann messbar, wenn eine Folge $f_n \in \mathcal{T}(\mathcal{A})$ existiert mit $f_n(t) \rightarrow f(t)$ bei $n \rightarrow \infty$ für alle $t \in \Omega$. Die Folge kann so gewählt werden, dass $|f_n(t)| \uparrow |f(t)|$ bei $n \rightarrow \infty$ für alle $t \in \Omega$ gilt.

Beweis: i) Für $\alpha \in \mathbb{R}$ gilt $[\sup\{f_n \mid n \in \mathbb{N}\} > \alpha] = \bigcup_{n=1}^{\infty} [f_n > \alpha]$. Damit folgt die Aussage für $\sup\{f_n\}$ und $\inf\{f_n\}$. Wegen

$$\limsup_{n \rightarrow \infty} f_n = \inf \left\{ \sup\{f_m \mid n \leq m \in \mathbb{N}\} \mid n \in \mathbb{N} \right\}$$

sind daher die weiteren Aussagen ebenfalls klar.

ii) Zunächst sei $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar; es reicht, den Zusatz zu zeigen. Für alle $n \in \mathbb{N}$ und alle $1 \leq k \leq n2^n$ seien

$$A_k = \left[f < \frac{k}{2^n} \right] \cap \left[f \geq \frac{k-1}{2^n} \right] \text{ und } B_k = \left[-f < \frac{k}{2^n} \right] \cap \left[-f \geq \frac{k-1}{2^n} \right].$$

Wir setzen dann

$$f_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} (\mathbb{1}_{A_k} - \mathbb{1}_{B_k}).$$

Offenbar gilt $f_n(t) \rightarrow f(t)$ bei $n \rightarrow \infty$ im weiteren Sinne für alle $t \in \Omega$.

Die Umkehrung der Aussage folgt aus Aussage i).

iii) Wegen $f + g, f \cdot g \in \mathcal{T}(\mathcal{A})$ für alle $f, g \in \mathcal{T}(\mathcal{A})$ folgt der erste Teil der Aussage direkt aus ii). Wegen

$$[\max\{f, g\} \leq \alpha] = [f \leq \alpha] \cap [g \leq \alpha], \quad [\min\{f, g\} \leq \alpha] = [f \leq \alpha] \cup [g \leq \alpha]$$

für alle $\alpha \in \mathbb{R}$ ist auch der zweite Teil von iii) klar.

iv) Ist f messbar, so folgt, dass $\operatorname{Re} f, \operatorname{Im} f$ messbar sind. Da die trigonometrischen Funktionen stetig sind, folgt

$$|\gamma| = \sup\{\cos(s)\operatorname{Re} \gamma + \sin(s)\operatorname{Im} \gamma \mid s \in \mathbb{Q}\}$$

für alle $\gamma \in \mathbb{C}$. Daher gilt auch

$$|f(t)| = \sup \left\{ \cos(s)\operatorname{Re}(f(t)) + \sin(s)\operatorname{Im}(f(t)) \mid s \in \mathbb{Q} \right\}$$

für alle $t \in \Omega$. Da \mathbb{Q} abzählbar ist, folgt die Aussage jetzt aus i).

iv) Durch Betrachtung von $\operatorname{Re} f$ und $\operatorname{Im} f$ folgt diese Aussage leicht aus ii). □

Bevor wir die Klasse der integrierbaren Funktionen einführen notieren wir noch eine einfache Bemerkung, die dann besagt, dass der neu eingeführte Integralbegriff für Funktionen $f \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$ und damit für alle $f \in \mathcal{T}_0(\mathcal{A})$ mit dem früher definierten Integralbegriff übereinstimmt.

0.3.7 Bemerkung: *Es sei $f \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$. Dann gilt*

$$\int f \, d\mu = \sup \left\{ \int g \, d\mu \mid g \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+ \text{ mit } g \leq f \, \mu\text{-fast überall} \right\}.$$

Beweis: Wir betrachten $g \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$ mit $g \leq f$ μ -fast überall. Es sei $A \in \mathcal{A}$ mit $\mu(A) = 0$ und mit $f(t) \geq g(t)$ für alle $t \in A^c$. Wegen $\mu(B) = \mu(A \cap B) + \mu(A^c \cap B) = \mu(A^c \cap B)$ für alle $B \in \mathcal{A}$ dürfen wir $\Omega = A^c$ annehmen. Die Aussage folgt aus 0.3.2. \square

0.3.8 Definition: $f : \Omega \rightarrow \overline{\mathbb{R}}$ sei messbar mit $f(t) \geq 0$ für alle $t \in \Omega$. Wir definieren das Integral durch die Festsetzung

$$\int f d\mu = \sup \left\{ \int g d\mu \mid g \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+ \text{ mit } g \leq f \mu\text{-fast überall} \right\} \in [0, \infty].$$

Eine messbare Funktion $f : \Omega \rightarrow [0, \infty]$ nennen wir **integrierbar**, wenn $\int f d\mu < \infty$ gilt und schreiben $f \in \mathcal{L}^1(\mu)_+$; ansonsten schreiben wir auch $\int f d\mu = \infty$.

Eine messbare Funktion $f : \Omega \rightarrow \mathbb{C}$ heißt **integrierbar**, wenn $f_1, f_2, f_3, f_4 \in \mathcal{L}^1(\mu, \mathbb{R})_+$ existieren mit $f = f_1 - f_2 + if_3 - if_4$. In diesem Fall schreiben wir $f \in \mathcal{L}^1(\mu)$ und

$$\int f d\mu = \int f_1 d\mu - \int f_2 d\mu + i \int f_3 d\mu - i \int f_4 d\mu.$$

Es bleibt jetzt zu zeigen, dass diese Definition des Integrals für komplexwertige Funktionen widerspruchsfrei ist, dass also dieser Begriff unabhängig von der speziellen Darstellung definiert ist. Zunächst zeigen wir aber den Konvergenzsatz von der monotonen Konvergenz, der eine zentrale Rolle in der gesamten Theorie und in vielen Anwendungen spielt:

0.3.9 Satz: (Beppo Levi) Es seien $f, f_n : \Omega \rightarrow [0, \infty]$ messbare Funktionen mit $f_n(t) \leq f_{n+1}(t)$ für alle n und $f_n(t) \uparrow f(t)$ bei $n \rightarrow \infty$ für μ -fast alle $t \in \Omega$. Dann gilt

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu \in [0, \infty].$$

Beweis: Der Beweis dieser Aussage erfolgt in mehreren Schritten:

(I): Sind $f_1, f_2 : \Omega \rightarrow [0, \infty]$ messbar, so gilt

$$\int f_1 d\mu + \int f_2 d\mu \leq \int (f_1 + f_2) d\mu.$$

Beweis von (I): Es seien $g_1, g_2 \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$ beliebig mit $g_1 \leq f_1$ und $g_2 \leq f_2$ μ -fast überall. Es folgt $(g_1 + g_2) \leq (f_1 + f_2)$ μ -fast überall und daher

$$\int g_1 d\mu + \int g_2 d\mu = \int (g_1 + g_2) d\mu \leq \int (f_1 + f_2) d\mu.$$

Durch Supremumbildung über alle $g_1, g_2 \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$ mit $g_1 \leq f_1$ und $g_2 \leq f_2$ μ -fast überall folgt Aussage (I).

(II): Aus der Definition des Integrals folgt unmittelbar, dass die Folge $(\int f_n d\mu)_{n=1}^\infty$ monoton wachsend ist; folglich existiert der Grenzwert $\lim_{n \rightarrow \infty} \int f_n d\mu \leq \int f d\mu \in [0, \infty]$.

(III): Es sei $h \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$ mit $h \leq f$ μ -fast überall. Dann gilt $\int h d\mu \leq \lim_{n \rightarrow \infty} \int f_n d\mu$. Wenn diese Aussage gezeigt ist, folgt aus dieser unmittelbar die Aussage des Satzes aus der Definition des Integrals.

Beweis von (III): Es sei $A \in \mathcal{A}$ mit $\mu(A) = 0$ und $f_n(t) \uparrow f(t)$ bei $n \rightarrow \infty$ und $h(t) \leq f(t)$ für alle $t \in A^c$. Weiter sei $h = \sum_{k=1}^r \alpha_k \mathbb{1}_{A_k}$ mit $\alpha_1, \dots, \alpha_r > 0$ und mit $A_1, \dots, A_r \in \mathcal{A}$ paarweise disjunkt und $\mu(A_1), \dots, \mu(A_r) < \infty$. Wegen $\mu(A_j) = \mu(A_j \cap A) + \mu(A_j \cap A^c) = \mu(A_j \cap A^c)$ dürfen wir $A_j \subset A^c$ für alle $j = 1, \dots, r$ annehmen. Es sei $\varepsilon > 0$ mit $2\varepsilon < \min\{\alpha_1, \dots, \alpha_r\}$. Wir fixieren ein $j \in \{1, \dots, r\}$ und setzen $B_n = A_j \cap [f_n \geq \alpha_j - \varepsilon]$ für alle $n \in \mathbb{N}$. Wegen $f_n \leq f_{n+1}$ folgt $B_n \subset B_{n+1}$ für alle $n \in \mathbb{N}$. Für jedes $t \in A_j$ gilt $f_n(t) \rightarrow f(t) \geq \alpha_j > \alpha_j - \varepsilon$ bei $n \rightarrow \infty$ und daher $t \in \bigcup_{n=1}^{\infty} B_n$; es folgt $A_j = \bigcup_{n=1}^{\infty} B_n$. Der Stetigkeitssatz für Maße 0.1.5 impliziert $\mu(B_n) \rightarrow \mu(A_j)$ bei $n \rightarrow \infty$. Daher existiert ein $n_j \in \mathbb{N}$ mit $(\alpha_j - \varepsilon)\mu(B_n) \geq (\alpha_j - 2\varepsilon)\mu(A_j)$ für alle $n \geq n_j$. Aus $(\alpha_j - \varepsilon)\mathbb{1}_{B_n} \leq f_n \mathbb{1}_{B_n} \leq f_n \mathbb{1}_{A_j}$ auf Ω folgt

$$(\alpha_j - 2\varepsilon)\mu(A_j) \leq (\alpha_j - \varepsilon)\mu(B_n) = \int (\alpha_j - \varepsilon)\mathbb{1}_{B_n} d\mu \leq \int f_n \mathbb{1}_{A_j} d\mu$$

für alle $n \geq n_j$. Für alle $n \geq \max\{n_1, \dots, n_r\}$ folgt damit schließlich wegen (I)

$$\begin{aligned} \int h d\mu - 2\varepsilon \sum_{j=1}^r \mu(A_j) &= \sum_{j=1}^r (\alpha_j - 2\varepsilon)\mu(A_j) \leq \sum_{j=1}^r \int f_n \mathbb{1}_{A_j} d\mu \leq \int f_n \sum_{j=1}^r \mathbb{1}_{A_j} d\mu \\ &\leq \int f_n d\mu \leq \lim_{n \rightarrow \infty} \int f_n d\mu. \end{aligned}$$

Beim Grenzübergang $\varepsilon \downarrow 0$ folgt damit unmittelbar die Behauptung. □

Wir zeigen jetzt die wichtigsten Eigenschaften des Integrals; die hier formulierten Aussagen entsprechen den bereit früherer für das Integral stetiger Funktionen auf einem kompakten Intervall hergeleiteten Regeln.

0.3.10 Satz: *Es seien $f, g \in \mathcal{L}^1(\mu)$.*

- i) *Gilt $f = g$ μ -fast überall, so folgt $\int f d\mu = \int g d\mu$.*
- ii) *Es sei $\alpha \in \mathbb{C}$. Dann gilt $(f + \alpha g) \in \mathcal{L}^1(\mu)$ mit*

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

- iii) *Sind f und g reellwertig mit $f \leq g$ μ -fast überall, so folgt $\int f d\mu \leq \int g d\mu$.*
- iv) *Ist $h : \Omega \rightarrow \mathbb{C}$ messbar, so ist $h \in \mathcal{L}^1(\mu)$ äquivalent zu $|h| \in \mathcal{L}^1(\mu, \mathbb{R})_+$.*
- v) *Es gilt*

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

Beweis: (I) ii) für $f, g \in \mathcal{L}_1(\mu)_+$ und $0 < \alpha \in \mathbb{R}$: Beweis von (I): Wegen Satz 0.3.6 existieren Folgen $(f_n)_{n=1}^{\infty}, (g_n)_{n=1}^{\infty} \subset \mathcal{T}(\mathcal{A}, \mathbb{R})_+$ mit $f_n \uparrow f, g_n \uparrow g$ bei $n \rightarrow \infty$. Wegen $(f_n + \alpha g_n) \uparrow (f + \alpha g)$ bei $n \rightarrow \infty$ folgt aus dem Konvergenzsatz von Beppo Levi 0.3.9

$$\int f_n d\mu \rightarrow \int f d\mu, \int g_n d\mu \rightarrow \int g d\mu, \int (f_n + \alpha g_n) d\mu \rightarrow \int (f + \alpha g) d\mu.$$

Wegen 0.3.6 gilt $\int (f_n + \alpha g_n) = \int f_n d\mu + \alpha \int g_n d\mu$ für alle $n \in \mathbb{N}$; es folgt $\int (f + \alpha g) d\mu = \int f d\mu + \alpha \int g d\mu$.

i) Wir betrachten zunächst $f, g \in \mathcal{L}(\mu, \mathbb{R})$ und wählen dann $f_1, f_2, g_1, g_2 \in \mathcal{L}^1(\mu, \mathbb{R})_+$ mit $f = f_1 - f_2$, $g = g_1 - g_2$. Aus $f = g$ μ -fast überall folgt $f_1 + g_2 = f_2 + g_1$ μ -fast überall und daher gilt wegen Teil (I) des Beweises

$$\int f_1 d\mu + \int g_2 d\mu = \int (f_1 + g_2) d\mu = \int (f_2 + g_1) d\mu = \int f_2 d\mu + \int g_1 d\mu.$$

Wir erhalten unmittelbar die behauptete Gleichheit der Integrale, also die behauptete Eindeutigkeit. Der Fall $f, g \in \mathcal{L}^1(\mu\mathbb{C})$ kann dann durch Betrachtung der Realteile und der Imaginärteile der Funktionen leicht auf den reellen Fall zurückgeführt werden.

ii) Diese Aussage kann wegen der Definition des Integrals für komplexwertige Funktionen unmittelbar auf Teil (I) zurückgeführt werden.

iii) $f \leq g$ ist äquivalent zu $g - f \geq 0$ μ -fast überall; aus der Definition des Integrals für nicht-negative Funktionen folgt unmittelbar $\int (f - g) d\mu \geq 0$ und damit wegen ii) $0 \leq \int (g - f) d\mu = \int g d\mu - \int f d\mu$.

iv) Zunächst sei h integrierbar. Dann existieren $h_1, h_2, h_3, h_4 \in \mathcal{L}^1(\mu, \mathbb{R})_+$ mit $h = h_1 - h_2 + ih_3 - ih_4$. Aus $|h| \leq h_1 + h_2 + h_3 + h_4$ folgt unmittelbar aus (I) $|h| \in \mathcal{L}^1(\mu, \mathbb{R})_+$. Für die Umkehrung setzen wir $g_1 = \max\{\operatorname{Re} h, 0\}$, $g_2 = \max\{-\operatorname{Re} h, 0\}$, $g_3 = \max\{\operatorname{Im} h, 0\}$, $g_4 = \max\{-\operatorname{Im} h, 0\}$ und wegen $|h| \geq g_1, g_2, g_3, g_4$ folgt unmittelbar $\int |h_j| d\mu < \infty$ für $j = 1, \dots, 4$ und damit wegen $h = g_1 - g_2 + ig_3 - ig_4$ die Aussage.

v) Wir setzen wie eben $g_1 = \max\{\operatorname{Re} f, 0\}$, $g_2 = \max\{-\operatorname{Re} f, 0\}$, $g_3 = \max\{\operatorname{Im} f, 0\}$, $g_4 = \max\{-\operatorname{Im} f, 0\}$. Wegen 0.3.6 existieren Folgen $(g_n^{(k)})_{n=1}^\infty \subset \mathcal{T}(\mathcal{A}, \mathbb{R})_+$ mit $g_n^{(k)} \uparrow g_k \leq |f|$ und $|g_1^{(n)} - g_2^{(n)} + ig_3^{(n)} - ig_4^{(n)}| \uparrow |f|$ bei $n \rightarrow \infty$. Für alle $n \in \mathbb{N}$ setzen wir $f_n = g_1^{(n)} - g_2^{(n)} + ig_3^{(n)} - ig_4^{(n)}$; wegen Lemma 0.3.2 folgt unmittelbar $|\int f_n d\mu| \leq \int |f_n| d\mu$. Aus dem Satz von Beppo Levi folgt $\int g_n^{(k)} d\mu \rightarrow \int g_k d\mu$ bei $n \rightarrow \infty$ für $k = 1, \dots, 4$ und $\int |f_n| d\mu \rightarrow \int |f| d\mu$ bei $n \rightarrow \infty$. Damit ist die Aussage gezeigt worden. \square

Bevor wir weitere Eigenschaften integrierbarer Funktionen behandeln, zeigen wir zunächst, dass bei stetigen Funktionen auf einem kompakten Intervall das so mit dem Lebesgue-Maß definierte Integral mit dem früheren Integrationsbegriff übereinstimmt, man vergleiche dazu etwa Abschnitt 3.5. Im weiteren Verlauf verwenden wir die folgende Notation: Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $A \in \mathcal{A}$ und $f : A \rightarrow \mathbb{C}$ eine Funktion mit $f \mathbb{1}_A \in \mathcal{L}^1(\mu)$, so definieren wir das **Integral über die Teilmenge A** durch

$$\int_A f d\mu = \int f \cdot \mathbb{1}_A d\mu.$$

Im Fall $f \in \mathcal{L}^1(\mu)$ gilt offenbar stets $\mathbb{1}_A f \in \mathcal{L}^1(\mu)$ für alle $A \in \mathcal{A}$ wegen 0.3.10; daher ist die Bildung dieses Teilintegrals in dieser Situation stets sinnvoll.

0.3.11 Satz: *Es seien $a, b \in \mathbb{R}$ mit $a < b$. Ist $f : [a, b] \rightarrow \mathbb{C}$ stetig, so gilt*

$$\int_{[a, b]} f d\lambda = \int_a^b f(t) dt.$$

Beweis: Wir dürfen offenbar annehmen, dass f reellwertig ist mit $f \geq 0$. Man beachte Definition des früher eingeführten Integralbegriffes. Da jede Intervall-Treppenfunktion auch eine Treppenfunktion ist, erhalten wir $U(f) \subset \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+$ und daher die folgende Ungleichung

$$\begin{aligned} \int_a^b f(t) dt &= \sup \left\{ \int_a^b g(t) dt \mid g \in U(f) \right\} \\ &\leq \sup \left\{ \int g d\mu \mid g \in \mathcal{T}_0(\mathcal{A}, \mathbb{R})_+ \text{ mit } g \leq f \mu\text{-fast überall} \right\} = \int_{[a,b]} f d\lambda. \end{aligned}$$

Ersetzen wir jetzt f durch $-f$ und wählen dann $c \in \mathbb{R}$ mit $c - f \geq 0$ so erhalten wir

$$c(b-a) - \int_a^b f(t) dt = \int_a^b (c-f)(t) dt \leq \int_{[a,b]} (c-f) d\lambda = c(b-a) - \int_{[a,b]} f d\lambda$$

und aus beiden Ungleichungen die behauptete Gleichheit. □

0.3.12 Bemerkung: $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ sei der Maßraum mit dem Zählmaß μ . Für eine Funktion $f : \mathbb{N} \rightarrow \mathbb{C}$ und $n \in \mathbb{N}$ setzen wir $a_n = f(n)$. Dann ist $f \in \mathcal{L}^1(\mu)$ äquivalent zu $\sum_{n=1}^\infty |a_n| < \infty$. In diesem Fall gilt $\int f d\mu = \sum_{n=1}^\infty a_n < \infty$.

Beweis: Wir dürfen offenbar annehmen, dass f reellwertig ist mit $f \geq 0$. Für jedes $m \in \mathbb{N}$ setzen wir $f_m = \sum_{n=1}^m a_n \mathbb{1}_{\{n\}}$. Dann gilt

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

Beim Grenzübergang $m \rightarrow \infty$ folgt

$$\int f_m d\mu \uparrow \int f d\mu \quad \text{und} \quad \sum_{n=1}^m a_n \uparrow \sum_{n=1}^\infty a_n.$$

Damit ist die Aussage gezeigt worden. □

Wir zeigen jetzt den für Anwendungen der Theorie sehr wichtigen Konvergenzsatz von Lebesgue über die **majorisierte Konvergenz**:

0.3.13 Satz: (Lebesgue) Es seien $f, f_n : \Omega \rightarrow \mathbb{C}$ (oder $\overline{\mathbb{R}}$) messbare Funktionen mit $f_n(t) \rightarrow f(t)$ für μ -fast alle $t \in \Omega$ bei $n \rightarrow \infty$. Es existiere ferner ein $h \in \mathcal{L}^1(\mu, \mathbb{R})_+$ mit $|f_n| \leq h$ μ -fast-überall für alle $n \in \mathbb{N}$. Dann gilt $f, f_n \in \mathcal{L}^1(\mu)$ für alle $n \in \mathbb{N}$ und

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu \rightarrow 0 \text{ bei } n \rightarrow \infty$$

Beweis: Da die Vereinigung von abzählbar vielen Nullmengen wieder eine Nullmenge ist, existiert eine Nullmenge $A \in \mathcal{A}$ mit $|f_n(t)| \leq h(t)$ für alle $t \in A^c$ und alle $n \in \mathbb{N}$ und mit $f_n(t) \rightarrow f(t)$ bei $n \rightarrow \infty$ für alle $t \in A^c$. Die Abschätzung folgt unmittelbar aus Satz 0.3.10. Für alle $t \in A^c$ und $n \in \mathbb{N}$ setzen wir

$$g_n(t) = \sup \{ |f_m(t) - f(t)| \mid m \in \mathbb{N} \text{ mit } m \geq n \} .$$

Wegen $|f_n(t) - f(t)| \rightarrow 0$ bei $n \rightarrow \infty$ folgt $g_n(t) \downarrow 0$ bei $n \rightarrow \infty$ für alle $t \in A^c$. Weiter gilt $0 \leq 2h(t) - g_n(t)$ für alle $n \in \mathbb{N}$ und $2h(t) - g_n(t) \uparrow 2h(t)$ bei $n \rightarrow \infty$ für alle $t \in A^c$. Aus dem Satz von Beppo Levi 0.3.9 folgt

$$2 \int h d\mu - \int g_n d\mu = \int (2h - g_n) d\mu \rightarrow 2 \int h d\mu$$

bei $n \rightarrow \infty$, also $\int g_n d\mu \rightarrow 0$ bei $n \rightarrow \infty$. Wegen $\int |f_n - f| d\mu \leq \int g_n d\mu$ für alle $n \in \mathbb{N}$ folgt unmittelbar die Aussage. \square

Wir behandeln jetzt die Situation stetiger Funktionen $f : I \rightarrow \mathbb{C}$ mit einem nicht-kompakten Definitionsintervall nicht. Dann kann unter Umständen durch eine Grenzwertbestimmung ein uneigentliches Integral gebildet werden. Wir diskutieren jetzt den Zusammenhang mit dem in diesem Abschnitt eingeführten allgemeinen Integrationsbegriff.

0.3.14 Definition: $I \subset \mathbb{R}$ sei ein Intervall, das aus mehr als einem Punkt besteht. Wir setzen $a = \inf(I)$, $b = \sup(I) \in \overline{\mathbb{R}}$. Eine Funktion $f : I \rightarrow \mathbb{C}$ heie **uneigentlich integrierbar** auf I , wenn bei beliebigem $c \in I$ und beliebigen $u, v \in I$ mit $a < u < c < v < b$ die Funktion f auf den Intervallen $[u, c]$ und $[c, v]$ integrierbar ist und die Grenzwerte

$$\lim_{v \uparrow b} \int_c^v f(t) dt \quad \text{und} \quad \lim_{u \downarrow a} \int_u^c f(t) dt$$

existieren. Wir setzen

$$\int_{a+}^{b-} f(t) dt = \lim_{v \uparrow b} \int_c^v f(t) dt + \lim_{u \downarrow a} \int_u^c f(t) dt.$$

Man erkennt unmittelbar, dass die Definition der uneigentlichen Integrierbarkeit unabhängig von der speziellen Wahl des in der Definition auftretenden $c \in I$ ist: Sind $c, c' \in I$ und dann $u, v \in I$ mit $a < u < c < c' < v < b$, so folgt die behauptete Unabhängigkeit unmittelbar aus

$$\int_u^c f(t) dt + \int_c^v f(t) dt = \int_u^{c'} f(t) dt + \int_{c'}^v f(t) dt .$$

Entsprechend einfach folgt auch, dass jede steige Funktion $f : [a, b] \rightarrow \mathbb{C}$ auch auf $[a, b]$ uneigentlich integrierbar ist bei gleichem Wert der Integrale. Weiter gilt:

0.3.15 Bemerkung: Es sei $I =]a, b[\subset \overline{\mathbb{R}}$ ein Intervall mit $a, b \in \overline{\mathbb{R}}$, und es sei $f : I \rightarrow \mathbb{R}$ eine nichtnegative, stetige Funktion. Dann ist f genau dann auf I uneigentlich integrierbar, wenn f auf I integrierbar ist; es gilt dann

$$\int_I f d\lambda = \int_{a+}^{b-} f(t) dt .$$

Beweis: Es seien $a_n, b_n \in \mathbb{R}$ mit $a_n < b_n$ für alle $n \in \mathbb{N}$ und mit $a_n \downarrow a$ und $b_n \uparrow b$ im uneigentlichen Sinn bei $n \rightarrow \infty$. Wegen 0.3.11 gilt

$$\int_{a_n}^{b_n} f(t) dt = \int_{[a_n, b_n]} f d\lambda$$

für alle $n \in \mathbb{N}$. Bei dem Grenzübergang $n \rightarrow \infty$ folgt aus dem Konvergenzsatz von Beppo Levi 0.3.9 unmittelbar

$$\int_{a_n}^{b_n} f(t) dt \rightarrow \int_a^b f(t) dt \text{ bei } n \rightarrow \infty \iff \int_{[a_n, b_n]} f d\lambda \rightarrow \int_{]a, b[} f d\lambda \text{ bei } n \rightarrow \infty$$

im weiteren Sinn. Damit ist die Aussage gezeigt worden. □

0.4 Produktmaße

In diesem Abschnitt gehen wir von den zwei Maßräumen $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ aus und konstruieren auf der Produktmenge $\Omega_1 \times \Omega_2$ die Produkt- σ -Algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ und auf dieser dann ein Produktmaß $\mu_1 \otimes \mu_2$. Durch vollständige Induktion kann man dann auch unmittelbar ein Produkt von endlich vielen Maßräumen bilden. Als eine der wichtigsten Anwendungen folgt dann unmittelbar die Konstruktion des n -dimensionalen Lebesgue-Maßes λ^n auf dem Messraum $(\mathbb{R}^n, \mathcal{B}^n)$.

Für die allgemeine Konstruktion fixieren wir zwei σ -endliche Maßräume $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$.

0.4.1 Definition: Wir setzen $\Omega = \Omega_1 \times \Omega_2$. Die kleinste σ -Algebra \mathcal{A} auf Ω , die alle Mengen der Form $A_1 \times A_2$ mit $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$ enthält, heißt die von den σ -Algebren $\mathcal{A}_1, \mathcal{A}_2$ erzeugte **Produkt- σ -Algebra**. Wir schreiben $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$.

Entsprechend bilden wir im Fall von endlich vielen σ -endlichen Maßräumen $(\Omega_i, \mathcal{A}_i, \mu_i)$ für $i = 1, \dots, n$ die Produkt- σ -Algebra $\bigotimes_{i=1}^n \mathcal{A}_i$, die von dem System

$$\left\{ A_1 \times \dots \times A_n \mid A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n \right\}$$

erzeugte σ -Algebra auf $\Omega = \Omega_1 \times \dots \times \Omega_n$, also der kleinsten σ -Algebra auf Ω , die dieses Mengensystem enthält.

Betrachten wir jetzt die Borel-Algebra \mathcal{B}^n auf dem Raum \mathbb{R}^n , so gilt offenbar

$$\mathcal{B}^n = \bigotimes_{i=1}^n \mathcal{B}^1,$$

als n -faches Produkt des Maßraumes $(\mathbb{R}, \mathcal{B}^1)$. Man beachte dabei 0.1.3. Weiter definieren wir noch die Klasse der endlichen **Zylindermengen**

$$\mathcal{Z}_0 = \left\{ B_1 \times B_2 \mid B_1 \in \mathcal{A}_1, B_2 \in \mathcal{A}_2, \mu_1(B_1), \mu_2(B_2) < \infty \right\}.$$

0.4.2 Bemerkung: Für $i = 1, 2$ seien $\mathcal{E}_i \subset \mathcal{A}_i$ Mengensysteme gegeben mit $\mathcal{A}_i = \mathcal{A}(\mathcal{E}_i)$.

Weiter existiere zu $i = 1, 2$ eine Folge $(A_i^{(n)})_{n=1}^\infty \subset \mathcal{E}_i$ mit $\bigcup_{n=1}^\infty A_i^{(n)} = \Omega_i$. Wir setzen

$$\mathcal{E} = \left\{ A_1 \times A_2 \mid A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2 \right\} \text{ und erhalten } \mathcal{A}(\mathcal{E}) = \mathcal{A}_1 \otimes \mathcal{A}_2.$$

Speziell gilt $\mathcal{A}(\mathcal{Z}_0) = \mathcal{A}_1 \otimes \mathcal{A}_2$.

Beweis: Aus der Konstruktion der Produkt- σ -Algebra folgt unmittelbar $\mathcal{A}(\mathcal{E}) \subset \mathcal{A}_1 \otimes \mathcal{A}_2$. Wir müssen $\mathcal{A}(\mathcal{E}) \supset \mathcal{A}_1 \otimes \mathcal{A}_2$ einsehen. Offenbar gilt

$$A_1^\sim = A_1 \times \Omega_2, A_2^\sim = \Omega_1 \times A_2 \in \mathcal{A}_1 \otimes \mathcal{A}_2$$

für alle $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$. Aus der Konstruktion folgt ebenfalls, dass

$$\mathcal{A}_i^\sim = \left\{ A_i^\sim \mid A_i \in \mathcal{A}_i \right\}$$

für $i = 1, 2$ σ -Algebren auf $\Omega = \Omega_1 \times \Omega_2$ sind. Da für $i = 1, 2$ eine Folge $(A_i^{(n)})_{n=1}^\infty \subset \mathcal{E}_i$ existiert mit $\bigcup_{n=1}^\infty A_i^{(n)} = \Omega_i$, erhalten wir damit

$$\mathcal{A}_i^\sim = \mathcal{A}\left(\left\{ A_i^\sim \mid A_i \in \mathcal{E}_i \right\}\right) \subset \mathcal{A}\left(\left\{ A_1 \times A_2 \mid A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2 \right\}\right).$$

Wegen $A_1 \times A_2 = A_1^\sim \cap A_2^\sim$ für alle $A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$ ist damit die Aussage bewiesen worden. Der Zusatz folgt unmittelbar, da die Maße σ -endlich sind. \square

In der folgenden Aussage zeigen wir die Existenz des Produktmaßes $\mu = \mu_1 \otimes \mu_2$ auf dem Raum $\Omega = \Omega_1 \times \Omega_2$ mit $\mu(B_1 \times B_2) = \mu_1(B_1) \cdot \mu_2(B_2)$ für alle $B_1 \times B_2 \in \mathcal{Z}_0$:

0.4.3 Satz: Es seien $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ zwei σ -endliche Maßräume. Auf $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ existiert ein eindeutig bestimmtes Maß $\mu = \mu_1 \otimes \mu_2$ mit $\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$ für alle $A_1 \times A_2 \in \mathcal{Z}_0$. Für alle $A \in \mathcal{A}$ gilt weiter

$$\begin{aligned} \mu(A) &= \int_{\Omega_2} \left(\int_{\Omega_1} \mathbb{1}_A(s_1, s_2) d\mu_1(s_1) \right) d\mu_2(s_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} \mathbb{1}_A(s_1, s_2) d\mu_2(s_2) \right) d\mu_1(s_1). \end{aligned}$$

Dabei sind alle vorkommenden inneren Integrale messbare Funktionen auf Ω_1 beziehungsweise auf Ω_2 , so dass die Integrale mit Werten in $[0, \infty]$ stets definiert sind.

Auf den Beweis dieser Aussage soll verzichtet werden.

Ist jetzt weiter $f : \Omega \rightarrow \mathbb{R}_+$ eine Treppenfunktion, also

$$f = \alpha_1 \mathbb{1}_{A_1} + \dots + \alpha_n \mathbb{1}_{A_n} \text{ mit } A_1, \dots, A_n \in \mathcal{A}_1 \otimes \mathcal{A}_2 = \mathcal{A},$$

so folgt aus der Aussage der vorstehenden Satzes unmittelbar

$$\int f d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} f d\mu_1 \right) d\mu_2 \in [0, \infty].$$

Betrachten wir jetzt eine allgemeine messbare Funktion $f : \Omega \rightarrow \mathbb{R}_+$, so existiert wegen 0.3.6 eine Folge von Treppenfunktionen $(f_n)_{n=1}^\infty$ mit $f_n(t_1, t_2) \uparrow f(t_1, t_2)$ bei $n \rightarrow \infty$ für alle $(t_1, t_2) \in \Omega = \Omega_1 \times \Omega_2$. Aus dem Konvergenzsatz von Beppo Levi 0.3.13 folgt dann

$$\begin{aligned} \int f_n d\mu &\rightarrow \int f d\mu \text{ bei } n \rightarrow \infty \\ \int_{\Omega_1} f_n(s_1, t_2) d\mu_1(s_1) &\rightarrow \int_{\Omega_1} f(s_1, t_2) d\mu_1(s_1) \text{ für alle } t_2 \in \Omega_2 \text{ und} \\ \int_{\Omega_2} f_n(t_1, s_2) d\mu_2(s_2) &\rightarrow \int_{\Omega_2} f(t_1, s_2) d\mu_2(s_2) \text{ für alle } t_1 \in \Omega_1 \end{aligned}$$

bei $n \rightarrow \infty$. Eine erneute Anwendung des Konvergenzsatzes von Beppo Levi 0.3.9 liefert

$$\begin{aligned} \int_{\Omega_2} \left(\int_{\Omega_1} f_n(s_1, s_2) d\mu_1(s_1) \right) d\mu_2(s_2) &\rightarrow \int_{\Omega_2} \left(\int_{\Omega_1} f(s_1, s_2) d\mu_1(s_1) \right) d\mu_2(s_2) \\ \int_{\Omega_1} \left(\int_{\Omega_2} f_n(s_1, s_2) d\mu_2(s_2) \right) d\mu_1(s_1) &\rightarrow \int_{\Omega_1} \left(\int_{\Omega_2} f(s_1, s_2) d\mu_2(s_2) \right) d\mu_1(s_1) \end{aligned}$$

bei $n \rightarrow \infty$. Da wir eben für Teppenfunktionen die Gleichheit der Integrale bewiesen hatten, folgt unmittelbar die folgende Aussage:

0.4.4 Satz: (Fubini) *Es seien $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ σ -endliche Maßräume. Wir setzen $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ und $\mu = \mu_1 \otimes \mu_2$. Weiter sei $f : \Omega \rightarrow \mathbb{R}_+$ messbar. Dann gilt*

i) *Die Funktionen*

$$\Omega_2 \ni t_2 \mapsto \int_{\Omega_1} f(s_1, t_2) d\mu_1(s_1), \quad \Omega_1 \ni t_1 \mapsto \int_{\Omega_2} f(t_1, s_2) d\mu_2(s_2)$$

sind messbar.

ii)

$$\int f d\mu = \int_{\Omega_2} \left(\int_{\Omega_1} f(s_1, s_2) d\mu_1(s_1) \right) d\mu_2(s_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(s_1, s_2) d\mu_2(s_2) \right) d\mu_1(s_1).$$

Betrachten wir jetzt für eine messbare Funktion $f : \Omega \rightarrow \mathbb{C}$ den Realteil und den Imaginärteil und dann jeweils die Positiv- und Negativteile, so erhalten wir die folgende Aussage:

0.4.5 Satz: (Fubini, Tonelli) *Es seien $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ σ -endliche Maßräume. Wir setzen $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ und $\mu = \mu_1 \otimes \mu_2$. Weiter sei $f : \Omega \rightarrow \mathbb{C}$ messbar. Die folgenden Aussagen sind äquivalent:*

i) $f \in \mathcal{L}^1(\mu)$.

ii) *Es gilt $f(t_1, \cdot) \in \mathcal{L}^1(\mu_2)$ für μ_1 -fast alle $t_1 \in \Omega_1$ und*

$$\int_{\Omega_2} |f(\cdot, s_2)| d\mu_2(s_2) \in \mathcal{L}^1(\mu_1).$$

iii) Es gilt $f(\cdot, t_2) \in \mathcal{L}^1(\mu_1)$ für μ_2 -fast alle $t_2 \in \Omega_2$ und

$$\int_{\Omega_1} |f(s_1, \cdot)| d\mu_1(s_1) \in \mathcal{L}^1(\mu_2).$$

Ist eine dieser Aussagen erfüllt, so folgt

$$\int f d\mu = \int_{\Omega_2} \left(\int_{\Omega_1} f(s_1, s_2) d\mu_1(s_1) \right) d\mu_2(s_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(s_1, s_2) d\mu_2(s_2) \right) d\mu_1(s_1).$$

Wir wenden uns jetzt wieder der Situation der Raumes \mathbb{R}^q mit der Borel-Algebra \mathcal{B}^q zu. Bekanntlich lässt sich \mathbb{R}^q als q -faches Produkt von \mathbb{R} schreiben. Da weiter $\mathcal{B}^q = \mathcal{A}(\mathcal{I}^q)$

gilt, folgt aus den durchgeführten Konstruktionen, dass $\mathcal{B}^q = \bigotimes_{n=1}^q \mathcal{B}^1$ gilt. Setzen wir jetzt

weiter $\lambda_q = \bigotimes_{n=1}^q \lambda$, das Produktmaß auf der σ -Algebra $\mathcal{B}^q = \bigotimes_{n=1}^q \mathcal{B}^1$, so folgt speziell aus den vorstehenden Resultaten die folgende Aussage:

0.4.6 Satz: Es existiert ein eindeutig bestimmtes Maß λ_q auf dem Messraum $(\mathbb{R}^q, \mathcal{B}^q)$ mit

$$\lambda_q\left(]a_1, b_1[\times \dots \times]a_q, b_q[\right) = (b_1 - a_1) \cdot \dots \cdot (b_q - a_q)$$

für alle offenen Intervalle $]a_1, b_1[\times \dots \times]a_q, b_q[\subset \mathbb{R}^q$. Für eine beliebige Menge $A \in \mathcal{B}^q$ gilt

$$\lambda_q(A) = \int \left(\int \left(\dots \left(\int \mathbb{1}_A(s_1, s_2, \dots, s_q) d\lambda(s_q) \right) \dots \right) d\lambda(s_2) \right) d\lambda(s_1).$$

Eine entsprechenden Aussage gilt natürlich auch für Integrale messbarer Funktionen:

0.4.7 Folgerung: $f : \mathbb{R}^q \rightarrow \mathbb{C}$ sei messbar mit

$$\int \left(\int \left(\dots \left(\int |f(s_1, s_2, \dots, s_q)| d\lambda(s_q) \right) \dots \right) d\lambda(s_2) \right) d\lambda(s_1) < \infty.$$

Dann gilt $f \in \mathcal{L}^1(\lambda_q)$ mit

$$\int_{\mathbb{R}^q} f d\lambda_q = \int \left(\int \left(\dots \left(\int f(s_1, s_2, \dots, s_q) d\lambda(s_q) \right) \dots \right) d\lambda(s_2) \right) d\lambda(s_1).$$

Dabei kann in den beiden vorstehenden Aussagen die Reihenfolge der Integration beliebig vertauscht werden. Ohne absolute Konvergenz bleibt diese Aussage nicht mehr richtig; man vergleiche dazu das folgende Beispiel:

0.4.8 Beispiel: Die Funktion $f : [0, 2\pi] \times [1, \infty[\rightarrow \mathbb{R}$ sei definiert durch $f(s, t) = \sin(s) \frac{1}{t}$. Dann gilt

$$\int_0^{2\pi} \sin(s) \frac{1}{t} ds = 0 \quad \text{und} \quad \int_1^{\infty} |\sin(s)| \frac{1}{t} dt = \infty$$

für alle $t \in [1, \infty[$ und alle $s \in]0, 2\pi[$, $s \neq \pi$. □

0.4.9 Beispiel: Man berechne

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \frac{1}{(1+t^2)(1+s^4)} \sin(s^3 t^2) dt \right) ds .$$

Die Reihenfolge der Integration darf wegen

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \left| \frac{\sin(s^3 t^2)}{(1+t^2)(1+s^4)} \right| dt \right) ds = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \frac{1}{(1+t^2)(1+s^4)} dt \right) ds = \int_{-\infty}^{\infty} \pi \frac{1}{(1+s^4)} ds < \infty$$

vertauscht werden. Da für alle fixierten $t \neq 0$ die Funktion $g: g(s) = \frac{\sin(s^3 t^2)}{(1+s^4)(1+t^2)}$ ungerade (d.h. $g(-s) = -g(s)$) ist, gilt $\int_{-\infty}^{\infty} g(s) ds = 0$ für alle $t \neq 0$. Es folgt

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \frac{\sin(s^3 t^2)}{(1+t^2)(1+s^4)} dt \right) ds = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \frac{\sin(s^3 t^2)}{(1+t^2)(1+s^4)} ds \right) dt = \int_{-\infty}^{\infty} 0 dt = 0 . \quad \square$$

0.4.10 Beispiel: Man berechne

$$\lambda_3(A) \text{ für } A = \left\{ (x, y, z) \in \mathbb{R}^3 \mid x, y, z \geq 0, x + y^2 + z^2 \leq 4 \right\} .$$

Wegen 0.4.6 gilt

$$\begin{aligned} \lambda_3(A) &= \int \mathbb{1}_A d\lambda_3 = \int_0^4 \left(\int_0^{\sqrt{4-x}} \left(\int_0^{\sqrt{4-x-y^2}} dz \right) dy \right) dx \\ &= \int_0^4 \left(\int_0^{\sqrt{4-x}} \sqrt{4-x-y^2} dy \right) dx \\ &= \int_0^4 \left(\int_0^{\pi/2} \left(\sqrt{(4-x) - (4-x)\sin^2(t)} \right) \sqrt{4-x} \cos(t) dt \right) dx \\ &= \int_0^4 \left(\int_0^{\pi/2} (4-x) \cos^2(t) dt \right) dx = \int_0^4 (4-x) \frac{\pi}{4} dx = 2\pi . \end{aligned} \quad \square$$

0.5 Der Transformationssatz

In diesem Abschnitt behandeln wir einige Begriffsbildungen für den konkreten Maßraum $(\mathbb{R}^q, \mathcal{B}^q, \lambda_q)$. So zeigen wir den Transformationssatz, der im Fall \mathbb{R}^q die bekannte Substitutionsregel ersetzt und die Berechnung einiger wichtiger Integrale ermöglicht.

0.5.1 Satz: (Transformationsatz) Es seien $\emptyset \neq U, W \subset \mathbb{R}^q$ offene Teilmengen und $\Phi : U \rightarrow W$ eine bijektive Abbildung, so dass Φ und Φ^{-1} stetig differenzierbar sind. Für alle $A \in \mathcal{B}^q(U) = \{A \in \mathcal{B}^q \mid A \subset U\}$ gilt dann

$$\lambda_q(\Phi(A)) = |\det(D\Phi(\cdot))| \cdot \lambda_q(A) = \int_A |\det(D\Phi(x))| d\lambda_q(x).$$

Ist jetzt $f : W = \Phi(U) \rightarrow \mathbb{C}$ messbar, so ist f genau dann integrierbar auf W , wenn $(f \circ \Phi) \cdot |\det(D\Phi)| : U \rightarrow \mathbb{C}$ auf U integrierbar ist. In diesem Fall gilt dann

$$\int_W f d\lambda_q = \int_U f(\Phi(x)) |\det(D\Phi(x))| d\lambda_q(x).$$

Aud den Beweis dieser Aussage soll verzichtet werden.

0.5.2 Beispiel: (Polarkoordinaten) Wir definieren

$$\Phi :]0, \infty[\times]0, 2\pi[\rightarrow \mathbb{R}^2 \setminus ([0, \infty[\times \{0\}) \text{ durch } \Phi(r, t) = (r \cos(t), r \sin(t))$$

für alle $0 < r \in \mathbb{R}$ und $0 < t < 2\pi$. Es gilt offenbar

$$\det(D\Phi(r, t)) = r \text{ für alle } 0 < r \in \mathbb{R}, 0 \leq t \leq 2\pi.$$

Daher folgt für jede messbare Funktion $f : \mathbb{R}^2 \rightarrow [0, \infty[$

$$\int_{\mathbb{R}^2} f(x) d\lambda_2(x) = \int_0^\infty \left(\int_0^{2\pi} r f(r \cos(t), r \sin(t)) dt \right) dr.$$

Eine entsprechende Aussage erhalten wir natürlich auch für integrierbare Funktionen.

0.5.3 Beispiel: Es gilt $\int_{-\infty}^\infty e^{-t^2} dt = \sqrt{\pi}$.

Wegen der vorstehenden Resultate erhalten wir

$$\begin{aligned} \left(\int_{-\infty}^\infty e^{-t^2} dt \right)^2 &= \left(\int_{-\infty}^\infty e^{-t^2} dt \right) \left(\int_{-\infty}^\infty e^{-s^2} ds \right) = \int_{-\infty}^\infty \left(\int_{-\infty}^\infty e^{-t^2-s^2} dt \right) ds \\ &= \int_{\mathbb{R}^2} e^{-|x|^2} d\lambda_2(x) = \int_0^\infty \left(\int_0^{2\pi} r e^{-r^2} dt \right) dr = 2\pi \int_0^\infty r e^{-r^2} dr = \pi. \end{aligned}$$

1. Grundlagen der Wahrscheinlichkeitstheorie

Das Ziel der ersten Abschnitte ist es den Begriff *Wahrscheinlichkeit* einzuführen und ausführlich zu diskutieren. Dabei setzen wir grundlegende Kenntnisse der allgemeinen Integrations-
theorie voraus, die hier die für eine schnelle und übersichtliche Darstellung der Wahrschein-
lichkeitstheorie unverzichtbar ist. Dabei benutzen wir hier lediglich die wichtigsten Resultate
und Begriffsbildungen ohne auf die teilweise komplizierten Konstruktionen einzugehen; das
gilt insbesondere für die doch umständlichen Konstruktionen der Maßfortsetzung. Als sehr
wichtige Aspekte treten als wichtiger Spezialfall die endlichen Wahrscheinlichkeitsräume auf;
um diese behandeln zu können, benötigen wir einige Aussagen über elementare Kombinatorik,
die hier kurz dargestellt werden sollen.

1.1 Elementare kombinatorische Probleme

1.1.1 Bemerkung: *Es seien $n \in \mathbb{N}$ und A eine Menge mit $\text{Anz}(A) = n$. Für $r \in \mathbb{N}$ sei A^r das r -fache kartesische Produkt*

$$A^r = \{(a_1, \dots, a_r) \mid a_1, \dots, a_r \in A\} .$$

Dann gilt $|A^r| = \text{Anz}(A^r) = n^r$.

Dabei bezeichnen wir jedes $a = (a_1, \dots, a_r) \in A^r$ als **r -Permutation mit Wiederholungen** aus der Menge $A = \{a_1, \dots, a_r\}$.

Diese Aussage kann sehr einfach durch vollständige Induktion nachgewiesen werden, auf die Durchführung dieses Beweises soll daher verzichtet werden.

1.1.2 Bemerkung: *Es seien $r, n \in \mathbb{N}$ und A, B Mengen mit $\text{Anz}(A) = r$, $\text{Anz}(B) = n$. Dann gilt*

$$\text{Anz}(\{\varphi \mid \varphi : A \rightarrow B\}) = n^r .$$

Beweis: Mit $A = \{a_1, \dots, a_r\}$ können wir jede Abbildung $\varphi : A \rightarrow B$ umkehrbar eindeutig mit dem r -Tupel

$$(\varphi(a_1), \dots, \varphi(a_r)) \in B^r$$

identifizieren. Die Aussage folgt also aus 1.1.1. □

1.1.3 Bemerkung: *Es seien $r, n \in \mathbb{N}$ mit $r \leq n$ und $A = \{a_1, \dots, a_n\}$ eine Menge. Wir setzen*

$$A^{(r)} = \{(b_1, \dots, b_r) \in A^r \mid b_i \neq b_k \text{ für alle } i \neq k\} .$$

- die Gesamtheit aller r -Permutationen ohne Wiederholungen von Elementen aus A . Dann gilt

$$\text{Anz}(A^{(r)}) = (n)_r := \frac{n!}{(n-r)!} .$$

Beweis: Die Aussage ist im Fall $r = n = 1$ offenbar richtig.

Es sei jetzt $m \in \mathbb{N}$, und die Aussage sei richtig für alle $r, n \in \mathbb{N}$ mit $r \leq n \leq m$.

Wir setzen $n = m + 1$. Im Fall $r = 1$ ist nichts zu zeigen. Es muss die Aussage im Fall $(r + 1) \leq m + 1$ nachgewiesen werden. Zu jedem beliebigen $b_0 \in A$ bilden wir $A_0 = A \setminus \{b_0\}$. Wegen $\text{Anz}(A_0) = n - 1 = m$ und der Induktionsvoraussetzung existieren $(n - 1)_r = \frac{(n-1)!}{(n-1-r)!}$ unterschiedliche r -Permutationen von Elementen aus A_0 ohne Wiederholungen, also $(n - 1)_r$ unterschiedliche $(r + 1)$ -Permutationen (c_0, c_1, \dots, c_r) von Elementen aus A mit $c_0 = b_0$. Da $b_0 \in A$ beliebig war, existieren

$$n \cdot \frac{(n - 1)!}{(n - 1 - r)!} = \frac{n!}{(n - (r + 1))!} = (n)_{r+1}$$

unterschiedliche Elemente $(b_0, \dots, b_r) \in A^{(r+1)}$, was zu zeigen war. □

1.1.4 Bemerkung: *Es seien $r, n \in \mathbb{N}$ mit $r \leq n$ und A, B Mengen mit $\text{Anz}(A) = r$ und $\text{Anz}(B) = n$.*

- i) *Es existieren genau $(n)_r = \frac{n!}{(n - r)!}$ injektive Abbildungen $\varphi : A \rightarrow B$.*
- ii) *Es existieren genau $n!$ bijektive Abbildungen $\varphi : A \rightarrow A$.*

Beweis: Es ist nur die Aussage i) zu zeigen; mit $B = A$ folgt ii) dann unmittelbar aus i). Im Fall $A = \{a_1, \dots, a_r\}$ kann jede injektive Abbildung $\varphi : A \rightarrow B$ durch

$$\varphi \mapsto (\varphi(a_1), \dots, \varphi(a_r)) \in B^{(r)}$$

umkehrbar eindeutig dargestellt werden. Die Aussage folgt daher aus 1.1.3. □

1.1.5 Satz: *Es seien $n \in \mathbb{N}$ und eine Menge mit $\text{Anz}(A) = n$. Zu jedem $r \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ mit $r \leq n$ existieren genau $\binom{n}{k}$ unterschiedliche Teilmengen B von A mit $\text{Anz}(B) = r$.*

Beweis: Die Aussage ist im Fall $r = n$ richtig. Die Aussage sei jetzt richtig für ein $n \in \mathbb{N}$ mit $r \leq n$. Es sei jetzt A eine Menge mit $\text{Anz}(A) = n + 1$. Wir fixieren $a_0 \in A$ und setzen $A_0 = A \setminus \{a_0\}$. Dann gilt $\text{Anz}(A_0) = n$. Im Fall $r = 0$ ist $B = \emptyset$ die einzige Teilmenge mit $\text{Anz}(B) = 0$, wegen $\binom{n}{0} = 1$ ist die Aussage richtig. Im Fall $r > 0$ existieren genau $\binom{n-1}{r-1}$ unterschiedliche Teilmengen $B \subset A$ mit $a_0 \in B$ und $\text{Anz}(B) = r$ und $\binom{n-1}{r}$ unterschiedliche Teilmengen $B \subset A$ mit $a_0 \notin B$ und $\text{Anz}(B) = r$. Wegen $\binom{n-1}{r-1} + \binom{n-1}{r} = \binom{n}{r}$ ist die Aussage gezeigt worden. □

Wir interessieren uns jetzt für die Anzahl unterschiedlicher ungeordneter Proben mit Wiederholungen auf einer Menge A mit $\text{Anz}(A) = n \in \mathbb{N}$. Zur Herleitung benötigen wir die folgende Identifizierung: Es sei $k \in \mathbb{N}$ und $a = (a_1, \dots, a_k), b = (b_1, \dots, b_k) \in A^k$. Wir schreiben

$$[a_1, \dots, a_k] = [b_1, \dots, b_k]$$

genau dann wenn eine bijektive Abbildung $\sigma = \{1, \dots, k\} \rightarrow \{1, \dots, k\}$ existiert mit $b_j = a_{\sigma(j)}$ für alle $j = 1, \dots, k$.

1.1.6 Satz: *A sei eine Menge mit $n = \text{Anz}(A) \in \mathbb{N}$ und es sei $k \in \mathbb{N}$. Dann existieren genau*

$$\binom{n+k-1}{k}$$

ungeordnete Proben mit Wiederholungen vom Umfang k .

Beweis: Wir dürfen offenbar $A = \{1, \dots, n\}$ annehmen. Ist jetzt $b = (b_1, \dots, b_k) \in A^k$, so bestimmen wir durch Umordnung ein $a = (a_1, \dots, a_k) \in A^k$ mit $[a_1, \dots, a_n] = [b_1, \dots, b_n]$ und mit

$$a_1 \leq a_2 \leq \dots \leq a_k.$$

Wir setzen jetzt $C = \{1, 2, \dots, n+k-1\}$ und $A_n^k = \{[a_1, \dots, a_k] \mid (a_1, \dots, a_k) \in A^k\}$ und definieren eine Abbildung

$$\varphi : A_n^k \rightarrow \{B \in \mathcal{P}(C) \mid \text{Anz}(B) = k\}$$

durch

$$\varphi([a_1, \dots, a_k]) = \{a_1, a_2 + 1, a_3 + 2, \dots, a_k + k - 1\}.$$

Offenbar gilt $\text{Anz}(\varphi([a_1, \dots, a_k])) = k$; weiter ist die Abbildung φ bijektiv. Aus 1.1.5 folgt daher

$$\text{Anz}(A_n^k) = \binom{n+k-1}{k}.$$

□

Mit der folgenden Aussage soll der Begriff der Binomialkoeffizienten verallgemeinert werden:

1.1.7 Bemerkung: *Es seien $r \in \mathbb{N}$, $n_1, \dots, n_r \in \mathbb{N}_0$ und $n = n_1 + \dots + n_r \in \mathbb{N}$. Ω sei eine Menge mit $\text{Anz}(\Omega) = n$. Es existieren genau*

$$\binom{n}{n_1, \dots, n_r} := \frac{n!}{n_1! \dots n_r!}$$

unterschiedliche Tupel (A_1, \dots, A_r) disjunkter Teilmengen $A_i \subset \Omega$ mit $\text{Anz}(A_i) = n_i$ für alle $i = 1, \dots, r$.

Dabei wird der Wert $\binom{n}{n_1, \dots, n_r}$ als **Multinomialkoeffizient** bezeichnet.

Beweis: Im Fall $r = 1$ ist nichts zu zeigen.

$(r \Rightarrow r+1)$: Es seien $n_0, n_1, \dots, n_r \in \mathbb{N}_0$ und $n = n_0 + n_1 + \dots + n_r$. Wir setzen $m = n_1 + \dots + n_r$ und dürfen $m \in \mathbb{N}$ und $m < n$ annehmen. Eine Menge $A_0 \subset \Omega$ mit $\text{Anz}(A_0) = n_0$ sei fixiert und es sei $\Omega_0 = \Omega \setminus A_0$. Es gilt $\text{Anz}(\Omega_0) = m$. Wegen der Induktionsvoraussetzung existieren $\binom{m}{n_1, \dots, n_r}$ unterschiedliche Tupel (A_1, \dots, A_r) disjunkter Teilmengen $A_i \subset \Omega_0$ mit $\text{Anz}(A_i) = n_i$ für alle $i = 1, \dots, r$. Da es $\binom{n}{n_0}$ unterschiedliche Teilmengen $A_0 \subset \Omega$ gibt mit $\text{Anz}(A_0) = n_0$ folgt die Aussage aus

$$\binom{n}{n_0, n_1, \dots, n_r} = \frac{n!}{n_0! n_1! \dots n_r!} = \frac{n!}{n_0! m!} \frac{m!}{n_1! \dots n_r!} = \binom{n}{n_0} \binom{m}{n_1, \dots, n_r}.$$

□

1.2 Wahrscheinlichkeitsräume

In diesem Abschnitt sollen Wahrscheinlichkeitsräume definiert werden. Dabei gehen wir von der allgemeinen Situation aus und behandeln dabei ausführlich die besonders wichtige Klasse der diskreten oder auch endlichen Wahrscheinlichkeitsräume ohne allerdings für diese die Resultate gesondert zu formulieren.

Gegeben sei eine nichtleere Menge Ω , deren Elemente ω werden als **Versuchsausgänge** oder **Elementarereignisse** bezeichnet. Die Grundmenge Ω heißt auch **Ereignisraum**, **Stichprobenraum** oder **Ergebnismenge**.

Oft ist es sinnvoll nicht Elementarereignisse zu betrachten: man fasst (gewisse) Elementarereignisse zusammen zu Ereignissen und fordert, dass dieses Mengensystem eine σ -Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$ ist – also ein Mengensystem mit den folgenden Eigenschaften:

- i) $\emptyset, \Omega \in \mathcal{A}$.
- ii) Für $A, B \in \mathcal{A}$ gilt $A \cup B, A \cap B, A \setminus B, A^c \in \mathcal{A}$.
- iii) Sind $A_n \in \mathcal{A}$ für alle $n \in \mathbb{N}$, so folgt $\bigcup_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

Zwei Ereignisse $A, B \in \mathcal{A}$ heißen **unvereinbar** (disjunkt), wenn $A \cap B = \emptyset$ gilt. Eine auf dieser σ -Algebra definierte Funktion $P : \mathcal{A} \rightarrow [0, 1]$ mit den Eigenschaften

$$P(\Omega) = 1, \quad P(\emptyset) = 0,$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad \text{für jede Folge } (A_n)_{n=1}^{\infty} \subset \mathcal{A} \text{ mit } A_n \cap A_m = \emptyset \text{ für alle } n \neq m$$

heißt ein **Wahrscheinlichkeitsmaß** auf (Ω, \mathcal{A}) . Dabei wird jedes $A \in \mathcal{A}$ mit $P(A) = 1$ als ein **sicheres Ereignis** und jedes $A \in \mathcal{A}$ mit $P(A) = 0$ als ein **unmögliches Ereignis** bezeichnet.

Um die Darstellung übersichtlich gestalten zu können treffen wir folgende **Verabredung**:

Im Fall $\text{Anz}(\Omega) \in \mathbb{N}$ - also einem endlichen Ereignisraum - oder im Fall einer abzählbaren Menge Ω sei stets $\mathcal{A} = \mathcal{P}(\Omega)$. In diesem Fall verzichten wir auf die explizite Erwähnung dieser Ereignismengen. Ein Wahrscheinlichkeitsmaß P soll in diesem Fall stets die folgende Darstellung besitzen: Für jedes $\omega \in \Omega$ sei $p_\omega = P(\{\omega\}) \in [0, 1]$, und es gelte $\sum_{\omega \in \Omega} p_\omega = 1$.

Für jedes $A \subset \mathcal{P}(\Omega)$ sei dann

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

Diese Darstellung ist stets im Fall $\text{Anz}(\Omega) \in \mathbb{N}$ gegeben.

Im Fall $\Omega = \mathbb{R}^n$ betrachten wir als Ereignismengensystem das System aller Borelmengen \mathcal{B}^n - die kleinste σ -Algebra $\subset \mathcal{P}(\mathbb{R}^n)$, die alle offenen (oder alle geschlossenen, oder alle Intervalle des \mathbb{R}^n) enthält, wenn es nicht anders vereinbart wird.

Im allgemeinen Fall ist die Restriktion auf eine σ -Algebra erforderlich aus dem folgenden Grund: Für den Fall $\Omega = \mathbb{R}$ kann man zeigen, dass kein Maß auf $\mathcal{P}(\mathbb{R})$ existiert, das jedem Intervall die Intervalllänge zuordnet. Das System $\mathcal{P}(\mathbb{R})$ enthält demnach zu viele Mengen - es ist daher aus mathematischen Gründen erforderlich, ein kleineres System zu

wählen. Dieses System ist das der Borel algebra \mathcal{B} . Auf \mathcal{B} existiert das Lebesgue-Maß λ : Ein Maß mit $\lambda([a, b]) = b - a$ für alle $a, b \in \mathbb{R}$, mit $a \leq b$. Der Vollständigkeit halber notieren wir einige elementare Eigenschaften allgemeinerer Wahrscheinlichkeitsmaße ohne Beweis. Die sehr einfachen Beweise folgen unmittelbar aus der Definition, sie finden sich außerdem in der Maßtheorie.

1.2.1 Bemerkung: (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum - also $\Omega \neq \emptyset$, $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine σ -Algebra, und $P : \mathcal{A} \rightarrow [0, 1]$ ein Wahrscheinlichkeitsmaß. Dann gilt

- i) $P(A^c) = 1 - P(A)$ für alle $A \in \mathcal{A}$.
- ii) $A, B \in \mathcal{A}$ mit $A \subset B$ impliziert $P(A) \leq P(B)$.
- iii) $P(B \setminus A) = P(B) - P(B \cap A)$ für alle $A, B \in \mathcal{A}$.
- iv) $P(A_1 \cup \dots \cup A_n) \leq \sum_{k=1}^n P(A_k)$ für alle $A_1, \dots, A_n \in \mathcal{A}$.
- v) $P(A \cup B) + P(A \cap B) = P(A) + P(B)$ für alle $A, B \in \mathcal{A}$.

Wir diskutieren jetzt einige der wichtigsten endlichen Wahrscheinlichkeitsräume, zunächst behandeln wir die **Laplace-Wahrscheinlichkeit**. Dabei besitzt jedes einzelne Elementarerereignis die gleiche Wahrscheinlichkeit:

1.2.2 Bemerkung: Es sei $n = \text{Anz}(\Omega) \in \mathbb{N}$. Für jedes $A \in \mathcal{P}(\Omega)$ setzen wir

$$P(A) = \frac{\text{Anz}(A)}{\text{Anz}(\Omega)}.$$

Dann ist P ein Wahrscheinlichkeitsmaß auf Ω .

Demnach gilt $p_\omega := P(\{\omega\}) = \frac{1}{n}$ für alle $\omega \in \Omega$. Auf den sehr einfachen Beweis dieser Bemerkung kann verzichtet werden.

1.2.3 Beispiel: Gegeben seien 4 Würfel. Man bestimme die Wahrscheinlichkeit P , dass die vier verschiedenen Augenzahlen alle verschieden sind.

Es sei $A = \{1, 2, \dots, 6\}$ - die Gesamtheit der Augenzahlen eines Würfels; weiter sei $\Omega = A^4$ der Ereignisraum des Werfens von 4 Würfeln. Nach 1.1.3 gilt mit

$$A^{(4)} = \{(b_1, \dots, b_n) \mid b_1, b_2, b_3, b_4 \in A \text{ paarweise verschieden}\}$$

Weiter gilt $\text{Anz}(A^{(4)}) = \frac{6!}{2!}$; für das gesuchte Ergebnis erhalten wir $P = \frac{6!}{2! 6^4} = \frac{5}{18}$. □

1.2.4 Beispiel: Zahlenlotto 6 aus 49: Das Tippen von 6 verschiedenen Zahlen zwischen 1 und 49 entspricht der Auswahl einer 6-elementigen Teilmenge $A \subset \{1, \dots, 49\}$. Wir setzen also $\Omega = \{A \subset \{1, \dots, 49\} \mid \text{Anz}(A) = 6\}$. Wegen 1.1.5 gilt $\text{Anz}(\Omega) = \binom{49}{6}$, man kann demnach auf $\binom{49}{6}$ verschiedene Arten ungeordnete Proben vom Umfang 6 entnehmen. Daher ist

$$p_6 = \frac{1}{\binom{49}{6}} = \frac{1}{13.983.816}$$

die Wahrscheinlichkeit für 6 "Richtige".

Wir wollen jetzt die Wahrscheinlichkeit p_j für j -“Richtige” bestimmen für ein $1 \leq j < 6$. Dazu sei $A \subset \{1, \dots, 49\}$ die Menge der getippten Zahlen. Ist jetzt $B \subset \{1, \dots, 49\}$ eine Menge mit $\text{Anz}(B) = 6$, die genau j der getippten Zahlen enthält, so gilt $\text{Anz}(A \cap B) = j$ und $\text{Anz}(A^c \cap B) = 6 - j$. Demnach existieren genau

$$\binom{6}{j} \binom{49-6}{6-j}$$

unterschiedliche Teilmengen $B \subset \{1, \dots, 49\}$, die jeweils genau j der getippten Zahlen enthalten. Da ein Laplace-Wahrscheinlichkeitsraum vorliegt, erhalten wir als Wahrscheinlichkeit p_j für genau j Richtige den Quotienten der Anzahl der günstigen zur Anzahl aller Fälle:

$$p_j = \frac{\binom{6}{j} \binom{43}{6-j}}{\binom{49}{6}}.$$

□

Wir wollen jetzt die eben behandelte Situation von j Richtigen allgemeiner formulieren und führen in diesem Zusammenhang die **hypergeometrische Verteilung** $h(m_1, m_2, n_1, n_2)$ ein; auf den sehr einfachen Beweis der folgenden Bemerkung unter Benutzung von 1.1.5 kann verzichtet werden:

1.2.5 Bemerkung: *Es seien $n, n_1, n_2 \in \mathbb{N}$ mit $n = n_1 + n_2$. Ω sei die Vereinigung zweier disjunkter Teilmengen $\Omega_1 \cup \Omega_2$ mit $\text{Anz}(\Omega_1) = n_1$, $\text{Anz}(\Omega_2) = n_2$. Weiter seien $m_1, m_2 \in \mathbb{N}$ mit $m_i \leq n_i$ für $i = 1, 2$.*

- i) *Es existieren genau $\binom{n_1}{m_1} \binom{n_2}{m_2}$ Teilmengen $A \subset \Omega$ mit $\text{Anz}(A \cap \Omega_i) = m_i$ für $i = 1, 2$.*
 ii) *Setzen wir $m = m_1 + m_2$, so ist bei Gleichverteilung die Wahrscheinlichkeit der Auswahl einer m -elementigen Teilmenge $A \subset \Omega$ mit $\text{Anz}(A \cap \Omega_i) = m_i$ für $i = 1, 2$ gleich*

$$h(m_1, m_2, n_1, n_2) = \frac{\binom{n_1}{m_1} \binom{n_2}{m_2}}{\binom{n}{m}}.$$

Der hypergeometrischen Verteilung entspricht das **Urnenmodell ohne Zurücklegen**, also einer Urne mit n_1 roten und n_2 schwarzen Kugeln. Nacheinander werden $m \leq n = n_1 + n_2$ Kugeln blind entnommen. Sind dann $m_1, m_2 \in \mathbb{N}_0$ mit $m = m_1 + m_2$, $m_1 \leq n_1$, und $m_2 \leq n_2$, so ist $h(m_1, m_2, n_1, n_2)$ die Wahrscheinlichkeit, dass sich in der entnommenen Probe m_1 rote und m_2 schwarze Kugeln befinden.

1.2.6 Beispiel: Die Wahrscheinlichkeit, dass ein bestimmter Spieler bei einem Kartenspiel (Skat) mit 32 Spielkarten, davon 4 Buben, genau 3 Buben hat ist $p = \frac{66}{899} = 0,0734$:

Diese Aussage folgt mit $n_1 = 4$, $n_2 = 28$, $m_1 = 3$, $m_2 = 7$ unmittelbar aus 1.2.5 ii). □

Wir wollen jetzt eine der wichtigsten Verteilungen - die **Binomialverteilung** - einführen. Dazu betrachten wir zunächst ein **Bernoulli-Experiment**: Einen Wahrscheinlichkeitsraum $\Omega_0 = \{a, b\}$ mit zwei möglichen Ausgängen. Es sei $p = P(\{a\})$ die Wahrscheinlichkeit des Eintreffens von a , diese wird auch als Wahrscheinlichkeit des Erfolges und $q = 1 - p = P_0(\{b\})$ die Wahrscheinlichkeit des Misserfolges bezeichnet. Oft wird auch die Bezeichnung $\Omega_0 = \{1, 0\}$ gewählt, und dann das Eintreffen von dem Elementarereignis 1 als Erfolg und das von 0 als Misserfolg bezeichnet.

Dieses Bernoulli-Experiment werde jetzt n -mal unabhängig wiederholt. Mathematisch gesehen liegt hier eine Produktsituation vor: Wir betrachten also

$$\Omega = \Omega_0 \times \dots \times \Omega_0 = \Omega_0^n \text{ mit dem Produktmaß } P = P_0 \otimes \dots \otimes P_0 .$$

Ist jetzt $c = (c_1, \dots, c_n) \in \Omega$ mit $\text{Anz}(\{i \mid c_i = a\}) = j$, so erhalten wir als Wahrscheinlichkeit des Eintreffens von c wegen der Konstruktion des Produktmaßes

$$P(\{c\}) = P_0(\{c_1\}) \cdot \dots \cdot P_0(\{c_n\}) = p^j q^{n-j} .$$

Bei fixiertem $j \in \{0, 1, \dots, n\}$ existieren wegen Satz 1.1.5 genau $\binom{n}{j}$ unterschiedliche Teilmengen $A \subset \{1, \dots, n\}$ mit $\text{Anz}(A) = j$, also äquivalent dazu $\binom{n}{j}$ unterschiedliche $c \in \Omega = \Omega_0^n$ mit $\text{Anz}(\{i \mid c_i = a\}) = j$. Ist demnach $A \subset \mathcal{P}(\Omega)$ das Ereignis des j -maligen Erfolges, also

$$A = \{c \in \Omega \mid \text{Anz}(\{i \mid c_i = a\}) = j\} ,$$

so folgt

$$P(A) = \binom{n}{j} p^j q^{n-j} .$$

Daher gilt die folgende Aussage:

1.2.7 Satz: Gegeben sei ein Bernoulli Experiment $\Omega_0 = \{a, b\}$ mit $p = P(\{a\})$ und $q = 1 - p = P(\{b\})$. Für $n \in \mathbb{N}$ und $0 \leq k \leq n$ ist

$$b_{n,p}(k) = \binom{n}{k} p^k q^{n-k} \text{ - die Binomialverteilung -}$$

die Wahrscheinlichkeit des Eintreffens des Ereignisses

$$A = \{c \in \Omega^n \mid \text{Anz}(\{i \mid c_i = a\}) = k\}$$

- des Ereignisses aller k -maligen Erfolge unter n Wiederholungen.

Anmerkung: Es gilt $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^n b_{n,p}(k) = (p+q)^n = 1$ als unmittelbare Konsequenz des "Binomischen Lehrsatzes".

Die Binomialverteilung enthält das **Urnenmodell mit Zurücklegen**, also einer Urne mit n_1 roten und n_2 schwarzen Kugeln. Es sei $n = n_1 + n_2$. Nacheinander werden m Kugeln blind entnommen und wieder in die Urne zurückgelegt. Sind dann $m_1, m_2 \in \mathbb{N}_0$ mit $m = m_1 + m_2$, $m_1 \leq n_1$, und $m_2 \leq n_2$, so ist mit $p = \frac{n_1}{n}$ und $q = \frac{n_2}{n}$ der Wert $b_{n,p}(m_1) = \binom{m}{m_1} p^{m_1} q^{m_2}$ die Wahrscheinlichkeit, dass m_1 rote und m_2 schwarze Kugeln entnommen wurden.

1.2.8 Bemerkung: Es seien $p_n \in [0, 1]$ für $n \in \mathbb{N}$ mit $\sum_{n=1}^{\infty} p_n = 1$. Dann ist

$$P : P(A) = \sum_{n \in A} p_n \text{ für } A \subset \mathcal{P}(\mathbb{N})$$

ein Wahrscheinlichkeitsmaß auf \mathbb{N} .

Der einfache Beweis dieser Aussage bleibt dem Leser überlassen.

Anmerkung: Ist Ω eine abzählbare unendliche Menge, so kann man die Existenz einer Funktion $Q : \mathcal{P}(\Omega) \rightarrow [0, 1]$ nachweisen mit den Eigenschaften

$$Q(\Omega) = 1, \quad Q(A) = 0 \quad \text{für alle } A \in \mathcal{P}(\Omega) \text{ mit } \text{Anz}(A) < \infty \\ \text{und mit } Q(A \cup B) = Q(A) + Q(B) \text{ für alle } A, B \in \mathcal{P}(\Omega) \text{ mit } A \cap B = \emptyset.$$

Derartige Funktionen lassen keine Summendarstellung wie oben zu, sie sind auch nicht in den hier betrachteten Zusammenhängen von Interesse.

Es soll noch kurz die Schreibweise $P(A) = \sum_{\omega \in A} p_\omega$ diskutiert werden für den Fall, dass Ω eine abzählbare Menge ist.

Es gilt $P(\emptyset) = 0$ – die Summation wird über die leere Indexmenge erstreckt.

Im Fall $\text{Anz}(A) < \infty$ erfolgt die Summation in beliebiger Reihenfolge.

Im Fall $\text{Anz}(A) = \infty$ setzen wir

$$P(A) = \sup \{ P(B) \mid B \subset A \text{ mit } \text{Anz}(B) \in \mathbb{N} \}.$$

1.2.9 Bemerkung: (Geometrische Verteilung) Gegeben sei das Bernoulli Experiment $\Omega_0 = \{a, b\}$, dieses werde unabhängig wiederholt bis zum ersten Mal „a“ auftritt. Die Wahrscheinlichkeit, dass bei der k -ten Wiederholung zum ersten Mal „a“ auftritt ist $p \cdot q^{k-1}$. Es gilt

$$\sum_{n=1}^{\infty} p \cdot q^{n-1} = p \sum_{j=0}^{\infty} q^j = \frac{p}{1-q} = 1.$$

Damit ist $P : P(\{k\}) = p \cdot q^{k-1}$ ein Wahrscheinlichkeitsmaß auf \mathbb{N} .

1.2.10 Bemerkung: (Negative Binomialverteilung, Pascal Verteilung): Es seien $r, k \in \mathbb{N}$ und $\Omega_0 = \{a, b\}$ ein Bernoulli-Experiment mit der Erfolgswahrscheinlichkeit p . $f(k, r, p)$ sei die Wahrscheinlichkeit, dass bei $n = r + k$ Wiederholungen des Experimentes dem r -ten Erfolg genau k Misserfolge vorangehen. Es gilt

$$f(k, r, p) = \binom{k+r-1}{k} p^r q^k = \binom{k+r-1}{r-1} p^r q^k = \binom{-r}{k} p^r (p-1)^k$$

mit $\binom{-r}{k} = \frac{(-r)(-r-1)\dots(-r-k+1)}{k!}$ als verallgemeinertem Binomialkoeffizient.

Diese Aussage folgt unmittelbar aus den vorstehenden Konstruktionen. Ebenso einfach folgt die nächste Aussage:

1.2.11 Bemerkung: (Poisson Verteilung) Es seien $\lambda \in \mathbb{R}$ und $\Omega = \mathbb{N}_0$, wir setzen $p(k|\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$, dann gilt $\sum_{k=0}^{\infty} p(k|\lambda) = 1$.

Die Poisson-Verteilung wird auch die Verteilung seltener Ereignisse genannt. Sie ersetzt im Fall einer kleinen Wahrscheinlichkeit p die Binomial-Verteilung; es gilt die folgende Approximationsaussage:

1.2.12 Bemerkung: Es seien $p \in [0, 1]$ und $k, n \in \mathbb{N}_0$, so dass k und $\lambda = p \cdot n$ „klein gegenüber“ n ist. Dann ist $p^k (1-p)^{n-k} \binom{n}{k}$ angenähert gleich $\frac{\lambda^k}{k!} e^{-\lambda}$.

Beweis: Es gilt

$$\begin{aligned} p^k(1-p)^{n-k} \binom{n}{k} &= p^k n^k \frac{1}{n^k} \frac{n(n-1)\dots(n-k+1)}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Für große n und kleine k, λ ist $\frac{n(n-1)\dots(n-k+1)}{n^k}$ und $\left(1 - \frac{\lambda}{n}\right)^{-k}$ angenähert gleich 1, während $\left(1 - \frac{\lambda}{n}\right)^n$ angenähert gleich $e^{-\lambda}$ ist. - Man beachte $\left(1 - \frac{a}{n}\right)^n \rightarrow e^{-a}$ für alle $a \in \mathbb{R}$ bei $n \rightarrow \infty$. \square

In der folgenden Aussage zeigen wir, dass die Werte einer hypergeometrischen Verteilung durch die entsprechenden Werte der Binomialverteilung bei Vorliegen großer Zahlen approximiert werden kann. Diese Tatsache ist besonders bei Praxisproblemen von Bedeutung und auch unmittelbar plausibel. Man beachte weiter, dass die Werte der Binomialverteilung in der Regel wesentlich besser zu berechnen sind als die der hypergeometrischen Verteilung. Bei größeren Werten des Umfangs n der Zahl der Wiederholungen eines Bernoulli-Experimentes oder des Umfangs einer Stichprobe ist das auch nicht mehr ausreichend. Man beachte dabei, dass wegen der Größe der Binomialkoeffizienten eine konkrete Berechnung nur sehr eingeschränkt möglich ist: Mit Hilfe des Binomischen Lehrsatzes kann man einfach $\binom{1000}{500} > 10^{297}$ zeigen.

1.2.13 Bemerkung: Es seien $p_1, p_2 \in [0, 1]$ mit $p_1 + p_2 = 1$. Weiter seien n_1, n_2, n mit $n_1 + n_2 = n$, und es gelte $\frac{n_j}{n} \rightarrow p_j$ bei $n \rightarrow \infty$ für $j = 1, 2$. Für fixierte $m, m_1, m_2 \in \mathbb{N}_0$ mit $m = m_1 + m_2$ gilt

$$h(m_1, m_2, n_1, n_2) \rightarrow \binom{m}{m_1} p_1^{m_1} p_2^{m_2} = \binom{m}{m_1} p_1^{m_1} (1 - p_1)^{m - m_1}$$

bei $n \rightarrow \infty$.

Beweis: Es gilt für $m_j \leq n_j$ die Beziehung

$$\begin{aligned} h(m_1, m_2, n_1, n_2) &= \frac{\binom{n_1}{m_1} \binom{n_2}{m_2}}{\binom{n}{m}} \\ &= \frac{n_1(n_1-1)\dots(n_1-m_1+1)}{m_1!} \cdot \frac{n_2(n_2-1)\dots(n_2-m_2+1)}{m_2!} \cdot \frac{m!}{n \dots (n-m+1)} \\ &\rightarrow \binom{m}{m_1} p_1^{m_1} p_2^{m_2} \end{aligned}$$

bei $n \rightarrow \infty$; dabei beachte man $\frac{n_j - k}{n - l} \rightarrow p_j$ bei $n \rightarrow \infty$ für alle $k, l \leq m$. \square

Anmerkung: Die gleiche Aussage gilt auch bei Approximation der Multivarianten (hypergeometrisch, Multinomialkoeffizienten).

1.2.14 Satz: $\psi : \mathbb{R} \rightarrow [0, \infty[$ sei integrierbar mit

$$\int_{\mathbb{R}} \psi d\lambda = \int_{-\infty}^{\infty} \psi(t) dt = 1.$$

Dann ist $\mathcal{B} \ni A \mapsto P(A) = \int_A \psi d\lambda$ ein Wahrscheinlichkeitsmaß - das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ mit der Dichte ψ .

Anmerkung: Bei dieser vorliegenden Situation sind hauptsächlich Mengen der Form $[a, b]$ beziehungsweise $] - \infty, a]$ von Interesse. Es gilt dann etwa $P([a, b]) = \int_a^b \Psi(t) dt$.

Beweis: Es seien $A \in \mathcal{B}$ und $(A(n))_{n=1}^\infty \subset \mathcal{B}$ eine Folge paarweise unvereinbarer Ereignisse (paarweise disjunkter Mengen) mit $A = \bigcup_{n=1}^\infty A(n)$. Für jedes $B \in \mathcal{B}$ sei $\mathbb{1}_B$ die Indikatorfunktion: $\mathbb{1}_B(\omega) = 1$ für alle $\omega \in B$ und $\mathbb{1}_B(\omega) = 0$ für alle $\omega \notin B$. Für jedes $n \in \mathbb{N}$ sei $h_n = \mathbb{1}_{A(1)} + \dots + \mathbb{1}_{A(n)}$. Wegen der Konstruktion gilt $h_n(\omega) \uparrow \mathbb{1}_A(\omega)$, bei $n \rightarrow \infty$ für alle $\omega \in \mathbb{R}$. Aus dem Konvergenzatz von Beppo Levi (Monotone Konvergenz) folgt

$$P(A(1)) + \dots + P(A(n)) = \int h_n \psi d\lambda \uparrow \int \mathbb{1}_A \cdot \psi d\lambda = P(A)$$

bei $n \rightarrow \infty$. Demnach ist P σ -additiv auf $(\mathbb{R}, \mathcal{B})$. □

1.2.15 Satz: Für $t \in \mathbb{R}$ sei $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ - die Dichte der Standard-Normalverteilung $N(0, 1)$. Dann ist $\mathcal{B} \ni A \mapsto \int_A \varphi d\lambda$ ein Wahrscheinlichkeitsmaß mit der Verteilungsfunktion

$$\Phi_{0,1}(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \int_{-\infty}^t \varphi(s) ds.$$

Beweis: Es ist $\int_{\mathbb{R}} \varphi d\lambda = 1$ zu zeigen. Dazu berechnen wir

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right)^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right) \left(\int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} ds \right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{t^2+s^2}{2}} dt ds = \int_0^{\infty} \int_0^{2\pi} r e^{-\frac{r^2}{2}} dr dx \quad (\text{Polarkoordinaten}) \\ &= 2\pi \left(e^{-\frac{r^2}{2}} \right) \Big|_0^{\infty} = 2\pi, \quad \text{was zu zeigen war.} \end{aligned}$$

□

Die Dichte der Normalverteilung $\varphi : \mathbb{R} \rightarrow [0, \infty[$ besitzt keine geschlossene Stammfunktion. Daher liegen die wichtigsten Funktionswerte dieser Stammfunktion $\Phi_{0,1}$ in Tabellen vor; dabei beschränkt man sich auf Werte im Intervall $[0, 3]$: Für $t > 3$ ist $\Phi_{0,1}(t)$ angenähert gleich 1. Wegen $\varphi(t) = \varphi(-t)$ für alle $t \in \mathbb{R}$ gilt $\phi(0) = \frac{1}{2}$ und $\phi_{0,1}(-t) = 1 - \phi_{0,1}(t)$ für alle $t \in \mathbb{R}$. Andererseits kann man leicht wegen der sehr guten Konvergenz der Exponentialfunktion

$e^s = \sum_{k=0}^{\infty} \frac{s^k}{k!}$ durch Auswertung der ersten Summanden die Werte für kleine $t > 0$ mit einem Rechner auswerten: Es gilt

$$\Phi_{0,1}(t) = \frac{1}{2} + \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{t^{(2k+1)}}{(2k+1)}.$$

Speziell für $t \leq 1$ müssen nur sehr wenige Summanden ausgewertet werden.

1.2.16 Satz: *Es sei $\Omega \subset \mathcal{B}^n \subset \mathcal{P}(\mathbb{R})$ mit $0 < \lambda^n(\Omega) < \infty$. Für $A \subset \Omega$ mit $A \in \mathcal{B}^n$ sei $m(A) = \frac{\lambda^n(A)}{\lambda^n(\Omega)}$. m ist ein W -Maß auf (Ω, \mathcal{B}^n) - die **Gleichverteilung**.*

Beweis: Die Aussage folgt aus der Darstellung

$$m(A) = \int \mathbb{1}_A d\lambda^n \cdot \frac{1}{\lambda^n(\Omega)} \text{ und 1.2.13.}$$

□

Im weiteren Verlauf diskutieren wir noch die Gamma-Verteilung: Es sei dazu $\alpha > 0$ eine Konstante. Zur Motivation betrachten wir die folgende Situation: Für $t > 0$ mit $\alpha t < 1$ sei αt die Wahrscheinlichkeit, dass eine gewisse vorgegebene aber unbekannte Zahl im Intervall $]0, t]$ liegt. Für großes $n \in \mathbb{N}$ zerlegen wir das Intervall $]0, t]$ in n Teilintervalle der Länge $\frac{t}{n}$; dann ist $\frac{\alpha t}{n}$ die Wahrscheinlichkeit, dass diese vorgegebene Zahl in diesem Teilintervall liegt.

Sind jetzt $k \in \mathbb{N}$ und derartige der Größe nach geordnete Zahlen (mit $k \ll n$) vorgegeben, so dürfen wir annehmen, dass jedes dieser kleinen Teilintervalle maximal eine der vorgegebenen Zahlen enthält. Setzen wir also $p = \frac{\alpha t}{n}$, so ist wegen der Definition der Binomialverteilung $b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$ die Wahrscheinlichkeit, dass genau k dieser Zahlen in dem Intervall $]0, t]$ liegen. Es gilt wegen 1.2.12 $b_{n,p}(k) \rightarrow p(k|\lambda)$ bei $n \rightarrow \infty$ mit $\lambda = \alpha t$. Die Wahrscheinlichkeit, dass mindestens r Punkte in $]0, t[$ liegen, ist demnach

$$P(r) = P(]0, t[) = 1 - e^{-\lambda} \sum_{k=0}^{r-1} \frac{\lambda^k}{k!} = 1 - e^{-\alpha t} \sum_{k=0}^{r-1} \frac{(\alpha t)^k}{k!} = g(t) = \int_0^t g'(s) ds$$

Wegen

$$g'(t) = \alpha e^{-\alpha t} \sum_{k=0}^{r-1} \frac{\alpha^k t^k}{k!} - e^{-\alpha t} \sum_{k=0}^{r-1} \frac{\alpha^k t^{k-1}}{(k-1)!} = \frac{\alpha^r e^{-\alpha t} t^{r-1}}{(r-1)!} = \gamma_{\alpha,r}(t)$$

gilt die Gamma-Verteilung: die Verteilung mit der Dichte $\gamma_{\alpha,r}$. Zur Definition in der allgemeinen Situation $r \in]0, \infty[$ setzen wir

$$\gamma_{\alpha,r}(t) = \frac{\alpha^r e^{-\alpha t} t^{r-1}}{\Gamma(r)}$$

mit der Gamma-Funktion $\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt$.

Mit Hilfe partieller Integration folgt unmittelbar $\Gamma(r+1) = r \cdot \Gamma(r)$ für alle $r > 0$ und damit folgt wegen $\Gamma(1) = 1$ sofort $\Gamma(n+1) = n!$ für alle $n \in \mathbb{N}$.

1.2.17 Definition: Das Wahrscheinlichkeitsmaß $\Gamma_{\alpha,r}$ auf $(]0, \infty[, \mathcal{B}^1)$ mit der Dichte $\gamma_{\alpha,r}$ für $\alpha, r > 0$ heißt die **Gamma-Verteilung** und im Fall $r = 1$ die **Exponential-Verteilung**.

Die Exponentialverteilung besitzt demnach die Dichte $\alpha e^{-\alpha t}$ für $t > 0$. Eine weitere bekannte Verteilung ist die Cauchy-Verteilung:

1.2.18 Bemerkung: Für $t \in \mathbb{R}$ und $0 < a \in \mathbb{R}$ sei

$$c_a(t) = \frac{a}{\pi(a^2 + t^2)}$$

die Dichte der **Cauchy-Verteilung** auf \mathbb{R} . Es gilt $\int_{\mathbb{R}} c_a(t) dt = 1$;

Diese Aussage kann sehr einfach nachgerechnet werden.

Im weiteren Verlauf sollen bedingte Wahrscheinlichkeiten diskutiert werden:

1.2.19 Definition: Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Es seien $A, B \in \mathcal{A}$ mit $P(B) > 0$. Wir definieren die **bedingte Wahrscheinlichkeit** von A unter der Hypothese B durch

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

1.2.20 Beispiel: Zwei gleiche Schränke mit jeweils 4 Schubladen seien gegeben. Im Schrank I befinden sich in drei der Schubladen jeweils eine Kugel; im Schrank II befindet sich nur in einer der Schubladen eine Kugel.

Es werde jetzt willkürlich eine Schublade eines der Schränke geöffnet, in dieser befand sich eine Kugel. Wie groß ist die Wahrscheinlichkeit, dass sich diese Schublade im Schrank I befand?

B sei das Ereignis, dass eine Schublade mit einer Kugel geöffnet wurde.

A sei das Ereignis, dass eine Schublade des Schrankes I geöffnet wurde.

Es gilt, da ein Laplace'scher Wahrscheinlichkeitsraum vorausgesetzt wurde

$$P(B) = \frac{1}{2}, \quad P(A \cap B) = \frac{3}{8} \quad \text{und damit} \quad P(A|B) = \frac{3}{4}.$$

□

1.2.21 Satz: i) Es sei $B \in \mathcal{A}$ mit $P(B) > 0$.

Dann ist $\mathcal{A} \ni A \mapsto P(A|B)$ ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) .

ii) Es seien $B_1, \dots, B_r \in \mathcal{A}$ paarweise unvereinbar (disjunkt) mit $\Omega = B_1 \cup \dots \cup B_r$ (eine disjunkte Zerlegung von Ω). Dann gilt die Formel von der **totalen Wahrscheinlichkeit**

$$P(A) = \sum_{k=1}^r P(A|B_k)P(B_k)$$

für alle $A \in \mathcal{A}$; dabei setzen wir $P(A|B_k) = 0$ im Fall $P(B_k) = 0$.

iii) Im Fall von $A \in \mathcal{A}$ mit $P(A) > 0$ und den Voraussetzungen aus ii) gilt die **Formel von Bayes**:

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{k=1}^r P(A|B_k)P(B_k)}.$$

Beweis: i) Es gilt $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$ und $P(\emptyset|B) = 0$.

Sind $A_1, A_2 \in \mathcal{A}$ disjunkt, so sind auch $(A_1 \cap B), (A_2 \cap B)$ disjunkt mit $P((A_1 \cap B) \cup (A_2 \cap B)) = P(A_1 \cap B) + P(A_2 \cap B)$. Es folgt unmittelbar $P(A_1|B) + P(A_2|B) = P(A_1 \cup A_2|B)$ und damit Aussage (i).

ii) Es gilt

$$A = A \cap \Omega = A \cap (B_1 \cup \dots \cup B_r) = (A \cap B_1) \cup \dots \cup (A \cap B_r)$$

für alle $A \in \mathcal{A}$. Wir erhalten also wegen $(A \cap B_i) \cap (A \cap B_h) = \emptyset$ für alle $i \neq h$ unmittelbar

$$P(A) = P(A \cap B_1) + \dots + P(A \cap B_r) = P(A|B_1) \cdot P(B_1) + \dots + P(A|B_r)P(B_r),$$

also die Formel von der totalen Wahrscheinlichkeit.

iii) Im Fall $P(A) > 0$ gilt für jedes $j \in \{1, \dots, r\}$ offenbar

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{h=1}^r P(A|B_h)P(B_h)}$$

und damit die Formel von Bayes. □

1.2.22 Folgerung: $B_1, \dots, B_r \in \mathcal{A}$ seien paarweise unvereinbar mit $P(B_1), \dots, P(B_r) > 0$, weiter sei $A \in \mathcal{A}$ mit $P(A|B_1) = \dots = P(A|B_r)$. Dann gilt $P(A|(B_1 \cup \dots \cup B_r)) = P(A|B_1)$.

Beweis: Setzen wir $B = B_1 \cup \dots \cup B_r$, so gilt

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B_1) + \dots + P(A \cap B_r)}{P(B_1) + \dots + P(B_r)} \\ &= \frac{P(A|B_1)P(B_1) + \dots + P(A|B_r)P(B_r)}{P(B_1) + \dots + P(B_r)} = P(A|B_1). \end{aligned}$$

□

1.2.23 Beispiel: Beim Skatspiel besitzt der erste Spieler 2 Buben. Wie groß ist dann die Wahrscheinlichkeit, dass die beiden anderen Spieler jeweils einen Buben haben?

A_1 sei das Ergebnis, dass der erste Spieler 2 Buben hat.

A_2 sei das Ergebnis, dass der zweite Spieler genau einen Buben hat.

A_3 sei das Ergebnis, dass der dritte Spieler genau einen Buben hat.

Es gilt

$$P(A_1) = \frac{\binom{4}{2} \binom{28}{8}}{\binom{32}{10}}, \quad P(A_2|A_1) = \frac{\binom{2}{1} \binom{20}{9}}{\binom{22}{10}}, \quad P(A_3|A_1 \cap A_2) = \frac{\binom{1}{1} \binom{11}{9}}{\binom{12}{10}}.$$

Damit erhalten wir die Wahrscheinlichkeit

$$P(A_2 \cap A_3|A_1) = P(A_3|A_1 \cap A_2)P(A_2|A_1) = \frac{2(20)!(11)!(10)!(12)!(10)!2!}{9!(11)!9!2!(22)!(12)!} = \frac{100}{11 \cdot 21} = 0,4329.$$

1.3 Zufallsvariable

In diesem Abschnitt sei stets (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum; wir führen im weiteren Verlauf den Begriff der Zufallsvariablen ein. Im mathematischen Sinn handelt es sich dabei um eine messbare Abbildung in einen anderen Maßraum. Besonders wichtig sind dabei die (reellen) Zufallsvariablen als messbare Abbildungen in den Raum $(\mathbb{R}, \mathcal{B})$ mit der üblichen Borelalgebra $\mathcal{B} \subset \mathcal{P}(\mathbb{R})$. Unter einer **reellen Zufallsvariablen** verstehen wir also eine messbare Funktion

$$X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}).$$

Wir erinnern, dass die Messbarkeit in diesem Fall definiert wird durch die Forderung $X^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}$.

Bekanntlich ist $X : \Omega \rightarrow \mathbb{R}$ genau dann messbar, wenn $[X \leq \alpha] \in \mathcal{A}$ gilt für alle $\alpha \in \mathbb{R}$. Für einen Beweis vergleiche man ein entsprechendes Resultat der Maßtheorie.

Entsprechend heißt auch eine messbare Funktion $X : \Omega \rightarrow \mathbb{R}^q$ (im Sinn $X^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}^q$) eine vektorwertige Zufallsvariable beziehungsweise ein **Zufallsvektor**.

Es sei weiter angemerkt, dass im Fall einer endlichen Menge Ω wegen der getroffenen Vereinbarung $\mathcal{A} = \mathcal{P}(\Omega)$, jede Abbildung $X : \Omega \rightarrow \mathbb{R}$ messbar ist.

Im allgemeinen Fall einer Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$ führen wir jetzt den Erwartungswert ein: Jede Zufallsvariable ist eine messbare Funktion; für messbare Funktionen $X : (\Omega, \mathcal{A})$ ist der Begriff der Integrierbarkeit definiert. Wir benutzen hier dafür (aus historischen Gründen) an Stelle des Integrals den Begriff **Erwartungswert**:

$$E(X) = \int_{\Omega} X(\omega) dP(\omega) \text{ im Fall } \int_{\Omega} |X(\omega)| dP(\omega) < \infty ,$$

und schreiben wir üblich $X \in \mathcal{L}^1(P)$ im Fall $\int_{\Omega} |X(\omega)| dP(\omega) < \infty$ beziehungsweise $X \in \mathcal{L}^2(P)$ im Fall $X^2 \in \mathcal{L}^1(P)$.

Wir diskutieren zunächst den Fall einer Zufallsvariablen $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{A})$, die nur endlich viele Werte annimmt: etwa die Werte $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. Wir setzen

$$A_j = [X = \alpha_j] = \{\omega \in \Omega \mid X(\omega) = \alpha_j\} .$$

Dann gilt $X = \sum_{j=1}^n \alpha_j \mathbb{1}_{A(j)}$ mit der Indikatorfunktion der Menge $A(j)$:

$$\mathbb{1}_{A(j)}(\omega) = 1 \text{ für } \omega \in A(j) \text{ und } \mathbb{1}_{A(j)}(\omega) = 0 \text{ für } \omega \notin A(j) .$$

X heißt daher auch eine Stufenfunktion. Es gilt daher die folgende Darstellung des Erwartungswertes in diesem Fall

$$E(X) = \sum_{j=1}^n \alpha_j P(A_j) .$$

Der Erwartungswert ist demnach ein gewichtetes Mittel der Werte der Zufallsvariablen wegen $P(A_j) \geq 0$ und $P(A_1) + \dots + P(A_n) = 1$.

Zufallsvariablen besitzen eine sehr große Bedeutung in der Wahrscheinlichkeitstheorie und

treten sehr oft unmittelbar in Zusammenhang mit elementarerer Problemstellungen auf. Besteht beispielsweise das wahrscheinlichkeitstheoretische Experiment in einer n -maligen Wiederholung eines Bernoulli-Experimentes, so kann X die Erfolge zählen:

$$X(\omega_1, \dots, \omega_n) = \text{Anz}(\{j \mid \omega_j = a\}) .$$

Wird dieses Bernoulli-Experiment beliebig oft wiederholt, so ist eine weitere Zufallsvariable Y definiert durch die Zahl der Versuche bis zum ersten Erfolg:

$$Y(\omega_1, \dots, \omega_n, \dots) = \min \{n \mid \omega_n = a, \omega_1 = \dots = \omega_{n-1} = b\} .$$

In der allgemeinen Situation besitzt der Erwartungswert die gleichen Eigenschaften wie das Integral auf allgemeinen Maßräumen im Sinn von Lebesgue, die hier nur kurz formuliert werden sollen. Ein Beweis findet sich in der Maßtheorie.

1.3.1 Bemerkung: $X : \Omega \rightarrow \mathbb{R}$ sei eine Funktion. X ist genau dann messbar, wenn eine Folge $(X_n)_{n=1}^\infty$ von Stufenfunktionen existiert mit $X_n(\omega) \rightarrow X(\omega)$ und $|X_n(\omega)| \uparrow |X(\omega)|$ bei $n \rightarrow \infty$ für jedes $\omega \in \Omega$.

1.3.2 Satz: Es seien $X, Y \in \mathcal{L}^1(P)$ und $\alpha \in \mathbb{R}$, dann gilt

- i) $X + \alpha Y \in \mathcal{L}^1(P)$ mit $E(X + \alpha Y) = E(X) + \alpha E(Y)$
- ii) Im Fall $|X| \leq |Y|$ gilt $|E(X)| \leq E(|Y|)$.
- iii) Im Fall $X \geq 0$ folgt $E(X) \geq 0$.

Wir bestimmen im weiteren Verlauf den Erwartungswert in wichtigen Spezialfällen.

1.3.3 Beispiel: (Erwartungswert der Binomialverteilung) Es sei $p \in]0, 1[$. Das Bernoulli-Experiment $\Omega_0 = \{a, b\}$ werde n -mal wiederholt mit $p = P(\{a\})$ als Erfolgswahrscheinlichkeit; wir erhalten also $\Omega = \Omega_0^n$ als Ereignisraum mit $b_{n,p}(k) = \binom{n}{k} p^k q^{n-k}$ gemäß 1.2.7 als die Wahrscheinlichkeit des Eintreffens von k Erfolgen. X sei die Zufallsvariable, die die Zahl der Erfolge zählt: $X(\omega_1, \dots, \omega_n) = k$ für alle $0 \leq k \leq n$. Es gilt demnach

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \\ &= \sum_{j=0}^{n-1} p \cdot n \frac{(n-1)!}{j!(n-1-j)!} p^j q^{n-1-j} = pn(p+q)^{n-1} = p \cdot n . \end{aligned}$$

Wir notieren an dieser Stelle eine weitere und dabei einfachere Methode zu Bestimmung des Erwartungswertes der Binomialverteilung: Für alle $j = 1, \dots, n$ sei X_j die Zufallsvariable auf Ω definiert durch

$$X_j(\omega_1, \dots, \omega_n) = \begin{cases} 1 & \text{im Fall } \omega_j = a \\ 0 & \text{im Fall } \omega_j = b \end{cases} .$$

Offenbar gilt $E(X_j) = p$ für alle $j = 1, \dots, n$ und $X = X_1 + \dots + X_n$, und damit folgt $E(X) = E(X_1) + \dots + E(X_n) = np$. □

1.3.4 Beispiel: (Erwartungswert der geometrischen Verteilung) Es gelte die Situation 1.2.9 mit $\Omega_0 = \{a, b\}$, und mit der Erfolgswahrscheinlichkeit $p = P(\{a\})$ und $q = 1 - p$. $X(k) = k$ sei die Zufallsvariable, die die Dauer bis zum ersten Erfolge angibt. Es gilt

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k p q^{k-1} = p \sum_{k=1}^{\infty} k q^{k-1} \\ &= p \sum_{k=0}^{\infty} \frac{d}{dq} (q^k) = p \cdot \frac{d}{dq} \left(\sum_{k=0}^{\infty} q^k \right) = p \frac{d}{dq} \left(\frac{1}{1-q} \right) = p \frac{1}{(1-q)^2} = \frac{1}{p}. \end{aligned}$$

□

1.3.5 Beispiel: (Erwartungswert der Poissonverteilung) In 1.2.11 wurde die Poissonverteilung eingeführt. X sei die Zufallsvariable, die die Zahl der Erfolge zählt, es gilt also $X(k) = k$ für alle $k \in \mathbb{N}_0$. Damit folgt

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

□

Im weiteren Verlauf sollen die Verteilungsfunktionen und die

1.3.6 Definition: $X : \Omega \rightarrow \mathbb{R}^q$ sei ein Zufallsvektor. Für $t \in \mathbb{R}^q$ sei

$$F_X(t) = P([X \leq t]) \text{ – die (gemeinsame) Verteilung,}$$

wobei die Ordnung „ \leq “ punktweise zu verstehen ist: $X(\omega) \leq t$ ist äquivalent zu $X_j(\omega) \leq t_j$ für alle $j = 1, \dots, q$.

Im Fall $q = 1$ sprechen wir von der **Verteilungsfunktion** $F_X = F$. Ist diese Funktion F differenzierbar mit der Ableitung $F' = f$, so heißt f die **Dichte** der Verteilungsfunktion oder der Verteilungsdichte.

1.3.7 Beispiel: In der Situation von 1.3.3 (Binomialverteilung) gilt $F_X(t) = 0$ für $t < 0$, $F_X(t) = 1$ für $t \geq n$, und $F_X(t) = \sum_{j=0}^k \binom{n}{j} p^j q^{n-j}$ für $k \leq t < k+1$ und $k \in \mathbb{N}_0$ mit $k \leq n$.

1.3.8 Satz: $X : \Omega \rightarrow \mathbb{R}$ sei eine reelle Zufallsvariable. Dann besitzt die Verteilungsfunktion $F = F_X : \mathbb{R} \rightarrow \mathbb{R}$ folgende Eigenschaften.

- i) F ist monoton wachsend.
- ii) F ist rechtsseitig stetig.
- iii) $F(t) \rightarrow 1$ bei $t \rightarrow \infty$, $F(t) \rightarrow 0$ bei $t \rightarrow -\infty$.

Beweis: i) folgt unmittelbar aus der Definition.

ii) Es sei $\tau \in \mathbb{R}$ und $(t_n)_{n=1}^{\infty}$ eine Folge mit $t_n \downarrow \tau$ bei $n \rightarrow \infty$. Wir setzen

$$A = [X \leq \tau] \text{ und } A_n = [X \leq t_n].$$

Dann gilt $A_n \supset A_{n+1}$ und $A = \bigcap_{n=1}^{\infty} A_n$. Der Stetigkeitssatz für Maße impliziert $P(A_n) = F(t_n) \rightarrow P(A) = F(\tau)$ bei $n \rightarrow \infty$, was zu zeigen war.

iii) folgt entsprechend.

□

Für viele Situationen ist eine andere Betrachtung der Verteilungsfunktion erforderlich. Wir fixieren eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$. Da X messbar ist, gilt $X^{-1}(A) \in \mathcal{A}$ für alle $A \in \mathcal{B}^1$. Wir definieren also das Bildmaß oder die Verteilung von X durch $P_X(A) = P(X^{-1}(A))$ für alle $A \in \mathcal{B}^1$. Ist jetzt die Verteilungsfunktion $F = F_X$ differenzierbar mit der Ableitung f , so gilt für alle $\tau, t \in \mathbb{R}$ mit $\tau < t$

$$F(t) - F(\tau) = \int_{\tau}^t f(s)ds = P(X \leq t) - P(X \leq \tau) = P(\tau < X \leq t) = P_X(] \tau, t]) .$$

f heißt die Dichte der Verteilungsfunktion. Da die Intervalle der Form $] \tau, t]$ ein Erzeugendensystem von \mathcal{B}^1 bilden, folgt aus dem Eindeutigkeitsatz für Maße unmittelbar

$$P_X(A) = \int_A f(t)dt \text{ für alle } A \in \mathcal{B}^1$$

1.3.9 Lemma: $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ sei eine Zufallsvariable $h : \mathbb{R} \rightarrow [0, \infty[$ messbar. Dann gilt

$$E(h(X)) = E(h \circ X) = \int h \circ X dP = \int h(t)dP_X(t) \in [0, \infty]$$

Besitzt die Verteilungsfunktion F von X zusätzlich eine Dichte f so gilt

$$E(h \circ X) = \int_{-\infty}^{\infty} h(t)f(t)dt.$$

Beweis: Es sei $A \in \mathcal{B}$ und $h = \mathbb{1}_A$. Dann gilt

$$E(h(X)) = P(X^{-1}(A)) = \int h(t)dP_X(t).$$

Damit gilt die behauptete Gleichheit und der Zusatz für Funktionen der Form $h = \mathbb{1}_A$, also auch für Funktionen $h = \alpha_1 \mathbb{1}_{A_1} + \dots + \alpha_r \mathbb{1}_{A_r}$. Da jedes messbare $h \geq 0$ durch eine Folge von Treppenfunktionen h_n approximierbar ist, folgt die Behauptung aus dem Satz von Beppo Levi. \square

1.3.10 Satz: $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ sei eine Zufallsvariable und $h : \mathbb{R} \rightarrow \mathbb{R}$ sei messbar. Dann gilt:

$h \circ X \in \mathcal{L}^1(P)$ ist äquivalent zu $h \in \mathcal{L}^1(P_X)$. In diesem Fall haben wir

$$E(h(X)) = \int_{\mathbb{R}} h(t)dP_X(t) .$$

Zusatz: Besitzt F eine Dichte f , so gilt

$$E(h(X)) = \int_{-\infty}^{\infty} h(t)f(t)dt .$$

(Speziell gilt $E(X) = \int_{-\infty}^{\infty} tf(t)dt$, wenn eine der Seiten definiert ist.)

1.3.11 Beispiel: Es sei Ω abzählbar mit dem Wahrscheinlichkeitsmaß

$$P(A_0) = \sum_{\omega \in A_0} p_\omega \text{ für } A_0 \in \Omega$$

$X : \Omega \rightarrow \mathbb{R}$ sei eine Zufallsvariable. Für eine Menge $A \in \mathcal{B}$ gilt dann

$$P_X(A) = P(X^{-1}(A)) = \sum_{\omega: X(\omega) \in A} p_\omega.$$

Ist jetzt $h : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion, mit $E(|h| \circ X) < \infty$, so folgt

$$E(h \circ X) = \sum_{\omega \in \Omega} h(X(\omega))p_\omega = \sum_{t \in X(\Omega)} h(t)P_X(\{t\})$$

1.3.12 Beispiel: Erwartungswert der Gamma-Verteilung Es seien $\alpha, r > 0$ und X sei eine Zufallsvariable, die gamma-verteilt ist bezüglich der Parameter α, r , es gilt also

$$F_X(t) = P(X \leq t) = \int_0^t \gamma_{\alpha, r}(s) ds$$

für alle $t > 0$ und $F_X(t) = 0$ für alle $t \leq 0$. Dann gilt

$$E(X) = \int_0^\infty t \gamma_{\alpha, r}(t) dt = \int_0^\infty t \frac{\alpha^r}{\Gamma(r)} e^{-\alpha t} t^{r-1} dt = \frac{r}{\alpha} \int_0^\infty \frac{\alpha^{r+1}}{\Gamma(r+1)} e^{-\alpha t} t^r dt = \frac{r}{\alpha}$$

- man beachte 1.2.17. Beim Spezialfall $r = 1$ - also bei der Exponentialverteilung - erhalten wir demnach als Erwartungswert $E(X) = \frac{1}{\alpha}$.

Es sei darauf hingewiesen, dass nicht jede Verteilungsfunktion einen endlichen Erwartungswert besitzen muss: dazu betrachte man etwa als Gegenbeispiel die Cauchy-Verteilung, man vergleiche 1.2.18

1.3.13 Bemerkung: Es sei $a > 0 \in \mathbb{R}$ und $c_a(t) = \frac{a}{\pi(a^2 + t^2)}$ für alle $t \in \mathbb{R}$. Dann ist die Funktion $\mathbb{R} \ni f(t) = t \mapsto t c_a(t)$ nicht integrierbar wegen $\int_{\mathbb{R}} |t| c_a(t) dt = \infty$.

Eine weitere wichtige Größe für reellwertige Zufallsvariable ist der Median der wie folgt definiert wird:

1.3.14 Definition: $X(\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ sei eine Zufallsvariable. $m \in \mathbb{R}$ heißt ein **Median** von X wenn $P(X \geq m) \geq \frac{1}{2}$ und $P(X \leq m) \geq \frac{1}{2}$ gilt.

Sehr häufig ist es günstiger an Stelle des Erwartungswertes den Median als Mittelwert anzugeben. Das gilt insbesondere für den Fall, dass einzelne relevante "Ausreißer" vorliegen, die eine starke Verzerrung des Erwartungswertes bewirken. Ein Median ist nicht notwendigerweise eindeutig bestimmt; er ist eindeutig, wenn die Verteilungsfunktion streng monoton wachsend und stetig ist.

Betrachten wir etwa ein Bernoulli-Experiment mit der Erfolgswahrscheinlichkeit p und X als Bernoulli-Zufallsvariablen, so impliziert $P(X \geq m), P(X \leq m) \geq \frac{1}{2}$ im Fall $p < \frac{1}{2}$ stets $m = 0$, im Fall $p > \frac{1}{2}$ stets $m = 1$. Im Fall $p = \frac{1}{2}$ erfüllt jedes $m \in [0, 1]$ diese obige Beziehung.

1.3.15 Beispiel: X sei eine exponentialverteilte Zufallsvariable zu dem Parameter $\alpha > 0$, also besitzt X eine Verteilungsdichte $f : f(t) = \alpha e^{-\alpha t}$ für $t > 0$ und $f(t) = 0$ für $t \leq 0$. Es gilt $E(X) = \frac{1}{\alpha}$ und wegen $F(t) = \int_0^t f(s) ds = e^{-\alpha t}$ für $t > 0$ folgt aus $F(m) = \frac{1}{2}$ unmittelbar $m = \frac{\ln(2)}{\alpha}$. Der Median ist also eindeutig bestimmt.

1.4 Unabhängige Zufallsvariablen

In diesem Abschnitt sei (Ω, \mathcal{A}, P) stets ein Wahrscheinlichkeitsmaßraum. Wir diskutieren jetzt den Begriff der Unabhängigkeit von Mengensystemen und damit zusammenhängend den Begriff unabhängiger Zufallsvariablen. Dieser Begriff spielt eine zentrale Rolle in der weiteren Theorie, speziell auch in Fragestellungen, die mit der Statistik zusammenhängen. Zu Motivation gehen wir von der folgenden Situation aus, die besonders bei Anwendungen sehr häufig auftritt: Es werden voneinander unbeeinflusst zwei (meistens bei Anwendungen mehr) stochastische Versuche durchgeführt. Mathematisch formuliert liegt also die folgende Situation vor:

Für zwei Wahrscheinlichkeitsräume $(\Omega_j, \mathcal{A}_j, P_j)$ $j = 1, 2$ bilden wir den Produktraum (Ω, \mathcal{A}, P) durch

$$\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, \text{ und } P = P_1 \otimes P_2.$$

Für $A_1 \in \mathcal{A}_1$ und $B_2 \in \mathcal{A}_2$ betrachten wir die folgenden Produktmengen $A = A_1 \times \Omega_2$ und $B = \Omega_1 \times B_2$; gehen wir also von zwei stochastischen Experimenten aus, so ist A die Menge, die nur vom Ausgang des ersten Experimentes abhängt und B die Mengen die nur vom Ausgang des zweiten Experimentes abhängt mit den Wahrscheinlichkeiten $P(A) = P_1(A_1)$ und $P(B) = P_2(B_2)$. Es folgt daher $P(A \cap B) = P(A) \cdot P(B)$. Diese Beziehung wählen wir jetzt für eine allgemeinere Definition; diese größere Allgemeinheit ist später bei Anwendungen in der Statistik nützlich, da diese etwa bei Drehungen des Koordinatensystems erhalten bleibt, bei denen aber die Produktsituation nicht mehr vorliegt.

1.4.1 Definition: A und $B \in \mathcal{A}$ heißen *unabhängig*, wenn $P(A \cap B) = P(A)P(B)$ gilt.

Entsprechend heißen Systeme von Teilmengen $\{\mathcal{F}_i \mid i \in I\}$ paarweise unabhängig, wenn für alle $k, j \in I$ mit $k \neq j$ und alle $A \in \mathcal{F}_j$, und $B \in \mathcal{F}_k$ stets $P(B \cap A) = P(B)P(A)$ gilt.

Im Fall von Ereignissen $A, B \in \mathcal{A}$ mit $P(B) > 0$ sind A und B genau dann unabhängig, wenn $P(A|B) = P(A)$ gilt. Diese Aussage folgt unmittelbar aus $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

1.4.2 Beispiel: Es wird mit zwei Würfeln gewürfelt, jeweils einmal. Wir betrachten die Ereignisse

- i) $A = \{\text{Augenzahl des ersten Würfels ist gerade}\}$
 $B = \{\text{Augenzahl des zweiten Würfels ist gleich } 1\}.$

Offenbar gilt $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{6}$, wegen $A \cap B = \{(2, 1), (4, 1), (6, 1)\}$ erhalten wir daher $P(A \cap B) = \frac{1}{12} = P(A) \cdot P(B)$. - die Ereignisse sind also unabhängig.

- ii) $A = \{\text{Summe der Augenzahlen beider Würfel ist gerade}\}$
 $B = \{\text{Augenzahl des ersten Würfels ist gerade}\}.$

Offenbar gilt $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$ und $P(A \cap B) = \frac{1}{4}$ - die Ereignisse sind also unabhängig, obwohl das Ereignis A das Ereignis B beeinflusst.

- iii) $A = \{\text{Summe der Augenzahlen beider Würfel ist gerade, } \leq 8\}$
 $B = \{\text{Augenzahl des ersten Würfels ist } 4 \text{ oder } 6\}$

Es gilt $P(A) = \frac{15}{36}$, $P(B) = \frac{1}{3}$, $P(A \cap B) = \frac{1}{12}$. - Die Ereignisse sind also nicht unabhängig. \square

Der eben für zwei Ereignisse definierte Begriff der Unabhängigkeit soll jetzt auf allgemeinere Systeme von Ereignissen übertragen werden. Dazu muss die Definition etwas abgewandelt werden. Man vergleiche dazu auch Beispiel 1.4.4; dieses Beispiel zeigt die Unterschiedlichkeit der beiden Begriffe.

1.4.3 Definition: Es seien $\mathcal{F}_i \subset \mathcal{A}$ ($i \in I$) Systeme von Teilmengen. Wir nennen diese Systeme unabhängig, wenn für jeweils endlich viele paarweise verschiedene $i(1), \dots, i(r) \in I$ und beliebige $A_{i(1)} \in \mathcal{F}_{i(1)}, \dots, A_{i(r)} \in \mathcal{F}_{i(r)}$ stets

$$P(A_{i(1)} \cap \dots \cap A_{i(r)}) = P(A_{i(1)}) \cdot \dots \cdot P(A_{i(r)})$$

gilt.

Aus der Definition folgt unmittelbar, dass jedes unabhängige System auch paarweise unabhängig ist - man wähle $r = 2$. Die Umkehrung ist allerdings nicht richtig:

1.4.4 Beispiel: Es wird einmal mit zwei Würfeln gewürfelt. Wir setzen

$$\begin{aligned} A_1 &= \{\text{Augenzahl des ersten Würfels ist gerade}\} \\ A_2 &= \{\text{Augenzahl des zweiten Würfels ist gerade}\} \\ A_3 &= \{\text{Augenzahlen beider Würfel stimmen überein}\} \end{aligned}$$

und dann $\mathcal{F}_i = \{A_j\}$ für $j = 1, 2, 3$. Offenbar gilt $P(A_1) = P(A_2) = \frac{1}{2}$, $P(A_3) = \frac{1}{6}$, $P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{12}$, $P(A_1 \cap A_2) = \frac{1}{4}$ und $P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_3) = \frac{1}{12}$.

Das System von Ereignissen $\{\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$ ist demnach paarweise unabhängig, aber nicht unabhängig.

1.4.5 Satz: Für $i \in I$ seien $\mathcal{F}_i \subset \mathcal{A}$ Mengensysteme mit $A \cap B \in \mathcal{F}_i$ für alle $A, B \in \mathcal{F}_i$. Sind die Mengensysteme \mathcal{F}_i für $i \in I$ unabhängig, so sind auch die erzeugten σ -Algebren $\mathcal{A}_i = \mathcal{A}(\mathcal{F}_i)$ für $i \in I$ unabhängig.

Beweis: Wegen der Definition der Unabhängigkeit von Mengensystemen dürfen wir $I = \{1, \dots, n\}$ für ein $n \in \mathbb{N}$ annehmen. Weiter dürfen wir offenbar $\emptyset, \Omega \in \mathcal{F}_i$ für alle $i = 1, \dots, n$ voraussetzen.

Behauptung I: Es seien $A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$ beliebig fixiert. Dann gilt

$$P(A \cap A_2 \cap \dots \cap A_n) = P(A)P(A_2) \cdot \dots \cdot P(A_n)$$

für alle $A \in \mathcal{A}_1$

Beweis von Behauptung I: Die Aussage der Behauptung I gilt offenbar für alle $A \in \mathcal{F}_1$. Definieren wir jetzt Maße Q_1, Q_2 auf (Ω, \mathcal{A}_1) durch

$$Q_1(A) = P(A \cap A_2 \cap \dots \cap A_n) \text{ und } Q_2(A) = P(A)P(A_2) \cdot \dots \cdot P(A_n),$$

so folgt $Q_1(A) = Q_2(A)$ für alle $A \in \mathcal{F}_1$. Wegen der Voraussetzungen über \mathcal{F}_1 folgt aus dem Eindeigkeitssatz unmittelbar $Q_1(A) = Q_2(A)$ für alle $A \in \mathcal{A}(\mathcal{F}_1) = \mathcal{A}_1$ und damit die Behauptung I.

Als unmittelbare Konsequenz der Behauptung I erhalten wir:

Behauptung II: $\mathcal{F}_2, \dots, \mathcal{F}_n, \mathcal{A}_1$ sind unabhängige Mengensysteme.

Die Aussage des Satzes folgt durch wiederholte Anwendungen von Behauptung II: Da $\mathcal{F}_2, \dots, \mathcal{F}_n, \mathcal{A}_1$ die Voraussetzungen des Satzes erfüllen, sind $\mathcal{F}_3, \dots, \mathcal{F}_n, \mathcal{A}_1, \mathcal{A}_2$ unabhängig und folgt schließlich durch n -fache Anwendung des Argumentes die Behauptung. \square

$X : \Omega \rightarrow \mathbb{R}$ sei eine Zufallsvariable. Zu Beginn des Abschnittes 3 wurden folgende Bezeichnungen eingeführt: Wir schreiben $X \in \mathcal{L}^1(P)$ genau dann, wenn $E(|X|) < \infty$ gilt und $X \in \mathcal{L}^2(P)$ genau dann, wenn $E(X^2) < \infty$ gilt. Für $X \in \mathcal{L}^1(P)$ sei weiter $\|X\|_1 = E(|X|)$ und $\|X\|_2 = \sqrt{E(X^2)}$ im Fall von $X \in \mathcal{L}^2(P)$. Es gelten die folgenden Aussagen

1.4.6 Satz: *Es seien $X, Y \in \mathcal{L}^1(P)$ und $\alpha \in \mathbb{R}$. Dann gilt $X + Y \in \mathcal{L}^1(P)$ und $\alpha Y \in \mathcal{L}^1(P)$ mit*

$$\|X + Y\|_1 \leq \|X\|_1 + \|Y\|_1 \quad \text{und} \quad \|\alpha Y\|_1 = |\alpha| \|Y\|_1$$

Auf den sehr einfachen Beweis dieser Aussage, die lediglich eine Umformulierung bekannter Resultate ist, soll verzichtet werden. Auch die folgende Aussage ist eine sehr einfache Konsequenz der Theorie der Skalarprodukträume. Dazu muss man lediglich beachten, dass mit $X, Y \in \mathcal{L}^2(P)$ stets $2|XY| \leq X^2 + Y^2$ und daher $XY \in \mathcal{L}^2(P)$ gilt. Daher ist die Abbildung $(X, Y) \mapsto E(XY)$ ein Skalarprodukt (bis auf die Eigenschaft $E(XX) = 0 \Leftrightarrow X = 0$.) Damit sind dann die Aussagen klar.

1.4.7 Satz: *Es seien $X, Y \in \mathcal{L}^2(P)$ und $\alpha \in \mathbb{R}$.*

i) *Es gilt $XY \in \mathcal{L}^1(P)$ mit*

$$|E(XY)| \leq E(|XY|) \leq \|X\|_2 \cdot \|Y\|_2$$

(*Ungleichung von Cauchy-Schwarz*)

ii) *Es gilt $X + Y, \alpha Y \in \mathcal{L}^2(P)$ mit*

$$\|X + Y\|_2 \leq \|X\|_2 + \|Y\|_2 \quad \text{und} \quad \|\alpha Y\|_2 = |\alpha| \|Y\|_2$$

iii) *Es gilt $X \in \mathcal{L}^1(P)$ mit $E(|X|) = \|X\|_1 \leq \min\{1 + E(X^2), \|X\|_2\}$.*

Wir wollen jetzt den Begriff der Unabhängigkeit auf den Fall von Zufallsvariablen übertragen. Als Motivation diskutieren wir zunächst die Produktsituation bei zwei Wahrscheinlichkeitsmaßräumen $(\Omega_j, \mathcal{A}_j, P_j)$ $j = 1, 2$ und dem Produktraum (Ω, \mathcal{A}, P) :

$$\Omega = \Omega_1 \times \Omega_2, \quad \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, \quad \text{und} \quad P = P_1 \otimes P_2.$$

Sind jetzt $X_j : \Omega_j \rightarrow \mathbb{R}$ für $j = 1, 2$ zwei Zufallsvariable, so können wir diese in trivialer Weise auf $\Omega = \Omega_1 \times \Omega_2$ definieren: Wir setzen

$$\hat{X}_1(\omega_1, \omega_2) = X_1(\omega_1) \quad \text{und} \quad \hat{X}_2(\omega_1, \omega_2) = X_2(\omega_2).$$

Man kann einfach nachrechnen, dass bei diesem Vorgang sich der Erwartungswert nicht verändert. Das folgt unmittelbar aus dem Satz von Fubini wegen $P_j(\Omega_j) = 1$.

Um den Begriff unabhängiger Zufallsvariablen einzuführen bilden wir die von einer Zufallsvariablen $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ erzeugten σ -Algebra $\mathcal{A}(X)$: Zunächst betrachten wir das System

$$\mathcal{E}(X) = \{[X \leq \alpha] \mid \alpha \in \mathbb{R}\}.$$

Da X messbar ist, folgt $\mathcal{E}(X) \subset \mathcal{A}$ und daher gilt

$$\mathcal{A}(X) := \mathcal{A}(\mathcal{E}(X)) \subset \mathcal{A}.$$

Damit ist $\mathcal{A}(X)$ die kleinste σ -Algebra, die das System $\mathcal{E}(X)$ enthält; sie heißt die von der Zufallsvariablen X erzeugte σ -Algebra.

1.4.8 Definition: Ein System $X_j : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ für $j \in I$ von Zufallsvariablen heißt **unabhängig** (paarweise unabhängig), wenn das System $\{\mathcal{A}(X_j) \mid j \in I\}$ der erzeugten σ -Algebren unabhängig (paarweise unabhängig) im Sinne von 1.4.3 ist.

Wegen $[X \leq \alpha] = [X \leq \alpha] \cap [X \leq \beta]$ für alle $\alpha, \beta \in \mathbb{R}$ mit $\alpha \leq \beta$ und jede Zufallsvariable $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ folgt unmittelbar aus 1.4.5 die folgende Charakterisierung der Unabhängigkeit.

1.4.9 Satz: $X_j : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ für $j \in I$ seien Zufallsvariable. Die folgenden Aussagen sind äquivalent:

- i) Die Zufallsvariablen $X_j : j \in I$ sind unabhängig.
- ii) Die Mengensysteme $\mathcal{E}(X_j) : j \in I$ sind unabhängig.

1.4.10 Folgerung: $X_1, \dots, X_n : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ seien Zufallsvariable. Dann gilt:

- i) X_1, \dots, X_n sind genau dann unabhängig, wenn für alle $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ die Ereignisse $[X_1 \leq \alpha_1], \dots, [X_n \leq \alpha_n]$ unabhängig sind.
- ii) Sind X_1, \dots, X_n unabhängige Zufallsvariable, so existieren Folgen von Stufenfunktionen $(X_{j,k})_{k=1}^\infty$ mit $X_{j,k} \rightarrow X_j$, $|X_{j,k}| \uparrow |X_j|$ bei $k \rightarrow \infty$ für $j = 1, \dots, n$ und mit $[X_{j,k} = \alpha] \in \mathcal{A}(X_j)$ für alle $\alpha \in \mathbb{R}$, $j = 1, \dots, n$, $k \in \mathbb{N}$.

Anmerkung: In der Situation von 1.4.10 ii) sagen wir auch, dass $X_{j,k}$ eine $\mathcal{A}(X_j)$ -Stufenfunktion ist.

Beweis: i) folgt unmittelbar aus 1.4.9.

ii) Für diese Aussage vergleiche man 1.4.2 und die nachfolgenden Bemerkungen. □

1.4.11 Satz: $X, Y : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ seien unabhängige Zufallsvariable. Dann sind X und Y **unkorrelierte Zufallsvariable** in dem Sinne, dass $E(XY) = E(X)E(Y)$ gilt.

Beweis: Der Einfachheit halber nehmen wir $X, Y \geq 0$ an. Wegen 1.4.10 ii) existiert eine Folge $(X_k)_{k=1}^\infty$ von $\mathcal{A}(X)$ -Stufenfunktionen und eine Folge $(Y_k)_{k=1}^\infty$ von $\mathcal{A}(Y)$ -Stufenfunktionen mit $0 \leq X_k \uparrow X$ und $0 \leq Y_k \uparrow Y$ bei $k \rightarrow \infty$. Wir fixieren jetzt $k, m \in \mathbb{N}$ und erhalten

$$X_k = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j} \quad \text{und} \quad Y_m = \sum_{r=1}^q \beta_r \mathbb{1}_{B_r}$$

mit geeigneten $\alpha_j, \beta_r \in \mathbb{R}$ und $A_j \in \mathcal{A}(X)$, $B_r \in \mathcal{A}(Y)$ für $j = 1, \dots, n$ und $r = 1, \dots, q$. Wegen

$$X_k Y_m = \sum_{j=1}^n \sum_{r=1}^q \alpha_j \beta_r \mathbb{1}_{A_j} \mathbb{1}_{B_r} = \sum_{j=1}^n \sum_{r=1}^q \alpha_j \beta_r \mathbb{1}_{A_j \cap B_r}$$

und der Unabhängigkeit der Mengen A_j und B_r folgt

$$E(X_k Y_m) = \sum_{j=1}^n \sum_{r=1}^q \alpha_j \beta_r P(A_j \cup B_r) = \sum_{j=1}^n \sum_{r=1}^q \alpha_j \beta_r P(A_j) P(B_r) = E(X_n) E(Y_m).$$

Beim Grenzübergang $k \rightarrow \infty$ folgt $X_k Y_m \uparrow XY_m$ und daher aus dem Konvergenzsatz von Beppo Levi

$$E(X_k Y_m) \rightarrow E(XY_m) \text{ und } E(X_k) \rightarrow E(X) \text{ bei } k \rightarrow \infty.$$

Demnach gilt $E(XY_m) = E(X)E(Y_m)$. Beim Grenzübergang $m \rightarrow \infty$ folgt dann entsprechend

$$E(XY) = E(X)E(Y).$$

□

1.4.12 Definition: $X, Y : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ seien Zufallsvariable mit $X, Y \in \mathcal{L}^2(P)$.

i) $V(X) = E((X - E(X))^2)$ heißt die **Varianz** von X und $\sigma(X) = \sqrt{V(X)}$ die **Streuung** oder **Standard-Abweichung** von X .

ii) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ heißt die **Kovarianz** von X und Y .

Bei einer Zufallsvariablen X ist die Varianz ein Maß für die Abweichung der Werte dieser Zufallsvariablen von dem Erwartungswert: je kleiner die Varianz ist, desto massierter liegen die Werte der Zufallsvariablen in der Nähe vom Erwartungswert.

Aus der Definition folgt weiter unmittelbar, dass die Zufallsvariablen X und Y genau dann unkorreliert sind, wenn $\text{Cov}(X, Y) = 0$ gilt.

1.4.13 Satz: Es seien $X, Y, X_1, \dots, X_n \in \mathcal{L}^2(P)$ und $a, b \in \mathbb{R}$. Dann gilt:

i) $V(X) = E(X^2) - E(X)^2$.

ii) $V(aX + b) = a^2 V(X)$.

iii) $\text{Cov}(X, Y) \leq \sqrt{V(X)V(Y)}$.

iv) Sind X_1, \dots, X_n (paarweise) unkorreliert, so folgt

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n).$$

Anmerkung: Sind $X_1, \dots, X_n \in \mathcal{L}^2(P)$ (paarweise) unabhängig, so sind sie wegen 1.4.11 auch unkorreliert. Demnach gilt hier auch Aussage (iv).

Beweis: i) Es gilt

$$\begin{aligned} V(X) &= E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2. \end{aligned}$$

ii) Aus $E(aX + b) = aE(X) + b$ folgt

$$\begin{aligned} V(aX + b) &= E((aX + b - aE(X) - b)^2) \\ &= E((a(X - E(X)))^2) = a^2 E((X - E(X))^2) = a^2 V(X). \end{aligned}$$

iii) Es seien zunächst $b, d \in \mathbb{R}$, dann gilt

$$\text{Cov}(X + b, Y + d) = E((X + b)(Y + d)) - E(X + b)E(Y + d) = \text{Cov}(X, Y).$$

Wir dürfen also $E(X) = E(Y) = 0$ annehmen und erhalten wegen 1.4.8 (i), der Cauchy-Schwarzschen Ungleichung

$$|\text{Cov}(X, Y)| = |E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)} = \sqrt{V(X)V(Y)}.$$

iv) Es gilt

$$\begin{aligned} V(X_1 + \dots + X_n) &= E((X_1 + \dots + X_n)^2) - (E(X_1) + \dots + E(X_n))^2 \\ &= E\left(\sum_{j=1}^n X_j^2 - \sum_{\substack{j,k=1 \\ j \neq k}}^n X_j X_k\right) - \sum_{j=1}^n (E(X_j))^2 - \sum_{\substack{j,k=1 \\ j \neq k}}^n E(X_j)E(X_k) \\ &= V(X_1) + \dots + V(X_n). \end{aligned}$$

□

Die folgende recht elementar herzuleitende Ungleichung ist unverzichtbar für viele Anwendungen.

1.4.14 Bemerkung: (Tschebyscheff Ungleichung) *Es seien $X \in \mathcal{L}^2(P)$ und $\varepsilon > 0$. Dann gilt*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} V(X).$$

Beweis: Wir setzen $A = [|X - E(X)| \geq \varepsilon]$. Dann gilt $A \in \mathcal{A}$, und wir erhalten wegen $\varepsilon \cdot \mathbb{1}_A \leq |X - E(X)|$ die verlangte Abschätzung:

$$P(A) = E(\mathbb{1}_A) = E(\mathbb{1}_A^2) \leq E\left(\frac{1}{\varepsilon^2} |X - E(X)|^2\right) \leq \frac{1}{\varepsilon^2} E(|X - E(X)|^2) = \frac{1}{\varepsilon^2} V(X).$$

□

Im Anschluss an das folgende Beispiel diskutieren wir in einem Spezialfall die Güte der Abschätzung der Wahrscheinlichkeit für ein bestimmtes Abweichen einer Verteilung vom Erwartungswert mit Hilfe dieser Tschebyscheff-Ungleichung.

1.4.15 Beispiel: (Binomial-Verteilung) Es sei wie früher $\Omega_0 = \{a, b\}$ ein Bernoulli-Experiment mit $p = P(\{a\})$ und $q = 1 - p$. Es sei weiter $X_0(a) = 1, X_0(b) = 0$, wir erhalten damit

$$E(X_0) = p, \quad V(X_0) = E(X_0^2) - (E(X_0))^2 = p - p^2 = pq.$$

Die Binomial-Verteilung ist darstellbar als $X = X_1 + \dots + X_n$ mit den unabhängigen Bernoulli-Variablen X_1, \dots, X_n . Aus 1.4.13 folgt daher

$$V(X) = V(X_1) + \dots + V(X_n) = npq.$$

□

In dem folgenden Beispiel soll die Aussagekraft der Tschebyscheff-Ungleichung demonstriert werden. Es zeigt sich, dass diese Ungleichung nur in Spezialfällen eine gute Berechnungsmöglichkeit für konkrete Aufgabenstellungen liefert.

1.4.16 Beispiel: i) Es sei $n = 20$. Im Fall $p = 0.05$ soll die Wahrscheinlichkeit des Ereignisses A abgeschätzt werden, dass sich der Wert der Zufallsvariablen X , also die Zahl der Erfolge, um mehr als 1 vom Erwartungswert unterscheiden.

Es gilt $E(X) = n \cdot p = 1$. Exakt ausgerechnet ergibt sich für die gesuchte Wahrscheinlichkeit

$$P(A) = \sum_{k=3}^{20} \binom{20}{k} p^k (1-p)^{20-k} = 1 - \sum_{k=0}^2 \binom{20}{k} p^k (1-p)^{20-k} = 0.07548 .$$

Mit Hilfe der Tschebyscheff-Ungleichung erhalten wir die Abschätzung

$$P(A) = P(|X - E(X)| \geq 2) \leq \frac{V(X)}{2^2} = \frac{20 \cdot p(1-p)}{2^2} = 0.2375 .$$

ii) Wir wählen jetzt $p = 0.5$ und lassen ansonsten das Beispiel unverändert. Dann erhalten wir wie eben $P(A) = 0.4966$. Mit der Tschebyscheff-Ungleichung folgt $P(A) \leq 1.25$, damit ist keine Aussage mehr möglich.

1.4.17 Beispiel: (Normal-Verteilung) Es seien $\sigma \in]0, \infty[$ und $\mu \in \mathbb{R}$. Die Zufallsvariable X heißt $N(\mu, \sigma)$ -verteilt, wenn die Verteilungsfunktion $\Phi_{\mu, \sigma}$ von X die Dichte $\varphi_{\mu, \sigma}$ besitzt mit

$$\varphi_{\mu, \sigma}(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

für alle $t \in \mathbb{R}$. Mit $N(\mu, \sigma)$ wird dabei das Bildmaß bezeichnet: $N(0, 1) = P_X$: $P_X(B) = P(X^{-1}(B))$ für alle $B \in \mathcal{B}^1$. Offenbar gilt $N(0, 1)(]-\infty, t]) = \Phi_{\mu, \sigma}(t)$ für alle $t \in \mathbb{R}$. Wir erhalten weiter

$$\begin{aligned} \int_{-\infty}^{\infty} \varphi_{\mu, \sigma}(t) dt &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{s^2}{2\sigma^2}} ds \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} \sigma du = 1. \end{aligned}$$

Es folgt daher

$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} t e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (s+\mu) e^{-\frac{s^2}{2\sigma^2}} ds \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} s e^{-\frac{s^2}{2\sigma^2}} ds + \frac{\mu}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{s^2}{2\sigma^2}} ds = 0 + \mu \quad \text{und} \end{aligned}$$

$$\begin{aligned}
E(X^2) &= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} t^2 e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (s+\mu)^2 e^{-\frac{s^2}{2\sigma^2}} ds \\
&= \frac{1}{\sqrt{2\pi\sigma}} \left(\int_{-\infty}^{\infty} s^2 e^{-\frac{s^2}{2\sigma^2}} ds + 2\mu \int_{-\infty}^{\infty} s e^{-\frac{s^2}{2\sigma^2}} ds + \mu^2 \int_{-\infty}^{\infty} e^{-\frac{s^2}{2\sigma^2}} ds \right) \\
&= \frac{1}{\sqrt{2\pi\sigma}} \sigma^3 \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du + 0 + \mu^2 \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(u \left(-e^{-\frac{u^2}{2}} \right) \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \right) + \mu^2 = \mu^2 + \sigma^2.
\end{aligned}$$

Wir erhalten damit unmittelbar $V(X) = E(X^2) - (E(X))^2 = \sigma^2$. \square

1.4.18 Beispiel: (geometrische Verteilung) X sei eine geometrisch verteilte Zufallsvariable. Dann gilt

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k \cdot q^{k-1} p = p \frac{d}{dq} \sum_{k=1}^{\infty} q^k = \frac{p}{(1-q)^2} = \frac{1}{p} \\
E(X^2) &= \sum_{k=1}^{\infty} k^2 \cdot q^{k-1} p = \sum_{k=1}^{\infty} (k(k-1)q^{k-1}p + kq^{k-1}p) \\
&= qp \sum_{k=1}^{\infty} k(k-1)q^{k-2} + p \sum_{k=1}^{\infty} kq^{k-1} \\
&= qp \left(\frac{d}{dq} \right)^2 \frac{1}{(1-q)} + \frac{1}{p} = 2 \frac{qp}{(1-q)^3} + \frac{1}{p} = \frac{2q}{p^2} + \frac{1}{p}.
\end{aligned}$$

Es folgt unmittelbar

$$V(X) = E(X^2) - E(X)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q}{p^2} + \frac{p}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}. \quad \square$$

1.4.19 Beispiel: (Poisson-Verteilung) X sei eine Poisson-verteilte Zufallsvariable. Dann gilt

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \\
E(X^2) &= \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k \cdot (k-1) \frac{e^{-\lambda} \lambda^k}{k!} + \sum_{k=1}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} \\
&= \lambda^2 \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-2}}{(k-1)!} + \lambda = \lambda^2 + \lambda \text{ und daher} \\
V(X) &= \lambda.
\end{aligned}$$

\square

1.4.20 Beispiel: (Gamma-Verteilung) Es seien $r, \alpha > 0$, und X sei eine $\gamma_{\alpha,r}$ -verteilte Zufallsvariable. Dann gilt $\gamma_{\alpha,r}(t) = \frac{\alpha^r}{\Gamma(r)} t^{r-1} e^{-\alpha t}$ für alle $t > 0$ und $\gamma_{\alpha,r}(t) = 0$ für $t \leq 0$ mit

$\Gamma(r) = \int_0^\infty s^{r-1} e^{-s} ds$. Wir erhalten

$$E(X) = \int_0^\infty t \cdot \frac{\alpha^r}{\Gamma(r)} t^{r-1} e^{-\alpha t} dt = \frac{r}{\alpha} \int_0^\infty \frac{\alpha^{r+1}}{\Gamma(r+1)} t^r e^{-\alpha t} dt = \frac{r}{\alpha} \text{ und}$$

$$E(X^2) = \int_0^\infty t^2 \frac{\alpha^r}{\Gamma(r)} t^{r-1} e^{-\alpha t} dt = \left(\int_0^\infty \frac{\alpha^{r+2}}{\Gamma(r+2)} t^{r+1} e^{-\alpha t} dt \right) \cdot \frac{r(r+1)}{\alpha^2} = \frac{r(r+1)}{\alpha^2}$$

$$V(X) = \frac{r(r+1)}{\alpha^2} - \frac{r^2}{\alpha^2} = \frac{r}{\alpha^2}.$$

□

Für den weiteren Verlauf benötigen wir noch Eigenschaften der gemeinsamen Verteilung eines Zufallsvektors; die in Definition 1.3.6 eingeführt wurde.

1.4.21 Satz: $X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ sei ein Zufallsvektor. Für $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ seien

$$F(t) = P(X \leq t) = P([X_1 \leq t_1] \cap \dots \cap [X_n \leq t_n])$$

und $F_j(t_j) = P(X_j \leq t_j)$ für $j = 1, \dots, n$.

i) Die Zufallsvariablen X_1, \dots, X_n sind genau dann unabhängig, wenn

$$F(t) = F_1(t_1) \cdot \dots \cdot F_n(t_n)$$

für alle $t = (t_1, \dots, t_n)$ gilt.

ii) Es seien X_1, \dots, X_n unabhängige Zufallsvariable mit Dichten f_1, \dots, f_n der Verteilungsfunktionen. Wir setzen $f(t) = f_1(t_1) \cdot \dots \cdot f_n(t_n)$ für alle $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ für alle $t = (t_1, \dots, t_n) \in \mathbb{R}^n$. Für alle $B \in \mathcal{B}^n$ gilt dann

$$P(X^{-1}(B)) = \int_B f(t) d\lambda^n(t)$$

Beweis: i) Diese Aussage folgt unmittelbar aus 1.4.5, da die Gesamtheit aller Mengen der Form $] - \infty, t_1] \times \dots \times] - \infty, t_n]$ ein Erzeugendensystem von \mathcal{B}^n bildet.

ii) Für Mengen der Form $B = [X \leq t]$ folgt die Aussage aus dem Satz von Fubini wegen (i). Wegen des Eindeutigkeitsatzes gilt sie daher für allgemeine $B \in \mathcal{B}^n$. □

Im weiteren Verlauf dieses Abschnittes soll die Verteilung der Summe unabhängiger Zufallsvariablen untersucht werden. Zunächst diskutieren wir ein einfaches Beispiel.

1.4.22 Beispiel: Mit einem Würfel werde beliebig oft gewürfelt. Wir notieren die Ergebnisfolge $\omega = (\omega_j)_{j=1}^\infty$ mit $\omega_j \in \{1, 2, \dots, 6\}$ für alle $j \in \mathbb{N}$. Die Zufallsvariablen X und Y seien definiert durch $X(\omega) = k$ falls $\omega_k = 6$ und $\omega_j \neq 6$ für alle $j < k$ gilt und $Y(\omega) = r$ falls $\omega_{k+r} = 6$ und $\omega_j \neq 6$ für alle j mit $k+r > j > k$ gilt. X gibt also die Zahl der Würfe bis

zur ersten “6” einschließlich und Y die Zahl der Würfe von der ersten “6” bis zur zweiten “6” an. Es seien $p = \frac{1}{6}$ und $q = 1 - p$; für alle $n \in \mathbb{N}$ mit $n > 2$ gilt dann

$$P(X + Y = n) = \sum_{k=1}^{n-1} pq^{k-1}pq^{n-k-1} = p^2q^{n-2}(n - 1)$$

und weiter

$$\begin{aligned} P(X + Y \leq n) &= \sum_{j=2}^n P(X + Y = j) = p^2 \sum_{j=2}^n q^{j-2}(j - 1) \\ &= p^2 \frac{d}{dq} \left(\sum_{j=0}^{n-1} q^j \right) = p^2 \frac{d}{dq} \left(\frac{1 - q^n}{1 - q} \right) = 1 - (np + q)q^{n-1} \end{aligned}$$

Speziell erhalten wir

$$\begin{aligned} P(X + Y \leq 4) &= 0,1319, & P(X + Y \leq 10) &= 0,5155 \\ P(X + Y \leq 20) &= 0,8913, & P(X + Y \leq 30) &= 0,9754 \end{aligned}$$

Definieren wir eine weitere Zufallsvariable Z entsprechend, so erhalten wir wegen der vorstehenden Resultate für $n \geq 3$

$$\begin{aligned} P(X + Y + Z = n) &= \sum_{k=2}^{n-1} p^2q^{k-2}(k - 1)pq^{n-k-1} \\ &= p^3q^{n-3} \sum_{k=2}^{n-1} (k - 1) = p^3q^{n-3} \frac{(n - 2)(n - 1)}{2}. \end{aligned}$$

Damit kann dann wie eben eine (komplizierte) Formel für $P(X + Y + Z \leq n)$ aufgestellt werden. □

Wir diskutieren jetzt die allgemeine Situation. (Ω, \mathcal{A}, P) sei ein beliebiger Wahrscheinlichkeitsraum und $X, Y : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ seien Zufallsvariable. Wir wollen jetzt die Verteilungsfunktion der Zufallsvariablen $X + Y$ bestimmen. Ist jetzt die Abbildung $\text{Add} : \mathbb{R}^2 \rightarrow \mathbb{R}$ definiert durch $\text{Add}(s, t) = s + t$ für alle $(s, t) \in \mathbb{R}^2$, so gilt offenbar $X + Y = \text{Add} \circ (X, Y)$. Für alle $t \in \mathbb{R}$ gilt daher

$$P(X + Y \leq t) = P((X, Y)^{-1} \circ \text{Add}^{-1}(] - \infty, t]) .$$

Für alle $B \in \mathcal{B}^1$ und seien wie früher die Verteilungsmaße P_X und entsprechend P_Y definiert durch

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega \mid X(\omega) \in B\}).$$

Für Intervalle der Form $] - \infty, t]$ gilt daher $F_X(t) = P_X(] - \infty, t])$. Wir betrachten jetzt das Produktmaß $P_X \otimes P_Y$ auf \mathcal{B}^2 und definieren die **Faltung** dieser Maße $P_X * P_Y$ durch

$$P_X * P_Y(B) = P_Y \otimes P_X(\text{Add}^{-1}(B))$$

für alle $B \in \mathcal{B}^1$. Man kann sehr einfach nachrechnen, dass wegen dieser Definition $P_X * P_Y$ ein Wahrscheinlichkeitsmaß auf \mathcal{B}^1 ist. Dabei benutzt man lediglich, dass die mengentheoretische Inverse einer Abbildung verträglich mit den Mengenoperationen ist. Um zu einer Herleitung einer Darstellung des Faltungsproduktes von zwei Maßen zu kommen genügt es das Erzeugendensystem aller Intervalle der Form $] - \infty, a]$ mit $a \in \mathbb{R}$ zu betrachten, da jedes Maß auf \mathcal{B}^1 durch die Werte auf diesem Erzeugendensystem eindeutig bestimmt ist. Für alle $t \in \mathbb{R}$ gilt

$$\text{Add}^{-1}(] - \infty, t]) = \{(s_1, s_2) \mid s_1 + s_2 \leq t\} \quad \text{und daher} \quad \mathbb{1}_{\text{Add}^{-1}(] - \infty, t])}(s_1, s_2) = \mathbb{1}_{] - \infty, t - s_1]}(s_2)$$

für alle $s_2 \in \mathbb{R}$. Weiter gilt wegen der Definition der Produktmaße für alle $t \in \mathbb{R}$

$$\begin{aligned} P_X * P_Y(] - \infty, t]) &= \int_{\mathbb{R}} \mathbb{1}_{] - \infty, t]} d(P_X * P_Y) = \int_{\mathbb{R}^2} \mathbb{1}_{\text{Add}^{-1}(] - \infty, t])} d(P_X \otimes P_Y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{\text{Add}^{-1}(] - \infty, t])}(s_1, s_2) dP_X(s_1) \right) dP_Y(s_2) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{] - \infty, t - s_1]}(s_2) dP_X(s_1) \right) dP_Y(s_2) \\ &= \int_{\mathbb{R}} F_X(t - s_2) dP_Y(s_2) . \end{aligned}$$

Besitzen die Verteilungsfunktionen F_X und F_Y Dichten f_X und f_Y , so gilt weiter

$$\begin{aligned} P_X * P_Y(] - \infty, t]) &= \int_{\mathbb{R}} F_X(t - s_2) f_Y(s_2) ds_2 \\ &= \int_{\mathbb{R}} \int_{-\infty}^{(t-s_2)} f_X(s_1) ds_1 f_Y(s_2) ds_2 \\ &= \int_{\mathbb{R}} \int_{-\infty}^t f_X(s - s_2) ds f_Y(s_2) ds_2 \\ &= \int_{-\infty}^t \left(\int_{\mathbb{R}} f_X(s - s_2) f_Y(s_2) ds_2 \right) ds = \int_{-\infty}^t f_1 * f_2(s) ds \end{aligned}$$

mit der **Faltung** der Dichten $f_X * f_Y(t) = \int_{-\infty}^{\infty} f_X(t - s) f_Y(s) ds$. Wegen der Konstruktion gilt

$$\int_{-\infty}^{\infty} f_X * f_Y(s) ds = 1 .$$

Demnach erhalten wir die folgende Aussage.

1.4.23 Satz: (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum $X, Y : \Omega \rightarrow \mathbb{R}$ seien unabhängige Zufallsvariablen mit den Dichten $f_X, f_Y : \mathbb{R} \rightarrow \mathbb{R}$. Für $t \in \mathbb{R}$ sei

$$f(t) = f_X * f_Y(t) = \int_{-\infty}^{\infty} f_X(s)f_Y(t-s)ds$$

- die Faltung von f_X und f_Y . Dann ist f die Dichte der Zufallsvariablen $X + Y$.

Besonders wichtig ist die folgende Aussage, die besagt, dass die Summe zweier unabhängiger, normalverteilter Zufallsvariablen wieder normalverteilt ist.

1.4.24 Satz: (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum, die Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{R}$ seien unabhängig, und normalverteilt mit den Erwartungswerten μ_X, μ_Y und den Streuungen σ_X, σ_Y . Dann ist die Zufallsvariable $Z = X + Y$ normalverteilt mit dem Erwartungswert $\mu_Z = \mu_X + \mu_Y$ und der Varianz $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

Beweis: Die Zufallsvariable $X - \mu_X$ ist $N(0, \sigma_X^2)$ -verteilt; wir dürfen also $\mu_X = \mu_Y = 0$ annehmen. X und Y besitzen daher Verteilungsfunktionen mit den Dichten

$$\varphi_{0, \sigma_X^2}(t) = \frac{1}{\sqrt{2\pi\sigma_X}} \exp\left(-\frac{t^2}{2\sigma_X^2}\right) \text{ und } \varphi_{0, \sigma_Y^2}(t) = \frac{1}{\sqrt{2\pi\sigma_Y}} \exp\left(-\frac{t^2}{2\sigma_Y^2}\right)$$

Wir setzen $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$. Wegen 1.4.23 besitzt die Zufallsvariable $Z = X + Y$ daher die Verteilungsfunktion mit der Dichte

$$\varphi_{0, \sigma_X^2} * \varphi_{0, \sigma_Y^2}(t) = \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left[\frac{(t-s)^2}{\sigma_X^2} + \frac{s^2}{\sigma_Y^2}\right]\right) ds.$$

Zur Auswertung dieses Integral fixieren wir ein $t \in \mathbb{R}$ und erhalten mit der Substitution $w = \frac{\sigma_Z}{\sigma_X\sigma_Y} s - \frac{\sigma_Y}{\sigma_X\sigma_Z} t$

$$ds = \frac{\sigma_X\sigma_Y}{\sigma_Z} dw, \quad s = \frac{\sigma_X\sigma_Y}{\sigma_Z} w + \frac{\sigma_Y^2}{\sigma_Z^2} t \quad s - t = \frac{\sigma_X\sigma_Y}{\sigma_Z} w - \frac{\sigma_X^2}{\sigma_Z^2} t \text{ und}$$

$$\left(\frac{t-s}{\sigma_X}\right)^2 + \left(\frac{s}{\sigma_Y}\right)^2 = \frac{1}{\sigma_Z^2} \left[\left(\sigma_Y w - \frac{\sigma_X}{\sigma_Z} t\right)^2 + \left(\sigma_X w - \frac{\sigma_Y}{\sigma_Z} t\right)^2 \right] = w^2 + \frac{1}{\sigma_Z^2} t^2.$$

Damit erhalten wir

$$\varphi_{0, \sigma_X^2} * \varphi_{0, \sigma_Y^2} = \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\frac{w^2}{2}\right) \exp\left(-\frac{t^2}{2\sigma_Z^2}\right) \frac{\sigma_X\sigma_Y}{\sigma} dw = \varphi_{0, \sigma_Z^2}(t). \quad \square$$

1.4.25 Beispiel: X_1 und X_2 seien exponentialverteilt zum Parameter $\alpha > 0$. Man bestimme die Dichte der Verteilungsfunktion f von $X_1 + X_2$ und die Verteilungsfunktion.

Es gilt $f(t) = 0$ für $t \leq 0$. Für $t > 0$ erhalten wir aus der Faltungsformel

$$f(t) = \int_0^t \alpha e^{-\alpha(t-s)} \alpha e^{-\alpha s} ds = \alpha^2 \int_0^t e^{-\alpha t} ds = \alpha^2 t e^{-\alpha t} = \gamma_{\alpha, 2}(t). \quad \square$$

1.5 Grenzwertsätze, der zentrale Grenzwertsatz

Das Ziel dieses Abschnittes ist es, den zentralen Grenzwertsatz herzuleiten, der eine fundamentale Bedeutung in der Stochastik besitzt. Wir geben hier einen relativ elementaren (aber nicht einfachen) Beweis dieses Satzes an, der sich in ähnlicher Form in dem Buch von Georgii befindet.

Wir betrachten in diesem Abschnitt das Grenzverhalten einer Folge von unabhängigen Zufallsvariablen. Motiviert wird diese Untersuchung durch die Aufgabenstellung aus einer Vielzahl unabhängig durchgeführter Wiederholungen des gleichen stochastischen Experimentes Aussagen über Erwartung und Verteilungsfunktion der Zufallsvariablen machen zu können, also einer Aufgabenstellung der Statistik. Ein weiterer Vorteil dieser Grenzwerttheorie ist die Möglichkeit viele Verteilungen – wie etwa die Binomialverteilung – für große n berechnen zu können, was elementar gar nicht oder kaum möglich ist.

Zunächst zeigen wir eine mehr technische Aussage über die Existenz unendlicher Produkte von Wahrscheinlichkeitsräumen. Diese Aussage liefert eine Motivation für die weiteren beweistechnischen Konstruktionen, die allerdings für konkrete Berechnungen eine untergeordnete Rolle spielen. Beim Beweis knüpfen dabei an die Produkttheorie endlich vieler Maßräumen an und benutzen diese dabei zur Durchführung einer mehr formalen Konstruktion.

1.5.1 Satz: Für alle $n \in \mathbb{N}$ seien $(\Omega_n, \mathcal{A}_n, P_n)$ Wahrscheinlichkeitsmaßräume. Wir setzen

$$\begin{aligned}\Omega &= \Omega_1 \times \dots \times \Omega_n \times \Omega_{n+1} \times \dots, \\ \mathcal{A}_0 &= \{A_1 \times \dots \times A_n \times \dots \mid A_j \in \mathcal{A}_j, \text{ mit } A_j = \Omega_j \text{ für fast alle } n \in \mathbb{N}\} \\ \mathcal{A} &= \mathcal{A}(\mathcal{A}_0),\end{aligned}$$

die von \mathcal{A}_0 erzeugte σ -Algebra. Dann existiert auf \mathcal{A} ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P mit

$$P(A_1 \times \dots \times A_n \times \Omega_{n+1} \times \dots) = \prod_{k=1}^n P_k(A_k)$$

für alle $n \in \mathbb{N}_0$ und alle $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$.

Zusatz: Sind $X_n : \Omega_n \rightarrow \mathbb{R}$ Zufallsvariable mit $X_n \in \mathcal{L}^1(P_n)$ für alle $n \in \mathbb{N}$, ist E_n der Erwartungswert bezüglich $(\Omega_n, \mathcal{A}_n, P_n)$, und ist $Q_n : \Omega \rightarrow \Omega_n$ die Projektion von Ω auf die n -te Komponente, so sind die Zufallsvariablen $X_n \circ Q_n$ unabhängig mit

$$E(X_n \circ Q_n) = E_n(X_n) \text{ für alle } n \in \mathbb{N}.$$

Beweis: Für den Beweis führen wir die folgenden Bezeichnungen ein: Wir setzen

$$\Omega^{[n]} = \Omega_1 \times \dots \times \Omega_n, \text{ und } \Omega^{[n \rightarrow]} = \Omega_{n+1} \times \Omega_{n+2} \times \dots$$

für alle $n \in \mathbb{N}$. Auf $\Omega^{[n]}$ sei weiter

$$\mathcal{A}^{[n]} = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n \text{ und } P^{[n]} = P_1 \otimes \dots \otimes P_n$$

die endliche Produkt- σ -Algebra mit dem Produktmaß $P^{[n]}$, und

$$P_0(A \times \Omega^{[n \rightarrow]}) = (P_1 \otimes \dots \otimes P_n)(A) = P^{[n]}(A)$$

für alle $A \in \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$, also für alle $A \times \Omega^{[n \rightarrow]} \in \mathcal{A}_0$, weiter setzen wir noch

$$P^{[m \rightarrow]}(A \times \Omega^{[n \rightarrow]}) = P_{m+1} \otimes \dots \otimes P_n(A)$$

für alle $A \in \mathcal{A}_{m+1} \otimes \dots \otimes \mathcal{A}_n$. Wegen

$$\begin{aligned} (A \times \Omega^{[n \rightarrow]}) \cap (B \times \Omega^{[n \rightarrow]}) &= (A \cap B) \times \Omega^{[n \rightarrow]} \text{ und} \\ (A \times \Omega^{[n \rightarrow]}) \cup (B \times \Omega^{[n \rightarrow]}) &= (A \cup B) \times \Omega^{[n \rightarrow]} \end{aligned}$$

für alle $n \in \mathbb{N}$ und alle $A, B \in \mathcal{A}^{[n]}$ ist \mathcal{A}_0 eine Algebra und P_0 eine endlich additive Mengenfunktion auf \mathcal{A}_0 . Wir wollen jetzt zeigen, dass $P^{[n]}$ die Eigenschaften des Stetigkeitssatzes für additive Mengenfunktionen erfüllt. Wegen des Fortsetzungssatzes der Maßtheorie und des Eindeutigkeitsatzes lässt sich dann P_0 eindeutig zu einem Wahrscheinlichkeitsmaß auf \mathcal{A} fortsetzen. Ist diese Aussage falsch, so existieren ein $\delta > 0$ und $A_0, A_k \in \mathcal{A}^{[0]}$ mit $A_k \supset A_{k+1}$ für alle $k \in \mathbb{N}$ mit $P_0(A_k) \geq P_0(A_0) + 2\delta$ für alle $k \in \mathbb{N}$ und mit $\bigcap_{k=1}^{\infty} A_k = A_0$. Wir setzen

$B_k = A_k \setminus A_0$ und erhalten $P_0(B_k) \geq 2\delta$, $B_k \supset B_{k+1}$ für alle $n \in \mathbb{N}$, und $\bigcap_{k=1}^{\infty} B_k = \emptyset$.

Da jedes B_k von der Form $B_k = B \times \Omega^{[n(k) \rightarrow]}$ mit $B \in \mathcal{A}^{[n(k)]}$ für ein geeignetes $n(k) \in \mathbb{N}$ können wir die Indikatorfunktion $\mathbb{1}_{B_k}$ bezüglich $P^{[n]}$ und $P^{[m \rightarrow]}$ integrieren für alle $n \geq n(k)$, da diese auf den endlichen Produkten Maße sind. Zur Vereinfachung der Schreibweise benutzen wir daher stets die Schreibweise $dP^{[m \rightarrow]}$ für die Integration der Indikatorfunktion bezüglich eines Maßes $d(P_{m+1} \otimes \dots \otimes P_n)$ für ein beliebiges $n > m$.

Zu Konstruktion des Widerspruchs nehmen wir an, dass zu einem $m \in \mathbb{N}$ im Fall $m > 1$ $\omega_1 \in \Omega_1, \dots, \omega_{m-1} \in \Omega_{m-1}$ existieren mit

$$\int_{\Omega^{[r \rightarrow]}} \mathbb{1}_{A_k}(\omega_1, \dots, \omega_r, \omega_{r+1}, \dots) dP^{[r \rightarrow]}(\omega_{r+1}, \dots) \geq \left(1 + \frac{1}{r+1}\right) \delta$$

für alle $n \in \mathbb{N}$ und für alle $0 \leq r \leq m-1$. Wir bestimmen jetzt ein $\omega_m \in \Omega_m$ mit der entsprechenden Eigenschaft: Dazu setzen wir setzen

$$Y_k(\omega_m) = \int_{\Omega^{[m \rightarrow]}} \mathbb{1}_{B_k}(\omega_1, \dots, \omega_{m-1}, \omega_m, \omega_{m+1} \dots) dP^{[m \rightarrow]}(\omega_{m+1}, \dots)$$

für alle $\omega_m \in \Omega_m$ und erhalten wegen des Satzes von Fubini

$$\begin{aligned} & \int_{\Omega^{[m-1 \rightarrow]}} \mathbb{1}_{B_k}(\omega_1, \dots, \omega_{m-1}, \omega_m, \dots) dP^{[m-1 \rightarrow]}(\omega_m, \dots) \\ &= \int_{\Omega_m} \left(\int_{\Omega^{[m \rightarrow]}} \mathbb{1}_{B_k}(\omega_1, \dots, \omega_{m-1}, \omega_m, \omega_{m+1} \dots) dP^{[m \rightarrow]}(\omega_{m+1}, \dots) \right) dP_m(\omega_m) \\ &= \int_{\Omega_m} Y_k(\omega_m) dP_m(\omega_m) \geq \left(1 + \frac{1}{m}\right) \delta \end{aligned}$$

für alle $k \in \mathbb{N}$. Wegen $B_{k+1} \subset B_k$ gilt

$$1 \geq Y_k(\omega_m) \geq Y_{k+1}(\omega_m)$$

für alle $k \in \mathbb{N}$ und alle $\omega_m \in \Omega_m$. Es sei $Y(\omega_m) = \lim_{k \rightarrow \infty} Y_k(\omega_m)$ für alle $\omega_m \in \Omega_m$. Aus dem Konvergenzsatz von Lebesgue folgt

$$\int_{\Omega_m} Y_k(\omega_m) dP_m(\omega_m) \rightarrow \int_{\Omega_m} Y(\omega_m) dP_m(\omega_m) \geq \left(1 + \frac{1}{m}\right)\delta.$$

Speziell existiert also ein $\omega_m \in \Omega_m$ mit $Y_k(\omega_m) \geq Y(\omega_m) \geq \left(1 + \frac{1}{m+1}\right)\delta$ für alle $k \in \mathbb{N}$. Damit ist die rekursive Konstruktion beendet. Wegen der Konstruktion gilt $(\omega_1, \dots, \omega_m, \dots) \in B_k$ für alle $k \in \mathbb{N}$. Das ist ein Widerspruch zur Voraussetzung.

Die im Zusatz behauptete Unabhängigkeit folgt unmittelbar aus der Produktsituation; man vergleiche dazu die Diskussion zu Beginn der Abschnittes 1.4. Die Gleichheit der Erwartungswerte $E(X_n \circ Q_n) = E_n(X_n)$ gilt im Fall $X_n = \mathbb{1}_{A_n}$ und damit auch allgemein. \square

Wir beginnen die Untersuchung des Grenzverhaltens einer Folge von Zufallsvariablen mit einer nützlichen mengentheoretischen Version einer Grenzwertaussage.

1.5.2 Lemma: (Borel-Cantelli) Für eine Folge $(A_n)_{n=1}^\infty \subset \mathcal{A}$ bilden wir den (mengentheoretischen) Limes superior dieser Folge:

$$A := \limsup_{n \rightarrow \infty} A_n = \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty A_n = \left\{ a \in \Omega \mid a \in A_n \text{ für unendlich viele } n \in \mathbb{N} \right\}.$$

i) Im Fall $\sum_{k=1}^\infty P(A_k) < \infty$ gilt $P(A) = 0$.

ii) Sind die Ereignisse $\{A_k \mid k \in \mathbb{N}\}$ unabhängig mit $\sum_{k=1}^\infty P(A_k) = \infty$, so folgt $P(A) = 1$.

Beweis: i) Wegen der Monotonie gilt $A \subset \bigcup_{m=n}^\infty A_m$ für alle $n \in \mathbb{N}$ und daher

$$P(A) \leq P\left(\bigcup_{m=n}^\infty A_m\right) \leq \sum_{m=n}^\infty P(A_m).$$

Aus der Voraussetzung folgt $\sum_{m=n}^\infty P(A_m) \rightarrow 0$ bei $n \rightarrow \infty$, da die Reihe konvergiert. Es gilt daher $P(A) = 0$.

ii) Wegen $A^c = \bigcup_{n=1}^\infty \bigcap_{m=n}^\infty A_m^c$ und $1 - t \leq e^{-t}$ für $t \in \mathbb{R}$ folgt aus dem Stetigkeitssatz für Maße

$$\begin{aligned} P(A^c) &\leq \sum_{n=1}^\infty P\left(\bigcap_{m=n}^\infty A_m^c\right) = \sum_{n=1}^\infty \lim_{r \rightarrow \infty} P\left(\bigcap_{m=n}^r A_m^c\right) \\ &= \sum_{n=1}^\infty \lim_{r \rightarrow \infty} \prod_{m=n}^r (1 - P(A_m)) \leq \sum_{n=1}^\infty \lim_{r \rightarrow \infty} \exp\left(-\sum_{m=n}^r P(A_m)\right) = 0 \end{aligned}$$

wegen der Unabhängigkeit der A_n^c , es gilt also $P(A) = 1$. □

Als erste Konsequenz dieser Aussage und der Tschebyscheff-Ungleichung 1.4.14 zeigen wir eine Aussage, die im Spezialfall einer Folge von gleichverteilten Zufallsvariablen besagt, dass das arithmetische Mittel dieser Folge gegen den Erwartungswert dieser Verteilung fast sicher konvergiert. Dabei konvergiert eine Folge von $(X_n)_{n=1}^\infty$ von reellen Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) fast sicher (P -fast-überall) gegen eine Zufallsvariable X_0 , wenn ein $A \in \mathcal{A}$ existiert mit $P(A) = 0$ und $X_n(\omega) \rightarrow X_0(\omega)$ bei $n \rightarrow \infty$ gilt für alle $\omega \in A^c$.

Die nachfolgende Aussage wird allerdings etwas allgemeiner formuliert. Im Anschluss an den Beweis dieses Satzes sollen diese Konvergenzarten dann näher beschrieben werden.

1.5.3 Satz: (Starkes Gesetz der großen Zahl) $(X_n)_{n=1}^\infty \subset \mathcal{L}^2(P)$ sei eine Folge unkorrelierter Zufallsvariablen mit

$$M := \sup_{n \in \mathbb{N}} V(X_n) < \infty.$$

Dann gilt

$$\frac{1}{n} \sum_{k=0}^n (X_k - E(X_k)) \rightarrow 0$$

bei $n \rightarrow \infty$ fast sicher (P -fast überall).

Zusatz: Für alle $\varepsilon > 0$ und alle $n \in \mathbb{N}$ gilt speziell

$$P\left(\left|\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k))\right| \geq \varepsilon\right) \leq \frac{M}{n\varepsilon^2}.$$

Beweis:(I) Wir zeigen zunächst den Zusatz: Für jedes $n \in \mathbb{N}$ sei

$$Z_n = \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)).$$

Da die Zufallsvariablen X_n unkorreliert sind, gilt $Z_n \in \mathcal{L}^2(P)$ mit $E(Z_n) = 0$ und

$$V(Z_n) = \frac{1}{n^2} \sum_{k=1}^n V(X_k) \leq \frac{M}{n}$$

für alle $n \in \mathbb{N}$. Die Aussage ist jetzt eine Konsequenz der Tschebyscheff Ungleichung 1.4.14.

(II) Offenbar dürfen wir $E(X_n) = 0$ für alle $n \in \mathbb{N}$ annehmen; sonst setzen wir $Y_k = X_k - E(X_k) \in \mathcal{L}^2(P)$; diese Folge $(Y_n)_{n=1}^\infty$ ist ebenfalls unkorreliert. Wir betrachten jetzt

$$Z_n = \frac{1}{n} \sum_{k=1}^n X_k \text{ und } A_n(\varepsilon) = [|Z_n| \geq \varepsilon]$$

für alle $\varepsilon > 0$ und $n \in \mathbb{N}$. Wegen (I) gilt $P(A_n(\varepsilon)) \leq \frac{M}{n^2\varepsilon^2}$ und daher $\sum_{n=1}^\infty P(A_n(\varepsilon)) < \infty$.

Aus den Lemmas von Borel Cantelli folgt $P(\limsup_{n \rightarrow \infty} A_n(\varepsilon)) = 0$ für alle $\varepsilon > 0$ und wegen

der Definition des Limes superior gilt

$$P(\{\omega \mid \exists k \in \mathbb{N} \text{ mit } |Z_{n^2}(\omega)| \geq \frac{1}{k} \text{ f\"ur unendlich viele } n \in \mathbb{N}\}) \\ \leq \sum_{k=1}^{\infty} P(|Z_{n^2}| \geq \frac{1}{k} \text{ f\"ur unendlich viele } n \in \mathbb{N}) = 0$$

Das bedeutet $Z_{n^2} \rightarrow 0$ fast sicher bei $n \rightarrow \infty$.

(III) Zu jedem $m \in \mathbb{N}$ bestimmen wir ein $n = n(m) \in \mathbb{N}$ mit $n^2 \leq m < (n+1)^2$. F\"ur $k \in \mathbb{N}$ sei $Y_k = k \cdot Z_k$. F\"ur $\varepsilon > 0$ gilt wegen der Tschebyscheff Ungleichung 1.4.14

$$P(|Y_m - Y_{n^2}| \geq \varepsilon n^2) \leq \frac{1}{\varepsilon^2 n^4} V(X_{n^2+1} + \dots + X_m) \leq \frac{M(m - n^2)}{\varepsilon^2 n^4}.$$

Daher gilt

$$\sum_{m=1}^{\infty} P(|Y_m - Y_{n(m)^2}| \geq \varepsilon n(m)^2) \leq \frac{M}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{m=n^2}^{(n+1)^2-1} \frac{m - n^2}{n^4} \\ = \frac{M}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{k=0}^{2n} \frac{k + n^2 - n^2}{n^4} = \frac{M}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{n(2n+1)}{n^4} < \infty$$

Wie im ersten Schritt folgt mit dem Lemma von Borel-Cantelli 1.5.2

$$P\left(\left[\left| \frac{1}{n(m)^2} Y_m - Z_{n(m)^2} \right| \rightarrow 0 \right]_{(m \rightarrow \infty)}\right) = 1$$

Wegen $|Z_m| = \frac{1}{m} |Y_m| \leq \frac{1}{n(m)^2} |Y_m|$ und Schritt II gilt $P([Z_m \rightarrow 0]_{(m \rightarrow \infty)}) = 1$. □

Anmerkung: Ist $(Y_n)_{n=1}^{\infty}$ eine Folge von Zufallsvariablen $Y_n : \Omega \rightarrow \mathbb{R}$ und Y eine Zufallsvariable, mit

$$P(|Y_n - Y| \geq \varepsilon) \rightarrow 0 \text{ bei } n \rightarrow \infty \text{ f\"ur alle } \varepsilon > 0,$$

so heit die Folge $(Y_n)_{n=1}^{\infty}$ **stochastisch konvergent** oder **konvergent dem Mae nach**. Ist die Folge $(Z_n)_{n=1}^{\infty}$ wie in 1.5.3 definiert, so besagt die Aussage des Zusatzes dann $Z_n \rightarrow 0$ stochastisch bei $n \rightarrow \infty$; diese stochastische Konvergenz ist allerdings schcher als die im vorstehenden Satz bewiesene fast sichere Konvergenz. Man vergleiche dazu die folgende Aussage, die mehr von theoretischem Interesse ist, ber den Zusammenhang einiger der unterschiedlichen Konvergenzbegriffe fr Folgen von Zufallsvariablen.

1.5.4 Bemerkung: Fr alle $n \in \mathbb{N}_0$ sei X_n eine reelle (oder komplexe) Zufallsvariable auf (Ω, \mathcal{A}, P) .

i) Gilt $X_n \rightarrow X_0$ fast sicher bei $n \rightarrow \infty$, so folgt $X_n \rightarrow X_0$ stochastisch bei $n \rightarrow \infty$.

ii) Gilt $X_n \rightarrow X_0$ fast sicher bei $n \rightarrow \infty$ und existiert eine Zufallsvariable $0 \leq Y \in \mathcal{L}^1(P)$ mit $|X_n| < Y$ fr alle $n \in \mathbb{N}$, so folgt

$$\|X_n - X_0\|_{1 \rightarrow 0} \text{ bei } n \rightarrow \infty.$$

iii) Gilt $\|X_n - X_0\|_{1 \rightarrow 0}$ bei $n \rightarrow \infty$, so existiert eine Teilfolge $(k(n))_{n=1}^{\infty}$ mit $X_{k(n)} \rightarrow X_0$ fast sicher bei $n \rightarrow \infty$.

iv) Gilt $\|X_n - X_0\|_{1 \rightarrow 0}$ bei $n \rightarrow \infty$, so folgt $X_n \rightarrow X_0$ stochastisch bei $n \rightarrow \infty$.

Beweis: Wir dürfen in allen Aussagen $X_0 = 0$ annehmen.

i) Für $\varepsilon > 0$ sei $A_n(\varepsilon) = \{\omega \in \Omega \mid \sup_{k \geq n} |X_k(\omega)| \geq \varepsilon\}$. Wegen $X_n(\omega) \rightarrow 0$ bei $n \rightarrow \infty$ für alle $\omega \in A$ für eine Menge $A \in \mathcal{A}$ mit $P(A) = 1$ folgt $A_n(\varepsilon) \supset A_{n+1}(\varepsilon)$ für alle $n \in \mathbb{N}$ und $\bigcap_{n=1}^{\infty} A_n(\varepsilon) \subset A^c$. Aus dem Stetigkeitssatz für Maße folgt $P(A_n(\varepsilon)) \rightarrow 0$ bei $n \rightarrow \infty$.

ii) Diese Aussage folgt unmittelbar aus dem Konvergenzsatz von Lebesgue.

iii) Wir wählen eine Teilfolge $(k(n))_{n=1}^{\infty}$ mit $\|X_{k(n)}\|_1 \leq 2^{-n}$ für alle $n \in \mathbb{N}$ und setzen dann $Z_n = |X_{k(1)}| + \dots + |X_{k(n)}|$ und $Z = \sum_{n=0}^{\infty} |X_{k(n)}|$. Aus dem Konvergenzsatz von Beppo Levi folgt $\int Z_n dP \rightarrow \int Z dP$, also $\|Z\|_1 \leq 1$ und damit $P(\{\omega \in \Omega \mid Z(\omega) = \infty\}) = 0$. Wir erhalten $X_{k(n)}(\omega) \rightarrow 0$ bei $n \rightarrow \infty$ für alle $\omega \in \Omega$ mit $Z(\omega) \neq \infty$.

iv) Für $\varepsilon > 0$ und $n \in \mathbb{N}$ sei

$$A_n(\varepsilon) = \{\omega \in \Omega \mid |X_n(\omega)| \geq \varepsilon\}.$$

Offenbar gilt $\mathbb{1}_{A_n(\varepsilon)} \leq \frac{1}{\varepsilon} |X_n|$; und wir erhalten daher

$$P(A_n(\varepsilon)) \leq \frac{1}{\varepsilon} \|X_n\|_1 \rightarrow 0 \text{ bei } n \rightarrow \infty.$$

□

Wir sind jetzt in der Lage eine Variante der Gesetzes der großen Zahl als eine einfache Konsequenz der Aussage 1.5.3 zu beweisen. Diese zeigt die Konvergenz in der \mathcal{L}^1 -Norm, wenn vorausgesetzt wird, dass die Folge der Zufallsvariablen unabhängig und identisch verteilt ist mit $X_1 \in \mathcal{L}^1(P)$. Es existieren allerdings Varianten, die die fast sichere Konvergenz auch unter diesen Voraussetzungen nachweisen.

1.5.5 Folgerung: $(X_n)_{n=1}^{\infty} \subset \mathcal{L}^1(P)$ sei eine Folge von identisch verteilten, unabhängigen Zufallsvariablen. Dann gilt

$$\left\| \frac{1}{n} \sum_{k=0}^n X_k - E(X_k) \right\|_1 \rightarrow 0 \text{ bei } n \rightarrow \infty.$$

Beweis: Wir nehmen $E(X_n) = 0$ an und setzen $Z = |X_1|$. Für alle $r \in \mathbb{N}$ sei $Z_r = \min\{Z, r\}$. Wir fixieren ein $\varepsilon > 0$ beliebig. Es gilt $Z_r \uparrow Z$ bei $r \rightarrow \infty$; wegen des Konvergenzsatzes von Beppo Levi 0.3.9 folgt daher $E(Z_r) \uparrow E(Z)$ bei $r \rightarrow \infty$. Speziell existiert also ein $r \in \mathbb{N}$ mit

$$0 \leq E(Z) - E(Z_r) = E(Z - Z_r) = E(|Z - Z_r|) = \|Z - Z_r\|_1 < \varepsilon.$$

Es sei jetzt $h : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch $h(t) = \max\{-r, \min\{t, r\}\}$ für alle $t \in \mathbb{R}$. Wir setzen weiter $Y_n = h \circ X_n = \max\{-r, \min\{X_n, r\}\}$ für alle $n \in \mathbb{N}$; damit schneiden wir die Zufallsvariable oben bei r und unten bei $-r$ ab. Die Folge $(Y_n)_{n=1}^{\infty}$ besteht offenbar aus unabhängigen, gleichverteilten Zufallsvariablen, und es gilt $V(Y_n) \leq r^2$ für alle $n \in \mathbb{N}$. Durch Betrachtung des Positivteils und des Negativteils erkennt man weiter

$$E(|X_n - Y_n|) = E(|Z - Z_r|) < \varepsilon \text{ für alle } n \in \mathbb{N}.$$

Die Voraussetzungen des Gesetzes der großen Zahl 1.5.3 sind erfüllt. Daher gilt

$$\frac{1}{n} \sum_{k=1}^n Y_k \rightarrow E(Y_1) \text{ fast sicher bei } n \rightarrow \infty .$$

Es folgt unmittelbar aus dem Konvergenzsatz von Lebesgue 0.3.13

$$\left\| \frac{1}{n} \sum_{k=1}^n Y_k - E(Y_1) \right\|_1 = E \left(\left| \frac{1}{n} \sum_{k=1}^n Y_k - E(Y_1) \right| \right) \rightarrow 0 \text{ bei } n \rightarrow \infty .$$

Wir erhalten weiter

$$\begin{aligned} E \left(\left| \frac{1}{n} \sum_{k=1}^n X_k \right| \right) &\leq E \left(\left| \frac{1}{n} \sum_{k=1}^n (X_k - Y_k) \right| \right) + E \left(\left| \frac{1}{n} \sum_{k=1}^n Y_k - E(Y_1) \right| \right) + E(|Y_1 - X_1|) \\ &\leq \frac{1}{n} \sum_{k=1}^n E(|X_k - Y_k|) + E \left(\left| \frac{1}{n} \sum_{k=1}^n Y_k - E(Y_1) \right| \right) + E(|Y_1 - X_1|) \\ &\leq 2E(|Y_1 - X_1|) + E \left(\left| \frac{1}{n} \sum_{k=1}^n Y_k - E(Y_1) \right| \right) \\ &\leq 2\varepsilon + E \left(\left| \frac{1}{n} \sum_{k=1}^n Y_k - E(Y_1) \right| \right) \rightarrow 2\varepsilon \text{ bei } n \rightarrow \infty . \end{aligned}$$

Da $\varepsilon > 0$ beliebig vorgegeben war, folgt unmittelbar die Aussage. \square

Im weiteren Verlauf dieses Abschnittes soll jetzt der zentrale Grenzwertsatz bewiesen werden. Dazu benötigen wir einen weiteren Konvergenzbegriff, der unabhängig von dem ursprünglichen Definitionsbereich der Zufallsvariablen ist und sich ausschließlich an den Verteilungsfunktionen orientiert.

1.5.6 Definition: (Ω, \mathcal{A}, P) und $(\Omega_1, \mathcal{A}_1, P_1)$ seien Wahrscheinlichkeitsräume, $F : \mathbb{R} \rightarrow \mathbb{R}$ sei die Verteilungsfunktion einer Zufallsvariablen $Y : \Omega_1 \rightarrow \mathbb{R}$, und es $(X_n)_{n=1}^\infty$ sei eine Folge von Zufallsvariablen auf Ω mit den Verteilungsfunktionen F_n . Die Folge $(X_n)_{n=1}^\infty$ heißt **verteilungskonvergent** gegen Y bei $n \rightarrow \infty$, wenn

$$F_n(\tau) \rightarrow F(\tau) \text{ bei } n \rightarrow \infty$$

gilt für alle $\tau \in \mathbb{R}$, in denen F stetig ist. Wir schreiben dann $X_n \xrightarrow{d} Y$ oder auch $X_n \xrightarrow{d} F$ bei $n \rightarrow \infty$. Häufig ist auch die Schreibweise $X_n \xrightarrow{\mathcal{L}} Y$ bei $n \rightarrow \infty$.

Im weiteren Verlauf behandeln wir die Verteilungskonvergenz gegen die Normalverteilung. Da die Normalverteilung eine stetige Verteilungsfunktion (sogar mit einer Dichte) besitzt, gilt die folgende Verschärfung:

1.5.7 Bemerkung: (Ω, \mathcal{A}, P) und $(\Omega_1, \mathcal{A}_1, P_1)$ seien Wahrscheinlichkeitsräume, die Zufallsvariable $Y : \Omega_1 \rightarrow \mathbb{R}$ besitze eine stetige Verteilungsfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$, und es sei $(X_n)_{n=1}^\infty$ eine Folge von Zufallsvariablen auf Ω mit den Verteilungsfunktionen F_n . Es gelte $X_n \xrightarrow{d} Y$ bei $n \rightarrow \infty$. Dann gilt $\|F_n - F\|_\infty \rightarrow 0$ bei $n \rightarrow \infty$.

Beweis: Wir fixieren ein $\varepsilon > 0$ und wählen ein $r \in \mathbb{N}$ mit $r \cdot \varepsilon \leq 1$. Da F eine stetige Verteilungsfunktion ist, existieren $t_1 < \dots < t_{r-1}$ mit $F(t_j) = \frac{j}{r}$ für alle $j = 1, \dots, r-1$. Wir setzen $t_0 = -\infty$, $t_r = \infty$, $F(t_0) = 0$, und $F(t_r) = 1$. Es sei $m \in \mathbb{N}$ mit $|F_n(t_j) - F(t_j)| \leq \frac{1}{2r}$ für alle $j = 1, \dots, r-1$ und alle $n \geq m$. Es sei jetzt $t \in \mathbb{R}$ und $1 \leq j \leq r$ mit $t_{j-1} \leq t \leq t_j$. Im Fall $F_n(t) \geq F(t)$ folgt

$$0 \leq F_n(t) - F(t) \leq F_n(t_j) - F(t) \leq F_n(t_j) + \frac{1}{2r} - F(t_j) \leq \frac{1}{r} \leq \varepsilon.$$

Im Fall $F_n(t) \leq F(t)$ folgt entsprechend $0 \leq F(t) - F_n(t) \leq \varepsilon$. Damit erhalten wir die verlangte Ungleichung $\|F_n - F\|_\infty \leq \varepsilon$ für alle $n \in \mathbb{N}$ mit $n \geq m$. □

Der folgende Satz ist die wohl wichtigste Aussage der gesamten Wahrscheinlichkeitstheorie; er besagt, dass sich sehr viele Verteilungen durch die Normalverteilung approximieren lassen. Mit Hilfe dieser Approximation können dann sehr viele kompliziertere Problemstellungen relativ einfach mit großer Genauigkeit bestimmt werden. In der Literatur finden sich sehr viele unterschiedliche Beweise. Der hier durchgeführte recht elementare Beweis benutzt wesentlich Ideen der Darstellung aus dem Buch von Georgi. Erste Anwendungen zeigen wir im Anschluss an den Beweis.

1.5.8 Satz: (Zentraler Grenzwertsatz) $(X_n)_{n=1}^\infty$ sei eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit $0 < \sigma^2 = V(X_1) < \infty$. Weiter sei $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ die Standard-Normalverteilung. Für jedes $n \in \mathbb{N}$ setzen wir

$$S_n^* = \frac{1}{\sigma\sqrt{n}}(X_1 + \dots + X_n - nE(X_1)).$$

Ist F_n die Verteilungsfunktion von S_n^* , so gilt $\|F_n - \Phi\|_\infty \rightarrow 0$ bei $n \rightarrow \infty$.

Beweis: Wir können offenbar $E(X_n) = 0$ und $V(X_n) = 1$ annehmen für alle $n \in \mathbb{N}$, ansonsten ersetzen wir X_n durch $X_n - E(X_1)$ und dann durch $\frac{1}{\sqrt{V(X_1)}}X_n$. Wegen 1.5.1 können wir durch Produktbildung erreichen, dass eine Folge unabhängiger standard-normalverteilter Zufallsvariablen $(Y_n)_{n=1}^\infty$ auf Ω existiert, die zu allen X_n unabhängig ist. Wir setzen jetzt

$$T_n = \sum_{k=1}^n \frac{1}{\sqrt{n}} X_k.$$

Wegen 1.4.24 ist T_n standard-normalverteilt. Wir zeigen jetzt, dass die Verteilungsfunktion von S_n^* punktweise auf \mathbb{R} gegen die Verteilungsfunktion von T_n , also gegen die Standard-Normalverteilung konvergiert; die Aussage folgt dann wegen der Stetigkeit der Normalverteilung aus 1.5.7. Der Beweis zerfällt in mehrere Schritte.

(I): Es seien $\tau, \tau' \in \mathbb{R}$ mit $\tau < \tau'$, und $f : \mathbb{R} \rightarrow [0, 1]$ zweimal stetig differenzierbar mit $f(t) = 0$ für alle $t \geq \tau'$ und mit $f(t) = 1$ für alle $t \leq \tau$. Weiter seien $X, Y, Z \in \mathcal{L}^2(P)$ unabhängige Zufallsvariablen mit $E(X) = E(Y) = E(Z) = 0$ und mit $V(X) = V(Y) = 1$. Für jedes $\delta > 0$ sei $N(\delta) = \sup \{|f''(s) - f''(t)| \mid |s - t| \leq \delta\}$ der Stetigkeitsmodul von f'' . Für $n \in \mathbb{N}$ gilt

$$\left| E\left(f \circ \left(\frac{1}{\sqrt{n}}X + Z\right)\right) - E\left(f \circ \left(\frac{1}{\sqrt{n}}Y + Z\right)\right) \right| \leq \frac{N(\delta)}{n} + \frac{\|f''\|_\infty}{n} \left(E(X^2 \cdot \mathbb{1}_{X^2 > \delta n}) + E(Y^2 \cdot \mathbb{1}_{Y^2 > \delta n}) \right).$$

Beweis von (I): Wir setzen $U_n = \frac{1}{\sqrt{n}} X$, $W_n = \frac{1}{\sqrt{n}} Y$. Wegen des Satzes von Taylor existiert punktweise auf Ω ein $0 < \vartheta_U < 1$ mit

$$\begin{aligned} f \circ (Z + U_n) &= f(Z) + f'(Z)U_n + \frac{1}{2}f''(Z + \vartheta_U U_n)U_n^2 \\ &= f(Z) + f'(Z)U_n + \frac{1}{2}f''(Z)U_n^2 + \frac{1}{2}\left(f''(Z + \vartheta_U U_n) - f''(Z)\right)U_n^2. \end{aligned}$$

Es gilt weiter die folgende Abschätzung

$$\left| \frac{1}{2}\left(f''(Z + \vartheta_U U_n) - f''(Z)\right)U_n^2 \right| \leq \frac{1}{2}U_n^2\left(N(\delta) + 2\|f''\|_\infty \mathbb{1}_{U_n^2 > \delta}\right).$$

Eine entsprechende Entwicklung gilt auch für $f \circ (Z + W_n)$. Damit erhalten wir aus $E(U_n) = E(W_n) = 0$ und da die Zufallsvariablen $f'(Z)$, U_n , W_n unabhängig sind, wegen 1.4.11 unmittelbar $E(f'(Z)U_n) = E(f'(Z))E(U_n) = 0$ und entsprechend $E(f'(Z)W_n) = 0$. Daher gilt

$$E(f''(Z)(U_n - W_n)) = E(f''(Z))(E(U_n) - E(W_n)) = 0.$$

Mit Hilfe der hergeleiteten Taylor-Darstellung erhalten wir daher

$$\begin{aligned} &|E(f \circ (Z + U_n)) - E(f \circ (Z + W_n))| \\ &= \frac{1}{2} \left| E\left(\left(f''(Z + \vartheta_U U_n) - f''(Z)\right)U_n^2 - \left(f''(Z + \vartheta_W W_n) - f''(Z)\right)W_n^2\right) \right| \\ &\leq \frac{1}{2}E\left(\left|f''(Z + \vartheta_U U_n) - f''(Z)\right|U_n^2\right) + \frac{1}{2}E\left(\left|f''(Z + \vartheta_W W_n) - f''(Z)\right|W_n^2\right) \\ &\leq \frac{1}{2}N(\delta)E(U_n^2) + \frac{1}{2}N(\delta)E(W_n^2) + \|f''\|_\infty E(U_n^2 \cdot \mathbb{1}_{U_n^2 > \delta}) + \|f''\|_\infty E(W_n^2 \cdot \mathbb{1}_{W_n^2 > \delta}) \\ &= \frac{1}{2n}N(\delta) + \frac{1}{2n}N(\delta) + \frac{\|f''\|_\infty}{n}E(X \cdot \mathbb{1}_{X^2 > n\delta}) + \frac{\|f''\|_\infty}{n}E(Y \cdot \mathbb{1}_{Y^2 > n\delta}). \end{aligned}$$

(II): Es seien $\tau, \tau' \in \mathbb{R}$ mit $\tau < \tau'$, und $f : \mathbb{R} \rightarrow [0, 1]$ zweimal stetig differenzierbar mit $f(t) = 0$ für alle $t \geq \tau'$ und mit $f(t) = 1$ für alle $t \leq \tau$. Dann gilt

$$E(f(S_n^*) - f(T_n)) \rightarrow 0 \text{ bei } n \rightarrow \infty.$$

Beweis von (II): Zunächst sei $n \in \mathbb{N}$ fixiert. Dann gilt

$$\begin{aligned} \left| E(f(S_n^*) - f(T_n)) \right| &= \left| E\left(f\left(\sum_{k=1}^n \frac{1}{\sqrt{n}} X_k\right) - f\left(\sum_{k=1}^n \frac{1}{\sqrt{n}} Y_k\right)\right) \right| \\ &= \left| E\left(f\left(\sum_{k=1}^n \frac{1}{\sqrt{n}} X_k\right) - f\left(\sum_{k=1}^{n-1} \frac{1}{\sqrt{n}} X_k + \frac{1}{\sqrt{n}} Y_n\right) \right. \right. \\ &\quad \left. \left. + f\left(\sum_{k=1}^{n-1} \frac{1}{\sqrt{n}} X_k + \frac{1}{\sqrt{n}} Y_n\right) - f\left(\sum_{k=1}^{n-2} \frac{1}{\sqrt{n}} X_k + \sum_{k=n-1}^n \frac{1}{\sqrt{n}} Y_k\right) + \dots \right. \right. \\ &\quad \left. \left. + f\left(\sum_{k=1}^1 \frac{1}{\sqrt{n}} X_k + \sum_{k=2}^n \frac{1}{\sqrt{n}} Y_k\right) - f\left(\sum_{k=1}^n \frac{1}{\sqrt{n}} Y_k\right)\right) \right| \\ &\leq \sum_{j=0}^{n-1} \left| E\left(f\left(\frac{1}{\sqrt{n}} \left[\sum_{k=1}^{n-j} X_k + \sum_{k=n-j+1}^n Y_k\right]\right) - f\left(\frac{1}{\sqrt{n}} \left[\sum_{k=1}^{n-j-1} X_k + \sum_{k=n-j}^n Y_k\right]\right)\right) \right|. \end{aligned}$$

Wir schätzen jetzt den j -ten Summanden ab mit Hilfe von (I). Dazu setzen wir

$$Z = \frac{1}{\sqrt{n}} \left[\sum_{k=1}^{n-j-1} X_k + \sum_{k=n-j+1}^n Y_k \right], \quad X = X_{n-j}, \quad \text{und} \quad Y = Y_{n-1} .$$

Da die Zufallsvariablen X_1, \dots, X_n und Y_1, \dots, Y_n jeweils gleichverteilt sind, erhalten wir als Abschätzung

$$\frac{N(\delta)}{n} + \frac{\|f''\|_\infty}{n} \left(E(X_1^2 \cdot \mathbb{1}_{X_1^2 > \delta n}) + E(Y_1^2 \cdot \mathbb{1}_{Y_1^2 > \delta n}) \right)$$

für jeden einzelnen Summanden. Insgesamt ergibt sich

$$\left| E(f(S_n^*) - f(T_n)) \right| \leq N(\delta) + \|f''\|_\infty \left(E(X_1^2 \cdot \mathbb{1}_{X_1^2 > \delta n}) + E(Y_1^2 \cdot \mathbb{1}_{Y_1^2 > \delta n}) \right) =: b_n .$$

Aus $X_1 \in \mathcal{L}^2$ folgt $X_1^2 \cdot \mathbb{1}_{X_1^2 > \delta n} \downarrow 0$ bei $n \rightarrow \infty$. Wegen des Konvergenzsatzes von Lebesgue gilt dann auch $E(X_1^2 \cdot \mathbb{1}_{X_1^2 > \delta n}) \rightarrow 0$ und entsprechend $E(Y_1^2 \cdot \mathbb{1}_{Y_1^2 > \delta n}) \rightarrow 0$ bei $n \rightarrow \infty$. Es folgt $b_n \rightarrow N(\delta)$ bei $n \rightarrow \infty$. Wegen $N(\delta) \downarrow 0$ bei $\delta \downarrow 0$ erhalten wir

$$\left| E(f(S_n^*) - f(T_n)) \right| \rightarrow 0 \text{ bei } \delta \downarrow 0 .$$

(III): Beweis des Satzes: F_n sei die Verteilungsfunktion der Zufallsvariablen S_n^* . Für alle $t \in \mathbb{R}$ gilt dann $F(t) = E(\mathbb{1}_{]-\infty, t]} \circ S_n^*)$. Wir fixieren jetzt ein $t \in \mathbb{R}$ beliebig und wählen dann $t_1, t_2 \in \mathbb{R}$ mit $t_1 < t < t_2$.

Wir setzen jetzt $\tau = t$ und $\tau' = t_2$ und wählen f gemäß (II) mit $\mathbb{1}_{]-\infty, \tau]} \leq f \leq \mathbb{1}_{]-\infty, \tau']}$. Wegen (II) gilt dann

$$\begin{aligned} \limsup_{n \rightarrow \infty} F_n(t) &= \limsup_{n \rightarrow \infty} E(\mathbb{1}_{]-\infty, \tau]} \circ S_n^*) \leq \limsup_{n \rightarrow \infty} E(f \circ S_n^*) \\ &= \limsup_{n \rightarrow \infty} E(f \circ T_n) \leq E(\mathbb{1}_{]-\infty, \tau']}) = \Phi(\tau') = \Phi(t_2) . \end{aligned}$$

Wir setzen jetzt $\tau = t_1$ und $\tau' = t$ und wählen f gemäß (II) mit $\mathbb{1}_{]-\infty, \tau]} \leq f \leq \mathbb{1}_{]-\infty, \tau']}$. Dann gilt

$$\begin{aligned} \liminf_{n \rightarrow \infty} F_n(t) &= \liminf_{n \rightarrow \infty} E(\mathbb{1}_{]-\infty, \tau']}) \geq \liminf_{n \rightarrow \infty} E(f \circ S_n^*) \\ &= \liminf_{n \rightarrow \infty} E(f \circ T_n) \geq E(\mathbb{1}_{]-\infty, \tau]}) = \Phi(\tau) = \Phi(t_1) . \end{aligned}$$

Mit den Grenzübergängen $t_1 \uparrow t$ und $t_2 \downarrow t$ folgt dann wegen der Stetigkeit von Φ

$$\Phi(t) \geq \limsup_{n \rightarrow \infty} F_n(t) \geq \liminf_{n \rightarrow \infty} F_n(t) \geq \Phi(t) ,$$

was zu zeigen war. □

Der zentrale Grenzwertsatz gestattet eine gute Approximation vieler Verteilungen, die durch unabhängige Wiederholungen des gleichen Experimentes entstanden sind. Als erstes Beispiel diskutieren wir die Approximation der Binomialverteilung. Entsprechend können auch andere Verteilungen behandelt werden; die unten geschilderte Vorgehensweise wird dabei nicht verändert.

Approximation der Binomialverteilung:

Der Zentrale Grenzwertsatz 1.5.8 gestattet eine schon für recht kleine n eine relative gute Approximation der Binomialverteilung: Es liege also jetzt eine Binomialverteilung mit dem Parameter p vom Umfang $n \in \mathbb{N}$ vor. Mit den Bezeichnungen des zentralen Grenzwertsatzes und den Zufallsvariablen X_1, \dots, X_n , die die Gewinne der einzelnen Bernoulli-Experimente zählen, gilt

$$\|S_n^* - \Phi\|_\infty \rightarrow 0 \text{ bei } n \rightarrow \infty.$$

Wegen 1.4.15 gilt $V(X_1) = p(1-p)$, $E(X_1) = p$, es folgt $\sigma = \sqrt{p(1-p)}$. Damit erhalten wir

$$S_n^* = \frac{1}{\sqrt{np(1-p)}}(X_1 + \dots + X_n - n \cdot p).$$

Interessieren wir uns jetzt speziell für die Wahrscheinlichkeit, dass die Zahl der Erfolg k zwischen zwei Grenzen k_a und k_b liegt mit $k_a < k_b$, so gilt

$$P([k_a \leq X_1 + \dots + X_n \leq k_b]) = \sum_{k=k_a}^{k_b} \binom{n}{k} p^k (1-p)^{n-k} \simeq (\Phi(b) - \Phi(a))$$

mit $a = \frac{1}{\sqrt{np(1-p)}}(k_a - np)$ $b = \frac{1}{\sqrt{np(1-p)}}(k_b - np)$.

Die Werte $\Phi(a)$, $\Phi(b)$ können dabei der Tabelle für die Normalverteilung entnommen werden.

Eine Konvergenzverbesserung erhält man noch durch die folgende Überlegung. Dazu interessieren wir uns jetzt für die Wahrscheinlichkeit

$$P([X_1 + \dots + X_n \geq k_0])$$

für ein $0 < k_0 < n$. Die Wahrscheinlichkeit des Komplementärereignisses ergibt sich entsprechend:

$$P([X_1 + \dots + X_n < k_0]) = P([X_1 + \dots + X_n \leq k_0 - 1]).$$

Formal liegt hier also eine gewisse Asymmetrie vor. Um diese zu beheben betrachten wir $k_0 - \frac{1}{2}$ an Stelle von k_0 . Bei der Binomialverteilung ergibt sich damit keine Veränderung, wohl aber bei der Approximation durch die Normalverteilung: Diese Veränderung liefert eine wesentliche Konvergenzverbesserung, sollte also vor allen Dingen bei kleineren n -Werten benutzt werden. Für je sehr große n ist sie hingegen bedeutungslos.

1.5.9 Beispiel: Es seien $p = \frac{1}{2}$, $n = 20$, $k_a = 7$, $k_b = 11$; Wir berechnen $P([k_a \leq k \leq k_b])$:

$$\text{Es gilt } P([k_a \leq k \leq k_b]) = \sum_{k=7}^{11} \binom{20}{k} \left(\frac{1}{2}\right)^{20} = 0,69062$$

Berechnung ohne Korrekturterm:

$$\text{Wir setzen } a = \frac{k_a - p \cdot n}{\sqrt{np(1-p)}} = \frac{-3}{\sqrt{5}} = -1,3416 \text{ und } b = \frac{k_b - p \cdot n}{\sqrt{np(1-p)}} = \frac{1}{\sqrt{5}} = 0,4472 :$$

Aus der Tabelle erhalten wir

$$\Phi(a) - \Phi(b) = 0,6727 - 1 + 0,9101 = \underline{0,5828}$$

Berechnung mit Korrekturterm:

Wir setzen $\tilde{a} = \frac{k_a - \frac{1}{2} - pn}{\sqrt{np(1-p)}} = \frac{-3,5}{\sqrt{5}} = -1,5652$ und $\tilde{b} = \frac{k_b + \frac{1}{2} - pn}{\sqrt{np(1-p)}} = \frac{1,5}{\sqrt{5}} = 0,6708$.

Aus der Tabelle erhalten wir

$$\Phi(\tilde{b}) - \Phi(\tilde{a}) = 0,7489 - 1 + 0,9412 = \underline{0,6901}.$$

Ein Vergleich zeigt, dass unter Berücksichtigung dieser Korrektur schon bei diesem kleinen n eine gute Approximation der Wahrscheinlichkeiten vorliegt. Ohne Korrektur ergibt sich allerdings ein wesentlich größerer Unterschied.

1.6 Mit der Normalverteilung zusammenhängende Verteilungen

Der vorstehende Abschnitt hat gezeigt, dass die Normalverteilung in sehr vielen Situationen wahrscheinlichkeitstheoretische Probleme wenigstens angenähert beschreibt. Aus diesem Grund sollen weitere Verteilungen, die mit der Normalverteilung in engem Zusammenhang stehen, behandelt werden. Benötigt werden diese Verteilungen später beispielsweise für die Überprüfung ob eine konkrete Situation durch eine Normalverteilung beschrieben werden kann. Um die Darstellung übersichtlich gestalten zu können führen wir dazu eine vektorwertige Variante der Normalverteilung ein – die so genannte multivariante Normalverteilung.

In diesem Abschnitt sei wieder (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Zunächst notieren wir als eine einfache Konsequenz des Transformationssatzes 0.5.2 für das n -dimensionale Lebesgue-Integral die folgende Aussage über die **multivariante Normalverteilung**, als vektorwertige Variante der Normalverteilung.

1.6.1 Lemma: $\emptyset \neq U, W \subset \mathbb{R}^n$ seien offen, $\psi : U \rightarrow W$ sei bijektiv, stetig differenzierbar mit differenzierbarer Umkehrabbildung $\vartheta = \psi^{-1} : W \rightarrow U$. Es sei $X : \Omega \rightarrow \mathbb{R}^n$ ein Zufallsvektor mit $X(\omega) \in U$ für alle $\omega \in \Omega$. Ferner besitze die gemeinsame Verteilung F_X von X eine Dichte f_X . Dann besitzt die gemeinsame Verteilung F_Y des Zufallsvektors $Y = \psi \circ X$ eine Dichte f_Y mit

$$f_Y(y) = f_X(\vartheta(y)) \left| \det(D\vartheta(y)) \right| \text{ für alle } y \in W.$$

Beweis: Wir bezeichnen wieder mit P_X das Verteilungsmaß von X , definiert durch $P_X(A) = P(X^{-1}(A))$ für alle $A \in \mathcal{B}^n$ mit $A \subset W$ und entsprechend P_Y . Dann gilt für alle $A \in \mathcal{B}^n$ mit $A \subset W$ die Beziehung

$$\begin{aligned} P_Y(A) &= P(Y^{-1}(A)) = P\left((\psi \circ X)^{-1}(A)\right) = P\left(X^{-1}(\vartheta(A))\right) = P_X(\vartheta(A)) \\ &= \int_{\vartheta(A)} f_X(x) d\lambda^n(x) = \int_A f_X(\vartheta(y)) \left| \det(D\vartheta(y)) \right| d\lambda^n(y), \end{aligned}$$

was zu zeigen war. □

1.6.2 Satz: Es sei $X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}$ ein Zufallsvektor, so dass X_1, \dots, X_n unabhängige standard normalverteilte Zufallsvariablen sind.

i) Die gemeinsame Verteilung von X besitzt die Dichte

$$\varphi_{0,I}(x) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}x^T x\right) \text{ für alle } x \in \mathbb{R}^n.$$

ii) Es seien $B \in \mathbb{R}^{n \times n}$ invertierbar und $m \in \mathbb{R}^n$. Setzen wir $C = BB^T$, so besitzt der Zufallsvektor $Y = BX + m$ die Verteilungsdichte

$$\varphi_{m,C}(y) = (2\pi)^{-\frac{n}{2}} \left| \det(C) \right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y-m)^T (C)^{-1}(y-m)\right)$$

für alle $y \in \mathbb{R}^n$. Für die Koordinaten-Funktionen Y_1, \dots, Y_n von Y gilt $E(Y_j) = m_j$ und $\text{Cov}(Y_k, Y_j) = (BB^T)_{k,j} = c_{k,j}$.

Anmerkung: Das Wahrscheinlichkeitsmaß $N_n(m, C)$ auf $(\mathbb{R}^n, \mathcal{B}^n)$ mit der Dichtefunktion $\varphi_{m,C}$ heißt die n -dimensionale oder **multivariate Normalverteilung** mit dem Erwartungswertvektor m und der Kovarianzmatrix C falls $C \in \mathbb{R}^{n \times n}$ eine positiv definite symmetrische Matrix ist; diese wird auch als **Kovarianzmatrix** bezeichnet.

Beweis: i) Da die Zufallsvariablen X_1, \dots, X_n unabhängig sind, besitzt X wegen 1.4.21 die gemeinsame Verteilung mit der Produktdichte

$$\varphi_{0,I}(x) := \varphi_{0,1}(x_1) \cdot \dots \cdot \varphi_{0,1}(x_n) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}x^T x\right)$$

für alle $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$.

ii) Wir setzen $\psi(x) = Bx + m$ und $\vartheta(y) = B^{-1}(y - m)$; dann gilt offenbar $X = \vartheta \circ Y$. Wegen 1.6.1 gilt

$$f_Y(y) = \varphi_{0,I}(\vartheta(y)) \left| \det(D\vartheta(y)) \right| = (2\pi)^{-\frac{n}{2}} |\det(C)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - m)^T (C)^{-1}(y - m)\right)$$

für alle $y \in \mathbb{R}^n$. Man beachte dabei $\det(BB^T) = (\det(B))^2 = \det(C)$. Weiter erhalten wir

$$E(Y_i) = E\left(\sum_{j=1}^n b_{i,j} X_j + m_i\right) = \sum_{j=1}^n b_{i,j} E(X_j) + m_i = m_i,$$

und wegen der Unabhängigkeit der Zufallsvariablen X_1, \dots, X_n folgt

$$\text{Cov}(Y_i, Y_j) = \text{Cov}\left(\sum_{k=1}^n b_{i,k} X_k, \sum_{l=1}^n b_{j,l} X_l\right) = \sum_{k,l=1}^n b_{i,k} b_{j,l} \text{Cov}(X_k, X_l) = \sum_{k=1}^n b_{i,k} b_{j,k} = c_{i,j}. \quad \square$$

Als unmittelbare Konsequenz von Aussage ii) des vorstehenden Satzes erhalten wir die folgende Aussage für eine orthogonale Matrix U :

1.6.3 Folgerung: $U \in \mathbb{R}^{n \times n}$ sei eine orthogonale Matrix. X sei $N_n(0, I)$ verteilt, und es sei $Y = UX$. Dann ist Y ebenfalls $N_n(0, I)$ verteilt ist.

1.6.4 Satz: $C \in \mathbb{R}^{n \times n}$ sei symmetrisch und positiv definit. Weiter sei der Zufallsvektor X $N_n(m, C)$ -verteilt für ein $m \in \mathbb{R}^n$. Für $A \in \mathbb{R}^{k \times n}$ mit $\text{Rang}(A) = k \leq n$ und $a \in \mathbb{R}^k$ sei $Z = AX + a$. Dann ist Z eine $N_k(Am + a, ACA^T)$ - verteilte Zufallsvariable.

Beweis: Wir können offenbar $a = 0$ und $m = 0$ annehmen. Da C positiv definit ist, existiert eine positiv definite symmetrische Matrix B mit $B^2 = C$. Wir setzen $Y = B^{-1}X$. Dann besitzt Y die Verteilung $N_n(0, I)$. Es sei W der von den Zeilenvektoren von $A \cdot B$ erzeugte Untervektorraum. Wir wählen eine Orthonormalbasis $u_1, \dots, u_k \in W \subset \mathbb{R}^n$, ergänzen diese zu einer Orthonormalbasis u^1, \dots, u^n des \mathbb{R}^n , und bilden dann $U = (u_1 : \dots : u_n)^T \in \mathbb{R}^{n \times n}$. Wegen der Konstruktion ist U eine orthogonale Matrix. Wir setzen $R = ABU^T \in \mathbb{R}^{k \times n}$, dann gilt $RU = ABU^T U = AB$. Da die Vektoren u^{k+1}, \dots, u^n orthogonal zu allen Zeilenvektoren der Matrix (AB) sind, gilt $r_{i,j} = 0$ für alle $j > k$ und alle $i = 1, \dots, k$. Wegen der vorstehenden Bemerkung ist der Zufallsvektor $\bar{Y} = UY$ wieder $N_n(0, I)$ -verteilt. Die Koordinaten $\bar{Y}_1, \dots, \bar{Y}_n$ sind folglich $N(0, 1)$ -verteilte unabhängige Zufallsvariable. Daher ist

$$Y \sim (\bar{Y}_1, \dots, \bar{Y}_k)^T$$

$N_k(0, I)$ -verteilt. Weiter besitzt $AX = ABY = RUY = R\bar{Y} = RY \sim$ die Verteilung $N_k(0, RR^T)$. Wegen $RR^T = RUU^T R^T = ACA^T$ ist die Aussage gezeigt worden. □

1.6.5 Satz: X sei eine $N(0, 1)$ -verteilte Zufallsvariable.

i) Die Zufallsvariable X^2 ist $\Gamma_{\frac{1}{2}, \frac{1}{2}}$ -verteilt, besitzt also die Verteilung mit der Dichte

$$\gamma_{\frac{1}{2}, \frac{1}{2}}(t) = \frac{1}{\Gamma(\frac{1}{2})\sqrt{2t}} \exp\left(-\frac{t}{2}\right) \quad \text{für } t > 0.$$

ii) Die Zufallsvariable $Y = \exp(X)$ ist **lognormal-verteilt**, besitzt also die Verteilungsfunktion mit der Dichte $f(t) = \frac{1}{t\sqrt{2\pi}} t^{-\ln(t)/2}$.

Beweis: F sei die Verteilungsfunktion von X^2 mit der Dichte $f = F'$. Dann gilt $F(t) = 0$ für alle $t \leq 0$; und für $t > 0$ erhalten wir

$$F(t) = P(X^2 \leq t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = \Phi_{0,1}(\sqrt{t}) - \Phi_{0,1}(-\sqrt{t}), \text{ also}$$

$$f(t) = F'(t) = \varphi_{0,1}(\sqrt{t}) \cdot \frac{1}{2\sqrt{t}} + \varphi_{0,1}(-\sqrt{t}) \cdot \frac{1}{2\sqrt{t}} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{t}} \exp\left(-\frac{t}{2}\right) = \gamma_{\frac{1}{2}, \frac{1}{2}}(t).$$

ii) Für $t > 0$ gilt $[Y \leq t] = [X \leq \ln(t)]$. Ist F die Verteilungsfunktion von Y , so gilt $F(t) = \Phi_{0,1}(\ln(t))$ und daher

$$F'(t) = \varphi_{0,1}(\ln(t)) \frac{1}{t} = \frac{1}{t\sqrt{2\pi}} \exp\left(-\frac{\ln(t)^2}{2}\right) = \frac{1}{t\sqrt{2\pi}} t^{-\ln(t)/2}.$$

□

Im weiteren Verlauf wird jetzt die **Beta-Verteilung** eingeführt. Für eine mögliche Motivation dieser Verteilung wird auf die Literatur verwiesen, etwa auf das Buch von Georgi, Abschnitt 2.5.3.

1.6.6 Bemerkung: Für $a, b > 0$ bilden wir die Beta-Funktion

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt.$$

Dann gilt $B(a+1, b) = \frac{a}{a+b} B(a, b)$.

Beweis: Wir erhalten mit Hilfe partieller Integration

$$\begin{aligned} a(B(a, b) - B(a+1, b)) &= \int_0^1 at^{a-1}(1-t)^{b-1} dt - \int_0^1 at^a(1-t)^{b-1} dt \\ &= a \int_0^1 (t^{a-1} - t^a)(1-t)^{b-1} dt = \int_0^1 at^{a-1}(1-t)^b ds \\ &= 0 + \int_0^1 t^a b(1-t)^{b-1} dt = bB(a+1, b) \end{aligned}$$

und damit die Behauptung. □

Es seien jetzt $a, b > 0$. Wir bezeichnen das Wahrscheinlichkeitsmaß $B_{a,b}$ auf $]0, 1[$ mit der Dichte

$$\beta_{a,b}(s) = \frac{1}{B(a,b)} s^{a-1}(1-s)^{b-1}$$

für $0 < s < 1$ als Beta-Verteilung zu a, b .

1.6.7 Satz: *Es seien $\alpha, r, k > 0$. Die unabhängigen Zufallsvariablen X_1 und X_2 seien $\Gamma_{\alpha, r}$ beziehungsweise $\Gamma_{\alpha, k}$ -verteilt. Es gelten die folgenden Aussagen:*

$X_1 + X_2$ ist $\Gamma_{\alpha, r+k}$ -verteilt.

$\frac{X_1}{X_1 + X_2}$ ist Beta-verteilt mit der Verteilung-Dichte $\beta_{r, k}$.

$X_1 + X_2$ und $\frac{X_1}{X_1 + X_2}$ sind unabhängig.

Beweis: Der Zufallsvektor $X = (X_1, X_2)$ besitzt die gemeinsame Dichte

$$f_X(s, t) = \gamma_{\alpha, r}(s)\gamma_{\alpha, k}(t) = \frac{\alpha^{r+k}}{\Gamma(r)\Gamma(k)} s^{r-1} t^{k-1} e^{-\alpha(s+t)}$$

für $s, t > 0$. Wir definieren $\Psi :]0, \infty[^2 \rightarrow]0, \infty[\times]0, 1[$ durch $\Psi(s, t) = (s + t, \frac{s}{s+t})^T$. Offenbar ist Ψ bijektiv mit der Umkehrabbildung

$$\vartheta(u, v) = \Psi^{-1}(u, v) = (uv, u(1-v))^T.$$

Diese Umkehrabbildung besitzt die Ableitungsmatrix

$$D\vartheta(u, v) = \begin{pmatrix} v & u \\ 1-v & -u \end{pmatrix}.$$

Wegen $|\det D\vartheta(u, v)| = u$ folgt aus der Folgerung des Transformationssatzes 1.6.1, dass der Zufallsvektor $Y = \Psi(X_1, X_2) = (X_1 + X_2, \frac{X_1}{X_1 + X_2})$ die Dichte

$$\begin{aligned} f_Y(u, v) &= f_X(uv, u(1-v)) \cdot u \\ &= \frac{\alpha^{r+k}}{\Gamma(r)\Gamma(k)} u^{r+k-1} e^{-\alpha u} v^{r-1} (1-v)^{k-1} \\ &= \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k)} B(r, k) \gamma_{\alpha, r+k}(u) \beta_{r, k}(v) \end{aligned}$$

für $(u, v) \in W =]0, \infty[\times]0, 1[$ besitzt. Integration über W liefert dann

$$\begin{aligned} 1 &= \int_W f_Y(u, v) \, dudv = \int_0^\infty \left(\int_0^1 \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k)} B(r, k) \beta_{r, k}(v) \, dv \right) \gamma_{\alpha, r+k}(u) \, du \\ &= \int_0^\infty \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k)} B(r, k) \gamma_{\alpha, r+k}(u) \, du = \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k)} B(r, k). \end{aligned}$$

Wir erhalten also $f_Y(u, v) = \gamma_{\alpha, r+k}(u) \cdot \beta_{r, k}(v)$ für alle $(u, v) \in W$. Damit sind die Aussagen wegen 1.4.21 bewiesen worden. □

Als eine unmittelbare Konsequenz ergibt sich die folgende Aussage

1.6.8 Folgerung: X_1, \dots, X_n seien unabhängige standard-normalverteilte Zufallsvariable. Dann besitzt die Zufallsvariable $Y = X_1^2 + \dots + X_n^2$ die Verteilung $f_Y = \Gamma_{\frac{1}{2}, \frac{n}{2}}$. Wir sprechen in diesem Fall von der **Chiquadrat-Verteilung** χ_n^2 mit n Freiheitsgraden:

$$\chi_n^2(t) = \gamma_{\frac{1}{2}, \frac{n}{2}}(t) = \frac{t^{\frac{n}{2}-1}}{\Gamma(\frac{1}{2})2^{\frac{n}{2}}} e^{-\frac{n}{2}t} \quad \text{für } t > 0.$$

1.6.9 Folgerung: $X_1, \dots, X_n, Y_1, \dots, Y_m$ seien unabhängige standard-normalverteilte Zufallsvariable. Dann besitzt die Zufallsvariable

$$Y = \frac{m}{n} \frac{X_1^2 + \dots + X_n^2}{Y_1^2 + \dots + Y_m^2}$$

die **Fisher-Verteilung** $F_{n,m}$ auf $]0, \infty[$ mit n und m Freiheitsgraden mit der Dichte

$$f_{n,m}(t) = \frac{n^{\frac{n}{2}} m^{\frac{m}{2}}}{B(\frac{n}{2}, \frac{m}{2})} \frac{t^{\frac{n}{2}-1}}{(m+nt)^{\frac{m+n}{2}}} \quad \text{für } t > 0.$$

Beweis: Wegen des vorstehenden Korollars besitzt $X = X_1^2 + \dots + X_n^2$ eine χ_n^2 -Verteilung und $Y = Y_1^2 + \dots + Y_m^2$ eine χ_m^2 -Verteilung. Weiter sind X und Y unabhängig. Daher besitzt $Z = \frac{X}{X+Y}$ wegen Satz 1.6.7 die Beta-Verteilung mit der Dichte $\beta_{\frac{n}{2}, \frac{m}{2}}$. Wir betrachten jetzt die bijektive Abbildung $\psi :]0, 1[\rightarrow]0, \infty[$, definiert durch $\psi(t) = \frac{n}{m} \frac{t}{1-t}$ mit der Umkehrabbildung $\psi^{-1}(x) = \frac{mx}{n+mx}$ für $x \in]0, \infty[$. Wegen

$$F_{n,m} = \frac{m}{n} \frac{X}{Y} = \frac{m}{n} \frac{Z}{1-Z} = \psi(Z)$$

und Lemma 1.6.1 besitzt $F_{n,m}$ die Verteilungsdichte

$$f_{n,m}(x) = \beta_{\frac{n}{2}, \frac{m}{2}} \left(\frac{mx}{n+mx} \right) \frac{mn}{(n+mx)^2} = \frac{n^{\frac{n}{2}} m^{\frac{m}{2}}}{B(\frac{n}{2}, \frac{m}{2})} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{m+n}{2}}} \quad \text{für } x > 0. \quad \square$$

1.6.10 Folgerung: X, Y_1, \dots, Y_n seien unabhängige standard-normal verteilte Zufallsvariable. Dann besitzt die Zufallsvariable

$$T = \frac{X}{\sqrt{\frac{1}{n}(Y_1^2 + \dots + Y_n^2)}}$$

die **Student t -Verteilung** mit n Freiheitsgraden mit der Dichtefunktion τ_n :

$$\tau_n(x) = \frac{\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}}{B(\frac{1}{2}, \frac{n}{2}) \sqrt{n}} \quad \text{für } x \in \mathbb{R}.$$

Beweis: T^2 besitzt die Verteilung $F_{1,n}$. Daher besitzt $|T| = \sqrt{T^2}$ eine Verteilung mit der Dichtefunktion $f_{1,n}(y^2) \cdot 2y$ für $y > 0$. Da T symmetrisch ist, besitzen T und $-T$ die gleiche Verteilung, also die Verteilung mit der Dichte

$$\tau_n(y) = f_{1,n}(y^2) \cdot |y|. \quad \square$$

Wir betrachten jetzt das Gaußsche Produktmodell

$$(\mathbb{R}, \mathcal{B}^n, N^{\otimes n}(m, v) \mid m \in \mathbb{R}, \quad v > 0)$$

und unabhängige $N(m, v)$ -verteilte Zufallsvariable X_1, \dots, X_n und dann die bekannten Schätzer der Statistik

$$M = \frac{1}{n}(X_1 + \dots + X_n), \quad V^* = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2.$$

1.6.11 Satz: *Unter den obigen Voraussetzungen sei $\vartheta = (m, v) \in \mathbb{R} \times]0, \infty[$. Bezüglich $P_\vartheta = N_{m,v}^{\otimes n}$ gilt*

- i) M und V^* sind unabhängig.
- ii) M ist $N(m, \frac{v}{n})$ -verteilt und $\frac{n-1}{v}V^*$ ist χ_{n-1}^2 -verteilt.
- iii) $T_m := \frac{\sqrt{n}(M-n)}{\sqrt{V^*}}$ ist τ_{n-1} -verteilt.

Beweis: Es sei $X = (X_1, \dots, X_n)^t : \mathbb{R}^n \rightarrow \mathbb{R}^n$, und $e = (1, \dots, 1)^t \in \mathbb{R}^n$. Wir betrachten eine beliebige orthogonale Matrix $U \in \mathbb{R}^{n \times n}$ mit $u_{1,j} = \frac{1}{\sqrt{n}}$ für alle $j = 1, \dots, n$. Wir setzen dann $Y = UX$. Es sei $Y = (Y_1, \dots, Y_n)^t$. Dann hat Y unter $N_n(me, vE) = N(m, nv)^{\otimes n}$ die Verteilung $N_n(mUe, vE) = N(m\sqrt{n}, v) \otimes N(0, v)^{\otimes n-1}$; dazu beachte man $mUe = (m\sqrt{n}, 0, \dots, 0)^t$. Da Y_1, \dots, Y_n unabhängig sind gilt

$$M = \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{1}{\sqrt{n}} X_j = \frac{1}{\sqrt{n}} Y_1.$$

Wegen $|Y| = |X|$ folgt

$$(n-1)V^* = \sum_{j=1}^n (X_j - M)^2 = \sum_{j=1}^n X_j^2 - nM^2 = Y^2 - Y_1^2 = \sum_{j=2}^n Y_j^2.$$

Die Aussagen folgen jetzt aus den vorstehenden Resultaten. □

2. Grundbegriffe der Statistik

Eine grundlegende Aufgabenstellung der Statistik ist es mit Hilfe von zufallsbedingten Beobachtungen auf zugrunde liegende Gesetzmäßigkeiten zu schließen. Im Regelfall werden die in Frage kommenden Gesetzmäßigkeiten durch bestimmte Wahrscheinlichkeitsmaße beschrieben. In vielen Fällen betrachtet man dabei Wahrscheinlichkeitsmaße die mit Normalverteilungen $N(\mu, \sigma^2)$ zusammenhängen, wobei μ und σ^2 unbekannt sind und durch geeignete Verfahren zu bestimmen, oder zu schätzen sind.

2.1 Lineare Regression

In diesem kurzen Abschnitt diskutieren wir einige Anfangsgründe der Regressionsanalyse. Als Beispiel diskutieren wir folgende Problemstellung. Es gelte eine theoretische Abhängigkeit einer Messreihe

$$y = g(x)$$

mit einer unbekanntem Funktion g . Gemessen seien die Werte

$$\begin{array}{c|cccc} x & x_1 & x_2 & \dots & x_n \\ \hline y & y_1 & y_2 & \dots & y_n \end{array},$$

die jeweils mit zufälligen Fehlern behaftet sind. Wir behandeln zunächst eine lineare Variante dieser Problemstellung und nehmen dazu für eine derartige Messreihe die Gültigkeit von

$$y = g(x) = \alpha + \beta x$$

mit unbekanntem $\alpha, \beta \in \mathbb{R}$ an und wollen diese Größen $\hat{\alpha}$ und $\hat{\beta}$ geeignet schätzen. Zur Bestimmung dieser so genannten Schätzer betrachten wir die Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$:

$$f(\alpha, \beta) = \sum_{j=1}^n (y_j - (\alpha + \beta x_j))^2$$

als quadratische Fehlerfunktion in α und β . Die Größen $\hat{\alpha}$ und $\hat{\beta}$ werden so bestimmt, dass $f(\hat{\alpha}, \hat{\beta})$ minimal wird. Da die Funktion f als Summe quadratischer Funktionen konvex ist, muss jede Nullstelle der Ableitung von f eine Minimalstelle sein. Es gilt

$$\begin{aligned} D_1 f(\alpha, \beta) &= - \sum_{j=1}^n 2(y_j - (\alpha + \beta x_j)) = 0 \\ D_2 f(\alpha, \beta) &= - \sum_{j=1}^n 2(y_j - (\alpha + \beta x_j))x_j = 0 \end{aligned}$$

Wir erhalten aus diesen Gleichungen

$$\begin{aligned}\sum_{j=1}^n y_j - n\alpha - \beta \sum_{j=1}^n x_j &= 0 \\ \sum_{j=1}^n x_j y_j - \alpha \sum_{j=1}^n x_j - \beta \sum_{j=1}^n x_j^2 &= 0\end{aligned}$$

Mit den Abkürzungen

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum_{j=1}^n x_j, & \bar{y} &= \frac{1}{n} \sum_{j=1}^n y_j, & S_{x,x} &= \sum_{j=1}^n (x_j - \bar{x})^2 \\ S_{y,y} &= \sum_{j=1}^n (y_j - \bar{y})^2, & S_{x,y} &= \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})\end{aligned}$$

ergibt sich weiter durch Ausmultiplikation

$$\begin{aligned}S_{x,x} &= \sum_{j=1}^n (x_j^2 - 2x_j\bar{x} + \bar{x}^2) = \sum_{j=1}^n x_j^2 - n\bar{x}^2 \\ S_{y,y} &= \sum_{j=1}^n (y_j^2 - 2y_j\bar{y} + \bar{y}^2) = \sum_{j=1}^n y_j^2 - n\bar{y}^2 \\ S_{x,y} &= \sum_{j=1}^n (x_j y_j - y_j\bar{x} - \bar{y}x_j + \bar{x}\bar{y}) = \sum_{j=1}^n x_j y_j - n\bar{x}\bar{y}.\end{aligned}$$

Einsetzen dieser Werte in die obige Gleichung ergibt

$$\begin{aligned}0 &= n\bar{y} - n\alpha - n\beta\bar{x} \\ 0 &= \sum_{j=1}^n x_j y_j - \alpha n \sum_{j=1}^n x_j - \beta \sum_{j=1}^n x_j^2 \\ &= S_{x,y} + n\bar{x}\bar{y} - \alpha n\bar{x} - \beta S_{x,x} - n\beta\bar{x}^2 \\ &= S_{x,y} + \bar{x}(n\bar{y} - n\alpha - n\beta\bar{x}) - \beta S_{x,x} = S_{x,y} - \beta S_{x,x}.\end{aligned}$$

Wir erhalten also

$$\hat{\beta} = \frac{S_{x,y}}{S_{x,x}}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Damit haben wir die folgende Bemerkung gezeigt:

2.1.1 Bemerkung: *Unter der Annahme, dass zwischen den Messgrößen x und y eine theoretische Beziehung $y = \alpha + \beta x$ gilt, liefert die Schätzung mit der Methode der kleinsten Quadrate die Schätzer*

$$\hat{\beta} = \frac{S_{x,y}}{S_{x,x}}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

mit $\bar{y} = \frac{1}{n}(y_1 + \dots + y_n)$, $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$,

$$S_{x,x} = x_1^2 + \dots + x_n^2 - n\bar{x}^2, \quad S_{x,y} = (x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}).$$

Die Methode der kleinsten Quadrate kann auch in anderen Situationen angewandt werden. Als Beispiel betrachten wir dazu wieder eine Messreihe für die Werte x und y vom Umfang

n und nehmen an, dass zwischen x und y eine Beziehung $y = \alpha + \beta e^{\gamma x}$ gilt. Wir betrachten wieder eine wie eben gebildete Funktion

$$f(\alpha, \beta, \gamma) = \sum_{j=1}^n (\alpha + \beta e^{\gamma x_j})^2$$

und lösen die Gleichungen

$$D_1 f(\alpha, \beta, \gamma) = 0, \quad D_2 f(\alpha, \beta, \gamma) = 0, \quad D_3 f(\alpha, \beta, \gamma) = 0.$$

Die Lösung $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ ist dann wieder ein Schätzer für α, β, γ .

Wir diskutieren eine Variante der Ausgangssituation und gehen davon aus, dass bei einer gegebenen Datenmenge für die Messwerte nur noch die Größe y mit einem zufälligen Fehler behaftet ist, während die Größe der Werte für x jeweils exakt bestimmt ist; wir betrachten y als Wert der Stochastischen Variablen Y_x . Ist also eine Messreihe für die Werte x und y vom Umfang n gegeben, so nehmen wir die Werte x_1, \dots, x_n als gegebene feste Werte auf mit den zugeordneten Zufallsvariablen Y_1, \dots, Y_n . Wir nehmen weiter eine theoretische Abhängigkeit der Form

$$y = \alpha + \beta x$$

an und treffen folgende Annahmen

- (I) $E(Y_x) = \alpha + \beta x$ und speziell $E(Y_i) = \alpha + \beta x_i$ für $i = 1, \dots, n$.
- (II) Y_1, \dots, Y_n sind unabhängig.
- (III) $V(Y_1) = \dots = V(Y_n) = \sigma^2$.

2.1.2 Bemerkung: Die nach der Methode der kleinsten Quadrate bestimmten Schätzer

$$\hat{\beta} = \frac{\sum_{j=1}^n (x_j - \bar{x}) Y_j}{S_{x,x}}, \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x}$$

sind erwartungstreu im Sinn von $E(\hat{\beta}) = \beta$ und $E(\hat{\alpha}) = \alpha$. Es gilt weiter

$$V(\hat{\beta}) = \frac{\sigma^2}{S_{x,x}}, \quad V(\hat{\alpha}) = \frac{(S_{x,x} + n(\bar{x})^2)\sigma^2}{nS_{x,x}}, \quad \text{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{x}\sigma^2}{S_{x,x}}.$$

Beweis: Wegen $\sum_{j=1}^n (x_j - \bar{x}) = 0$ folgt

$$\hat{\beta} = \frac{S_{x,Y}}{S_{x,x}} = \frac{\sum_{j=1}^n (x_j - \bar{x})(Y_j - \bar{Y})}{S_{x,x}} = \frac{\sum_{j=1}^n (x_j - \bar{x}) Y_j}{S_{x,x}} = \sum_{j=1}^n \frac{(x_j - \bar{x})}{S_{x,x}} Y_j$$

und damit

$$\begin{aligned} E(\hat{\beta}) &= \frac{1}{S_{x,x}} \sum_{j=1}^n (x_j - \bar{x}) E(Y_j) = \frac{1}{S_{x,x}} \sum_{j=1}^n (x_j - \bar{x})(\alpha + \beta x_j) \\ &= \frac{\alpha}{S_{x,x}} \sum_{j=1}^n (x_j - \bar{x}) + \frac{\beta}{S_{x,x}} \sum_{j=1}^n (x_j - \bar{x}) x_j \\ &= \frac{\beta}{S_{x,x}} \sum_{j=1}^n (x_j - \bar{x})(x_j - \bar{x}) = \beta. \end{aligned}$$

Weiter gilt:

$$\begin{aligned} E(\hat{\alpha}) &= E(\bar{Y} - \hat{\beta}\bar{x}) = E(\bar{Y}) - E(\hat{\beta})\bar{x} \\ &= \frac{1}{n} \sum_{j=1}^n E(Y_j) - \beta\bar{x} = \frac{1}{n} \sum_{j=1}^n (\alpha + \beta x_j) = \alpha \end{aligned}$$

Da die Zufallsvariablen Y_1, \dots, Y_n unabhängig sind folgt

$$\begin{aligned} V(\hat{\beta}) &= V\left(\sum_{j=1}^n \frac{(x_j - \bar{x})}{S_{x,x}} Y_j\right) = \sum_{j=1}^n V\left(\frac{(x_j - \bar{x})}{S_{x,x}} Y_j\right) \\ &= \sum_{j=1}^n \left(\frac{(x_j - \bar{x})}{S_{x,x}}\right)^2 V(Y_j) = \frac{1}{(S_{x,x})^2} \sum_{j=1}^n (x_j - \bar{x})^2 \sigma^2 = \frac{\sigma^2}{S_{x,x}}. \end{aligned}$$

und entsprechend

$$\begin{aligned} V(\hat{\alpha}) &= V(\bar{Y} - \hat{\beta}\bar{x}) = V\left(\sum_{j=1}^n \left(\frac{1}{n} Y_j - \frac{(x_j - \bar{x})}{S_{x,x}} \bar{x} Y_j\right)\right) \\ &= \sum_{j=1}^n \left(\frac{1}{n} - \frac{(x_j - \bar{x})\bar{x}}{S_{x,x}}\right)^2 V(Y_j) = \sigma^2 \sum_{j=1}^n \left(\frac{1}{n^2} - \frac{2(x_j - \bar{x})\bar{x}}{n S_{x,x}} + \frac{(x_j - \bar{x})^2 \bar{x}^2}{(S_{x,x})^2}\right) \\ &= \sigma^2 \sum_{j=1}^n \left(\frac{1}{n^2} + \frac{(x_j - \bar{x})^2 \bar{x}^2}{(S_{x,x})^2}\right) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{x,x}}\right) = \frac{\sigma^2 (S_{x,x} + n\bar{x}^2)}{nS_{x,x}}. \end{aligned}$$

Da Y_1, \dots, Y_n unabhängig sind, folgt

$$\text{Cov}(Y_i, \hat{\beta}) = \text{Cov}\left(Y_i, \sum_{j=1}^n \frac{(x_j - \bar{x})}{S_{x,x}} Y_j\right) = \sum_{j=1}^n \frac{(x_j - \bar{x})}{S_{x,x}} \text{Cov}(Y_i, Y_j) = \frac{x_i - \bar{x}}{S_{x,x}} \sigma^2$$

für alle $i = 1, \dots, n$ und damit weiter

$$\begin{aligned} \text{Cov}(\hat{\alpha}, \hat{\beta}) &= \text{Cov}(\bar{Y} - \hat{\beta}\bar{x}, \hat{\beta}) = \text{Cov}(\bar{Y}, \hat{\beta}) - \bar{x} \text{Cov}(\hat{\beta}, \hat{\beta}) \\ &= \frac{1}{n} \sum_{j=1}^n \text{Cov}(Y_j, \hat{\beta}) - \bar{x} V(\hat{\beta}) = \frac{\sigma^2}{nS_{x,x}} \sum_{j=1}^n (x_j - \bar{x}) - \frac{\bar{x}\sigma^2}{S_{x,x}} = -\frac{\bar{x}\sigma^2}{S_{x,x}}. \end{aligned}$$

□

2.2 Parameterabschätzung der Statistik

Eine Aufgabenstellung der Statistik besteht darin aus gegebenen Daten oder anderen Gelegenheiten Rückschlüsse auf theoretische Zusammenhänge zu machen und diese gegebenenfalls zu überprüfen. In dem vorliegenden Abschnitt sollen zunächst Parameter auf Grund von realen Beobachtungen bestimmt oder geschätzt werden.

2.2.1 Beispiel: Kontrolle der Produktionsqualität. Von $N = 100000$ Schrauben wird eine Stichprobe von $n = 100$ Stück genommen, von diesen ist eine zufällige Zahl x defekt. Welche Rückschlüsse kann man auf die wahre Zahl w der defekten Schrauben ziehen.

1. *Ansatz:* Wir setzen als Schätzung an

$$W(x) = x \cdot \frac{N}{n}$$

Dabei ist $W(x)$ eine von dem Zufall der gewählten Stichprobe abhängige Größe.

Wir behandeln zunächst diesen ersten Ansatz allgemein. Wir gehen dabei von einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) aus und einer Zufallsvariablen

$$X : \Omega \rightarrow \mathcal{X}$$

wobei \mathcal{X} die wirklich beobachtbaren Ereignisse enthält. Dieser Raum \mathcal{X} wird als **Stichprobenraum** bezeichnet, die Gesamtheit aller möglichen Beobachtungen x . Da Ω und X keine weitere Rollen spielen, sondern nur die Verteilung von X , wird dieser theoretische Zusammenhang nicht weiter beachtet. Die Aufgabe besteht darin das Wahrscheinlichkeitsmaß auf \mathcal{X} , das dieser Verteilung zugeordnet ist, aus den Beobachtungswerten zu ermitteln.

2.2.2 Definition: Ein statistisches Modell ist ein Tripel $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ mit dem Stichprobenraum \mathcal{X} , einer σ -Algebra $\mathcal{F} \subset \mathcal{P}(\mathcal{X})$, Wahrscheinlichkeitsmaßen P_ϑ auf \mathcal{F} für alle $\vartheta \in \Theta$ und $\text{Anz}(\Theta) \geq 2$. Für $\vartheta \in \Theta$ sei E_ϑ der Erwartungswert bezüglich des Maßes ϑ , V_ϑ entsprechend die Varianz bezüglich ϑ .

Wir nennen das statistische Modell $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ ein **parametrisches Modell**, wenn $\Theta \subset \mathbb{R}^q$ für ein $q \in \mathbb{N}$ gilt. Im Fall $q = 1$ heißt es ein **einparametrisches Modell**.

Ist \mathcal{X} diskret und besitzt jedes P_ϑ eine Zähldichte ρ_ϑ , so heißt $(\mathcal{X}, P_\vartheta \mid \vartheta \in \Theta)$ ein **diskretes Modell**. In diesem Fall gilt natürlich $\mathcal{F} = \mathcal{P}(\mathcal{X})$.

Im Fall $\mathcal{X} \subset \mathcal{B}^q$ für ein $q \in \mathbb{N}$ und $\mathcal{F} = \mathcal{B}_{\mathcal{X}}^q$, der auf \mathcal{X} eingeschränkte Borel algebra, heißt $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ ein **stetiges Modell**, wenn jedes P_ϑ eine Dichtefunktion ρ_ϑ besitzt.

In den beiden letztgenannten Fällen sprechen wir von dem Vorliegen eines **Standardmodells**.

Im Fall einer abzählbaren Indexmenge Θ gehen wir der Einfachheit halber von $\Theta \subset \mathbb{R}$ aus, und es liegt daher der einparametrische Fall vor.

In der Praxis werden viele Aussagen der Statistik aus Stichproben gewonnen, die in endlich vielen unabhängigen Versuchen oder Auswertungen gewonnen werden. Mathematisch gesehen liegt damit die Situation einer Bildung eines mathematischen Produktes vor.

2.2.3 Definition: $(\mathcal{E}, \mathcal{G}, Q_\vartheta \mid \vartheta \in \Theta)$ sei ein statistisches Modell und $2 \leq n \in \mathbb{N}$. Wir setzen $\mathcal{X} = \mathcal{E}^n$, $\mathcal{F} = \mathcal{G}^{\otimes n}$, $P_\vartheta = Q_\vartheta^{\otimes n}$ für alle $\vartheta \in \Theta$ und betrachten dann das endliche Produktmodell

$$(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta) = (\mathcal{E}^n, \mathcal{G}^{\otimes n}, Q_\vartheta^{\otimes n} \mid \vartheta \in \Theta).$$

Offenbar ist das Produktmodell eines parametrischen Modells wieder parametrisch und das eines Standardmodells ist wieder eines. Wir geben jetzt die Definition eines Schätzers an; darunter verstehen wir speziell die Angabe einer Vorschrift aus dem Datenmaterial einen gewissen Parameter zu bestimmen.

2.2.4 Definition: $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ sei ein statistisches Modell und (Σ, \mathcal{S}) ein Ereignisraum.

i) Jede Abbildung $S : (\mathcal{X}, \mathcal{F}) \rightarrow (\Sigma, \mathcal{S})$ heißt dann eine Statistik.

ii) Ist $\tau : \Theta \rightarrow \Sigma$ eine Abbildung, so heißt jede Statistik $T : (\mathcal{X}, \mathcal{F}) \rightarrow (\Sigma, \mathcal{S})$ dann ein Schätzer für τ .

Anmerkung: Man beachte bei dieser Definition, dass eine Statistik eine Zufallsvariable ist mit allerdings einer anderen Interpretation: Es ist eine konstruierte Größe, die auf Beobachtungen beruht. Es sei an dieser Stelle darauf hingewiesen, dass die obige Definition wegen der großen Allgemeinheit nicht in jeder Situation sinnvoll ist.

2.2.5 Beispiel: Aus $[0, \vartheta]$ für ein $\vartheta \in \mathbb{R}$ werden n Zahlen x_1, \dots, x_n ausgewählt. Dabei ist ϑ unbekannt und soll geschätzt werden:

Statistisches Modell: $\mathcal{X} = [0, \infty[^n$, $\mathcal{F} = \mathcal{B}^n$, $\Theta =]0, \infty[$; für $\vartheta > 0$ sei Q_ϑ die Gleichverteilung auf $[0, \vartheta]$, also $P_\vartheta = Q_\vartheta^{\otimes n}$. Für die weitere Betrachtung fassen wir in Gedanken die Werte x_1, \dots, x_n als Werte von unabhängigen Zufallsvariablen X_1, \dots, X_n auf, die auf dem Intervall $[0, \vartheta]$ gleichverteilt sind. Dann gilt offenbar

$$E(Q_\vartheta) = E_\vartheta(X_k) = \frac{\vartheta}{2} = \int_0^\vartheta t \cdot \frac{1}{\vartheta} dt.$$

1. Fall: Wir betrachten den Schätzer

$$T_n := 2M = \frac{2}{n} \sum_{k=1}^n X_k \quad \text{-- doppelter Mittelwert.}$$

Das Gesetz der großen Zahl impliziert

$$P_\vartheta(|T_n - \vartheta| > \varepsilon) \rightarrow 0 \quad \text{bei } n \rightarrow \infty \text{ für alle } \varepsilon > 0.$$

2. Fall: Als Schätzer betrachten wir jetzt $\tilde{T}_n = \max\{X_1, \dots, X_n\}$: Für alle $\varepsilon > 0$ gilt

$$\begin{aligned} P_\vartheta(\tilde{T}_n \leq \vartheta - \varepsilon) &= P_\vartheta([X_1 \leq \vartheta - \varepsilon] \cap \dots \cap [X_n \leq \vartheta - \varepsilon]) \\ &= P_\vartheta(X_1 \leq \vartheta - \varepsilon) \cdot \dots \cdot P_\vartheta(X_n \leq \vartheta - \varepsilon) \\ &= \left(\frac{\vartheta - \varepsilon}{\vartheta}\right)^n \rightarrow 0 \quad \text{bei } n \rightarrow \infty. \end{aligned}$$

Vergleich dieser Schätzer:

(I) Beide Schätzer sind **konsistent** in dem Sinne, dass

$$T_n \rightarrow \vartheta \text{ und } \tilde{T}_n \rightarrow \vartheta \text{ bei } n \rightarrow \infty \text{ bezüglich } P_\vartheta$$

Die Schätzer besitzen gleiches asymptotisches Verhalten.

(II) Der Schätzer T_n ist **erwartungstreu** in dem Sinne, dass

$$E_\vartheta(T_n) = \frac{2}{n} \sum_{k=1}^n E_\vartheta(X_k) = \vartheta$$

für alle $\vartheta \in \Theta$ gilt.

(III) \tilde{T}_n ist nicht erwartungstreu sondern **asymptotisch erwartungstreu** im Sinne von

$$E_\vartheta(\tilde{T}_n) \rightarrow \vartheta \text{ bei } n \rightarrow \infty .$$

Man beachte: Es gilt $P_\vartheta(\tilde{T}_n \leq x) = \left(\frac{x}{\vartheta}\right)^n$ für alle $x \in [0, \vartheta]$, also besitzt die Verteilungsfunktion die Dichte $g(t) : g(t) = \frac{n}{\vartheta} \left(\frac{x}{\vartheta}\right)^{n-1}$ für $t \in [0, \vartheta]$ und $g(t) = 0$ sonst. Damit gilt

$$E_\vartheta(\tilde{T}_n) = \int_0^\vartheta t g(t) dt = \frac{n}{n+1} \vartheta .$$

(IV) Betrachtung der Varianz:

$$V_\vartheta(T_n) = \left(\frac{2}{n}\right)^2 V_\vartheta(X_1 + \dots + X_n) = \frac{4}{n} V_\vartheta(X_1) = \frac{4}{n\vartheta} \int_0^\vartheta \left(t - \frac{\vartheta}{2}\right)^2 dt = \frac{\vartheta^2}{3n}$$

Wegen der Berechnung in III gilt

$$\begin{aligned} V_\vartheta(\tilde{T}_n) &= E_\vartheta(\tilde{T}_n^2) - E_\vartheta(\tilde{T}_n)^2 = \int_0^\vartheta t^2 n t^{n-1} \vartheta^{-n} dt - \left(\frac{n\vartheta}{n+1}\right)^2 \\ &= \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) \vartheta^2 = \frac{n}{(n+1)^2(n+2)} \vartheta^2 . \end{aligned}$$

Setzen wir $T_n^* = \frac{n+1}{n} \tilde{T}_n$ so ist T_n^* erwartungstreu mit

$$V_\vartheta(T_n^*) = \frac{\vartheta^2}{n(n+2)} .$$

Da der Schätzer \tilde{T}_n nicht erwartungstreu ist, wird auch der mittlere quadratische Fehler für \tilde{T}_n bestimmt:

$$E_\vartheta((\tilde{T}_n - \vartheta)^2) = V_\vartheta(\tilde{T}_n) + \left(E_\vartheta(\tilde{T}_n) - \vartheta\right)^2 = \frac{2\vartheta^2}{(n+1)(n+2)}$$

Der beste erwartungstreue Schätzer unter den oben behandelten Schätzern ist in diesem Fall der Schätzer T_n^* .

Im weiteren Verlauf diskutieren wir jetzt **Maximum-Likelihood-Schätzer**. Dieser Schätzer wird dabei für diskrete und für stetige Modelle, also für Standardmodelle gebildet.

Der diskrete Fall: Wir gehen von einem diskreten statistischen Modell aus:

$$(\mathcal{X}, P_\vartheta | \vartheta \in \Theta)$$

sei gegeben und es werde eine Größe x beobachtet. Ist ϑ der richtige Parameter, so können wir die Wahrscheinlichkeit

$$\rho_\vartheta(x) = P_\vartheta(\{x\})$$

berechnen, mit der dieses Ereignis $x \in \mathcal{X}$ eintritt. In diesem Fall ist ρ_ϑ die Zähldichte.

Der stetige Fall: Es sei jetzt $(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta)$ ein stetiges Modell, und jedes P_ϑ besitzt daher eine Dichtefunktion ρ_ϑ . Es werden jetzt ebenfalls eine Größe $x \in \mathcal{X}$ beobachtet. Wir betrachten jetzt ebenfalls $\rho_{\vartheta(x)}$.

In beiden Fällen ist es nun unwahrscheinlich, dass ϑ der richtige Parameter ist, wenn $\rho_\vartheta(x)$ sehr klein ist: Das würde bedeuten, dass das beobachtete Ereignis ein seltenes Ausnahmereignis wäre. Daher bestimmen wir den Schätzer $T(x)$ zu dem beobachteten Ereignis $x \in \mathcal{X}$, so dass die folgende Bedingung erfüllt ist:

$$\rho_{T(x)}(x) = \max \{ \rho_\vartheta(x) | \vartheta \in \Theta \} .$$

Wir kommen zu der folgenden Definition:

2.2.6 Definition: $(\mathcal{X}, \mathcal{F}, P_\vartheta | \vartheta \in \Theta)$ sei ein statistisches Standardmodell. Die Funktion

$$\begin{aligned} \rho : \mathcal{X} \times \Theta &\rightarrow [0, \infty[\\ (x, \vartheta) &\mapsto \rho(x, \vartheta) = \rho_\vartheta(x) \end{aligned}$$

heißt die zugeordnete **Likelihoodfunktion**, und $\Theta \ni \vartheta \mapsto \rho(x, \vartheta)$ heißt die **Likelihoodfunktion zum Beobachtungswert x** . Ein Schätzer $T : \mathcal{X} \rightarrow \Theta$ mit der Eigenschaft $\rho(x, T(x)) = \max \{ \rho(x, \vartheta) | \vartheta \in \Theta \}$ heißt der **Maximum-Likelihood-Schätzer**.

Anmerkung: Man beachte, dass ein Maximum-Likelihood-Schätzer nicht eindeutig bestimmt sein muss.

2.2.7 Beispiel: Die Qualität einer Lieferung von N gleichen Produkten soll durch eine Stichprobe vom Umfang n überprüft werden. Diese enthalte x defekte Produkte. Es soll die Zahl ϑ aller defekten Produkte geschätzt werden. Damit ergibt sich folgendes statistische Modell: $(\mathcal{X}, \mathcal{P}(\mathcal{X}), P_\vartheta | \vartheta \in \Theta)$ mit $\mathcal{X} = \{0, \dots, n\}$, $\Theta = \{0, \dots, N\}$ und

$$\rho_\vartheta(x) = \frac{\binom{\vartheta}{x} \binom{N-\vartheta}{n-x}}{\binom{N}{n}}$$

- die hypergeometrische Verteilung. Wir bestimmen jetzt den Schätzer $T(x)$, so dass $\rho_\vartheta(x)$ maximal ist im Fall $\vartheta = T(x)$. Für alle $\vartheta \in \mathbb{N}$ mit $\vartheta \leq N$ gilt

$$\frac{\rho_\vartheta(x)}{\rho_{\vartheta-1}(x)} = \frac{\binom{\vartheta}{x} \binom{N-\vartheta}{n-x}}{\binom{\vartheta-1}{x} \binom{N-\vartheta+1}{n-x}} = \frac{\vartheta(N-\vartheta-(n-x)+1)}{(\vartheta-x)(N-\vartheta+1)}$$

Offenbar ist $\frac{\rho_{\vartheta}(x)}{\rho_{\vartheta-1}(x)} \geq 1$ äquivalent zu

$$\begin{aligned} \vartheta N - \vartheta^2 - \vartheta n + \vartheta x + \vartheta &\geq \vartheta N - \vartheta^2 + \vartheta - xN + \vartheta x - x \text{ also zu} \\ -\vartheta n &\geq -xN - x \text{ und damit zu} \\ \vartheta &\leq x \cdot \frac{N+1}{n}. \end{aligned}$$

Wir setzen daher als Schätzer an: $T(x) = \lfloor x \frac{N+1}{n} \rfloor$ - den ganzzahligen Anteil dieser Zahl. Offenbar ist die Funktion $\Theta \ni \vartheta \mapsto \rho_{\vartheta}(x)$ monoton wachsend auf $\{0, \dots, T(x)\}$ und auf $\{T(x), \dots, N\}$ monoton fallend, sie nimmt also in $T(x)$ das Maximum an.

2.2.8 Beispiel: (Schätzung der Zahl von Kugeln in einer Urne) Eine Urne enthalte eine unbekannte (relativ große) Anzahl ϑ von Kugeln. Diese Zahl $\vartheta \in \mathbb{N}$ soll geschätzt werden. Dazu werden ω dieser Kugeln entnommen, markiert und wieder mit den anderen Kugeln vermischt. Danach werden wieder n Kugeln entnommen, von denen x markiert sind. Wie groß ist ϑ ?

Das statistische Modell ist

$$\mathcal{X} = \{0, \dots, n\}, \quad \Theta = \{\omega, \omega + 1, \dots\} \subset \mathbb{N}$$

P_{ϑ} sei definiert durch

$$\rho_{\vartheta}(x) = \frac{\binom{\omega}{x} \binom{\vartheta - \omega}{n - x}}{\binom{\vartheta}{n}}$$

Analog wie im ersten Beispiel können wir schließen, dass

$$T(x) = \lfloor n \frac{\omega}{x} \rfloor$$

der Maximum-Likelihood-Schätzer für ϑ ist.

2.2.9 Beispiel: (Schätzung der Erfolgswahrscheinlichkeit) Es liege die Situation von Bernoulli-Experimenten mit einer zu schätzenden Erfolgswahrscheinlichkeit $p = \vartheta \in [0, 1]$ vor: Bei n Experimenten trete x -mal Erfolg auf.

Das statistische Modell wird gegeben durch

$$\mathcal{X} = \{0, 1, \dots, n\}, \quad \Theta = [0, 1]$$

mit den Wahrscheinlichkeitsmaßen P_{ϑ} definiert durch

$$\rho_{\vartheta}(x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x},$$

und mit $\rho_{\vartheta}(x)$ als Maximum-Likelihood-Funktion.

Zur Bestimmung des Maximalwertes dieser Funktion gehen wir auf die so genannte **Log-Likelihood-Funktion** $\ln \rho_{\vartheta}(x)$ über, die bekanntlich wegen der Monotonie von $t \mapsto \ln(t)$ die gleichen Extremstellen aufweist wie die Funktion $\vartheta \mapsto \rho_{\vartheta}(x)$. Es gilt

$$\ln \rho_{\vartheta}(x) = \ln \binom{n}{x} + x \ln(\vartheta) + (n - x) \ln(1 - \vartheta).$$

Durch Differentiation nach ϑ und Nullstellenbestimmung der Ableitung folgt

$$\frac{x}{\vartheta} - \frac{(n-x)}{1-\vartheta} = 0, \text{ also } \vartheta = \frac{x}{n}$$

$T : T(x) = \frac{x}{n}$ ist offenbar der (eindeutig bestimmte) Maximum-Likelihood-Schätzer für ϑ .

2.2.10 Beispiel: Wir wollen in dem schon früher behandelten Beispiel 2.2.5 den Bereich von Zufallszahlen schätzen:

Für unbekanntes $\vartheta > 0$ und n Zufallszahlen $x_1, \dots, x_n \in [0, \vartheta]$ soll ϑ geschätzt werden:

Wir haben das Produktmodell

$$([0, \infty[^n, \mathcal{B}^n, Q_{\vartheta}^{\otimes n} \mid \vartheta \in]0, \infty[)$$

mit der Gleichverteilung Q_{ϑ} auf $[0, \vartheta]$. Die Likelihoodfunktion ist

$$\rho_x(\vartheta) = \begin{cases} \vartheta^{-n} & \text{falls } x_1, \dots, x_n \leq \vartheta \\ 0 & \text{sonst} \end{cases}$$

Also ist der Schätzer $\tilde{T}_n(x) = \max\{x_1, \dots, x_n\}$ der Maximum-Likelihood-Schätzer. - Man vergleiche das frühere Beispiel 2.2.1.

2.2.11 Beispiel: (n -fach normalverteiltes - das **Gaußsche Produktmodell**) Der Ausgang eines Experimentes sei $N(\mu, v)$ -verteilt (normal-verteilt mit Mittelwert μ , Varianz $\sigma^2 = v$) mit unbekanntem Parametern $\mu \in \mathbb{R}$, $v \geq 0$. Es liegen n Ergebnisse x_1, \dots, x_n vor. Dieser Situation entspricht folgendes n -fach-normalverteilte (Gaußsche) Produktmodell

$$(\mathcal{X}, \mathcal{F}, P_{\vartheta} \mid \vartheta \in \Theta) = (\mathbb{R}^n, \mathcal{B}^n, N_{\mu, v}^{\otimes n} \mid (\mu, v) \in \mathbb{R} \times]0, \infty[) .$$

Die entsprechende Likelihoodfunktion hat die Form

$$\rho(x, \mu, v) = \prod_{j=1}^n \varphi_{\mu, v}(x_j) = \left(\frac{1}{2\pi v}\right)^{n/2} \exp\left(-\sum_{j=1}^n \frac{(x_j - \mu)^2}{2v}\right)$$

für $x_1 = (x_1, \dots, x_n) \in \mathbb{R}^n$ und $\vartheta = (\mu, v) \in \Theta = \mathbb{R} \times]0, \infty[$.

Die notwendigen Bedingungen für das Vorliegen eines (lokalen) Maximums sind

$$D_2\rho(x, \mu, v) = 0, \quad D_3\rho(x, \mu, v) = 0.$$

Aus der ersten Forderung erhalten wir

$$\frac{1}{v} \sum_{j=1}^n (x_j - \mu) = 0, \text{ also } \mu = \frac{1}{n}(x_1 + \dots + x_n).$$

Aus der zweiten Forderung erhalten wir

$$-\frac{n}{2}v^{-\frac{n}{2}-1} + v^{-\frac{n}{2}} \frac{1}{2v^2} \sum_{j=1}^n (x_j - \mu)^2 = 0 \text{ oder äquivalent dazu}$$

$$v = \frac{1}{n} \sum_{j=1}^n (x_j - \mu)^2.$$

Durch Bildung der Hesse Matrix kann man mit etwas Rechnung zeigen, dass in dem eben bestimmten Werten ein lokales Maximum vorliegt. Man erkennt weiter, dass dieses auch ein globales Maximum ist. \square

2.2.12 Beispiel: (Produktmodell mit Exponentialverteilung zum Parameter $\alpha > 0$) Dieser Parameter soll geschätzt werden. Der Maximum-Likelihood-Schätzer für α ergibt sich

$$T(x) = \frac{n}{x_1 + \dots + x_n}$$

für $x = (x_1, \dots, x_n)$. $S(x) = \frac{1}{T(x)}$ ist dann ein erwartungstreuer Schätzer für $\frac{1}{\alpha}$. In diesem Fall ist die Kenngröße gegeben durch $\tau(\alpha) = \frac{1}{\alpha}$.

2.2.13 Satz: *Der Maximum-Likelihood-Schätzer im n -fachen Produktmodell einer $N(\mu, v)$ -Verteilung (Gaußsches Produktmodell) wird gegeben durch*

$$T(X) = (\mu(X), v(X)) = \left(\frac{1}{n} \sum_{j=1}^n X_j, \frac{1}{n} \sum_{j=1}^n (X_j - \mu(X))^2 \right).$$

- Stichprobenmittel und Stichprobenvarianz.

Es soll jetzt die Qualität von Schätzern näher untersucht werden. Dazu kommen wir auf den bereits eingeführten Begriff der Erwartungstreue zurück.

2.2.14 Definition: $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ sei ein statistisches Modell. $\tau : \Theta \rightarrow \mathbb{R}$ eine Abbildung (oder reelle Kenngröße). $T : \mathcal{X} \rightarrow \mathbb{R}$ heißt ein erwartungstreuer Schätzer für τ , wenn

$$E_\vartheta(T) = \tau(\vartheta)$$

gilt für alle $\vartheta \in \Theta$. Ansonsten heißt $B_\vartheta(T) = E_\vartheta(T) - \tau(\vartheta)$ der **Bias** oder der **systematische Fehler** von T .

Dabei werden in der obigen Definition die Existenz der Erwartungswerte stillschweigend vorausgesetzt. Im Gaußschen Produktmodell gilt die folgende Aussage.

2.2.15 Satz: *Es sei $2 \leq n \in \mathbb{N}$ und $(\mathbb{R}^n, \mathcal{B}^n, Q_\vartheta^{\otimes n} \mid \vartheta \in \Theta)$ ein n -faches Produktmodell, so dass für jedes $\vartheta \in \Theta$ die Werte $m(\vartheta) = E(Q_\vartheta)$ und $v(\vartheta) = V(Q_\vartheta)$ definiert sind. Dann sind Stichprobenmittelwert und korrigierte Stichprobenvarianz*

$$M = \frac{1}{n} \sum_{j=1}^n X_j, \quad V^* = \frac{1}{n-1} \sum_{j=1}^n (X_j - M)^2$$

erwartungstreue Schätzer für m und v .

Beweis: Da X_j bezüglich Q_ϑ verteilt ist für $j = 1, \dots, n$ folgt aus der Linearität unmittelbar

$$E_\vartheta(M) = \frac{1}{n} \sum_{j=1}^n E_\vartheta(X_j) = m(\vartheta).$$

Es sei weiter $V = \frac{1}{n} \sum_{j=1}^n (X_j - M)^2$. Man rechnet unmittelbar nach, dass

$$E_\vartheta(V) = \frac{n-1}{n} v(\vartheta) \text{ gilt.}$$

Dabei beachte man, dass X_1, \dots, X_n unabhängig sind. \square

Eine weitere wichtige Größe von Schätzern ist neben der Erwartungstreue auch der mittlere quadratische Fehler eines Schätzers T für τ von Bedeutung. Dieser ist unabhängig von der Erwartungstreue, es existieren wie im nachfolgenden Beispiel nicht erwartungstreue Schätzer, also Schätzer mit einem Bias, mit einem geringeren quadratischen Fehler. Dieser **mittlere quadratische Fehler** wird definiert durch

$$F_\vartheta(T) = E_\vartheta((T - \tau(\vartheta))^2) = V_\vartheta(T) + (B_\vartheta(T))^2.$$

2.2.16 Beispiel: Ein guter Schätzer mit Bias im Binomialmodell

$$\mathcal{X} = \{0, \dots, n\}, \quad \Theta = [0, 1], \quad P_\vartheta(k) = \binom{n}{m} \vartheta^k (1 - \vartheta)^{n-k}.$$

Der Maximumschätzer $T : T(x) = \frac{x}{n}$ ist erwartungstreu. Sein quadratischer Fehler ist

$$F_\vartheta(T) = \frac{1}{n^2} V(b_{n,\vartheta}) + 0 = \frac{1}{n} \vartheta(1 - \vartheta).$$

Wir betrachten jetzt den Schätzer $S : S(x) = \frac{x+1}{n+2}$ mit dem Bias

$$B_\vartheta(S) = \frac{n\vartheta + 1}{n+2} - \vartheta = \frac{1 - 2\vartheta}{n+2}$$

und dem quadratischen Fehler

$$F_\vartheta(S) = V_\vartheta(S) + (B_\vartheta(S))^2 = \frac{n\vartheta(1 - \vartheta) + (1 - 2\vartheta)^2}{(n+2)^2}.$$

Weiter gilt

$$F_\vartheta(T) - F_\vartheta(S) = \frac{(4n+4)\vartheta(1 - \vartheta) - n(1 - 2\vartheta)^2}{n(n+2)^2} \geq 0.$$

Diese Aussage ist äquivalent zu

$$4(n+1)\vartheta(1 - \vartheta) - n(1 - 2\vartheta)^2 \geq 0$$

also zu $1 + \frac{1}{n} \geq \frac{(\vartheta - \frac{1}{2})^2}{\vartheta(1 - \vartheta)}$.

Wenn diese Ungleichung für alle $n \in \mathbb{N}$ gelten soll, erhalten wir durch den Grenzübergang $n \rightarrow \infty$ die Beziehung $\vartheta(1 - \vartheta) \geq (\vartheta - \frac{1}{2})^2$ und dann durch Umrechnen $(\vartheta - \frac{1}{2})^2 \leq \frac{1}{8}$, also $0.1464 \leq \vartheta \leq 0.8536$. Für diese ϑ liefert der Schätzer S einen kleineren mittleren Fehler.

2.2.17 Definition: $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ sei ein statistisches Modell. $\tau : \Theta \rightarrow \mathbb{R}$ sei eine Kenngröße und $T : \mathcal{X} \rightarrow \mathbb{R}$ ein erwartungstreuer Schätzer für τ . T heißt **varianzminimierend** oder **besten Schätzer**, wenn für jeden erwartungstreuen Schätzer $S : \mathcal{X} \rightarrow \mathbb{R}$

$$V_\vartheta(T) \leq V_\vartheta(S) \text{ gilt.}$$

Im weiteren Verlauf sollen beste Schätzer konstruiert werden. Dazu nennen wir ein statistisches Modell $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ **regulär**, wenn folgende Eigenschaften erfüllt sind:

- i) $\Theta \subset \mathbb{R}$ ist ein offenes Intervall.
- ii) Die Likelihood Funktion $\rho : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ ist strikt positiv und stetig nach $\vartheta \in \Theta$ differenzierbar mit der so genannten **Scorefunktion**

$$U_\vartheta(x) = D_\vartheta \log(\rho(x, \vartheta)) = \frac{D_\vartheta \rho(x, \vartheta)}{\rho(x, \vartheta)} .$$

- iii) $I(\vartheta) := V_\vartheta(U_\vartheta) > 0$ für alle $\vartheta \in \Theta$.
- iv) $\int D_\vartheta \rho(x, \vartheta) dx = D_\vartheta \int \rho(x, \vartheta) dx$.

$I(\vartheta) : I : \Theta \rightarrow \mathbb{R}$ heißt die **Fisher-Information** des Modells. Als unmittelbare Konsequenz dieser Definition erhalten wir

$$\begin{aligned} E_\vartheta(U_\vartheta) &= \int D_\vartheta \rho(x, \vartheta) dx = D_\vartheta \int \rho(x, \vartheta) dx = D_\vartheta 1 = 0. \\ I(\vartheta) &= E_\vartheta(U_\vartheta^2) . \end{aligned}$$

Wir diskutieren kurz, warum der Begriff Fisher-Information eingeführt wurde: Die Bedingung $I(\vartheta) = 0$ für alle $\vartheta \in \Theta$ bei einem Intervall $\Theta_0 \subset \Theta$ ist äquivalent zu $U_\vartheta(x) = 0$ für alle $\vartheta \in \Theta_0$ und fast alle $x \in \mathcal{X}$. Damit können durch keine Beobachtung die Parameter aus Θ_0 unterschieden werden. Es gilt die folgende Bemerkung

2.2.18 Bemerkung: $\mathcal{M} = (\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ sei regulär mit der Fisher-Information $I : \Theta \rightarrow \mathbb{R}$. Dann besitzt das Produktmodell

$$\mathcal{M}^{\otimes n} = (\mathcal{X}^n, \mathcal{F}^{\otimes n}, P_\vartheta^{\otimes n} \mid \vartheta \in \Theta)$$

die Fisher-Information $I^{\otimes n} = nI$.

Beweis: $\mathcal{M}^{\otimes n}$ hat die Likelihoodfunktion $\rho_\vartheta^{\otimes n} \prod_{k=1}^n \rho_\vartheta \circ X_k$ also gilt

$$U_\vartheta^{\otimes n} = \sum_{k=1}^n U_\vartheta \circ X_k .$$

Wobei $X_k : \mathcal{X}^n \rightarrow \mathcal{X}$ die Projektion auf die k -te Komponente ist. Damit folgt aus der Unabhängigkeit der X_k :

$$I_\vartheta^{\otimes n} = V_\vartheta(U_\vartheta^{\otimes n}) = \sum_{k=1}^n V_\vartheta(U_\vartheta \circ X_k) = n I(\vartheta) .$$

□

Ein erwartungstreuer Schätzer T für τ heißt regulär, wenn

$$\int T x \frac{d}{d\vartheta} \rho(x, \vartheta) dx = \frac{d}{d\vartheta} \int T x \rho(x, \vartheta) dx$$

erfüllt ist

2.2.19 Satz: (*Informationsungleichung*) $(\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ sei ein reguläres statistisches Modell, $\tau : \Theta \rightarrow \mathbb{R}$ eine zu schätzende stetige differenzierbare Funktion, und $T : \mathcal{X} \rightarrow \mathbb{R}$ ein regulärer erwartungstreuer Schätzer für τ . Dann gilt

$$V_\vartheta(T) \geq \frac{\tau'(\vartheta)^2}{I(\vartheta)} \text{ für alle } \vartheta \in \Theta.$$

Die Gleichheit für alle $\vartheta \in \Theta$ ist äquivalent zu

$$T - \tau(\vartheta) = \tau'(\vartheta) \frac{U_\vartheta}{I(\vartheta)}$$

für alle $\vartheta \in \Theta$. Also wenn das Modell die Likelihoodfunktion

$$\rho(x, \vartheta) = \exp[a(\vartheta)T(x) - b(\vartheta)]h(x)$$

mit einer Stammfunktion $a : \Theta \rightarrow \mathbb{R}$ der Funktion $\frac{I}{\tau'}$, und $h : \mathcal{X} \rightarrow]0, \infty[$ messbar mit der Normierungsfunktion

$$b(\vartheta) = \ln \left(\int_{\mathcal{X}} e^{a(\vartheta)T(x)} h(x) dx \right).$$

Beweis: Wegen $E_\vartheta(U_\vartheta) = 0$, Regularität und Erwartungstreue folgt

$$\begin{aligned} \text{Cov}_\vartheta(T, U_\vartheta) &= E_\vartheta(TU_\vartheta) = \int_{\mathcal{X}} T(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx \\ &= \frac{d}{d\vartheta} \int_{\mathcal{X}} T(x) \rho(x, \vartheta) dx = \frac{d}{d\vartheta} E_\vartheta(T) = \tau'(\vartheta) \end{aligned}$$

für alle $\vartheta \in \Theta$. Mit $c(\vartheta) = \frac{\tau'(\vartheta)}{I(\vartheta)}$ erhalten wir weiter

$$\begin{aligned} 0 \leq V_\vartheta(T - c(\vartheta)U_\vartheta) &= V_\vartheta(T) + c(\vartheta)^2 V_\vartheta(U_\vartheta) - 2c(\vartheta) \text{Cov}_\vartheta(T, U_\vartheta) \\ &= V_\vartheta(T) - \frac{\tau'(\vartheta)^2}{I(\vartheta)}. \end{aligned}$$

Gleichheit liegt vor, wenn $T - c(\vartheta)U_\vartheta$ bezüglich P_ϑ fast sicher konstant ist. Aus der Erwartungstreue folgt dann weiter

$$T = c(\vartheta)U_\vartheta = \tau(\vartheta).$$

Ist μ das Lebesguemaß oder ein Zählmaß auf \mathcal{X} und besitzt P_ϑ eine positive Dichte bezüglich μ so folgt aus Stetigkeitgründen

$$\mu \left([T - \tau(\vartheta)] \frac{1}{c(\vartheta)} = U_\vartheta \right) = 1.$$

□

2.2.20 Definition: Ein einparametrisches Standardmodell $\mathcal{M} = (\mathcal{X}, \mathcal{F}, P_\vartheta \mid \vartheta \in \Theta)$ heißt ein **exponentielles Modell**, wenn $\Theta \subset \mathbb{R}$ ein Intervall ist, $T : \mathcal{X} \rightarrow \mathbb{R}$ eine Statistik, und

$$\rho(x, \vartheta) = h(x) \exp(a(\vartheta)T(x) - b(\vartheta)) \text{ mit } b(\vartheta) = \ln \left(\int_{\mathcal{X}} \exp(a(\vartheta)T(x)) h(x) dx \right)$$

mit einer differenzierbaren Funktion $a : \Theta \rightarrow \mathbb{R}$ für die $a' \neq 0$ gilt, und $h : \mathcal{X} \rightarrow]0, \infty[$ messbar ist.

2.3 Aussagen zur Testtheorie, Nichtparametrische Tests

Das Ziel dieses Abschnittes ist es einige allgemeine Grundlagen der Testtheorie zu entwickeln und sich dann mit speziellen nichtparametrischen Tests auseinanderzusetzen. Diese können unter sehr allgemeinen Umständen angewandt werden. Allerdings geht dieses nur auf Kosten der Aussagekraft: Bei speziellen Situationen, etwa wenn eine Normalverteilung angenommen werden kann, stehen dann oft bessere Tests zur Verfügung.

Bevor wir allgemeine Grundsätze zu Testtheorie aufstellen, soll das Problem an Hand einer Wahlvorhersage kurz diskutiert werden:

Bei einer großen Grundgesamtheit soll durch eine Befragung von 1 000 Wählern ermittelt werden, wie groß die Wahrscheinlichkeit ist, dass eine Partei F die 5%–Hürde schafft – man beachte, dass diese Forderung umgangssprachlich ist, und daher unbedingt präzisiert werden muss.

Ermittelt wurden bei der dann durchgeführten Befragung $k = 62$ Personen, die sich für diese Partei aussprachen. Fassen wir diese Aufgabestellung als ein Schätzproblem auf, so würden wir als Schätzwert $p = 6,2\%$ als Wert des gesuchten Anteils der Partei A erhalten. Diese Aussage gibt allerdings noch keine Information über die Sicherheit einer Prognose an. Wir müssen daher anders vorgehen: Anstatt eine nicht mögliche Bestimmung der Sicherheit des Überspringens der 5%–Hürde an Hand des Resultates der Befragung zu bestimmen berechnen wir die Wahrscheinlichkeit des Ereignisses A , dass sich 62 oder mehr Personen der Stichprobe für diese Partei F aussprechen unter der Hypothese des vorausgesetzten Stimmenanteils von $p_0 = 5\%$. Die Werte $P(k)$ in Prozentangaben bestimmen wir mit Hilfe der Normalapproximation:

k	52	52	54	55	56	57	58	59	60	61	62	63	64	65
$P(k)$	42,0	33,2	28,0	23,4	19,1	15,0	12,3	8,8	7,3	5,5	4,1	3,0	2,1	1,5

Das Ergebnis der Auswertung lässt sich folgendermaßen formulieren:

Wenn der reale Stimmenanteil der Partei F 5% oder weniger wäre, besitzt das Ereignis bei einer Meinungsumfrage bei 1 000 Wählern ein Ergebnis von 62 oder mehr Entscheidungen für diese Partei zu bekommen, eine Wahrscheinlichkeit von 4,1% oder weniger; es ist daher ein unwahrscheinliches Ereignis. Umgangssprachlich wird in falscher Formulierung dann oft gesagt, dass die Partei F mit einer Sicherheit von 96,9% die 5%–Hürde schafft.

Vergleichen wir in der obigen Tabelle noch die k –Werte, so sehen wir, dass für $k \geq 59$ eine 10%–Marke unterschritten wird. Für k –Werte darunter werden Aussagen dagegen immer ungenauer. Wir kommen in allgemeiner Situation zu folgender Definition; dabei diskutieren wir einige Grundlagen für das Vorgehen bei Entscheidungssituationen:

2.3.1 Definition: Ein statistisches Entscheidungsverfahren besteht aus

(I) Formulierung des statistischen Modells $(\mathcal{X}, \mathcal{F}, \Theta, P_\theta \mid \theta \in \Theta)$.

(II) Formulierung der Null-Hypothese H_0 und der Alternative H_1 durch Zerlegung von $\Theta = \Theta_0 \cup \Theta_1$ in disjunkte Teilmengen nach dem Prinzip, dass Θ_0 alle akzeptablen $\theta \in \Theta$ enthält und dass Θ_1 aus allen problematischen $\theta \in \Theta$ besteht. Man sagt, dass die Null-Hypothese gegen die Alternative getestet werden soll.

(III) Wahl eines Irrtums-Niveaus $\alpha \in]0, 1[$ – etwa $\alpha = 0.1$ oder $\alpha = 0.05$. Man fordert von dem noch zu spezifizierenden Entscheidungsverfahren dass

$$P(H_1|H_0) \leq \alpha$$

gilt: die Wahrscheinlichkeit einer Entscheidung für die Alternative H_1 wenn in Wirklichkeit H_0 gilt, also für den Fehler 1. Art.

(IV) Bestimmung einer Entscheidungsregel: Man wähle eine Statistik $\varphi : \mathcal{X} \rightarrow [0, 1]$ nach folgenden Prinzipien: Für ein beobachtetes $x \in \mathcal{X}$ sei $\varphi(x)$ der Grad mit dem zur Entscheidung für die Alternative tendiert wird in dem Sinne

$\varphi(x) = 0$ – Festhalten an der Null-Hypothese

$\varphi(x) = 1$ – Annahme des Vorliegens der Alternative

$0 < \varphi(x) < 1$ – Unklarheit über eine richtige Entscheidung, Durchführung eines zusätzlichen Zufalls-Experimentes, das mit Wahrscheinlichkeit $\varphi(x)$ eine Entscheidung für die Alternative liefert.

(V) Durchführung des Statistischen Experimentes.

Anmerkung: Es dürfen nie auf Grund des Ausganges eines statistischen Experimentes die Regeln bestimmt werden.

Die Statistik $\varphi : \mathcal{X} \rightarrow [0, 1]$ heißt ein **nichtrandomisierter Test**, wenn $\varphi(x)$ nur die Werte 0 oder 1 annimmt und ansonsten ein **randomisierter Test**.

Die Menge $\{x \in \mathcal{X} \mid \varphi(x) = 1\}$ heißt der **Ablehnungsbereich** oder auch der **Verwerfungsbereich**.

$$\sup \{E_\theta(\varphi) \mid \theta \in \Theta\}$$

ist die im ungünstigsten Fall vorliegende Wahrscheinlichkeit für einen Fehler 1. Art und wird als **Umfang** oder **Niveau** von φ bezeichnet.

Ein Test $\varphi : \mathcal{X} \rightarrow [0, 1]$ mit

$$\sup \{E_\theta(\varphi) \mid \theta \in \Theta\} \leq \alpha$$

heißt ein Test zum Niveau α oder Irrtums-Niveau α .

Wir stellen jetzt die allgemeine Behandlung von Tests zurück und behandeln zunächst einige nichtparametrische Tests.

Der Wilcoxonsche Vorzeichentest

Voraussetzung: Gegeben sei eine Population mit einer stetigen Verteilungsfunktion. Weiter nehmen wir an, dass diese Verteilungsfunktion in dem betrachteten Teilintervall $\subset \mathbb{R}$, in dem die betrachteten Werte liegen, streng monoton wachsend ist. Diese Verteilung besitzt dann dort einen eindeutig bestimmten Median m , der unbekannt sei.

Es werde unabhängig eine Stichprobe x_1, \dots, x_n als Stichprobe gewonnen. Weiter sei $m_0 \in \mathbb{R}$ ein gegebener Wert. Es soll getestet werden, ob $m = m_0$ gilt oder nicht. Wir betrachten also die Nullhypothese

$H_0 : m = m_0$ gegen die Alternative in den jeweiligen Fällen

i) $H_1 : m \neq m_0$ oder

ii) $H_1 : m < m_0$ oder

iii) $H_1 : m > m_0$.

Da die Verteilungsfunktion stetig und streng monoton wachsend ist, folgt

$$P(X_i = m_0) = 0 \quad \text{für alle } i = 1, \dots, n.$$

Wir nehmen also $X_i \neq m_0$ für alle $i = 1, \dots, n$. Im Fall $x_i > m_0$ wählen wir das Symbol +, im Fall $X_i < m_0$ wählen wir das Symbol -. Es sei weiter

$$\begin{aligned} T^+ &= \text{Anz}(\{i \mid X_i > m_0\}) \\ T^- &= \text{Anz}(\{i \mid X_i < m_0\}) \quad \text{und} \\ R^+(X_i) &= \begin{cases} 1 & \text{falls } X_i > m_0 \\ 0 & \text{falls } X_i < m_0, \end{cases} \\ R^-(X_i) &= \begin{cases} 0 & \text{falls } X_i > m_0 \\ 1 & \text{falls } X_i < m_0. \end{cases} \end{aligned}$$

2.3.2 Bemerkung: Die Zufallsvariablen $R^+(X_1), \dots, R^+(X_n)$ und $R^-(X_1), \dots, R^-(X_n)$ sind jeweils unabhängig. Weiter gilt

$$T^+ = R^+(X_1) + \dots + R^+(X_n), \quad T^- = R^-(X_1) + \dots + R^-(X_n).$$

2.3.3 Bemerkung: Unter der Nullhypothese $H_0 : m = m_0$ gilt

- i) $P(X_i > m_0) = P(X_i < m_0) = \frac{1}{2}$.
- ii) $E(R^+(X_i)) = \frac{1}{2}$, $V(R^+(X_i)) = \frac{1}{4}$, $E(R^-(X_i)) = \frac{1}{2}$, und $V(R^-(X_i)) = \frac{1}{4}$ für alle $i = 1, \dots, n$.

Beweis: i) Da die Verteilungsfunktion der Zufallsvariablen X_i stetig ist und m_0 der Median ist folgt unmittelbar Aussage i).

ii) $R^+(X_i)$ ist die charakteristische Funktion der Menge $[X_i > m_0]$ und $R^-(X_i)$ die von $[X_i < m_0]$; daher ist die Aussage direkt nach zurechnen. \square

Aus dieser Bemerkung folgt unmittelbar die nachstehende Aussage:

2.3.4 Satz: Unter der Nullhypothese $H_0 : m = m_0$ besitzen die Zufallsvariablen T^+ und T^- eine Binomialverteilung mit den Parametern n und $p = \frac{1}{2}$. Folglich gilt

$$E(T^+) = E(T^-) = \frac{n}{2} \quad \text{und} \quad V(T^+) = V(T^-) = \frac{n}{4}.$$

(I) Wir diskutieren jetzt die Durchführung eines Tests der Nullhypothese $H_0 : m = m_0$ gegen die Alternative H_1 aus ii): $m < m_0$ mit einem Signifikanzwert oder **Quantil** α mit $0 < \alpha < 1$. Dazu bestimmen wir ein $g = g_L \in \{1, \dots, n\}$ mit

$$\begin{aligned} P(T^+ = 0) + \dots + P(T^+ = g) &\leq \alpha \quad \text{und} \\ P(T^+ = 0) + \dots + P(T^+ = g + 1) &> \alpha \quad \text{also mit} \\ \sum_{i=0}^g \binom{n}{i} 2^{-n} &\leq \alpha \quad \text{und} \quad \sum_{i=1}^{g+1} \binom{n}{i} 2^{-n} > \alpha. \end{aligned}$$

Die Hypothese H_0 wird dann im Fall $T^+ \leq g_L$ zugunsten der Alternative $m < m_0$ verworfen.

(II) Jetzt werde die Hypothese $H_0; m = m_0$ gegen die Alternative H_1 aus (iii): $m > m_0$ mit einem Signifikanzwert $\alpha : 0 < \alpha < 1$ getestet. Dazu bestimmen wir $g = g_R = \{1, \dots, n\}$ mit

$$\begin{aligned}
 P(T^+ = g) + \dots + P(T^+ = n) &\leq \alpha \quad \text{und} \\
 P(T^+ = g - 1) + \dots + P(T^+ = n) &> \alpha \quad \text{also mit} \\
 \sum_{i=g}^n \binom{n}{i} 2^{-n} &\leq \alpha \quad \text{und} \quad \sum_{i=g-1}^n \binom{n}{i} 2^{-n} > \alpha.
 \end{aligned}$$

Die Hypothese H_0 wird dann im Fall $T^+ \geq g_R$ zugunsten der Hypothese $m > m_0$ verworfen.

(III) Die Nullhypothese $H_0 : m = m_0$ werde gegen die Alternative H_1 aus (i): $m \neq m_0$ getestet, es sei weiter $\alpha : 0 < \alpha < 1$ der Signifikanzwert. Gemäß (I) und (II) bilden wir zu $\frac{\alpha}{2}$ die Werte $g_L = g_L(\frac{\alpha}{2})$ und $g_R = g_R(\frac{\alpha}{2})$ und verwerfen die Nullhypothese $H_0 : m = m_0$ zugunsten der Alternative $m \neq m_0$ in den Fällen $T^+ \geq g_R$ und $T^+ \leq g_L$.

2.3.5 Beispiel: Gegeben sei eine Population mit einer stetigen Verteilungsfunktion und einem eindeutig bestimmten Median. Gegeben sei die folgende Stichprobe vom Umfang $n = 15$:

$$\begin{aligned}
 &169, 1/168, 9/173, 4/175, 1/163, 5/167, 3/166, 9/170, 3/166, 7/ \\
 &165, 4/166, 1/168, 7/164, 9/169, 9/168, 5
 \end{aligned}$$

Es sei $\alpha = 0,10$ - der Signifikanzwert. Man erkennt unmittelbar $T^+ = 3$. Es werde die Nullhypothese H_0 zugunsten der Hypothese H_1 aus (i): $m \neq m_0 = 170$ getestet. Die kritische Region für T^+ ist

$$h = \{0, 1, 2, 3\} \cup \{12, 13, 14, 15\}; \quad \text{wegen}$$

$3 \in h$ wird demnach die Hypothese $H_0 : m = 170$ zugunsten der Hypothese $H_1 : m \neq 170$ verworfen.

Wir behandeln jetzt die folgenden Problemstellung: n Objekte werden vor und nach einer durchgeführten Maßnahme gemessen. Es soll geprüft werden, ob eine Verbesserung oder Verschlechterung durch diese Maßnahme erreicht wurde. Wir betrachten also die folgende Situation: Gegeben seien n unabhängige Paare von Messungen oder Zufallsvariablen.

$$(X_1, Y_1), \dots, (X_n, Y_n).$$

Wir bilden die Differenzen

$$D_1 = Y_1 - X_1, \dots, D_n = Y_n - X_n.$$

Diese Differenzen D_1, \dots, D_n sind ebenfalls unabhängig. Weiter setzen wir voraus, dass die Gesamtheit der Differenzen eine stetige Verteilungsfunktion mit einem eindeutig bestimmten Median m besitzt. Die Nullhypothese $H_0 : m = 0$ soll getestet werden zur Alternative.

- (i) $H_1 : m \neq 0$ oder
- (ii) $H_1 : m > 0$ oder
- (iii) $H_1 : m < 0$.

2.3.6 Beispiel: Das Signifikanzniveau sei $\alpha = 0,10$. Gegeben seien die Differenzen $D_i = Y_i - X_i$ für $i = 1, \dots, 10$.

10, 3/ - 0, 9/11, 2/12, 0/1, 6/11, 8/ - 1, 8/5, 4/ - 0, 5/ - 7, 4.

Es gilt offenbar $T^+ = 6$. Wegen $P(T^+ \geq 6 \mid H_0) = 0,377$ wird die Hypothese H_0 gegenüber der Alternative $H_1 : m > 0$ angenommen, obwohl der mittlere Anstieg $\bar{D} = 4,17$ ist.

Dieser Vorzeichentest hängt nicht vom Größenunterschied ab und ist daher ein sehr grober Test.

Im Fall, dass alle D_i normalverteilt sind, erhalten wir später mit der t -Verteilung (9 Freiheitsgrade) einen Test, bei dem die Nullhypothese abgelehnt wird.

Eine Verfeinerung der obigen Methode ist der folgende Test:

Der Signierte Rang Test von Wilcoxon

Voraussetzungen: Gegeben sei eine Population mit einer stetigen Verteilungsfunktion. Weiter nehmen wir an, dass diese Verteilungsfunktion in dem betrachteten Teilintervall $\subset \mathbb{R}$, in dem die betrachteten Werte liegen, streng monoton wachsend ist. Diese Verteilung besitzt dann dort einen eindeutig bestimmten Median m , der unbekannt sei. Weiter sei die Verteilungssymmetrie in Bezug auf den Median m . Die Aufgabe besteht wieder darin, einen angenommenen Wert m_0 zu testen.

Die Nullhypothesen $H_0 : m = m_0$ soll getestet werden gegen die

- i) Alternative $H_1 : m \neq m_0$ oder
- ii) Alternative $H_1 : m < m_0$ oder
- iii) Alternative $H_1 : m > m_0$.

Gegeben seien unabhängige Testgrößen X_1, \dots, X_n .

Wir setzen $D_i : X_i - m_0$ für $i = 1, \dots, n$.

2.3.7 Bemerkung: Unter den obigen Voraussetzungen gilt:

- i) $P(D_i = a) = 0$ für $i = 1, \dots, n$ und alle $a \in \mathbb{R}$.
- ii) $P(|D_i| = |D_j|) = 0$ für alle $i \neq j$, $i, j = 1, \dots, n$.

Beweis: i) Da die Verteilungsfunktion F_X stetig und in dem betrachteten Intervall streng monoton wachsend ist, gilt $P(X_i = a) = 0$ für alle $i \in \{1, \dots, n\}$ und $a \in \mathbb{R}$. Damit ist i) gezeigt worden.

ii) Es sei $A = \{(s, t) \in \mathbb{R}^2 \mid s = t \text{ oder } s = 2m_0 - t\}$. Wegen i) gilt dann für alle $i \neq j$ die Beziehung

$$\begin{aligned} P(|D_i| = |D_j|) &= P((X_i, X_j) \in A) \\ &= P_{X_i} \otimes P_{X_j}(A) = \int \int \mathbb{1}_A(s, t) dP_{X_i}(t) dP_{X_j}(s) \\ &= \int [P_{X_i}(\{s\}) + P_{X_i}(-s + 2m_0)] dP_{X_j}(s) = 0. \end{aligned}$$

□

Als Konsequenz dieser Aussage können wir annehmen, dass $|D_i|$ $i = \{1, \dots, n\}$ paarweise verschieden sind. Wir ordnen die $|D_i|$ der Größe nach, die Rangzahlen seien $R(1), \dots, R(n)$.

Es gilt etwa $R(i) = 1$, falls $|D_i|$ das kleinste der Element ist. Demnach ist die Abbildung $R : \{1, \dots, \} \rightarrow \{1, \dots, n\}$ bijektiv. Es sei weiter

$$T^+ = \sum_{D_i > 0} R(i), \quad T^- = \sum_{D_i < 0} R(i).$$

Wegen der Bijektivität von R erhalten wir unmittelbar

$$T^+ + T^- = \frac{n(n+1)}{2}$$

als Summe der ersten n natürlichen Zahlen. Der Test besteht jetzt darin, dass für große oder kleine T^+ , die Nullhypothese verworfen wird. Diese Eigenschaft muss jetzt noch quantifiziert werden. Man beachte ferner, dass die Abbildung R invertierbar ist.

2.3.8 Satz: *Unter der Nullhypothese H_0 $m = m_0$ erhalten wir:*

- i) $P(D_i > 0) = P(D_i < 0) = \frac{1}{2}$ für alle $i = 1, \dots, n$.
- ii) Die Zufallsvariablen T^+ und T^- besitzen die gleiche Verteilungsfunktion.
- iii) Mit der Abkürzung $m = \frac{n(n+1)}{2}$ folgt

$$P(T^+ = k) = P(T^+ = m - k) \text{ für alle } 0 \leq k \leq m.$$

- iv) $E(T^+) = E(T^-) = \frac{n(n+1)}{4}$.
- v) $V(T^+) = \frac{n(n+1)(2n+1)}{24}$.

Beweis: i) und ii) folgen unmittelbar aus der Definition.

iii) Wegen $T^+ + T^- = m$ gilt

$$[T^+ = k] = [T^- = m - k] \text{ für alle } 0 \leq k \leq m.$$

Aussage ii) impliziert daher $P(T^+ = k) = P(T^- = k)$ für alle $k = 1, \dots, n$; damit gilt weiter $P(T^+ = k) = P(T^- = m - k) = P(T^+ = m - k)$.

iv) Wegen ii) gilt $E(T^+) = E(T^-)$; aus $T^+ + T^- = m$ folgt daher $E(T^+) = \frac{m}{2} = \frac{n(n+1)}{4}$.

v) Wir setzen $B_i = \mathbb{1}_{[D_i > 0]}$ für alle $i = 1, \dots, n$; dann gilt offenbar

$$P(B_i = 1) = P(B_i = 0) = \frac{1}{2} \text{ und}$$

$$T^+ = \sum_{i=1}^n R(i)B_i = \sum_{i=1}^n iB_{R^{-1}(i)}.$$

Wegen der Voraussetzung sind die Zufallsvariablen B_1, \dots, B_n unabhängig, aus der obigen Darstellung folgt weiter

$$V(T^+) = \sum_{k=1}^n V(kB_{R(k)}) = \sum_{k=1}^n k^2 V(B_{R(k)}) = \sum_{k=1}^n \frac{1}{4} \cdot k^2 = \frac{1}{4} \frac{n(n+1)(2n+1)}{6},$$

was zu zeigen war. □

Anmerkung: Für große $n \in \mathbb{N}$ ist die Statistik T^+ angenähert normalverteilt mit den Parametern $\mu = \frac{n(n+1)}{4}$ und $\sigma^2 = \frac{n(n+1)(2n+1)}{24}$.

Wir diskutieren jetzt noch einmal das Beispiel 2.3.5

2.3.9 Beispiel: Es sei

Nr.	1	2	3	4	5	6	7	8	9	10
D_i	10.3	-0.9	11.2	12.0	1.6	11.8	-1.8	5.4	-0.5	-7.4
R_i	7	2	8	10	3	9	4	5	1	6

Wir haben $T^+ = 7 + 8 + 10 + 3 + 9 + 5 = 42$. Weiter sei die Vertrauensgrenze $\alpha = 0,10$, das kritische Gebiet für T^+ besteht aus $\{41, \dots, 55\}$ - die Nullhypothese wird daher abgelehnt. Dabei wurde die Nullhypothese $H_0 : m = 0$ gegen die Alternative $H_1 : m > 0$ getestet.

Der Runs Test

Der so genannte Runs-Test dient dazu Zufälligkeiten bei der Anordnung zweier unterschiedlicher Symbole zu untersuchen. Man kann diese Situation mit der n -fachen Durchführung von Bernoulli-Experimenten vergleichen; im Gegensatz zur Binomialverteilung werden aber nicht die Erfolge gezählt, sondern es wird auf die Anordnung geachtet. Gegeben seien also

n_1 Symbole a und n_2 Symbole b , die in irgendeiner Reihenfolge angeordnet sind.

Ein a -Run besteht aus einem maximalen ununterbrochenen Teil dieser Anordnung, der nur aus den Symbolen a besteht; T_1 sei die Anzahl der a -Runs und T_2 die Anzahl alle b -Runs. Wir setzen $T = T_1 + T_2$ und erhalten unmittelbar $|T_1 - T_2| \leq 1$.

Weiter beachte man, dass die Zahl aller möglichen Anordnungen der Elemente a und b genau $\binom{n_1 + n_2}{n_1} = \binom{n_1 + n_2}{n_2}$ beträgt.

Wir wollen die Nullhypothese testen:

H_0 : Alle Anordnungen erscheinen unter der gleichen Wahrscheinlichkeit.

gegen eine der folgenden Alternativen:

i) H_1 : Die Anordnung der Symbole a und b ist nicht zufällig.

ii) H_1 : Die Anordnung tendiert zur Cluster-Bildung.

iii) H_1 : Die Anordnung tendiert zu einer starken Vermischung.

Die Funktion T , die oben definiert wurde, soll dabei als Teststatistik dienen. Dazu bestimmen wir zunächst die Zähldichte des zugeordneten Wahrscheinlichkeitsmaßes; dabei setzen wir die Gültigkeit der Nullhypothese voraus. Die Gesamtzahl aller möglichen Anordnungen der Symbole a und b beträgt $\binom{n_1 + n_2}{n_1} = \binom{n_1 + n_2}{n_2}$. Wir bestimmen jetzt $P(T = m)$ für alle $m \in \mathbb{N}_0$ mit $m \leq n = n_1 + n_2$. Wir unterscheiden die Fälle $m = 2k$ und $m = 2k + 1$ für ein $k \in \mathbb{N}_0$. In beiden Fällen muss $k \leq \min\{n_1, n_2\}$ gelten.

Der Fall $m = 2k + 1$: Wir nehmen an, dass mit dem Symbol a begonnen wird. Dann liegen $k + 1$ a -Runs vor und k b -Runs. (Im Fall des Beginns mit b ist es genau umgekehrt.) Zur Bestimmung der Anzahl der Möglichkeiten die Menge $\{1, \dots, n_1\}$ in genau $k + 1$ Teilmengen aufeinanderfolgender Elemente aufzuteilen, identifizieren wir jede dieser Teilmengen mit ihrem ersten Element und erhalten so Zahlen $r(1) = 1 < r(2) < \dots < r(k + 1) \leq n_1$. Da $r(1) = 1$ fest gewählt wurde, ergibt sich als die gesuchte Anzahl $\binom{n_1 - 1}{k}$ Möglichkeiten der

Zerlegung. Entsprechend ergeben sich $\binom{n_2-1}{k-1}$ Möglichkeiten der Zerlegung für b . Es folgt

$$P(T = 2k + 1) = \frac{\binom{n_1-1}{k} \binom{n_2-1}{k-1} + \binom{n_2-1}{k} \binom{n_1-1}{k-1}}{\binom{n_1+n_2}{n_1}}$$

wegen der oben erwähnten Symmetrie.

Der Fall $m = 2k$: Wie eben erhalten wir

$$P(T = 2k) = \frac{2 \binom{n_1-1}{k-1} \binom{n_2-1}{k-1}}{\binom{n_1+n_2}{n_1}}.$$

Ohne den technisch aufwendigeren Beweis durchzuführen notieren wir:

2.3.10 Satz: i) *Unter der Nullhypothese H_0 gilt*

$$E(T) = \frac{2n_1n_2+1}{n_1+n_2} + 1 \quad \text{und} \quad V(T) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}.$$

ii) *Unter der Nullhypothese H_0 konvergiert die standardisierte Verteilungsfunktion von T gegen die Standard-Normalverteilung bei $n_1 + n_2 \rightarrow \infty$.*

Durchführung des Testes: Bei der Durchführung hängt es wesentlich von der Augenstellung ab, welche der drei Alternativen abgetestet werden sollen. Wird beispielsweise Cluster-Bildung als Alternative abgetestet, so wird die Nullhypothese verworfen, wenn T besonders kleine Werte annimmt. Ist also $\alpha \in]0, 1[$ das Quantil – etwa $\alpha = 0.1$, so wird $m \in \mathbb{N}_0$ bestimmt mit

$$\sum_{k=0}^m P(T = k) \leq \alpha \quad \text{und} \quad \sum_{k=0}^{m+1} P(T = k) > \alpha.$$

Im Fall $T \leq m$ wird dann die Nullhypothese abgelehnt und stattdessen die Cluster-Bildung angenommen.

2.4 Mit der Normalverteilung zusammenhängende Tests

Das Ziel dieses Abschnittes ist es einige spezielle Testverfahren zu behandeln, die mit der Normalverteilung und mit einigen aus dieser konstruierten Verteilungen zusammenhängen; aus der Vielzahl wohl bekannter Tests werden hier nur einige wenige vorgestellt um Prinzipien der Testtheorie zu diskutieren; dabei wird natürlich ein Anspruch auf Vollständigkeit schon wegen des enormen Umfangs der Testtheorie nicht erhoben. Diese werden sehr ausführlich etwa in dem Buch von W.R. Pestman behandelt.

In der Regel sind die in diesen Abschnitt diskutierten Tests aussagekräftiger als die vorher behandelten nichtparametrischen Tests aber dafür weniger allgemein anwendbar. Die hier diskutierten Tests beruhen auf Verteilungen, die in Abschnitt 1.6 behandelt wurden. Im Unterschied zu den im letzten Abschnitt behandelten Tests werden in dem jetzigen Abschnitt die Testbedingungen an Hand der Dichten direkt mit der Methode der so genannten Likelihood-Quotienten bestimmt und nicht wie im letzten Abschnitt aus abgeleiteten Eigenschaften. Bevor wir mit der eigentlichen Problemstellung beginnen, behandeln wir zunächst noch einige weitere allgemeine Aspekte der Testtheorie, die wir im ersten Abschnitt noch nicht angesprochen haben. Dazu gehören insbesondere einige Aspekte der Gütebestimmung von Tests.

Es sei also $(\mathcal{X}, \mathcal{F}, \Theta, P_\vartheta \mid \vartheta \in \Theta)$ ein statistisches Modell und $\varphi : \mathcal{X} \rightarrow [0, 1]$ eine Statistik. Man beachte, dass im Fall $\varphi(x) = 0$ an der Nullhypothese festgehalten wird und im Fall $\varphi(x) = 1$ die Alternative gewählt wird. Die Funktion

$$\Theta \ni \vartheta \mapsto G_\varphi(\vartheta) := E_\vartheta(\varphi) \in [0, 1]$$

heißt **Güte-Funktion** des Tests. Wir setzen

$$\alpha(\varphi) = \sup \{ G_\varphi(\vartheta) \mid \vartheta \in \Theta_0 \}$$

und bezeichnen $\alpha(\varphi)$ auch als maximale Wahrscheinlichkeit für einen **Fehler erster Art** oder Alpha-Fehler: Das Vorliegen der Alternative wird angenommen, obwohl die Null-Hypothese vorliegt. Weiter bezeichnen wir für $\vartheta \in \Theta_1$ mit $G_\varphi(\vartheta)$ die **Macht** oder auch **Schärfe** des Tests φ bei ϑ – die Wahrscheinlichkeit mit der die Alternative erkannt wird, wenn sie vorliegt und setzen weiter $\beta_\varphi(\vartheta) = 1 - G_\varphi(\vartheta)$. Demnach gilt

$$\beta(\varphi) = \sup \{ 1 - G_\varphi(\vartheta) \mid \vartheta \in \Theta_1 \}$$

als maximale Wahrscheinlichkeit für einen **Fehler zweiter Art** oder Beta-Fehler: Das Vorliegen der Alternative wird nicht erkannt, und die Null-Hypothese wird fälschlicherweise akzeptiert. Es sei angemerkt, dass aus den gemachten Voraussetzungen zumindest nicht unmittelbar folgt, dass α und β Maße im üblichen Sinn sind. Damit ergeben sich die folgenden

Forderungen an einen Test φ :

$G_\varphi(\vartheta) \leq \alpha$ für alle $\vartheta \in \vartheta_0$ – die Wahrscheinlichkeit für einen Fehler 1. Art ist $\leq \alpha$.

$G_\varphi(\vartheta)$ soll maximal sein für alle $\vartheta \in \vartheta_1$ – ein Fehler 2. Art soll eine möglichst kleine Wahrscheinlichkeit besitzen.

Ein Test φ von Θ_0 gegen Θ_1 zum Niveau $\alpha \in]0, 1[$ heißt ein **bester Test zum Niveau α** , wenn für jeden weiteren Test ψ von Θ_0 gegen Θ_1 zum Niveau α stets $G_\varphi(\vartheta) \geq G_\psi(\vartheta)$ für

alle $\vartheta \in \Theta_1$ gilt, wenn also der Fehler zweiter Art am kleinsten ist.

Weiter nennen wir einen Test φ von Θ_0 gegen Θ_1 zum Niveau α **unverfälscht** genau dann, wenn

$$G_\varphi(\vartheta_0) \leq G_\varphi(\vartheta_1) \text{ für alle } (\vartheta_0, \vartheta_1) \in \Theta_0 \times \Theta_1$$

gilt: man entscheidet sich mit größerer Wahrscheinlichkeit für die Alternative H_1 , wenn diese richtig ist, als wenn diese falsch ist.

Im Fall $\text{Anz}(\vartheta_0) = \text{Anz}(\vartheta_1) = 1$ heißt $(\mathcal{X}, \mathcal{F}, P_0, P_1)$ ein **einfaches statistisches Modell**.

Für die weiteren Überlegungen betrachten wir die folgende Situation eines einfachen statistischen Modells $(\mathcal{X}, \mathcal{F}, P_0, P_1)$, so dass zusätzlich die Wahrscheinlichkeitsmaße P_0 und P_1 durch geeignete Zähldichten oder aber alternativ durch geeignete Dichte-Funktionen auf \mathcal{X} gegeben sind. In beiden Fällen bezeichnen wir diese mit ρ_0 und ρ_1 und können durch etwaige Verkleinerung von \mathcal{X} erreichen, dass $\rho_0 + \rho_1 > 0$ auf \mathcal{X} gilt. Gesucht wird ein bester Test φ von P_0 gegen P_1 zu einem festen Niveau α . Dazu definieren wir den **Likelihood-Quotienten**

$$R(x) = \frac{\rho_1(x)}{\rho_0(x)} \text{ für } x \in \mathcal{X}$$

mit der Verabredung $\frac{a}{0} = \infty$ für alle $a > 0$.

2.4.1 Satz: Neyman-Pearcy Für jedes $0 < \alpha < 1$ gilt unter den obigen Voraussetzungen

i) Ein bester Test ψ von $H_0 : P = P_0$ gegen die Alternative $H_1 : P = P_1$ hat die Gestalt

$$\psi(x) = \begin{cases} 1 & \text{für } R(x) > c \\ 0 & \text{für } R(x) < c \end{cases}$$

für ein geeignetes $c = c(\alpha) \geq 0$. Derartige Tests werden als **Neyman-Pearcy-Tests** bezeichnet.

ii) Es existiert ein Neyman-Pearcy-Test φ mit $E_0(\varphi) = \alpha$.

iii) Jeder Neyman-Pearcy-Test φ mit $E_0(\varphi) = \alpha$ ist ein bester Test zum Niveau α .

Beweis: Wir nehmen an, dass stetige Dichten vorliegen. Der andere Fall von Zähldichten kann entsprechend behandelt werden.

ii) $c \in \mathbb{R}$ sei ein α -Fraktile von $P_0 \circ R^{-1}$ in dem Sinn, dass $P_0(R \geq c) \geq \alpha$ und $P_0(R > c) \leq \alpha$ gilt. Damit erhalten wir

$$\alpha - P_0(R > c) \leq P_0(R \geq c) - P_0(R > c) = P_0(R = c).$$

Wir setzen weiter $\gamma = 0$ im Fall $P_0(R = c) = 0$ und $\gamma = \frac{\alpha - P_0(R > c)}{P_0(R = c)}$ im Fall $P_0(R = c) \neq 0$. Daher definieren wir die Teststatistik φ durch

$$\varphi(x) = \begin{cases} 1 & \text{für } R(x) > c \\ 0 & \text{für } R(x) < c \\ \gamma & \text{für } R(x) = c \end{cases} .$$

Mit dieser Festsetzung gilt offenbar $E_0(\varphi) = P_0(R > c) + P_0(R = c) \cdot \gamma = \alpha$.

iii) Es sei jetzt φ ein Neyman-Pearcy-Test mit $E_0(\varphi) = \alpha$ zum Schwellenwert c und ψ ein beliebiger Test zum Niveau α . Wir setzen jetzt

$$f = (\rho_1 - \rho_0 c)(\varphi - \psi) .$$

Im Fall $R(x) = \frac{\rho_1(x)}{\rho_0(x)}$ folgt $(\rho_1(x) - \rho_0(x)c) > 0$ und $\varphi(x) = 1$. Wegen $\psi(x) \leq 1$ gilt daher $f(x) \geq 0$.

Im Fall $R(x) = c$ gilt $f(x) = 0$.

Im Fall $R(x) < c$ gilt $\varphi(x) = 0$ und $(\rho_1(x) - \rho_0(x)c) < 0$, also $f(x) \geq 0$. Wegen $E_0(\varphi) = \alpha$ und $E_0(\psi) \leq \alpha$ erhalten wir damit

$$\begin{aligned} 0 \leq \int_{\mathcal{X}} f(x)dx &= E_1(\varphi - \psi) - cE_0(\varphi - \psi) \\ &= E_1(\varphi) - E_1(\psi) - c(E_0(\varphi) - E_0(\psi)) \leq E_1(\varphi) - E_1(\psi) . \end{aligned}$$

φ ist also ein bester Test zum Niveau α .

i) mit den Bezeichnungen des Beweises von iii) sei ψ ein beliebiger Test zum Niveau α . Dann gilt offenbar $E_1(\varphi) = G_\varphi(1) = G_\psi(1) = E_1(\psi)$. Aus

$$0 \leq \int_{\mathcal{X}} f(x)dx \leq E_1(\varphi) - E_1(\psi) = 0$$

und $f(x) \geq 0$ für alle $x \in \mathcal{X}$ folgt $f = 0$ fast überall, wir erhalten demnach $\varphi = \psi$ auf der Menge $[R < c] \cup [R > c]$, was zu zeigen war. □

Bevor wir die mit der Normalverteilung zusammenhängenden Tests behandeln zeigen wir noch eine Aussage über die Macht des Tests bei einem Produktmodell einfacher alternativen. Dazu benötigen wir die folgende Begriffsbildung:

2.4.2 Bemerkung: Gegeben sei ein einfaches statistisches Standardmodell $(\mathcal{X}, \mathcal{F}, P_0, P_1)$, die Wahrscheinlichkeitsmaße P_0 und P_1 seien also durch geeignete Zähl-dichten oder aber alternativ durch geeignete Dichte-Funktionen auf \mathcal{X} gegeben. Wir definieren die relative Entropie $H(P_0; P_1)$ durch $H(P_0; P_1) = \infty$ im Fall $P_0(\rho_1 = 0) > 0$ und im Fall $P_0(\rho_1 = 0) = 0$ sei

$$H(P_0; P_1) := E_{P_0} \left(\ln \left(\frac{\rho_0}{\rho_1} \right) \right) \in [0, \infty]; .$$

Diese Entropie ist widerspruchsfrei definiert, und es gilt $H(P_0; P_1) \geq 0$ und $H(P_0; P_1) = 0$ ist äquivalent zu $\rho_0 = \rho_1$ fast überall.

Anmerkung: Bezeichnen wir in der Formulierung der obigen Bemerkung mit f die Funktion $f(t) = \ln \left(\frac{\rho_0(t)}{\rho_1(t)} \right)$ für $t \in \mathbb{R}$, so sei $f^-(t) = \max\{-f(t), 0\}$ und $f^+(t) = \max\{f(t), 0\}$. Dann gilt stets $f^- \in \mathcal{L}^1(P_0)$, folglich ist der Erwartungswert $E_{P_0} \left(\ln \left(\frac{\rho_0}{\rho_1} \right) \right) \in [0, \infty]$ stets wohldefiniert. Dieser Zusatz folgt unmittelbar mit den Methoden des folgenden Beweises, wenn statt über \mathcal{X} über die Menge $f[f < 0]$ integriert wird.

Beweis: Wir zeigen diese Aussage im Fall der Existenz von Dichten. Wir können weiter voraussetzen, dass $P_0(\rho_1 = 0) = 0$ und $\rho_0 > 0$ auf \mathcal{X} gilt. Die Funktion $g :]0, \infty[\rightarrow \mathbb{R}$ sei definiert durch

$$g(s) = 1 - s + s \ln(s).$$

Wegen $g'(s) = \ln(s)$ und $g''(s) > 0$ besitzt g in $s = 1$ ein Minimum. Daher gilt speziell $0 = g(1) \leq g(s)$ für alle $s > 0$. Es folgt daher

$$\begin{aligned} 0 \leq E_{P_1} \left(g \circ \left(\frac{\rho_0}{\rho_1} \right) \right) &= \int_{\mathcal{X}} \left(1 - \frac{\rho_0(x)}{\rho_1(x)} + \frac{\rho_0(x)}{\rho_1(x)} \ln \left(\frac{\rho_0(x)}{\rho_1(x)} \right) \right) \rho_1(x) dx \\ &= 1 - \int_{\mathcal{X}} \frac{\rho_0(x)}{\rho_1(x)} \rho_0(x) dx + \int_{\mathcal{X}} \frac{\rho_0(x)}{\rho_1(x)} \ln \left(\frac{\rho_0(x)}{\rho_1(x)} \right) \rho_1(x) dx \\ &= 1 - 1 + \int_{\mathcal{X}} \ln \left(\frac{\rho_0(x)}{\rho_1(x)} \right) \rho_0(x) dx = H(P_0; P_1). \end{aligned}$$

Es sei jetzt $H(P_0; P_1) = 0$. Wegen der eben hergeleiteten Beziehung gilt $0 = H(P_0; P_1) = E_{P_1} \left(g \circ \left(\frac{\rho_0}{\rho_1} \right) \right)$. Damit folgt $P_1 \left(g \left(\frac{\rho_0}{\rho_1} \right) = 0 \right) = P_1 \left(\left(\frac{\rho_0}{\rho_1} \right) = 1 \right)$ und damit $\rho_0 = \rho_1$. \square

Für ein unendliches Produktmodell gemäß 1.5.1 eines Standardmodells mit einfachen Alternativen soll jetzt die Macht des Neyman-Pearcy Tests in Anhängigkeit vom Umfang der Stichprobe beschrieben werden. Dazu sei $(\mathcal{E}, \mathcal{G}, Q_0, Q_1)$ ein einfaches Standardmodell. Es sei

$$(\mathcal{X}, \mathcal{F}, P_0, P_1) = (\mathcal{E}^{\mathbb{N}}, \mathcal{G}^{\otimes \mathbb{N}}, Q_0^{\otimes \mathbb{N}}, Q_1^{\otimes \mathbb{N}})$$

das unendliche Produkt des Standardmodells. Weiter sei $X_j : \mathcal{X} \rightarrow \mathcal{E}$ die Projektion auf die j -te Komponente. Die Dichten ρ_0 und ρ_1 des ursprünglichen Modells seien beide strikt positiv.

2.4.3 Satz: *Unter den obigen Voraussetzungen sei $\alpha \in]0, 1[$. Zu jedem $n \in \mathbb{N}$ sei φ_n ein Neyman-Pearcy Test zum Niveau α , der nur von den ersten n Werten der X_1, \dots, X_n abhängt. Dann gilt $E_{P_1}(\varphi_n) \rightarrow 1$ bei $n \rightarrow \infty$ exponentiell im Sinne von*

$$\begin{aligned} \frac{1}{n} \ln(1 - E_{P_1}(\varphi_n)) &\rightarrow -H(Q_0; Q_1) \text{ bei } n \rightarrow \infty \text{ im Fall } H(Q_0; Q_1) < \infty \text{ und} \\ \lim_{n \rightarrow \infty} \frac{1}{n} \ln(1 - E_{P_1}(\varphi_n)) &\leq -m \text{ für alle } m \in \mathbb{N} \text{ im Fall } H(Q_0; Q_1) = \infty. \end{aligned}$$

Beweis: Für alle $n \in \mathbb{N}$ sei $\rho_{0,n} = \rho_0(x_1) \cdot \dots \cdot \rho_0(x_n)$ die n -fache Produktdichte von ρ_0 , und entsprechend sei $\rho_{1,n}$ definiert. Weiter sei $R_n = \frac{\rho_{1,n}}{\rho_{0,n}}$ der Likelihoodquotient der Stufe $n \in \mathbb{N}$. Zur Abkürzung setzen wir $f = \ln \left(\frac{\rho_0}{\rho_1} \right)$, und $f_n :$

$$f_n(x_1, \dots, x_n) = -\frac{1}{n} \ln(R_n(x)) = \frac{1}{n} \sum_{j=1}^n f(x_j).$$

Es gelte zunächst $f \in \mathcal{L}^1(P_1)$, aus der Variante des Gesetzes der großen Zahl 1.5.5 folgt dann wegen der stochastischen Konvergenz gemäß 1.5.4

$$P_0(|f_n - E_{P_0}(f)| \geq c) \rightarrow 0 \text{ bei } n \rightarrow \infty \text{ für alle } c > 0.$$

Mit den Bezeichnungen von 2.4.2 gilt $E_0(f) = H(P_0, P_1)$.

Es sei $\varepsilon > 0$ und $a = E_{P_0}(f) - \varepsilon$. Als Neyman-Pearcy Test besitzt die Teststatistik φ_n die Gestalt $\varphi_n = 1$ im Fall $R_n > c_n$ und $\varphi_n = 0$ im Fall $R_n < c_n$ für geeignete $c_n \in]0, \infty[$. Setzen wir $a_n = -\ln(c_n)$, so folgt $\varphi_n = 1$ im Fall $f_n < a_n$, $\varphi_n = 0$ im Fall $f_n > a_n$, und $P_0(f_n \geq a_n) \geq E_{P_0}(\varphi_n) = \alpha$ als Test zum Niveau α .

Daher gilt $a_n > a = E_{P_0}(f) - \varepsilon$ für fast alle $n \in \mathbb{N}$ wegen

$$P_0(f_n \leq a) \leq P_0(|f_n - E_{P_0}(f)| \geq a) \rightarrow 0 \text{ bei } n \rightarrow \infty .$$

Aus $\mathbb{1}_{[(1-\varphi_n)>0]} \left(\frac{\rho_{1,n}}{\rho_{0,n}}\right) = \mathbb{1}_{[f_n \geq a_n]} \left(\frac{\rho_{1,n}}{\rho_{0,n}}\right) \leq e^{-na_n} \mathbb{1}_{[f_n \geq a_n]}$ folgt damit weiter

$$\begin{aligned} \frac{1}{n} \ln(1 - E_{P_1}(\varphi_n)) &= \frac{1}{n} \ln(E_{P_1}(1 - \varphi_n)) = \frac{1}{n} \ln\left(E_{P_0}\left((1 - \varphi_n) \frac{\rho_{1,n}}{\rho_{0,n}}\right)\right) \\ &\leq \frac{1}{n} \ln\left(e^{-na_n} E_{P_0}((1 - \varphi_n))\right) \\ &= -a_n + \frac{1}{n} \ln(1 - \alpha) < -E_{P_0}(f) + \varepsilon \end{aligned}$$

für fast alle $n \in \mathbb{N}$.

Es sei jetzt umgekehrt $\varepsilon > 0$ und $a = E_{P_0}(f) + \frac{\varepsilon}{2}$. Dann folgt

$$P_0(f_n \leq a) = P_0(f_n - E_{P_0}(f) \leq \frac{\varepsilon}{2}) \geq P_0(|f_n - E_{P_0}(f)| \leq \frac{\varepsilon}{2}) \rightarrow 1 \text{ bei } n \rightarrow \infty .$$

Wegen $e^{-nf_n} = \frac{\rho_{1,n}}{\rho_{0,n}} \geq \mathbb{1}_{[f_n \leq a]} e^{-na}$ für alle $n \in \mathbb{N}$ gilt

$$\begin{aligned} \frac{1}{n} \ln\left(E_{P_1}(1 - \varphi_n)\right) &= \frac{1}{n} \ln\left(E_{P_0}\left((1 - \varphi_n) \frac{\rho_{1,n}}{\rho_{0,n}}\right)\right) \\ &\geq \frac{1}{n} \ln\left(E_{P_0}\left((1 - \varphi_n) \mathbb{1}_{[f_n \leq a]} e^{-na}\right)\right) \\ &\geq \frac{1}{n} \ln\left(e^{-na} E_{P_0}(\mathbb{1}_{[f_n \leq a]} - \varphi_n)\right) \\ &= -a + \frac{1}{n} \ln(P_0(f_n \leq a) - \alpha) \\ &\rightarrow -a \text{ bei } n \rightarrow \infty ; \end{aligned}$$

es folgt daher $\frac{1}{n} \ln\left(E_{P_1}(1 - \varphi_n)\right) \geq -E_{P_0}(f) - \varepsilon$ für fast alle $n \in \mathbb{N}$, und mit dem ersten Teil des Beweises erhalten wir $\left|\frac{1}{n} \ln\left(E_{P_1}(1 - \varphi_n)\right) + E_{P_0}(f)\right| \leq \varepsilon$ für fast alle $n \in \mathbb{N}$ im Fall $f \in \mathcal{L}^1(P_0)$. Der Fall $f \notin \mathcal{L}^1(P_0)$ kann wie im ersten Teil des Beweises behandelt werden; auf die explizite Durchführung soll verzichtet werden. \square

2.4.4 Beispiel: Es soll ein übersichtlicher Spezialfall des vorstehenden Satzes diskutiert werden: Dabei gehen wir von zwei Normalverteilungen mit der gleichen Streuung $\sigma = 1$ aus. Daher unterscheiden sich diese zwei Verteilungen nur durch die unterschiedlichen Erwartungswerte. Wir können also annehmen, dass $\mu_0 = 0$ und $\mu_1 = \mu > 0$ gilt. Damit ergibt sich

für den Likelihoodquotienten der Stufe $n \in \mathbb{N}$

$$\begin{aligned} R_n(x) &= \frac{\rho_{1,n}(x)}{\rho_{0,n}(x)} = \frac{\exp\left(-\frac{1}{2}((x_1 - \mu)^2 + \dots + (x_n - \mu)^2)\right)}{\exp\left(-\frac{1}{2}(x_1^2 + \dots + x_n^2)\right)} \\ &= \exp\left(-\frac{1}{2}((-2\mu x_1 + \mu^2) + \dots + (2\mu x_n - \mu^2))\right) = k_1 \exp(\mu(x_1 + \dots + x_n)) \end{aligned}$$

für alle $x = (x_1, \dots, x_n)$ und ein geeignetes $k_1 > 0$. Es sei jetzt $0 < \alpha < 1$, und es werde ein $k_2 > 0$ bestimmt mit $P_{0,n}(R_n \geq k_2) = \alpha$, wenn $P_{0,n}$ das n -fache Produktmaß $P_0 \otimes \dots \otimes P_0$ ist. Wegen der vorstehenden Ungleichung gilt

$$R_n \geq k_2 \iff (x_1 + \dots + x_n) \geq \frac{1}{\mu} (\ln(k_2) - \ln(k_1)) .$$

Wir müssen daher ein $k \in \mathbb{R}$ bestimmen mit $P_{0,n}(A) = \alpha$ mit $A = [(x_1 + \dots + x_n) \geq k]$.

Zur Berechnung der Wahrscheinlichkeiten gemäß 1.6.2 bestimmen wir eine beliebige orthogonale Matrix U mit der ersten Zeile $U_1^Z = \frac{1}{\sqrt{n}}(1, \dots, 1)$. Für jedes $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ sei $y = (y_1, \dots, y_n)^T = Ux$ und $c_0 = \frac{1}{\sqrt{n}}c$; dann folgt $y_1 = \frac{1}{\sqrt{n}}(x_1 + \dots + x_n)$ und wegen $U^T U = I$ weiter

$$\begin{aligned} P_{0,n}(A) &= \int_A \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}x^T x\right) dx = \int_{[y_1 \geq c_0]} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}y^T y\right) dy \\ &= \int_{\mathbb{R}} \left(\dots \int_{\mathbb{R}} \left(\int_{k_0}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}y^T y\right) dy_1\right) dy_2 \dots\right) dy_n = 1 - \Phi_{0,1}(k_0) . \end{aligned}$$

Dabei sei wie üblich $\Phi_{0,1}$ die Verteilungsfunktion der Standard-Normalverteilung. Entsprechend erhalten wir unter der Berücksichtigung der Koordinatenverschiebung gemäß 1.6.2

$$P_{1,n}(A^c) = \Phi_{\mu,1}(k_0 - \sqrt{n} \mu) = \Phi_{0,1}(k_0 - \sqrt{n} \mu) .$$

Es ist zu erwähnen, dass wir das gleiche Resultat auch erhalten mit einer vereinfachten direkten Modellbildung: Sind die unabhängigen Zufallsvariablen $N(0, 1)$ -verteilt, so ist die Zufallsvariable $Y = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$ wieder $N(0, 1)$ -verteilt. Sind die Zufallsvariablen hingegen $N(\mu, 1)$ -verteilt, so ist Y $N(\sqrt{n} \mu, 1)$ -verteilt. Mit wachsendem $n \in \mathbb{N}$ werden die Verteilungen also wesentlich getrennt.

Wir fixieren jetzt einen α -Wert als vorgegebene Wahrscheinlichkeit für einen Fehler erster Art und bestimmen dann bei einigen n -Werten die zugeordneten β -Fehler.

$\alpha = 0.05 :$	n		5	10	20
	β		0.276	0.058	0.002
$\alpha = 0.01 :$	n		5	10	20
	β		0.533	0.201	0.016

Bei dieser sehr übersichtlichen Situation liegt auch eine andere Wahl der Grenze c für den Likelihood-Quotienten nahe als bei der eben geschilderten üblichen Vorgehensweise. Wir

wählen etwa $R(x) = 1$. Diese Vorgehensweise hat als Konsequenz, dass hier in dieser Situation die beiden Fehler: der ersten und der zweiten Art, gleich groß werden. Dazu bestimmen wir $\alpha, z \in \mathbb{R}$ mit

$$\Phi(z) = 1 - \alpha, \quad \Phi(z - \sqrt{n}) = \alpha,$$

so folgt $2z = \sqrt{n}$. Wir erhalten die folgenden Zahlenwerte:

n	6	10	12	20	30
$\alpha = \beta$	0.111	0.057	0.042	0.013	0.003

Wir wenden uns jetzt der ursprünglich vorgesehenen Aufgabenstellung der Behandlung einige mit der Normalverteilung zusammenhängender Tests zu beginnen mit dem folgenden Problem:

Testen des Mittelwertes bei bekannter Streuung $\sigma_0 > 0$:

Es seien also x_1, \dots, x_n eine Stichprobe einer $N(\mu, \sigma_0)$ -verteilten Population. Dann gilt offenbar

$$\Theta = \{(\mu, \sigma_0) \in \mathbb{R}^2 : \mu \in \mathbb{R}\}.$$

Weiter sei $\mu_0 \in \mathbb{R}$ ein vorgegebener Wert. Die Aufgabe bestehe darin die Nullhypothese

$$H_0: \mu = \mu_0 \quad \text{zu testen gegen die Alternative}$$

$$H_1: \mu \neq \mu_0.$$

In Beispiel 2.2.11 hatten wir für die vorliegende Situation die Likelihood-Funktion

$$\rho(x, \mu, v) = \sqrt{(2\pi v)^{-n}} \exp\left(-\sum_{j=1}^n \frac{(x_j - \mu)^2}{2v}\right)$$

eingeführt. Dort hatten wir auch nachgewiesen, dass diese Funktion in $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$ ihr Maximum annimmt. $\mu_1 = \bar{x}$ ist demnach der Maximum-Likelihood Schätzer von μ . Wir setzen jetzt $L_{\mu_0}(x) = \rho(x, \mu_0, \sigma_0^2)$ und $L_{\mu}(x) = \rho(x, \mu, \sigma_0^2)$. Mit diesen Größen bilden wir den Likelihood-Quotienten

$$\begin{aligned} R(x) &= R(x_1, \dots, x_n) = \frac{L_{\mu_0}(x)}{\sup\{L_{\mu}(x) \mid \mu \in \mathbb{R}\}} = \frac{L_{\mu_0}(x)}{L_{\mu_1}(x)} \\ &= \exp\left(-\frac{1}{2\sigma_0^2} \sum_{j=1}^n (x_j - \mu)^2 - (x_j - \bar{x})^2\right) \\ &= \exp\left(-\frac{1}{2\sigma_0^2} \sum_{j=1}^n ((x_j - \bar{x})^2 + 2(x_j - \bar{x})(\bar{x} - \mu_0) + (\bar{x} - \mu_0)^2 - (x_j - \bar{x})^2)\right) \\ &= \exp\left(-\frac{n(\bar{x} - \mu_0)^2}{2\sigma_0^2}\right). \end{aligned}$$

Für $0 < \delta < 1$ setzen wir $c = \sigma_0 \sqrt{\frac{-2 \ln(\delta)}{n}}$; wegen des vorstehenden Satzes wird der Ablehnungsbereich $G(\delta)$ gegeben durch

$$G(\delta) = \{x \in \mathbb{R}^n \mid R(x) \leq \delta\} = \{x \in \mathbb{R}^n \mid |\bar{x} - \mu_0| \geq c\}.$$

Wir haben die folgende Aussage nachgewiesen:

2.4.5 Satz: *Es liege eine $N(\mu, \sigma_0^2)$ -verteilte Population vor mit bekanntem $\sigma_0 > 0$ und unbekanntem $\mu \in \mathbb{R}$. Es sei $\mu_0 \in \mathbb{R}$.*

Die Nullhypothese $H_0 : \mu = \mu_0$ soll gegen $H_1 : \mu \neq \mu_0$ getestet werden. Dann wird der Likelihood-Quotient gegeben durch

$$R(x) = R(x_1, \dots, x_n) = \exp\left(-\frac{n(\bar{x} - \mu_0)^2}{2\sigma_0^2}\right).$$

Der Ablehnungsbereich $G(\delta)$ für $0 < \delta < 1$ wird gegeben durch

$$G(\delta) = \{x \in \mathbb{R}^n \mid |\bar{x} - \mu_0| \geq c\} \quad \text{mit } c = \sigma_0 \sqrt{\frac{-2 \ln(\delta)}{n}}.$$

Anmerkung: In diesem Zusammenhang soll noch auf die einseitigen Alternativen hingewiesen werden. Unter den obigen Voraussetzungen soll

die Nullhypothese $H_0 : \mu = \mu_0$ soll gegen $H_1 : \mu > \mu_0$

bei der rechtsseitigen Alternative getestet werden. Der Ablehnungsbereich $G(\delta)$ für $0 < \delta < 1$ wird in diesem Fall gegeben durch

$$G(\delta) = \{x \in \mathbb{R}^n \mid \bar{x} \geq \mu_0 + c\} \quad \text{mit } c = \sigma_0 \sqrt{\frac{-2 \ln(\delta)}{n}}.$$

Bei der linksseitigen Alternative $H_1 : \mu < \mu_0$ wird der Ablehnungsbereich $G(\delta)$ für $0 < \delta < 1$ gegeben durch

$$G(\delta) = \{x \in \mathbb{R}^n \mid \bar{x} \leq \mu_0 - c\} \quad \text{mit } c = \sigma_0 \sqrt{\frac{-2 \ln(\delta)}{n}}.$$

Im weiteren Verlauf geben wir einige Tests an, die mit ähnlichen Methoden mit Hilfe eines Likelihood-Quotienten konstruiert werden; wir verzichten hier auf explizite Konstruktionen, da diese sich nicht substanziiell von den eben durchgeführten unterscheiden. Zunächst behandeln wir einen Test für den Erwartungswert einer Normalverteilung, wenn die Varianz nicht bekannt ist. Es gilt die folgende Aussage:

2.4.6 Satz: *Es liege eine $N(\mu, \sigma_0^2)$ -verteilte Population vor mit unbekanntem $\sigma_0 > 0$ und unbekanntem $\mu \in \mathbb{R}$. Es sei $\mu_0 \in \mathbb{R}$.*

Die Nullhypothese $H_0 : \mu = \mu_0$ soll gegen $H_1 : \mu \neq \mu_0$ getestet werden. Dann wird der Likelihood-Quotient gegeben durch

$$R(x) = R(x_1, \dots, x_n) = \left[1 + \frac{1}{n-1} \left(\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)^2\right]^{-n/2} \quad \text{mit } s = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2.$$

Der Ablehnungsbereich wird beschrieben durch Mengen der Form

$$G = \left[(x_1, \dots, x_n) \in \mathbb{R}^n \mid \left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| \geq c \right].$$

Unter der Nullhypothese $\mu = \mu_0$ ist R Student t -verteilt mit $n - 1$ Freiheitsgraden.

Entsprechend ergeben sich natürlich auch einseitige Tests; in diesen Fällen wird dann der Ablehnungsbereich gegeben durch

$$G = \left[(x_1, \dots, x_n) \in \mathbb{R}^n \mid \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \geq c \right] \text{ beziehungsweise}$$

$$G = \left[(x_1, \dots, x_n) \in \mathbb{R}^n \mid \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \leq -c \right].$$

Entsprechend ergibt sich der **Chiquadrat**test bei normalverteilten Zufallsvariablen zum Abtesten der Varianz:

2.4.7 Satz: *Es liege eine $N(\mu, \sigma_0^2)$ -verteilte Population vor mit unbekanntem $\sigma_0 > 0$ und unbekanntem $\mu \in \mathbb{R}$. Die Nullhypothese $H_0 : \sigma = \sigma_0$ soll gegen $H_1 : \sigma \neq \sigma_0$*

getestet werden. Dann wird der Likelihood-Quotient gegeben durch

$$R(x) = R(x_1, \dots, x_n) = e^{n/2} n^{-n/2} \left[\frac{(n-1)s^2}{\sigma_0^2} \right]^{n/2} \exp\left(-\frac{(n-1)s^2}{2\sigma_0^2} \right) \text{ mit } s = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2.$$

Der Ablehnungsbereich wird beschrieben durch Mengen der Form

$$G = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \left| \frac{(n-1)s^2}{\sigma_0^2} \right| \leq c_1 \right\} \cup \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \left| \frac{(n-1)s^2}{\sigma_0^2} \right| \geq c_2 \right\}.$$

Unter der Nullhypothese $\sigma = \sigma_0$ ist R Chiquadrat-verteilt mit $n - 1$ Freiheitsgraden.

Wie eben können auch die einseitigen Tests formuliert werden; dabei beschränkt man sich dann auf eine der obigen Vereinigung vorkommenden Mengen. Die Werte c_1 und c_2 werden mit Hilfe der Chiquadrat-Verteilung ermittelt. Man vergleiche etwa die Darstellung in W.R. Pestman, Abschnitt III.2. Dort finden sich auch viele Beispiele für die genannten Tests und eine Vielzahl von weiteren Testverfahren.

2.5 Verteilungstest

In dem vorliegenden Abschnitt soll der Test von Kolmogorov-Smirnov vorgestellt werden, der überprüft, ob eine Datenreihe einer fest vorgegebenen Verteilungsfunktion entspricht. Der Einfachheit halber gehen wir dabei von einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und einer Folge $(X_n)_{n=1}^\infty$ unabhängiger Zufallsvariablen mit der gleichen Verteilungsfunktion $F: \mathbb{R} \rightarrow [0, 1]$ aus, die in den Anwendungen als unbekannt angenommen wird. Zu diesen Testgrößen X_1, \dots, X_n, \dots bilden wir für jedes $n \in \mathbb{N}$ die **empirische Verteilungsfunktion**:

2.5.1 Bemerkung: *Es sei $(X_n)_{n=1}^\infty$ eine Folge unabhängiger Zufallsvariablen mit der Verteilungsfunktion F . Für alle $n \in \mathbb{N}$ sei die empirische Verteilungsfunktion $\hat{F}(X_1, \dots, X_n)$ definiert durch*

$$\hat{F}(X_1, \dots, X_n)(t) = \frac{1}{n} \text{Anz}\{j \in \{1, \dots, n\} \mid X_j \leq t\}.$$

Dann gilt

$$\hat{F}(X_1, \dots, X_n)(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{]-\infty, t]} \circ X_j \text{ und}$$

$$\hat{F}(X_1, \dots, X_n)(t) \rightarrow F(t) \text{ fast sicher bei } n \rightarrow \infty$$

$$\hat{F}(X_1, \dots, X_n)(t-) \rightarrow F(t-) \text{ fast sicher bei } n \rightarrow \infty \text{ für alle } t \in \mathbb{R}.$$

Beweis: Wegen $[\mathbb{1}_{]-\infty, t]} \circ X_j = 1] = [X_j \leq t]$ für alle $t \in \mathbb{R}$ erhalten wir

$$E(\mathbb{1}_{]-\infty, t]} \circ X_j) = P(X_j \leq t) = F(t)$$

für alle $t \in \mathbb{R}$ und alle $j \in \mathbb{N}$. Wegen $\mathbb{1}_{]-\infty, t]} \circ X_j \in \mathcal{L}^2(P)$ folgt aus dem Gesetz der großen Zahl 1.5.3

$$\hat{F}(X_1, \dots, X_n)(t) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{]-\infty, t]} \circ X_j \rightarrow E(\mathbb{1}_{]-\infty, t]} \circ X_1) = F(t)$$

fast sicher bei $n \rightarrow \infty$. Die weitere Aussage folgt entsprechend bei Benutzung der Zufallsvariablen $\mathbb{1}_{]-\infty, t]} \circ X_j$. \square

Der Test von Kolmogorov-Smirnov bewertet den maximalen Abstand von der empirischen Verteilungsfunktion zu der angenommenen Verteilungsfunktion. Diese Differenz der Verteilungsfunktionen in der Supremumsnorm wird ja als Supremum des Betrages der Differenz über alle $t \in \mathbb{R}$, also über eine nicht mehr abzählbare Teilmenge gebildet. Wir müssen daher noch die Messbarkeit dieses Supremums nachweisen:

2.5.2 Bemerkung: *Es sei $(X_n)_{n=1}^\infty$ eine Folge unabhängiger Zufallsvariablen mit der Verteilungsfunktion F . Für alle $n \in \mathbb{N}$ ist die Funktion*

$$\|\hat{F}(X_1, \dots, X_n) - F\|_\infty = \sup \{ |\hat{F}(X_1, \dots, X_n)(t) - F(t)| \mid t \in \mathbb{R} \}$$

eine Zufallsvariable.

Beweis: Da die beiden betrachteten Funktionen Verteilungsfunktionen sind, müssen sie von rechts stetig sein. Man rechnet daher sehr einfach nach dass

$$\|\hat{F}(X_1, \dots, X_n) - F\|_\infty = \sup \{ |\hat{F}(X_1, \dots, X_n)(t) - F(t)| \mid t \in \mathbb{Q} \}$$

gilt. Daher ist $\|\hat{F}(X_1, \dots, X_n) - F\|_\infty$ als Supremum über abzählbar viele Zufallsvariablen, wieder messbar und damit ebenfalls eine Zufallsvariable. \square

Die folgende Aussage ist eine wesentliche Voraussetzung für die Motivation des Test-Verfahrens von Kolmogorov–Smirnov, der Beweis ist umfangreicher aber doch recht elementar; er benutzt lediglich die Eigenschaften der Verteilungsfunktionen als rechtseitig stetige, monoton wachsende, und beschränkte Funktionen.

2.5.3 Satz: (Satz von Glivenko–Cantelli) *Es sei $(X_n)_{n=1}^\infty$ eine Folge unabhängiger Zufallsvariablen mit der Verteilungsfunktion F . Dann gilt*

$$\|\hat{F}(X_1, \dots, X_n) - F\|_\infty \rightarrow 0 \text{ fast sicher bei } n \rightarrow \infty .$$

Beweis: Wir zeigen in mehreren Schritten, dass

$$\limsup_{n \rightarrow \infty} \|\hat{F}(X_1, \dots, X_n) - F\|_\infty < 5\varepsilon$$

fast sicher gilt für alle vorgegebenen $\varepsilon > 0$.

Schritt (I): Es sei $\varepsilon > 0$ beliebig. Dann existieren endlich viele $\tau_0 < \tau_1 < \dots < \tau_k$ mit $F(\tau_0) < \varepsilon$, $F(\tau_k) > 1 - \varepsilon$ und mit $F(t) - F(s) < F(\tau_j-) - F(\tau_{j-1}) < \varepsilon$ für alle $j = 1, \dots, k$ und alle $s, t \in]\tau_{j-1}, \tau_j[$.

Beweis von (I): Die Aussage ist evident, da F monoton wachsend, rechtseitig stetig ist mit $F(t) \rightarrow 0$ bei $t \rightarrow -\infty$ und $F(t) \rightarrow 1$ bei $t \rightarrow \infty$.

Schritt (II): Es sei $\varepsilon > 0$, und es seien $\tau_0 < \tau_1 < \dots < \tau_k$ gemäß (I) gewählt. Dann existiert zu jedem $j \in \{1, \dots, k\}$ ein $A_j \in \mathcal{A}$ mit $P(A_j) = 0$ und mit

$$\limsup_{n \rightarrow \infty} \|\hat{F}(X_1(\omega), \dots, X_n(\omega)) - F\|_{\infty,]\tau_{j-1}, \tau_j[} \leq 3\varepsilon$$

für alle $\omega \notin A_j$.

Beweis von (II): Wegen 2.5.1 gilt

$$\hat{F}(X_1, \dots, X_n)(\tau_{j-1}) \rightarrow F(\tau_{j-1}) \text{ und } \hat{F}(X_1, \dots, X_n)(\tau_j-) \rightarrow F(\tau_j-)$$

fast sicher bei $n \rightarrow \infty$. Es existiert also eine Menge $A_j \in \mathcal{A}$ mit $P(A_j) = 0$ und mit

$$\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-1}) \rightarrow F(\tau_{j-1}) \text{ und } \hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_j-) \rightarrow F(\tau_j-)$$

bei $n \rightarrow \infty$ für alle $\omega \notin A_j$. Für jedes $\omega \notin A_j$ existiert also ein $m = m(\omega) \in \mathbb{N}$ mit

$$\begin{aligned} |\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-1}) - F(\tau_{j-1})| &< \varepsilon \text{ und} \\ |\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_j-) - F(\tau_j-)| &< \varepsilon \text{ für alle } n \geq m(\omega). \end{aligned}$$

Daher gilt für alle $\omega \notin A_j$, alle $n \geq m(\omega)$, und alle $s \in [\tau_{j-1}, \tau_j[$ die Abschätzung

$$\begin{aligned} & |\hat{F}(X_1(\omega), \dots, X_n(\omega))(s) - F(s)| \\ & \leq |\hat{F}(X_1(\omega), \dots, X_n(\omega))(s) - \hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-1})| \\ & \quad + |\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-1}) - F(\tau_{j-1})| + |F(\tau_{j-1}) - F(s)| \\ & \leq |\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-}) - \hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-1})| + 2\varepsilon \\ & \leq |\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-}) - F(\tau_{j-})| + |F(\tau_{j-}) - F(\tau_{j-1})| \\ & \quad + |F(\tau_{j-1}) - \hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_{j-1})| + 2\varepsilon \\ & \leq 5\varepsilon. \end{aligned}$$

Es folgt daher $\|\hat{F}(X_1(\omega), \dots, X_n(\omega)) - F\|_{\infty, [\tau_{j-1}, \tau_j[} \leq 5\varepsilon$ für alle $\omega \notin A_j$ und alle $n \geq m(\omega)$, was zu zeigen war.

Schritt (III): Es sei $\varepsilon > 0$, und es seien $\tau_0 < \tau_1 < \dots < t_k$ gemäß (I) gewählt. Dann existieren $A_0, A_{k+1} \in \mathcal{A}$ mit $P(A_0) = P(A_{k+1}) = 0$, und mit

$$\limsup_{n \rightarrow \infty} \hat{F}(X_1(\omega), \dots, X_n(\omega))(s) < \varepsilon \text{ und } \limsup_{n \rightarrow \infty} \hat{F}(X_1(\omega), \dots, X_n(\omega))(t) > 1 - \varepsilon$$

für alle $s < \tau_0, t > \tau_k$ und $\omega \notin A_0 \cup A_{k+1}$.

Beweis von Schritt (III): Wegen 2.5.1 gilt $\hat{F}(X_1, \dots, X_n)(\tau_0) \rightarrow F(\tau_0)$ fast sicher bei $n \rightarrow \infty$. Es existiert also ein $A_0 \in \mathcal{A}$ mit $P(A_0) = 0$ und

$$\hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_0) \rightarrow F(\tau_0)$$

bei $n \rightarrow \infty$ für alle $\omega \notin A_0$. Wegen $F(t) \leq F(\tau_0) < \varepsilon$ und $\hat{F}(X_1(\omega), \dots, X_n(\omega))(t) \leq \hat{F}(X_1(\omega), \dots, X_n(\omega))(\tau_0)$ für alle $t > \tau_0$ folgt der erste Teil der Aussage. Der zweite Teil folgt entsprechend.

Schritt (IV) Wir setzen $A(\varepsilon) = A_0 \cup \dots \cup A_{k+1}$. Wegen (ii) und (III) folgt daher

$$\limsup_{n \rightarrow \infty} \|\hat{F}(X_1(\omega), \dots, X_n(\omega)) - F\|_{\infty} \leq 5\varepsilon$$

für alle $\omega \notin A$. Die Aussage des Satzes folgt durch Betrachtung von $\varepsilon = \frac{1}{n}$ und von $A = A(1) \cup A(\frac{1}{2}) \cup \dots$. □

Wir sind jetzt in der Lage den Kolmogorov–Smirnov–Test zu formulieren. Dazu sei wieder $(X_n)_{n=1}^{\infty}$ eine Folge unabhängiger Zufallsvariablen mit einer unbekanntem Verteilungsfunktion F , und es sei F_0 eine vorgegebene Verteilungsfunktion. Es soll die Nullhypothese

$$H_0 : F = F_0 \text{ gegen die Alternative } H_1 : F \neq F_0$$

getestet werden. Dazu wählen wir die Teststatistik

$$D_{n, F_0} = \|\hat{F}(X_1, \dots, X_n) - F_0\|_{\infty}.$$

Mit dem Satz von Glivenko–Cantelli 2.5.3 haben wir nachgewiesen, dass unter der Voraussetzung der Nullhypothese $F = F_0$

$$D_{n, F_0} \rightarrow 0 \text{ fast sicher bei } n \rightarrow \infty$$

gilt. Wir akzeptieren also bei einer gegebenen Stichprobe x_1, \dots, x_n die Nullhypothese, wenn D_{n, F_0} klein ist und verwerfen diese, wenn D_{n, F_0} groß ist. Die folgende Aussage zeigt die Unabhängigkeit der Teststatistik von der zu Grunde liegenden Verteilungsfunktion:

2.5.4 Satz: (Satz von Kolmogorov–Smirnov) *Es sei $U : \mathbb{R} \rightarrow [0, 1]$ die Verteilungsfunktion der Gleichverteilung auf $[0, 1]$: $U(t) = 0$ für $t > 0$, $U(t) = t$ für $0t < 1$ und $U(t) = 1$ für $t \geq 1$. Es sei weiter $F : \mathbb{R} \rightarrow [0, 1]$ eine stetige Verteilungsfunktion. Fast sicher gilt dann*

$$D_{n, U} = D_{n, F} \text{ für alle } n \in \mathbb{N}.$$

Beweis: Zu der Verteilungsfunktion f bilden wir eine Pseudo-Inverse $g : [0, 1] \rightarrow \overline{\mathbb{R}}$, definiert durch

$$\begin{aligned} g(0) &= \sup \{t \in \mathbb{R} \mid F(t) = 0\} \in [-\infty, \infty[\\ g(1) &= \inf \{t \in \mathbb{R} \mid F(t) = 1\} \in]-\infty, \infty] \\ g(s) &= \min \{t \in \mathbb{R} \mid F(t) = s\} \text{ für alle } 0 < s < 1 \end{aligned}$$

mit der üblichen Verabredung $\sup\{\emptyset\} = -\infty$ und $\inf\{\emptyset\} = \infty$. Speziell folgt wegen der Stetigkeit von F die Beziehung $s = F(g(s))$ für alle $s \in]0, 1[$.

Es sei jetzt $(X_n)_{n=1}^\infty$ eine Folge von unabhängigen Zufallsvariablen mit der Verteilungsfunktion F . Für alle $j \in \mathbb{N}$ sei $Y_j = F \circ X_j$. Dann ist $(Y_n)_{n=1}^\infty$ eine Folge von unabhängigen Zufallsvariablen. Wir bestimmen jetzt die Verteilungsfunktion von Y_j : Für alle $0 < s < 1$ gilt

$$P([Y_j \leq s]) = P([F \circ X_j \leq s]) = P(X_j \leq g(s)) = F(g(s)) = s = U(s).$$

Entsprechend erhalten wir $P([Y_j \leq s]) = 0$ für alle $s \leq 0$ und $P([Y_j \leq s]) = 1$ für alle $s \geq 1$. Wenn die Verteilungsfunktion F auf einem Intervall $[a, b]$ konstant ist, folgt $P(a \leq X_j < b) = F(b-) - F(a) = 0$. Daher ist auch die empirische Verteilungsfunktion auf $[a, b]$ fast sicher konstant. Für alle $0 < s < 1$ gilt weiter

$$\hat{F}(Y_1, \dots, Y_n)(s) = \frac{1}{n} \text{Anz}\{j \mid F(j) \leq s\} = \frac{1}{n} \text{Anz}\{j \mid j \leq g(s)\} = \hat{F}(X_1, \dots, X_n)(g(s))$$

fast sicher. Daher erhalten wir

$$\begin{aligned} D_{n, U} &= \|\hat{F}(Y_1, \dots, Y_n) - U\|_\infty = \sup \left\{ |\hat{F}(Y_1, \dots, Y_n)(s) - U(s)| \mid s \in \mathbb{R} \right\} \\ &= \sup \left\{ |\hat{F}(Y_1, \dots, Y_n)(s) - s| \mid 0 < s < 1 \right\} \\ &= \sup \left\{ |\hat{F}(X_1, \dots, X_n)(g(s)) - F(g(s))| \mid s \in \mathbb{R} \right\} \\ &= \sup \left\{ |\hat{F}(X_1, \dots, X_n)(t) - F(t)| \mid t \in \mathbb{R} \right\} = D_{n, F} \end{aligned}$$

fast sicher auf Ω . □

Die Bedeutung dieses Satzes liegt in der Tatsache, dass eine beliebige stetige Verteilungsfunktion F die gleiche Teststatistik besitzt wie die Gleichverteilung auf dem Einheitsintervall. Für diese lassen sich aber bei vorgegebenem $\alpha \in]0, 1[$ die Ablehnungsbereiche der Nullhypothese durch die Forderung bestimmen, dass zu jedem $n \in \mathbb{N}$ ein $C > 0$ existiert mit $P(D_{n, U} \geq C) = \alpha$. Diese Werte liegen in tabellierter Form vor.