Ruprecht-Karls-Universität Heidelberg
Institut für Mathematik
Professor Dr. Jan JOHANNES

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Skript zur Vorlesung

Einführung in die Wahrscheinlichkeitstheorie und Statistik

Wintersemester 2025/26

Fassung Stand 11. Dezember 2025

Falls Sie **Fehler im Skript** finden, teilen Sie mir diese bitte per eMail an johannes@math.uni-heidelberg.de mit.

MΛTHEMΛTIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
Telefon: +49 6221 54.14.190 – Fax: +49 6221 54.14.101
eMail: johannes@math.uni-heidelberg.de
Webseite: sip.math.uni-heidelberg.de

Inhaltsverzeichnis

Kapitel 1

Prolog

Beispiel I: Geschlecht von Konsumierenden

Einem amerikanischen Produzenten eines *kalorienarmen, koffeinhaltigen Erfrischungsgetränkes* wurde das Gerücht zugetragen, dass es mehr weibliche als nicht weibliche Konsumierende des Getränkes gibt. Um diesen Rumor zu überprüfen, hat der firmeneigene Verbraucherservice das Geschlecht ({Weiblich,Nicht weiblich}) von 1000 Konsumierenden des Getränkes erhoben. Mit folgendem Resultat:

W W W W N W W W N W N W W N W W W N W N W W N W N N W W W W W W W W W W N W W N
W W W N W W W W W W N W W N W W N W W W W N N N W W W W W N W N N W W W N W W N
W W N W W N W N W W W W W N W N W W W N N W W W W W W W N W W W N W W W W W W W
N W W W N W W N N W W W N W N N W W W W W N W W N W N W W W W W W W N W W W W W
N W W N N W W N W W N N N W W W W W N W N W W W N N N W W N W W W W W W W W W W
N W N W W W W W W W W N N W W N W W W W W N W W W W N N W N W N W N W N W W W W N
N W W W W N N W N W W W W W N W W W N W N W W W W W W W W N W W W N W N W W W W
W N N N N W N N W W W N W W W W W W N W W W W W W W W N W W W W W N W N W W W W N
W W W N N W W N W N W N W W W N W W N W N W W W W N N W N W N W W N W W W W W W
N W W W N W W N W W W N W N W N W W N N W W W W N W N W N W N N N W W N W N N W W
W W W W W W W W W W N W W W W W N W W W W N N W W N W W W W W N N N W W W N N W
W W W W W N W W N W N N W W W W W W N W W W N W N W W W W W N N W W W N W W N W N
W W W W W W N N N W W W N W W N W W N W W W W N N W W W W N W W W W N W N W W W W
N W W W N N W W N W N N W W W N W N N W N W W W W N W N N N N W W N W W W W W N N W
N W W W N W W N W N N N N N N W W W W W W W W N W W W W N W W W W W W W W N W N N
W W W N W N W W W W N N W N N W N W W N N W N N W W W W N W W W W N W W W W N N W
N N W N N W N N W W W N W W N W W W N W N W N W W W W N W N N W W N W N W W W W W
N W W W W N W W W W N W W W N W W W W W W N W W W N N W W W N W W W W W W W N W
W W W W W W N W W W W N W W W W N W W W N W W W N N W W W W W W W N N N W W W W
W N N N W W W W N N W N W W W W W W W W N W W W W N W W W W W W W W N W W W W N
W W W W W W N W W N W W N W W W W W W W W W W W N N N W N W W W W W N W W W W N
N N W W W W W W N N N N N W W W W N W N W N W N N W W N N N W W W W N W W W W N N W W
W W N W W W W W W W W W W W W W W W W W N N W W W N W W N W N W W N W N N N N W N W
W N W N N W W N W N N N W W W W W W N W W N W N W W W W N W W N W W W W W N W N N
W W W W W W W W W N W W W W N W N W W W W

(Anzahl W=699)

Der amerikanische Produzent hat daraufhin entschieden, seine Produktpalette um ein koffeinhaltiges Erfrischungsgetränk ohne Zucker zu erweitern. Dabei war das Ziel, eine *neue Marke speziell für Männer* zu kreieren. Zur Kontrolle des Ziels hat der firmeneigene Verbraucherservice wieder das Geschlecht ({Männlich,Nicht männlich}) von 1000 Konsumierenden des Getränkes erhoben. Mit folgendem Resultat:

M M M M M M M M N M N M N N M N M M M M M N M M N N N M M M M N N M M M M M M M M
N M N N M M M M M N N M M N M N N N N N N M M N N M M N N N N N M M M M M M M M M M

*M M M M N M M M M M M N M M N M M N N N M N N M N M N N M M N M N N N N M N M M M
M N M M N N N M M M M M M N M M N M N M M M N N N M N M N N N N M N N M M N N M M N
M N M M M M M M N M M M M M M M M M N M M N N N M N M M M N M M M N N M M M N M M
M M N M M N M M M M M M M N M M M N M M M N M M M M N M M M N N N N M M N M M N M
M M N N M N M M M M M M N N M N M M N M N N M M M N N M N N M M M N M M M M M M N
M M N M N N M M N M N M M N M N N N N M M N M M N N M N M N M M N N M M M N N M M M
M N M M N M N M M N M M M N N M M M N M N M M M M M M M N M M M M M M M M N M M N
N M M M M M N N N M M M M M M N N M N M M M N N N M M M N N N M M M N N N M M M N
M M M M N M M M N M N M M M M N M N N M N N M M M N N M N N N M M M N N M N M M M
M M N N N N M M M M N M M N M N M N M M N M M M M M N N N M M N M M M M M N N M N
M M M M N N M M M M N M M M M N M M N M M M N N M M M N M M M M N M N N M M M M N
M M N N M M N M M N M M N M N N N N M M N N M M M M N M N N M M M M N M N M N M M
M N M N N N N M N M M M N M N N N N M M M M M N M N M M M N M M M M N M M N M M N
M N N M M N M N M N M N N M M N N N M M N N M M M M M N M N M M N M N N N M M M N
M M M N N M M N M M M N M M M M N N N M N M N M M M M N N N M M M M N N N M N M N N
M M M M M M N M N M N M M N N N N M N N M N N M M M M N N M M N N M N M M M M M N M
M M M M N M M N N M M M N M M M M M M M N M M N N M M M N N M M N N M M M M M M M
M M N M M N M N M N N M M M M N N N M M N M M M M N M N M M M M N M M M M N M N M
M M N N N N N N N N N M M N M M N M M M N M N M N M M M M M M M M N M N M N M N N
M N N N N M M M M M M M M M M M N N N M N N N M M N M M M N M N N M M M N M M M M
M M N N N N M M M M N N N M N N N M M N M M N N N N N M N N N M M N N M N M N M N N M
M N N N M M N N M M M N M N M M N M M M M M N N M N M N M N M M M N N M M N N N M M
N M N N M M M*

(Anzahl M=611)

Beispiel II: Qualitätskontrolle

Eine Schraubenherstellerin hat einen Vertrag mit einem Kunden abgeschlossen, in dem sie sich verpflichtet, dass in der nächsten Lieferung im Mittel *höchstens 1 von 100 Schrauben beschädigt* ist. Der Kunde hat 10.000 Beutel a je 50 Schrauben bestellt. Zur Kontrolle der Qualität ihrer Lieferung, zählte die Herstellerin in 100 Beuteln der Lieferung die beschädigten Schrauben. Mit dem folgenden Resultat:

*0 2 0 0 0 1 0 1 0 1 0 0 2 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 2 0 1 0 0 0 0 0 0 0 2 0 0 0 0 2 1 0 0 0 1 1 2 0
0 0 1 1 2 0 0 0 0 1 0 4 0 0 1 0 0 0 0 2 1 0 0 1 0 0 0 0 1 0 0 1 0 0 2 0 0 1 0 1 1 0 1*

(Summe=46)

Auf Grund der Nachfrage hat die Schraubenherstellerin einen zweiten Produktionsstandort eingerichtet. Zur Kontrolle der Produktionsqualität zählte die Herstellerin in 100 Beuteln a je 50 Schrauben die beschädigten Schrauben. Mit dem folgenden Resultat:

*0 1 1 4 2 0 1 2 1 1 3 3 3 0 2 0 2 0 2 4 3 4 0 3 1 4 0 2 1 1 1 3 3 2 0 1 1 1 2 1 1 4 1 2 1 1 2 4 4 2 2 0 3 1 1 1 1
3 1 4 5 2 3 0 2 2 2 0 0 0 7 1 2 1 0 1 1 4 0 1 2 1 1 1 3 4 1 1 1 1 4 1 3 1 1 3 1 0 3 0*

(Summe=172)

Beispiel III: Anzahl

Ein Unternehmen bietet seinen Klienten eine Telefonhotline von Montag bis Freitag zwischen 8:00 Uhr und 18:00 Uhr an. Um den Kundenservice zu verbessern, zählt das Unternehmen eine Woche lang *die Anrufe innerhalb einer Viertelstunde*. Mit dem folgenden Resultat:

18 17 16 8 19 8 12 15 14 11 15 11 13 8 12 14 15 13 8 17 16 15 12 13 10 15 15 13 13 16 13 15 13 16 8 9 11 10 14 15 8 8 15 13 8 15 10 18 8 11 7 16 13 8 15 9 14 9 7 13 10 8 16 15 13 9 14 15 14 18 6 10 6 14 9 7 14 6 14 10 11 16 13 9 14 11 13 11 12 12 11 11 9 15 11 7 20 10 8 12 24 14 11 8 14 13 11 10 12 7 13 11 10 13 14 10 7 16 19 12 6 13 9 15 10 22 12 15 14 16 13 12 14 6 11 13 10 6 11 13 5 15 9 14 12 12 11 9 7 13 13 9 12 9 11 12 19 13 9 14 10 10 11 10 9 11 12 10 10 13 10 11 21 14 12 10 8 11 16 16 18 13 12 11 13 13 14 17 11 9 13 11 11 9 15 15 7 15 20 5 18 14 10 11 20 19 12 10 14 7 9 9 12 7 8 8 6 14 12 13 9 12 12 14 9 10 7 11 14 18 10 12 11 12 13 11 16 14 10 11 15 10 10 14 9 14 15 12 10 13 15 15 8 13 11 10 12 11 11 12 9 3 5 13 13 13 17 13 16 13 9 15 13 17 12 9 7 13 9

(Summe=3348)

Beispiel IV: Wartezeit

Ein Wissenschaftler fährt jeden Tag mit der U-Bahn in Berlin. Während drei Monaten misst er seine tägliche *Wartezeit an der U-Bahn Haltestelle*. Mit dem folgenden Resultat (in Minuten):

7.44 6.21 6.03 8.08 7.07 6.33 9.00 1.28 5.92 2.38 5.89 0.09 1.59 5.57 8.56 7.01 1.54 1.21 0.99 2.30 1.00 7.82 8.53 0.18 2.33 7.81 7.77 9.05 6.86 3.82 1.36 7.39 2.06 0.96 2.28 3.54 4.01 7.59 7.06 7.58 6.59 1.35 8.64 1.01 8.88 2.27 5.96 1.25 6.45 2.45 9.02 5.83 3.80 1.95 9.28 0.29 0.73 3.67 2.74 0.17 9.32 0.02 4.76 7.49 8.49 2.36 2.15 2.06 3.73 2.86 5.35 1.27 6.08 9.50 4.70 5.44 4.11 9.20 2.59 5.98 1.17 5.98 8.40 2.97 2.66 5.25 0.97 3.74 2.79 8.81

(Maximum=9.5)

Beispiel V: Lebensdauer

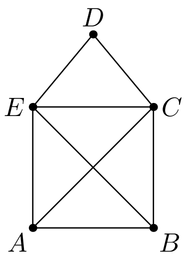
Eine Stadtverwaltung benutzt in ihren Gebäuden 35.000 Glühlampen. Die Herstellerin der Glühlampen garantiert, dass die Lebensdauer der Glühlampe im Mittel mindestens 1000 Stunden beträgt. Um diese Aussage zu überprüfen, führt die Einkaufsabteilung eine Studie mit 100 Glühlampen durch. Mit dem folgenden Resultat (in Stunden):

2109.35 1074.86 28.83 1279.98 156.99 5019.57 1996.70 478.79 999.25 2253.01 465.35 530.50 624.27 4167.61 721.24 134.58 1292.09 174.07 504.60 83.62 2824.01 881.01 42.30 227.83 156.20 13.34 1740.43 23.56 908.03 271.18 23.28 2191.34 357.66 1917.95 357.95 1281.17 183.21 98.11 700.71 820.09 739.14 23.79 923.54 17.19 108.47 467.33 761.10 15.48 5635.45 2714.75 457.12 271.27 155.30 1396.21 644.90 393.85 382.58 1087.93 4547.50 1241.29 807.20 2291.24 2027.31 150.71 5031.31 811.09 1049.09 988.20 1003.60 1264.92 1488.36 1603.13 1923.11 204.41 1765.73 224.21 1011.45 587.16 888.23 1274.92 1222.33 295.25 631.28 48.22 109.15 692.58 11.14 1855.80 377.98 200.43 731.42 823.03 106.78 2233.35 409.38 115.18 1542.88 112.48 1278.54 2345.72

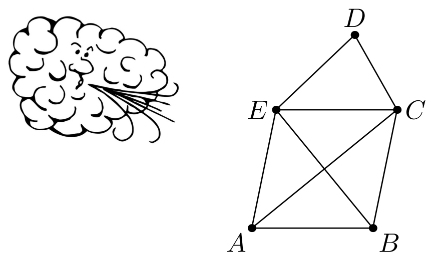
(Mittelwert=1026.37)

Beispiel VI: Messfehler

Nach einem starken Sturm fürchtet der alte Fürst G. Őmetry dass sein Schloss (siehe Bild unten) beschädigt ist. Genauer, er hat den Eindruck dass die Fassade, die vor dem Sturm ein Rechteck war, sich in ein nicht-rechteckiges Parallelogram verformt hat.



intaktes Schloss



beschädigtes Schloss

Um diesen Eindruck objektiv zu überprüfen, entschließt sich der Fürst G. Őmetry die Segmente $\overrightarrow{AC}$ und $\overrightarrow{EB}$ 31 mal zu messen, mit folgendem Resultat für das Segmente $\overrightarrow{AC}$ (in m):

9.60 9.35 9.57 8.65 10.11 10.05 10.82 9.17 11.02 9.02 9.48 9.34 9.27 9.96 9.50 9.31 9.21 10.11 9.96 9.00 7.77 9.44 9.08 9.97 9.68 10.55 8.66 9.14 9.06 9.43 9.56

(Mittelwert=9.51)

und für das Segment $\overrightarrow{EB}$ (in m)

10.38 11.32 9.01 9.51 10.19 10.94 10.26 10.12 10.57 9.93 12.17 11.88 10.30 11.33 10.95 10.23 8.72 10.22 11.42 11.84 10.29 11.55 8.43 9.06 9.45 9.99 10.40 11.00 12.68 9.73 11.36

(Mittelwert=10.49)

Die belgische Abbaye de Rochefort ist berühmt für ihr selbst gebrautes Bier. Die Mönche möchten in einen neuen Abfüllautomaten für 33cl Bierflaschen investieren. Der Hersteller des Automaten gibt eine Genauigkeit der Abfüllung von 0.8 an. Um diese Herstellerangabe zu überprüfen, messen die Mönche das Volumen von 42 zufällig ausgewählten 33cl Bierflaschen, mit folgendem Resultat (in Zentiliter):

32.68 32.19 33.50 33.78 33.95 33.04 33.17 30.62 33.92 32.99 32.13 32.99 33.64 34.41 33.88 32.42 33.08 32.48 32.49 33.52 33.86 33.95 32.80 32.66 32.57 33.93 31.02 32.71 31.89 34.28 33.76 32.79 31.74 33.03 33.57 33.72 31.55 32.57 32.87 32.03 32.47 32.87

(Standardabweichung=0.86)

Laut Hersteller kann der Abfüllautomat mit derselben Genauigkeit auch 75cl Bierflaschen abfüllen. Die Mönche messen zusätzlich das Volumen von 42 zufällig ausgewählten 75cl Bierflaschen, mit folgendem Resultat (in Zentiliter):

75.71 76.28 75.33 75.37 74.84 73.13 76.34 75.61 75.15 74.56 74.30 74.35 75.43 73.70 74.49 75.18 74.23 76.62 73.29 74.76 75.12 74.12 75.57 75.89 75.67 75.68 74.66 73.93 76.08 75.50 75.52 75.59 75.36 74.95 72.84 76.13 75.37 75.53 73.80 74.56 74.44 73.98

(Standardabweichung=0.9)

Kapitel 2

Wahrscheinlichkeitsraum

§01 Stichprobenraum

Ein zufälliges Experiment ist ein Experiment, in dem das Ergebnis nicht mit Sicherheit vorhergesagt werden kann. Um ein zufälliges Experiment mathematisch zu untersuchen, ist es notwendig alle Versuchsausgänge beschreiben zu können.

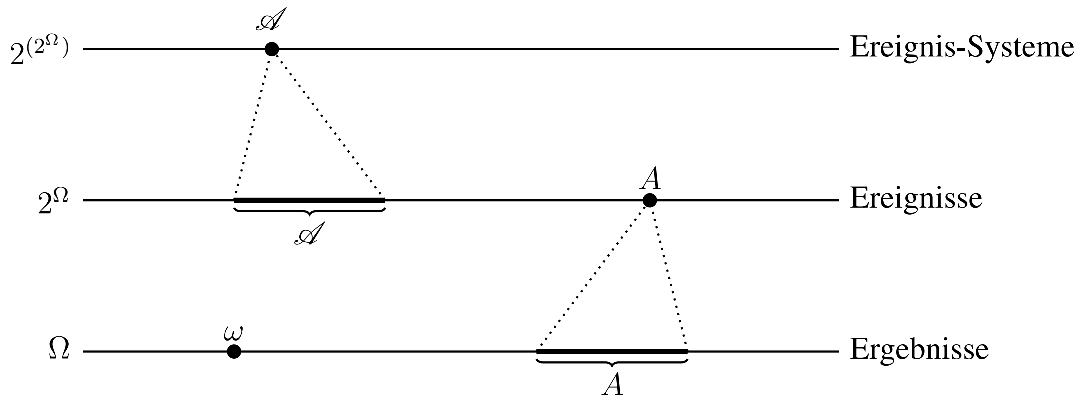
§01.01 **Definition**. Eine nicht-leere Menge Ω aller möglichen Ausgänge eines zufälligen Experiments wird *Ergebnismenge* (*Grundmenge* oder *Stichprobenraum*) genannt. Ein möglicher Versuchsausgang ω des zufälligen Experiments, also ein Element von Ω, kurz $\omega \in \Omega$ heißt *Ergebnis* (*Stichprobe*). Im Folgenden sei stets Ω eine nicht-leere Menge. □

Ein Ereignis ist typischerweise in Form einer Frage gegeben, deren Antwort nur vom Versuchsausgang des zufälligen Experiments abhängt. Von Interesse ist dabei insbesondere, ob das Ereignis eingetreten ist (sich realisiert hat) oder eben nicht.

§01.02 **Definition**. Ein *Ereignis* ist eine Teilmenge der Grundmenge Ω, also ein Element der Potenzmenge 2^Ω von Ω. Für einen Versuchsausgang $\omega \in \Omega$ ist ein Ereignis $A \in 2^\Omega$ eingetreten, wenn $\omega \in A$ gilt. Wir bezeichnen mit $A^c := \Omega \backslash A \in 2^\Omega$ das Komplement von A in Ω. □

§01.03 **Sprechweise**. Die Mengen $\emptyset$, Ω und $\{\omega\}$, $\omega \in \Omega$, heißen das (absolut) *unmögliche Ereignis*, das (absolut) *sichere Ereignis* bzw. ein *Elementarereignis*; Ereignisse A und B mit $A \cap B = \emptyset$ werden *unvereinbar* oder *unverträglich* genannt. Wir schreiben kurz $2^\Omega_{\neq\emptyset} := 2^\Omega \backslash \{\emptyset\}$. □

§01.04 **Skizze**. Die folgende Abbildung stellt drei begriffliche Ebenen der Stochastik dar.



Die Abbildung wurde auf der Grundlage von Georgii (2015, Abb.1.1, S.11) erstellt. □

§01.05 **Definition**. Ein Teilmengensystem $\mathscr{A} \subseteq 2^\Omega$ heißt *σ-Algebra* über Ω, wenn es folgenden Bedingungen genügt:

(σA1) $\Omega \in \mathscr{A}$;

(σA2) $A^c \in \mathscr{A}$ für alle $A \in \mathscr{A}$;

(σA3) $\bigcup_{n \in \mathbb{N}} A_n \in \mathscr{A}$ für alle $A_n \in \mathscr{A}$, $n \in \mathbb{N}$.

Das Paar $(\Omega, \mathscr{A})$ wird *messbarer Raum* genannt und $\mathscr{A}$ wird als *Menge der interessierenden Ereignisse* bezeichnet. □

§01.06 **Lemma**. *Es sei $\mathscr{E} \subseteq 2^\Omega$ ein System von Teilmengen von Ω. Dann ist*

$$\sigma(\mathscr{E}) := \bigcap \{\mathscr{A} : \mathscr{A} \text{ ist } \sigma\text{-Algebra auf } \Omega \text{ und } \mathscr{E} \subseteq \mathscr{A}\}$$

die kleinste σ-Algebra auf Ω, die $\mathscr{E}$ enthält. $\mathscr{E}$ heißt Erzeuger *von $\sigma(\mathscr{E})$ und $\sigma(\mathscr{E})$ die von $\mathscr{E}$* erzeugte σ-Algebra *auf Ω.*

§01.07 **Beweis** von Lemma §01.06. Übungsaufgabe. □

§01.08 **Beispiel**.

(a) Auf jeder nicht-leeren Ergebnismenge Ω existieren die *triviale σ-Algebra* $\{\emptyset, \Omega\}$ sowie die Potenzmenge 2^Ω als σ-Algebren.

(b) Für jedes nicht-leeres $A \subseteq \Omega$ gilt $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$.

(c) Für eine höchstens abzählbar unendliche (d.h. endlich oder abzählbar unendliche), kurz *abzählbare*, Indexmenge $\mathcal{I}$ sei $\mathscr{E} = \{A_i \in 2^\Omega_{\neq\emptyset} : i \in \mathcal{I}, \text{ paarweise disjunkt und } \biguplus_{i\in\mathcal{I}} A_i = \Omega\}$ eine Partition von Ω. Dann ist $\sigma(\mathscr{E}) = \{\biguplus_{j\in\mathcal{J}} A_j : \mathcal{J} \subseteq \mathcal{I}\}$, wobei wir $\emptyset = \biguplus_{j\in\emptyset} A_j$ vereinbaren. □

§01.09 **Bemerkung**. Wie im letzten Beispiel §01.08, bezeichnen wir zur optischen Verdeutlichung die Vereinigung paarweiser disjunkter Mengen mit dem Symbol $\biguplus$. □

§01.10 **Lemma**. *Für jede σ-Algebra $\mathscr{A}$ über Ω gilt:*

(i) *$\emptyset \in \mathscr{A}$;*

(ii) *Für $A, B \in \mathscr{A}$ sind $A \cup B$, $A \cap B$, $A \backslash B := A \cap B^c$ und $A \Delta B := (A\backslash B) \cup (B \backslash A)$ Elemente von $\mathscr{A}$;*

(iii) *Für eine abzählbare Indexmenge $\mathcal{I}$ und $\{A_i : i \in \mathcal{I}\} \subseteq \mathscr{A}$ gilt $\bigcap_{i\in\mathcal{I}} A_i \in \mathscr{A}$.*

§01.11 **Beweis** von Lemma §01.10. Übungsaufgabe. □

§01.12 **Schreibweise**. Für $x, y \in \mathbb{R}$ vereinbaren wir $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \in (-\infty, x]\}$ (ganzzahliger Anteil), $x \vee y = \max(x, y)$ (Maximum), $x \wedge y = \min(x, y)$ (Minimum), $x^+ = \max(x, 0)$ (positiver Teil) und $x^- = \max(-x, 0)$ (negativer Teil) so dass $|x| = x^+ + x^-$.

(a) Für $c \in \mathbb{R}$ und $\mathbb{A} \subseteq \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\} = [-\infty, \infty]$ setzen wir $\mathbb{A}_{\geqslant c} := \mathbb{A} \cap [c, \infty]$, $\mathbb{A}_{\leqslant c} := \mathbb{A} \cap [-\infty, c]$, $\mathbb{A}_{>c} := \mathbb{A} \cap (c, \infty]$, $\mathbb{A}_{<c} := \mathbb{A} \cap [\infty, c)$, $\mathbb{A}_{\neq c} := \mathbb{A} \backslash \{c\}$, und $\overline{\mathbb{A}} := \mathbb{A} \cup \{\pm\infty\}$.

(b) Für $b \in \overline{\mathbb{R}}$ and $a \in \overline{\mathbb{R}}_{<b}$ schreiben wir $\llbracket a, b \rrbracket := [a, b] \cap \overline{\mathbb{Z}}$, $\llbracket a, b)\!) := [a, b) \cap \overline{\mathbb{Z}}$, $(\!(a, b \rrbracket := (a, b] \cap \overline{\mathbb{Z}}$, und $(\!(a, b)\!) := (a, b) \cap \overline{\mathbb{Z}}$. Insbesondere setzen wir $\llbracket n \rrbracket := \llbracket 1, n \rrbracket$ und $\llbracket n)\!) := \llbracket 1, n)\!)$ für $n \in \mathbb{N} = \mathbb{Z}_{>0}$. □

§01.13 **Erinnerung**. Ein Folge $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ heißt *monoton wachsend*, wenn $x_n \leqslant x_{n+1}$ (bzw. *monoton fallend*, wenn $x_{n+1} \leqslant x_n$) für alle $n \in \mathbb{N}$ gilt. Ist eine monotone wachsende (bzw. fallende) Folge konvergent, etwa $x = \lim_{n\to\infty} x_n$, so schreiben wir kurz $x_n \uparrow x$ (bzw. $x_n \downarrow x$). □

§01.14 **Definition**. Eine Folge $(A_n)_{n\in\mathbb{N}}$ in 2^Ω heißt *monoton wachsend*, wenn $A_n \subseteq A_{n+1}$ (bzw. *monoton fallend*, wenn $A_{n+1} \subseteq A_n$) für alle $n \in \mathbb{N}$ gilt. Weiterhin heißen

$$\boxed{A_\star := \liminf_{n\to\infty} A_n} := \bigcup_{n\in\mathbb{N}} \bigcap_{m\in\mathbb{N}_{\geqslant n}} A_m := \bigcup\{\bigcap\{A_m : m\in\mathbb{N}_{\geqslant n}\} : n\in\mathbb{N}\} \quad \text{und}$$

$$\boxed{A^\star := \limsup_{n\to\infty} A_n} := \bigcap_{n\in\mathbb{N}} \bigcup_{m\in\mathbb{N}_{\geqslant n}} A_m$$

Limes inferior bzw. *Limes superior* der Folge $(A_n)_{n\in\mathbb{N}}$. Die Folge $(A_n)_{n\in\mathbb{N}}$ heißt *konvergent*, wenn $A_\star = A^\star =: A$ gilt. In diesem Fall schreiben wir kurz $\boxed{\lim_{n\to\infty} A_n} := A$. □

§01.15 **Bemerkung**.

(a) Jede *monoton wachsende* (bzw. *fallende*) Folge $(A_n)_{n\in\mathbb{N}}$ von Teilmengen von Ω ist konvergent mit $A := \lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n$ (bzw. $A := \lim_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n$). In diesem Fall schreiben wir kurz $A_n \uparrow A$ (bzw. $A_n \downarrow A$).

(b) Für den Limes inferior bzw. superior einer Folge $(A_n)_{n\in\mathbb{N}}$ gilt

$$A_\star = \liminf_{n\to\infty} A_n = \{\omega\in\Omega : |\{n\in\mathbb{N} : \omega\notin A_n\}| \in \mathbb{N}\} \quad \text{bzw.}$$

$$A^\star = \limsup_{n\to\infty} A_n = \{\omega\in\Omega : |\{n\in\mathbb{N} : \omega\in A_n\}| = \infty\}.$$

In Worten, $A_\star$ ist das Ereignis, dass *schließlich alle der A_n eintreten*. $A^\star$ ist das Ereignis, dass *unendlich viele der A_n eintreten*. Es gilt $A_\star = \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n = A^\star$. □

§01.16 **Lemma**. *Sei $(\Omega, \mathscr{A})$ ein messbarer Raum und $(A_n)_{n\in\mathbb{N}}$ eine Folge von Teilmengen aus $\mathscr{A}$. Dann gilt:*

(i) $\liminf_{n\to\infty} A_n \in \mathscr{A}$ *und* $\limsup_{n\to\infty} A_n \in \mathscr{A}$*;*

(ii) *Falls $A := \lim_{n\to\infty} A_n$ existiert, dann ist $A \in \mathscr{A}$.*

§01.17 **Beweis** von Lemma §01.16. Übungsaufgabe. □

§01.18 **Definition**. Es sei $\mathcal{S}$ ein metrischer (oder topologischer) Raum und $\mathscr{O}$ das System der offenen Teilmengen von $\mathcal{S}$. Dann heißt die von den offen Mengen $\mathscr{O}$ erzeugte σ-Algebra $\boxed{\mathscr{B}_\mathcal{S}} := \sigma(\mathscr{O})$ die *Borel-σ-Algebra* über $\mathcal{S}$. Die Elemente von $\mathscr{B}_\mathcal{S}$ heißen *Borel-Mengen*. □

§01.19 **Bemerkung**. Häufig sind wir an der Borel-σ-Algebra $\boxed{\mathscr{B}^n} := \mathscr{B}_{\mathbb{R}^n}$ über $\mathbb{R}^n$ interessiert, wobei $\mathbb{R}^n$ versehen ist mit dem euklidischen Abstand $d(x,y) = \|x-y\| = \sqrt{\sum_{i\in[\![n]\!]}(x_i-y_i)^2}$ für $x = (x_i)_{i\in[\![n]\!]}$, $y = (y_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$. □

§01.20 **Schreibweise**. Für $a, b \in \overline{\mathbb{R}}^n$ schreiben wir $a < b$, wenn $a_i < b_i$ für alle $i \in [\![n]\!]$ gilt. Für $a < b$, definieren wir den offenen *Quader* als das Kartesische Produkt $(a,b) := \bigtimes_{i\in[\![n]\!]}(a_i, b_i) := (a_1,b_1)\times(a_2,b_2)\times\cdots\times(a_n,b_n)$. Analog, sind $[a,b]$, $(a,b]$ sowie $[a,b)$ definiert. Weiterhin, sei $(-\infty, b) := \bigtimes_{i\in[\![n]\!]}(-\infty, b_i)$ und analog $(-\infty, b]$ definiert. □

§01.21 **Bemerkung**. Wie in Schreibweise §01.20 bezeichnet $\bigtimes_{i\in\mathcal{I}} \mathcal{S}_i$ das Kartesische Produkt der Mengen $\mathcal{S}_i$, $i\in\mathcal{I}$, und für $s \in \bigtimes_{i\in\mathcal{I}} \mathcal{S}_i$ bezeichnen s_i, $i\in\mathcal{I}$, stets die Komponenten von $s = (s_i)_{i\in\mathcal{I}}$. □

§01.22 **Satz.** *Die Borel-σ-Algebra $\mathscr{B}^n$ über $\mathbb{R}^n$ wird auch erzeugt von folgenden Mengensystemen:*
(i) $\mathscr{E}_1 := \{A \subseteq \mathbb{R}^n : A \text{ ist abgeschlossen}\}$; (ii) $\mathscr{E}_2 := \{A \subseteq \mathbb{R}^n : A \text{ ist kompakt}\}$;
(iii) $\mathscr{E}_3 := \{(a,b) : a,b \in \mathbb{Q}^n, a < b\}$; (iv) $\mathscr{E}_4 := \{[a,b] : a,b \in \mathbb{Q}^n, a < b\}$;
(v) $\mathscr{E}_5 := \{(a,b] : a,b \in \mathbb{Q}^n, a < b\}$; (vi) $\mathscr{E}_6 := \{[a,b) : a,b \in \mathbb{Q}^n, a < b\}$;
(vii) $\mathscr{E}_7 := \{(-\infty,b] : b \in \mathbb{Q}^n\}$; (viii) $\mathscr{E}_8 := \{(-\infty,b) : b \in \mathbb{Q}^n\}$;
(ix) $\mathscr{E}_9 := \{(a,\infty) : a \in \mathbb{Q}^n\}$ *und* (x) $\mathscr{E}_{10} := \{[a,\infty) : a \in \mathbb{Q}^n\}$.

§01.23 **Beweis** von Satz §01.22. In der Vorlesung Wahrscheinlichkeitstheorie 1. □

§01.24 **Lemma.** *Seien $(\Omega, \mathscr{A})$ ein messbarer Raum und $B \in 2^{\Omega}_{\neq\emptyset}$ eine nicht-leere Teilmenge. Dann ist*

$$\mathscr{A}_B := \mathscr{A}\big|_B := \mathscr{A} \cap B := \{A \cap B : A \in \mathscr{A}\}$$

eine σ-Algebra über B, die sogenannte Spur oder Einschränkung von $\mathscr{A}$ auf B. Des Weiteren, für $\mathscr{E} \subseteq 2^{\Omega}$ gilt $\sigma(\mathscr{E})\big|_B = \sigma(\mathscr{E}|_B)$.

§01.25 **Beweis** von Lemma §01.24. Übungsaufgabe. □

§01.26 **Schreibweise.** Wir bezeichnen mit $\overline{\mathscr{B}} := \mathscr{B}_{\overline{\mathbb{R}}}$ die Borel-σ-Algebra über der kompaktifizierten Zahlengerade $\overline{\mathbb{R}} := [-\infty, \infty]$, wobei in $\overline{\mathbb{R}}$ die Mengen $\{-\infty\}$ und $\{\infty\}$ abgeschlosssen und $\mathbb{R}$ offen, also Borel-Mengen sind. Insbesondere, ist $\mathscr{B} := \mathscr{B}_{\mathbb{R}} = \overline{\mathscr{B}} \cap \mathbb{R}$ die Borel-σ-Algebra über $\mathbb{R}$. Für $c \in \mathbb{R}$ und σ-Algebra $\mathscr{A}$ über einer Teilmenge von $\overline{\mathbb{R}}$ schreiben wir weiterhin $\mathscr{A}_{\geqslant c} := \mathscr{A} \cap \overline{\mathbb{R}}_{\geqslant c}$, $\mathscr{A}_{>c} := \mathscr{A} \cap \overline{\mathbb{R}}_{>c}$, $\mathscr{A}_{\leqslant c} := \mathscr{A} \cap \overline{\mathbb{R}}_{\leqslant c}$, und $\mathscr{A}_{<c} := \mathscr{A} \cap \overline{\mathbb{R}}_{<c}$ □

§02 Wahrscheinlichkeit

In einem zufälligem Experiment, das durch einen messbaren Raum $(\Omega, \mathscr{A})$ beschrieben wird, möchten wir den interessierenden Ereignissen eine Wahrscheinlichkeit zu ordnen. Die folgende Definition wurde 1933 von dem russischen Mathematiker A.N. Kolmogorov eingeführt.

§02.01 **Definition.** Sei $(\Omega, \mathscr{A})$ ein messbarer Raum. Eine Abbildung $\mathbb{P} : \mathscr{A} \to [0,1]$ heißt *Wahrscheinlichkeitsmaß* (*Wahrscheinlichkeitsverteilung* oder kurz *Verteilung*) auf $\mathscr{A}$, wenn sie folgenden Bedingungen genügt:

(Wm1) $\mathbb{P}(\Omega) = 1$; (*normiert*)

(Wm2) $\mathbb{P}(\biguplus_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ für jede Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{A}$ paarweise disjunkter Ereignisse, d.h. $A_n \cap A_m = \emptyset$ für alle $n \in \mathbb{N}$ und $m \in \mathbb{N}_{\neq n}$. (*$\sigma$-additiv*) □

Ein Tripel $(\Omega, \mathscr{A}, \mathbb{P})$ bestehend aus einer Ergebnismenge Ω, einer σ-Algebra $\mathscr{A}$ interessierender Ereignisse sowie einem Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathscr{A}$ wird *Wahrscheinlichkeitsraum* genannt. Weiterhin bezeichnen wir mit $\mathcal{W}(\mathscr{A})$ die Menge aller Wahrscheinlichkeitsmaße auf $\mathscr{A}$. □

§02.02 **Sprechweise.** Ereignisse $A, B \in \mathscr{A}$ mit $\mathbb{P}(A) = 0$ und $\mathbb{P}(B) = 1$ werden *Nullmenge* bzw. *Einsmenge* genannt. □

§02.03 **Beispiel.** Sei $(\Omega, \mathscr{A})$ ein messbarer Raum. Für $A \in 2^{\Omega}$ bezeichnet

$$\mathbb{1}_A : \Omega \to \{0,1\} \quad \text{mit} \quad \mathbb{1}_A^{-1}(\{1\}) := A \quad \text{und} \quad \mathbb{1}_A^{-1}(\{0\}) := A^c$$

die *Indikatorfunktion* auf A. Für jedes $\omega \in \Omega$ ist das *Einpunkt*- oder *Diracmaß*

$$\delta_\omega : \mathscr{A} \to \{0,1\} \quad \text{mit} \quad A \mapsto \delta_\omega(A) := \mathbb{1}_A(\omega)$$

ein Wahrscheinlichkeitsmaß in $\mathcal{W}(\mathscr{A})$. Ist $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaße in $\mathcal{W}(\mathscr{A})$, so ist auch jede Konvexkombination $\sum_{n\in\mathbb{N}} \mathrm{w}_n\mathbb{P}_n$ in $\mathcal{W}(\mathscr{A})$ mit $\mathrm{w}_n \in \mathbb{R}_{\geqslant 0}$, $n \in \mathbb{N}$, und $\sum_{n\in\mathbb{N}} \mathrm{w}_n = 1$. Die Diracmaße bilden Extremalpunkte der konvexen Menge aller Wahrscheinlichkeitsmaße. □

§02.04 **Definition**. Für ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{A})$ auf $(\Omega, \mathscr{A})$ heißt ein Element $\omega \in \Omega$ mit $\{\omega\} \in \mathscr{A}$ und $\mathbb{P}(\{\omega\}) \in (0, 1]$ *Atom*. Die Wahrscheinlichkeit $\mathbb{P}(\{\omega\})$ wird *Masse* des Atoms ω genannt. □

§02.05 **Lemma**. *Ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathscr{A})$ besitzt höchstens abzählbar viele Atome.*

§02.06 **Beweis** von Lemma §02.05. In der Vorlesung. □

§02.07 **Lemma**. *Für jedes Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{A})$ gilt:*

(i) $\mathbb{P}(\emptyset) = 0$; *(unmögliches Ereignis)*

(ii) *Für jedes $A \in \mathscr{A}$ gilt $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;*

(iii) *Für alle $A, B \in \mathscr{A}$ gilt* *(Boole)*
$\mathbb{P}(A\backslash B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$,
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
$\mathbb{P}(A\Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$;
für disjunkte A und B, d.h. $A \cap B = \emptyset$, also $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$; *(additiv)*

(iv) *Für $n \in \mathbb{N}$ und $A_i \in \mathscr{A}$, $i \in [\![n]\!]$, gilt* *(Poincaré und Sylvester)*
$\mathbb{P}(\bigcup_{i\in[\![n]\!]} A_i) = \sum_{\mathcal{I}\in 2^{[\![n]\!]}_{\neq\emptyset}} (-1)^{|\mathcal{I}|-1}\mathbb{P}(\bigcap_{i\in\mathcal{I}} A_i)$;

(v) *Für alle $A, B \in \mathscr{A}$ mit $A \subseteq B$ gilt $0 \leqslant \mathbb{P}(A) \leqslant \mathbb{P}(B) \leqslant 1$* *(monotone);*

(vi) *Für $A, B \in \mathscr{A}$ gilt:* *(Bonferroni)*
$\mathbb{P}(A) \vee \mathbb{P}(B) \leqslant \mathbb{P}(A \cup B) \leqslant \{\mathbb{P}(A) + \mathbb{P}(B)\} \wedge 1$;
$\{\mathbb{P}(A) + \mathbb{P}(B) - 1\} \vee 0 \leqslant \mathbb{P}(A \cap B) \leqslant \mathbb{P}(A) \wedge \mathbb{P}(B)$;
$|\mathbb{P}(A) - \mathbb{P}(B)| \leqslant \mathbb{P}(A\Delta B)$.

§02.08 **Beweis** von Lemma §02.07. Übungsaufgabe. □

§02.09 **Proposition**. *Für jedes $\mathbb{P} \in \mathcal{W}(\mathscr{A})$, $A \in \mathscr{A}$ und jede Folge $(A_n)_{n\in\mathbb{N}}$ von Teilmengen aus $\mathscr{A}$ gilt:*

(i) $\mathbb{P}(\bigcup_{n\in\mathbb{N}} A_n) \leqslant \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$; *($\sigma$-subadditiv)*

(ii) *Falls $A_n \downarrow \emptyset$, dann gilt $\mathbb{P}(A_n) \downarrow 0$;* *($\sigma$-stetig)*
falls $A_n \uparrow A$, dann gilt $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$;

(iii) *Falls $\lim_{n\to\infty} A_n = A$, dann gilt $\lim_{n\to\infty} \mathbb{P}(A_n\Delta A) = 0$ und somit $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.*

§02.10 **Beweis** von Proposition §02.09. (iii) in der Vorlesung und (i)-(ii) Übungsaufgabe. □

§02.11 **Lemma**. *Jede normierte, additive Mengenfunktion $\mathbb{P} : \mathscr{A} \to [0, 1]$, die σ-stetig ist, ist auch σ-additiv und damit ein Wahrscheinlichkeitsmaß, also $\mathbb{P} \in \mathcal{W}(\mathscr{A})$.*

§02.12 **Beweis** von Lemma §02.11. In der Vorlesung. □

§02.13 **Bemerkung**. Die letzte Aussage in Lemma §02.11 erlaubt eine alternative aber zu Definition §02.01 äquivalente Definition eines Wahrscheinlichkeitsmaßes, d.h. eine Mengenfunktion $\mathbb{P} : \mathscr{A} \to [0, 1]$ ist ein Wahrscheinlichkeitsmaß wenn es (a) normiert, (b) additiv und (c) σ-stetig ist. □

§02.14 **Korollar** *(Ungleichungen von Boole)*. *Für jedes $\mathbb{P} \in \mathcal{W}(\mathscr{A})$ und jede Folge $(A_n)_{n\in\mathbb{N}}$ von Teilmengen aus $\mathscr{A}$ gilt:*

$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) \leqslant \inf_{n\in\mathbb{N}} \mathbb{P}(A_n) \leqslant \sup_{n\in\mathbb{N}} \mathbb{P}(A_n) \leqslant \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leqslant \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

$$\mathbb{P}(\liminf_{n\to\infty} A_n) \leqslant \liminf_{n\to\infty} \mathbb{P}(A_n) \leqslant \limsup_{n\to\infty} \mathbb{P}(A_n) \leqslant \mathbb{P}(\limsup_{n\to\infty} A_n).$$

§02.15 **Beweis** von Lemma §02.11. Übungsaufgabe. □

§02.16 **Definition**. Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathscr{B})$ wird

$$\boxed{\mathbb{F}} : \mathbb{R} \to [0,1] \text{ mit } x \mapsto \mathbb{F}(x) := \mathbb{P}((-\infty, x]) = \mathbb{P}(\mathbb{R}_{\leqslant x})$$

die zugehörige *Verteilungsfunktion* genannt. □

§02.17 **Lemma**. *Für jede Verteilungsfunktion $\mathbb{F}$ eines Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ gelten:*

(i) *$\mathbb{F}$ ist monoton wachsend, d.h., $\mathbb{F}(x) \leqslant \mathbb{F}(y)$ gilt für alle $y \in \mathbb{R}$, $x \in \mathbb{R}_{\leqslant y}$.*

(ii) *$\mathbb{F}$ ist rechtsstetig, d.h., $\mathbb{F}(x_n) \downarrow \mathbb{F}(x)$ gilt für alle $x_n \downarrow x$ in $\mathbb{R}$.*

(iii) *$\mathbb{F}$ besitzt einen linksseitigen Grenzwert, d.h. $\mathbb{F}(x_n) \uparrow \mathbb{F}(x-) := \mathbb{P}((-\infty, x)) = \mathbb{P}(\mathbb{R}_{<x})$ gilt für alle $x_n \uparrow x$ in $\mathbb{R}$.*

(iv) *Die Sprünge von $\mathbb{F}$ korrespondieren mit den Atomen von $\mathbb{P}$, d.h. $\mathbb{F}(x) - \mathbb{F}(x-) = \mathbb{P}(\{x\})$ für alle $x \in \mathbb{R}$.*

(v) *Es gilt $\lim_{x\to-\infty} \mathbb{F}(x) = 0$ und $\lim_{x\to\infty} \mathbb{F}(x) = 1$.*

§02.18 **Beweis** von Lemma §02.17. Übungsaufgabe. □

§02.19 **Bemerkung**. Aufgrund von Lemma §02.05 und Lemma §02.17 (iv) ist die Menge der Unstetigkeitsstellen einer Verteilungsfunktion $\mathbb{F}$ abzählbar. □

§03 Dynkin'scher π-λ-Satz

Seien $(\Omega, \mathscr{A})$ ein messbarer Raum und $\mathscr{E} \subseteq 2^\Omega$ ein Erzeuger von $\mathscr{A}$, also $\mathscr{A} = \sigma(\mathscr{E})$, so stellt sich die Frage: Ist ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathscr{A}$ schon eindeutig festgelegt durch seine Werte auf $\sigma(\mathscr{E})$? Im Allgemeinen ist die Antwort „nein". Die Antwort ist aber „ja", wenn das Mengensystem schnittstabil ist. Dies ist eine Schlussfolgerung aus dem Dynkin'schen π-λ-Satz.

§03.01 **Definition**. Ein Teilmengensystem $\mathscr{E}$ aus 2^Ω heißt

$\cap$-stabil (sprich: *schnittstabil*) oder ein *π-System*, wenn $A \cap B \in \mathscr{E}$ für je zwei Mengen $A, B \in \mathscr{E}$ gilt;

σ-$\cap$-stabil (sprich: *sigma-schnittstabil*), wenn $\bigcap_{n\in\mathbb{N}} A_n \in \mathscr{E}$ für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Teilmengen aus $\mathscr{E}$ gilt;

$\cup$-stabil (sprich: *vereinigungsstabil*), wenn $A \cup B \in \mathscr{E}$ für je zwei Mengen $A, B \in \mathscr{E}$ gilt;

σ-$\cup$-stabil (sprich: *sigma-vereinigungsstabil*), wenn $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{E}$ für jede Folge $(A_n)_{n\in\mathbb{N}}$ von Teilmengen aus $\mathscr{E}$ gilt;

$\backslash$-stabil (sprich: *differenzmengenstabil*), wenn $A \backslash B \in \mathscr{E}$ für je zwei Mengen $A, B \in \mathscr{E}$ gilt;

komplementstabil, wenn mit jeder Menge $A \in \mathscr{E}$ auch $A^c \in \mathscr{E}$ gilt. □

§03.02 **Bemerkung.**

(a) Ein Teilmengensystem $\mathscr{E}$ aus 2^Ω mit $\Omega \in \mathscr{E}$, das komplementstabil und σ-$\cup$-stabil ist, ist somit eine σ-Algebra.

(b) Für ein komplementstabiles Teilmengensystem $\mathscr{E}$ aus 2^Ω folgen aus den de Morgan'schen Regeln die Äquivalenzen von $\cup$-stabil und $\cap$-stabil als auch von σ-$\cup$-stabil und σ-$\cap$-stabil. □

§03.03 **Definition.** Ein Teilmengensystem $\mathscr{D}$ aus 2^Ω heißt Dynkin-System (oder *λ-System*), wenn es folgenden Bedingungen genügt:

(DS1) $\Omega \in \mathscr{D}$;

(DS2) $\mathscr{D}$ ist komplementstabil;

(DS3) Für jede Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{D}$ paarweise disjunkter Mengen gilt $\biguplus_{n\in\mathbb{N}} A_n \in \mathscr{D}$. □

§03.04 **Bemerkung.** Die Bedingung (DS2), dass $\mathscr{D}$ komplementstabil ist; kann äquivalent ersetzt werden durch die scheinbar stärkere Bedingung

(DS2') für alle $A, B \in \mathscr{D}$ mit $A \subseteq B$ gilt $B\backslash A \in \mathscr{D}$.

Da jedes Dynkin-System auch (DS2') erfüllt. Denn für $A, B \in \mathscr{D}$ mit $A \subseteq B$ sind A und B^c disjunkt und es gilt $B\backslash A = (A \uplus B^c)^c \in \mathscr{D}$. □

§03|00.05 **Anmerkung.** Jede σ-Algebra ist ein Dynkin-System. Die Umkehrung gilt nicht, da (DS3) nur für Folgen paarweise disjunkter Ereignisse gefordert ist. Zum Beispiel für $\Omega = \{1, 2, 3, 4\}$ ist $\mathscr{D} = \{\emptyset, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \Omega\}$ ein Dynkin-System aber keine σ-Algebra. Der Unterschied ist allerdings nicht sehr groß. □

§03.06 **Lemma.** *Ein Dynkin-System $\mathscr{D} \subseteq 2^\Omega$ ist genau dann $\cap$-stabil, wenn es eine σ-Algebra ist.*

§03.07 **Beweis** von Lemma §03.06. In der Vorlesung. □

§03.08 **Korollar.** *Für ein Teilmengensystem $\mathscr{E}$ aus 2^Ω ist*

$$\delta(\mathscr{E}) := \bigcap \{\mathscr{D} : \mathscr{D} \text{ ist Dynkin-System auf } \Omega \text{ und } \mathscr{E} \subseteq \mathscr{D}\}$$

das kleinstes Dynkin-System auf Ω, das $\mathscr{E}$ enthält. $\mathscr{E}$ heißt Erzeuger von $\delta(\mathscr{E})$ und $\delta(\mathscr{E})$ das von $\mathscr{E}$ erzeugte Dynkin-System auf Ω.

§03.09 **Beweis** von Korollar §03.08. Der Beweis §01.07 lässt sich direkt auf Dynkin-Systeme übertragen. □

§03.10 **Bemerkung.** Da jede σ-Algebra ein Dynkin-System ist, gilt stets $\delta(\mathscr{E}) \subseteq \sigma(\mathscr{E})$. □

§03.11 **π-λ-Satz.** *Für jedes $\cap$-stabile $\mathscr{E}$ ist das erzeugte Dynkin-System $\delta(\mathscr{E})$ auch $\cap$-stabil.*

§03.12 **Beweis** von Satz §03.11. In der Vorlesung. □

§03.13 **Korollar** *(Dynkin'scher π-λ-Satz). Für jedes $\cap$-stabile $\mathscr{E}$ gilt $\sigma(\mathscr{E}) = \delta(\mathscr{E})$.*

§03.14 **Beweis** von Korollar §03.13. In der Vorlesung. □

§03.15 **Korollar.** *Seien $\mathscr{D}$ ein Dynkin-System und $\mathscr{E} \subseteq \mathscr{D}$ $\cap$-stabil. Dann gilt $\sigma(\mathscr{E}) \subseteq \mathscr{D}$.*

§03.16 **Beweis** von Korollar §03.15. In der Vorlesung. □

§03.17 **Beweisstrategie**. Möchten wir zeigen, dass alle Ereignisse einer σ-Algebra eine bestimmte Eigenschaft, sagen wir (R) besitzen, so ist eine häufig angewendete Beweisstrategie:

(Schritt 1) Zeige, dass $\{A \in \mathscr{A} : A \text{ erfüllt } (R)\}$ ein Dynkin-System ist;

(Schritt 2) Finde einen $\cap$-stabilen Erzeuger $\mathscr{E}$ von $\mathscr{A}$, d.h. ein π-System $\mathscr{E}$ mit $\mathscr{A} = \sigma(\mathscr{E})$;

(Schritt 3) Zeige, dass alle Elemente aus $\mathscr{E}$ die Eigenschaft (R) besitzen.

Nach Korollar §03.15 besitzen dann alle Ereignisse aus $\mathscr{A} = \sigma(\mathscr{E})$ die Eigenschaft (R). □

§03.18 **Satz** *(Eindeutigkeit eines Wahrscheinlichkeitsmaßes). Seien $(\Omega, \mathscr{A})$ ein messbarer Raum, $\mathscr{E}$ ein $\cap$-stabiler Erzeuger von $\mathscr{A}$ und $\mathbb{P}$, $\widetilde{\mathbb{P}}$ Wahrscheinlichkeitsmaße auf $\mathscr{A}$. Falls die Einschränkungen $\mathbb{P}|_{\mathscr{E}}$ und $\widetilde{\mathbb{P}}|_{\mathscr{E}}$ von $\mathbb{P}$ bzw. $\widetilde{\mathbb{P}}$ auf $\mathscr{E}$ übereinstimmen, d.h. $\mathbb{P}(E) = \widetilde{\mathbb{P}}(E)$ für alle $E \in \mathscr{E}$ gilt, dann stimmen auch $\mathbb{P}$ und $\widetilde{\mathbb{P}}$ überein, $\mathbb{P} = \widetilde{\mathbb{P}}$.*

§03.19 **Beweis** von Satz §03.18. Übungsaufgabe unter Verwendung der Beweisstrategie §03.17 mit

$$(R) = \text{„}\mathbb{P}(A) = \widetilde{\mathbb{P}}(A)\text{“}.$$ □

§03.20 **Korollar** *(Eindeutigkeit der Verteilungsfunktion). Ist $\mathbb{F} : \mathbb{R} \to [0,1]$ monoton wachsend, rechtsstetig mit $\lim_{x\to-\infty} \mathbb{F}(x) = 0$ und $\lim_{x\to\infty} \mathbb{F}(x) = 1$, so existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ auf $(\mathbb{R}, \mathscr{B})$ mit $\mathbb{P}((x,y]) = \mathbb{F}(y) - \mathbb{F}(x)$ für alle $y \in \mathbb{R}$ und $x \in \mathbb{R}_{<y}$. Insbesondere ist $\mathbb{F}$ die Verteilungsfunktion von $\mathbb{P}$.*

§03.21 **Beweis** von Korollar §03.20. Die *Existenz* wird in der Vorlesung Wahrscheinlichkeitstheorie gezeigt, die *Eindeutigkeit* folgt direkt aus Satz §03.18 mit $\cap$-stabilen $\mathscr{E} = \{(-\infty, x] : x \in \mathbb{R}\}$. □

§04 Diskreter Wahrscheinlichkeitsraum

§04.01 **Definition**. Ist Ω eine abzählbare Menge, so wird ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf 2^Ω *diskret* genannt und $(\Omega, 2^\Omega, \mathbb{P})$ heißt *diskreter Wahrscheinlichkeitsraum*. Die Abbildung

$$\boxed{\mathbb{p}} : \Omega \to [0,1] \quad \text{mit} \quad \omega \mapsto \mathbb{p}(\omega) := \mathbb{P}(\{\omega\})$$

wird *Zähldichte* des Wahrscheinlichkeitsmaßes $\mathbb{P} \in \mathcal{W}(2^\Omega)$ genannt. □

§04.02 **Lemma**.

(i) *Ist $(\Omega, 2^\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum, so ist $\mathbb{P}$ eindeutig durch seine Zähldichte $\mathbb{p}$ festgelegt.*

(ii) *Ist andererseits Ω eine abzählbare Menge und besitzt $\mathbb{p} : \Omega \to [0,1]$ die Eigenschaft $\sum_{\omega\in\Omega} \mathbb{p}(\omega) = 1$, so wird durch $A \mapsto \mathbb{P}(A) := \sum_{\omega\in A} \mathbb{p}(\omega)$ ein diskretes Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(2^\Omega)$ definiert, dessen Zähldichte gerade $\mathbb{p}$ ist.*

§04.03 **Beweis** von Lemma §04.02. Übungsaufgabe. □

§04.04 **Lemma**. *In einer Urne liegen $N \in \mathbb{N}$ Kugeln mit den Aufschriften $1, 2, \dots, N$. Es werden $n \in \mathbb{N}$ Kugeln gezogen. Dann gilt für die Grundmenge und die Anzahl verschiedenen Ergebnisse:*

Ziehen mit Zurücklegen, mit Betrachtung der Reihenfolge:
$\Omega_1 = \{(k_i)_{i\in[\![n]\!]} : k_i \in [\![N]\!], i \in [\![n]\!]\} = [\![N]\!]^n$, $|\Omega_1| = N^n$;

Ziehen ohne Zurücklegen, mit Betrachtung der Reihenfolge für $n \in [\![N]\!]$:
$\Omega_2 = \{(k_i)_{i\in[\![n]\!]} \in [\![N]\!]^n : k_1, \dots, k_n \text{ paarweise verschieden}\}$, $|\Omega_2| = \frac{N!}{(N-n)!}$;

Ziehen ohne Zurücklegen, ohne Betrachtung der Reihenfolge für $n \in [\![N]\!]$*:*
$\Omega_3 = \{A \subseteq [\![N]\!] : |A| = n\}$, $|\Omega_3| = \binom{N}{n}$*;*

Ziehen mit Zurücklegen, ohne Betrachtung der Reihenfolge:
$\Omega_4 = \{(k_i)_{i\in[\![n]\!]} \in [\![N]\!]^n : k_1 \leqslant k_2 \leqslant \dots \leqslant k_n\}$, $|\Omega_4| = \binom{N+n-1}{n}$.

§04.05 **Beweis** von Lemma §04.04. In der Vorlesung. □

§04.06 **Beispiel.**

(a) Die Menge aller Ergebnisse im Lotto „6 aus 49“ entspricht Ω_3 mit $N = 49$ und $n = 6$, so dass es $\binom{49}{6} = 13983816$ verschiedene Ergebnisse gibt.

(b) Wie groß ist die Wahrscheinlichkeit, dass in dem Hörsaal keine zwei Personen am selben Tag Geburtstag haben? Nehmen wir an, dass Jahr hat 365 Tage, so ist die Menge aller Geburtstagskombinationen für $n \in \mathbb{N}$ Personen gerade $\Omega_1 = [\![N]\!]^n$ mit $N = 365$. Das Ereignis „keine zwei Personen haben am selben Tag Geburtstag“ entspricht $\Omega_2 = \{(k_i)_{i\in[\![n]\!]} \in [\![N]\!]^n : k_1, \dots, k_n \text{ paarweise verschieden}\}$. Unter der Annahme einer Gleichverteilung folgt $\frac{|\Omega_2|}{|\Omega_1|} = \frac{N!}{(N-n)!N^n} = 1 \cdot (1 - \frac{1}{N}) \cdots (1 - \frac{n-1}{N})$. Für $n = 25$, $n = 50$ und $n = 100$ ergibt sich approximativ 0.431, 0.03 und $3.1 \cdot 10^{-7}$. □

§04|00.07 **Anmerkung.** Sei $\emptyset \neq \Omega \subseteq \mathbb{R}$ abzählbar und $\mathbb{p}$ eine Zähldichte auf Ω. Das Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ auf $(\mathbb{R}, \mathscr{B})$ mit

$$\mathbb{P}(B) := \sum_{\omega\in\Omega} \mathbb{p}(\omega)\delta_\omega(B), \quad \forall B \in \mathscr{B}$$

und die zugehörige Verteilungsfunktion $\mathbb{F} : \mathbb{R} \to [0,1]$ mit

$$\mathbb{F}(x) = \mathbb{P}(\mathbb{R}_{\leqslant x}) = \sum_{\omega\in\Omega} \mathbb{p}(\omega)\delta_\omega(\mathbb{R}_{\leqslant x}) = \sum_{\omega\in\Omega} \mathbb{1}_{(-\infty,x]}(\omega)\mathbb{p}(\omega) =: \sum_{\omega\leqslant x} \mathbb{p}(\omega), \quad \forall x \in \mathbb{R}$$

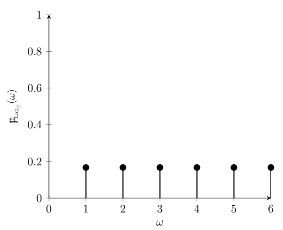
werden ebenfalls *diskret* genannt. □

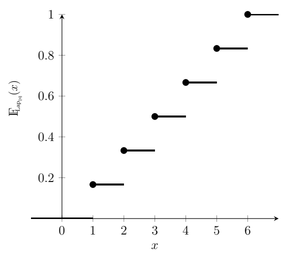
§04.08 **Beispiel.** Folgende Zähldichten beschreiben häufig auftretende diskrete Verteilungen:

(a) *Laplace-/Gleich-Verteilung*, kurz Lap_Ω, mit $|\Omega| \in \mathbb{N}$:

$$\mathbb{p}_{\mathrm{Lap}_\Omega}(\omega) = \frac{1}{|\Omega|} \text{ für } \omega \in \Omega.$$

Betrachten wir den *Wurf eines Würfels*, so ist $\Omega = [\![6]\!]$. Ist der Würfel fair, so erhalten wir die Zähldichte $\mathbb{p}_{\mathrm{Lap}_{[6]}}(\omega) = 1/6$, $\omega \in [\![6]\!]$, mit der zugehörigen Verteilungsfunktion $\mathbb{F}_{\mathrm{Lap}_{[6]}}(x) = \frac{1}{6}\mathbb{1}_{[1,2)}(x) + \frac{2}{6}\mathbb{1}_{[2,3)}(x) + \frac{3}{6}\mathbb{1}_{[3,4)}(x) + \frac{4}{6}\mathbb{1}_{[4,5)}(x) + \frac{5}{6}\mathbb{1}_{[5,6)}(x) + \mathbb{1}_{\mathbb{R}_{\geqslant 6}}(x)$, $x \in \mathbb{R}$:

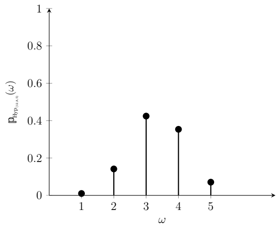


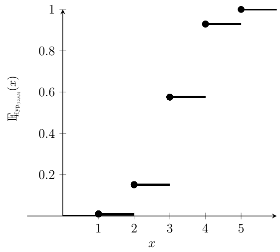


(b) *hypergeometrische Verteilung*, kurz $\mathrm{Hyp}_{(N,K,n)}$, mit Parametern $N \in \mathbb{N}$, $n, K \in [\![0, N]\!]$ und $\Omega = [\![0 \vee (n + K - N), n \wedge K]\!]$:

$$\mathbb{p}_{\mathrm{Hyp}_{(N,K,n)}}(\omega) = \frac{\binom{N-K}{n-\omega}\binom{K}{\omega}}{\binom{N}{n}} \text{ für } \omega \in [\![0 \vee (n + K - N), n \wedge K]\!].$$

Ziehen wir n Kugeln aus einer Urne mit K weißen und $N - K$ schwarzen Kugeln, so ist die Wahrscheinlichkeit, dass davon ω weiß sind, gerade durch $\mathbb{p}_{\mathrm{Hyp}_{(N,K,n)}}(\omega)$ gegeben. Für $N = 12$, $K = 8$ und $n = 5$ erhalten wir die Zähldichte und die zugehörigen Verteilungsfunktion:

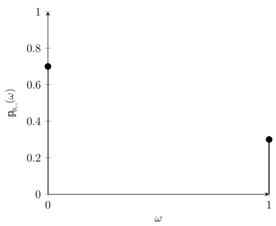


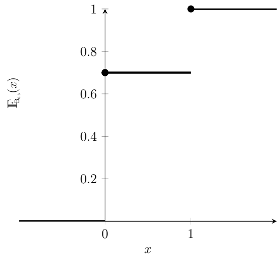


(c) *Bernoulli-Schema*, kurz $\mathrm{B}_{\mathrm{p}}^{n}$, mit Erfolgswahrscheinlichkeit $\mathrm{p} \in [0, 1]$, Länge $n \in \mathbb{N}$ und $\Omega = \{0, 1\}^n$:

$$\mathbb{p}_{\mathrm{B}_{\mathrm{p}}^{n}}(\omega) = \mathrm{p}^{\sum_{i \in [\![n]\!]} \omega_i}(1 - \mathrm{p})^{n - \sum_{i \in [\![n]\!]} \omega_i} \text{ für } \omega = (\omega_i)_{i \in [\![n]\!]} \in \{0, 1\}^n.$$

Im Fall $n = 1$, beschreibt eine *Bernoulli-Verteilung*, kurz B_{p}, also $\mathbb{p}_{\mathrm{B}_{\mathrm{p}}}(\omega) = \mathrm{p}^{\omega}(1 - \mathrm{p})^{1-\omega}$ für $\omega \in \{0, 1\}$, gerade den Wurf einer Münze, wobei wir die Ergebnisse 1=„Kopf“ als Erfolg und 0=„Zahl“ als Misserfolg auffassen. Für $\mathrm{B}_{0.3}$ (die Münze ist nicht fair, da keine Laplace-Verteilung vorliegt), erhalten wir die Zähldichte und die zugehörigen Verteilungsfunktion:

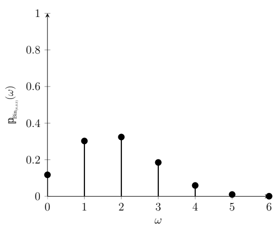


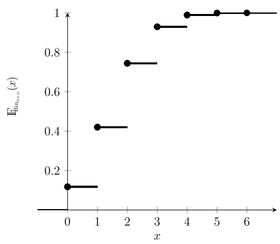


(d) *Binomial-Verteilung*, kurz $\mathrm{Bin}_{(n,\mathrm{p})}$, mit Erfolgswahrscheinlichkeit $\mathrm{p} \in [0, 1]$, Länge $n \in \mathbb{N}$ und $\Omega := [\![0, n]\!]$:

$$\mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p})}}(\omega) = \binom{n}{\omega} \mathrm{p}^{\omega}(1 - \mathrm{p})^{n-\omega} \text{ für } \omega \in [\![0, n]\!].$$

Die Anzahl „Zahl“ bei 6 Würfen einer Münze (wie in (c)) kann mit einer $\mathrm{Bin}_{(6,0.3)}$ beschrieben werden, für die wir folgende Zähldichte und zugehörige Verteilungsfunktion erhalten:



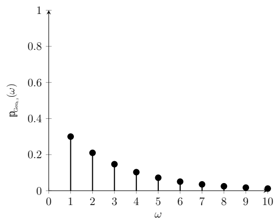


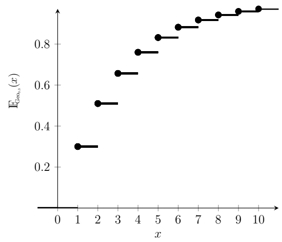
Ziehen wir wie in (b) Kugeln aus einer Urne aber mit Zurücklegen, so erhalten wir für die Anzahl der gezogenen weißen Kugeln eine $\mathrm{Bin}_{(n,K/N)}$-Verteilung. Die Binomialapproximation einer Hypergeometrischen Verteilung ist eine Übungsaufgabe.

(e) *geometrische Verteilung*, kurz $\mathrm{Geo}_{\mathrm{p}}$, mit Erfolgswahrscheinlichkeit $\mathrm{p} \in [0, 1]$ und $\Omega = \mathbb{N}$:

$$\mathbb{p}_{\mathrm{Geo}_{\mathrm{p}}}(\omega) = (1 - \mathrm{p})^{\omega - 1}\mathrm{p} \text{ für } \omega \in \mathbb{N}.$$

Werfen wir eine Münze wie in (c) solange bis das erste Mal „Kopf“ fällt, so kann die Anzahl der benötigten Würfe durch eine $\mathrm{Geo}_{0.3}$-Verteilung beschrieben werden, für die wir folgende Zähldichte und zugehörige Verteilungsfunktion erhalten:

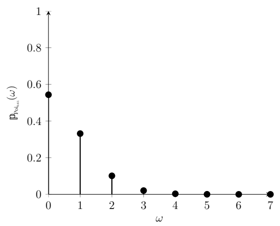


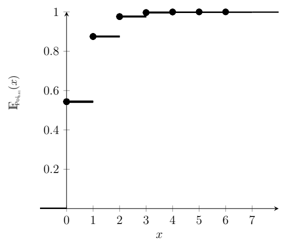


(f) *Poissonverteilung*, kurz Poi_λ, mit Parameter $\lambda \in \mathbb{R}_{>0}$ und $\Omega = \mathbb{Z}_{\geqslant 0}$:

$$\mathbb{p}_{\mathrm{Poi}_\lambda}(\omega) = \exp(-\lambda)\frac{\lambda^\omega}{\omega!} \text{ für } \omega \in \mathbb{Z}_{\geqslant 0}.$$

Ladislau von Bortkewitsch (1898) beschreibt die Anzahl an Todesfällen in der preußischen Kavallerie durch den Schlag eines Pferdes (vgl. Beispiel §04.12) durch eine $\mathrm{Poi}_{0.61}$-Verteilung, für die wir folgende Zähldichte und zugehörigen Verteilungsfunktion erhalten:





□

§04.09 **Poissonscher Grenzwertsatz**. *Es sei eine Folge von reellen Zahlen* $(\mathrm{p}_n)_{n\in\mathbb{N}}$ *aus dem Intervall* $[0,1]$ *mit* $\lambda := \lim_{n\to\infty} n\mathrm{p}_n \in \mathbb{R}_{>0}$ *gegeben. Dann gilt für alle* $\omega \in \mathbb{Z}_{\geqslant 0}$*:*

$$\lim_{n\to\infty} \mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p}_n)}}(\omega) = \mathbb{p}_{\mathrm{Poi}_\lambda}(\omega).$$

§04.10 **Beweis** von Satz §04.09. In der Vorlesung. □

§04.11 **Bemerkung**. Damit kann für hinreichend großes n und $n\mathrm{p}$ „mittelgroß“ die $\mathrm{Bin}_{(n,\mathrm{p})}$- Verteilung durch eine Poi_λ-Verteilung approximiert werden (zur Güte der Approximation vgl. Satz 5.9 in Krengel (2005)). Das heißt insbesondere, dass die Erfolgswahrscheinlichkeit p klein sein muss. Die Abhängigkeit von $\mathrm{p} = \mathrm{p}_n$ von n wird nur für die mathematische Beschreibung verwendet. Sie ist nicht so gemeint, dass p wirklich von n abhängt. □

§04.12 **Beispiel**. Ladislaus von Bortkewitsch (1898) gibt für 20 Jahre (1875-1894) und 10 Armeekorps der preußischen Kavallerie, folgende Anzahlen an Todesfällen durch den Schlag eines Pferdes pro $20 \times 10 = 200$ Korpsjahr an:

Todesfälle ω	0	1	2	3	4	$\geqslant 5$
Anzahl Korpsjahre mit ω Todesfällen	109	65	22	3	1	0
$200\mathbb{p}_{\mathrm{Poi}_{0.61}}(\omega)$	109	66	20	4	1	0

Der Parameter λ der Poissonverteilung Poi_λ wurde als mittlere Anzahl an Todesfällen pro Korpsjahr gewählt $\lambda = (1\cdot 65 + 2\cdot 22 + 3\cdot 3 + 1\cdot 4)/200 = 0.61$. □

§04.13 **Definition**. Für $i \in [\![n]\!]$ sei $\mathbb{P}_i \in \mathcal{W}(2^\Omega)$ ein diskretes Wahrscheinlichkeitsmaß mit Zähldichte $\mathbb{p}_i$ auf Ω, dann heißt die Zähldichte

$$\bigotimes_{i\in[\![n]\!]} \mathbb{p}_i : \Omega^n \ni \omega = (\omega_i)_{i\in[\![n]\!]} \mapsto \big(\bigotimes_{i\in[\![n]\!]} \mathbb{p}_i\big)(\omega) := \prod_{i\in[\![n]\!]} \mathbb{p}_i(\omega_i) \in [0,1]$$

die *Produktzähldichte* der $(\mathbb{p}_i)_{i\in[\![n]\!]}$ auf Ω^n. Das diskrete Wahrscheinlichkeitsmaß

$$\bigotimes_{i\in[\![n]\!]} \mathbb{P}_i : 2^{\Omega^n} \ni A \mapsto \big(\bigotimes_{i\in[\![n]\!]} \mathbb{P}_i\big)(A) = \sum_{\omega\in A} \big(\bigotimes_{i\in[\![n]\!]} \mathbb{p}_i\big)(\omega) \in [0,1]$$

mit Zähldichte $\otimes_{i\in[\![n]\!]} \mathbb{p}_i$ heißt *Produktmaß* der $(\mathbb{P}_i)_{i\in[\![n]\!]}$ auf $(\Omega^n, 2^{\Omega^n})$. □

§04.14 **Bemerkung**. Für $i \in [\![n]\!]$ sei $\mathbb{P}_i \in \mathcal{W}(2^\Omega)$ ein diskretes Wahrscheinlichkeitsmaß mit Zähldichte $\mathbb{p}_i$ auf Ω, dann ist die *Produktzähldichte* $\otimes_{i\in[\![n]\!]} \mathbb{p}_i$ in der Tat eine Zähldichte auf Ω^n, da

$$\sum_{\omega\in\Omega^n} \big(\bigotimes_{i\in[\![n]\!]} \mathbb{p}_i\big)(\omega) = \sum_{\omega_1\in\Omega} \cdots \sum_{\omega_n\in\Omega} \prod_{i\in[\![n]\!]} \mathbb{p}_i(\omega_i) = \prod_{i\in[\![n]\!]} \sum_{\omega\in\Omega} \mathbb{p}_i(\omega) = \prod_{i\in[\![n]\!]} 1 = 1.$$

Nach Lemma §04.02 (ii) ist das *Produktmaß* $\bigotimes_{i\in[\![n]\!]} \mathbb{P}_i$ mit Produktzähldichte $\otimes_{i\in[\![n]\!]} \mathbb{p}_i$ ein diskretes Wahrscheinlichkeitsmaß $\bigotimes_{i\in[\![n]\!]} \mathbb{P}_i \in \mathcal{W}(2^{\Omega^n})$, wobei

$$\big(\bigotimes_{i\in[\![n]\!]} \mathbb{P}_i\big)\big(\bigtimes_{i\in[\![n]\!]} A_i\big) = \sum_{\omega_1\in A_1} \cdots \sum_{\omega_n\in A_n} \big(\bigotimes_{i\in[\![n]\!]} \mathbb{p}_i\big)(\omega_{i}: i\in[\![n]\!]) = \prod_{i\in[\![n]\!]} \sum_{\omega\in A_i} \mathbb{p}_i(\omega) = \prod_{i\in[\![n]\!]} \mathbb{P}_i(A_i)$$

für alle $\bigtimes_{i\in[\![n]\!]} A_i \in 2^{\Omega^n}$ gilt. □

§04.15 **Beispiel**. Die Zähldichte eines Bernoulli-Schema wie Beispiel §04.08 (d) ist die Produktzähldichte von n Zähldichten zu identischen Bernoulli-Verteilungen $\mathrm{B_p}$, d.h. $\mathrm{B}_\mathrm{p}^n = \bigotimes_{i\in[\![n]\!]} \mathrm{B_p}$. Im Fall $\mathrm{p} = 0.5$ gilt für die Zähldichte

$$\mathbb{p}_{\mathrm{B}^n_{0.5}}(\omega) = (\tfrac{1}{2})^{\sum_{i\in[\![n]\!]}\omega_i}(1-\tfrac{1}{2})^{n-\sum_{i\in[\![n]\!]}\omega_i} = (\tfrac{1}{2})^n \quad \forall\, \omega = (\omega_i)_{i\in[\![n]\!]}\{0,1\}^n.$$

Betrachte zum Zeitpunkt $j \in [\![n]\!]$ einen Flip

$$T_j : \{0,1\}^n \ni \omega = (\omega_i)_{i\in[\![n]\!]} \mapsto T_j(\omega) := (\omega_1,\dots,\omega_{j-1}, 1-\omega_j, \omega_{j+1},\dots,\omega_n) \in \{0,1\}^n,$$

der das Ergebnis des j-ten Münzwurfes umdreht. Für das Bild $T_j(A) := \{T_j(\omega) : \omega \in A\}$ von $A \subseteq \{0,1\}^n$ unter T_j gilt dann die Invarianzeigenschaft

$$\mathrm{B}^n_{0.5}(T_j(A)) = \sum_{\omega\in A} \mathbb{p}_{\mathrm{B}^n_{0.5}}(T_j(\omega)) = \sum_{\omega\in A} (\tfrac{1}{2})^n = \sum_{\omega\in A} \mathbb{p}_{\mathrm{B}^n_{0.5}}(\omega) = \mathrm{B}^n_{0.5}(A).$$

Für jedes $n \in \mathbb{N}$ genügt $\mathrm{B}^n_{0.5} \in \mathcal{W}(2^{\{0,1\}^n})$ somit der Invarianzeigenschaft $\mathrm{B}^n_{0.5}(T_j(A)) = \mathrm{B}^n_{0.5}(A)$ für alle $A \in 2^{\{0,1\}^n}$ und für alle $j \in [\![n]\!]$. (Dies drückt die Fairness der Münze und die Unabhängigkeit der Würfe aus.) □

§04.16 **Satz** *(Vitali (1905))*. *Sei $\Omega := \{0,1\}^{\mathbb{N}}$ der Ergebnisraum eines unendlich oft wiederholten Münzwurfes. Für $n \in \mathbb{N}$ sei der Flip*

$$T_n : \Omega \ni \omega = (\omega_n)_{n\in\mathbb{N}} \mapsto T_n(\omega) := (\omega_1,\dots,\omega_{n-1}, 1-\omega_n, \omega_{n+1},\dots) \in \Omega$$

die Abbbildung, welche das Ergebnis des n-ten Münzwurfes umdreht. Für $A \in 2^\Omega$ bezeichne $T_n(A) := \{T_n(\omega) : \omega \in A\}$ das Bild von A unter T_n. Dann gibt es kein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(2^\Omega)$ auf der Potenzmenge 2^Ω, das der Invarianzeigenschaft $\mathbb{P}(T_n(A)) = \mathbb{P}(A)$ für alle $A \in 2^\Omega$ und $n \in \mathbb{N}$ genügt.

§04.17 **Beweis** von Satz §04.16. In der Vorlesung. □

§05 Stetiger Wahrscheinlichkeitsraum

§05.01 **Definition**. Ist $\boxed{\mathbb{f}}: \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ eine (Lebesgue-)integrierbare Funktion mit $\int_{\mathbb{R}^n} \mathbb{f}(x)dx = 1$, so heißt $\mathbb{f}$ *Wahrscheinlichkeitsdichte* (kurz *Dichte*) auf $\mathbb{R}^n$. □

§05.02 **Satz**. *Jede Wahrscheinlichkeitsdichte $\mathbb{f}$ auf $\mathbb{R}^n$ erzeugt mittels*

$$\mathbb{P}((a,b]) := \int_{(a,b]} \mathbb{f}(x)dx = \int_{(a_1,b_1]} \cdots \int_{(a_n,b_n]} \mathbb{f}(x_1,\dots,x_n)dx_n \cdots dx_1 \text{ für } a,b \in \mathbb{R}^n,\ a < b,$$

ein eindeutig bestimmtes Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{B}^n)$ auf $(\mathbb{R}^n, \mathscr{B}^n)$ und es gilt

$$\mathbb{P}(B) = \int_B \mathbb{f}(x)dx \quad \text{für alle } B \in \mathscr{B}^n.$$

§05.03 **Beweis** von Satz §05.02. In der Vorlesung Wahrscheinlichkeitstheorie 1. □

§05.04 **Definition**. Ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{B}^n)$ auf $(\mathbb{R}^n, \mathscr{B}^n)$ heißt *stetig*, wenn eine Wahrscheinlichkeitsdichte $\mathbb{f}$ auf $\mathbb{R}^n$ mit $\mathbb{P}(B) = \int_B \mathbb{f}(x)dx$ für alle $B \in \mathscr{B}^n$ existiert. $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P})$ wird dann *stetiger Wahrscheinlichkeitsraum* genannt. □

§05.05 **Lemma**.

(i) *Ist $\mathbb{f}$ eine Dichte eines Wahrscheinlichkeitsmaßes $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ mit Verteilungsfunktion $\mathbb{F}$, so gilt $\mathbb{F}(x) = \mathbb{P}(\mathbb{R}_{\leqslant x}) = \int_{\mathbb{R}_{\leqslant x}} \mathbb{f}(y)dy$ für alle $x \in \mathbb{R}$.*

(ii) *Ist die Verteilungsfunktion $\mathbb{F}$ eines Wahrscheinlichkeitsmaßes $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ (schwach) differenzierbar, so ist $\mathbb{f} := \mathbb{F}'$ die zugehörige Wahrscheinlichkeitsdichte.*

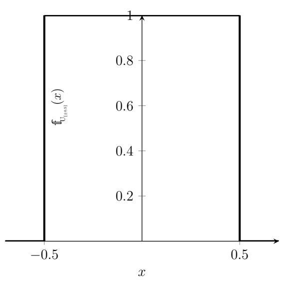
§05.06 **Beweis** von Lemma §05.05. In der Vorlesung Analysis 3. □

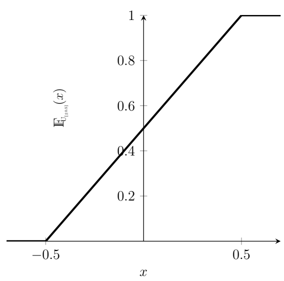
§05.07 **Beispiel**. Folgende Wahrscheinlichkeitsdichten beschreiben häufig auftretende stetige Verteilungen auf $(\mathbb{R}, \mathscr{B})$:

(a) *Gleich-/Uniformverteilung*, kurz $\boxed{\mathrm{U}_G}$, auf $G \in \mathscr{B}$ mit Lebesgue-Maß $\lambda(G) \in \mathbb{R}_{>0}$:

$$\mathbb{f}_{\mathrm{U}_G}(x) = \tfrac{1}{\lambda(G)} \mathbb{1}_G(x) \text{ für } x \in \mathbb{R}.$$

Runden wir Messungen reeller Größen auf die jeweils nächstgelegene ganze Zahl hin auf bzw. ab, so kann der Rundungsfehler durch eine Gleichverteilung $\mathrm{U}_{[\pm 0.5]}$ auf dem Intervall $[\pm 0.5] := [-0.5, 0.5]$ mit Dichte $\mathbb{f}_{\mathrm{U}_{[\pm 0.5]}}(x) = \mathbb{1}_{[\pm 0.5]}(x)$ und Verteilungsfunktion $\mathbb{F}_{\mathrm{U}_{[\pm 0.5]}}(x) = (x + 0.5)\mathbb{1}_{[\pm 0.5]}(x) + \mathbb{1}_{\mathbb{R}_{>0.5}}(x)$ für $x \in \mathbb{R}$ beschrieben werden:

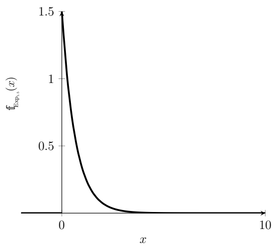


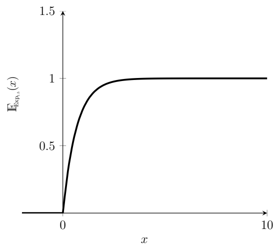


(b) *Exponentialverteilung*, kurz $\boxed{\mathrm{Exp}_\lambda}$, mit Parametern $\lambda \in \mathbb{R}_{>0}$:

$$\mathbb{f}_{\mathrm{Exp}_\lambda}(x) = \lambda \exp(-\lambda x) \mathbb{1}_{\mathbb{R}_{\geqslant 0}}(x) \text{ für } x \in \mathbb{R}.$$

Häufig wird in Telefonzentralen die Dauer eines Telefongesprächs durch eine Exponentialverteilung beschrieben. Nimmt eine Servicehotline zum Beispiel für die Gesprächsdauer eine $\mathrm{Exp}_{1.5}$-Verteilung an, so besitzt diese die Dichte $\mathbb{f}_{\mathrm{Exp}_{1.5}}(x) = 1.5\exp(-1.5x)\mathbb{1}_{\mathbb{R}_{\geqslant 0}}(x)$ und die zugehörige Verteilungsfunktion $\mathbb{F}_{\mathrm{Exp}_{1.5}}(x) = (1-\exp(-1.5x))\mathbb{1}_{\mathbb{R}_{\geqslant 0}}(x)$ für $x \in \mathbb{R}$:





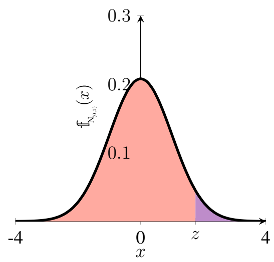
(c) *Normalverteilung*, kurz $\mathrm{N}_{(\mu,\sigma^2)}$, mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 \in \mathbb{R}_{>0}$:

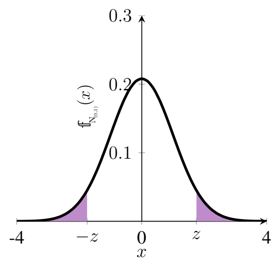
$$\mathbb{f}_{\mathrm{N}_{(\mu,\sigma^2)}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\tfrac{1}{2\sigma^2}(x-\mu)^2) \text{ für } x \in \mathbb{R}.$$

Die Verteilung gemittelter zufälliger Größen (z. Bsp. der jährliche Wasserverbrauch eines Haushaltes) kann häufig durch eine Normalverteilung gut approximiert werden (vgl. Abschnitt §32). Eine $\mathrm{N}_{(0,1)}$-Verteilung heißt *Standardnormalverteilung*. Wir bezeichnen mit Φ die Verteilungsfunktion einer Standardnormalverteilung, d.h., für alle $z \in \mathbb{R}$ gilt

$$\Phi(z) = \mathrm{N}_{(0,1)}(\mathbb{R}_{\leqslant z}) = \int_{\mathbb{R}_{\leqslant z}} \frac{1}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}x^2)dx.$$

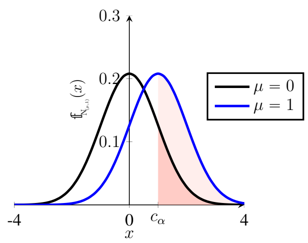
Für alle $z \in \mathbb{R}_{\geqslant 0}$ gilt dann $\Phi(-z) = 1 - \Phi(z)$:

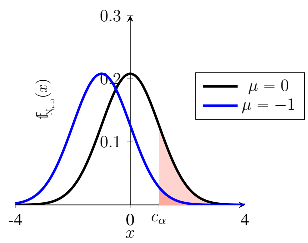




Weiterhin heißt für $\alpha \in (0,1)$ der eindeutig bestimmte Wert z_α mit $\alpha = \Phi(\mathrm{z}_\alpha)$ das α-Quantil einer Standardnormalverteilung. Typische Werte für Quantile der Standardnormalverteilung sind tabellarisiert, siehe Seite 117 im Anhang. Wir halten weiterhin fest, dass für $\alpha \in (0,1)$ und $\mathrm{c}_\alpha \in \mathbb{R}$ mit $\mathrm{N}_{(\mu,\sigma^2)}(\mathbb{R}_{\geqslant \mathrm{c}_\alpha}) = \alpha$ gilt

$$\mathrm{N}_{(\mu+\mathrm{c},\sigma^2)}(\mathbb{R}_{\geqslant \mathrm{c}_\alpha}) > \alpha \quad \text{und} \quad \mathrm{N}_{(\mu-\mathrm{c},\sigma^2)}(\mathbb{R}_{\geqslant \mathrm{c}_\alpha}) < \alpha \quad \text{für alle } \mathrm{c} \in \mathbb{R}_{>0}:$$





□

§05.08 **Definition**. Für $i \in [\![n]\!]$ sei $\mathbb{P}_i \in \mathcal{W}(\mathscr{B})$ ein stetiges Wahrscheinlichkeitsmaß mit Wahrscheinlichkeitsdichte $\mathbb{f}_i$ auf $\mathbb{R}$. Dann heißt die Wahrscheinlichkeitsdichte

$$\bigotimes_{i\in[\![n]\!]} \mathbb{f}_i : \mathbb{R}^n \ni x = (x_i)_{i\in[\![n]\!]} \mapsto (\bigotimes_{i\in[\![n]\!]} \mathbb{f}_i)(x) := \prod_{i\in[\![n]\!]} \mathbb{f}_i(x_i) \in \mathbb{R}_{\geqslant 0}$$

die *Produktdichte* der $(\mathbb{f}_i)_{i\in[\![n]\!]}$ auf $\mathbb{R}^n$. Das stetige Wahrscheinlichkeitsmaß

$$\bigotimes_{i\in[\![n]\!]}\mathbb{P}_i : \mathscr{B}^n \ni B \mapsto (\bigotimes_{i\in[\![n]\!]}\mathbb{P}_i)(B) = \int_B (\bigotimes_{i\in[\![n]\!]}\mathbb{f}_i)(x)dx \in [0,1]$$

mit Wahrscheinlichkeitsdichte $\otimes_{i\in[\![n]\!]}\mathbb{f}_i$ heißt *Produktmaß* der $(\mathbb{P}_i)_{i\in[\![n]\!]}$ auf $(\mathbb{R}^n, \mathscr{B}^n)$. □

§05.09 **Bemerkung**. Für $i \in [\![n]\!]$ sei $\mathbb{P}_i \in \mathcal{W}(\mathscr{B})$ ein stetiges Wahrscheinlichkeitsmaß mit Wahrscheinlichkeitsdichte $\mathbb{f}_i$ auf $\mathbb{R}$, dann ist die *Produktdichte* $\otimes_{i\in[\![n]\!]}\mathbb{f}_i : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ eine Wahrscheinlichkeitsdichte auf $\mathbb{R}^n$, da

$$\int_{\mathbb{R}^n} (\bigotimes_{i\in[\![n]\!]}\mathbb{f}_i)(x)dx = \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} \prod_{i\in[\![n]\!]}\mathbb{f}_i(x_i)dx_n\cdots dx_1 = \prod_{i\in[\![n]\!]}\int_{\mathbb{R}}\mathbb{f}_i(y)dy = \prod_{i\in[\![n]\!]} 1 = 1.$$

Nach Satz §05.02 ist das *Produktmaß* $\bigotimes_{i\in[\![n]\!]}\mathbb{P}_i$ mit Produktdichte $\otimes_{i\in[\![n]\!]}\mathbb{f}_i$ ein stetiges Wahrscheinlichkeitsmaß $\bigotimes_{i\in[\![n]\!]}\mathbb{P}_i \in \mathcal{W}(\mathscr{B}^n)$, wobei

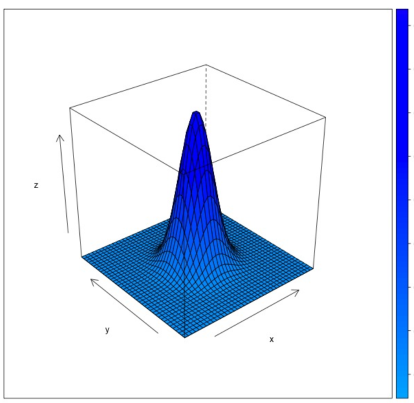
$$(\bigotimes_{i\in[\![n]\!]}\mathbb{P}_i)(\bigtimes_{i\in[\![n]\!]}A_i) = \int_{A_1}\cdots\int_{A_n}(\bigotimes_{i\in[\![n]\!]}\mathbb{f}_i)(x_i\colon i\in[\![n]\!])dx_n\cdots dx_1 = \prod_{i\in[\![n]\!]}\int_{A_i}\mathbb{f}_i(y)dy = \prod_{i\in[\![n]\!]}\mathbb{P}_i(A_i)$$

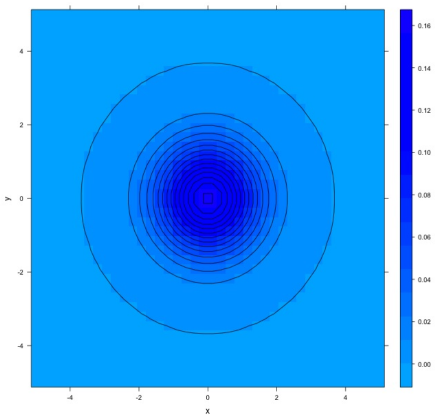
für alle $\bigtimes_{i\in[\![n]\!]}A_i \in \mathscr{B}^n$ gilt. □

§05.10 **Beispiel**. Die *n-dimensionale Standard-Normalverteilung* $\mathrm{N}_{(\mathbb{0}_n, \mathrm{I}_{n\times n})} \in \mathcal{W}(\mathscr{B}^n)$ auf $(\mathbb{R}^n, \mathscr{B}^n)$, wobei $\mathrm{I}_{n\times n} \in \mathbb{R}^{n\times n}$ die n-dimensionale Einheitsmatrix und $\mathbb{0}_n \in \mathbb{R}^n$ den n-dimensionalen Nullvektor bezeichnet, ist definiert durch die Dichte

$$\mathbb{f}_{\mathrm{N}_{(\mathbb{0}_n,\mathrm{I}_{n\times n})}}(x) := \prod_{i\in[\![n]\!]}\mathbb{f}_{\mathrm{N}_{(0,1)}}(x_i) = (2\pi)^{-n/2}\exp(-\tfrac{1}{2}\sum_{i\in[\![n]\!]}x_i^2) \quad \text{für } x = (x_i)_{i\in[\![n]\!]} \in \mathbb{R}^n;$$

Für $n = 2$ erhalten wir die Dichte $\mathbb{f}_{\mathrm{N}_{(\mathbb{0}_2,\mathrm{I}_{2\times 2})}}(x_1,x_2) = (2\pi)^{-1}\exp(-\frac{1}{2}(x_1^2+x_2^2))$ für $x_1, x_2 \in \mathbb{R}$:





□

§06 Statistisches Modell

§06.01 **Erinnerung**. $\mathcal{W}(\mathscr{X})$ bezeichnet die Menge aller Wahrscheinlichkeitsmaße auf einem messbaren Raum $(\mathcal{X}, \mathscr{X})$. Für eine nicht-leere Indexmenge Θ wird eine Familie $(\mathbb{P}_\theta)_{\theta\in\Theta}$ von Wahrscheinlichkeitsmaßen auf $\mathscr{X}$ formal durch die Abbildung $\Theta \to \mathcal{W}(\mathscr{X})$ mit $\theta \mapsto \mathbb{P}_\theta$ definiert. □

§06.02 **Definition**. Sei $\mathbb{P}_\Theta := (\mathbb{P}_\theta)_{\theta\in\Theta}$ eine Familie von Wahrscheinlichkeitsmaßen auf einem messbaren Raum (*Stichprobenraum*) $(\mathcal{X}, \mathscr{X})$. Die nicht-leere Indexmenge Θ wird *Parametermenge* genannt. Wir bezeichnen das Tripel $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ als *statistisches Experiment* oder *statistisches Modell*. Ein statistisches Experiment $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ heißt *adäquat* für ein zufälliges Experiment, wenn dieses durch den Wahrscheinlichkeitsraum $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\theta)$ für ein $\theta \in \Theta$ beschrieben wird. In diesem Fall wird der Parameter $\theta \in \Theta$ auch *wahrer Parameter* genannt. □

§06.03 **Sprechweise**. Wir nennen ein statistisches Experiment $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$

diskret, wenn der Stichprobenraum $\mathcal{X}$ abzählbar und $\mathscr{X} = 2^{\mathcal{X}}$ ist. In diesem Fall bezeichnet $\mathbb{p}_\Theta := (\mathbb{p}_\theta)_{\theta\in\Theta}$ die zur Familie $\mathbb{P}_\Theta$ von diskreten Wahrscheinlichkeitsmaßen gehörige Familie von Zähldichten;

stetig, wenn $(\mathbb{R}^n, \mathscr{B}^n)$ der Stichprobenraum ist und für jeden Parameter $\theta \in \Theta$ ist $\mathbb{P}_\theta$ ein stetiges Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathscr{B}^n)$. In diesem Fall bezeichnet $\mathbb{f}_\Theta := (\mathbb{f}_\theta)_{\theta\in\Theta}$ die zur Familie $\mathbb{P}_\Theta$ von stetigen Wahrscheinlichkeitsmaßen gehörige Familie von Wahrscheinlichkeitsdichten. □

§06.04 **Beispiel**. Im Folgenden geben wir statistische Modelle für jeweils ein Ergebnis der im Kapitel 1 Prolog vorgestellten Beispiele an.

Beispiel I: Setzen wir Eins für das weibliche Geschlecht und Null für kein weibliches Geschlecht eines Konsumierenden, so beschreiben wir das zufällige Geschlecht eines Konsumierenden durch ein *Bernoulliverteilungsmodell* $\big(\{0,1\}, 2^{\{0,1\}}, (\mathrm{B}_p)_{p\in[0,1]}\big)$, vgl. Beispiel §04.08 (c).

Beispiel II: Die zufällige Anzahl beschädigter Schrauben in einem Beutel mit 50 Schrauben beschreiben wir durch ein *Binomialverteilungsmodell* $\big([\![0,50]\!], 2^{[\![0,50]\!]}, (\mathrm{Bin}_{(50,p)})_{p\in[0,1]}\big)$, vgl. Beispiel §04.08 (d).

Beispiel III Die zufällige Anzahl der eingegangen Anrufe innerhalb einer Viertelstunde beschreiben wir durch ein *Poissonverteilungsmodell* $\big(\mathbb{Z}_{\geqslant 0}, 2^{\mathbb{Z}_{\geqslant 0}}, (\mathrm{Poi}_\lambda)_{\lambda\in\mathbb{R}_{>0}}\big)$, vgl. Beispiel §04.08 (f).

Beispiel IV Die zufällige Wartezeit an einer Haltestelle an einem Tag (in Minuten) beschreiben wir durch ein *Uniformverteilungsmodell* $\big(\mathbb{R}, \mathscr{B}, (\mathrm{U}_{[0,\theta]})_{\theta\in\mathbb{R}_{>0}}\big)$, vgl. Beispiel §05.07 (a).

Beispiel V Die zufällige Lebensdauer einer Glühlampe (in Stunden) beschreiben wir durch ein *Exponentialverteilungsmodell* $\big(\mathbb{R}, \mathscr{B}, (\mathrm{Exp}_\lambda)_{\lambda\in\mathbb{R}_{>0}}\big)$, vgl. Beispiel §05.07 (b).

Beispiel VI Die zufälligen Fehler in gemessenen Werten beschreiben wir durch ein *Normalverteilungsmodell* $\big(\mathbb{R}, \mathscr{B}, (\mathrm{N}_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}\big)$, vgl. Beispiel §05.07 (c). □.

§06|01 Testen einfacher Hypothesen

§06.05 **Beispiel** *(Binomialverteilungsmodell)*. Im Beispiel I im Kapitel 1 Prolog zählte der Verbraucherservice 699 Konsumentinnen unter den 1000 befragten Personen. Dieses zufällige Experiment lässt sich durch ein *Binomialverteilungsmodell* $\big([\![0,1000]\!], 2^{[\![0,1000]\!]}, (\mathrm{Bin}_{(1000,p)})_{p\in[0,1]}\big)$ beschreiben (formale Begründung in Beispiel §20.03, (a)). Zur (unrealistischen) Vereinfachung geht der Verbraucherservice zunächst davon aus, dass $\{0.5, 0.7\}$ die Parametermenge ist, das heißt, das statistische Modell $\big([\![0,1000]\!], 2^{[\![0,1000]\!]}, (\mathrm{Bin}_{(1000,0.5)}, \mathrm{Bin}_{(1000,0.7)})\big)$ beschreibt adäquat das zufällige Experiment. In anderen Worten, entweder ist $\mathbb{P}_0 = \mathrm{Bin}_{(1000,0.5)}$ oder $\mathbb{P}_1 = \mathrm{Bin}_{(1000,0.7)}$ die wahre Verteilung. □

§06.06 **Definition**. Sei $\big(\mathcal{X}, \mathscr{X}, (\mathbb{P}_0, \mathbb{P}_1)\big)$ ein (binäres) statistisches Experiment. Die Entscheidung,

ob die *Nullhypothese* $H_0 : \{\mathbb{P}_0\}$ oder die *Alternative* $H_1 : \{\mathbb{P}_1\}$ vorliegt, wird *(statistisches) Testproblem mit einfachen Hypothesen* genannt. Eine Entscheidungsfunktion

$$\varphi : \mathcal{X} \to \{0, 1\},$$

also Entscheidungen nur anhand einer Stichprobe x aus dem Stichprobenraum $\mathcal{X}$, mit

$\varphi^{-1}(\{1\}) \in \mathscr{X}$ dem Ereignis aller Stichproben, die zu einer *Entscheidung gegen die Nullhypothese H_0*, also für die Alternative H_1 führen, sowie

$\varphi^{-1}(\{0\}) \in \mathscr{X}$ dem Ereignis aller Stichproben, die zu einer *Entscheidung für die Nullhypothese H_0* führen,

heißt *(statistischer) Hypothesentest*, kurz *Test*. □

§06.07 **Sprechweise**. Die Sprechweise Testproblem mit *einfachen* Hypothesen bedeutet in diesem Zusammenhang, dass sowohl die Nullhypothese als auch die Alternative jeweils einelementig ist. Üblicherweise bezeichnen wir eine Entscheidung für die Alternative H_1 als *Ablehnen der Nullhypothese*, wogegen eine Entscheidung für die Nullhypothese H_0 *nicht Ablehnen der Nullhypothese* genannt wird. Für einen Test φ bezeichnen wir

$\mathcal{A} := \varphi^{-1}(\{1\}) \in \mathscr{X}$ das Ereignis aller Stichproben, die zum Ablehnen der Nullhypothese H_0 führen, als *Ablehnbereich* des Tests φ.

Da wir hier nur die zwei Möglichkeiten Ablehnen oder nicht Ablehnen für einen Test zulassen, ist $\varphi = \mathbb{1}_{\mathcal{A}}$ eine Indikatorfunktion und

$\mathcal{A}^c = \varphi^{-1}(\{0\}) \in \mathscr{X}$ das Ereignis aller Stichproben, die nicht zu einem Ablehnen der Nullhypothese führen, genannt *Annahmebereich*,

wobei definitionsgemäß $\mathcal{A} \biguplus \mathcal{A}^c = \mathcal{X}$ gilt. □

§06.08 **Beispiel** *(Binomialverteilungsmodell §06.05 fortgesetzt)*. Wir sind an einer Entscheidung unter Vorliegen der Stichprobe $x = 677$ aus dem Stichprobenraum $\mathcal{X} = [\![0, 1000]\!]$ interessiert. Wir legen *vor* der Erhebung der Stichprobe einen Test $\varphi : [\![0, 1000]\!] \to \{0, 1\}$ fest. Ergibt ein Auswerten des Testes $\varphi(677) = 1$, so lehnen wir die Nullhypothese $H_0 : \{\mathrm{Bin}_{(1000,0.5)}\}$ gegen die Alternative $H_1 : \{\mathrm{Bin}_{(1000,0.7)}\}$ ab. Nimmt der Test dagegen den Wert $\varphi(677) = 0$ an, so lehnen wir die Nullhypothese H_0 nicht ab. □

§06.09 **Bemerkung**. Durch Angabe des Ablehnbereiches $\mathcal{A}$ ist ein Test $\varphi = \mathbb{1}_{\mathcal{A}}$ eindeutig festgelegt. Offensichtlich können in einem statistisches Testproblem mit einfachen Hypothesen nur zwei Fehlentscheidungen auftreten, die Nullhypothese $H_0 : \{\mathbb{P}_0\}$ wird abgelehnt, also der Ablehnbereich $\mathcal{A}$ tritt ein, obwohl $\mathbb{P}_0$ vorliegt, oder die Nullhypothese wird nicht gegen die Alternative $H_1 : \{\mathbb{P}_1\}$ abgelehnt, also der Annahmebereich $\mathcal{A}^c$ tritt ein, obwohl $\mathbb{P}_1$ vorliegt. Man beachte, dass wir für einen Test gefordert haben, dass der Ablehnbereich $\mathcal{A}$ und somit auch der Annahmebereich $\mathcal{A}^c$ messbar ist, also $\mathcal{A} \in \mathscr{X}$ gilt. Damit können wir den Ereignissen $\mathcal{A}$ und $\mathcal{A}^c$ Wahrscheinlichkeiten zuordnen. □

§06.10 **Definition**. Sei $\big(\mathcal{X}, \mathscr{X}, (\mathbb{P}_0, \mathbb{P}_1)\big)$ ein (binäres) statistisches Experiment und $\mathcal{A} \in \mathscr{X}$ der Ablehnbereich eines Tests $\varphi = \mathbb{1}_{\mathcal{A}}$ der einfachen Nullhypothese $H_0 : \{\mathbb{P}_0\}$ gegen die einfache Alternative $H_1 : \{\mathbb{P}_1\}$. Dann bezeichnet

Fehler 1. Art: die Wahrscheinlichkeit $\mathbb{P}_0(\mathcal{A})$ die Nullhypothese abzulehnen, sich also für $\mathbb{P}_1$ zu entscheiden, obwohl $\mathbb{P}_0$ vorliegt;

Fehler 2. Art: die Wahrscheinlichkeit $\mathbb{P}_1(\mathcal{A}^c)$ die Nullhypothese nicht abzulehnen, sich also für $\mathbb{P}_0$ zu entscheiden, obwohl $\mathbb{P}_1$ vorliegt. □

§06|01.11 **Anmerkung**. Möchten wir zwei Tests miteinander vergleichen, so erscheint es sinnvoll, die Fehler 1. Art und 2. Art zu vergleichen. Offensichtlich würden wir gern einen Test finden der sowohl den Fehler 1. Art als auch den Fehler 2. Art minimiert. Betrachten wir den konstanten Test $\varphi = \mathbb{1}_{\mathcal{X}}$ mit Ablehnbereich $\mathcal{A} = \varphi^{-1}(\{1\}) = \mathcal{X}$, das heißt, wir lehnen immer die Nullhypothese ab. Dann ist $\mathbb{P}_0(\mathcal{A}) = 1$ aber auch $\mathbb{P}_1(\mathcal{A}^c) = \mathbb{P}_1(\emptyset) = 0$. Im Allgemeinen können wir nicht beide Fehler gleichzeitig minimieren. Es gibt verschiedene Möglichkeiten, dieses Problem zu umgehen. Zum Beispiel könnte eine gewichtete Summe der beiden Fehler minimiert werden. Da die beiden Fehler häufig unterschiedliche Konsequenzen haben, betrachten wir im Folgenden nur Testfunktionen, die eine Obergrenze für den Fehler 1. Art einhalten. Innerhalb dieser Klasse von Tests suchen wir dann denjenigen, der den Fehler 2. Art minimiert. □

§06.12 **Definition**. Sei $(\mathcal{X}, \mathscr{X}, (\mathbb{P}_0, \mathbb{P}_1))$ ein (binäres) statistisches Experiment. Ein Test $\varphi = \mathbb{1}_{\mathcal{A}}$ mit Ablehnbereich $\mathcal{A} \in \mathscr{X}$ der einfachen Nullhypothese $H_0 : \{\mathbb{P}_0\}$ gegen die einfache Alternative $H_1 : \{\mathbb{P}_1\}$ hält das (*Signifikanz-*) *Niveau* $\alpha \in [0, 1]$ ein (oder kurz ist ein *α-Test*), wenn der Fehler 1. Art $\mathbb{P}_0(\mathcal{A}) \leqslant \alpha$ erfüllt. Ein Test $\varphi = \mathbb{1}_{\mathcal{A}}$ mit Ablehnbereich $\mathcal{A} \in \mathscr{X}$ heißt *bester Test zum Niveau $\alpha \in [0, 1]$*, falls er das Niveau $\alpha \in [0, 1]$ einhält und der Fehler 2. Art $\mathbb{P}_1(\tilde{\mathcal{A}}^c)$ eines jeden anderen α-Tests $\varphi = \mathbb{1}_{\tilde{\mathcal{A}}}$ mit Ablehnbereich $\tilde{\mathcal{A}} \in \mathscr{X}$ nicht kleiner ist, dass heißt, $\mathbb{P}_1(\mathcal{A}^c) \leqslant \mathbb{P}_1(\tilde{\mathcal{A}}^c)$ gilt. □

§06.13 **Bemerkung**. In der Definition eines besten α-Tests werden die Fehler 1. und 2. Art nicht symmetrisch einbezogen. Dies ist eine gewollte Eigenschaft, da so sicher gestellt wird, dass ein Fehler 1. Art klein ist. Andererseits sollte somit die Festlegung der Nullhypothese und Alternative dies widerspiegeln. Vereinfacht gesprochen, das Ziel ist es, die Nullhypothese abzulehnen, da nur dann die Wahrscheinlichkeit sich zu irren, immer klein ist. □

§06.14 **Definition**. Sei $(\mathcal{X}, \mathscr{X}, (\mathbb{P}_0, \mathbb{P}_1))$ ein (binäres) diskretes statistisches Experiment mit entsprechenden Zähldichten $\mathbb{p}_0$ und $\mathbb{p}_1$. Jeder Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$ mit Ablehnbereich der Form

$$\mathcal{A}_k := \{\mathbb{p}_1 > k\, \mathbb{p}_0\} = \{x \in \mathcal{X} : \mathbb{p}_1(x) > k\, \mathbb{p}_0(x)\}$$

für einen *kritischen Wert* $k \in \mathbb{R}_{\geqslant 0}$ heißt *Neyman-Pearson-Test*. □

§06.15 **Beispiel** *(Binomialverteilungsmodell §06.08 fortgesetzt)*. Wir betrachten das Testproblem der einfachen Nullhypothese $H_0 : \{\mathrm{Bin}_{(1000,0.5)}\}$ gegen die einfache Alternative $H_1 : \{\mathrm{Bin}_{(1000,0.7)}\}$ im binären diskreten statistischen Experiment $([\![0, 1000]\!], 2^{[\![0,1000]\!]}, (\mathrm{Bin}_{(1000,p)})_{p \in \{0.5, 0.7\}})$. Setzen wir

$$p_0 := 0.5 \text{ und } p_1 := 0.7,$$

so ist für einen kritischem Wert $k \in \mathbb{R}_{\geqslant 0}$ ein Neyman-Pearson-Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$ gegeben durch den Ablehnbereich

$$\begin{aligned}\mathcal{A}_k &= \{x \in [\![0, 1000]\!] : \tbinom{1000}{x} p_1^x (1 - p_1)^{1000-x} > k \tbinom{1000}{x} p_0^x (1 - p_0)^{1000-x}\} \\ &= \{x \in [\![0, 1000]\!] : \left(\tfrac{p_1/(1-p_1)}{p_0/(1-p_0)}\right)^x \left(\tfrac{1-p_1}{1-p_0}\right)^{1000} > k\}\end{aligned}$$

Da $p_1 > p_0$ gilt, folgt $\frac{p_1/(1-p_1)}{p_0/(1-p_0)} > 1$ und die Funktion

$$L_{p_1,p_0} : [\![0, 1000]\!] \ni x \mapsto L_{p_1,p_0}(x) := \left(\tfrac{p_1/(1-p_1)}{p_0/(1-p_0)}\right)^x \left(\tfrac{1-p_1}{1-p_0}\right)^{1000} \in \mathbb{R}_{\geqslant 0}$$

ist somit streng monoton wachsend. Damit gibt es zu jedem kritischen Wert $k \in \mathbb{R}_{\geqslant 0}$ einen Wert $x^* \in [\![0, 1000]\!]$, derart dass $\mathcal{A}_k = [\![x^*, 1000]\!]$ gilt, so dass jeder Neyman-Pearson-Test durch einen Ablehnbereich dieser Form gegeben ist. □

§06.16 **Neyman-Pearson Lemma**. *Sei $(\mathcal{X}, \mathscr{X}, (\mathbb{P}_0, \mathbb{P}_1))$ ein (binäres) diskretes statistisches Experiment. Für das Testproblem der einfachen Nullhypothese $H_0 : \{\mathbb{P}_0\}$ gegen die einfache Alternative*

H_1 : $\{\mathbb{P}_1\}$ ist jeder Neyman-Pearson-Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$ mit kritischem Wert $k \in \mathbb{R}_{\geqslant 0}$ und Ablehnbereich $\mathcal{A}_k \in \mathscr{X}$ wie in Definition §06.14 ein bester Test zum Niveau $\mathbb{P}_0(\mathcal{A}_k) \in [0,1]$.

§06.17 **Beweis** von Satz §06.16. In der Vorlesung. □

§06.18 **Bemerkung**. Variieren wir den kritischen Wert $k \in \mathbb{R}_{\geqslant 0}$ im Neyman-Pearson-Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$, so ändert sich auch der entsprechende Fehler 1. Art $\mathbb{P}_0(\mathcal{A}_k)$. Für jedes $\tilde{k} \in \mathbb{R}_{\geqslant k}$ gilt

$$\{\mathbb{p}_1 > 0\} = \mathcal{A}_0 \supseteq \mathcal{A}_k = \{\mathbb{p}_1 > k\,\mathbb{p}_0\} \supseteq \mathcal{A}_{\tilde{k}} \supseteq \{\mathbb{p}_1 > 0\} \cap \{\mathbb{p}_0 = 0\}.$$

sowie $\mathcal{A}_k \downarrow \{\mathbb{p}_1 > 0\} \cap \{\mathbb{p}_0 = 0\}$ für $k \to \infty$ und $\mathcal{A}_k \uparrow \{\mathbb{p}_1 > 0\}$ für $k \to 0$. Damit folgt für den

Fehler 1. Art

- $\mathbb{P}_0(\varphi^\star_k = 1) = \mathbb{P}_0(\mathcal{A}_k) \downarrow \mathbb{P}_0(\{\mathbb{p}_1 > 0\} \cap \{\mathbb{p}_0 = 0\}) = 0$ für $k \to \infty$ (da $\mathbb{P}_0(\mathbb{p}_0 = 0) = 0$) und
- $\mathbb{P}_0(\varphi^\star_k = 1) = \mathbb{P}_0(\mathcal{A}_k) \uparrow \mathbb{P}_0(\mathbb{p}_1 > 0) \in [0,1]$ für $k \to 0$.

Fehler 2. Art

- $\mathbb{P}_1(\varphi^\star_k = 0) = \mathbb{P}_1(\mathcal{A}^c_k) \uparrow \mathbb{P}_1(\{\mathbb{p}_1 = 0\} \cup \{\mathbb{p}_0 > 0\}) = \mathbb{P}_1(\mathbb{p}_0 > 0) \in [0,1]$ für $k \to \infty$ und
- $\mathbb{P}_1(\varphi^\star_k = 1) = \mathbb{P}_1(\mathcal{A}^c_k) \downarrow \mathbb{P}_1(\mathbb{p}_1 = 0) \in [0,1]$ für $k \to 0$.

Diese Änderung ist aber im Allgemeinen nicht stetig, so dass zu einem vorgegebenen Niveau $\alpha \in (0,1)$ nicht immer ein kritischer Wert k_α und entsprechender Neyman-Pearson-Test $\varphi^\star_{k_\alpha} = \mathbb{1}_{\mathcal{A}_{k_\alpha}}$ mit $\mathbb{P}_0(\varphi^\star_k = 1) = \mathbb{P}_0(\mathcal{A}_{k_\alpha}) = \alpha$ gewählt werden kann. In diesem Fall wählen wir zu einem vorgegebenen Niveau $\alpha \in (0,1)$ den kritischen Wert

$$k_\alpha := \min\{k \in \mathbb{R}_{\geqslant 0} : \ \alpha \geqslant \mathbb{P}_0(\mathcal{A}_k)\}.$$

Dabei ist die definierende Menge nichtleer, da $\mathbb{P}_0(\mathcal{A}_k) \downarrow 0$ für $k \to \infty$ gilt. Offensichtlich gilt dann auch $\mathbb{P}_0(\mathcal{A}_{k_\alpha}) \leqslant \alpha$ und der entsprechende Neyman-Pearson-Test $\varphi^\star_{k_\alpha} = \mathbb{1}_{\mathcal{A}_{k_\alpha}}$ mit Ablehnbereich $\mathcal{A}_{k_\alpha}$ ist ein α-Test, der aber im Allgemeinen nicht mehr ein bester α-Test ist, wie wir in der Vorlesung Statistik I sehen werden. □

§06.19 **Beispiel** *(Binomialverteilungsmodell §06.15 fortgesetzt)*. Für jedes $x^* \in [\![0, 1000]\!]$ ist ein Neyman-Pearson-Test $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,1000]\!]}$ mit Ablehnbereich $[\![x^*, 1000]\!]$ ein bester Test zum Niveau

$$\mathrm{Bin}_{(1000,0.5)}([\![x^*, 1000]\!]) = \sum_{x \in [\![x^*,1000]\!]} \binom{1000}{x} 0.5^x (1-0.5)^{1000-x} = 0.5^{1000} \sum_{x \in [\![x^*,1000]\!]} \binom{1000}{x}$$

wobei der entsprechende Fehler 2. Art

$$\mathrm{Bin}_{(1000,0.7)}([\![0, x^*)\!)) = \sum_{x \in [\![0,x^*)\!)} \binom{1000}{x} 0.7^x (1-0.7)^{1000-x}$$

minimal ist. Zum Beispiel für $x^* = 538$ erhalten wir $\mathrm{Bin}_{(1000,0.5)}([\![538, 1000]\!]) \approx 0.0088$ und $\mathrm{Bin}_{(1000,0.7)}([\![0, 537]\!]) \approx 0$ sowie für $x^* = 537$ berechnen wir $\mathrm{Bin}_{(1000,0.5)}([\![537, 1000]\!]) \approx 0.0105$ und $\mathrm{Bin}_{(1000,0.7)}([\![0, 536]\!]) \approx 0$. Sind wir an einem Test zu einem vorgegebenen Niveau $\alpha \in (0,1)$ interessiert, so wählen wir

$$x_\alpha := \min\{x^* \in [\![0, 1000]\!] : \ \alpha \geqslant \mathrm{Bin}_{(1000,0.5)}([\![x^*, 1000]\!])\}.$$

Dann ist der entsprechende Neyman-Pearson-Test $\varphi^\star_{x_\alpha} = \mathbb{1}_{[\![x_\alpha,1000]\!]}$ mit Ablehnbereich $[\![x_\alpha, 1000]\!]$ ein α-Test. Ein typischer Wert für das Niveau ist $\alpha = 0.01$, so dass mit der obigen Rechnung

$x_{0.01} = 538$ und der entsprechende Neyman-Pearson-Test $\varphi^\star_{538} = \mathbb{1}_{[\![538,1000]\!]}$ den Ablehnbereich $[\![538, 1000]\!]$ besitzt. Im Beispiel I im Prolog §1 zählte der firmeneigene Verbraucherservice 699 Konsumentinnen unter den 1000 befragten Personen, so dass die Nullhypothese, der Anteil der Konsumentinnen beträgt 50%, gegen die Alternative, der Anteil der Konsumentinnen beträgt 70%, zum Niveau $\alpha = 0.01$ abgelehnt werden kann. Offensichtlich hängt ein Neyman-Pearson-Test $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,1000]\!]}$ mit Ablehnbereich $[\![x^*, 1000]\!]$ nicht von dem Parameter $\mathrm{p}_1 \in (\mathrm{p}_0, 1]$ der Alternative ab. Damit ist der Neyman-Pearson-Test mit Ablehnbereich $[\![538, 1000]\!]$ auch bester Test zum Niveau $\alpha = \mathrm{Bin}_{(1000,0.5)}([\![538, 1000]\!])$ der einfachen Nullhypothese $H_0 : \{\mathrm{Bin}_{(1000,0.5)}\}$ gegen die einfache Alternative $H_1 : \{\mathrm{Bin}_{(1000,0.55)}\}$ im binären diskreten statistischen Experiment $\big([\![0, 1000]\!], 2^{[\![0,1000]\!]}, (\mathrm{Bin}_{(1000,\mathrm{p})})_{\mathrm{p}\in\{0.5,0.55\}}\big)$. Beachte, dass der minimale Fehler 2. Art dann $\mathrm{Bin}_{(1000,0.55)}([\![0, 537]\!]) \approx 0.213$ ist, und dieser hängt natürlich von der Alternative ab.
Geben wir uns das Niveau $\alpha \in (0, 1)$ vor, so können wir auch zu jedem $n \in \mathbb{N}$ im binären diskreten statistischen Experiment $\big([\![0, n]\!], 2^{[\![0,n]\!]}, (\mathrm{Bin}_{(n,\mathrm{p})})_{\mathrm{p}\in\{0.5,0.55\}}\big)$ für das Testproblem der einfachen Nullhypothese $H_0 : \{\mathrm{Bin}_{(n,0.5)}\}$ gegen die einfache Alternative $H_1 : \{\mathrm{Bin}_{(n,0.55)}\}$ den Neyman-Pearson-Test $\varphi^\star_{x_{\alpha,n}} = \mathbb{1}_{[\![x_{\alpha,n},n]\!]}$ mit Ablehnbereich $[\![x_{\alpha,n}, n]\!]$, derart dass

$$x_{\alpha,n} := \arg\min\,\{x^* \in [\![0, n]\!] : \ \alpha \geqslant \mathrm{Bin}_{(n,0.5)}([\![x^*, n]\!])\},$$

wählen. Der Neyman-Pearson-Test $\varphi^\star_{x_{\alpha,n}} = \mathbb{1}_{[\![x_{\alpha,n},n]\!]}$ ist dann ein α-Test, sowie der beste Test zum Niveau $\mathrm{Bin}_{(n,0.5)}([\![x_{\alpha,n}, n]\!]) \leqslant \alpha$. Für $\alpha = 0.01$ erhalten wir zum Beispiel die kritischen Werte $x_{0.01,500} = 277$, $x_{0.01,1000} = 538$ und $x_{0.01,2000} = 1053$ mit entsprechendem Fehler 1. Art

$$\mathrm{Bin}_{(500,0.5)}([\![277, 500]\!]) \approx 0.0088, \quad \mathrm{Bin}_{(1000,0.5)}([\![538, 1000]\!]) \approx 0.0088 \quad \text{und} \quad \mathrm{Bin}_{(2000,0.5)}([\![1053, 2000]\!]) \approx 0.0094,$$

sowie Fehler 2. Art $\mathrm{Bin}_{(500,0.55)}([\![0, 276]\!]) \approx 0.553$, $\mathrm{Bin}_{(1000,0.55)}([\![0, 537]\!]) \approx 0.213$ und $\mathrm{Bin}_{(2000,0.55)}([\![0, 1052]\!]) \approx 0.016$, der offensichtlich von $n \in \mathbb{N}$ abhängt. Wir können uns also auch eine obere Schranke $\beta \in (0, 1)$ für den Fehler 2. Art vorgegeben, und nach dem kleinsten Wert für n fragen, so dass der Neyman-Pearson-Test $\varphi^\star_{x_{\alpha,n}} = \mathbb{1}_{[\![x_{\alpha,n},n]\!]}$ mit Ablehnbereich $[\![x_{\alpha,n}, n]\!]$ diese einhält, das heißt,

$$n_{\alpha,\beta} := \arg\min\,\{n \in \mathbb{N} : \ \beta \geqslant \mathrm{Bin}_{(n,0.55)}([\![0, x_{\alpha,n}[\![)\}.$$

Für $\alpha = 0.01$ und $\beta = 0.01$ erhalten wir $n_{0.01,0.01} = 2170$, wobei $x_{0.01,2170} = 1140$, der Fehler 1. Art $\mathrm{Bin}_{(2170,0.5)}([\![1140, 2170]\!]) \approx 0.0096$ und der Fehler 2. Art $\mathrm{Bin}_{(2170,0.55)}([\![0, 1139]\!]) \approx 0.00997$ ist. Zusammenfassend, im Beispiel I im Prolog §1 müsste der firmeneigene Verbraucherservice mindestens 2170 Konsumierende befragen, um sicherzustellen, dass der Fehler 1. Art und der Fehler 2. Art des entsprechenden Neyman-Pearson-Tests nicht größer als 0.01 ist. □

§06.20 **Ausblick.** Betrachten wir für ein beliebiges $\mathrm{p}_0 \in [0, 1)$, $\mathrm{p}_1 \in (\mathrm{p}_0, 1]$ und $n \in \mathbb{N}$ das Testproblem der einfachen Nullhypothese $H_0 : \{\mathrm{Bin}_{(n,\mathrm{p}_0)}\}$ gegen die einfache Alternative $H_1 : \{\mathrm{Bin}_{(n,\mathrm{p}_1)}\}$ im diskreten statistischen Experiment $\big([\![0, n]\!], 2^{[\![0,n]\!]}, (\mathrm{Bin}_{(n,\mathrm{p}_0)}, \mathrm{Bin}_{(n,\mathrm{p}_1)})\big)$, so hängt der Ablehnbereich $[\![x^*, n]\!]$ mit $x^* \in [\![0, n]\!]$ eines Neyman-Pearson-Tests $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,n]\!]}$ nicht von dem Wert $\mathrm{p}_1 \in (\mathrm{p}_0, 1]$ der Alternative ab. Damit ist für jedes $\mathrm{p}_1 \in (\mathrm{p}_0, 1]$ ein Neyman-Pearson-Test $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,n]\!]}$ mit Ablehnbereich $[\![x^*, n]\!]$ bester Test zum Niveau $\mathrm{Bin}_{(n,\mathrm{p}_0)}([\![x^*, n]\!])$ der *einfachen Nullhypothese* $H_0 : \{\mathrm{Bin}_{(n,\mathrm{p}_0)}\}$ gegen die *einfache Alternative* $H_1 : \{\mathrm{Bin}_{(n,\mathrm{p}_1)}\}$. Dies erlaubt uns, das diskrete statistische Experiment $\big([\![0, n]\!], 2^{[\![0,n]\!]}, (\mathrm{Bin}_{(n,\mathrm{p})})_{\mathrm{p}\in[\mathrm{p}_0,1]}\big)$ zu betrachten. Für jedes $x^* \in [\![0, n]\!]$ ist der Neyman-Pearson Test $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,n]\!]}$ mit Ablehnbereich $[\![x^*, n]\!]$ dann *gleichmäßig bester Test* zum Niveau $\alpha = \mathrm{Bin}_{(n,\mathrm{p}_0)}([\![x^*, n]\!])$ der *einfachen Nullhypothese* $H_0 : \{\mathrm{Bin}_{(n,\mathrm{p}_0)}\}$ gegen die *zusammengesetzte Alternative* $H_1 : \{\mathrm{Bin}_{(n,\mathrm{p}_1)} : \mathrm{p}_1 \in (\mathrm{p}_0, 1]\}$, da der Fehler 2. Art für jeden anderen Test $\varphi = \mathbb{1}_{\mathcal{A}}$ zum Niveau α mit Ablehnbereich $\mathcal{A}$ gleichmäßig nicht kleiner ist, das heißt, $\mathrm{Bin}_{(n,\mathrm{p}_1)}([\![x^*, n]\!]^c) \leqslant$

$\mathrm{Bin}_{(n,\mathrm{p}_1)}(\mathcal{A}^c)$ für alle $\mathrm{p}_1 \in (\mathrm{p}_0, 1]$ gilt. Wir halten fest, dass diese Schlussfolgerung möglich ist, da für jedes $\mathrm{p}_0 \in [0, 1)$ und $\mathrm{p}_1 \in (\mathrm{p}_0, 1]$ die Funktion

$$\mathrm{L}_{\mathrm{p}_1,\mathrm{p}_0} : [\![0, n]\!] \ni x \mapsto \mathrm{L}_{\mathrm{p}_1,\mathrm{p}_0}(x) := \left(\tfrac{\mathrm{p}_1/(1-\mathrm{p}_1)}{\mathrm{p}_0/(1-\mathrm{p}_0)}\right)^x \left(\tfrac{1-\mathrm{p}_1}{1-\mathrm{p}_0}\right)^n \in \mathbb{R}_{\geqslant 0},$$

streng monoton wachsend ist. Bezeichnet $\mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p})}}$ die Zähldichte der Verteilung $\mathrm{Bin}_{(n,\mathrm{p})}$ für $\mathrm{p} \in [0, 1]$. $\mathrm{L}_{\mathrm{p}_1,\mathrm{p}_0}$ wird *Likelihood-Quotient* (oder *Dichtequotient*) genannt, da offensichtlich gilt

$$\mathrm{L}_{\mathrm{p}_1,\mathrm{p}_0}(\mathrm{x}) = \tfrac{\mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p}_1)}}(x)}{\mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p}_0)}}(x)}, \quad x \in [\![0, n]\!].$$

Die Verteilungsfamilie $(\mathrm{Bin}_{(n,\mathrm{p})})_{\mathrm{p}\in[0,1]}$ besitzt damit einen *monotonen Likelihood-Quotienten*. Im Beispiel I im Prolog §1 kann also der firmeneigene Verbraucherservice die einfache Nullhypothese, der Anteil der Konsumentinnen beträgt 50%, gegen die zusammengesetzte Alternative, der Anteil der Konsumentinnen beträgt mehr als 50%, zum Niveau $\alpha = 0.01$ ablehnen. □

Kapitel 3

Zufallsvariablen

§07 Zufallsvariable

§07.01 **Definition**. Es seien $(\Omega, \mathscr{A})$ und $(\mathcal{X}, \mathscr{X})$ zwei messbare Räume. Eine Funktion $X : \Omega \to \mathcal{X}$ heißt *$\mathscr{A}$-$\mathscr{X}$-messbar* (kurz *messbar*), wenn

$$\sigma(X) := X^{-1}(\mathscr{X}) := \{X^{-1}(B) : B \in \mathscr{X}\} \subseteq \mathscr{A}$$

gilt. Jede solche messbare Funktion wird ($(\mathcal{X}, \mathscr{X})$-wertige) *Zufallsvariable* genannt und wir bezeichnen mit $\mathcal{M}(\mathscr{A}, \mathscr{X})$ die Menge aller $\mathscr{A}$-$\mathscr{X}$-messbarer Funktionen. □

§07.02 **Schreibweise**. Für $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$, $B \subseteq \mathcal{X}$ und $x \in \mathcal{X}$ schreiben wir kurz

$$\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \text{ und}$$
$$\{X = x\} := X^{-1}(\{x\}) = \{\omega \in \Omega : X(\omega) = x\}.$$
□

§07.03 **Lemma**. *Es sei $(\mathcal{X}, \mathscr{X})$ ein messbarer Raum und $X : \Omega \to \mathcal{X}$ eine Abbildung. Dann ist das Teilmengensystem $\sigma(X) = X^{-1}(\mathscr{X}) \subseteq 2^{\Omega}$ eine σ-Algebra auf Ω, die sogenannte* von X erzeugte σ-Algebra.

§07.04 **Beweis** von Lemma §07.03. Übungsaufgabe. □

§07.05 **Bemerkung**. Nach Lemma §07.03 ist $\sigma(X) = X^{-1}(\mathscr{X})$ die kleinste σ-Algebra auf Ω, sodass die Funktion $X : \Omega \to \mathcal{X}$ eine Zufallsvariable ist, d.h. $X \in \mathcal{M}(X^{-1}(\mathscr{X}), \mathscr{X})$. □

§07.06 **Lemma**. *Es seien $X : \Omega \to \mathcal{X}$ eine Abbildung und $\mathscr{E} \subseteq 2^{\mathcal{X}}$ ein Teilmengensystem. Dann gilt $\sigma(X^{-1}(\mathscr{E})) = X^{-1}(\sigma(\mathscr{E}))$.*

§07.07 **Beweis** von Lemma §07.06. In der Vorlesung. □

§07.08 **Lemma**. *Seien $(\Omega, \mathscr{A})$ und $(\mathcal{X}, \mathscr{X})$ zwei messbare Räume und $\mathscr{E}$ ein Erzeuger von $\mathscr{X}$, d.h. $\mathscr{X} = \sigma(\mathscr{E})$. Jede Funktion $X : \Omega \to \mathcal{X}$ mit $X^{-1}(\mathscr{E}) \subseteq \mathscr{A}$ ist $\mathscr{A}$-$\mathscr{X}$-messbar, also eine $(\mathcal{X}, \mathscr{X})$-wertige Zufallsvariable, d.h. $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$.*

§07.09 **Beweis** von Lemma §07.08. Übungsaufgabe. □

§07.10 **Korollar**. *Jede stetige Funktion $g : \mathcal{S} \to \mathcal{T}$ zwischen metrischen Räumen $\mathcal{S}$ und $\mathcal{T}$ ist $\mathscr{B}_{\mathcal{S}}$-$\mathscr{B}_{\mathcal{T}}$-messbar, kurz* Borel-messbar.

§07.11 **Beweis** von Korollar §07.10. In der Vorlesung. □

§07.12 **Proposition**. *Seien $(\Omega, \mathscr{A})$, $(\mathcal{X}, \mathscr{X})$ und $(\mathcal{S}, \mathscr{S})$ drei messbare Räume. Für jedes $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$ und $h \in \mathcal{M}(\mathscr{X}, \mathscr{S})$ ist die Hintereinanderausführung $h \circ X : \Omega \to \mathcal{S}$ eine $\mathscr{A}$-$\mathscr{S}$-messbare Abbildung, also $Y = h(X) \in \mathcal{M}(\mathscr{A}, \mathscr{S})$ eine $(\mathcal{S}, \mathscr{S})$-wertige Zufallsvariable.*

$$\begin{tikzcd} (\Omega,\mathscr{A}) \arrow[r, "X"] \arrow[dr, "Y=h(X)"'] & (\mathcal{X},\mathscr{X}) \arrow[d, "h"] \\ & (\mathcal{S},\mathscr{S}) \end{tikzcd}$$

§07.13 **Beweis** von Proposition §07.12. In der Vorlesung. □

§08 Numerische und reelle Zufallsvariablen

§08.01 **Definition**. Eine Zufallsvariable $X : (\Omega, \mathscr{A}) \to (\mathcal{X}, \mathscr{X})$ heißt

numerisch, wenn $(\mathcal{X}, \mathscr{X}) = (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, kurz $X \in \overline{\mathcal{M}}(\mathscr{A}) := \mathcal{M}(\mathscr{A}, \overline{\mathscr{B}})$;

positiv numerisch, wenn $(\mathcal{X}, \mathscr{X}) = (\overline{\mathbb{R}}_{\geqslant 0}, \overline{\mathscr{B}}_{\geqslant 0})$, kurz $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) := \mathcal{M}(\mathscr{A}, \overline{\mathscr{B}}_{\geqslant 0})$;

reell, wenn $(\mathcal{X}, \mathscr{X}) = (\mathbb{R}, \mathscr{B})$, kurz $X \in \mathcal{M}(\mathscr{A}) := \mathcal{M}(\mathscr{A}, \mathscr{B})$;

positiv reell, wenn $(\mathcal{X}, \mathscr{X}) = (\mathbb{R}_{\geqslant 0}, \mathscr{B}_{\geqslant 0})$, kurz $X \in \mathcal{M}_{\geqslant 0}(\mathscr{A}) := \mathcal{M}(\mathscr{A}, \mathscr{B}_{\geqslant 0})$;

Zufallsvektor, wenn $(\mathcal{X}, \mathscr{X}) = (\mathbb{R}^n, \mathscr{B}^n)$.

Falls der Urbildraum $(\Omega, \mathscr{A})$ keine Rolle spielt, schreiben wir auch kurz $\overline{\mathcal{M}} := \overline{\mathcal{M}}(\mathscr{A})$, $\overline{\mathcal{M}}_{\geqslant 0} := \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, $\mathcal{M} := \mathcal{M}(\mathscr{A})$ und $\mathcal{M}_{\geqslant 0} := \mathcal{M}_{\geqslant 0}(\mathscr{A})$. □

§08.02 **Bemerkung**. Ist speziell $X \in \mathcal{M}$, so ist X in kanonischer Weise auch in $\overline{\mathcal{M}}$. Eine detaillierte Diskussion findet man zum Beispiel in Klenke (2012), Abschnitt 1.4. □

§08.03 **Beispiel**.

(a) Seien $(\Omega, \mathscr{A})$ ein messbarer Raum und $\mathbb{1}_A$ eine Indikatorfunktion mit $A \subseteq \Omega$. Dann gilt $\mathbb{1}_A \in \mathcal{M}(\mathscr{A})$, d.h. $\mathbb{1}_A$ ist eine reelle Zufallsvariable, genau dann, wenn $A \in \mathscr{A}$ gilt.

(b) Sei $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P})$ ein stetiger Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsdichte $\mathbb{f} : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$, dann ist $\mathbb{f} \in \mathcal{M}_{\geqslant 0}(\mathscr{B}^n)$ eine positive reelle Zufallsvariable. □

§08.04 **Lemma**. *Sei $(\Omega, \mathscr{A})$ ein messbarer Raum.*

(i) *Eine Funktion $X : \Omega \to \overline{\mathbb{R}}$ ist eine numerische Zufallsvariable, also $X \in \overline{\mathcal{M}}(\mathscr{A})$, genau dann, wenn für alle $x \in \mathbb{R}$ gilt $\{X \leqslant x\} = X^{-1}(\overline{\mathbb{R}}_{\leqslant x}) \in \mathscr{A}$.*

(ii) *Eine Funktion $X : \Omega \to \mathbb{R}^n$ ist ein Zufallsvektor, also $X \in \mathcal{M}(\mathscr{A}, \mathscr{B}^n)$, genau dann, wenn jede Komponente des Vektors eine reelle Zufallsvariable ist.*

§08.05 **Beweis** von Lemma §08.04. In der Vorlesung. □

Die Familie numerischer (bzw. reeller) Zufallsvariablen ist stabil für fast alle vorstellbaren Operationen.

§08.06 **Lemma**. *Für numerische Zufallsvariablen $X, Y \in \overline{\mathcal{M}}(\mathscr{A})$ gilt:*

(i) *Für alle $a \in \mathbb{R}$ gilt $aX \in \overline{\mathcal{M}}(\mathscr{A})$ (mit der Konvention $0 \times \infty = 0$);*

(ii) *$X \vee Y, X \wedge Y \in \overline{\mathcal{M}}(\mathscr{A})$ und $X^+ = X \vee 0$, $X^- = (-X)^+$, $|X| = X^+ + X^- \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$;*

(iii) *$\{X < Y\}, \{X \leqslant Y\}, \{X = Y\} \in \mathscr{A}$.*

§08.07 **Beweis** von Lemma §08.06. Übungsaufgabe. □

§08.08 **Lemma**. *Es seien $X_i \in \mathcal{M}(\mathscr{A})$, $i \in [\![n]\!]$, $n \in \mathbb{N}$, reelle Zufallsvariablen und es sei $h \in \mathcal{M}(\mathscr{B}^n, \mathscr{B}^m)$ Borel-messbar. Dann ist $h((X_i)_{i\in[\![n]\!]}) \in \mathcal{M}(\mathscr{A}, \mathscr{B}^m)$ ein Zufallsvektor. Insbesondere gilt $(X_i)_{i\in[\![n]\!]} \in \mathcal{M}(\mathscr{A}, \mathscr{B}^n)$ und $X_1 + X_2$, $X_1 - X_2$, $X_1 X_2 \in \mathcal{M}(\mathscr{A})$ sowie, falls überall wohldefiniert, $X_1/X_2 \in \mathcal{M}(\mathscr{A})$.*

§08.09 **Beweis** von Lemma §08.08. In der Vorlesung. □

§08.10 **Definition**. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge numerischer (bzw. reeller) Zufallsvariablen auf $(\Omega, \mathscr{A})$. Die Folge $(X_n)_{n\in\mathbb{N}}$ heißt *(punktweise) monoton wachsend*, wenn $X_n(\omega) \leqslant X_{n+1}(\omega)$ (bzw. *(punktweise) monoton fallend*, wenn $X_{n+1}(\omega) \leqslant X_n(\omega)$) für alle $\omega \in \Omega$ und für alle $n \in \mathbb{N}$ gilt. Für jedes $\omega \in \Omega$ definiere

$$[\liminf_{n\to\infty} X_n](\omega) := \sup_{n\in\mathbb{N}} \inf_{m\in\mathbb{N}_{\geqslant n}} X_m(\omega) := \sup\{\inf\{X_m(\omega) : m \in \mathbb{N}_{\geqslant n}\} : n \in \mathbb{N}\};$$

$$[\limsup_{n\to\infty} X_n](\omega) := \inf_{n\in\mathbb{N}} \sup_{m\in\mathbb{N}_{\geqslant n}} X_m(\omega) := \inf\{\sup\{X_m(\omega) : m \in \mathbb{N}_{\geqslant n}\} : n \in \mathbb{N}\}.$$

Dann heißen $X_\star := \boxed{\liminf_{n\to\infty} X_n}$ und $X^\star := \boxed{\limsup_{n\to\infty} X_n}$ *Limes inferior* bzw. *Limes superior* der Folge $(X_n)_{n\in\mathbb{N}}$. Die Folge $(X_n)_{n\in\mathbb{N}}$ heißt *(punktweise) konvergent*, wenn $X_\star = X^\star =: X$ gilt, d.h. der punktweise Grenzwert existiert überall. In diesem Fall schreiben wir kurz $\boxed{\lim_{n\to\infty} X_n} := X$. □

§08.11 **Bemerkung**.

(a) Jede *monoton wachsende* (bzw. *fallende*) Folge $(X_n)_{n\in\mathbb{N}}$ von numerischen Zufallsvariablen ist konvergent mit $X := \lim_{n\to\infty} X_n = \sup_{n\in\mathbb{N}} X_n$ (bzw. $X := \lim_{n\to\infty} X_n = \inf_{n\in\mathbb{N}} X_n$). In diesem Fall schreiben wir kurz $X_n \uparrow X$ (bzw. $X_n \downarrow X$).

(b) Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge von Teilmengen aus Ω. Für $A_\star$ und $A^\star$ wie in Definition §01.14 gilt dann $\mathbb{1}_{A_\star} = \liminf_{n\to\infty} \mathbb{1}_{A_n}$ und $\mathbb{1}_{A^\star} = \limsup_{n\to\infty} \mathbb{1}_{A_n}$. □

§08.12 **Lemma**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\overline{\mathcal{M}}$. Dann gilt:*

(i) $\sup_{n\in\mathbb{N}} X_n \in \overline{\mathcal{M}}$ *und* $\inf_{n\in\mathbb{N}} X_n \in \overline{\mathcal{M}}$;

(ii) $\liminf_{n\to\infty} X_n \in \overline{\mathcal{M}}$ *und* $\limsup_{n\to\infty} X_n \in \overline{\mathcal{M}}$;

(iii) $\lim_{n\to\infty} X_n \in \overline{\mathcal{M}}$, *falls* $(X_n)_{n\in\mathbb{N}}$ *konvergent ist.*

§08.13 **Beweis** von Lemma §08.12. Übungsaufgabe. □

§09 Einfache Zufallsvariable

Eine reelle Zufallsvariable Y, die nur die Werte 0 oder 1 annimmt, wird Beispiel §04.08 (c) entsprechend *Bernoulli-Zufallsvariable* genannt. Diese entspricht gerade der Indikatorfunktion $\mathbb{1}_A$ auf dem Ereignis $A = \{Y = 1\}$, da $Y = \mathbb{1}_{\{Y=1\}}$ gilt.

§09.01 **Lemma**.

(i) *Seien $A, B \subseteq \Omega$, dann gilt: $\mathbb{1}_{A\cap B} = \mathbb{1}_A \wedge \mathbb{1}_B = \mathbb{1}_A \mathbb{1}_B$, $\mathbb{1}_{A\cup B} = \mathbb{1}_A \vee \mathbb{1}_B = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B$ und $\mathbb{1}_{A\Delta B} = |\mathbb{1}_A - \mathbb{1}_B|$. Insbesondere ist $A \subseteq B$ genau dann, wenn $\mathbb{1}_A \leqslant \mathbb{1}_B$.*

(ii) *Für $X : \Omega \to \mathcal{X}$ und $B \subseteq \mathcal{X}$ gilt $\mathbb{1}_{\{X\in B\}} = \mathbb{1}_{X^{-1}(B)} = \mathbb{1}_B \circ X = \mathbb{1}_B(X)$.*

§09.02 **Beweis** von Lemma §09.01. Übung. □

§09.03 **Definition**. Eine reelle Zufallsvariable $X \in \mathcal{M}(\mathscr{A})$ auf $(\Omega, \mathscr{A})$ heißt

einfach oder *elementar*, wenn sie nur endlich viele reelle Werte annimmt.

Für $X(\Omega) = \{x_i : i \in [\![n]\!]\} \subseteq \mathbb{R}$ mit $n \in \mathbb{N}$ besitzt X eine Darstellung der Form $X = \sum_{i\in[\![n]\!]} x_i \mathbb{1}_{A_i}$ mit $A_i := X^{-1}(\{x_i\}) = \{X = x_i\} \in \mathscr{A}$, $i \in [\![n]\!]$. Wir bezeichnen mit $\mathcal{M}^{\text{einf}}(\mathscr{A})$ und $\mathcal{M}^{\text{einf}}_{\geqslant 0}(\mathscr{A})$ die Menge aller einfachen bzw. positiven einfachen Zu fallsvariablen auf $(\Omega, \mathscr{A})$. □

Einfache Zufallsvariablen erlauben numerische (bzw. reelle) Zufallsvariablen zu approximieren. Zur Erinnerung für $a \in \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ setzen wir $\lfloor a \rfloor := \sup\{z \in \mathbb{Z} : z \leqslant a\}$, d.h. für $a \in \mathbb{R}$ ist $\lfloor a \rfloor$ die nächstliegende nicht größere ganze Zahl.

§09.04 **Lemma**. *Für jedes $X \in \overline{\mathcal{M}}_{\geqslant 0}$ ist $(X_n)_{n\in\mathbb{N}}$ mit $X_n := (2^{-n}\lfloor 2^n X \rfloor) \wedge n$ für $n \in \mathbb{N}$ eine Folge einfacher Zufallsvariablen aus $\mathcal{M}_{\geqslant 0}$, derart dass gilt*

(i) *$X_n \uparrow X$;*

(ii) *$X_n \leqslant X \wedge n$, d.h. $X_n(\omega) \leqslant X(\omega) \wedge n$ für alle $\omega \in \Omega$;*

(iii) *Für jedes $c \in \mathbb{R}_{\geqslant 0}$ gilt $\lim_{n\to\infty} X_n = X$ gleichmäßig auf $\{X \leqslant c\}$, d.h. $\lim_{n\to\infty} \sup_{\{\omega : X(\omega)\leqslant c\}} |X(\omega) - X_n(\omega)| = 0$.*

§09.05 **Beweis** von Lemma §09.04. In der Vorlesung. □

§09.06 **Beweistrategie.**. Möchten wir zeigen, dass jede numerische Zufallsvariable Y eine bestimmte Eigenschaft, sagen wir (R), besitzt, so ist eine häufig angewendete Beweisstrategie:

(Schritt 1) Zeige, dass Bernoulli-Zufallsvariablen die Eigenschaft (R) erfüllen;

(Schritt 2) Zeige, dass einfache Zufallsvariablen die Eigenschaft (R) erfüllen;

(Schritt 3) Zeige, dass die Eigenschaft (R) für den Grenzwert einer monoton wachsenden Folge von elementaren Zufallsvariablen gilt, sodass nach Lemma §09.04 auch positive numerische Zufallsvariablen die Eigenschaft (R) besitzen;

(Schritt 4) Zeige die Eigenschaft (R) für Y mittels der Zerlegung $Y = Y^+ - Y^-$. □

Bevor wir uns das folgende Resultat anschauen, wollen wir an Proposition §07.12 erinnern.

§09.07 **Proposition**. *Seien $(\Omega, \mathscr{A})$, $(\mathcal{X}, \mathscr{X})$ zwei messbare Räume und $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$, $Y : \Omega \to \overline{\mathbb{R}}$ zwei Abbildungen. Dann sind die folgenden zwei Aussagen äquivalent:*

(A1) *Y ist messbar bzgl. $\sigma(X) = X^{-1}(\mathscr{X})$, kurz $Y \in \overline{\mathcal{M}}(\sigma(X))$;*

(A2) *Es existiert eine messbare Funktion $h : (\mathcal{X}, \mathscr{X}) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, kurz $h \in \overline{\mathcal{M}}(\mathscr{X})$, derart dass $Y = h(X)$ gilt.*

Falls Y reell oder beschränkt oder positiv ist, so erbt h diese Eigenschaft.

$$\begin{tikzcd}
(\Omega, \mathscr{A}) \arrow[r, "X"] \arrow[dr, "Y = h(X) \in \overline{\mathcal{M}}(\sigma(X))"'] & (\mathcal{X}, \mathscr{X}) \arrow[d, "h \in \overline{\mathcal{M}}(\mathscr{X})"] \\
& (\overline{\mathbb{R}}, \overline{\mathscr{B}})
\end{tikzcd}$$

§09.08 **Beweis** von Proposition §09.07. In der Vorlesung. □

§10 Verteilung einer Zufallsvariablen

§10.01 **Definition**. Die *Wahrscheinlichkeitsverteilung* (kurz: *Verteilung*) einer Zufallsvariable $X : (\Omega, \mathscr{A}, \mathbb{P}) \to (\mathcal{X}, \mathscr{X})$ bezeichnet die Abbildung $\boxed{\mathbb{P}^X} := \mathbb{P} \circ X^{-1} : \mathscr{X} \to [0,1]$, d.h. für alle $B \in \mathscr{X}$ gilt: $\mathbb{P}^X(B) := \mathbb{P} \circ X^{-1}(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$. □

§10.02 **Schreibweise**. $\mathbb{P}(X \in S) := \mathbb{P}(\{X \in S\}) = \mathbb{P}(X^{-1}(S))$, $\mathbb{P}(X = x) := \mathbb{P}(\{X = x\})$ etc. □

§10.03 **Lemma**. *Die Verteilung $\mathbb{P}^X$ einer $(\mathcal{X}, \mathscr{X})$-wertigen Zufallsvariable X ist ein Wahrscheinlichkeitsmaß auf $(\mathcal{X}, \mathscr{X})$, also $\mathbb{P}^X \in \mathcal{W}(\mathscr{X})$.*

§10.04 **Beweis** von Lemma §10.03. Übungsaufgabe. □

§10.05 **Definition**. Die Verteilung $\mathbb{P}^X \in \mathcal{W}(\mathscr{X})$ von $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$, kurz $\boxed{X \sim \mathbb{P}^X}$, wird auch Bildmaß von $\mathbb{P}$ unter X genannt. Mit der Verteilungsfunktion $\boxed{\mathbb{F}^X}$ (Dichte $\boxed{\mathbb{f}^X}$, Zähldichte $\boxed{\mathbb{p}^X}$) von X werden wir stets die zu $\mathbb{P}^X$ gehörigen Größen bezeichnen. X heißt *diskret-verteilte Zufallsvariable*, wenn $\mathbb{P}^X$ ein diskretes Wahrscheinlichkeitsmaß auf $(\mathcal{X}, \mathscr{X})$ ist. Eine reelle Zufallsvariable (Zufallsvektor) X mit stetigem Bildmaß $\mathbb{P}^X$ wird *stetig-verteilt* genannt. □

§10.06 **Schreibweise**.

(a) $\mathbb{F}^X(x) = \mathbb{P}^X(\mathbb{R}_{\leqslant x}) = \mathbb{P}(X \in \mathbb{R}_{\leqslant x}) = \mathbb{P}(X \leqslant x)$

(b) Bei Zufallsvariablen spielt der Urbildraum $(\Omega, \mathscr{A}, \mathbb{P})$ häufig keine Rolle und wird daher dann auch nicht angegeben.

(c) Ist die Verteilung einer reellen Zufallsvariable X zum Beispiel eine Normalverteilung $\mathrm{N}_{(\mu,\sigma^2)}$ (vgl. Beispiel §04.08), so schreiben wir kurz $X \sim \mathrm{N}_{(\mu,\sigma^2)}$. □

§10.07 **Definition**. $(\mathcal{X}, \mathscr{X})$-wertige Zufallsvariablen $\{X_i : i \in \mathcal{I}\}$ heißen *identisch verteilt* (kurz *i.v.*), wenn für alle $i \in \mathcal{I}$, $\mathbb{P}^{X_i} = \mathbb{P}$ für ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{X})$ gilt. □

§10.08 **Dichtetransformationssatz**. *Sei X ein stetig-verteilter n-dimensionaler Zufallsvektor mit stetiger Dichte $\mathbb{f}^X$, sodass $\mathbb{F}^X(x) = \int_{\mathbb{R}_{\leqslant x_1}} \cdots \int_{\mathbb{R}_{\leqslant x_n}} \mathbb{f}^X(y_1, \dots, y_n) dy_n \cdots dy_1$ für jedes $x \in \mathbb{R}^n$ gilt. Sei $A \subseteq \mathbb{R}^n$ eine offene (oder abgeschlossene) Menge mit $\mathbb{P}^X(\mathbb{R}^n \backslash A) = 0$. Ferner sei $B \subseteq \mathbb{R}^n$ offen oder abgeschlossen sowie $h : A \to B$ bijektiv und stetig differenzierbar mit Ableitung h'. Dann ist $Y := h(X)$ auch stetig-verteilt mit die Dichte $\mathbb{f}^Y(y) = \frac{\mathbb{f}^X(h^{-1}(y))}{|\det(h'(h^{-1}(y)))|}$ für $y \in B$ und $\mathbb{f}^Y(y) = 0$ für $y \in \mathbb{R}^n \backslash B$.*

§10.09 **Beweis** von Satz §10.08. In der Vorlesung Analysis 3. □

§10.10 **Korollar**.

(i) *Ist X eine reelle Zufallsvariable mit Dichte $\mathbb{f}^X$, so besitzt $Y = aX + b$ für $a \in \mathbb{R}_{\neq 0}$ und $b \in \mathbb{R}$ die Dichte $\mathbb{f}^Y(y) = \frac{1}{|a|}\mathbb{f}^X(a^{-1}(y - b))$, $y \in \mathbb{R}$.*

(ii) *Ist X aufgefasst als Spaltenvektor ein n-dimensionaler Zufallsvektor mit Dichte $\mathbb{f}^X$, so besitzt $Y = AX + b$ für reguläres $A \in \mathbb{R}^{n \times n}$ und $b \in \mathbb{R}^n$ die Dichte $\mathbb{f}^Y(y) = \frac{\mathbb{f}^X(A^{-1}(y-b))}{|\det A|}$.*

§10.11 **Beweis** von Korollar §10.10. Direktes Anwenden von Satz §10.08. □

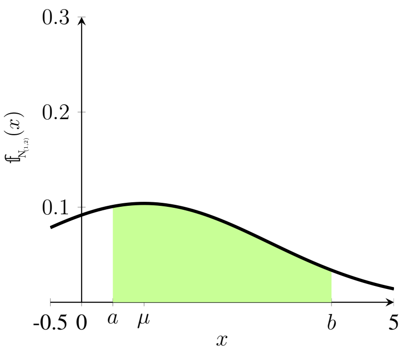
§10.12 **Beispiel**.

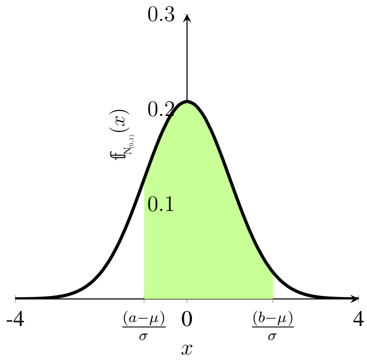
(a) Ist $Z \sim \mathrm{N}_{(0,1)}$, $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{\neq 0}$, so ist $X = \mu + \sigma Z$ eine $\mathrm{N}_{(\mu,\sigma^2)}$-verteilte Zufallsvariable. Ausgehend von $X \sim \mathrm{N}_{(\mu,\sigma^2)}$ erhalten wir die standardisierte Zufallsvariable

$$Z = (X - \mu)/\sigma \sim \mathrm{N}_{(0,1)}.$$

Für $b \in \mathbb{R}$ und $a \in \mathbb{R}_{<b}$ gilt (vgl. Beispiel §04.08 (c)):

$$\begin{aligned}\mathrm{N}_{(\mu,\sigma^2)}((a,b]) &= \mathbb{P}(X \in (a,b]) = \mathbb{P}(\tfrac{a-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} \leqslant \tfrac{b-\mu}{\sigma}) \\ &= \mathbb{P}(Z \in (\tfrac{a-\mu}{\sigma}, \tfrac{b-\mu}{\sigma}]) = \mathrm{N}_{(0,1)}((\tfrac{a-\mu}{\sigma}, \tfrac{b-\mu}{\sigma}]) = \Phi(\tfrac{b-\mu}{\sigma}) - \Phi(\tfrac{a-\mu}{\sigma})\end{aligned}$$





(b) Für $X \sim \mathrm{N}_{(0,1)}$ ist $Y = X^2$ eine χ^2_1-verteilte Zufallsvariable, wobei die *χ^2-Verteilung mit einem Freiheitsgrad* gegeben ist durch die Dichte $\mathbb{f}_{\chi^2_1}(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \mathbb{1}_{\mathbb{R}_{>0}}(y)$, $y \in \mathbb{R}$.

(c) Ist X aufgefasst als Spaltenvektor ein n-dimensionaler standard-normalverteilter Zufallsvektor wie in Beispiel §05.10, so ist $Y = \mu + AX$ mit $\mu \in \mathbb{R}^n$ und regulärem $A \in \mathbb{R}^{n\times n}$ ein n-dimensionaler zufälliger Spaltenvektor mit Dichte

$$\mathbb{f}_{\mathrm{N}_{(\mu,\Sigma)}}(y) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp(-\tfrac{1}{2}\langle \Sigma^{-1}(y-\mu), (y-\mu)\rangle) \quad \forall y \in \mathbb{R}^n,$$

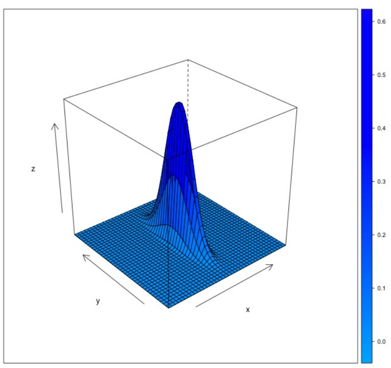
wobei $\Sigma = AA^{\mathrm{T}}$ eine symmetrische positive definite Matrix ist. Y heißt *normalverteilt* mit Vektor $\mu \in \mathbb{R}^n$ und Matrix $\Sigma \in \mathbb{R}^{n\times n}$, kurz $Y \sim \mathrm{N}_{(\mu,\Sigma)}$.

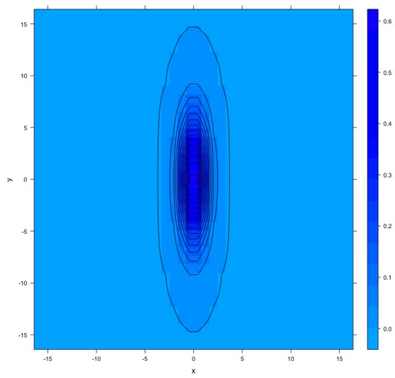
(d) Ein stetig-verteilter Zufallsvektor (X_1, X_2) wird *bivariat normalverteilt* mit Parametern $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$ und $\rho \in (-1,1)$, genannt, wenn für alle $x, y \in \mathbb{R}$ die gemeinsame Dichte durch

$$\mathbb{f}^{(X_1,X_2)}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big(-\frac{(x-\mu_1)^2}{2(1-\rho^2)\sigma_1^2}\Big) \exp\Big(\frac{2\rho(x-\mu_1)(y-\mu_2)}{2(1-\rho^2)\sigma_1\sigma_2}\Big) \exp\Big(-\frac{(y-\mu_2)^2}{2(1-\rho^2)\sigma_2^2}\Big)$$

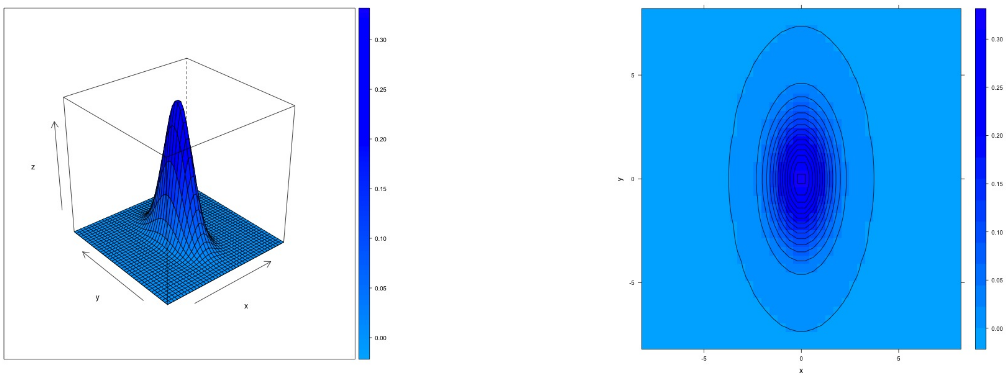
gegeben ist. Setzen wir $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ und $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, so entspricht dies gerade dem Fall $n = 2$ aus (c). Die nächsten Graphiken stellen die Dichte für verschiedene Werte der Parameter dar.

Für $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1$, $\sigma_2 = 4$ *und* $\rho = 0$:

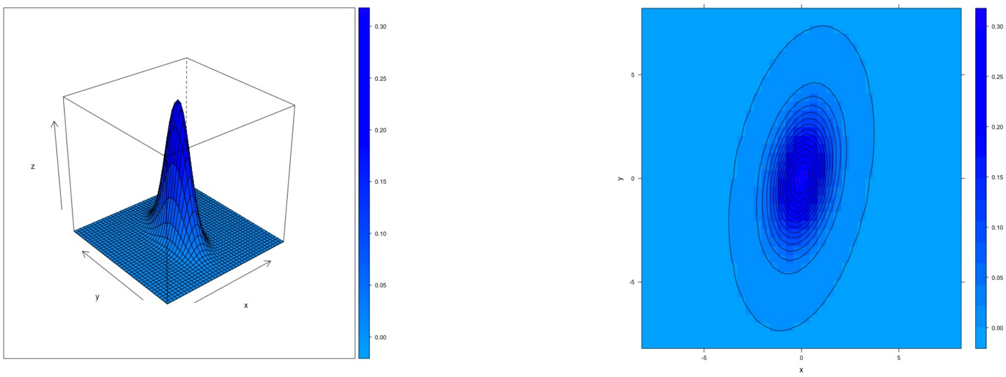




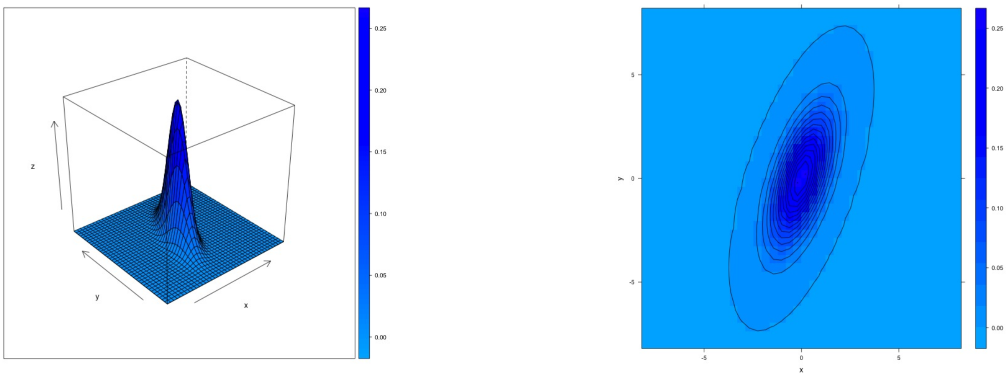
Für $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1$, $\sigma_2 = 2$ *und* $\rho = 0$:



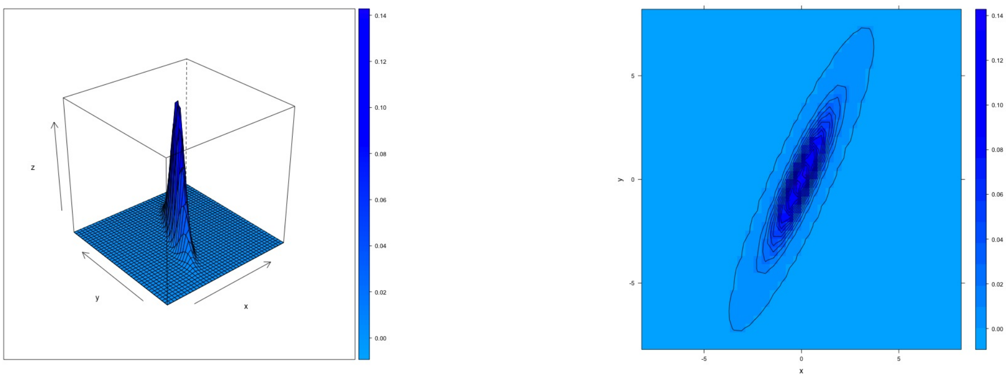
Für $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1$, $\sigma_2 = 2$ *und* $\rho = 0,3$:



Für $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1$, $\sigma_2 = 2$ *und* $\rho = 0,6$:



Für $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1$, $\sigma_2 = 2$ *und* $\rho = 0,9$:



$\square$

§11 Verteilung einer Familie von Zufallsvariablen

Im Folgenden seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{I}$ eine beliebige nicht-leere Indexmenge, $((\mathcal{X}_i, \mathscr{X}_i))_{i\in\mathcal{I}}$ eine Familie messbarer Räume und für jedes $i \in \mathcal{I}$, $X_i \in \mathcal{M}(\mathscr{A}, \mathscr{X}_i)$ eine $(\mathcal{X}_i, \mathscr{X}_i)$-wertige Zufallsvariable.

§11.01 **Definition**. Für eine Familie $(\mathscr{A}_i)_{i\in\mathcal{I}}$ von Teil-σ-Algebren aus $\mathscr{A}$ bezeichnet

$$\bigwedge_{i\in\mathcal{I}} \mathscr{A}_i := \bigcap_{i\in\mathcal{I}} \mathscr{A}_i$$

die *größte σ-Algebra*, die in allen $\mathscr{A}_i$, $i \in \mathcal{I}$, enthalten ist, und

$$\bigvee_{i\in\mathcal{I}} \mathscr{A}_i := \sigma(\bigcup_{i\in\mathcal{I}} \mathscr{A}_i)$$

die *kleinste σ-Algebra*, die alle $\mathscr{A}_i$, $i \in \mathcal{I}$ enthält. □

§11.02 **Erinnerung**. Die Menge

$$\boxed{\mathcal{X}_\mathcal{I}} := \times_{i\in\mathcal{I}} \mathcal{X}_i := \{(x_i)_{i\in\mathcal{I}} : (\forall i \in \mathcal{I} : x_i \in \mathcal{X}_i)\}$$
$$= \{(x : \mathcal{I} \to \cup_{i\in\mathcal{I}} \mathcal{X}_i) : (\forall i \in \mathcal{I} : x(i) \in \mathcal{X}_i)\}$$

heißt *Produktraum* oder kartesisches Produkt der Familie $(\mathcal{X}_i)_{i\in\mathcal{I}}$, wobei wir die Familie $(x_i)_{i\in\mathcal{I}}$ und die Abbildung $x : i \mapsto x_i$ identifizieren. Sind alle $\mathcal{X}_i$ gleich, etwa $\mathcal{X}_i = \mathcal{X}$, dann schreiben wir $\boxed{\mathcal{X}^\mathcal{I}} := \mathcal{X}_\mathcal{I}$, im Fall $n := |\mathcal{I}| \in \mathbb{N}$, auch nur kurz $\boxed{\mathcal{X}^n} := \mathcal{X}^\mathcal{I}$. □

§11.03 **Definition**. Für $\mathcal{J} \in 2^\mathcal{I}_{\neq\emptyset}$ wird die Abbildung

$$\Pi_\mathcal{J} : \mathcal{X}_\mathcal{I} \ni x = (x_i)_{i\in\mathcal{I}} \mapsto \Pi_\mathcal{J}(x) := (x_j)_{j\in\mathcal{J}} \in \mathcal{X}_\mathcal{J}$$

kanonische Projektion genannt. Für jedes $j \in \mathcal{I}$ bezeichne

$$\Pi_j := \Pi_{\{j\}} : \mathcal{X}_\mathcal{I} \ni x = (x_i)_{i\in\mathcal{I}} \mapsto \Pi_j(x) := x_j \in \mathcal{X}_j$$

die *Koordinantenabbildung*. Die Produkt-σ-Algebra $\boxed{\mathscr{X}_\mathcal{I}} := \bigotimes_{i\in\mathcal{I}} \mathscr{X}_i$ auf dem Produktraum $\mathcal{X}_\mathcal{I}$ ist die kleinste σ-Algebra, sodass für jedes $i \in \mathcal{I}$ die Koordinantenabbildung $\Pi_i \in \mathcal{M}(\mathscr{X}_\mathcal{I}, \mathscr{X}_i)$ $\mathscr{X}_\mathcal{I}$-$\mathscr{X}_i$-messbar ist, d.h.

$$\mathscr{X}_\mathcal{I} = \bigotimes_{i\in\mathcal{I}} \mathscr{X}_i := \bigvee_{i\in\mathcal{I}} \sigma(\Pi_i) = \bigvee_{i\in\mathcal{I}} \Pi_i^{-1}(\mathscr{X}_i).$$

Sind alle $(\mathcal{X}_i, \mathscr{X}_i)$ gleich, etwa $(\mathcal{X}_i, \mathscr{X}_i) = (\mathcal{X}, \mathscr{X})$, dann schreiben wir $\boxed{\mathscr{X}^\mathcal{I}} := \mathscr{X}_\mathcal{I}$, im Fall $n := |\mathcal{I}| \in \mathbb{N}$, auch nur kurz $\boxed{\mathscr{X}^n} := \mathscr{X}^\mathcal{I}$. Ist für jedes $i \in \mathcal{I}$ weiterhin $\mathbb{P}_i \in \mathcal{W}(\mathscr{X}_i)$ ein Wahrscheinlichkeitsmaß auf $(\mathcal{X}_i, \mathscr{X}_i)$, dann heißt ein Wahrscheinlichkeitsmaß $\boxed{\mathbb{P}_\mathcal{I}} := \bigotimes_{i\in\mathcal{I}} \mathbb{P}_i$ auf $\bigotimes_{i\in\mathcal{I}} \mathscr{X}_i$ *Produktmaß*, wenn für alle endlichen $\mathcal{J} \in 2^\mathcal{I}_{\neq\emptyset}$ und für alle $B_j \in \mathscr{X}_j$ für $j \in \mathcal{J}$ gilt:

$$\mathbb{P}_\mathcal{I}(\bigcap_{j\in\mathcal{J}} \Pi_j^{-1}(B_j)) = \prod_{j\in\mathcal{J}} \mathbb{P}_j(B_j).$$

Sind alle $(\mathcal{X}_i, \mathscr{X}_i, \mathbb{P}_i)$ gleich, etwa $(\mathcal{X}_i, \mathscr{X}_i, \mathbb{P}_i) = (\mathcal{X}, \mathscr{X}, \mathbb{P})$, dann schreiben wir kurz $\boxed{\mathbb{P}^\mathcal{I}} := \mathbb{P}_\mathcal{I}$ und im Fall $n := |\mathcal{I}| \in \mathbb{N}$ auch $\boxed{\mathbb{P}^n} := \mathbb{P}^\mathcal{I}$. □

§11.04 **Lemma**. *Die Abbildung $X_{\mathcal{I}} := (X_i)_{i\in\mathcal{I}} : \Omega \to \mathcal{X}_{\mathcal{I}}$ ist $\mathscr{A}$-$\mathscr{X}_{\mathcal{I}}$-messbar, also $X_{\mathcal{I}} \in \mathcal{M}(\mathscr{A}, \mathscr{X}_{\mathcal{I}})$ eine $(\mathcal{X}_{\mathcal{I}}, \mathscr{X}_{\mathcal{I}})$-wertige Zufallsvariable.*

§11.05 **Beweis** von Lemma §11.04. In der Vorlesung. □

§11.06 **Bemerkung**. Sei $\mathcal{I}$ abzählbar, für jedes $i \in \mathcal{I}$ sei $\mathcal{X}_i$ ein separabler und vollständiger metrischer Raum (polnisch) versehen mit der Borel-σ-Algebra $\mathscr{B}_i := \mathscr{B}_{\mathcal{X}_i}$ und sei $\mathscr{B}_{\mathcal{X}_{\mathcal{I}}}$ die Borel-σ-Algebra bzgl. der Produkttopologie auf $\mathcal{X}_{\mathcal{I}} = \times_{i\in\mathcal{I}} \mathcal{X}_i$. Dann gilt $\mathscr{B}_{\mathcal{X}_{\mathcal{I}}} = \mathscr{B}_{\mathcal{I}} = \bigotimes_{i\in\mathcal{I}} \mathscr{B}_i$, also insbesondere $\mathscr{B}_{\mathbb{R}^n} = \mathscr{B}^n$ (vgl. Klenke (2012) Satz 14.8). □

§11.07 **Satz**. *Sei $\mathcal{I}$ eine beliebige nicht-leerer Indexmenge. Für jedes $i \in \mathcal{I}$ sei $\mathcal{X}_i$ ein separabler und vollständiger metrischer Raum (polnisch) versehen mit der Borel-σ-Algebra $\mathscr{B}_i := \mathscr{B}_{\mathcal{X}_i}$ und $\mathbb{P}_i \in \mathcal{W}(\mathscr{X}_i)$ ein beliebiges Wahrscheinlichkeitsmaß. Dann existiert stets ein eindeutig bestimmtes Produktmaß $\mathbb{P}_{\mathcal{I}} = \bigotimes_{i\in\mathcal{I}} \mathbb{P}_i \in \mathcal{W}(\mathscr{B}_{\mathcal{I}})$ auf dem Produktraum $\mathcal{X}_{\mathcal{I}}$ versehen mit der Produkt-σ-Algebra $\mathscr{B}_{\mathcal{I}} = \bigotimes_{i\in\mathcal{I}} \mathscr{B}_i$.*

§11.08 **Beweis** von Satz §11.07. In der Vorlesung Wahrscheinlichkeitstheorie I. □

§11.09 **Definition**. Das Bildmaß $\boxed{\mathbb{P}^{X_{\mathcal{I}}}} := \mathbb{P} \circ X_{\mathcal{I}}^{-1}$ unter $X_{\mathcal{I}} := (X_i)_{i\in\mathcal{I}}$ auf $(\mathcal{X}_{\mathcal{I}}, \mathscr{X}_{\mathcal{I}})$ heißt *gemeinsame Verteilung* der Familie $(X_i)_{i\in\mathcal{I}}$. Für jedes $i \in \mathcal{I}$ wird das Bildmaß

$$\boxed{\mathbb{P}^{X_i}} := \mathbb{P} \circ X_i^{-1} = \mathbb{P} \circ (\Pi_i \circ X_{\mathcal{I}})^{-1} = \mathbb{P}^{X_{\mathcal{I}}} \circ \Pi_i^{-1}$$

Randverteilung (*marginale Verteilung*) von X_i bzgl. $\mathbb{P}^{X_{\mathcal{I}}}$ genannt. □

§11.10 **Satz**. *Sei $\mathcal{I} = [\![n]\!]$ und $X := (X_i)_{i\in[\![n]\!]}$.*

(i) *Sind alle $X_i \in \overline{\mathcal{M}}(\mathscr{A})$ numerische Zufallsvariablen mit gemeinsamer Verteilungsfunktion*

$$\mathbb{F}^X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X_1 \leqslant x_1, \dots, X_n \leqslant x_n) \quad \forall x = (x_i)_{i\in[\![n]\!]} \in \overline{\mathbb{R}}^n,$$

dann ist für jedes X_i die Randverteilungsfunktion

$$\mathbb{F}^{X_i}(x_i) = \mathbb{F}^X(\infty, \dots, \infty, x_i, \infty, \dots, \infty) \quad \forall x_i \in \overline{\mathbb{R}}.$$

(ii) *Ist $X \in \mathcal{M}(\mathscr{A}, 2^{\mathcal{X}_{[\![n]\!]}})$ eine diskret-verteilte Zufallsvariable mit gemeinsamer Zähldichte $\mathbb{p}^X : \times_{i\in[\![n]\!]} \mathcal{X}_i \to [0,1]$, dann ist jedes $X_i \in \mathcal{M}(\mathscr{A}, 2^{\mathcal{X}_i})$ diskret-verteilt mit Randzähldichte*

$$\mathbb{p}^{X_i}(x_i) = \sum_{x_1\in\mathcal{X}_1} \cdots \sum_{x_{i-1}\in\mathcal{X}_{i-1}} \sum_{x_{i+1}\in\mathcal{X}_{i+1}} \cdots \sum_{x_n\in\mathcal{X}_n} \mathbb{p}^X(x_1, \dots, x_n) \quad \forall x_i \in \mathcal{X}_i.$$

(iii) *Ist $X \in \mathcal{M}^n(\mathscr{A})$ ein stetig-verteilter Zufallsvektor mit gemeinsamer Dichte $\mathbb{f}^X : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$, dann ist jedes X_i stetig-verteilt mit Randdichte*

$$\mathbb{f}^{X_i}(x_i) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \mathbb{f}^X(x_1, \dots, x_n) dx_n \cdots dx_{i+1} dx_{i-1} \cdots dx_1 \quad \forall x_i \in \mathbb{R}.$$

§11.11 **Beweis** von Satz §11.10. In der Vorlesung. □

§11.12 **Beispiel.**

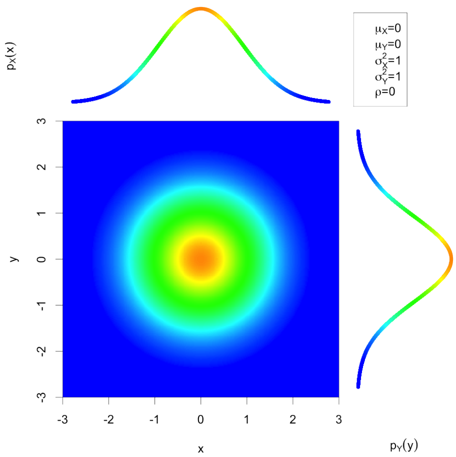
(a) Betrachten wir den Wurf von zwei fairen Würfeln (mit Laplaceverteilung). Dann sei X der Absolutbetrag der Differenz der Augenzahlen und Y die Summe der Augenzahlen. Dann ist $Z = (X, Y)$ diskret-verteilt mit gemeinsamer Zähldichte und Randzähldichten:

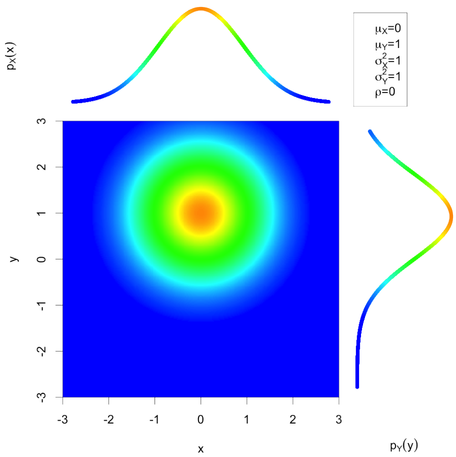
	$\mathbb{p}^Z(x,y)$	y: 2	3	4	5	6	7	8	9	10	11	12	$\mathbb{p}^X(x)$
	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{6}{36}$
	1	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{10}{36}$
x	2	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	$\frac{8}{36}$
	3	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	$\frac{6}{36}$
	4	0	0	0	0	$\frac{2}{36}$	0	$\frac{2}{36}$	0	0	0	0	$\frac{4}{36}$
	5	0	0	0	0	0	$\frac{2}{36}$	0	0	0	0	0	$\frac{2}{36}$
	$\mathbb{p}^Y(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

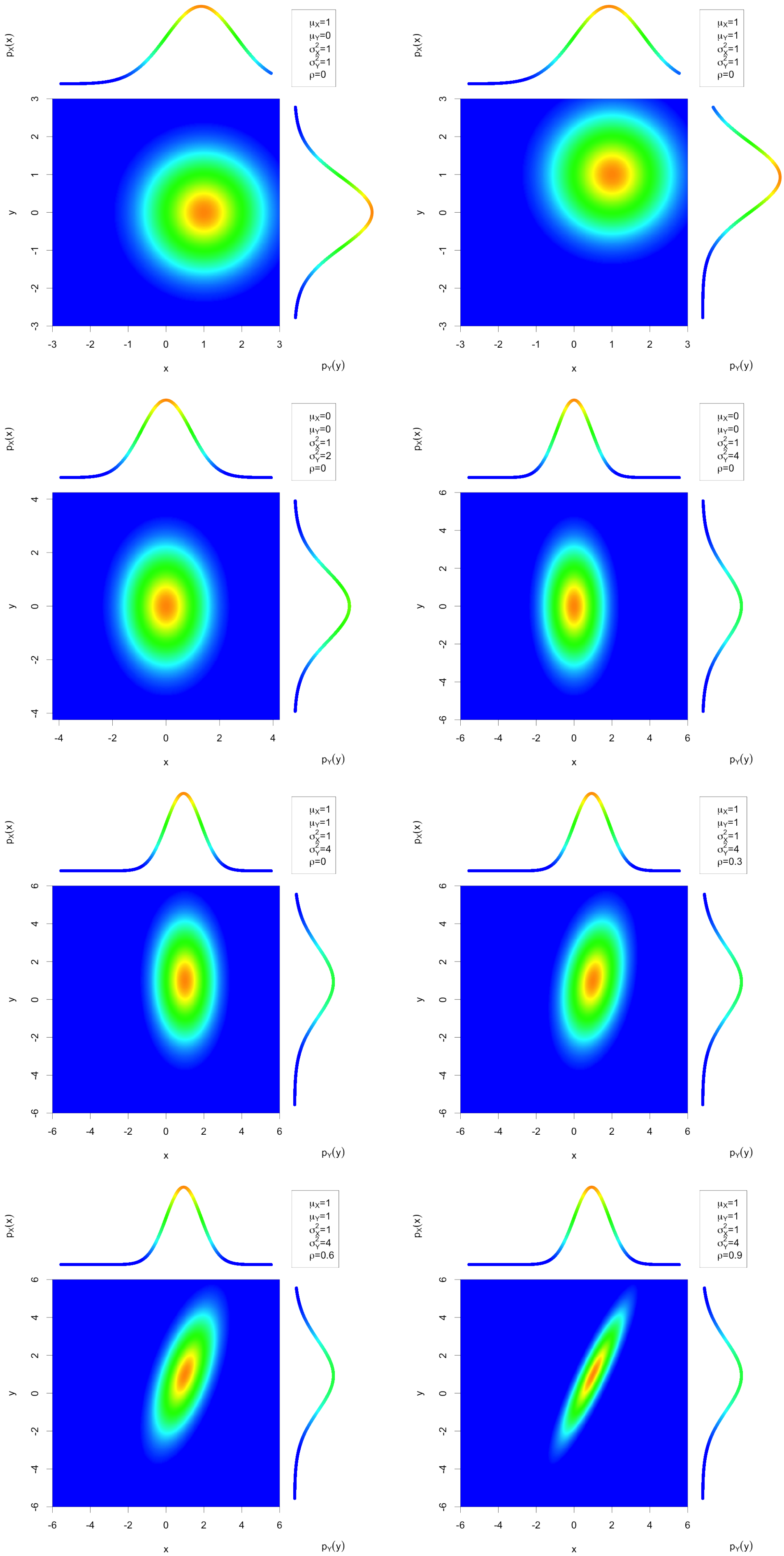
sowie gemeinsamer Verteilungsfunktion und Randverteilungsfunktionen:

	$\mathbb{F}^Z(x,y)$	y: $(-\infty,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,7)$	$[7,8)$	$[8,9)$	$[9,10)$	$[10,11)$	$[11,12)$	$[12,\infty)$	$\mathbb{F}^X(x)$
	$(-\infty,0)$	0	0	0	0	0	0	0	0	0	0	0	0	0
	$[0,1)$	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{6}{36}$
x	$[1,2)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{6}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{10}{36}$	$\frac{12}{36}$	$\frac{13}{36}$	$\frac{15}{36}$	$\frac{16}{36}$	$\frac{16}{36}$
	$[2,3)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{8}{36}$	$\frac{11}{36}$	$\frac{13}{36}$	$\frac{16}{36}$	$\frac{18}{36}$	$\frac{21}{36}$	$\frac{23}{36}$	$\frac{24}{36}$	$\frac{24}{36}$
	$[3,4)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{13}{36}$	$\frac{17}{36}$	$\frac{20}{36}$	$\frac{24}{36}$	$\frac{27}{36}$	$\frac{29}{36}$	$\frac{30}{36}$	$\frac{30}{36}$
	$[4,5)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{19}{36}$	$\frac{24}{36}$	$\frac{28}{36}$	$\frac{31}{36}$	$\frac{33}{36}$	$\frac{34}{36}$	$\frac{34}{36}$
	$[5,\infty)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$	$\frac{36}{36}$
	$\mathbb{F}^Y(y)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$	

(b) Ist (X, Y) bivariat normalverteilt (vgl. Beispiel §10.12 (c)) mit Parametern $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y \in \mathbb{R}_{>0}$ und $\rho \in (-1, 1)$ dann sind die Randverteilungen $X \sim \mathrm{N}_{(\mu_X,\sigma_X^2)}$ und $Y \sim \mathrm{N}_{(\mu_Y,\sigma_Y^2)}$. Die nächsten Graphiken stellen die gemeinsamen sowie die marginalen Dichten für verschiedene Werte der Parameter dar.






$\mu_X=1$
$\mu_Y=0$
$\sigma_X^2=1$
$\sigma_Y^2=1$
$\rho=0$
$\mu_X=1$
$\mu_Y=1$
$\sigma_X^2=1$
$\sigma_Y^2=1$
$\rho=0$
$\mu_X=0$
$\mu_Y=0$
$\sigma_X^2=1$
$\sigma_Y^2=2$
$\rho=0$
$\mu_X=0$
$\mu_Y=0$
$\sigma_X^2=1$
$\sigma_Y^2=4$
$\rho=0$
$\mu_X=1$
$\mu_Y=1$
$\sigma_X^2=1$
$\sigma_Y^2=4$
$\rho=0$
$\mu_X=1$
$\mu_Y=1$
$\sigma_X^2=1$
$\sigma_Y^2=4$
$\rho=0.3$
$\mu_X=1$
$\mu_Y=1$
$\sigma_X^2=1$
$\sigma_Y^2=4$
$\rho=0.6$
$\mu_X=1$
$\mu_Y=1$
$\sigma_X^2=1$
$\sigma_Y^2=4$
$\rho=0.9$
$p_X(x)$
$p_Y(y)$
x
y

(c) Für $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{>0}$ besitzt das Produktmaß $\mathrm{N}^n_{(\mu,\sigma^2)} \in \mathcal{W}(\mathscr{B}^n)$ auf $(\mathbb{R}^n, \mathscr{B}^n)$ die Dichte

$$\prod_{i\in[\![n]\!]} \mathbb{f}_{\mathrm{N}_{(\mu,\sigma^2)}}(x_i) = (2\pi\sigma^2)^{-n/2} \exp(-\tfrac{1}{2\sigma^2} \sum_{i\in[\![n]\!]} (x_i - \mu)^2), \quad x \in \mathbb{R}^n.$$

Andererseits, seien $\mathbb{1}_n := (1)_{i\in[\![n]\!]} \in \mathbb{R}^n$ der Einsvektor und $\mathrm{I}_{n\times n} \in \mathbb{R}^{n\times n}$ die Einheitsmatrix, sodass $\mu\mathbb{1}_n = (\mu)_{i\in[\![n]\!]} \in \mathbb{R}^n$ und $\sigma^2\mathrm{I}_{n\times n} = \mathrm{Diag}[\sigma^2]$ die Diagonalmatrix mit Diagonaleinträgen $(\sigma^2)_{i\in[\![n]\!]}$ ist. Die Normalverteilung $\mathrm{N}_{(\mu\mathbb{1}_n,\sigma^2\mathrm{I}_{n\times n})} \in \mathcal{W}(\mathscr{B}^n)$ auf $(\mathbb{R}^n, \mathscr{B}^n)$ besitzt dann die Dichte (vgl. Beispiel §10.12 (c))

$$\mathbb{f}_{\mathrm{N}_{(\mu\mathbb{1}_n,\sigma^2\mathrm{I}_{n\times n})}}(x) = (2\pi\sigma^2)^{-n/2} \exp(-\tfrac{1}{2\sigma^2} \sum_{i\in[\![n]\!]} (x_i - \mu)^2), \quad x \in \mathbb{R}^n,$$

sodass $\mathrm{N}_{(\mu\mathbb{1}_n,\sigma^2\mathrm{I}_{n\times n})} = \mathrm{N}^n_{(\mu,\sigma^2)}$, da sie dieselbe Dichte besitzen. □

§12 Statistische Inferenz

§12.01 **Definition**. Sei $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ ein statistisches Experiment. Ist X eine $(\mathcal{X}, \mathscr{X})$-wertige Zufallsvariable, so schreiben wir abkürzend $X \odot \mathbb{P}_\Theta$, wenn $X \sim \mathbb{P}_\theta$ für ein $\theta \in \Theta$ gilt. In diesem Fall heißt das statistische Experiment $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ *adäquat* für die Zufallsvariable X. Ist $(\mathcal{S}, \mathscr{S})$ ein messbarer Raum, so wird jede $(\mathcal{S}, \mathscr{S})$-wertige Zufallsvariable $S \in \mathcal{M}(\mathscr{X}, \mathscr{S})$ auf $(\mathcal{X}, \mathscr{X})$, also $\mathscr{X}$-$\mathscr{S}$-messbare Funktion $S : \mathcal{X} \to \mathcal{S}$, *Beobachtung* oder *Statistik* genannt. Wir bezeichnen mit $\mathbb{P}^S_\Theta := (\mathbb{P}^S_\theta)_{\theta\in\Theta}$ die durch S auf $(\mathcal{S}, \mathscr{S})$ induzierte Familie von Wahrscheinlichkeitsmaßen und mit $(\mathcal{S}, \mathscr{S}, \mathbb{P}^S_\Theta)$ das induzierte statistische Modell. Eine Abbildung $\gamma : \Theta \to \Gamma$ heißt *abgeleiteter* oder *interessierender Parameter* und für jedes $\theta \in \Theta$ wird $\gamma(\theta)$ *abgeleiteter* oder *interessierender Parameterwert* genannt. Ein abgeleiteter Parameter $\gamma : \Theta \to \Gamma$ heißt *identifizierbar*, wenn für beliebige $\theta_1, \theta_2 \in \Theta$ aus $\gamma(\theta_1) \neq \gamma(\theta_2)$ folgt $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$. □

§12.02 **Bemerkung**. Häufig benutzen wir das Symbol γ sowohl für den abgeleiteten Parameter, also die Abbildung $\Theta \to \Gamma$, als auch für die Elemente von Γ, also die Parameterwerte. □

§12.03 **Beispiel** *(Normalverteilungsmodell)*. Wir betrachten wie in Beispiel §11.12 (c) eine Familie von Normalverteilungen $(\mathrm{N}^n_{(\mu,\sigma^2)})_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_{>0}}$ auf $(\mathbb{R}^n, \mathscr{B}^n)$. Wir sagen ein $\mathbb{R}^n$-wertiger Spaltenvektor X ist *normalverteilt mit unbekannten Parametern μ und σ*, wenn $X \odot (\mathrm{N}^n_{(\mu,\sigma^2)})_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_{>0}}$. In dieser Situation sind typischerweise $\mathbb{R}\times\mathbb{R}_{>0} \ni (\mu,\sigma) \mapsto \mu \in \mathbb{R}$ mit oder $\mathbb{R}\times\mathbb{R}_{>0} \ni (\mu,\sigma) \mapsto \sigma \in \mathbb{R}_{>0}$ interessierende Parameter. Andererseits sagen wir, X ist *normalverteilt mit unbekanntem Parameter μ und bekanntem Parameter σ*, wenn $X \odot (\mathrm{N}^n_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$, oder auch *normalverteilt mit unbekanntem Parameter σ und bekanntem Parameter μ*, wenn $X \odot (\mathrm{N}^n_{(\mu,\sigma^2)})_{\sigma\in\mathbb{R}_{>0}}$. In jedem dieser Fälle ist das arithmetische Mittel $\overline{X}_n := \frac{1}{n}\sum_{i\in[\![n]\!]} X_i$ eine reelle Statistik, und für $X \sim \mathrm{N}^n_{(\mu,\sigma^2)}$ bezeichnen wir mit $\mathbb{P}^{\overline{X}_n}_{\mu,\sigma}$ die von $\overline{X}_n$ auf $(\mathbb{R}, \mathscr{B})$ induzierte Verteilung. □

Im Folgenden seien stets $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ ein statistisches Experiment, $(\Gamma, \mathscr{G})$ ein messbarer Raum und $\gamma : \Theta \to \Gamma$ ein identifizierbarer interessierender Parameter.

§12|01 Hypothesentest

§12.04 **Definition**. Sei $\{\Gamma^0, \Gamma^1\}$ eine Partition der Menge der interessierenden Parameterwerte Γ, also $\Gamma^0 \biguplus \Gamma^1 = \Gamma$. Für $X \odot \mathbb{P}_\Theta$ sagen wir, die *Nullhypothese* $H_0 : \Gamma^0$ ist erfüllt, wenn das statistische Experiment $(\mathcal{X}, \mathscr{X}, \mathbb{P}_{\gamma^{-1}(\Gamma^0)})$ für X adäquat ist, das heißt, für ein $\theta \in \Theta$ mit $\gamma(\theta) \in \Gamma^0$ gilt $X \sim \mathbb{P}_\theta$, andernfalls ist die *Alternative* $H_1 : \Gamma^1$ erfüllt. Ein Test wie in Definition §06.06 für das *Testproblem* der Nullhypothese $H_0 : \Gamma^0$ gegen die Alternative $H_1 : \Gamma^1$ ist somit eine $(\{0,1\}, 2^{\{0,1\}})$-wertige Zufallsvariable (Statistik). Im Fall $\Gamma = (a,b) \subseteq \overline{\mathbb{R}}$ mit

(a) $\Gamma^0_{\gamma_0} = (a, \gamma_0]$ und $\Gamma^1_{\gamma_0} = (\gamma_0, b)$ (bzw. $\Gamma^0_{\gamma_0} = [\gamma_0, b)$ und $\Gamma^1_{\gamma_0} = (a, \gamma_0)$) für ein $\gamma_0 \in \Gamma$ wird das Testproblem *einseitig* genannt und wir schreiben es auch in der Form: Nullhypothese $H_0 : \gamma \leqslant \gamma_0$ gegen Alternative $H_1 : \gamma > \gamma_0$ (bzw. $H_0 : \gamma \geqslant \gamma_0$ gegen $H_1 : \gamma < \gamma_0$).

(b) $\Gamma^0_{\gamma_0} = \{\gamma_0\}$ und $\Gamma^1_{\gamma_0} = (a, \gamma_0) \cup (\gamma_0, b)$ für ein $\gamma_0 \in \Gamma$ wird das Testproblem *zweiseitig* genannt und wir schreiben es auch in der Form: Nullhypothese $H_0 : \gamma = \gamma_0$ gegen Alternative $H_1 : \gamma \neq \gamma_0$. □

§12.05 **Sprechweise**. Wir folgen der Konvention, dass $\mathcal{A} = \{\varphi = 1\} = \varphi^{-1}(\{1\})$ der Ablehnbereich eines Tests φ ist, also $\varphi = \mathbb{1}_{\mathcal{A}}$ ist und die *Nullhypothese abgelehnt* wird, wenn φ den Wert eins annimmt. Die Messbarkeit eines Tests garantiert dabei die Messbarkeit des assoziierten Ablehnbereiches. □

§12.06 **Beispiel** *(Normalverteilungsmodell §12.03 fortgesetzt)*. Sei $X \odot (\mathrm{N}^n_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$, also ein normalverteilter $\mathbb{R}^n$-wertiger Spaltenvektor mit unbekanntem Parameter μ und bekanntem Parameter σ. Für $\mu_0 \in \mathbb{R}$ betrachten wir das einseitige Testproblem der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$. □

§12.07 **Erinnerung**. Durch Angabe des Ablehnbereiches $\mathcal{A}$ ist ein Test $\varphi = \mathbb{1}_{\mathcal{A}}$ eindeutig festgelegt. Offensichtlich können in dieser Situation nur zwei Fehlentscheidungen auftreten, die Nullhypothese $H_0 : \Gamma^0$ wird abgelehnt, also der Ablehnbereich $\mathcal{A}$ tritt ein, obwohl das statistisches Experiment $(\mathcal{X}, \mathscr{X}, \mathbb{P}_{\gamma^{-1}(\Gamma^0)})$ adäquat ist, oder die Nullhypothese wird nicht gegen die Alternative $H_1 : \Gamma^1$ abgelehnt, also der Annahmebereich $\mathcal{A}^c$ tritt ein, obwohl $(\mathcal{X}, \mathscr{X}, \mathbb{P}_{\gamma^{-1}(\Gamma^1)})$ adäquat ist. □

§12.08 **Definition**. Sei $\{\Gamma^0, \Gamma^1\}$ eine Partition der Menge der interessierenden Parameterwerte Γ, und $\varphi = \mathbb{1}_{\mathcal{A}}$ ein Test der Nullhypothese $H_0 : \Gamma^0$ gegen die Alternative $H_1 : \Gamma^0$. Dann bezeichnet

Fehler 1. Art: die Wahrscheinlichkeit $\mathbb{P}^{\varphi}_{\theta}(\{1\}) = \mathbb{P}_{\theta}(\varphi = 1) = \mathbb{P}_{\theta}(\mathcal{A})$ die Nullhypothese abzulehnen, sich also für die Alternative zu entscheiden, obwohl $\mathbb{P}_{\theta}$ mit $\gamma(\theta) \in \Gamma^0$ vorliegt, also die Nullhypothese adäquat ist;

Fehler 2. Art: die Wahrscheinlichkeit $\mathbb{P}^{\varphi}_{\theta}(\{0\}) = \mathbb{P}_{\theta}(\varphi = 0) = \mathbb{P}_{\theta}(\mathcal{A}^c)$ die Nullhypothese nicht abzulehnen, sich also für Nullhypothese zu entscheiden, obwohl $\mathbb{P}_{\theta}$ mit $\gamma(\theta) \in \Gamma^1$ vorliegt, also die Alternative adäquat ist. □

Ein Test φ hält das (*Signifikanz-*) *Niveau* $\alpha \in [0, 1]$ ein, wenn für jedes $\theta \in \gamma^{-1}(\Gamma^0)$ der Fehler 1. Art $\mathbb{P}_{\theta}(\varphi = 1) \leqslant \alpha$ erfüllt. In diesem Fall wird φ kurz *α-Test* genannt. Ein Test φ heißt *gleichmäßig bester Test zum Niveau $\alpha \in [0, 1]$*, falls er das Niveau $\alpha \in [0, 1]$ einhält und für jedes $\theta \in \gamma^{-1}(\Gamma^1)$ der Fehler 2. Art $\mathbb{P}_{\theta}(\widetilde{\varphi} = 0)$ eines jeden anderen α-Tests $\widetilde{\varphi}$ nicht kleiner ist, dass heißt, $\mathbb{P}_{\theta}(\varphi = 0) \leqslant \mathbb{P}_{\theta}(\widetilde{\varphi} = 0)$ gilt. □

§12.09 **Beispiel** *(Binomialverteilungsmodell §06.19 fortgesetzt)*. Betrachte für beliebige $\mathrm{p}_0 \in (0, 1)$ und $n \in \mathbb{N}$ erneut das *einseitigen Testproblem* der Nullhypothese $H_0 : \mathrm{p} \leqslant \mathrm{p}_0$ gegen die Alternative $H_1 : \mathrm{p} > \mathrm{p}_0$ im Binomialverteilungsmodell $\big([\![0, n]\!], 2^{[\![0,n]\!]}, (\mathrm{Bin}_{(n,\mathrm{p})})_{\mathrm{p}\in(0,1)}\big)$, so ist $[\![x^*, n]\!]$ mit $x^* \in [\![0, n]\!]$ der Ablehnbereich eines Neyman-Pearson Tests $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,n]\!]}$, der für das Testproblem der *einfachen Nullhypothese $H_0 : \mathrm{p} = \mathrm{p}_0$* gegen die *zusammengesetzte Alternative $H_1 : \mathrm{p} > \mathrm{p}_0$* ein gleichmäßig bester Test zum Niveau $\alpha = \mathrm{Bin}_{(n,\mathrm{p}_0)}([\![x^*, n]\!])$ ist (vgl. Ausblick §06.20). Wir zeigen in der Vorlesung Statistik I, dass für alle $\mathrm{p} < \mathrm{p}_0$ auch gilt $\mathrm{Bin}_{(n,\mathrm{p})}([\![x^*, n]\!]) \leqslant \alpha$. Somit ist ein Neyman-Pearson-Test $\varphi^\star_{x^*} = \mathbb{1}_{[\![x^*,n]\!]}$ mit Ablehnbereich $[\![x^*, n]\!]$ auch ein *gleichmäßig bester Test* zum Niveau $\alpha = \mathrm{Bin}_{(n,\mathrm{p}_0)}([\![x^*, n]\!])$ des *einseitigen Testproblems*. Im Beispiel I im Prolog §1 kann also der firmeneigene Verbraucherservice die Nullhypothese, der Anteil der Konsumentinnen beträgt höchstens 50%, gegen die Alternative, der Anteil der Konsumentinnen beträgt mehr als 50%, zum Niveau $\alpha = 0.01$ ablehnen. □

§12.10 **Definition**. Sei $(\mathbb{R}^n, \mathscr{B}^n, (\mathbb{P}_0, \mathbb{P}_1))$ ein (binäres) stetiges statistisches Experiment mit entsprechenden Wahrscheinlichkeitsdichten $\mathfrak{f}_0$ und $\mathfrak{f}_1$. Jeder Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$ mit Ablehnbereich der Form

$$\mathcal{A}_k := \{x \in \mathbb{R}^n : \mathfrak{f}_1(x) > k\mathfrak{f}_0(x)\} = \{\mathfrak{f}_1 > k\mathfrak{f}_0\}$$

für einen *kritischen Wert* $k \in \mathbb{R}_{\geqslant 0}$ heißt *Neyman-Pearson-Test*. □

§12.11 **Bemerkung**. Da $\mathfrak{f}_0, \mathfrak{f}_1 \in \mathcal{M}(\mathscr{B}^n)$ zwei reelle Zufallsvariablen sind, gilt $\mathcal{A}_k = \{\mathfrak{f}_1 > k\mathfrak{f}_0\} \in \mathscr{B}^n$ nach Lemma §08.06 (iii) und $\mathcal{A}_k$ ist somit ein Ablehnbereich eines Tests. Mit anderen Worten ein *Neyman-Pearson-Test* ist in der Tat ein Test. □

§12.12 **Beispiel** *(Normalverteilungsmodell §12.06 fortgesetzt)*. Für $\sigma \in \mathbb{R}_{>0}$, $\mu_0 \in \mathbb{R}$ und $\mu_1 \in \mathbb{R}_{>\mu_0}$ betrachte im binären stetigen statistischen Experiment $(\mathbb{R}^n, \mathscr{B}^n, (\mathrm{N}^n_{(\mu,\sigma^2)})_{\mu \in \{\mu_0,\mu_1\}})$ zunächst das Testproblem der einfachen Nullhypothese $H_0 : \{\mathrm{N}^n_{(\mu_0,\sigma^2)}\}$ gegen die einfache Alternative $H_1 : \{\mathrm{N}^n_{(\mu_1,\sigma^2)}\}$. Setzen wir $\overline{x_n} := \frac{1}{n}\sum_{i \in [\![n]\!]} x_i$ für $x \in \mathbb{R}^n$, so ist ein Neyman-Pearson-Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$ für einen kritischem Wert $k \in \mathbb{R}_{\geqslant 0}$ gegeben durch den Ablehnbereich

$$\begin{aligned}\mathcal{A}_k &= \{x \in \mathbb{R}^n : \mathfrak{f}_{\mathrm{N}^n_{(\mu_1,\sigma^2)}}(x) > k\mathfrak{f}_{\mathrm{N}^n_{(\mu_0,\sigma^2)}}(x)\} \\ &= \{x \in \mathbb{R}^n : \exp\left(\tfrac{1}{2\sigma^2}\textstyle\sum_{i\in[\![n]\!]}(x_i-\mu_0)^2 - \tfrac{1}{2\sigma^2}\sum_{i\in[\![n]\!]}(x_i-\mu_1)^2\right) > k\} \\ &= \{x \in \mathbb{R}^n : \exp\left(\tfrac{n(\mu_1-\mu_0)}{\sigma^2}\overline{x_n} + \tfrac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) > k\}.\end{aligned}$$

Für $\mu_1 \in \mathbb{R}_{>\mu_0}$ ist die Funktion $\mathrm{L}_{\mu_1,\mu_0}(\overline{x_n}) := \exp\left(\frac{n(\mu_1-\mu_0)}{\sigma^2}\overline{x_n} + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right)$, $\overline{x_n} \in \mathbb{R}$, streng monoton wachsend in $\overline{x_n}$. Damit gibt es zu jedem kritischen Wert $k \in \mathbb{R}_{\geqslant 0}$ einen Wert $c^\star \in \mathbb{R}$ derart, dass

$$\mathcal{A}_k = \{x \in \mathbb{R}^n : \overline{x_n} > c^\star\} = \{\overline{X_n} > c^\star\} = \{\overline{X_n} \in \mathbb{R}_{>c^\star}\}$$

für die reelle Statistik $\overline{X_n} : \mathbb{R}^n \ni x \mapsto \overline{X_n}(x) := \overline{x_n} \in \mathbb{R}$ gilt. Wir halten fest, dass in einem Normalverteilungsmodell ein Neyman-Pearson-Test $\varphi^\star_{c^\star} = \mathbb{1}_{\{\overline{X_n} > c^\star\}} = \mathbb{1}_{\mathbb{R}_{>c^\star}}(\overline{X}_n)$ durch einen Ablehnbereich der Form $\{\varphi^\star_{c^\star} = 1\} = \{\overline{X_n} > c^\star\} = \{\overline{X_n} \in \mathbb{R}_{>c^\star}\}$ für ein $c^\star \in \mathbb{R}$ gegeben ist. Insbesondere, gilt also $\mathrm{N}^n_{(\mu_0,\sigma^2)}(\varphi^\star_{c^\star} = 1) = \mathbb{P}^{\overline{X_n}}_{(\mu_0,\sigma^2)}((c^\star, \infty)) = \mathbb{P}^{\overline{X_n}}_{(\mu_0,\sigma^2)}(\mathbb{R}_{>c^\star})$. □

§12.13 **Neyman-Pearson Lemma**. *Sei $(\mathbb{R}^n, \mathscr{B}^n, (\mathbb{P}_0, \mathbb{P}_1))$ ein (binäres) stetiges statistisches Experiment. Für das Testproblem der einfachen Nullhypothese $H_0 : \{\mathbb{P}_0\}$ gegen die einfache Alternative $H_1 : \{\mathbb{P}_1\}$ ist jeder Neyman-Pearson-Test $\varphi^\star_k = \mathbb{1}_{\mathcal{A}_k}$ mit kritischem Wert $k \in \mathbb{R}_{\geqslant 0}$ und Ablehnbereich $\mathcal{A}_k \in \mathscr{B}^n$ wie in Definition §12.10 ein bester Test zum Niveau $\mathbb{P}_0(\mathcal{A}_k) \in [0,1]$.*

§12.14 **Beweis** von Satz §12.13. Übungsaufgabe. □

§12.15 **Beispiel** *(Normalverteilungsmodell §12.12 fortgesetzt)*. Betrachten wir für $\sigma \in \mathbb{R}_{>0}$ (und $n = 1$) im stetigen statistischen Experiment $X \odot (\mathrm{N}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$ mit beliebigem $\mu_0 \in \mathbb{R}$ das *einseitige Testproblem* der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$, so hängt der Ablehnbereich $\{\varphi^\star_{c^\star} = 1\} = \{X \geqslant c^\star\} = \mathbb{R}_{\geqslant c^\star}$ mit $c^\star \in \mathbb{R}$ eines Neyman-Pearson-Test $\varphi^\star_{c^\star} = \mathbb{1}_{\{X > c^\star\}} = \mathbb{1}_{\mathbb{R}_{>c^\star}}$ nicht von der Alternative ab (vgl. Beispiel §12.12), da die Verteilungsfamilie $(\mathrm{N}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$ einen *monotonen Likelihood-Quotienten* besitzt. Damit ist der Neyman-Pearson-Test $\varphi^\star_{c^\star}$ für das Testproblem der *einfachen Nullhypothese* $H_0 : \mu = \mu_0$ gegen die *zusammengesetzte Alternative* $H_1 : \mu > \mu_0$ ein gleichmäßig bester Test zum Niveau

$$\mathrm{N}_{(\mu_0,\sigma^2)}(\varphi^\star_{c^\star} = 1) = \mathrm{N}_{(\mu_0,\sigma^2)}(\mathbb{R}_{>c^\star}) = \mathrm{N}_{(0,1)}\left(\left(\tfrac{c^\star-\mu_0}{\sigma}, \infty\right)\right) = 1 - \Phi\left(\tfrac{c^\star-\mu_0}{\sigma}\right) = \Phi\left(\tfrac{-c^\star+\mu_0}{\sigma}\right)$$

(vgl. Beispiel §10.12 (a)). Dies erlaubt uns für jedes $\alpha \in (0,1)$ einen kritischen Wert $c_\alpha \in \mathbb{R}$ zu bestimmen, derart dass $\alpha = \Phi\left(\frac{-c_\alpha+\mu_0}{\sigma}\right)$ gilt. In der Tat bezeichnet z_α das α-Quantil einer

Standardnormalverteilung (vgl. Beispiel §05.07 (c)), also $\alpha = \Phi(z_\alpha)$, so ist $c_\alpha = \mu_0 - \sigma z_\alpha$ der gesuchte kritische Wert. Weiterhin gilt $N_{(\mu,\sigma^2)}((c_\alpha,\infty)) = N_{(\mu,\sigma^2)}(\mathbb{R}_{>c_\alpha}) < \alpha$ für alle $\mu \in \mathbb{R}_{<\mu_0}$ (vgl. Beispiel §05.07 (c)). Somit ist $\varphi^\star_{c_\alpha} = \mathbb{1}_{\mathbb{R}_{>c_\alpha}} = \mathbb{1}_{\{X>\mu_0-\sigma z_\alpha\}}$ für jedes $\alpha \in (0,1)$ ein Neyman-Pearson-Test mit Ablehnbereich $\mathbb{R}_{>c_\alpha} = \{X > \mu_0 - \sigma z_\alpha\}$, der ein *gleichmäßig bester Test* zum Niveau α des *einseitigen Testproblems* der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$ ist. □

§12|02 Schätzfunktion

§12.16 **Definition.** Jede Statistik $\widehat{\gamma} : (\mathcal{X}, \mathscr{X}) \to (\Gamma, \mathscr{G})$, also $\mathscr{X}$-$\mathscr{G}$-messbare Abbildung $\widehat{\gamma} \in \mathcal{M}(\mathscr{X},\mathscr{G})$, heißt *Schätzfunktion*, kurz *Schätzer*, für den abgeleiteten Parameter γ. Für eine Stichprobe $x \in \mathcal{X}$ wird $\widehat{\gamma}(x)$ der zugehörige *Schätzwert* genannt. □

§12.17 **Beispiel.** In einer Tombola enthält eine Urne $N \in \mathbb{N}$ Lose (nummeriert von 1 bis N). Um die Gewinnchance abzuschätzen, möchte die Spielerin die Anzahl der Lose schätzen, dazu kauft sie sich ein Los. Wir nehmen an, dass die zufällige Losnummer X adäquat durch ein Laplaceverteilungsmodell $(\mathbb{N}, 2^{\mathbb{N}}, (\mathrm{Lap}_{[\![N]\!]})_{N\in\mathbb{N}})$ mit zugehöriger Familie von Zähldichten $(\mathbb{p}_N)_{N\in\mathbb{N}}$ und dem messbaren Parameterraum $(\mathbb{N}, 2^{\mathbb{N}})$ beschrieben wird. Eine plausible Schätzmethode für die Gesamtanzahl N, bei Vorliegen einer Stichprobe $x \in \mathbb{N}$ als Schätzwert $\widehat{N}(x)$ denjenigen Parameter $N \in \mathbb{N}$ zu wählen, für den die Wahrscheinlichkeit $\mathbb{p}_N(x)$ des Eintreten von x maximiert wird, d.h.

$$\widehat{N}(x) \in \underset{N\in\mathbb{N}}{\arg\max}\{\mathbb{p}_N(x)\} := \{N \in \mathbb{N} : \mathbb{p}_N(x) \geqslant \mathbb{p}_{\widetilde{N}}(x)\ \forall\, \widetilde{N} \in \mathbb{N}\}$$

$$= \underset{N\in\mathbb{N}}{\arg\max}\{\tfrac{1}{N}\mathbb{1}_{[\![1,N]\!]}(x)\} = \{x\}.$$

Die Statistik $\widehat{N} : \mathbb{N} \to \mathbb{N}$ mit $x \mapsto \widehat{N}(x) := x$ wird *Maximum-Likelihood-Schätzer* (MLS) genannt. □

§12.18 **Erinnerung.** Ist $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ ein diskretes bzw. stetiges statistisches Experiment mit zugehöriger Familie von Zähldichten bzw. Wahrscheinlichkeitsdichten $\mathbb{p}_\Theta := (\mathbb{p}_\theta)_{\theta\in\Theta}$, so ist für jedes $\theta \in \Theta$ die Funktion $\mathbb{p}_\theta : (\mathcal{X}, \mathscr{X}) \to (\mathbb{R}_{\geqslant 0}, \mathscr{B}_{\geqslant 0})$ messbar, sodass wir $\mathbb{p}_\theta$ im Folgenden als (positive) reelle Statistik auffassen, also $\mathbb{p}_\theta \in \mathcal{M}_{\geqslant 0}(\mathscr{X})$. □

§12.19 **Definition.** Sei $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ ein diskretes bzw. stetiges statistisches Experiment mit zugehöriger Familie von Zähldichten bzw. Wahrscheinlichkeitsdichten $(\mathbb{p}_\theta)_{\theta\in\Theta}$. Für jedes $\theta \in \Theta$ betrachten wir die reelle Statistik

$$\mathrm{L}(\theta) := \mathrm{L}(\theta, \bullet) : \mathcal{X} \to \mathbb{R}_{\geqslant 0} \text{ mit } x \mapsto [\mathrm{L}(\theta)](x) := \mathrm{L}(\theta, x) := \mathbb{p}_\theta(x),$$

also $\mathrm{L}(\theta) \in \mathcal{M}_{\geqslant 0}(\mathscr{X})$. Die (zufällige) Funktion

$$\mathrm{L} : \Theta \to \mathcal{M}_{\geqslant 0}(\mathscr{X}) \text{ mit } \theta \mapsto \mathrm{L}(\theta)$$

wird *Likelihood-Funktion* genannt. □

§12.20 **Sprechweise.** In der Maßtheorie werden *dominierte* Familie von Wahrscheinlichkeitsmaßen eingeführt, die als Spezialfälle die in Definition §12.19 betrachteten diskreten bzw. stetigen Familien von Wahrscheinlichkeitsmaßen umfassen. Wir bezeichnen im Folgenden daher die in Definition §12.19 betrachteten diskreten bzw. stetigen statistischen Experimente kurz als *dominierte statistische Experimente mit Likelihood-Funktion* L. □

§12.21 **Definition**. Seien $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ ein dominiertes statistisches Experiment mit Likelihood-Funktion L und $(\Theta, \mathscr{T})$ ein messbarer Raum. Eine Statistik $\widehat{\theta} \in \mathcal{M}(\mathscr{X}, \mathscr{T})$ auf $(\mathcal{X}, \mathscr{X})$ mit Werten in $(\Theta, \mathscr{T})$ wird *Maximum-Likelihood-Schätzer (MLS)* für θ genannt, wenn $\mathrm{L}(\widehat{\theta}) = \sup_{\theta\in\Theta} \mathrm{L}(\theta)$ gilt. Ist weiterhin $\gamma(\widehat{\theta}) \in \mathcal{M}(\mathscr{X}, \mathscr{G})$ eine Statistik auf $(\mathcal{X}, \mathscr{X})$ mit Werten in $(\Gamma, \mathscr{G})$, so wird $\gamma(\widehat{\theta})$ *Maximum-Likelihood-Schätzer (MLS)* für den abgeleiteten Parameter $\gamma : \Theta \to \Gamma$ genannt. □

§12.22 **Bemerkung**. $\mathrm{L}(\widehat{\theta}) = \sup_{\theta\in\Theta} \mathrm{L}(\theta)$ meint $\mathrm{L}(\widehat{\theta}(x), x) = \sup_{\theta\in\Theta} \mathrm{L}(\theta, x)$ für alle $x \in \mathcal{X}$. Weiterhin sind weder Existenz noch Eindeutigkeit eines MLS ohne Weiteres garantiert. Bei Mehrdeutigkeit wird üblicherweise ein maximierender Parameter θ nach Belieben gewählt. □

§12.23 **Beispiel**.

(a) *Binomialverteilungsmodell*, $\big([\![0, n]\!], 2^{[\![0,n]\!]}, (\mathrm{Bin}_{(n,\mathrm{p})})_{\mathrm{p}\in[0,1]}\big)$ mit Zähldichten $(\mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p})}})_{\mathrm{p}\in[0,1]}$:

$$\underset{\mathrm{p}\in[0,1]}{\arg\max}\,\{\mathbb{p}_{\mathrm{Bin}_{(n,\mathrm{p})}}(x)\} = \underset{\mathrm{p}\in[0,1]}{\arg\max}\,\{\binom{n}{x}\mathrm{p}^x(1-\mathrm{p})^{n-x}\} = \{\frac{x}{n}\}$$

Die Statistik $\widehat{\mathrm{p}} : [\![0, n]\!] \to [0, 1]$ mit $x \mapsto \widehat{\mathrm{p}}(x) = \frac{x}{n}$ ist der MLS für p.

(b) *Poissonverteilungsmodell*, $\big(\mathbb{Z}^n_{\geqslant 0}, 2^{\mathbb{Z}^n_{\geqslant 0}}, (\mathrm{Poi}^n_\lambda)_{\lambda\in\mathbb{R}_{>0}}\big)$ mit Produktzähldichten $(\mathbb{p}^n_{\mathrm{Poi}_\lambda})_{\lambda\in\mathbb{R}_{>0}}$:

$$\underset{\lambda\in\mathbb{R}_{>0}}{\arg\max}\{\mathbb{p}^n_{\mathrm{Poi}_\lambda}(x)\} = \underset{\lambda\in\mathbb{R}_{>0}}{\arg\max}\,\{\frac{\lambda^{\sum_{i\in[\![n]\!]} x_i}}{\prod_{i\in[\![n]\!]}(x_i!)}e^{-n\lambda}\} = \{\tfrac{1}{n}\sum_{i\in[\![n]\!]} x_i\}.$$

Die Statistik $\widehat{\lambda} : \mathbb{Z}^n_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ mit $x \mapsto \widehat{\lambda}(x) = \frac{1}{n}\sum_{i\in[\![n]\!]} x_i =: \overline{x_n}$ ist der MLS für λ.

(c) *Uniformverteilungsmodell*, $\big(\mathbb{R}^n_{\geqslant 0}, \mathscr{B}^n_{\geqslant 0}, (\mathrm{U}^n_{[0,\theta]})_{\theta\in\mathbb{R}_{>0}}\big)$ mit Produktdichten $(\mathbb{f}^n_{\mathrm{U}_{[0,\theta]}})_{\theta\in\mathbb{R}_{>0}}$:

$$\underset{\theta\in\mathbb{R}_{>0}}{\arg\max}\,\{\mathbb{f}^n_{\mathrm{U}_{[0,\theta]}}(x)\} = \underset{\theta\in\mathbb{R}_{>0}}{\arg\max}\,\{\tfrac{1}{\theta^n}\prod_{i\in[\![n]\!]}\mathbb{1}_{[0,\theta]}(x_i)\} = \{\max_{i\in[\![n]\!]} x_i\}.$$

Die Statistik $\widehat{\theta} : \mathbb{R}^n_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ mit $x \mapsto \widehat{\theta}(x) = \max\{x_i : i \in [\![n]\!]\}$ ist der MLS für θ.

(d) *Exponentialverteilungsmodell*, $\big(\mathbb{R}^n_{\geqslant 0}, \mathscr{B}^n_{\geqslant 0}, (\mathrm{Exp}^n_\lambda)_{\lambda\in\mathbb{R}_{>0}}\big)$ mit Produktdichten $(\mathbb{f}^n_{\mathrm{Exp}_\lambda})_{\lambda\in\mathbb{R}_{>0}}$:

$$\underset{\lambda\in\mathbb{R}_{>0}}{\arg\max}\,\{\mathbb{f}^n_{\mathrm{Exp}_\lambda}(x)\} = \underset{\lambda\in\mathbb{R}_{>0}}{\arg\max}\,\{\lambda^n \exp(-\lambda\sum_{i\in[\![n]\!]} x_i)\} = \{\frac{1}{\overline{x_n}}\}.$$

Die Statistik $\widehat{\lambda} : \mathbb{R}^n_{\geqslant 0} \to \overline{\mathbb{R}}_{\geqslant 0}$ mit $x \mapsto \widehat{\lambda}(x) = (\overline{x_n})^{-1}$ ist der MLS für λ.

(e) *Normalverteilungsmodell* $\big(\mathbb{R}^n, \mathscr{B}^n, (\mathrm{N}^n_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}\big)$, $\sigma \in \mathbb{R}_{>0}$, mit Produktdichten $(\mathbb{f}^n_{\mathrm{N}_{(\mu,\sigma^2)}})_{\mu\in\mathbb{R}}$:

$$\underset{\mu\in\mathbb{R}}{\arg\max}\,\{\mathbb{f}^n_{\mathrm{N}_{(\mu,\sigma^2)}}(x)\} = \underset{\mu\in\mathbb{R}}{\arg\max}\,\{\tfrac{1}{(2\pi\sigma^2)^{n/2}}\exp\big(\tfrac{-1}{2\sigma^2}\sum_{i\in[\![n]\!]}(x_i-\mu)^2\big)\} = \{\overline{x_n}\}.$$

Die Statistik $\widehat{\mu} : \mathbb{R}^n \to \mathbb{R}$ mit $x \mapsto \widehat{\mu}(x) = \overline{x_n}$ ist der MLS für μ. □

§12.24 **Bemerkung**. Im Binomialverteilungsmodell $X \odot (\mathrm{Bin}_{(n,\mathrm{p})})_{\mathrm{p}\in[0,1]}$ nimmt der MLS $\widehat{\mathrm{p}} = X/n$ für p (vgl. Beispiel §12.23 (a)) nur Werte in $\widehat{\mathrm{p}}([\![0, n]\!]) = \{\frac{i}{n} : i \in [\![0, n]\!]\}$ an. Für jeden Parameter $\mathrm{p} \in [0, 1]\backslash\widehat{\mathrm{p}}([\![0, n]\!])$ gilt somit $\mathrm{Bin}_{(n,\mathrm{p})}(\widehat{\mathrm{p}} = \mathrm{p}) = 0$. Analog, für $\sigma \in \mathbb{R}_{>0}$ und $n = 1$ im Normalverteilungsmodell $X \odot (\mathrm{N}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$ ist $\widehat{\mu} = X$ der MLS für μ (vgl. Beispiel §12.23 (e)), sodass für jeden Parameterwert $\mu \in \mathbb{R}$ gilt $\mathrm{N}_{(\mu,\sigma^2)}(\widehat{\mu} = \mu) = \mathrm{N}_{(\mu,\sigma^2)}(\{\mu\}) = 0$. Im Allgemeinen können wir nicht davon ausgehen, dass der Schätzwert des MLS gleich dem wahren Wert ist. Andererseits im Normalverteilungsmodell, gilt

$$\{\mu \in [\widehat{\mu} - c, \widehat{\mu} + c]\} = \{x \in \mathbb{R} : \mu \in [\widehat{\mu}(x) - c, \widehat{\mu}(x) + c]\} = \{\widehat{\mu} \in [\mu - c, \mu + c]\}$$

für jedes $c \in \mathbb{R}_{>0}$, sodass

$$\mathrm{N}_{(\mu,\sigma^2)}(\mu \in [\widehat{\mu} - c, \widehat{\mu} + c]) = \mathrm{N}_{(\mu,\sigma^2)}(\widehat{\mu} \in [\mu - c, \mu + c]) = \mathrm{N}_{(\mu,\sigma^2)}([\mu - c, \mu + c]) \in \mathbb{R}_{>0}.$$

Für eine Stichprobe x ist es somit sinnvoller einen Bereich $[\widehat{\mu}(x) - c, \widehat{\mu}(x) + c]$ anzugeben als nur den Schätzwert $\widehat{\mu}(x)$. □

§12|03 Konfidenzbereich

§12.25 **Definition**. Eine Abbildung $\mathrm{B} : \mathcal{X} \to 2^{\Gamma}$ heißt *Bereichsschätzfunktion (BSF)* für Γ, wenn $\{\gamma \in \mathrm{B}\} := \{x \in \mathcal{X} : \gamma \in \mathrm{B}(x)\} \in \mathscr{X}$ für alle Parameterwerte $\gamma \in \Gamma$ gilt. Für jedes $\gamma \in \Gamma$ und $\theta \in \Theta$ wird $\mathbb{P}_\theta(\gamma \in \mathrm{B})$ *Überdeckungswahrscheinlichkeit* von γ genannt. □

§12.26 **Bemerkung**. Für jedes $\gamma \in \Gamma$ sei $\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\}$ eine Partition der Menge der interessierenden Parameterwerte Γ, also $\Gamma^{\mathrm{r}}_\gamma \biguplus \Gamma^{\mathrm{f}}_\gamma = \Gamma$. Für jedes $\theta \in \Theta$ fassen wir $\Gamma^{\mathrm{r}}_{\gamma(\theta)}$ und $\Gamma^{\mathrm{f}}_{\gamma(\theta)}$ als Menge der „richtigen“ bzw. der „falschen“ abgeleiteten Parameterwerte auf. Wir suchen eine BSF B, die für alle $\theta \in \Theta$ eine möglichst große Überdeckungswahrscheinlichkeit $\mathbb{P}_\theta(\widetilde{\gamma} \in \mathrm{B})$ für jeden richtigen Parameterwert $\widetilde{\gamma} \in \Gamma^{\mathrm{r}}_{\gamma(\theta)}$ besitzt, und gleichzeitig aber eine möglichst präzise Überdeckung besitzt, das heißt, für jeden falschen Parameterwert $\widetilde{\gamma} \in \Gamma^{\mathrm{f}}_{\gamma(\theta)}$ soll die Überdeckungswahrscheinlichkeit $\mathbb{P}_\theta(\widetilde{\gamma} \in \mathrm{B})$ möglichst klein ist. Wie im Fall des Hypothesentest ist aber ein gleichzeitiges Maximieren und Minimieren dieser Überdeckungswahrscheinlichkeit im Allgemeinen nicht möglich. □

§12.27 **Beispiel**. Im Fall $\Gamma \subseteq \mathbb{R}$ sind insbesondere von Interesse

(a) $\Gamma^{\mathrm{r}}_\gamma = \{\gamma\}$ und $\Gamma^{\mathrm{f}}_\gamma = \{\gamma\}^c = \Gamma_{\neq\gamma} = \Gamma \setminus \{\gamma\}$;

(b) $\Gamma^{\mathrm{r}}_\gamma = \Gamma_{\leqslant\gamma} = \Gamma \cap (-\infty, \gamma]$ und $\Gamma^{\mathrm{f}}_\gamma = \Gamma_{>\gamma} = \Gamma^c_{\leqslant\gamma} = \Gamma \cap (-\infty, \gamma]^c = \Gamma \cap (\gamma, \infty)$;

(c) $\Gamma^{\mathrm{r}}_\gamma = \Gamma_{\geqslant\gamma} = \Gamma \cap [\gamma, \infty)$ und $\Gamma^{\mathrm{f}}_\gamma = \Gamma_{<\gamma} = \Gamma^c_{\geqslant\gamma} = \Gamma \cap [\gamma, \infty)^c = \Gamma \cap (-\infty, \gamma)$. □

§12.28 **Definition**. Sei $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$ eine Familie von Partitionen der Menge der interessierenden Parameterwerte Γ. Für ein $\alpha \in (0,1)$ heißt eine Bereichsschätzfunktion B *Konfidenzbereich zum Niveau $1-\alpha$ für $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$*, wenn $\mathbb{P}_\theta(\widetilde{\gamma} \in \mathrm{B}) \geqslant 1 - \alpha$ für alle $\widetilde{\gamma} \in \Gamma^{\mathrm{r}}_{\gamma(\theta)}$ und für alle $\theta \in \Theta$ gilt. In diesem Fall wird B kurz *(1-α)-Konfidenzbereich* (für $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$) genannt. Eine Bereichsschätzfunktion B^* für Γ heißt *gleichmäßig bester Konfidenzbereich zum Niveau $1-\alpha$ für $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$*, wenn sie ein (1-$\alpha$)-Konfidenzbereich ist, und jeder andere (1-α)-Konfidenzbereich B (für $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$) keine kleinere Überdeckungswahrscheinlichkeit der falschen abgeleiteten Parameterwerte besitzt, das heißt, für alle $\theta \in \Theta$ und $\widetilde{\gamma} \in \Gamma^{\mathrm{f}}_{\gamma(\theta)}$ gilt $\mathbb{P}_\theta(\widetilde{\gamma} \in \mathrm{B}^*) \leqslant \mathbb{P}_\theta(\widetilde{\gamma} \in \mathrm{B})$. □

§12.29 **Beispiel** *(Normalverteilungsmodell §12.15 fortgesetzt)*. Betrachte für $\sigma \in \mathbb{R}_{>0}$ das stetige statistische Experiment $X \odot (\mathrm{N}_{(\theta,\sigma^2)})_{\theta\in\mathbb{R}}$, wobei die Menge der interessierenden Parameter $\Gamma = \Theta = \mathbb{R}$ ist. Für jedes $c \in \mathbb{R}_{>0}$ ist $\mathrm{B}_c : \mathbb{R} \to 2^{\mathbb{R}}$ mit $\mathrm{B}_c(x) := (x \pm c) = (x - c, x + c)$, kurz $\mathrm{B}_c = (X \pm c)$, eine BSF, da für jedes $\widetilde{\theta} \in \mathbb{R}$ gilt

$$\{\widetilde{\theta} \in \mathrm{B}_c\} = \{x \in \mathbb{R} : \widetilde{\theta} \in (x \pm c)\} = \{x \in \mathbb{R} : x \in (\widetilde{\theta} \pm c)\} = (\widetilde{\theta} \pm c) \in \mathscr{B}.$$

Für alle $\theta \in \mathbb{R}$ ist die Überdeckungswahrscheinlichkeit von $\widetilde{\theta} \in \mathbb{R}$ (vgl. Beispiel §10.12 (a))

$$\mathrm{N}_{(\theta,\sigma^2)}(\widetilde{\theta} \in \mathrm{B}_c) = \mathrm{N}_{(\theta,\sigma^2)}((\widetilde{\theta} \pm c)) = \Phi(\tfrac{\widetilde{\theta}-\theta+c}{\sigma}) - \Phi(\tfrac{\widetilde{\theta}-\theta-c}{\sigma}).$$

Interessieren wir uns nur für den wahren Parameter $\Gamma^{\mathrm{r}}_\theta = \{\theta\}$ und dementsprechend der Menge der falschen Parameterwerte $\Gamma^{\mathrm{f}}_\theta = \mathbb{R}_{\neq\theta}$. Dann ist B_c ein (1-α)-Konfidenzbereich für $(\{\{\theta\}, \mathbb{R}_{\neq\theta}\})_{\theta\in\mathbb{R}}$, wenn für alle $\theta \in \mathbb{R}$ gilt

$$\mathrm{N}_{(\theta,\sigma^2)}(\theta \in \mathrm{B}_c) = \Phi(\tfrac{c}{\sigma}) - \Phi(\tfrac{-c}{\sigma}) \geqslant 1 - \alpha.$$

Wählen wir $\mathrm{c}_\alpha := \sigma \mathrm{z}_{1-\alpha/2}$ und $\mathrm{B}_{\mathrm{c}_\alpha} = (X \pm \sigma \mathrm{z}_{1-\alpha/2})$ mit 1-$\alpha/2$-Quantil $\mathrm{z}_{1-\alpha/2}$ der Standardnormalverteilung, das heißt $1 - \alpha/2 = \Phi(\mathrm{z}_{1-\alpha/2}) = 1 - \Phi(-\mathrm{z}_{1-\alpha/2})$, so gilt

$$\mathrm{N}_{(\theta,\sigma^2)}(\theta \in \mathrm{B}_{\mathrm{c}_\alpha}) = \Phi(\mathrm{z}_{1-\alpha/2}) - \Phi(-\mathrm{z}_{1-\alpha/2}) = 1 - \alpha/2 - (1 - (1 - \alpha/2)) = 1 - \alpha$$

für alle $\theta \in \mathbb{R}$. Damit ist $(X \pm \sigma \mathrm{z}_{1-\alpha/2})$ ein (1-α)-Konfidenzbereich für den wahren Parameter θ und die falschen Parameter $\mathbb{R}_{\neq\theta}$, $\theta \in \mathbb{R}$. □

§12.30 **Definition**.

(a) Ist $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$ eine Familie von Partitionen der Menge der interessierenden Parameterwerte Γ, so definiert $\Gamma^0_\gamma := \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma^{\mathrm{r}}_{\widetilde{\gamma}}\}$ und $\Gamma^1_\gamma := \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma^{\mathrm{f}}_{\widetilde{\gamma}}\}$ für $\gamma \in \Gamma$ eine *assoziierte Familie* $(\{\Gamma^0_\gamma, \Gamma^1_\gamma\})_{\gamma\in\Gamma}$ *von Partitionen* der Parametermenge Γ, dabei fassen wir Γ^0_γ als Nullhypothese und Γ^1_γ als Alternative eines Testproblems auf.

(b) Für eine Bereichsschätzfunktion B definiert $(\varphi_\gamma = \mathbb{1}_{\{\gamma\notin \mathrm{B}\}})_{\gamma\in\Gamma}$ eine *assoziierte Familie von Tests* mit Ablehnbereich $\{\varphi_\gamma = 1\} =: \mathcal{A}_\gamma$, da definitionsgemäß $\mathcal{A}_\gamma = \{\gamma \notin \mathrm{B}\} \in \mathscr{X}$ für jeden abgeleiteten Parameterwert $\gamma \in \Gamma$ gilt.

(c) Für eine Familie $(\varphi_\gamma = \mathbb{1}_{\mathcal{A}_\gamma})_{\gamma\in\Gamma}$ von Tests mit Ablehnbereich $\mathcal{A}_\gamma = \{\varphi_\gamma = 1\} \in \mathscr{F}$ definiert $\mathrm{B} : \mathcal{X} \to 2^\Gamma$ mit $\mathrm{B}(x) := \{\gamma \in \Gamma : x \notin \mathcal{A}_\gamma\}$ eine *assoziierte Bereichsschätzfunktion*, da in der Tat $\{\gamma \in \mathrm{B}\} = \{x \in \mathcal{X} : \gamma \in \mathrm{B}(x)\} = \{x \in \mathcal{X} : x \notin \mathcal{A}_\gamma\} = \mathcal{A}^c_\gamma \in \mathscr{X}$ für jeden abgeleiteten Parameterwert $\gamma \in \Gamma$ gilt. □

§12.31 **Beispiel** *(Beispiel §12.27 fortgesetzt)*. Im Folgenden geben wir zu typischen Familien $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$ von Partitionen der Menge der interessierenden Parameterwerte $\Gamma \subseteq \mathbb{R}$ die assoziierte Familie $(\{\Gamma^0_\gamma, \Gamma^1_\gamma\})_{\gamma\in\Gamma}$ von Partitionen der interessierenden Parametermenge $\Gamma \subseteq \mathbb{R}$ an.

(a) Für $\Gamma^{\mathrm{r}}_\gamma = \{\gamma\}$ und $\Gamma^{\mathrm{f}}_\gamma = \Gamma_{\neq\gamma}$ sind $\Gamma^0_\gamma = \{\gamma\}$ und $\Gamma^1_\gamma = \Gamma_{\neq\gamma}$;

(b) Für $\Gamma^{\mathrm{r}}_\gamma = \Gamma_{\leqslant\gamma}$ und $\Gamma^{\mathrm{f}}_\gamma = \Gamma_{>\gamma}$ sind

$$\Gamma^0_\gamma = \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma^{\mathrm{r}}_{\widetilde{\gamma}}\} = \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma_{\leqslant\widetilde{\gamma}}\} = \Gamma_{\geqslant\gamma}$$

und $\Gamma^1_\gamma = \Gamma_{<\gamma}$;

(c) Für $\Gamma^{\mathrm{r}}_\gamma = \Gamma_{\geqslant\gamma}$ und $\Gamma^{\mathrm{f}}_\gamma = \Gamma_{<\gamma}$ sind $\Gamma^0_\gamma = \Gamma_{\leqslant\gamma}$ und $\Gamma^1_\gamma = \Gamma_{>\gamma}$. □

§12.32 **Beispiel** *(Normalverteilungsmodell §12.15 fortgesetzt)*. Betrachten wir für $\sigma \in \mathbb{R}_{>0}$ im stetigen statistischen Experiment $X \odot (\mathrm{N}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$ mit beliebigem $\mu_0 \in \mathbb{R}$ das *einseitige Testproblem* der Nullhypothese $H_0 : \mu \leqslant \mu_0$, also $\Gamma^0_{\mu_0} = \mathbb{R}_{\leqslant\mu_0}$, gegen die Alternative $H_1 : \mu > \mu_0$, also $\Gamma^1_{\mu_0} = \mathbb{R}_{>\mu_0}$, so ist für jedes $\alpha \in (0, 1)$ ein Neyman-Pearson-Test $\varphi_{\mu_0} = \mathbb{1}_{\{X\geqslant\mu_0-\sigma z_\alpha\}}$ mit Ablehnbereich $\{\varphi_{\mu_0} = 1\} = \{X \geqslant \mu_0 - \sigma z_\alpha\}$ ein *gleichmäßig bester Test* zum Niveau α. Dann ist $\mathrm{B} : \mathbb{R} \to 2^{\mathbb{R}}$ mit

$$x \mapsto \mathrm{B}(x) = \{\mu_0 \in \mathbb{R} : x \in \{\varphi_{\mu_0} = 0\}\} = \{\mu_0 \in \mathbb{R} : x < \mu_0 - \sigma z_\alpha\} = (x + \sigma z_\alpha, \infty) = \mathbb{R}_{>x+\sigma z_\alpha}$$

die zu der Familie $(\varphi_{\mu_0})_{\mu_0\in\mathbb{R}}$ *assoziierte Bereichsschätzfunktion*. □

§12.33 **Satz.** *Seien* $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ *eine Familie von Partitionen in richtige und falsche interessierende Parameterwerte und* $(\{\Gamma_\gamma^0, \Gamma_\gamma^1\})_{\gamma\in\Gamma}$ *die assoziierte Familie von Partitionen in Nullhypothen und Alternativen. Dann gilt für eine Familie* $(\varphi_\gamma)_{\gamma\in\Gamma}$ *von Tests, dass genau dann* φ_γ *ein (gleichmäßig bester)* α*-Test der Nullhypothese* $H_0 : \Gamma_\gamma^0$ *gegen die Alternative* $H_1 : \Gamma_\gamma^1$ *für jedes* $\gamma \in \Gamma$ *ist, wenn die assoziierte Bereichsschätzfunktion* B *für* $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ *ein (gleichmäßig bester) (1-*α*)-Konfidenzbereich ist.*

§12.34 **Beweis** von Satz §12.33. In der Vorlesung. □

§12.35 **Beispiel** *(Normalverteilungsmodell §12.32 fortgesetzt).* Sei $X \odot (\mathrm{N}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R}}$ mit $\sigma \in \mathbb{R}_{>0}$. Dann ist $(X + \sigma z_\alpha, \infty)$ der assoziierte Konfidenzbereich zu einer Familie gleichmäßig bester α-Tests (vgl. Beispiel §12.32). Nach Satz §12.33 ist $(X + \sigma z_\alpha, \infty)$ damit auch ein gleichmäßig bester 1-α-Konfidenzbereich für die Mengen der richtigen Parameter $\Gamma_{\mu_0}^{\mathrm{r}} = \mathbb{R}_{\geqslant\mu_0}$ und der falschen Parameter $\Gamma_{\mu_0}^{\mathrm{f}} = \mathbb{R}_{<\mu_0}$ mit $\mu_0 \in \mathbb{R}$. Andererseits, ist $(X \pm \sigma z_{1-\alpha/2})$ ein (1-α)-Konfidenzbereich für die Menge der richtigen Parameter $\Gamma_{\mu_0}^{\mathrm{r}} = \{\mu_0\}$ und der falschen Parameter $\Gamma_{\mu_0}^{\mathrm{f}} = \mathbb{R}_{\neq\mu_0}$ mit $\mu_0 \in \mathbb{R}$ (vgl. Beispiel §12.29). Nach Satz §12.33 ist damit der assoziierte Test φ_{μ_0} mit Ablehnbereich

$$\{\varphi_{\mu_0} = 1\} = \{\mu_0 \notin (X \pm \sigma z_{1-\alpha/2}) = \{|X - \mu_0| \geqslant \sigma z_{1-\alpha/2}\}$$

ein α-Test für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$. □

Kapitel 4

Bedingte Wahrscheinlichkeit und Unabhängigkeit

§13 Bedingte Wahrscheinlichkeit und Bayes-Formel

Im Folgenden sei $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Weiterhin bezeichnen wir eine Partition $\mathcal{P}$ von Ω als ($\mathscr{A}$-)messbar, wenn $\mathcal{P} \subseteq \mathscr{A}$ gilt.

§13.01 **Definition**. Es seien A und B Ereignisse aus $\mathscr{A}$ mit $\mathbb{P}(B) \in \mathbb{R}_{>0}$. Dann wird mit

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

die *bedingte Wahrscheinlichkeit* von A gegeben (oder unter) B bezeichnet. □

§13.02 **Satz**. *Sei $B \in \mathscr{A}$ mit $\mathbb{P}(B) \in \mathbb{R}_{>0}$ und $\mathcal{I}$ eine abzählbare Indexmenge. Dann gilt:*

(i) $\widetilde{\mathbb{P}} : \mathscr{A} \ni A \mapsto \widetilde{\mathbb{P}}(A) := \mathbb{P}(A|B) \in [0,1]$ *ist ein Wahrscheinlichkeitsmaß auf* $(\Omega, \mathscr{A})$ *und die Einschränkung* $\widetilde{\mathbb{P}}\big|_B : \mathscr{A}\big|_B \ni A \mapsto \widetilde{\mathbb{P}}\big|_B(A) := \widetilde{\mathbb{P}}(A) \in [0,1]$ *ein Wahrscheinlichkeitsmaß auf der Einschränkung* $(B, \mathscr{A}\big|_B)$. *Der Wahrscheinlichkeitsraum* $(B, \mathscr{A}\big|_B, \widetilde{\mathbb{P}}\big|_B)$ *wird die* Spur *oder* Einschränkung *von* $(\Omega, \mathscr{A}, \widetilde{\mathbb{P}})$ *auf* B *genannt.*

(ii) *Es sei* $\{B_i : i \in \mathcal{I}\}$ *eine messbare Partition von* B *mit* $\mathbb{P}(B_i) \in \mathbb{R}_{>0}$ *für alle* $i \in \mathcal{I}$. *Für alle* $A \in \mathscr{A}$ *gilt dann die* *Formel von der totalen Wahrscheinlichkeit:*

$$\mathbb{P}(A \cap B) = \sum_{i \in \mathcal{I}} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

(iii) *Für jedes* $A \in \mathscr{A}$ *mit* $\mathbb{P}(A) \in \mathbb{R}_{>0}$ *und jede messbare Partition* $\{B_i : i \in \mathcal{I}\}$ *von* Ω *mit* $\mathbb{P}(B_i) \in \mathbb{R}_{>0}$ *für alle* $i \in \mathcal{I}$ *gilt die* *Formel von Bayes:*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i)\mathbb{P}(A|B_i)}{\sum_{j \in \mathcal{I}} \mathbb{P}(A|B_j)\mathbb{P}(B_j)} \quad \forall i \in \mathcal{I}.$$

§13.03 **Beweis** von Satz §13.02. (i) Übungsaufgabe und (ii)-(iii) in der Vorlesung. □

§13.04 **Beispiel**. Am Bahnhof Südkreuz in Berlin ist eine Videoüberwachung installiert. Zur Evaluation der Gesichtserkennungssoftware haben sich $0,1\%$ der täglichen Passagiere registrieren lassen. Der Hersteller der Gesichtserkennungssoftware gibt an, dass $99,9\%$ der registrierten Personen als registriert erkannt werden, aber auch $0,2\%$ der nicht-registrierten Personen fälschlich als registriert erkannt werden. Die Wahrscheinlichkeit, dass ein zufällig ausgewählter Passagier, der als registriert erkannt wird, nicht auf der Liste der Registrierten ist, beträgt dann nach der Formel von Bayes: $\frac{0{,}002 \bullet 0{,}999}{0{,}001 \bullet 0{,}999 + 0{,}999 \bullet 0{,}002} = 2/3$, d.h. über 66%. □

§13.05 **Lemma**. *Für* $A_i \in \mathscr{A}$, $i \in [\![n+1]\!]$ *mit* $\mathbb{P}(\bigcap_{i \in [\![n]\!]} A_i) \in \mathbb{R}_{>0}$ *gilt*

$$\mathbb{P}(\bigcap_{i \in [\![n+1]\!]} A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_{n+1}|\bigcap_{i \in [\![n]\!]} A_i).$$

§13.06 **Beweis** von Lemma §13.05. In der Vorlesung. □

§13.07 **Beispiel**. Beim Ziehen aus einer Urne mit W weißen und S schwarzen Kugeln ohne Zurücklegen und mit Beachtung der Reihenfolge ergibt sich mit $N = W+S$ für das Ergebnis „SSW“, also 1. und 2. Kugel schwarz, 3. Kugel weiß, nach der Pfadregel die Wahrscheinlichkeit $\frac{S}{N}\bullet\frac{S-1}{N-1}\bullet\frac{W}{N-2}$. Dieselbe Wahrscheinlichkeit besitzen die Versuchsausgänge „SWS“ und „WSS“ (man nennt die Verteilung *austauschbar*). □

§13.08 **Beispiel**. In einem Unternehmen werden von $825/560/325$ männlichen Bewerbern in den Abteilungen $A/B/C$ jeweils $62\%/63\%/34\%$ eingestellt, von $108/25/593$ weiblichen Bewerberinnen hingegen $82\%/68\%/37\%$. Obwohl die Einstellungsquote in jeder Abteilung für Frauen höher war, ergibt sich nach der Formel von der totalen Wahrscheinlichkeit insgesamt eine Einstellungsquote von ca. 57% für Männer und von ca. 45% für Frauen, weil letztere sich stärker für Abteilung C mit schwieriger Einstellung beworben haben. □

§14 Unabhängige Ereignisse

Im Folgenden seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathcal{I}$ eine beliebige nicht-leere Indexmenge.

§14.01 **Definition**.

(a) Zwei Ereignisse $A, B \in \mathscr{A}$ heißen *(stochastisch) unabhängig* (unter $\mathbb{P}$), kurz $A \perp\!\!\!\perp B$, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ gilt.

(b) Eine Familie $(A_i)_{i\in\mathcal{I}}$ von Ereignissen aus $\mathscr{A}$ heißt *(stochastisch) unabhängig* (unter $\mathbb{P}$), kurz $\perp\!\!\!\perp_{i\in\mathcal{I}} A_i$, wenn für jede endliche Teilmenge $\mathcal{J} \in 2^{\mathcal{I}}_{\neq\emptyset}$ gilt $\mathbb{P}(\bigcap_{j\in\mathcal{J}} A_j) = \prod_{j\in\mathcal{J}} \mathbb{P}(A_j)$. □

§14.02 **Bemerkung**.

(a) Für jedes $A \in \mathscr{A}$ und jede $\mathbb{P}$-Nullmenge $B \in \mathscr{A}$ (bzw. $\mathbb{P}$-Einsmenge) gilt $A \perp\!\!\!\perp B$. Im Fall $\mathbb{P}(B) \in \mathbb{R}_{>0}$ sind die Aussagen $A \perp\!\!\!\perp B$ und $\mathbb{P}(A|B) = \mathbb{P}(A)$ äquivalent.

(b) Es gilt $\perp\!\!\!\perp_{i\in\mathcal{I}} A_i$ genau dann, wenn $\perp\!\!\!\perp_{j\in\mathcal{J}} A_j$ für jede endliche Teilmenge $\mathcal{J} \in 2^{\mathcal{I}}_{\neq\emptyset}$ gilt.

(c) Für $|\mathcal{I}| \in \overline{\mathbb{N}}_{\geqslant 3}$ ist eine Familie $(A_i)_{i\in\mathcal{I}}$ unabhängig, so sind die Ereignisse paarweise unabhängig, d.h. $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ für alle $i \in \mathcal{I}$ und $j \in \mathcal{I}_{\backslash i}$, aber die Umkehrung gilt nicht. □

§14.03 **Beispiel**. Beim Würfeln mit zwei Würfeln haben die Ereignisse „Ist die Augensumme 7?“ und „Ist die erste Augenzahl 6?“ jeweils Wahrscheinlichkeit $1/6$ unter einer Gleichverteilung. Der Schnitt der beiden Ereignisse ist „Ist die erste Augenzahl 6 und die zweite Augenzahl 1?“ und hat die Wahrscheinlichkeit $1/36$, so dass die beiden Ereignisse (unter Gleichverteilung) unabhängig sind. □

§14.04 **Lemma**. *Seien $(A_i)_{i\in\mathcal{I}}$ unabhängige Ereignisse aus $\mathscr{A}$. Für $A_i^\star \in \sigma(\{A_i\}) = \{\emptyset, A_i, A_i^c, \Omega\}$, $i \in \mathcal{I}$, ist die Familie $(A_i^\star)_{i\in\mathcal{I}}$ von Ereignissen aus $\mathscr{A}$ unabhängig.*

§14.05 **Beweis** von Lemma §14.04. In der Vorlesung. □

§14.06 **Satz von Borel-Cantelli**. *Für eine Folge $(A_n)_{n\in\mathbb{N}}$ von Ereignissen aus $\mathscr{A}$ gilt:*

(i) *Aus $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) \in \mathbb{R}_{\geqslant 0}$ folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$;*

(ii) *Gilt zusätzlich $\perp\!\!\!\perp_{n\in\mathbb{N}} A_n$, so folgt aus $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ auch $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.*

§14.07 **Beweis** von Satz §14.06. In der Vorlesung. □

§14.08 **Beispiel.**

(a) Ist A ein Ereignis mit $\mathbb{P}(A) \in (0,1)$, so gilt $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ für $A_n := A$, jedoch $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}(A) < 1$. Damit kann also auf die Unabhängigkeit in Teil (ii) von Satz §14.06 nicht verzichtet werden.

(b) Im unendlichen Münzwurfexperiment bezeichne $A_n^M := \{\omega \in \Omega : \omega_n = \cdots = \omega_{n+M-1} = 1\}$ das Ereignis eines M-runs von Einsen („Kopf") ab Wurf n. Dann ist die Familie $(A_{kM}^M)_{k\in\mathbb{N}}$ unabhängig (auf dem entsprechenden Münzwurfmodell in den Übungen) mit $\mathbb{P}(A_{kM}^M) = 2^{-M}$. Aus Teil (ii) von Satz §14.06 folgt daher $\mathbb{P}(\limsup_{k\to\infty} A_{kM}^M) = 1$ für jedes $M \in \mathbb{N}$. Dies impliziert $\mathbb{P}(\limsup_{n\to\infty} A_n^M) = 1$. Es gilt sogar $\mathbb{P}(\bigcap_{M\in\mathbb{N}} \limsup_{n\to\infty} A_n^M) = 1$, da der abzählbare Schnitt von Einsmengen wieder eine Einsmenge ist. □

§15 Unabhängige σ-Algebren

Im Folgenden sei $\mathcal{I}$ eine beliebige nicht-leere Indexmenge und für jedes $i \in \mathcal{I}$ sei $\mathscr{A}_i \subseteq \mathscr{A}$ eine σ-Algebra. Die Definition der Unabhängigkeit der Familie $(\mathscr{A}_i)_{i\in\mathcal{I}}$ von Teil-σ-Algebren von $\mathscr{A}$ ist ein Spezialfall der folgenden Definition.

§15.01 **Definition.** Eine Familie $(\mathscr{E}_i)_{i\in\mathcal{I}}$ von Teilmengensystemen aus $\mathscr{A}$, d.h. $\mathscr{E}_i \subseteq \mathscr{A}$ für alle $i \in \mathcal{I}$, heißt *(stochastisch) unabhängig* (unter $\mathbb{P}$), kurz $\perp\!\!\!\perp_{i\in\mathcal{I}}\mathscr{E}_i$, wenn $\perp\!\!\!\perp_{i\in\mathcal{I}} A_i$ für alle $A_i \in \mathscr{E}_i$ und $i \in \mathcal{I}$ gilt. □

§15.02 **Erinnerung.** Nach Lemma §14.04 folgt aus $\perp\!\!\!\perp_{i\in\mathcal{I}} A_i$ dass $\perp\!\!\!\perp_{i\in\mathcal{I}} \sigma(\{A_i\})$ ist. Da $\{A_i\}$ ein π-System, also ein $\cap$-stabiles Teilmengensystem, aus $\mathscr{A}$ ist, ist dies ein Spezialfall der nächsten Aussage, welches wir mit Hilfe des π-λ Satz §03.11 zeigen. □

§15.03 **Satz.** *Seien $(\mathscr{E}_i)_{i\in\mathcal{I}}$ aus $\mathscr{A}$ $\cap$-stabil und unabhängig, also $\perp\!\!\!\perp_{i\in\mathcal{I}}\mathscr{E}_i$, dann gilt $\perp\!\!\!\perp_{i\in\mathcal{I}}\sigma(\mathscr{E}_i)$.*

§15.04 **Beweis** von Satz §15.03. In der Vorlesung. □

§15.05 **Lemma.** *Für eine Familie $(\mathscr{A}_i)_{i\in\mathcal{I}}$ von Teil-σ-Algebren aus $\mathscr{A}$ sei*

$$\mathscr{E} = \{\bigcap_{i\in\mathcal{J}} A_i : A_i \in \mathscr{A}_i,\ i \in \mathcal{J},\ \mathcal{J} \subseteq \mathcal{I},\ |\mathcal{J}| \in \mathbb{N}\}.$$

Dann ist $\mathscr{E}$ $\cap$-stabil und $\bigvee_{i\in\mathcal{I}} \mathscr{A}_i = \sigma(\mathscr{E})$, also $\mathscr{E}$ ist ein $\cap$-stabiler Erzeuger von $\bigvee_{i\in\mathcal{I}} \mathscr{A}_i$.

§15.06 **Beweis** von Lemma §15.05. In der Vorlesung. □

§15.07 **Satz.** *Für jede unabhängige Familie $(\mathscr{A}_i)_{i\in\mathcal{I}}$ von Teil-σ-Algebren aus $\mathscr{A}$, also $\perp\!\!\!\perp_{i\in\mathcal{I}}\mathscr{A}_i$, und jede Partition $\{\mathcal{I}_k : k \in \mathcal{K}\}$ von $\mathcal{I}$ ist $(\bigvee_{i\in\mathcal{I}_k} \mathscr{A}_i)_{k\in\mathcal{K}}$ eine unabhängige Familie von Teil-σ-Algebren aus $\mathscr{A}$, also $\perp\!\!\!\perp_{k\in\mathcal{K}} \bigvee_{i\in\mathcal{I}_k} \mathscr{A}_i$.*

§15.08 **Beweis** von Satz §15.07. In der Vorlesung. □

§15.09 **Definition.** Sei $(\mathscr{A}_n)_{n\in\mathbb{N}}$ eine Folge von Teil-σ-Algebren aus $\mathscr{A}$, dann heißt

$$\mathscr{A}_\infty := \bigwedge_{n\in\mathbb{N}} \bigvee_{m\in\mathbb{N}_{\geqslant n}} \mathscr{A}_m = \bigcap_{n\in\mathbb{N}} \bigvee_{m\in\mathbb{N}_{\geqslant n}} \mathscr{A}_m = \bigcap_{n\in\mathbb{N}} \sigma\big(\bigcup\{\mathscr{A}_m : m \in \mathbb{N}_{\geqslant n}\}\big)$$

die *asymptotische (terminale) σ-Algebra*. Ein Ereignis $A \in \mathscr{A}_\infty$ wird *asymptotisch (terminal)* bzgl. $(\mathscr{A}_n)_{n\in\mathbb{N}}$ genannt. □

§15.10 **Beispiel**. Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge von Ereignissen aus $\mathscr{A}$ und $\mathscr{A}_\infty$ die asymptotische σ-Algebra der Folge $(\sigma(\{A_n\}))_{n\in\mathbb{N}}$ von Teil-σ-Algebren aus $\mathscr{A}$. Dann sind $A_\star = \liminf_{n\to\infty} A_n$ und $A^\star = \limsup_{n\to\infty} A_n$ (vgl. Definition §01.14) asymptotische Ereignisse bzgl. $(\sigma(\{A_n\}))_{n\in\mathbb{N}}$. Setzen wir nämlich $B_n := \bigcap_{m\in\mathbb{N}_{\geqslant n}} A_m$ für $n \in \mathbb{N}$, dann gilt $B_n \uparrow A_\star$ und $B_n \in \bigvee_{m\in\mathbb{N}_{\geqslant N}} \sigma(\{A_m\})$ für jedes $n \in \mathbb{N}_{\geqslant N}$, sodass $A_\star \in \bigvee_{m\in\mathbb{N}_{\geqslant N}} \sigma(\{A_m\})$ für alle $N \in \mathbb{N}$ und damit auch

$$A_\star \in \bigwedge_{N\in\mathbb{N}} \bigvee_{m\in\mathbb{N}_{\geqslant N}} \sigma(\{A_m\}) = \mathscr{A}_\infty.$$

Für $A^\star$ geht dies analog. Insbesondere ist die asymptotische σ-Algebra $\mathscr{A}_\infty$ nicht leer. □

Betrachte eine Folge unabhängiger Ereignisse $(A_n)_{n\in\mathbb{N}}$, so ist auf Grund des Satzes von Borel-Cantelli §14.06 das asymptotische Ereignis $A^\star = \limsup_{n\to\infty} A_n$ bzgl. $(\sigma(\{A_n\}))_{n\in\mathbb{N}}$ entweder eine Nullmenge oder eine Einsmenge. Die nächste Aussage zeigt nun, dass dies für jedes asymptotische Ereignis gilt.

§15.11 **0-1-Gesetz von Kolmogorov**. *Seien $(\mathscr{A}_n)_{n\in\mathbb{N}}$ unabhängige Teil-σ-Algebren aus $\mathscr{A}$. Dann ist die Wahrscheinlichkeit für jedes bzgl. $(\mathscr{A}_n)_{n\in\mathbb{N}}$ asymptotische Ereigniss entweder 0 oder 1, also*

$$\mathscr{A}_\infty \subseteq \overline{\mathscr{T}} := \{A \in \mathscr{A} : \mathbb{P}(A) \in \{0, 1\}\}.$$

§15.12 **Beweis** von Satz §15.11. In der Vorlesung. □

§15.13 **Bemerkung**. $\overline{\mathscr{T}}$ ist eine σ-Algebra, da eine abzählbare Vereinigung von Nullmengen wieder eine Nullmenge ist. $\overline{\mathscr{T}}$ wird *$\mathbb{P}$-triviale* σ-Algebra genannt. □

§15.14 **Beispiel** *(Beispiel §15.10 forgesetzt.)*. Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge von *unabhängigen* Ereignissen aus $\mathscr{A}$. Da $A_\star = \liminf_{n\to\infty} A_n$ und $A^\star = \limsup_{n\to\infty} A_n$ (vgl. Beispiel §15.10) asymptotische Ereignisse bzgl. der Folge unabhängiger σ-Algebren $(\sigma(\{A_n\}))_{n\in\mathbb{N}}$ (vgl. Erinnerung §15.02) sind, folgt aus Satz §15.11 $\mathbb{P}(A_\star) \in \{0, 1\}$ und $\mathbb{P}(A^\star) \in \{0, 1\}$. □

§16 Unabhängige Zufallsvariablen

Im Folgenden seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $((\mathcal{X}_i, \mathscr{X}_i))_{i\in\mathcal{I}}$ eine Familie messbarer Räume mit beliebiger nicht-leerer Indexmenge $\mathcal{I}$ und für jedes $i \in \mathcal{I}$, $X_i \in \mathcal{M}(\mathscr{A}, \mathscr{X}_i)$ eine $(\mathcal{X}_i, \mathscr{X}_i)$-wertige Zufallsvariable.

§16.01 **Erinnerung**. Für jedes $i \in \mathcal{I}$ ist $\sigma(X_i) = X_i^{-1}(\mathscr{X}_i) = \{X_i(B) : B \in \mathscr{X}_i\} \subseteq \mathscr{A}$ die von X_i erzeugte σ-Algebra. □

§16.02 **Definition**. Die Familie $(X_i)_{i\in\mathcal{I}}$ von Zufallsvariablen heißt *unabhängig*, kurz $\perp\!\!\!\perp_{i\in\mathcal{I}} X_i$, wenn die Familie $(\sigma(X_i))_{i\in\mathcal{I}}$ von Teil-σ-Algebren von $\mathscr{A}$ unabhängig ist. □

§16.03 **Bemerkung**. Definitionsgemäß gilt $\perp\!\!\!\perp_{i\in\mathcal{I}} X_i$ genau dann, wenn $\perp\!\!\!\perp_{j\in\mathcal{J}} X_j$ für jede endliche Teilmenge $\mathcal{J} \in 2^{\mathcal{I}}_{\neq\emptyset}$ gilt. Weiterhin für jede endliche Familie $(B_j)_{j\in\mathcal{J}}$ von Ereignissen, also für $B_j \in \mathscr{X}_j$, $j \in \mathcal{J} \subseteq \mathcal{I}$, $|\mathcal{J}| \in \mathbb{N}$, gilt $\mathbb{P}(\bigcap_{j\in\mathcal{J}}\{X_j \in B_j\}) = \prod_{j\in\mathcal{J}} \mathbb{P}(X_j \in B_j)$. □

§16.04 **Beispiel**. Ist $(A_i)_{i\in\mathcal{I}}$ eine Familie von Ereignissen, so gilt $\sigma(\{A_i\}) = \sigma(\mathbb{1}_{A_i})$ für jede Bernoulli-Zufallsvariable $\mathbb{1}_{A_i}$, $i \in \mathcal{I}$. Nach Lemma §14.04 gilt $\perp\!\!\!\perp_{i\in\mathcal{I}} A_i$ genau dann wenn $\perp\!\!\!\perp_{i\in\mathcal{I}} \mathbb{1}_{A_i}$. □

§16.05 **Lemma**. *Sei $((\mathcal{S}_i, \mathscr{S}_i))_{i\in\mathcal{I}}$ eine Familie messbarer Räume und für jedes $i \in \mathcal{I}$ sei $h_i \in \mathcal{M}(\mathscr{X}_i, \mathscr{S}_i)$ eine $\mathscr{X}_i$-$\mathscr{S}_i$-messbare Funktion. Ist $\perp\!\!\!\perp_{i\in\mathcal{I}} X_i$ erfüllt, so gilt auch $\perp\!\!\!\perp_{i\in\mathcal{I}} h_i(X_i)$. Für jede Partition*

$\{\mathcal{I}_k : k \in \mathcal{K}\}$ *von* $\mathcal{I}$ *ist* $((h_i(X_i))_{i\in\mathcal{I}_k})_{k\in\mathcal{K}}$ *eine Familie von unabhängigen Zufallsvariablen, also* $\perp\!\!\!\perp_{k\in\mathcal{K}}(h_i(X_i))_{i\in\mathcal{I}_k}$.

§16.06 **Beweis** von Lemma §16.05. In der Vorlesung. □

§16.07 **Korollar.**

(i) *Eine Familie* $(X_i)_{i\in\mathcal{I}}$ *numerischer Zufallsvariablen ist genau dann unabhängig, wenn für jede endliche Teilmenge* $\mathcal{J} \in 2^{\mathcal{I}}_{\neq\varnothing}$ *und für alle* $x_j \in \mathbb{R}$, $j \in \mathcal{J}$, *gilt:*

$$\mathbb{P}\Big(\bigcap_{j\in\mathcal{J}}\{X_j \leqslant x_j\}\Big) = \prod_{j\in\mathcal{J}} \mathbb{P}(X_j \leqslant x_j).$$

Für $\mathcal{I} = [\![n]\!]$ *sind also* $X := (X_i)_{i\in[\![n]\!]}$ *genau dann unabhängig, wenn für alle* $x \in \mathbb{R}^n$ *gilt*

$$\mathbb{F}^X(x) = \prod_{i\in[\![n]\!]} \mathbb{F}^{X_i}(x_i).$$

(ii) *Diskret-verteilte Zufallsvariablen* $X := (X_i)_{i\in[\![n]\!]}$ *sind genau dann unabhängig, wenn für alle* $x \in \bigtimes_{i\in[\![n]\!]}\mathcal{X}_i$ *gilt*

$$\mathbb{p}^X(x) = \prod_{i\in[\![n]\!]} \mathbb{p}^{X_i}(x_i).$$

(iii) *Stetig-verteilte Zufallsvariablen* $X := (X_i)_{i\in[\![n]\!]}$ *sind genau dann unabhängig, wenn für Lebesgue-fast alle* $x \in \mathbb{R}^n$ *gilt*

$$\mathbb{f}^X(x) = \prod_{i\in[\![n]\!]} \mathbb{f}^{X_i}(x_i).$$

§16.08 **Beweis** von Korollar §16.07. In der Vorlesung. □

§16.09 **Beispiel.**

(a) Beim Würfelwurf mit zwei Würfeln sind $W_i : \Omega \to [\![6]\!]$ mit $W_i(\omega_1,\omega_2) = \omega_i$ für $i \in [\![2]\!]$ unabhängige Zufallsvariablen (unter Laplaceverteilung). Dazu genügt es, für $\omega_1,\omega_2 \in [\![6]\!]$ die entsprechenden Zähldichten $\mathbb{p}^{W_1}(\omega_1) = \mathbb{p}^{W_2}(\omega_2) = 1/6$ sowie $\mathbb{p}^W(\omega_1,\omega_2) = 1/36$ mit $W = (W_1, W_2)$ nachzuprüfen und Korollar §16.07 (ii) anzuwenden.

(b) Sei $X = |W_1 - W_2|$ und $Y = W_1 + W_2$ der Absolutbetrag der Differenz bzw. die Summe der Augenzahlen von zwei unabhängigen fairen Würfeln (W_1, W_2) (vgl. Beispiel §11.12 (a)). Betrachten wir die entsprechenden Zähldichten $\mathbb{p}^X$ und $\mathbb{p}^Y$, so gilt zum Beispiel $\mathbb{p}^{(X,Y)}(1,4) = 0 \neq \frac{10}{36}\cdot\frac{3}{36} = \mathbb{p}^X(1)\cdot\mathbb{p}^Y(4)$. Somit sind X und Y nicht unabhängig.

(c) Für $n \in \mathbb{N}$ sei die gemeinsame Verteilung von $(X_i)_{i\in[\![n]\!]}$ ein Bernoulli-Schema mit Parameter $\mathrm{p} \in [0,1]$ (vgl. Beispiel §04.08 (c)). Dann sind $(X_i)_{i\in[\![n]\!]}$ unabhängige und identisch B_{p}-verteilte Bernoulli-Zufallsvariablen mit identischer Randzähldichte $\mathbb{p}(x) = \mathrm{p}^x(1-\mathrm{p})^{1-x}$, $x \in \{0,1\}$, da für die gemeinsame Zähldichte $\mathrm{p}^{\sum_{i\in[\![n]\!]}x_i}(1-\mathrm{p})^{n-\sum_{i\in[\![n]\!]}x_i} = \prod_{i\in[\![n]\!]}\mathbb{p}(x_i)$ für alle $(x_i)_{i\in[\![n]\!]} \in \{0,1\}^n$ gilt.

(d) Sind $(\mathbb{f}_i)_{i\in[\![n]\!]}$ Dichten auf $\mathbb{R}$, so erzeugt die Produktdichte $\mathbb{f}(x) = \prod_{i\in[\![n]\!]}\mathbb{f}_i(x_i)$ ein Produktmaß auf $(\mathbb{R}^n, \mathscr{B}^n)$ (vgl. Definition §11.03). Für jedes $i \in [\![n]\!]$ ist die Koordinatenabbildungen $X_i : \mathbb{R}^n \to \mathbb{R}$ mit $x \mapsto X_i(x) := x_i$ Borel-messbar, also eine Zufallsvariable $X_i \in \mathcal{M}(\mathscr{B}^n)$. Ihre Randverteilung $\mathbb{P}^{X_i}$ ist gegeben durch die Randdichte $\mathbb{f}_i$, ihre gemeinsame Verteilung $\mathbb{P}^{(X_1,\dots,X_n)}$ durch die gemeinsame Dichte $\mathbb{f}$. Wir haben damit einen Wahrscheinlichkeitsraum konstruiert mit unabhängigen Zufallsvariablen $(X_i)_{i\in[\![n]\!]}$ deren Randverteilung $\mathbb{P}^{X_i}$ jeweils durch $\mathbb{f}_i$ bestimmt ist.

(e) Betrachten wir die gemeinsame Dichte $\mathbb{f}^{(X,Y)}$ eines bivariat normalverteilten Zufallsvektors (X, Y) wie in Beispiel §10.12 (d). So sind X und Y stetig-verteilt mit Randdichten $\mathbb{f}^X$ und $\mathbb{f}^Y$ (vgl. Beispiel §11.12 (b)). Die gemeinsame Dichte $\mathbb{f}^{(X,Y)}$ faktorisiert sich in das Produkt der Randdichten $\mathbb{f}^X$ und $\mathbb{f}^Y$ genau dann, wenn $\rho = 0$ gilt. Somit sind X und Y genau dann unabhängig, wenn $\rho = 0$ gilt. □

§16.10 **Erinnerung.** Nach Satz §11.07 existiert zu jeder Familie $(\mathcal{X}_i, \mathscr{B}_i, \mathbb{P}_i)_{i\in\mathcal{I}}$ von Wahrscheinlichkeitsräumen (polnische Räume versehen mit Borel-σ-Algebra) ein eindeutig bestimmtes Produktmaß $\bigotimes_{i\in\mathcal{I}} \mathbb{P}_i$ auf dem zugehörigen Produktraum $\mathcal{X}_\mathcal{I} = \bigtimes_{i\in\mathcal{I}} \mathcal{X}_i$ versehen mit der Produkt-σ-Algebra $\mathscr{B}_\mathcal{I} = \bigotimes_{i\in\mathcal{I}} \mathscr{B}_i$. □

§16.11 **Korollar.** *Für jede Familie $((\mathcal{X}_i, \mathscr{B}_i, \mathbb{P}_i))_{i\in\mathcal{I}}$ von Wahrscheinlichkeitsräumen (polnische Räume versehen mit Borel-σ-Algebra) existiert eine Familie $(X_i)_{i\in\mathcal{I}}$ unabhängiger $(\mathcal{X}_i, \mathscr{B}_i)$-wertiger Zufallsvariablen definiert auf einem gemeinsamen Wahrscheinlichkeitsraum mit Randverteilung $\mathbb{P}_i$ und dem Produktmaß $\bigotimes_{i\in\mathcal{I}} \mathbb{P}_i$ als gemeinsamer Verteilung auf dem Produktraum $(\mathcal{X}_\mathcal{I}, \mathscr{B}_\mathcal{I})$.*

§16.12 **Beweis** von Korollar §16.11. In der Vorlesung. □

§16.13 **Definition.** Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von $(\mathcal{X}_n, \mathscr{X}_n)$-wertigen Zufallsvariablen auf $(\Omega, \mathscr{A}, \mathbb{P})$. Dann heißt

$$\boxed{\mathscr{A}_X} := \bigwedge_{n\in\mathbb{N}} \bigvee_{m\in\mathbb{N}_{\geqslant n}} \sigma(X_m) = \bigcap_{n\in\mathbb{N}} \bigvee_{m\in\mathbb{N}_{\geqslant n}} \sigma(X_m) = \bigcap_{n\in\mathbb{N}} \sigma\big(\bigcup \{\sigma(X_m) : m \in \mathbb{N}_{\geqslant n}\}\big)$$

die *asymptotische σ-Algebra*. Ein Ereignis $A \in \mathscr{A}_X$ wird *asymptotisch* bzgl. $(X_n)_{n\in\mathbb{N}}$ genannt, d.h. A hängt für alle $n \in \mathbb{N}$ nur von $(X_m)_{m\in\mathbb{N}_{\geqslant n}}$ ab. □

§16.14 **Beispiel.**

(a) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge numerischer Zufallsvariablen, so sind die Abbildungen (vgl. Definition §08.10) $X_\star := \liminf_{n\to\infty} X_n$ und $X^\star := \limsup_{n\to\infty} X_n$ messbar bzgl. $\mathscr{A}_X$. In der Tat: setzen wir $Y_n := \sup\{X_m : m \in \mathbb{N}_{\geqslant n}\}$ mit $Y_n \downarrow X^\star$, so ist für jedes $N \in \mathbb{N}$ die Zufallsvariable $X^\star = \inf\{Y_n : n \in \mathbb{N}\} = \inf\{Y_n : n \in \mathbb{N}_{\geqslant N}\}$ messbar bzgl. $\bigvee_{n\in\mathbb{N}_{\geqslant N}} \sigma(X_n)$, also auch bzgl. $\bigwedge_{N\in\mathbb{N}} \bigvee_{n\in\mathbb{N}_{\geqslant N}} \sigma(X_n)$. Für $X_\star$ geht dies analog.

(b) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reeller Zufallsvariablen sowie A das Ereignis, dass der Grenzwert $\lim_{n\to\infty} \sum_{k\in[\![n]\!]} \frac{X_k}{k}$ existiert. Setzen wir für $\varepsilon \in \mathbb{Q}_{>0}$ und $N \in \mathbb{N}$

$$B_{N,\varepsilon} := \bigcap \Big\{\Big\{ \sum_{k\in[\![n,m]\!]} \frac{X_k}{k} \in (-\varepsilon, \varepsilon)\Big\} : n \in \mathbb{N}_{\geqslant N}, m \in \mathbb{N}_{>n}\Big\},$$

so gilt nach dem Cauchy-Kriterium $A = \bigcap_{\varepsilon\in\mathbb{Q}_{>0}} \bigcup_{N\in\mathbb{N}} B_{N,\varepsilon}$. Nun ist $(B_{N,\varepsilon})_{N\in\mathbb{N}}$ monoton wachsend, $B_{N,\varepsilon} \in \bigvee_{k\in\mathbb{N}_{\geqslant N}} \sigma(X_k)$ für alle $N \in \mathbb{N}$ und für jedes $n \in \mathbb{N}$ und alle $m \in \mathbb{N}_{\geqslant n}$ gilt $\bigvee_{k\in\mathbb{N}_{\geqslant m}} \sigma(X_k) \subseteq \bigvee_{k\in\mathbb{N}_{\geqslant n}} \sigma(X_k)$, so dass

$$\bigcup_{N\in[\![m]\!]} B_{N,\varepsilon} = B_{m,\varepsilon} \in \bigvee_{k\in\mathbb{N}_{\geqslant n}} \sigma(X_k) \quad \forall m \in \mathbb{N}_{\geqslant n}.$$

Folglich gilt $\bigcup_{N\in\mathbb{N}} B_{N,\varepsilon} \in \bigvee_{k\in\mathbb{N}_{\geqslant n}} \sigma(X_k)$ für alle $n \in \mathbb{N}$ und damit $C_\varepsilon := \bigcup_{N\in\mathbb{N}} B_{N,\varepsilon} \in \mathscr{A}_X$. Da $\mathscr{A}_X$ als σ-Algebra σ-$\cap$-stabil ist, folgt auch $A = \bigcap_{\varepsilon\in\mathbb{Q}_{>0}} C_\varepsilon \in \mathscr{A}_X$. □

§16.15 **0-1-Gesetz von Kolmogorov.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen auf $(\Omega, \mathscr{A}, \mathbb{P})$. Dann ist die Wahrscheinlichkeit für jedes bzgl. $(X_n)_{n\in\mathbb{N}}$ asymptotischen Ereignisses entweder 0 oder 1, also*

$$\mathscr{A}_X \subseteq \overline{\mathscr{T}} := \{A \in \mathscr{A} : \mathbb{P}(A) \in \{0,1\}\}.$$

§16.16 **Beweis** von Korollar §16.15. Direktes Anwenden von Satz §15.11. □

§16.17 **Beispiel** *(Beispiel §16.14 fortgesetzt).*

(a) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge *unabhängiger* numerischer Zufallsvariablen auf $(\Omega, \mathscr{A}, \mathbb{P})$, so sind die Abbildungen $X_\star := \liminf_{n\to\infty} X_n$ und $X^\star := \limsup_{n\to\infty} X_n$ fast sicher konstant, das heißt, es gibt $x_\star, x^\star \in \overline{\mathbb{R}}$ mit $\mathbb{P}(X_\star = x_\star) = 1$ und $\mathbb{P}(X^\star = x^\star) = 1$. In der Tat, da $X_\star$ messbar bzgl. der asymptotischen σ-Algebra ist (vgl. Beispiel §16.14 (a)), ist für jedes $x \in \overline{\mathbb{R}}$ das Ereignis $\{X_\star \leqslant x\}$ asymptotisch bzgl. $(X_n)_{n\in\mathbb{N}}$ und also $\mathbb{P}(X_\star \leqslant x) \in \{0,1\}$. Setze

$$x_\star := \min\{x \in \overline{\mathbb{R}} : \mathbb{P}(X_\star \leqslant x) = 1\} \in \overline{\mathbb{R}}.$$

Ist $x_\star = \infty$, und also $\mathbb{P}(X_\star \leqslant n) = 0$ für alle $n \in \mathbb{N}$, so gilt

$$\mathbb{P}(X_\star < \infty) = \lim_{n\to\infty} \mathbb{P}(X_\star \leqslant n) = 0,$$

also $\mathbb{P}(X_\star = \infty) = 1$. Ist $x_\star \in \mathbb{R}$, so ist

$$\mathbb{P}(X_\star \leqslant x_\star) = \lim_{n\to\infty} \mathbb{P}(X_\star \leqslant x_\star + \tfrac{1}{n}) = 1 \text{ und}$$
$$\mathbb{P}(X_\star < x_\star) = \lim_{n\to\infty} \mathbb{P}(X_\star \leqslant x_\star - \tfrac{1}{n}) = 0,$$

also $\mathbb{P}(X_\star = x_\star) = 1$. Ist $x_\star = -\infty$, so gilt $\mathbb{P}(X_\star > -\infty) = \lim_{n\to\infty} \mathbb{P}(X_\star > n) = 0$, also auch $\mathbb{P}(X_\star = -\infty) = 1$. Für den Limes superior geht dies analog.

(b) Ist $(X_n)_{n\in\mathbb{N}}$ wie in Beispiel §16.14 (b) eine Folge unabhängiger $\{-1,1\}$-wertiger Zufallsvariablen, so konvergiert die harmonische Reihe mit zufälligem Vorzeichen $\sum_{k\in\mathbb{N}} X_k \frac{1}{k}$ entweder mit Wahrscheinlichkeit 1 oder sie divergiert mit Wahrscheinlichkeit 1. Diese Aussage gilt für beliebige Wahl der Randverteilung $\mathbb{P}^{X_k}$ der Zufallsvariable X_k, $k \in \mathbb{N}$. □

§17 Faltung

§17|00.01 **Vorbemerkung.** Wir beenden diesen Abschnitt mit einem Begriff, der eng mit dem des Produktmaßes verknüpft ist, nämlich dem der Faltung. Zur Motivation seien X, Y zwei unabhängige, reelle Zufallsvariablen mit Verteilungen $\mathbb{P}^X$ und $\mathbb{P}^Y$ auf $(\mathbb{R}, \mathscr{B})$. Gemäß Satz §11.07 hat dann das Paar (X, Y) die Verteilung $\mathbb{P}^X \bigotimes \mathbb{P}^Y$ auf $(\mathbb{R}^2, \mathscr{B}^2)$. Andererseits ist $X + Y$ eine reelle Zufallsvariable, und für die Borel-messbare Additionsabbildung

$$\oplus : \mathbb{R}^2 \ni (x,y) \mapsto \oplus(x,y) := x + y \in \mathbb{R}$$

gilt $X + Y = \oplus \circ (X, Y)$. Damit hat $X + Y$ die Verteilung $(\mathbb{P}^X \bigotimes \mathbb{P}^Y) \circ \oplus^{-1}$ auf $(\mathbb{R}, \mathscr{B})$. Aufgrund der Kommutativität der Addition ist $(\mathbb{P}^Y \bigotimes \mathbb{P}^X) \circ \oplus^{-1}$ auch die Verteilung von $Y + X$. □

§17.02 **Definition.** Seien $\mathbb{P}, \widetilde{\mathbb{P}}$ zwei Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathscr{B})$. Dann heißt das Wahrscheinlichkeitsmaß $\mathbb{P} \star \widetilde{\mathbb{P}} := (\mathbb{P} \bigotimes \widetilde{\mathbb{P}}) \circ \oplus^{-1}$ auf $(\mathbb{R}, \mathscr{B})$ die *Faltung* von $\mathbb{P}$ und $\widetilde{\mathbb{P}}$. □

§17.03 **Satz**. *Es seien $X \sim \mathbb{P}^X$ und $Y \sim \mathbb{P}^Y$ unabhängige reelle Zufallsvariablen. Dann besitzt die reelle Zufallsvariable $X+Y$ die Verteilung $\mathbb{P}^{X+Y} = \mathbb{P}^X \star \mathbb{P}^Y$.*

§17.04 **Beweis** von Satz §17.03. Die Aussage folgt aus der Vorbemerkung §17|00.01. □

§17.05 **Korollar**. *Die Faltung ist kommutativ und assoziativ.*

§17.06 **Beweis** von Korollar §17.05. Die Kommutativität folgt aus der Vorbemerkung §17|00.01. Die Assoziativität der Addition vererbt sich analog auf die Faltung. □

§17.07 **Lemma** *(Diskreter Fall).*

(i) *Besitzen $\mathbb{P}$ und $\widetilde{\mathbb{P}}$ Zähldichten $\mathbb{p}$ und $\widetilde{\mathbb{p}}$ auf $\mathbb{Z}$, so besitzt $\mathbb{P} \star \widetilde{\mathbb{P}}$ die Zähldichte*

$$[\mathbb{p} \star \widetilde{\mathbb{p}}](z) := \sum_{k\in\mathbb{Z}} \mathbb{p}(z-k)\widetilde{\mathbb{p}}(k), \quad z \in \mathbb{Z};$$

(ii) *Besitzen $\mathbb{P}$ und $\widetilde{\mathbb{P}}$ Zähldichten $\mathbb{p}$ und $\widetilde{\mathbb{p}}$ auf $\mathbb{Z}_{\geqslant 0}$, so besitzt $\mathbb{P} \star \widetilde{\mathbb{P}}$ die Zähldichte*

$$[\mathbb{p} \star \widetilde{\mathbb{p}}](n) := \sum_{k\in[\![0,n]\!]} \mathbb{p}(n-k)\widetilde{\mathbb{p}}(k), \quad n \in \mathbb{Z}_{\geqslant 0}.$$

§17.08 **Beweis** von Lemma §17.07. In der Vorlesung. □

§17.09 **Beispiel**.

(a) Für $n \in \mathbb{N}$ sei die gemeinsame Verteilung von $(X_i)_{i\in[\![n]\!]} \sim \mathrm{B}_\mathrm{p}^n$ ein Bernoulli-Schema mit Parameter $\mathrm{p} \in [0,1]$ (vgl. Beispiel §04.08 (c)), das heißt, $(X_i)_{i\in[\![n]\!]}$ sind unabhängige, identisch B_p-verteilte Bernoulli-Zufallsvariablen (vgl. Beispiel §16.09 (c)). Dann ist $X_1 + X_2$ Binomial $\mathrm{Bin}_{(2,\mathrm{p})}$-verteilt (vgl. Beispiel §04.08 (d)), da $\mathrm{B}_\mathrm{p} \star \mathrm{B}_\mathrm{p} = \mathrm{Bin}_{(2,\mathrm{p})}$:

$$\begin{aligned}
[\mathbb{p}_{\mathrm{B}_\mathrm{p}} \star \mathbb{p}_{\mathrm{B}_\mathrm{p}}](0) &= \mathbb{p}_{\mathrm{B}_\mathrm{p}}(0)\mathbb{p}_{\mathrm{B}_\mathrm{p}}(0) = \mathrm{p}^0(1-\mathrm{p})^1\mathrm{p}^0(1-\mathrm{p})^1 = \mathrm{p}^0(1-\mathrm{p})^{2-0},\\
[\mathbb{p}_{\mathrm{B}_\mathrm{p}} \star \mathbb{p}_{\mathrm{B}_\mathrm{p}}](1) &= \mathbb{p}_{\mathrm{B}_\mathrm{p}}(1)\mathbb{p}_{\mathrm{B}_\mathrm{p}}(0) + \mathbb{p}_{\mathrm{B}_\mathrm{p}}(0)\mathbb{p}_{\mathrm{B}_\mathrm{p}}(1) = 2\mathrm{p}^1(1-\mathrm{p})^{2-1},\\
[\mathbb{p}_{\mathrm{B}_\mathrm{p}} \star \mathbb{p}_{\mathrm{B}_\mathrm{p}}](2) &= \mathbb{p}_{\mathrm{B}_\mathrm{p}}(1)\mathbb{p}_{\mathrm{B}_\mathrm{p}}(1) = \mathrm{p}^2(1-\mathrm{p})^{2-2}.
\end{aligned}$$

Per Induktion erhalten wir $\sum_{i\in[\![n]\!]} X_i \sim \mathrm{Bin}_{(n,\mathrm{p})}$. Für ein Bernoulli-Schema $(X_i)_{i\in[\![n+m]\!]} \sim \mathrm{B}_\mathrm{p}^{n+m}$ mit $n,m \in \mathbb{N}$ sind $\sum_{i\in[\![n]\!]} X_i \sim \mathrm{Bin}_{(n,\mathrm{p})}$ und $\sum_{i\in[\![n+1,n+m]\!]} X_i \sim \mathrm{Bin}_{(m,\mathrm{p})}$ unabhängig (Lemma §16.05), und $\sum_{i\in[\![n+m]\!]} X_i \sim \mathrm{Bin}_{(n+m,\mathrm{p})}$, also $\mathrm{Bin}_{(n,\mathrm{p})} \star \mathrm{Bin}_{(m,\mathrm{p})} = \mathrm{Bin}_{(n+m,\mathrm{p})}$.

(b) Seien $X \sim \mathrm{Poi}_\lambda$ und $Y \sim \mathrm{Poi}_\mu$ unabhängig mit $\lambda, \mu \in \mathbb{R}_{>0}$, so gilt $X+Y \sim \mathrm{Poi}_{\lambda+\mu}$ (*Übungsaufgabe!*). Setzt man zusätzlich $\mathrm{Poi}_0 := \delta_0$, so bildet $(\mathrm{Poi}_\lambda)_{\lambda\in\mathbb{R}_{\geqslant 0}}$ eine *Faltungshalbgruppe*, d.h. es gilt $\mathrm{Poi}_\lambda \star \mathrm{Poi}_\mu = \mathrm{Poi}_{\lambda+\mu}$ für alle $\lambda, \mu \in \mathbb{R}_{\geqslant 0}$. □

§17.10 **Lemma** *(Stetiger Fall). Seien X und Y unabhängige stetig-verteilte Zufallsvariablen mit Dichte $\mathbb{f}^X$ bzw. $\mathbb{f}^Y$. Dann ist $X+Y$ stetig-verteilt mit Dichte*

$$\mathbb{f}^{X+Y}(z) = [\mathbb{f}^X \star \mathbb{f}^Y](z) := \int_\mathbb{R} \mathbb{f}^X(z-y)\mathbb{f}^Y(y)dy, \quad z \in \mathbb{R}.$$

§17.11 **Beweis** von Lemma §17.10. In der Vorlesung. □

§17.12 **Beispiel**. Für $\lambda, p \in \mathbb{R}_{>0}$ definiere die *Gamma-Verteilung* $\Gamma_{(\lambda,p)}$ über die Dichte

$$\mathbb{f}_{\Gamma_{(\lambda,p)}}(x) = \tfrac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbb{1}_{\mathbb{R}_{>0}}(x), \quad x \in \mathbb{R},$$

mit Gamma-Funktion $\Gamma(p) = \int_0^\infty t^{p-1}e^{-t}dt$, sodass $\Gamma(n+1) = n!$ für $n \in \mathbb{Z}_{\geqslant 0}$ und $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

(a) Spezialfälle sind $\Gamma_{(\lambda,1)} = \mathrm{Exp}_\lambda$ und $\Gamma_{(1/2,1/2)} = \chi^2_1$.

(b) Setzt man zusätzlich $\Gamma_{(\lambda,0)} := \delta_0$, so bildet $(\Gamma_{(\lambda,p)})_{p\in\mathbb{R}_{\geqslant 0}}$ für festes $\lambda \in \mathbb{R}_{>0}$ eine *Faltungshalbgruppe*, d.h. es gilt $\Gamma_{(\lambda,p_1)} \star \Gamma_{(\lambda,p_2)} = \Gamma_{(\lambda,p_1+p_2)}$.

(c) Insbesondere ist die *Summe von k unabhängigen χ^2_1-verteilten Zufallsvariablen* (äquivalent dazu die Summe der Quadrate, $\sum_{i\in[\![k]\!]} Z_i^2$, von k unabhängigen standard-normalverteilten Zufallsvariablen $(Z_i)_{i\in[\![k]\!]}$, oder äquivalent $\|Z\|^2$ eines standard-normalverteilten Zufallsvektors $Z = (Z_i)_{i\in[\![k]\!]} \sim \mathrm{N}_{(0_{n\times 1},\mathrm{I}_{n\times n})}$ im $\mathbb{R}^k$, vgl. Beispiel §05.10) gemäß $\boxed{\chi^2_k} := \Gamma_{(1/2,k/2)}$*-verteilt* (Sprechweise: *χ^2-Verteilung mit k Freiheitsgraden*). □

§17.13 **Beispiel**. Für $X \sim \mathrm{N}_{(\mu_X,\sigma_X^2)}$ und $Y \sim \mathrm{N}_{(\mu_Y,\sigma_Y^2)}$ unabhängig ist $X + Y \sim \mathrm{N}_{(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2)}$. Setzen wir $\mathrm{N}_{(\mu,0)} := \delta_\mu$ so bildet $(\mathrm{N}_{(t\mu,t\sigma^2)})_{t\in\mathbb{R}_{\geqslant 0}}$ eine Faltungshalbgruppe für alle $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{\geqslant 0}$. Insbesondere, für $(X_i)_{i\in[\![n]\!]} \sim \mathrm{N}^n_{(\mu,\sigma^2)}$ gilt $\sum_{i\in[\![n]\!]} X_i \sim \mathrm{N}_{(n\mu,n\sigma^2)}$ und folglich $\overline{X_n} \sim \mathrm{N}_{(\mu,\sigma^2/n)}$ (vgl. Beispiel §10.12 (a)). □

§18 Multivariate Normalverteilung

§18.01 **Vorbemerkung.**. Im Folgenden fassen wir Vektoren $a = (a_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$ als Spaltenvektoren

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix} = [a_1 \cdots a_i \cdots a_n]^{\mathrm{T}} \in \mathbb{R}^n$$

auf. Wir bezeichnen mit $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt auf $\mathbb{R}^n$, das heißt,

$$\langle a, b\rangle = \mathrm{b}^{\mathrm{T}} a = \sum_{i\in[\![n]\!]} a_i b_i \quad \text{und} \quad \langle a, a\rangle = \|a\|^2 \quad \text{für alle } a, b \in \mathbb{R}^n.$$

Sei X ein $\mathbb{R}^n$-wertiger Zufallsvektor mit Dichte $\mathbb{f}^X$. Für $\mathrm{b} \in \mathbb{R}^n$ und regulärem $\mathrm{A} \in \mathbb{R}^{n\times n}$, also $\det(\mathrm{A}) \neq 0$, besitzt nach dem Dichtetransformationssatz §10.08 dann der $\mathbb{R}^n$-wertige Zufallsvektor $Y = \mathrm{A}X + b$ die Dichte

$$\mathbb{f}^Y(y) = \frac{\mathbb{f}^X(\mathrm{A}^{-1}(y-b))}{|\det(\mathrm{A})|} \quad \text{für alle } y \in \mathbb{R}^n.$$

Sind insbesondere die Komponenten von

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{bmatrix} = [X_1 \cdots X_i \cdots X_n]^{\mathrm{T}}$$

unabhängig und ist $\mathrm{Diag}[a_i]$ eine Diagonalmatrix mit Diagonaleinträgen $(a_i)_{i\in[\![n]\!]}$, kurz $\mathrm{A} = \mathrm{Diag}[a_i]$, mit $a_i \in \mathbb{R}_{\neq 0}$ für alle $i \in [\![n]\!]$, dann gilt $\det(\mathrm{A}) = \prod_{i\in[\![n]\!]} a_i \neq 0$ und

$$Y = \mathrm{A}X + b = \begin{bmatrix} a_1 X_1 + b_1 \\ \vdots \\ a_n X_n + b_n \end{bmatrix}$$

besitzt die Dichte

$$\mathbb{f}^Y(y) = \prod_{i \in [\![n]\!]} \tfrac{1}{|a_i|} \mathbb{f}^{X_i}(\tfrac{y_i - b_i}{a_i}) \qquad \text{für alle } y \in \mathbb{R}^n.$$

In Beispiel §10.12 (c) haben wir weiterhin auf $(\mathbb{R}^n, \mathscr{B}^n)$ die multivariate Normalverteilung $\mathrm{N}_{(\mu,\Sigma)}$ mit Vektor $\mu \in \mathbb{R}^n$ und positiv definiter, also symmetrischer und regulärer, Matrix $\Sigma \in \mathbb{R}^{n\times n}$ über ihre Dichte $\mathbb{f}_{\mathrm{N}_{(\mu,\Sigma)}}$ eingeführt. Im Fall einer nicht regulären Matrix Σ wird eine Normalverteilung *degeneriert* genannt. Im Folgenden werden wir einen allgemeineren Zugang betrachten, der es erlaubt, auch degenerierte Normalverteilungen einzuführen, die nicht über eine Lebesgue-Dichte auf $\mathbb{R}^{n\times 1}$ definiert werden können. □

§18.02 **Satz von Cramér-Wold**. *Die Verteilung eines $\mathbb{R}^n$-wertigen Zufallsvektors X ist eindeutig festgelegt durch die Verteilungen der linearen Formen $\langle X, a\rangle$ für alle $a \in \mathbb{R}^n$.*

§18.03 **Beweis** von Satz §18.02. Die Aussage kann zum Beispiel unter Zuhilfenahme von multivariaten charakteristischen Funktionen (zum Beispiel Satz 15.5 in Klenke (2012)) gezeigt werden. □

§18.04 **Korollar**. *Die Koordinaten eines $\mathbb{R}^n$-wertigen Zufallsvektors X sind genau dann unabhängig und identisch (standardnormal) $\mathrm{N}_{(0,1)}$-verteilt, wenn für jedes $a \in \mathbb{R}^n$ die reelle Zufallsvariable $\langle X, a\rangle$ eine $\mathrm{N}_{(0,\|a\|^2)}$-Verteilung besitzt. Für $a \neq \mathbb{0}_n$ ist $\langle X, a\rangle$ also stetig verteilt mit Dichte*

$$\mathbb{f}^{\langle X,a\rangle}(y) = \mathbb{f}_{\mathrm{N}_{(0,\|a\|^2)}}(y) = \frac{1}{\sqrt{2\pi\|a\|^2}} \exp\Big(-\frac{y^2}{2\|a\|^2}\Big), \quad y \in \mathbb{R}.$$

§18.05 **Beweis** von Korollar §18.04. In der Vorlesung. □

§18.06 **Definition**. Ein $\mathbb{R}^n$-wertiger Zufallsvektor X besitzt eine *multivariate Normalverteilung* $\mathrm{N}_{(\mu,\Sigma)}$ mit $\mu \in \mathbb{R}^n$ und positiv semi-definiter Matrix $\Sigma \in \mathbb{R}^{n\times n}$, falls für alle $a \in \mathbb{R}^n$ die reelle Zufallsvariable $\langle X, a\rangle$ eine $\mathrm{N}_{(\langle\mu,a\rangle,\langle\Sigma a,a\rangle)}$-Verteilung besitzt. Das Produktmaß $\mathrm{N}_{(\mathbb{0}_n,\mathrm{I}_{n\times n})} = \bigotimes_{i\in[\![n]\!]} \mathrm{N}_{(0,1)} = \mathrm{N}^n_{(0,1)}$ heißt insbesondere *(n-dimensionale) Standardnormalverteilung*, wobei $\mathrm{I}_{n\times n}$ die n-dimensionale Einheitsmatrix und $\mathbb{0}_n \in \mathbb{R}^n$ der n-dimensionale Nullvektor ist. □

§18|00.07 **Vorbemerkung**. Für eine Matrix $\mathrm{A} = [a_{\bullet 1} \cdots a_{\bullet n}] \in \mathbb{R}^{n\times m}$ mit Spaltenvektoren $(a_{\bullet i})_{i\in[\![m]\!]}$ bezeichnet $\mathrm{Bild}(\mathrm{A}) = \langle a_{\bullet i} : i \in [\![m]\!]\rangle \subseteq \mathbb{R}^n$ die lineare Hülle der Spaltenvektoren, also das Bild der linearen Abbildung $\mathbb{R}^m \to \mathbb{R}^n$ mit $x \mapsto \mathrm{A}x$. Für einen linearen Unterraum $\mathcal{U} \subseteq \mathbb{R}^n$ bezeichnet $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$ die direkte orthogonale Summe, das heißt, $\mathcal{U}$ und $\mathcal{U}^\perp$ sind orthogonal, also für alle $u \in \mathcal{U}$ und $u^\perp \in \mathcal{U}^\perp$ gilt $\langle u, u^\perp\rangle = 0$, und jedes Element $x \in \mathbb{R}^n$ hat eine eindeutige Darstellung $x = u + u^\perp$ mit $u \in \mathcal{U}$ und $u^\perp \in \mathcal{U}^\perp$. Wir bezeichnen mit $\Pi_{\mathcal{U}}$ die Darstellungsmatrix der orthogonalen Projektion von $\mathbb{R}^n$ auf $\mathcal{U}$, also $\mathcal{U} \oplus \mathcal{U}^\perp \to \mathcal{U}$ mit $x = u + v \mapsto u = \Pi_{\mathcal{U}} x$. Eine Matrix $\mathrm{U} \in \mathbb{R}^{n\times m}$ heißt partielle Isometrie, falls $\mathrm{U}\mathrm{U}^T = \Pi_{\mathrm{Bild}(\mathrm{U})}$ und $\mathrm{U}^T\mathrm{U} = \Pi_{\mathrm{Bild}(\mathrm{U}^T)}$. □

§18.08 **Lemma**. *Seien $X \sim \mathrm{N}_{(\mathbb{0}_m,\mathrm{I}_{m\times m})}$ und $Y \sim \mathrm{N}_{(\mathbb{0}_k,\mathrm{I}_{k\times k})}$, dann gelten die folgenden Aussagen*

(i) *Für $\mathrm{A} \in \mathbb{R}^{n\times m}$ und $\mathrm{B} \in \mathbb{R}^{n\times k}$ mit $\mathrm{A}\mathrm{A}^T = \mathrm{B}\mathrm{B}^T$ sind die $\mathbb{R}^n$-wertigen Zufallsvektoren AX und BY identisch verteilt.*

(ii) *Falls $U \in \mathbb{R}^{n\times m}$ eine partielle Isometrie ist, dann gilt $UX \sim \mathrm{N}_{(\mathbb{0}_{n\times 1},\Pi_{\mathrm{Bild}(\mathrm{U})})}$.*

(iii) *Für $\mathrm{A} \in \mathbb{R}^{m\times n}$ und $\mathrm{B} \in \mathbb{R}^{m\times k}$ mit $\mathrm{A}^T\mathrm{B} = \mathbb{0}_{n\times k}$ sind $\Pi_{\mathrm{Bild}(\mathrm{A})}X \sim \mathrm{N}_{(\mathbb{0}_m,\Pi_{\mathrm{Bild}(\mathrm{A})})}$ und $\Pi_{\mathrm{Bild}(\mathrm{B})}X \sim \mathrm{N}_{(\mathbb{0}_m,\Pi_{\mathrm{Bild}(\mathrm{B})})}$ unabhängig.*

§18.09 **Beweis** von Lemma §18.08. In der Vorlesung. □

§18.10 **Korollar**. *Sei* $X = (X_i)_{i\in[\![n]\!]} \sim \mathrm{N}_{(\mu,\Sigma)}$ *mit* $\mu = (\mu_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$ *und* $\Sigma = [\Sigma_{ij}] \in \mathbb{R}^{n\times n}$ *positiv semi-definit, dann gelten die folgenden Aussagen:*

(i) *Für alle* $i \in [\![n]\!]$ *gilt* $X_i \sim \mathrm{N}_{(\mu_i,\Sigma_{ii})}$.

(ii) *Für alle* $i \in [\![n]\!]$ *und* $j \in [\![n]\!]_{\setminus i}$ *sind* X_i *und* X_j *genau dann unabhängig, wenn* $\Sigma_{ij} = 0$ *gilt.*

(iii) *Für* $\mathrm{A} \in \mathbb{R}^{m\times n}$ *und* $b \in \mathbb{R}^m$ *gilt* $Y = AX + b \sim \mathrm{N}_{(\mathrm{A}\mu+b,\mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}})}$.

(iv) *Ist* Σ *positiv definit, dann ist* X *stetig verteilt mit Lebesgue-Dichte*

$$f(x) = \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}\langle\Sigma^{-1}(x-\mu),(x-\mu)\rangle\right\}, \qquad x \in \mathbb{R}^n.$$

§18.11 **Beweis** von Korollar §18.10. Übungsaufgabe. □

§19 Klassifikation

§19.01 **Beispiel**. Im Beispiel I im Kapitel 1 Prolog stellt eine Schraubenherstellerin an zwei Produktionsstandorten Schrauben her. Die Schrauben werden an beiden Standorten in Kartons mit Beuteln a je 50 Schrauben ausgeliefert. Im Lager findet die Herstellerin eine unbeschriftete Lieferung. Die Herstellerin fragt sich, ob die Anzahl der beschädigten Schrauben in einem Beutel genügt, den Produktionsstandort der Lieferung zu bestimmen. Die Herstellerin geht zunächst von den (unrealistischen) Annahmen aus, dass der zufällige Produktionsstandort Y mit einem *Bernoulliverteilungsmodell* $(\{0,1\}, 2^{\{0,1\}}, \mathrm{B}_{\mathrm{q}})$ mit bekanntem Parameter $\mathrm{q} \in [0,1]$ und die zufällige Anzahl X der beschädigten Schrauben in einem Beutel am Standort $i \in \{0,1\}$ mit einem *Binomialverteilungsmodell* $([\![0,50]\!], 2^{[\![0,50]\!]}, \mathrm{Bin}_{(50,\mathrm{p}_i)})$ mit bekanntem Parameter $\mathrm{p}_i \in [0,1]$ adäquat beschrieben werden. Der zufällige Vektor (Y,X) nimmt damit Werte in $\{0,1\} \times [\![0,50]\!]$ an und wir bezeichen mit $(Y,X) \sim \mathbb{P}^{(Y,X)} \in \mathcal{W}(2^{\{0,1\}\times[\![0,50]\!]})$ die gemeinsame diskrete Verteilung mit gemeinsamer Zähldichte

$$\begin{aligned}\mathbb{p}^{(Y,X)}(y,x) = \mathbb{P}^{(Y,X)}(\{(y,x)\}) &= \mathbb{P}(Y=y, X=x)\\ &= \mathbb{P}(Y=y)\mathbb{P}(X=x\,|\,Y=y)\\ &= \mathrm{B}_{\mathrm{q}}(\{y\})\mathrm{Bin}_{(50,\mathrm{p}_y)}(\{x\})\end{aligned}$$

für alle $(y,x) \in \{0,1\} \times [\![0,50]\!]$. □

§19.02 **Definition**. Es sei $(\mathcal{X},\mathscr{X})$ ein messbarer Raum, $\mathcal{Y}$ eine endliche Menge mit $|\mathcal{Y}| \in \mathbb{N}$ *Labels* und (Y,X) ein Zufallsvektor auf einen Wahrscheinlichkeitsraum $(\Omega,\mathscr{A},\mathbb{P})$ mit Werten in der Menge $\mathcal{Y}\times\mathcal{X}$ sowie gemeinsamer Verteilung $\mathbb{P}^{(Y,X)} \in \mathcal{W}(2^{\mathcal{Y}}\otimes\mathscr{X})$. Die Vorhersage des *Labels* Y mit Hilfe des *Features* X wird *Klassifikationsproblem mit* $|\mathcal{Y}|$ *Labels* genannt. Eine Entscheidungsfunktion, d.h. $\mathscr{X}$-$2^{\mathcal{Y}}$-messbare Abbildung,

$$\mathfrak{l} : \mathcal{X} \ni \mathrm{x} \mapsto \mathfrak{l}(\mathrm{x}) \in \mathcal{Y},$$

also Entscheidungen nur anhand einer Stichprobe x aus dem Stichprobenraum $\mathcal{X}$, mit $\mathfrak{l}^{-1}(\{y\}) \in \mathscr{X}$ dem Ereignis aller Stichproben, die zu einer *Wahl des Labels* $y \in \mathcal{Y}$ führt, heißt *Klassifizierer*. Ein Klassifizierer $\mathfrak{l}_\star \in \mathcal{M}(\mathscr{X},2^{\mathcal{Y}})$ wird *Bayes-Klassifizierer* genannt, wenn

$$\mathbb{P}(Y \neq \mathfrak{l}_\star(X)) = \min\{\mathbb{P}(Y \neq \mathfrak{l}(X)) : \mathfrak{l} \in \mathcal{M}(\mathscr{X},2^{\mathcal{Y}})\}$$

gilt, also wenn es keinen Klassifizierer mit echt kleinerer Wahrscheinlichkeit einer Fehlklassifizierung gibt. □

§19.03 **Erinnerung**. Es sei $\mathcal{X}$ ein abzählbare Menge, $\mathcal{Y}$ eine endliche Menge mit $|\mathcal{Y}| \in \mathbb{N}$ Labels, (Y, X) ein Zufallsvektor mit Werten in der Menge $\mathcal{Y} \times \mathcal{X}$ und $\mathbb{P}^{(Y,X)} \in \mathcal{W}(2^{\mathcal{Y}\times\mathcal{X}})$ die gemeinsame diskrete Verteilung von (Y, X) mit gemeinsamer Zähldichte $\mathbb{p}^{(Y,X)} : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ und Randzähldichte $\mathbb{p}^X : \mathcal{X} \to [0, 1]$ der diskreten Randverteilung $\mathbb{P}^X \in \mathcal{W}(2^{\mathcal{X}})$ von X. Für jedes Atom $x \in \{\mathbb{p}^X > 0\} \subseteq \mathcal{X}$ von $\mathbb{P}^X$ und jedes Label $y \in \mathcal{Y}$ bezeichnet

$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}^{(Y,X)}(\{(y, x)\})}{\mathbb{P}^X(\{x\})} = \frac{\mathbb{p}^{(Y,X)}(y, x)}{\mathbb{p}^X(x)}$$

die bedingte Wahrscheinlichkeit des Ereignis $\{Y = y\}$ gegeben das Ereignis $\{X = x\}$. □

§19.04 **Lemma**. *Es sei $\mathcal{X}$ ein abzählbare Menge, $\mathcal{Y}$ eine endliche Menge mit $|\mathcal{Y}| \in \mathbb{N}$ Labels, (Y, X) ein $\mathcal{Y} \times \mathcal{X}$-wertiger Zufallsvektor mit gemeinsamer diskreter Verteilung $\mathbb{P}^{(Y,X)} \in \mathcal{W}(2^{\mathcal{Y}\times\mathcal{X}})$ und Randzähldichte $\mathbb{p}^X$ der Randverteilung $\mathbb{P}^X$ von X.*

(i) *Für jede Abbildung $\mathfrak{l} : \mathcal{X} \to \mathcal{Y}$ gilt*

$$\mathbb{P}(Y \neq \mathfrak{l}(X)) = \sum_{x\in\{\mathbb{p}^X>0\}} \mathbb{p}^X(x)(1 - \mathbb{P}(Y = \mathfrak{l}(x)|X = x)).$$

(ii) *Jede Abbildung $\mathfrak{l}_\star : \mathcal{X} \to \mathcal{Y}$, die für jedes Atom $x \in \{\mathbb{p}^X > 0\} \subseteq \mathcal{X}$ die Bedingung*

$$\mathbb{P}(Y = \mathfrak{l}_\star(x)|X = x) = \max\{\mathbb{P}(Y = y|X = x) : y \in \mathcal{Y}\}$$

erfüllt, ist ein Bayes-Klassifizierer.

§19.05 **Beweis** von Lemma §19.04. In der Vorlesung. □

§19.06 **Beispiel** *(Beispiel §19.01 fortgesetzt.)*. Betrachte $(Y, X) \sim \mathbb{P}^{(Y,X)} \in \mathcal{W}(2^{\{0,1\}\times[\![0,50]\!]})$ mit gemeinsamer Zähldichte $\mathbb{p}^{(Y,X)}(y, x) = \mathrm{B}_{\mathrm{q}}(\{y\})\mathrm{Bin}_{(50,\mathrm{p}_y)}(\{x\})$ für alle $(y, x) \in \{0, 1\} \times [\![0, 50]\!]$ und fest gewählten Parameter $\mathrm{q}, \mathrm{p}_0, \mathrm{p}_1 \in [0, 1]$. Die diskrete Randverteilung $\mathbb{P}^X \in \mathcal{W}(2^{[\![0,50]\!]})$ von X besitzt die Randzähldichte

$$\mathbb{p}^X = (1 - \mathrm{q})\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_0)}} + \mathrm{q}\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_1)}} : [\![0, 50]\!] \to [0, 1].$$

Für jedes Atom $x \in \{\mathbb{p}^X > 0\} \subseteq [\![0, 50]\!]$ und $y \in \{0, 1\}$ ist die bedingte Wahrscheinlichkeit des Ereignis $\{Y = y\}$ gegeben das Ereignis $\{X = x\}$ gegeben durch

$$\begin{aligned}\mathbb{P}(Y = y|X = x) &= \frac{\mathbb{P}^{(Y,X)}(\{(y, x)\})}{\mathbb{P}^X(\{x\})} = \frac{\mathbb{p}^{(Y,X)}(y, x)}{\mathbb{p}^X(x)}\\ &= \frac{\mathbb{p}_{\mathrm{B}_{\mathrm{q}}}(y)\,\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_y)}}(x)}{\mathbb{p}_{\mathrm{B}_{\mathrm{q}}}(0)\,\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_0)}}(x) + \mathbb{p}_{\mathrm{B}_{\mathrm{q}}}(1)\,\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_1)}}(x)}\\ &= \frac{\mathrm{q}^y(1 - \mathrm{q})^{1-y}\mathrm{p}_y^x(1 - \mathrm{p}_y)^{50-x}}{(1 - \mathrm{q})\mathrm{p}_0^x(1 - \mathrm{p}_0)^{50-x} + \mathrm{q}\mathrm{p}_1^x(1 - \mathrm{p}_1)^{50-x}}.\end{aligned}$$

Damit ist jede Abbildung $\mathfrak{l}_\star : [\![0, 50]\!] \to \{0, 1\}$, die die Bedingungen

$$\{(1 - \mathrm{q})\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_0)}} > \mathrm{q}\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_1)}}\} \subseteq \mathfrak{l}_\star^{-1}(\{0\}) \quad \text{und} \quad \{(1 - \mathrm{q})\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_0)}} < \mathrm{q}\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_1)}}\} \subseteq \mathfrak{l}_\star^{-1}(\{1\})$$

erfüllt, ein Bayes-Klassifizierer mit Wahrscheinlichkeit einer Fehlklassifizierung

$$\begin{aligned}\mathbb{P}(Y \neq \mathfrak{l}_\star(X)) &= \sum_{x\in\{\mathbb{p}^X>0\}} \mathbb{p}^X(x)\big(1 - \mathbb{P}(Y = \mathfrak{l}_\star(x)|X = x)\big)\\ &= \sum_{x\in\{\mathbb{p}^X>0\}} \mathbb{p}^X(x)\big(\mathbb{P}(Y = 0|X = x) \wedge \mathbb{P}(Y = 1|X = x)\big)\\ &= \tfrac{1}{2}\Big(1 - \sum_{x\in\{\mathbb{p}^X>0\}} \mathbb{p}^X(x)\big|\mathbb{P}(Y = 0|X = x) - \mathbb{P}(Y = 1|X = x)\big|\Big) \in [0, \tfrac{1}{2}].\end{aligned}$$

Im Spezialfall

$\mathrm{q} \in (0,1)$ und $\mathrm{p}_0 = 0 = 1 - \mathrm{p}_1$ erhalten wir $\{\mathbb{p}^X > 0\} = \{0, 50\}$ sowie

$$\mathbb{P}(Y = 0|X = 0) = 1 = \mathbb{P}(Y = 1|X = 50).$$

Damit ist jede Abbildung $\mathfrak{l}_\star : [\![0, 50]\!] \to \{0, 1\}$ mit $\mathfrak{l}_\star(0) = 0$ und $\mathfrak{l}_\star(50) = 1$ ein Bayes-Klassifizierer mit Wahrscheinlichkeit $\mathbb{P}(Y \neq \mathfrak{l}_\star(X)) = 0$ einer Fehlklassifizierung.

$\mathrm{p}_0 = \mathrm{p}_1$ erhalten wir $\mathbb{p}^X = \mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_0)}}$ sowie für jedes Atom $x \in \{\mathbb{p}_{\mathrm{Bin}_{(50,\mathrm{p}_0)}} > 0\} \subseteq [\![0, 50]\!]$ und $y \in \{0, 1\}$

$$\mathbb{P}(Y = y|X = x) = \mathrm{q}^y(1 - \mathrm{q})^{1-y}.$$

Damit ist die konstante Abbildung $\mathfrak{l}_\star : [\![0, 50]\!] \to \{0, 1\}$ mit

$$x \mapsto \mathfrak{l}_\star(x) := \begin{cases} 1 & \text{falls } \mathrm{q} \geqslant (1 - \mathrm{q}) \\ 0 & \text{falls } \mathrm{q} < (1 - \mathrm{q}) \end{cases}$$

ein Bayes-Klassifizierer mit Wahrscheinlichkeit $\mathbb{P}(Y \neq \mathfrak{l}_\star(X)) = \mathrm{q} \wedge (1 - \mathrm{q}) \in [0, \frac{1}{2}]$ einer Fehlklassifizierung. □

§20 Beispiele statistischer Modelle

§20.01 **Definition**. Sei $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ ein adäquates statistisches Experiment für eine $(\mathcal{X}, \mathscr{X})$-wertige Zufallsvariable X und sei $(X_i)_{i \in \mathcal{I}}$ eine Familie von $(\mathcal{X}, \mathscr{X})$-wertige Zufallsvariablen mit nicht-leerer abzählbarer Indexmenge $\mathcal{I}$. $(X_i)_{i \in \mathcal{I}}$ werden *unabhängige und identisch-verteilte (u.i.v.) Kopien* von X genannt, wenn für $\theta \in \Theta$ mit $X \sim \mathbb{P}_\theta$ das Produktmaß $\mathbb{P}_\theta^{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathbb{P}_\theta$ auf dem Produktraum $(\mathcal{X}^{\mathcal{I}}, \mathscr{X}^{\mathcal{I}})$ die gemeinsame Verteilung von $(X_i)_{i \in \mathcal{I}}$ ist. In diesem Fall ist das statistische Produktexperiment $(\mathcal{X}^{\mathcal{I}}, \mathscr{X}^{\mathcal{I}}, \mathbb{P}_\Theta^{\mathcal{I}})$ also adäquat für $(X_i)_{i \in \mathcal{I}}$. $(\mathcal{X}^{\mathcal{I}}, \mathscr{X}^{\mathcal{I}}, \mathbb{P}_\Theta^{\mathcal{I}})$ wird auch adäquates *statistisches u.i.v. Modell* für $(X_i)_{i \in \mathcal{I}}$ genannt. □

§20.02 **Schreibweise**. $(X_i)_{i \in \mathcal{I}} \odot \mathbb{P}_\Theta^{\mathcal{I}}$ und im Fall $\mathcal{I} = [\![n]\!]$ auch $(X_i)_{i \in [\![n]\!]} \odot \mathbb{P}_\Theta^n$. □

§20.03 **Beispiel**. Im Folgenden geben wir statistische Modelle für die im Kapitel 1 Prolog vorgestellten Beispiele an.

(a) Beispiel I: Setzen wir Eins für das weibliche Geschlecht und Null für das nicht weibliche Geschlecht eines Konsumierenden, so beschreiben wir das Geschlecht der $n = 1000$ befragten Konsumierenden als Stichprobe eines *Bernoulli-Schemas* $\big(\{0,1\}^n, 2^{\{0,1\}^n}, (\mathrm{B}_\mathrm{p}^n)_{p \in [0,1]}\big)$, vgl. Beispiel §04.08 (c). Sei $(X_i)_{i \in [\![n]\!]} \odot (\mathrm{B}_\mathrm{p}^n)_{p \in [0,1]}$. Nach Beispiel §17.09 (a) ist dann ein *Binomialverteilungsmodell* $\big([\![0, n]\!], 2^{[\![0,n]\!]}, (\mathrm{Bin}_{(n,p)})_{p \in [0,1]}\big)$ ein adäquates statistisches Experiment für die Anzahl $\sum_{i \in [\![n]\!]} X_i$ der Konsumentinnen unter den befragten Konsumierenden.

(b) Beispiel II: Die zufällige Anzahl beschädigter Schrauben in den $n = 100$ Beuteln mit 50 Schrauben beschreiben wir wie in Beispiel §04.08 (d) durch ein *Binomialverteilungsmodell* $\big([\![0, 50]\!]^n, 2^{[\![0,50]\!]^n}, (\mathrm{Bin}_{(50,p)}^n)_{p \in [0,1]}\big)$. Sei $(X_i)_{i \in [\![n]\!]} \odot (\mathrm{Bin}_{(50,p)}^n)_{p \in [0,1]}$. Nach §17.09 (a) ist dann ein *Binomialverteilungsmodell* $\big([\![0, 50n]\!], 2^{[\![0,50n]\!]}, (\mathrm{Bin}_{(50n,p)})_{p \in [0,1]}\big)$ ein adäquates statistisches Experiment für die Gesamtanzahl $\sum_{i \in [\![n]\!]} X_i$ an beschädigten Schrauben.

(c) Beispiel III: Die $n = 280$ zufälligen Anzahlen eingegangener Anrufe innerhalb einer Woche beschreiben wir durch ein *Poissonverteilungsmodell* $\big(\mathbb{Z}_{\geqslant 0}^n, 2^{\mathbb{Z}_{\geqslant 0}^n}, (\mathrm{Poi}_\lambda^n)_{\lambda \in \mathbb{R}_{>0}}\big)$, vgl. Beispiel §04.08 (f). Sei $(X_i)_{i \in [\![n]\!]} \odot (\mathrm{Poi}_\lambda^n)_{\lambda \in \mathbb{R}_{>0}}$. Nach Beispiel §17.09 (b) ist dann ein *Poissonverteilungsmodell* $\big(\mathbb{Z}_{\geqslant 0}, 2^{\mathbb{Z}_{\geqslant 0}}, (\mathrm{Poi}_{n\lambda})_{\lambda \in \mathbb{R}_{>0}}\big)$ ein adäquates statistisches Experiment für die Gesamtanzahl $\sum_{i \in [\![n]\!]} X_i$ eingegangener Anrufe.

(d) Beispiel IV: Die zufälligen Wartezeiten an der U-Bahn Haltestelle an den $n = 90$ Tag (in Minuten) beschreiben wir durch ein *Uniformverteilungsmodell* $\big(\mathbb{R}^n, \mathscr{B}^n, (\mathrm{U}^n_{[0,\theta]})_{\theta\in\mathbb{R}_{>0}}\big)$, vgl. Beispiel §05.07 (a). Sei $(X_i)_{i\in[\![n]\!]} \circledsim (\mathrm{U}^n_{[0,\theta]})_{\theta\in\mathbb{R}_{>0}}$. Bezeichne für $\theta \in \mathbb{R}_{>0}$ mit $\mathbb{P}_\theta$ das stetige Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B})$ mit Wahrscheinlichkeitsdichte $\mathbb{f}_\theta(x) = \frac{nx^{n-1}}{\theta^n}\mathbb{1}_{[0,\theta]}(x)$, $x \in \mathbb{R}$. Das stetige statistische Experiment $\big(\mathbb{R}, \mathscr{B}, (\mathbb{P}_\theta)_{\theta\in\mathbb{R}_{>0}}\big)$ ist dann adäquat für die maximale Wartezeit $\max\{X_i : i \in [\![n]\!]\}$.

(e) Beispiel V: Die zufällige Lebensdauer der $n = 100$ Glühlampen (in Stunden) beschreiben wir durch ein *Exponentialverteilungsmodell* $\big(\mathbb{R}^n, \mathscr{B}^n, (\mathrm{Exp}^n_\lambda)_{\lambda\in\mathbb{R}_{>0}}\big)$, vgl. Beispiel §05.07 (b). Sei $(X_i)_{i\in[\![n]\!]} \circledsim (\mathrm{Exp}^n_\lambda)_{\lambda\in\mathbb{R}_{>0}}$. Nach Beispiel §17.12 ist dann ein *Gammaverteilungsmodell* $\big(\mathbb{R}, \mathscr{B}, (\Gamma_{(\lambda,n)})_{\lambda\in\mathbb{R}_{>0}}\big)$ ein adäquates statistisches Experiment für die kumulierte Lebensdauer $\sum_{i\in[\![n]\!]} X_i$ der Glühlampen.

(f) Beispiel VI: Die zufälligen Messwerte beschreiben wir durch ein *Normalverteilungsmodell* $\big(\mathbb{R}^n, \mathscr{B}^n, (\mathrm{N}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}\big)$, vgl. Beispiel §05.07 (c). Ein *Normalverteilungsmodell* $\big(\mathbb{R}, \mathscr{B}, (\mathrm{N}_{(n\mu,n\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}\big)$ ist dann nach Beispiel §17.13 ein adäquates statistisches Experiment für den kumulierten Messwert $\sum_{i\in[\![n]\!]} X_i$ und nach Beispiel §10.12 (a) für den mittleren Messwert $\frac{1}{n}\sum_{i\in[\![n]\!]} X_i = \overline{X}_n$ ist $\big(\mathbb{R}, \mathscr{B}, (\mathrm{N}_{(\mu,\sigma^2/n)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}\big)$ adäquat. □

Kapitel 5

Erwartungswert

§21 Positive numerische Zufallsvariablen

§21.01 **Satz.** *Für jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf einem messbarem Raum $(\Omega, \mathscr{A})$ existiert ein eindeutig bestimmtes Funktional* $\boxed{\mathbb{E}}$ $: \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ *derart, dass die folgenden Bedingungen erfüllt sind:*

(i) *Für alle $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ und $a, b \in \overline{\mathbb{R}}_{\geqslant 0}$ gilt $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$;* *(linear)*

(ii) *Für alle $(X_n)_{n\in\mathbb{N}} \uparrow X$ in $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ gilt $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$;* *(monoton konvergent)*

(iii) *Für jedes $A \in \mathscr{A}$ gilt $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.* *(normiert)*

Das Funktional $\mathbb{E}$ wird Erwartung bzgl. $\mathbb{P}$ genannt und für jedes $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ heißt $\boxed{\mathbb{E}(X)}$ *der Erwartungswert von X.*

§21.02 **Beweis** von Satz §21.01. Der Satz fasst die Hauptaussage dieses Abschnittes zusammen, der Beweis der Aussage erfolgt in mehreren Schritten. Wir zeigen zuerst in Satz §21.04 die Eindeutigkeitsaussage, und geben dann in Definition §21.06 ein Funktional $\widetilde{\mathbb{E}} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ explizit an, für das wir in mehreren Schritten die Bedingungen (i)-(iii) nachweisen. Zusammenfassend zeigen wir damit dann in Satz §21.10 auch die Existenzaussage. □

§21.03 **Bemerkung.** Der Beweis der nächsten Aussage wendet die Beweisstrategie §09.06 an. □

§21.04 **Eindeutigkeitssatz.** *Die Erwartung ist eindeutig bestimmt.*

§21.05 **Beweis** von Satz §21.04. In der Vorlesung. □

Im Folgenden sei stets $\mathbb{P} \in \mathcal{W}(\mathscr{A})$ ein Wahrscheinlichkeitsmaß auf einem messbaren Raum $(\Omega, \mathscr{A})$. Eine messbare Partition $\mathcal{P} := \{A_i : i \in \mathcal{I}\} \subseteq \mathscr{A}_{\neq\varnothing}$ von Ω nennen wir endlich, wenn $|\mathcal{I}| \in \mathbb{N}$ gilt, also die nicht-leere Indexmenge $\mathcal{I}$ endlich ist. Für jedes $A \in \mathcal{P}$ gilt somit $A \in \mathscr{A}_{\neq\varnothing}$. Für den weiteren Verlauf setzen wir

$$\mathscr{P} := \{\mathcal{P} \subseteq \mathscr{A}_{\neq\varnothing} : \mathcal{P} \text{ endliche, messbare Partition von } \Omega\}.$$

§21.06 **Definition.** Sei $\boxed{\widetilde{\mathbb{E}}}$ $: \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ mit

$$X \mapsto \widetilde{\mathbb{E}}(X) := \sup\Big\{ \sum_{\substack{A\in\mathcal{P}:\\ \mathbb{P}(A)\in\mathbb{R}_{>0}}} \big(\inf_{\omega\in A} X(w)\big)\mathbb{P}(A) : \mathcal{P} \in \mathscr{P}\Big\}.$$

□

§21.07 **Erinnerung.** Für $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ mit $X(\omega) \leqslant Y(\omega)$ für alle $\omega \in \Omega$ schreiben wir $X \leqslant Y$. □

§21.08 **Lemma.**

(i) *Sei $X \in \mathcal{M}^{\text{einf}}_{\geqslant 0}(\mathscr{A})$ eine positive einfache Zufallsvariable mit Darstellung $X = \sum_{j\in\mathcal{J}} x_j \mathbb{1}_{A_j}$, wobei $\{A_j : j \in \mathcal{J}\} \in \mathscr{P}$ eine endliche, messbare Partition von Ω ist. Dann gilt*

$$\widetilde{\mathbb{E}}(X) = \sum_{j\in\mathcal{J}} x_j \mathbb{P}(A_j).$$

Insbesondere, erfüllt $\widetilde{\mathbb{E}}$ also die Bedingung Satz §21.01 (iii). *(normiert)*

(ii) *Für alle $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ mit $X \leqslant Y$ gilt $\widetilde{\mathbb{E}}(X) \leqslant \widetilde{\mathbb{E}}(Y)$.* *(monoton)*

(iii) *Für alle $(X_n)_{n\in\mathbb{N}} \uparrow X$ in $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ gilt $\widetilde{\mathbb{E}}(X_n) \uparrow \widetilde{\mathbb{E}}(X)$.* *(monoton konvergent)*
$\widetilde{\mathbb{E}}$ erfüllt also die Bedingung Satz §21.01 (ii).

(iv) *Für alle $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ und $a, b \in \overline{\mathbb{R}}_{\geqslant 0}$ gilt $\widetilde{\mathbb{E}}(aX + bY) = a\widetilde{\mathbb{E}}(X) + b\widetilde{\mathbb{E}}(Y)$.* *(linear)*
$\widetilde{\mathbb{E}}$ erfüllt die Bedingung Satz §21.01 (i).

§21.09 **Beweis** von Lemma §21.08. In der Vorlesung. □

§21.10 **Existenzsatz**. *Das Funktional $\widetilde{\mathbb{E}} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ in Definition §21.06 ist eine Erwartung bzgl. $\mathbb{P}$, das heißt, es erfüllt die Bedingungen (i)-(iii) aus Satz §21.01.*

§21.11 **Beweis** von Satz §21.10. Die Aussage folgt direkt aus Lemma §21.08 (i), (iii) und (iv). □

Damit haben wir Satz §21.01 nachgewiesen, so dass die Erwartung $\mathbb{E}$ eindeutig ist und durch die explizite Form in Definition §21.06 gegeben ist.

§21.12 **Schreibweise**. Wenn es uns wichtig ist, den Stichprobenraum und das Wahrscheinlichkeitsmaß zu betonen, so schreiben wir auch

$$\mathbb{P}(X) = \mathbb{E}(X) = \int_\Omega X\, d\mathbb{P} = \int_\Omega X(\omega)\mathbb{P}(d\omega) \quad \text{sowie}$$

$$\mathbb{P}(X\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A) = \int_\Omega \mathbb{1}_A(\omega)X(\omega)\mathbb{P}(d\omega) = \int_A X\, d\mathbb{P} \quad \text{für } A \in \mathscr{A}.$$

Die Begründung für diese Schreibweise wird in der Maßtheorie gegeben, in der gezeigt wird, dass der Erwartungswert von X gerade das Lebesgue-Integral von X bzgl. des Wahrscheinlichkeitsmaßes $\mathbb{P}$ ist. □

§21|01 Weitere Eigenschaften der Erwartung

§21.13 **Lemma**.

(i) *Für $X \in \overline{\mathcal{M}}_{\geqslant 0}$ gilt $\mathbb{P}(X = 0) = 1$ genau dann, wenn $\mathbb{E}(X) = 0$ gilt.*

(ii) *Sei $X \in \overline{\mathcal{M}}_{\geqslant 0}$ mit $\mathbb{E}(X) \in \mathbb{R}_{\geqslant 0}$. Dann gilt $\mathbb{P}(X = \infty) = 0$.* *(endlich)*

(iii) *Für $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ gilt $\mathbb{P}(X \leqslant Y) = 1$ genau dann, wenn $\mathbb{E}(X\mathbb{1}_A) \leqslant \mathbb{E}(Y\mathbb{1}_A)$ für alle $A \in \mathscr{A}$ gilt. Insbesondere, aus $\mathbb{P}(X \leqslant Y) = 1$ folgt $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$.*

(iv) *Für $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ gilt $\mathbb{P}(X = Y) = 1$ genau dann, wenn $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$ für alle $A \in \mathscr{A}$ gilt.*

§21.14 **Beweis** von Lemma §21.13. In der Vorlesung. □

§21.15 **Schreibweise**. Sei $X \in \overline{\mathcal{M}}_{\geqslant 0}$. Für $\mathbb{P}(X = 0) = 1$ schreiben wir auch kurz $X = 0$ $\mathbb{P}$-f.s. (Sprechweise: $\mathbb{P}$ *fast sicher*), oder $X = 0$ $\mathbb{P}$-f.ü. (Sprechweise: $\mathbb{P}$ *fast überall*). Für $\mathbb{P}(X = \infty) = 0$ schreiben wir auch $X \in \mathbb{R}_{\geqslant 0}$ $\mathbb{P}$-f.s.. Falls $\mathbb{P}$ aus dem Kontext klar ist, wird dies häufig auch einfach weggelassen. □

§21.16 **Lemma von Fatou**. *Für jede Folge $(X_n)_{n\in\mathbb{N}}$ in $\overline{\mathcal{M}}_{\geqslant 0}$ gilt*

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leqslant \liminf_{n\to\infty} \mathbb{E}(X_n).$$

§21.17 **Beweis** von Lemma §21.16. In der Vorlesung. □

§21.18 **Beispiel**. Auf dem Wahrscheinlichkeitsraum $(\mathbb{R}, \mathscr{B}, \mathrm{U}_{[0,1]})$ betrachte für $n \in \mathbb{N}$ die reelle Zufallsvariable $X_n := n\mathbb{1}_{(0,1/n]} \in \mathcal{M}(\mathscr{B})$. Für jedes $n \in \mathbb{N}$ gilt dann $\mathrm{U}_{[0,1]}(X_n) = n\mathrm{U}_{[0,1]}((0, \frac{1}{n}]) = 1$ und somit $\liminf_{n\to\infty} \mathrm{U}_{[0,1]}(X_n) = 1$. Andererseits, für jedes $\omega \in \mathbb{R}$ ist $\lim_{n\to\infty} X_n(\omega) = 0$, sodass $\mathrm{U}_{[0,1]}(\liminf_{n\to\infty} X_n) = 0 < 1 = \liminf_{n\to\infty} \mathrm{U}_{[0,1]}(X_n)$ gilt. □

§21.19 **Lemma**. *Für jede Folge $(X_n)_{n\in\mathbb{N}}$ in $\overline{\mathcal{M}}_{\geqslant 0}$ gilt $\sum_{n\in\mathbb{N}} X_n \in \overline{\mathcal{M}}_{\geqslant 0}$ und*

$$\mathbb{E}\Big(\sum_{n\in\mathbb{N}} X_n\Big) = \sum_{n\in\mathbb{N}} \mathbb{E}(X_n).$$

§21.20 **Beweis** von Lemma §21.19. Übungsaufgabe unter Verwendung von Satz §21.01 (i) (linear) und (ii) (monoton konvergent). □

§21.21 **Lemma**. *Eine Familie $(\mathscr{A}_i)_{i\in\mathcal{I}}$ von Teil-σ-Algebren aus $\mathscr{A}$ mit beliebiger nicht-leerer Indexmenge $\mathcal{I}$ ist unabhängig, also $\perp\!\!\!\perp_{i\in\mathcal{I}}\mathscr{A}_i$, genau dann, wenn für jede Familie $(X_i)_{i\in\mathcal{I}}$ positiver numerischer Zufallsvariablen mit $X_i \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}_i)$, $i \in \mathcal{I}$, und jede endliche nicht-leere Teilmenge $\mathcal{J} \subseteq \mathcal{I}$ gilt $\mathbb{E}(\prod_{j\in\mathcal{J}} X_j) = \prod_{j\in\mathcal{J}} \mathbb{E}(X_j)$.*

§21.22 **Beweis** von Lemma §21.21. Übungsaufgabe unter Verwendung der Beweisstrategie §09.06. □

§22 Integrierbare Zufallsvariablen

§22.01 **Vorbemerkung**. Sei $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X \in \overline{\mathcal{M}}(\mathscr{A})$ eine numerische Zufallsvariable. Dann definieren wir den Erwartungswert von X mit Hilfe der Zerlegung von $X = X^+ - X^-$ in die positiven numerischen Zufallsvariablen $X^+ = X \vee 0 \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ und $X^- = (-X) \vee 0 \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. Zur Erinnerung, es gilt: $0 \leqslant X^+$, $0 \leqslant X^-$, $|X| = X^+ + X^-$ sowie $X^+X^- = 0$. □

§22.02 **Definition**. Seien $(\Omega, \mathscr{A})$ ein messbarer Raum, $\mathbb{P} \in \mathcal{W}(\mathscr{A})$ ein Wahrscheinlichkeitsmaß und $X \in \overline{\mathcal{M}}(\mathscr{A})$ eine numerische Zufallsvariable.

(a) Ist höchstens einer der beiden Erwartungswerte $\mathbb{E}(X^+)$ und $\mathbb{E}(X^-)$ nicht endlich, das heißt, $\mathbb{E}(X^+) \wedge \mathbb{E}(X^-) \in \mathbb{R}_{\geqslant 0}$, so definiert $\boxed{\mathbb{E}(X)} := \mathbb{E}(X^+) - \mathbb{E}(X^-)$ den *Erwartungswert* von X mit den üblichen Rechenregeln $\infty + x = \infty$ und $-\infty + x = -\infty$ für alle $x \in \mathbb{R}$. In diesem Fall wird X *$\mathbb{P}$-quasiintegrierbar* genannt. Der Erwartungswert von X ist nicht definiert, wenn $\mathbb{E}(X^+) = \infty = \mathbb{E}(X^-)$ gilt.

(b) Falls $\mathbb{E}(|X|) \in \mathbb{R}_{\geqslant 0}$, also falls $\mathbb{E}(X^+) \in \mathbb{R}_{\geqslant 0}$ und $\mathbb{E}(X^-) \in \mathbb{R}_{\geqslant 0}$, gilt, dann heißt X *$\mathbb{P}$-integrierbar*. Die Menge aller $\mathbb{P}$-integrierbaren numerischen Zufallsvariablen bezeichnen wir mit $\boxed{\mathcal{L}_1} := \boxed{\mathcal{L}_1(\mathbb{P})} := \mathcal{L}_1(\Omega, \mathscr{A}, \mathbb{P}) := \{X \in \overline{\mathcal{M}}(\mathscr{A}) : \mathbb{E}(|X|) \in \mathbb{R}_{\geqslant 0}\}$. □

Sei $\mathrm{id}_\Omega : \Omega \to \Omega$ mit $\omega \mapsto \mathrm{id}_\Omega(\omega) := \omega$. Das Wahrscheinlichkeitsmaß $\mathbb{P}$ besitzt ein *endliches erstes Moment*, wenn $\mathrm{id}_\Omega \in \mathcal{L}_1(\Omega, \mathscr{A}, \mathbb{P})$ gilt. In diesem Fall wird $\mathbb{E}(\mathrm{id}_\Omega) = \mathbb{P}(\mathrm{id}_\Omega)$ der *Erwartungswert des Wahrscheinlichkeitsmaßes $\mathbb{P}$* genannt. Wir bezeichnen mit $\mathcal{W}_1(\mathscr{A})$ die Menge aller Wahrscheinlichkeitsmaße auf $(\Omega, \mathscr{A})$, die ein endliches erstes Moment besitzen. □

§22.03 **Bemerkung**. Sei $X \in \mathcal{L}_1(\mathbb{P})$. Auf Grund der Endlichkeit in Lemma §21.13 (ii) gilt somit auch $\mathbb{P}(|X| \in \mathbb{R}_{\geqslant 0}) = \mathbb{P}(X \in \mathbb{R}) = 1$, das heißt, X ist fast sicher gleich der reellen Zufallsvariablen $X\mathbb{1}_{\mathbb{R}}(X) \in \mathcal{M}(\mathscr{A})$ (mit der Konvention $\infty \times 0 = 0 = -\infty \times 0$). Wie in Schreibweise §21.12 schreiben wir den Erwartungswert von X auch als Lebesgue-Integral von X bzgl. des Wahrscheinlichkeitsmaßes $\mathbb{P}$, das heißt, $\mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(d\omega)$. In der Maßtheorie werden weiterhin die folgenden zwei Darstellungen gezeigt:

(a) Sei $\mathbb{P}$ ein *stetiges* Wahrscheinlichkeitsmaß auf $\mathbb{R}^n$ mit Dichte $\mathbb{f}$. Dann gilt $X \in \mathcal{L}_1(\mathbb{R}^n, \mathscr{B}^n, \mathbb{P})$ genau dann, wenn $\int_{\mathbb{R}^n} |X(\omega)|\mathbb{f}(\omega)d\omega \in \mathbb{R}_{\geqslant 0}$ gilt. In diesem Fall ist

$$\mathbb{E}(X) = \mathbb{P}(X) = \int_{\mathbb{R}^n} X(\omega)\mathbb{f}(\omega)d\omega.$$

(b) Sei $\mathbb{P}$ ein *diskretes* Wahrscheinlichkeitsmaß mit Zähldichte $\mathbb{p}$. Dann gilt $X \in \mathcal{L}_1(\Omega, \mathscr{A}, \mathbb{P})$ genau dann wenn $\sum_{\omega\in\Omega} |X(\omega)|\mathbb{p}(\omega) \in \mathbb{R}_{\geqslant 0}$ gilt. In diesem Fall ist

$$\mathbb{E}(X) = \mathbb{P}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{p}(\omega).$$

□

§22.04 **Proposition**. *$\mathcal{L}_1$ ist ein Vektorraum. Für das Funktional $\mathbb{E} : \mathcal{L}_1 \to \mathbb{R}$ mit $X \mapsto \mathbb{E}(X)$ gilt:*

(i) *Für alle $X, Y \in \mathcal{L}_1$ und $a, b \in \mathbb{R}$ gilt $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$;* (linear)

(ii) *Für alle $X \in \mathcal{L}_1$ mit $X \geqslant 0$ gilt $\mathbb{E}(X) \geqslant 0$;* (positiv)

(iii) *Für alle $X \in \mathcal{L}_1$ gilt $|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|)$;*

(iv) *Für alle $X, Y \in \mathcal{L}_1$ mit $X \leqslant Y$ gilt $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$;* (monoton)

(v) *Für alle $X \in \mathcal{M}$ mit $\sup_{\omega\in\Omega}|X(\omega)| \in \mathbb{R}_{\geqslant 0}$ gilt $X \in \mathcal{L}_1$. In diesem Fall heißt X* beschränkt.

(vi) *Seien $X \in \mathcal{L}_1$ und $Y \in \overline{\mathcal{M}}$ mit $\mathbb{P}(X = Y) = 1$, dann gilt $Y \in \mathcal{L}_1$ und $\mathbb{E}(X) = \mathbb{E}(Y)$.*

(vii) *Falls $X, Y \in \mathcal{L}_1$ unabhängig sind, so gilt $XY \in \mathcal{L}_1$ und $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

§22.05 **Beweis** von Proposition §22.04. In der Vorlesung sowie (ii)-(vi) Übungsaufgabe, wobei für (vi) zuerst gezeigt werden sollte, dass $\mathbb{E}(|X - Y|) = 0$ gilt. □

§22.06 **Beispiel**.

(a) Betrachte $(\mathbb{Z}_{\geqslant 0}, 2^{\mathbb{Z}_{\geqslant 0}}, \mathrm{Poi}_\lambda)$ mit $\lambda \in \mathbb{R}_{>0}$ und Zähldichte $\mathbb{p}_{\mathrm{Poi}_\lambda}(k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k \in \mathbb{Z}_{\geqslant 0}$ (vgl. Beispiel §04.08 (f)). Dann ist $\mathrm{id}_{\mathbb{Z}_{\geqslant 0}} \in \mathcal{L}_1(\mathrm{Poi}_\lambda)$ mit

$$\mathrm{Poi}_\lambda(\mathrm{id}_{\mathbb{Z}_{\geqslant 0}}) = \sum_{k\in\mathbb{Z}_{\geqslant 0}} k\mathbb{p}_{\mathrm{Poi}_\lambda}(k) = \sum_{k\in\mathbb{Z}_{\geqslant 0}} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda \sum_{k\in\mathbb{N}} \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda \sum_{k\in\mathbb{Z}_{\geqslant 0}} \mathbb{p}_{\mathrm{Poi}_\lambda}(k) = \lambda.$$

(b) Betrachte ein Bernoulli-Schema $\left(\{0,1\}^n, 2^{\{0,1\}^n}, \mathrm{B}_p^n\right)$ mit $n \in \mathbb{N}$ und $p \in [0,1]$. Für jedes $i \in [\![n]\!]$ ist die Koordinatenprojektion $\Pi_i : \{0,1\}^n \to \{0,1\}$ B_p-verteilt, also $\Pi_i \sim \mathrm{B}_p$ mit $\mathrm{B}_p^n(\Pi_i = 1) = p$ und $\mathrm{B}_p^n(\Pi_i = 0) = 1 - p$, so dass $\Pi_i \in \mathcal{L}_1(\mathrm{B}_p^n)$ mit $\mathrm{B}_p^n(\Pi_i) = p$ gilt. Offensichtlich, folgt somit $\sum_{i\in[\![n]\!]} \Pi_i \in \mathcal{L}_1(\mathrm{B}_p^n)$ und $\mathrm{B}_p^n(\sum_{i\in[\![n]\!]} \Pi_i) = \sum_{i\in[\![n]\!]} \mathrm{B}_p^n(\Pi_i) = np$.

(c) Auf Grund der symmetrischen Dichte und $\mathbb{f}'_{\mathrm{N}_{(0,1)}}(x) = -x\mathbb{f}_{\mathrm{N}_{(0,1)}}(x)$ gilt

$$\mathrm{N}_{(0,1)}(\mathrm{id}_{\mathbb{R}}^+) = \mathrm{N}_{(0,1)}(\mathrm{id}_{\mathbb{R}}^-) = \int_{\mathbb{R}_{\geqslant 0}} x\mathbb{f}_{\mathrm{N}_{(0,1)}}(x)dx = \int_{\mathbb{R}_{\geqslant 0}} -\mathbb{f}'_{\mathrm{N}_{(0,1)}}(x)dx = \tfrac{1}{\sqrt{2\pi}},$$

also $\mathrm{id}_{\mathbb{R}} \in \mathcal{L}_1(\mathbb{R}, \mathscr{B}, \mathrm{N}_{(0,1)})$ und $\mathrm{N}_{(0,1)}(\mathrm{id}_{\mathbb{R}}) = 0$. □

§22|01 Weitere Eigenschaften der Erwartung

§22.07 **Lemma**. *Für $X \in \mathcal{L}_1(\mathbb{P})$ gilt $\mathbb{P}(X = 0) = 1$ genau dann, wenn $\mathbb{E}(X\mathbb{1}_A) = 0$ für alle $A \in \mathscr{A}$.*

§22.08 **Beweis** von Lemma §22.07. In der Vorlesung. □

§22.09 **Lemma** *(Stetigkeit)*. *Sei $X \in \mathcal{L}_1(\mathbb{P})$. Dann gilt*

$$\forall \varepsilon \in \mathbb{R}_{>0} \;:\; \exists \delta_\varepsilon \in \mathbb{R}_{>0} \;:\; \forall A \in \mathscr{A} \;:\; \mathbb{P}(A) \leqslant \delta_\varepsilon \Rightarrow \mathbb{E}(|X|\mathbb{1}_A) \leqslant \varepsilon.$$

Insbesondere, für alle $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{A}$ mit $\lim_{n\to\infty} \mathbb{P}(A_n) = 0$ gilt $\lim_{n\to\infty} \mathbb{E}(|X|\mathbb{1}_{A_n}) = 0$.

§22.10 **Beweis** von Lemma §22.09. In der Vorlesung. □

§22.11 **Erinnerung**. Eine Funktion $\phi : \mathbb{R} \to \mathbb{R}$ heißt *konvex*, wenn

$$\phi\big(\lambda x + (1-\lambda)y\big) \leqslant \lambda\phi(x) + (1-\lambda)\phi(y)$$

für alle $x, y \in \mathbb{R}$ und für alle $\lambda \in [0,1]$ gilt. □

§22.12 **Ungleichung von Jensen**. *Seien $X \in \mathcal{L}_1$ reell und $\phi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion mit $\phi(X) \in \mathcal{L}_1$. Dann gilt $\phi\big(\mathbb{E}(X)\big) \leqslant \mathbb{E}(\phi(X))$.*

§22.13 **Beweis** von Satz §22.12. In der Vorlesung. □

§22.14 **Beispiel**. Für $\mathrm{r} \in \mathbb{R}_{>0}$ und $\mathrm{s} \in \mathbb{R}_{>\mathrm{r}}$ seien $|X|^{\mathrm{r}}, |X|^{\mathrm{s}} \in \mathcal{L}_1$. Da $\phi(x) = |x|^{\mathrm{s}/\mathrm{r}}$ konvex ist, folgt aus Satz §22.12 die *Ungleichung von Lyapounov* $(\mathbb{E}(|X|^{\mathrm{r}}))^{\mathrm{s}/\mathrm{r}} \leqslant \mathbb{E}(|X|^{\mathrm{s}})$. In Korollar §24.07 zeigen wir, dass $|X|^{\mathrm{s}} \in \mathcal{L}_1$ auch $|X|^{\mathrm{r}} \in \mathcal{L}_1$ impliziert. □

§22|02 Konvergenzsätze

§22.15 **Satz von der monotonen Konvergenz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1$, derart dass:*

(mK1) *$(X_n)_{n\in\mathbb{N}}$ ist monoton wachsend (fallend), also $X_n \uparrow X$ (bzw. $X_n \downarrow X$) für ein $X \in \overline{\mathcal{M}}$;*

(mK2) *$(|\mathbb{E}(X_n)|)_{n\in\mathbb{N}}$ ist beschränkt, das heißt, $\sup_{n\in\mathbb{N}}|\mathbb{E}(X_n)| \in \mathbb{R}_{\geqslant 0}$.*

Dann gilt $X \in \mathcal{L}_1$ und $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ (bzw. $\mathbb{E}(X_n) \downarrow \mathbb{E}(X)$).

§22.16 **Beweis** von Satz §22.15. In der Vorlesung. □

§22.17 **Satz von der dominierten Konvergenz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1$ derart, dass gelten:*

(dK1) *$(X_n)_{n\in\mathbb{N}}$ ist (punktweise) konvergent, also $\lim_{n\to\infty} X_n = X$ für ein $X \in \overline{\mathcal{M}}$;*

(dK2) *Es existiert $Y \in \mathcal{L}_1$ mit $\sup_{n\in\mathbb{N}}|X_n| \leqslant Y$, also $\sup_{n\in\mathbb{N}}|X_n| \in \mathcal{L}_1$.*

Dann gilt $X \in \mathcal{L}_1$, $\lim_{n\to\infty} \mathbb{E}(|X - X_n|) = 0$ und $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

§22.18 **Beweis** von Satz §22.17. In der Vorlesung. □

§22.19 **Satz von der konvergenten Reihe**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1(\mathbb{P})$ mit $\sum_{n\in\mathbb{N}} \mathbb{E}(|X_n|) \in \mathbb{R}_{\geqslant 0}$. Dann gilt:*

(i) *Die Reihe $\sum_{n\in\mathbb{N}} X_n$ konvergiert absolut fast sicher, also $\mathbb{P}(\sum_{n\in\mathbb{N}}|X_n| \in \mathbb{R}_{\geqslant 0}) = 1$;*

(ii) *$\sum_{n\in\mathbb{N}} X_n \in \mathcal{L}_1(\mathbb{P})$ und*

(iii) *$\mathbb{E}(\sum_{n\in\mathbb{N}} X_n) = \sum_{n\in\mathbb{N}} \mathbb{E}(X_n)$.*

§22.20 **Beweis** von Satz §22.19. Übungsaufgabe. Hinweis: Benutzen Sie Lemma §21.19 und die Endlichkeitsaussage aus Lemma §21.13 (ii) um zu zeigen, dass $Y := \sum_{n\in\mathbb{N}}|X_n| \in \mathbb{R}_{\geqslant 0}$ $\mathbb{P}$-f.s., und dass $Y \in \mathcal{L}_1$. Anschließend wenden Sie den Beweis §22.18 von der dominierten Konvergenz auf die Folge der Partialsummen an. □

§22.21 **Bemerkung**. Die Reihe $\sum_{n\in\mathbb{N}} X_n$ bleibt nicht definiert auf dem Ereignis $\{\sum_{n\in\mathbb{N}}|X_n| = \infty\}$. Da die Wahrscheinlichkeit dieses Ereignisses Null ist, hat es keinen Einfluss auf den Wert von $\mathbb{E}(\sum_{n\in\mathbb{N}} X_n)$ (vgl. Proposition §22.04 (vi)). □

§23 Variablentransformation

§23.01 **Satz.** *Sei X eine $(\mathcal{X},\mathscr{X})$-wertige Zufallsvariable auf $(\Omega,\mathscr{A},\mathbb{P})$, sei $\mathbb{P}^X=\mathbb{P}\circ X^{-1}$ die Verteilung von X auf $(\mathcal{X},\mathscr{X})$ und sei $h\in\overline{\mathcal{M}}(\mathscr{X})$ eine numerische Zufallsvariable auf $(\mathcal{X},\mathscr{X})$.*

(i) *Für $h\geqslant 0$, also $h\in\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X})$, gilt*

$$\mathbb{E}(h(X))=\int_\Omega h(X)d\mathbb{P}=\mathbb{P}(h(X))=\mathbb{P}^X(h)=\int_{\mathcal{X}} hd\mathbb{P}^X. \tag{23.01}$$

(ii) *$h(X)\in\mathcal{L}_1(\Omega,\mathscr{A},\mathbb{P})$ gilt genau dann, wenn $h\in\mathcal{L}_1(\mathcal{X},\mathscr{X},\mathbb{P}^X)$. In diesem Fall ist (23.01) ebenfalls erfüllt.*

§23.02 **Skizze.**

$$\begin{tikzcd}
(\Omega,\mathscr{A}) \arrow[r, "X"] \arrow[dr, "h(X)\in\mathcal{L}_1(\mathbb{P})"'] & (\mathcal{X},\mathscr{X}) \arrow[d, "h\in\mathcal{L}_1(\mathbb{P}^X)"] \\
& (\overline{\mathbb{R}},\overline{\mathscr{B}})
\end{tikzcd}$$

□

§23.03 **Bemerkung.** Analog zu Bemerkung §22.03 werden in der Maßtheorie weiterhin die folgenden zwei Darstellungen gezeigt:

(a) Sei X ein *stetig-verteilter* Zufallsvektor im $\mathbb{R}^n$ mit Dichte $\mathbb{f}^X$ und sei $h:\mathbb{R}^n\to\mathbb{R}$ Borel-messbar. Dann gilt $h\in\mathcal{L}_1(\mathbb{R}^n,\mathscr{B}^n,\mathbb{P}^X)$ genau dann, wenn $\int_{\mathbb{R}^n}|h(x)|\mathbb{f}^X(x)dx\in\mathbb{R}_{\geqslant 0}$. In diesem Fall ist

$$\mathbb{E}(h(X))=\int_{\mathbb{R}^n} h(x)\mathbb{f}^X(x)dx=\mathbb{P}^X(h).$$

(b) Seien X eine *diskret-verteilte* $\mathcal{X}:=X(\Omega)$-wertige Zufallsvariable mit Zähldichte $\mathbb{p}^X$ und $h:\mathcal{X}\to\mathbb{R}$, so gilt $h\in\mathcal{L}_1(\mathcal{X},2^{\mathcal{X}},\mathbb{P}^X)$ genau dann, wenn $\sum_{x\in\mathcal{X}}|h(x)|\mathbb{p}^X(x)<\infty$. In diesem Fall ist

$$\mathbb{E}(h(X))=\sum_{x\in\mathcal{X}} h(x)\mathbb{p}^X(x)=\mathbb{P}^X(h).$$

Nach Satz §23.01 gilt für reelle $X\in\mathcal{L}_1(\Omega,\mathscr{A},\mathbb{P})$ genau dann, wenn $\mathrm{id}_{\mathbb{R}}\in\mathcal{L}_1(\mathbb{R},\mathscr{B},\mathbb{P}^X)$, und somit $\mathbb{E}(X)=\mathbb{P}(X)=\int_\Omega X(\omega)\mathbb{P}(d\omega)=\int_{\mathbb{R}} x\mathbb{P}^X(dx)=\mathbb{P}^X(\mathrm{id}_{\mathbb{R}})$. In anderen Worten $\mathbb{E}(X)$ ist gerade der Erwartungswert der durch X induzierten Verteilung $\mathbb{P}^X$. □

§23.04 **Beispiel.**

(a) Sei $X=|W_1-W_2|$ und $Y=W_1+W_2$ der Absolutbetrag der Differenz bzw. die Summe der Augenzahlen von zwei unabhängigen fairen Würfeln $(W_1,W_2)\sim\mathrm{Lap}^2_{[\![6]\!]}$ (vgl. Beispiel §11.12 (a)). Dann gilt

$$\mathbb{E}(|W_1-W_2|)=\sum_{(i,j)\in[\![6]\!]^2}|i-j|\tfrac{1}{36}=\sum_{i\in[\![6]\!]}\sum_{j\in[\![6]\!]}|i-j|\tfrac{1}{36}=\tfrac{35}{18}$$

oder alternativ $\mathbb{E}(X)=0\frac{6}{36}+1\frac{10}{36}+2\frac{8}{36}+3\frac{6}{36}+4\frac{4}{36}+5\frac{2}{36}=\frac{35}{18}$. Weiterhin erhalten wir $\mathbb{E}(Y)=7$, $\mathbb{E}(XY)=\frac{245}{18}$, $\mathbb{E}(X^2)=\frac{35}{6}$ und $\mathbb{E}(Y^2)=\frac{987}{18}$.

(b) Sei $X:=(X_i)_{i\in[\![n]\!]}\sim\mathrm{B}^n_{\mathrm{p}}$ mit $n\in\mathbb{N}$ und $\mathrm{p}\in[0,1]$, in anderen Worten der von X induzierte Wahrscheinlichkeitsraum ist das Bernoulli-Schema $\big(\{0,1\}^n,2^{\{0,1\}^n},\mathrm{B}^n_{\mathrm{p}}\big)$. Unter Verwendung von Beispiel §22.06 (b) für $h := \sum_{i\in[\![n]\!]} \Pi_i \in \mathcal{M}(2^{\{0,1\}^n}, 2^{[\![0,n]\!]})$ gilt $h \in \mathcal{L}_1(\mathrm{B}^n_{\mathrm{p}})$ und $\mathrm{B}^n_{\mathrm{p}}(h) = np$. Nach Satz §23.01 für $h(X) = \sum_{i\in[\![n]\!]} X_i$ folgt also $\mathbb{E}(h(X)) = \mathrm{B}^n_{\mathrm{p}}(h) = np$. Andererseits, da $(X_i)_{i\in[\![n]\!]} \sim \mathrm{B}^n_{\mathrm{p}}$ gilt auch $h(X) = \sum_{i\in[\![n]\!]} X_i \sim \mathrm{Bin}_{(n,\mathrm{p})}$ (vgl. Beispiel §17.09 (a)) und somit besitzt eine $\mathrm{Bin}_{(n,\mathrm{p})}$-Verteilung den Erwartungswert $\mathrm{Bin}_{(n,\mathrm{p})}(\mathrm{id}_{[\![0,n]\!]}) = \mathbb{E}(h(X)) = np$.

(c) Sei $X \sim \mathrm{Poi}_\lambda$ mit $\lambda \in \mathbb{R}_{>0}$ und Zähldichte $\mathbb{p}_{\mathrm{Poi}_\lambda}(k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k \in \mathbb{Z}_{\geqslant 0}$ (vgl. Beispiel §22.06 (a)). Dann gilt

$$\mathbb{E}(X(X-1)) = \sum_{k\in\mathbb{Z}_{\geqslant 0}} k(k-1)\mathbb{p}_{\mathrm{Poi}_\lambda}(k) = \lambda^2 \sum_{k\in\mathbb{Z}_{\geqslant 2}} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} = \lambda^2 \sum_{k\in\mathbb{Z}_{\geqslant 0}} \mathbb{p}_{\mathrm{Poi}_\lambda}(k) = \lambda^2$$

und somit $\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \lambda^2 + \lambda$.

(d) Für $Z \sim \mathrm{N}_{(0,1)}$ mit partieller Integration unter Benutzung von $\mathbb{f}'_{\mathrm{N}_{(0,1)}}(z) = -z\mathbb{f}_{\mathrm{N}_{(0,1)}}(z)$ sowie $0 = \mathbb{E}(Z) = \int_{\mathbb{R}} z\mathbb{f}_{\mathrm{N}_{(0,1)}}(z)dz$ (vgl. Beispiel §22.06 (c)) gilt

$$\mathbb{E}(Z^2) = \int_{\mathbb{R}} z^2\mathbb{f}_{\mathrm{N}_{(0,1)}}(z)dz = -\int_{\mathbb{R}} z\mathbb{f}'_{\mathrm{N}_{(0,1)}}(z)dz = \int_{\mathbb{R}} \mathbb{f}_{\mathrm{N}_{(0,1)}}(z)dz = 1.$$

Alternativ, da $Y = Z^2 \sim \chi^2_1$ gilt auch $\mathbb{E}(Z^2) = \int_0^\infty y(2\pi y)^{-1/2}e^{-y/2}dy = 1$.

(e) Betrachten wir wie in Beispiel §10.12 (d) einen bivariat normalverteilten Zufallsvektor (Z_1, Z_2) mit Parametern $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1 = \sigma_2$ und $\rho \in (-1,1)$, dann gilt $Z_1 \sim \mathrm{N}_{(0,1)}$ und $Z_2 \sim \mathrm{N}_{(0,1)}$ (vgl. Beispiel §11.12 (b)). Weiterhin für $Z \sim \mathrm{N}_{(0,1)}$ ist $V = \sqrt{1-\rho^2}Z \sim \mathrm{N}_{(0,(1-\rho^2))}$ (vgl. Beispiel §10.12 (a)) mit Dichte $\mathbb{f}^V(v) = \frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp(-\frac{v^2}{2(1-\rho^2)})$ und

$$0 = \sqrt{1-\rho^2}\mathbb{E}(Z) = \mathbb{E}(V) = \int_{\mathbb{R}} v\mathbb{f}^V(v)dv = \int_{\mathbb{R}} \frac{v}{\sqrt{2\pi(1-\rho^2)}}\exp(-\frac{v^2}{2(1-\rho^2)})dv.$$

Somit gilt, unter Verwendung der gemeinsamem Dichte $\mathbb{f}^{(Z_1,Z_2)}$ von (Z_1, Z_2) und der Substitution von $z_2 - \rho z_1$ durch v,

$$\begin{aligned}\mathbb{E}(Z_1(Z_2-\rho Z_1)) &= \int_{\mathbb{R}}\int_{\mathbb{R}} z_1(z_2-\rho z_1)\mathbb{f}^{(Z_1,Z_2)}(z_1,z_2)dz_1dz_2\\ &= \int_{\mathbb{R}} \frac{z_1}{\sqrt{2\pi}}\exp(-\frac{(1-\rho^2)z_1^2}{2(1-\rho^2)})\int_{\mathbb{R}} \frac{(z_2-\rho z_1)}{\sqrt{2\pi(1-\rho^2)}}\exp(-\frac{(z_2-\rho z_1)^2}{2(1-\rho^2)})dz_2dz_1\\ &= \int_{\mathbb{R}} z_1\mathbb{f}^{Z_1}(z_1)\int_{\mathbb{R}} v\mathbb{f}^V(v)dvdz_1 = \mathbb{E}(Z_1)\mathbb{E}(V) = 0.\end{aligned}$$

Folglich gilt $\mathbb{E}(Z_1Z_2) = \mathbb{E}(Z_1(Z_2-\rho Z_1)) + \rho\mathbb{E}(Z_1^2) = \rho$, da $\mathbb{E}(Z_1^2) = 1$ nach (c). □

§23.05 **Proposition**. *Sei $X \in \overline{\mathcal{M}}_{\geqslant 0}$, dann gilt $\mathbb{E}(X) = \int_{\mathbb{R}_{\geqslant 0}} \mathbb{P}(X > y)dy$.*

§23.06 **Beweis** von Proposition §23.05. Übungsaufgabe. □

§24 $\mathcal{L}_s$-integrierbare Zufallsvariablen

§24.01 **Definition**. Seien $(\Omega, \mathscr{A})$ ein messbarer Raum, $\mathbb{P} \in \mathcal{W}(\mathscr{A})$ ein Wahrscheinlichkeitsmaß und $X \in \overline{\mathcal{M}}(\mathscr{A})$ eine numerische Zufallsvariable.

(a) Für $s \in \overline{\mathbb{R}}_{>0}$ definiere

$$\boxed{\|X\|_{\mathcal{L}_s}} := \|X\|_{\mathcal{L}_s(\mathbb{P})} := \begin{cases} (\mathbb{E}(|X|^s))^{1/s} & , \quad \text{für } s \in \mathbb{R}_{>0}, \\ \inf\{x \in \mathbb{R}_{\geqslant 0} : \mathbb{P}(|X| > x) = 0\} & , \quad \text{für } s = \infty. \end{cases}$$

X heißt *$\mathcal{L}_s$-integrierbar*, wenn $\|X\|_{\mathcal{L}_s} \in \mathbb{R}_{\geqslant 0}$. Die Menge aller $\mathcal{L}_s$-integrierbaren Zufallsvariablen bezeichnen wir mit $\boxed{\mathcal{L}_s} := \mathcal{L}_s(\mathbb{P}) := \mathcal{L}_s(\Omega, \mathscr{A}, \mathbb{P}) := \{X \in \overline{\mathcal{M}}(\mathscr{A}) : \|X\|_{\mathcal{L}_s(\mathbb{P})} \in \mathbb{R}_{\geqslant 0}\}$.

(b) Für $X \in \mathcal{L}_s$ und $s \in \mathbb{N}$ heißt $\mathbb{E}(X^s)$ das *s-te Moment* von X; für $X \in \mathcal{L}_s$ und $s \in \mathbb{R}_{>0}$ heißt $\mathbb{E}(|X|^s)$ das *s-te absolute Moment* von X.

Das Wahrscheinlichkeitsmaß $\mathbb{P}$ besitzt ein *endliches s-tes Moment*, wenn $\mathrm{id}_\Omega \in \mathcal{L}_s(\Omega, \mathscr{A}, \mathbb{P})$ gilt. In diesem Fall wird $\mathbb{E}(\mathrm{id}_\Omega^s) = \mathbb{P}(\mathrm{id}_\Omega^s)$ das *s-te Moment des Wahrscheinlichkeitsmaßes* $\mathbb{P}$ genannt. Wir bezeichnen mit $\mathcal{W}_s(\mathscr{A})$ die Menge aller Wahrscheinlichkeitsmaße auf $(\Omega, \mathscr{A})$, die ein endliches s-tes Moment besitzen. □

§24.02 **Bemerkung**.

(a) Für $s = 1$ stimmt die letzte Definition mit Definition §22.02 (b) überein.

(b) Für $s \in \mathbb{R}_{>0}$ gilt $X \in \mathcal{L}_s$ genau dann, wenn $|X|^s \in \mathcal{L}_1$ gilt.

(c) Für $s = \infty$ gilt $\mathbb{P}(|X| > \|X\|_{\mathcal{L}_\infty}) = 0$, da die Funktion $x \mapsto \mathbb{P}(|X| > x)$ rechtsstetig ist. □

§24.03 **Lemma**. *Seien $X \in \overline{\mathcal{M}}(\mathscr{A})$ und $s \in \overline{\mathbb{R}}_{>0}$.*

(i) $\|X\|_{\mathcal{L}_s} = 0$ *gilt genau dann, wenn* $\mathbb{P}(X = 0) = 1$ *gilt;*

(ii) *Für* $a \in \mathbb{R}$ *gilt* $\|aX\|_{\mathcal{L}_s} = |a| \cdot \|X\|_{\mathcal{L}_s}$*;*

(iii) $\|1\|_{\mathcal{L}_s} = 1$*;*

(iv) *Für* $X \in \mathcal{L}_s$ *gilt* $\mathbb{P}(|X| \in \mathbb{R}_{\geqslant 0}) = 1$*;*

(v) *Für* $Y \in \overline{\mathcal{M}}(\mathscr{A})$ *mit* $\mathbb{P}(X = Y) = 1$ *gilt* $\|X\|_{\mathcal{L}_s} = \|Y\|_{\mathcal{L}_s}$.

§24.04 **Beweis** von Lemma §24.03. Übungsaufgabe. □

§24.05 **Hölder Ungleichung**. *Seien $X, Y \in \overline{\mathcal{M}}(\mathscr{A})$ und $s, r \in \overline{\mathbb{R}}_{\geqslant 1}$ mit $\frac{1}{s} + \frac{1}{r} = 1$. Dann gilt:*

$$\mathbb{E}(|XY|) \leqslant \|X\|_{\mathcal{L}_s} \|Y\|_{\mathcal{L}_r}.$$

§24.06 **Beweis** von Satz §24.05. In der Vorlesung. □

§24.07 **Korollar**. *Für $s \in \overline{\mathbb{R}}_{>0}$, $r \in \overline{\mathbb{R}}_{\geqslant s}$ und $X \in \mathcal{L}_r$ gilt $\|X\|_{\mathcal{L}_s} \leqslant \|X\|_{\mathcal{L}_r}$ und somit $\mathcal{L}_r \subseteq \mathcal{L}_s$.*

§24.08 **Beweis** von Korollar §24.07. In der Vorlesung. □

§24.09 **Cauchy-Schwarz Ungleichung**. *Für $X, Y \in \mathcal{L}_2$ gilt $XY \in \mathcal{L}_1$ und*

$$|\mathbb{E}(XY)|^2 \leqslant \|X\|_{\mathcal{L}_2}^2 \|Y\|_{\mathcal{L}_2}^2 = \mathbb{E}(|X|^2)\mathbb{E}(|Y|^2).$$

§24.10 **Beweis** von Satz §24.09. Die Aussage folgt direkt aus Satz §24.05. □

§24.11 **Minkowski Ungleichung**. *Für $X, Y \in \mathcal{L}_s$ mit $s \in \overline{\mathbb{R}}_{\geqslant 1}$ gilt $X + Y \in \mathcal{L}_s$ und*

$$\|X + Y\|_{\mathcal{L}_s} \leqslant \|X\|_{\mathcal{L}_s} + \|Y\|_{\mathcal{L}_s}.$$

§24.12 **Beweis** von Satz §24.11. In der Vorlesung. □

§24.13 **Bemerkung**. Für $s \in \overline{\mathbb{R}}_{\geqslant 1}$ ist $\|\cdot\|_{\mathcal{L}_s}$ eine Pseudonorm auf $\mathcal{L}_s$, d.h., für $X, Y \in \mathcal{L}_s$ und $a \in \mathbb{R}$ gilt:

(a) $\|aX\|_{\mathcal{L}_s} = |a| \cdot \|X\|_{\mathcal{L}_s}$; (Lemma §24.03 (ii))

(b) $\|X + Y\|_{\mathcal{L}_s} \leqslant \|X\|_{\mathcal{L}_s} + \|Y\|_{\mathcal{L}_s}$; (Minkowski-Ungleichung §24.11)

(c) $\|X\|_{\mathcal{L}_s} \in \mathbb{R}_{\geqslant 0}$ und $X = 0$ $\mathbb{P}$-f.s.. impliziert $\|X\|_{\mathcal{L}_s} = 0$. (Lemma §24.03 (i)) □

§25 Varianz, Kovarianz und Korrelation

§25.01 **Erinnerung**. Für $X, Y \in \mathcal{L}_2$ folgt aus Satz §24.05, dass $X, Y, XY \in \mathcal{L}_1$ gilt. □

§25.02 **Definition**. Für Zufallsvariablen $X, Y \in \mathcal{L}_2$ bezeichnet

$$\boxed{\mathbb{K}\mathrm{ov}(X, Y)} := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

die *Kovarianz* zwischen X und Y.

$$\boxed{\mathbb{V}\mathrm{ar}(X)} := \mathbb{K}\mathrm{ov}(X, X) = \mathbb{E}(|X - \mathbb{E}(X)|^2) = \mathbb{E}(X^2) - |\mathbb{E}(X)|^2 \quad \text{und}$$
$$\boxed{\mathbb{S}\mathrm{ta}(X)} := \sqrt{\mathbb{V}\mathrm{ar}(X)}$$

heißt *Varianz* bzw. *Standardabweichung* von X (und $\mathbb{P}^X$). □

§25.03 **Lemma**. *Seien $X, Y, Z \in \mathcal{L}_2$ und $a, b \in \mathbb{R}$.*

(i) *Es gilt $\mathbb{V}\mathrm{ar}(X) = 0$ genau dann, wenn $\mathbb{P}(X = \mathbb{E}(X)) = 1$, also wenn eine Konstante $c \in \mathbb{R}$ mit $X = c$ $\mathbb{P}$-f.s. existiert;*

(ii) $\mathbb{K}\mathrm{ov} : \mathcal{L}_2 \times \mathcal{L}_2 \to \mathbb{R}$ *ist eine positiv semi-definite, symmetrische Bilinearform, das heißt*

(ii-a) $\mathbb{K}\mathrm{ov}(X, Y) = \mathbb{K}\mathrm{ov}(Y, X)$ (symmetrisch)

(ii-b) $\mathbb{K}\mathrm{ov}(aX + bY, Z) = a\mathbb{K}\mathrm{ov}(X, Z) + b\mathbb{K}\mathrm{ov}(Y, Z)$ (linear)

(ii-c) $\mathbb{K}\mathrm{ov}(X, X) \in \mathbb{R}_{\geqslant 0}$ (positiv semi-definit)

und es gilt weiterhin $\mathbb{K}\mathrm{ov}(a, X) = 0$.

(iii) $\mathbb{V}\mathrm{ar} : \mathcal{L}_2 \to \mathbb{R}_{\geqslant 0}$ *ist die von $\mathbb{K}\mathrm{ov}$ induzierte quadratische Form, sodass*

(iii-a) $\mathbb{V}\mathrm{ar}(aX + b) = a^2\mathbb{V}\mathrm{ar}(X)$ *und*

(iii-b) $\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + 2\mathbb{K}\mathrm{ov}(X, Y)$ *gelten.*

§25.04 **Beweis** von Lemma §25.03. Übung □

§25.05 **Beispiel** *(Beispiel §23.04 fortgesetzt.)*.

(a) Sei $X \sim \mathrm{Poi}_\lambda$ mit $\lambda \in \mathbb{R}_{>0}$. Dann gilt $\mathbb{E}(X) = \lambda$ und $\mathbb{E}(X^2) = \lambda^2 + \lambda$ (vgl. Beispiel §23.04 (c)), sodass $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - |\mathbb{E}(X)|^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ gilt.

(b) Für $Z \sim \mathrm{N}_{(0,1)}$ gilt $\mathbb{E}(Z) = 0$ und $\mathbb{E}(Z^2) = 1$ (vgl. Beispiel §23.04 (d)), so dass $\mathbb{V}\mathrm{ar}(Z) = \mathbb{E}(Z^2) - |\mathbb{E}(Z)|^2 = 1$ gilt. Für $\mu \in \mathbb{R}$ und $\sigma \in \mathbb{R}_{>0}$ ist $X = \mu + \sigma Z \sim \mathrm{N}_{(\mu,\sigma^2)}$ (vgl. Beispiel §10.12 (a)), sodass $\mathbb{E}(X) = \sigma\mathbb{E}(Z) + \mu = \mu$ und $\mathbb{V}\mathrm{ar}(X) = \sigma^2\mathbb{V}\mathrm{ar}(Z) = \sigma^2$ gilt.

(c) Betrachten wir wie in Beispiel §10.12 (d) einen bivariat normalverteilten Zufallsvektor (X, Y) mit Parametern $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$ und $\rho \in (-1, 1)$, dann gilt $X \sim \mathrm{N}_{(\mu_1, \sigma_1^2)}$ und $Y \sim \mathrm{N}_{(\mu_2, \sigma_2^2)}$ (vgl. Beispiel §11.12 (b)), sowie $Z_1 := \sigma_1^{-1}(X - \mu_1) \sim \mathrm{N}_{(0,1)}$ und $Z_2 := \sigma_2^{-1}(Y - \mu_2) \sim \mathrm{N}_{(0,1)}$ (vgl. Beispiel §10.12 (a)). Der Zufallsvektor (Z_1, Z_2) ist ebenfalls bivariat normalverteilt mit Parametern $\mu_1 = 0 = \mu_2$, $\sigma_1 = 1 = \sigma_2$ und $\mathbb{E}(Z_1 Z_2) = \rho \in (-1, 1)$ (vgl. Beispiel §23.04 (e)). Insbesondere, gilt damit $\mathbb{K}\mathrm{ov}(X, Y) = \mathbb{E}((X - \mu_1)(Y - \mu_2)) = \mathbb{E}((\sigma_1 Z_1)(\sigma_2 Z_2)) = \sigma_1 \sigma_2 \rho$. □

§25.06 **Beste konstante Vorhersage**. *Für $Y \in \mathcal{L}_2$ und $a \in \mathbb{R}$ heißt $\mathbb{E}(Y - a)$ Bias und es gilt die Bias^2-Varianz-Zerlegung*

$$\mathbb{E}(|Y - a|^2) = |\mathbb{E}(Y) - a|^2 + \mathbb{V}\mathrm{ar}(Y).$$

Die Abbildung $a \mapsto \mathbb{E}(|Y - a|^2)$ nimmt ihr Minimum auf $\mathbb{R}$ genau bei $a^* = \mathbb{E}(Y)$ an, sodass

$$\min_{a \in \mathbb{R}} \mathbb{E}(|Y - a|^2) = \mathbb{V}\mathrm{ar}(Y) \quad \text{und} \quad \{\mathbb{E}(Y)\} = \arg\min_{a \in \mathbb{R}} \mathbb{E}(|Y - a|^2)$$

gilt und $a^* = \mathbb{E}(Y)$ beste konstante Vorhersage von Y genannt wird. □

§25.07 **Markov Ungleichung**. *Für $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ und $s, \varepsilon \in \mathbb{R}_{>0}$ gilt $\varepsilon^s \mathbb{1}_{\{Y \geqslant \varepsilon\}} \leqslant Y^s$, sodass Proposition §22.04 (iv) impliziert $\mathbb{P}(Y \geqslant \varepsilon) \leqslant \varepsilon^{-s}\mathbb{E}(Y^s)$.* □

§25.08 **Tschebischeff Ungleichung**. *Für $X \in \mathcal{L}_2$ und $\varepsilon \in \mathbb{R}_{>0}$ impliziert das Anwenden von Satz §25.07 auf $Y = |X - \mathbb{E}(X)|$, dass $\mathbb{P}(|X - \mathbb{E}(X)| \geqslant \varepsilon) \leqslant \varepsilon^{-2}\mathbb{V}\mathrm{ar}(X)$.* □

§25.09 **Bemerkung**. Der Erwartungswert $\mu := \mathbb{E}(X)$ und die Varianz $\sigma^2 := \mathbb{V}\mathrm{ar}(X)$ sind Kenngrößen der von X induzierten Verteilung $\mathbb{P}^X$, insbesondere hängen diese nicht direkt von X sondern nur von $\mathbb{P}^X$ ab. Dies erklärt auch, warum in der Statistik der Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ häufig nicht angegeben wird, da die Beobachtung von X nur Rückschlüsse auf die Verteilung $\mathbb{P}^X$ von X erlaubt. Betrachten wir die Varianz, so ist diese ein Maß der Streuung von X um seinen Erwartungswert, da nach Satz §25.08 $\mathbb{P}(|X - \mu|| \geqslant k\sigma) \leqslant k^{-2}$ für alle $k \in \mathbb{N}$ gilt. Damit erhalten wir zum Beispiel $\mathbb{P}(|X - \mu| \geqslant \sigma) \leqslant 1$, $\mathbb{P}(|X - \mu| \geqslant 2\sigma) \leqslant 1/4$ sowie $\mathbb{P}(|X - \mu| \geqslant 3\sigma) \leqslant 1/9$. Die Abschätzung Satz §25.08 ist sehr grob, da sie für alle Zufallsvariablen in $\mathcal{L}_2$ gilt. Kennt man die Verteilung von X, so kann man die Wahrscheinlichkeit genau ausrechnen. Zum Beispiel für $X \sim \mathrm{N}_{(\mu, \sigma^2)}$, gilt $\mathbb{P}(|X - \mu| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$ (vgl. Beispiel §10.12 (a)), sodass $\mathbb{P}(|X - \mu| \geqslant k\sigma) = 2 - 2\Phi(k)$ und insbesondere $\mathbb{P}(|X - \mu| \geqslant \sigma) \approx 0.32$, $\mathbb{P}(|X - \mu| \geqslant 2\sigma) \approx 0.04$ sowie $\mathbb{P}(|X - \mu| \geqslant 3\sigma) \approx 0.0026$. □

§25.10 **Erinnerung.**. Für $X, Y \in \mathcal{L}_2$ gilt $|\mathbb{K}\mathrm{ov}(X, Y)| \leqslant \mathbb{S}\mathrm{ta}(X)\mathbb{S}\mathrm{ta}(Y)$ durch Anwenden der Cauchy-Schwarz Ungleichung §24.09. □

§25.11 **Definition**. Für $X, Y \in \mathcal{L}_2$ mit $\mathbb{S}\mathrm{ta}X, \mathbb{S}\mathrm{ta}Y \in \mathbb{R}_{>0}$ bezeichnet

$$\boxed{\mathbb{K}\mathrm{orr}(X, Y)} := \frac{\mathbb{K}\mathrm{ov}(X, Y)}{\mathbb{S}\mathrm{ta}(X)\mathbb{S}\mathrm{ta}(Y)} \in [-1, 1]$$

die *Korrelation* zwischen X und Y. Falls $\mathbb{K}\mathrm{ov}(X, Y) = 0$ gilt, heißen X und Y *unkorreliert*.

§25.12 **Korollar**.

(i) *Für unkorrellierte Zufallsvariablen $X, Y \in \mathcal{L}_2$ gilt $\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$.*

(ii) *Unabhängige Zufallsvariablen $X, Y \in \mathcal{L}_2$ sind unkorreliert. Die Umkehrung gilt nicht.*

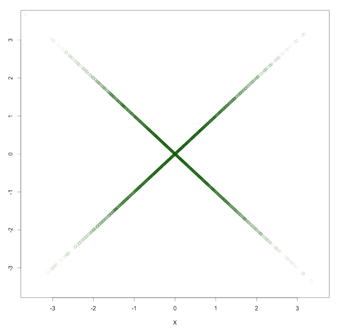
§25.13 **Beweis** von Korollar §25.12. Die Aussage (i) folgt aus Lemma §25.03 (iii-b) und die Aussage (ii) direkt aus Proposition §22.04 (vii) und Beispiel §25.14. □

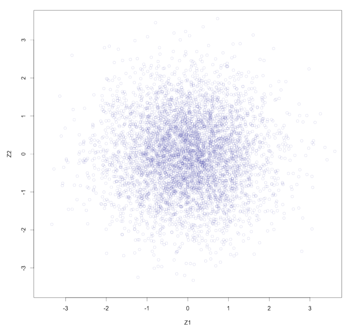
§25.14 **Beispiel.**

(a) Sei $(X_i)_{i\in[\![n]\!]} \sim \mathrm{B}^n_{\mathrm{p}}$, $n \in \mathbb{N}$, $\mathrm{p} \in [0,1]$, ein Bernoulli-Schema. Für jedes $i \in [\![n]\!]$ ist $X_i \sim \mathrm{B}_{\mathrm{p}}$ mit $\mathbb{P}(X_i = 1) = \mathrm{p}$ und $\mathbb{P}(X_i = 0) = 1-\mathrm{p}$, sodass $X_i \in \mathcal{L}_2$ mit $\mathbb{E}(X_i^2) = \mathrm{p}$ und $\mathbb{V}\mathrm{ar}(X_i) = \mathrm{p}(1-\mathrm{p})$ gilt. Da $X = \sum_{i\in[\![n]\!]} X_i \sim \mathrm{Bin}_{(n,\mathrm{p})}$ (vgl. Beispiel §23.04 (b)) und $\perp\!\!\!\perp_{i\in[\![n]\!]} X_i$ ist auch $X \in \mathcal{L}_2$ mit $\mathbb{V}\mathrm{ar}(X) = \sum_{i\in[\![n]\!]} \mathbb{V}\mathrm{ar}(X_i) = n\mathrm{p}(1-\mathrm{p})$. Im Fall $\mathrm{p} \in \{0,1\}$ gilt also $\mathbb{V}\mathrm{ar}(X) = 0$ sowie für beliebige $\mathrm{p} \in [0,1]$ stets $\mathbb{V}\mathrm{ar}(X) \leqslant n/4$. Die relative Häufigkeit von Erfolgen $\widehat{\mathrm{p}} := X/n$ erfüllt $\mathbb{E}(\widehat{\mathrm{p}}) = \mathrm{p}$ und $\mathbb{V}\mathrm{ar}(\widehat{\mathrm{p}}) = \frac{\mathrm{p}(1-\mathrm{p})}{n} \leqslant \frac{1}{4n}$.

(b) Sei $X = |W_1 - W_2|$ und $Y = W_1 + W_2$ der Absolutbetrag der Differenz bzw. die Summe der Augenzahlen von zwei unabhängigen fairen Würfeln $(W_1, W_2) \sim \mathrm{Lap}^2_{[\![6]\!]}$ (vgl. Beispiel §11.12 (a)). Dann sind X und Y nicht unabhängig (vgl. Beispiel §16.09 (b)), aber X und Y sind unkorreliert, da $\mathbb{K}\mathrm{ov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{245}{18} - \frac{35}{18} \cdot 7 = 0$ (vgl. Beispiel §23.04 (a)).

(c) Sei (X,Y) bivariat normalverteilt wie in Beispiel §10.12 (d) mit Parametern $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$ und $\rho \in (-1,1)$. Dann gilt $\mathbb{S}\mathrm{ta}(X) = \sigma_1$ und $\mathbb{S}\mathrm{ta}(Y) = \sigma_2$ sowie $\mathbb{K}\mathrm{ov}(X,Y) = \sigma_1\sigma_2\rho$ (vgl. Beispiel §25.05 (b) bzw. (c)) und somit $\mathbb{K}\mathrm{orr}(X,Y) = \rho$. Da X und Y genau dann unabhängig sind, wenn $\rho = 0$ gilt (vgl. Beispiel §16.09 (e)), sind also X und Y genau dann unabhängig, wenn sie unkorreliert sind. *Achtung*, es ist natürlich möglich, dass $X \sim \mathrm{N}_{(\mu_1,\sigma_1^2)}$ und $Y \sim \mathrm{N}_{(\mu_2,\sigma_2^2)}$ unkorreliert sind, aber der Vektor (X,Y) nicht bivariat normalverteilt ist. Betrachte dazu zwei unabhängige Zufallsvariablen X und V, wobei $X \sim \mathrm{N}_{(0,1)}$ und V ist eine Rademacher-Zufallsvariable, d.h. $V \in \{-1,1\}$ mit $\mathbb{P}(V = -1) = 1/2 = \mathbb{P}(V = 1)$. Es ist nun leicht zu zeigen, dass die Zufallsvariablen $Y := VX$ und X unkorreliert sind und dass $Y \sim \mathrm{N}_{(0,1)}$ (Übung!). Die Zufallsvariablen X und Y sind somit standardnormalverteilt und unkorreliert, aber ihre gemeinsame Verteilung ist keine Normalverteilung (warum?). Die nächsten Graphiken zeigen 5000 Realisierungen von (X,Y) (in grün) und zum Vergleich 5000 Realisierungen einer bivariaten Standardnormalverteilung.





□

§25.15 **Vorbemerkung.**. Seien $X, Y \in \mathcal{L}_2$ Zufallsvariablen. Welche affine Funktion $aX + b$ mit $a, b \in \mathbb{R}$ von X sagt Y am Besten vorher, im Sinne, dass der mittlere quadratische Vorhersagefehler (Mean Squared Prediction Error, kurz MSPE) $\mathbb{E}(|Y - (aX+b)|^2)$ minimal ist? Ist $\mathbb{V}\mathrm{ar}(X) = 0$, so befinden wir uns im Fall der *besten konstanten Vorhersage* §25.06, sodass wir im Folgenden $\mathbb{V}\mathrm{ar}(X) \in \mathbb{R}_{>0}$ annehmen und mit

$$L_X := \{aX + b : a, b \in \mathbb{R}\} \subseteq \mathcal{L}_2$$

die Menge der Zufallsvariablen, die affine Funktionen von X sind, bezeichnen. □

§25.16 **Beste lineare Vorhersage.** *Seien $X, Y \in \mathcal{L}_2$ mit $\mathbb{V}\mathrm{ar}(X) \in \mathbb{R}_{>0}$. Für jedes $Z = aX + b \in L_X$ gilt die Bias^2+Varianz-Zerlegung*

$$\mathbb{E}(|Y - Z|^2) = |\mathbb{E}(Y) - \mathbb{E}(Z)|^2 + \mathbb{V}\mathrm{ar}(Y - Z) = |\mathbb{E}(Y) - a\mathbb{E}(X) - b|^2 + \mathbb{V}\mathrm{ar}(Y - aX).$$

Da für $a^* := \frac{\mathbb{K}\mathrm{ov}(X,Y)}{\mathbb{V}\mathrm{ar}(X)}$ *gilt* $\mathbb{K}\mathrm{ov}(Y - a^*X, X) = \mathbb{K}\mathrm{ov}(Y, X) - a^*\mathbb{V}\mathrm{ar}(X) = 0$*, folgt*

$$\begin{aligned}\mathbb{V}\mathrm{ar}(Y - aX) &= \mathbb{V}\mathrm{ar}(Y - a^*X + (a^* - a)X) = \mathbb{V}\mathrm{ar}(Y - a^*X) + (a^* - a)^2\mathbb{V}\mathrm{ar}(X);\\ \mathbb{V}\mathrm{ar}(Y - a^*X) &= \mathbb{V}\mathrm{ar}(Y) + |a^*|^2\mathbb{V}\mathrm{ar}(X) - 2\mathbb{K}\mathrm{ov}(Y, a^*X)\\ &= \mathbb{V}\mathrm{ar}(Y) - \frac{(\mathbb{K}\mathrm{ov}(X,Y))^2}{\mathbb{V}\mathrm{ar}(X)} = \mathbb{V}\mathrm{ar}(Y)(1 - (\mathbb{K}\mathrm{orr}(X,Y))^2).\end{aligned}$$

Für $b^* := \mathbb{E}(Y) - a^*\mathbb{E}(X)$ *und* $Z^* = a^*X + b^*$ *gilt somit*

$$\mathbb{E}(|Y - Z|^2) \geqslant \mathbb{V}\mathrm{ar}(Y - aX) \geqslant \mathbb{V}\mathrm{ar}(Y - a^*X) = \mathbb{E}(|Y - Z^*|^2)$$

wobei Gleichheit genau dann gilt, wenn $a = a^*$ *und* $b = b^*$ *gilt. Zusammenfassend nimmt die Abbildung* $Z \mapsto \mathbb{E}(|Y - Z|^2)$ *damit ihr Minimum auf* L_X *genau bei* Z^* *an, sodass*

$$\min_{Z\in L_X} \mathbb{E}(|Y - Z|^2) = \mathbb{E}(|Y - Z^*|^2) = \mathbb{V}\mathrm{ar}(Y)(1 - (\mathbb{K}\mathrm{orr}(X,Y))^2)$$
$$\text{und} \quad \{\mathbb{E}(Y) + \tfrac{\mathbb{K}\mathrm{ov}(X,Y)}{\mathbb{V}\mathrm{ar}(X)}(X - \mathbb{E}(X)) = Z^*\} = \arg\min_{Z\in L_X} \mathbb{E}(|Y - Z|^2)$$

gilt. $Z^* = a^*X + b^*$ *wird* beste lineare Vorhersage *von* Y *durch* X *genannt.* □

§25.17 **Bemerkung**. Da $\min_{Z\in L_X} \mathbb{E}(|Y - Z|^2) = \mathbb{V}\mathrm{ar}(Y)(1 - (\mathbb{K}\mathrm{orr}(X,Y))^2)$ gilt, wird $\mathbb{K}\mathrm{orr}(X,Y)$ auch als Maß für die Stärke der *linearen* Abhängigkeit zwischen X und Y interpretiert. Zur Erinnerung $\mathbb{K}\mathrm{orr}(X,Y) \in [-1,1]$ und im Extremfall $|\mathbb{K}\mathrm{orr}(X,Y)| = 1$ gilt somit $\mathbb{E}(|Y - Z^*|^2) = 0$ für die beste lineare Vorhersage $Z^* = a^*X + b^*$ von Y, also $Y = a^*X + b^*$ $\mathbb{P}$-f.s. nach Lemma §21.13 (i), sodass bis auf einer Nullmenge Y gerade gleich einer affinen Funktion von X ist. □.

§26 Hauptkomponentenanalyse

§26.01 **Erinnerung**. Im Folgenden fassen wir wieder Vektoren $a = (a_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$ als Spaltenvektoren auf, das heißt $a = [a_1 \cdots a_n]^{\mathrm{T}} \in \mathbb{R}^n$. Wir bezeichnen mit $\|\bullet\|$ die vom Standardskalarprodukt $\langle\bullet,\bullet\rangle$ auf $\mathbb{R}^n$ induzierte Norm, das heißt, $\|a\| = \sqrt{\langle a, a\rangle} = \sqrt{a^{\mathrm{T}}a} = (\sum_{i\in[\![n]\!]} a_i^2)^{1/2}$ für alle $a \in \mathbb{R}^n$. Weiterhin setzen wir $|a| := \sum_{i\in[\![n]\!]} |a_i|$. □

§26.02 **Definition**. Sei $X = [X_1 \cdots X_n]^{\mathrm{T}}$ ein Zufallsvektor (aufgefasst als Spaltenvektor).

(a) Falls $|X| = \sum_{i\in[\![n]\!]}|X_i| \in \mathcal{L}_1$, also $X_i \in \mathcal{L}_1$, $i \in [\![n]\!]$, kurz $X \in \mathcal{L}_1$, dann heißt

$$\mathbb{E}(X) = [\mathbb{E}(X_1) \cdots \mathbb{E}(X_n)]^{\mathrm{T}} = \begin{bmatrix}\mathbb{E}(X_1)\\ \vdots\\ \mathbb{E}(X_n)\end{bmatrix} \in \mathbb{R}^n$$

Erwartungswertvektor von X.

(b) Falls $\|X\|^2 = \sum_{i\in[\![n]\!]} |X_i|^2 \in \mathcal{L}_1$, also $X_i \in \mathcal{L}_2$, $i \in [\![n]\!]$, kurz $X \in \mathcal{L}_2$, dann heißt

$$\begin{aligned}\mathbb{K}\mathrm{ov}(X) &= (\mathbb{K}\mathrm{ov}(X_i, X_j))_{i,j\in[\![n]\!]} = (\mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))))_{i,j\in[\![n]\!]}\\ &= \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^{\mathrm{T}}) = \mathbb{E}(XX^{\mathrm{T}}) - \mathbb{E}(X)(\mathbb{E}(X))^{\mathrm{T}} \in \mathbb{R}^{n\times n}\end{aligned}$$

Kovarianzmatrix von X. □

§26.03 **Bemerkung**. Für die Kovarianzmatrix $\Sigma := \mathbb{K}\mathrm{ov}(X)$ eines $\mathbb{R}^n$-wertigen Zufallsvektors X, falls sie existiert, gilt

$$\mathbb{K}\mathrm{ov}(\langle a, X\rangle, \langle X, b\rangle) = \sum_{i\in[\![n]\!]}\sum_{j\in[\![n]\!]} a_i b_j \mathbb{K}\mathrm{ov}(X_i, X_j) = b^{\mathrm{T}}\Sigma a = \langle \Sigma a, b\rangle$$

für alle $a, b \in \mathbb{R}^n$. Insbesondere ist $\Sigma \in \mathbb{R}^{n\times n}_{\geqslant}$ *positiv semidefinit*, da $\mathbb{K}\mathrm{ov}(X_i, X_j) = \mathbb{K}\mathrm{ov}(X_j, X_i)$ für alle $i, j \in [\![n]\!]$ und $\langle \Sigma a, a\rangle = \mathbb{V}\mathrm{ar}(\langle X, a\rangle) \in \mathbb{R}_{\geqslant 0}$ für alle $a \in \mathbb{R}^n$ gilt. □

§26.04 **Lemma**. *Sei X ein $\mathbb{R}^m$-wertiger Zufallsvektor in $\mathcal{L}_2$. Für alle $b \in \mathbb{R}^n$ und $\mathrm{A} \in \mathbb{R}^{n\times m}$ ist dann $Y = \mathrm{A}X + b$ ein $\mathbb{R}^n$-wertiger Zufallsvektor in $\mathcal{L}_2$. Bezeichnen wir weiterhin mit $\mu := \mathbb{E}(X) \in \mathbb{R}^m$ und $\Sigma := \mathbb{K}\mathrm{ov}(X) \in \mathbb{R}^{m\times m}_{\geqslant}$ den Erwartungswertvektor und die Kovarianzmatrix von X, dann gilt $\mathbb{E}(Y) = \mathrm{A}\mu + b$ und $\mathbb{K}\mathrm{ov}(Y) = \mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}}$.*

§26.05 **Beweis** von Lemma §26.04. In der Vorlesung. □

§26.06 **Bemerkung**. Sei X ein $\mathbb{R}^m$-wertiger Zufallsvektor in $\mathcal{L}_2$. Setzen wir $\mu := \mathbb{E}(X) \in \mathbb{R}^m$ und $\Sigma := \mathbb{K}\mathrm{ov}(X) \in \mathbb{R}^{m\times m}_{\geqslant}$. Dann gilt $\mathbb{E}(\langle X, a\rangle) = \langle \mu, a\rangle$ und $\mathbb{K}\mathrm{ov}(\langle a, X\rangle, \langle X, b\rangle)) = \langle \Sigma a, b\rangle$ für alle $a, b \in \mathbb{R}^n$. □

§26.07 **Beispiel**. Für $X \sim \mathrm{N}_{(\mu,\Sigma)}$ ist $\mu = \mathbb{E}(X) \in \mathbb{R}^n$ der Erwartungswertvektor und $\Sigma = \mathbb{K}\mathrm{ov}(X) \in \mathbb{R}^{n\times n}_{\geqslant}$ die Kovarianzmatrix von X. □

§26.08 **Erinnerung**. Sei $\Sigma \in \mathbb{R}^{n\times n}_{\geqslant}$ eine symmetrische, positiv semidefinite Matrix. Da Σ symmetrisch ist, ist Σ insbesondere diagonalisierbar, also, ähnlich zu einer Diagonalmatrix $\mathrm{Diag}[\lambda_i]$ mit reellen Diagonaleinträgen $\lambda = (\lambda_i)_{i\in[\![n]\!]}$, genannt *Eigenwerte* von Σ. Es existiert also eine orthogonale Matrix $\mathrm{U} \in \mathrm{O}(n)$, also $\mathrm{U}^{\mathrm{T}}\mathrm{U} = \mathrm{I}_{n\times n} = \mathrm{U}\mathrm{U}^{\mathrm{T}}$, deren Spalten $(u_{\bullet i})_{i\in[\![m]\!]}$, genannt *Eigenvektoren*, eine Orthonormalbasis des $\mathbb{R}^n$ bilden, derart dass $\mathrm{U}^{\mathrm{T}}\Sigma\mathrm{U} = \mathrm{Diag}[\lambda_i]$ gilt. Da Σ positiv semidefinit ist, sind alle Eigenwerte positiv, also $\lambda_i \in \mathbb{R}_{\geqslant 0}$, $i \in [\![n]\!]$. Im Folgenden nehmen wir an, dass die orthogonale Matrix U so gewählt ist, dass die Eigenwerte der Größe nach angeordnet sind, d.h. $\lambda_1 \geqslant \lambda_2 \geqslant \dots \lambda_n \geqslant 0$ gilt. □

§26.09 **Definition**. Sei X ein $\mathbb{R}^n$-wertiger Zufallsvektor in $\mathcal{L}_2$ mit positiv semidefiniter Kovarianzmatrix $\Sigma = \mathbb{K}\mathrm{ov}(X) \in \mathbb{R}^{n\times n}_{\geqslant}$. Weiterhin seien $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_n \geqslant 0$ die der Größe nach geordneten Eigenwerte der Matrix Σ und $u_{\bullet i} \in \mathbb{R}^n$, $i \in [\![n]\!]$, die zugehörigen Eigenvektoren. Dann heißen $(u_{\bullet i})_{i\in[\![n]\!]}$ *Hauptachsen* von Σ und die reellen Zufallsvariablen $(\langle X, u_{\bullet i}\rangle)_{i\in[\![n]\!]}$ *Hauptkomponenten* von X. □

§26.10 **Lemma**. *Seien $u_{\bullet i} \in \mathbb{R}^n$, $i \in [\![n]\!]$, Hauptachsen zu den der Größe nach angeordneten Eigenwerte $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_n \geqslant 0$ der Kovarianzmatrix $\Sigma = \mathbb{K}\mathrm{ov}(X)$ eines $\mathbb{R}^n$-wertigen Zufallsvektors X in $\mathcal{L}_2$. Dann sind die Hauptkomponenten $(\langle X, u_{\bullet i}\rangle)_{i\in[\![n]\!]}$ unkorreliert. Es gilt*

$$\mathbb{K}\mathrm{ov}(\langle X, u_{\bullet i}\rangle, \langle X, u_{\bullet j}\rangle) = \langle \Sigma u_{\bullet i}, u_{\bullet j}\rangle = \lambda_i \mathbb{1}_{\{i\}}(j),$$

also $\mathbb{V}\mathrm{ar}(\langle X, u_{\bullet i}\rangle) = \lambda_i$ für alle $i, j \in [\![n]\!]$ und $\mathbb{V}\mathrm{ar}(\langle X, u_{\bullet 1}\rangle) \geqslant \mathbb{V}\mathrm{ar}(\langle X, u_{\bullet 2}\rangle) \geqslant \dots \geqslant \mathbb{V}\mathrm{ar}(\langle X, u_{\bullet n}\rangle)$.

§26.11 **Beweis** von Lemma §26.10. Direktes Anwenden der Definition. □

§26.12 **Beispiel**. Seien $X \sim \mathrm{N}_{(\mu,\Sigma)}$, $\lambda = (\lambda_i)_{i\in[\![n]\!]}$ die Eigenwerte von Σ und die Hauptachsen $(u_{\bullet i})_{i\in[\![n]\!]}$ von Σ die Spalten der orthogonaler Matrix U. Dann gilt $\mathrm{U}^{\mathrm{T}}X \sim \mathrm{N}_{(\mathrm{U}^{\mathrm{T}}\mu, \mathrm{Diag}[\lambda])}$ und die Hauptkomponenten $(\langle X, u_{\bullet i}\rangle)_{i\in[\![n]\!]}$ sind somit unabhängig. □

§26.13 **Definition**. Seien X und Y ein $\mathbb{R}^k$-wertiger bzw. $\mathbb{R}^n$-wertiger Zufallsvektor in $\mathcal{L}_2$. Für $\mathrm{A}^\star \in \mathbb{R}^{n\times k}$ und $b^\star \in \mathbb{R}^n$ heißt $Z^\star = \mathrm{A}^\star X + b^\star$ eine *lineare Vorhersage* von Y durch X. $Z^\star$ wird *beste lineare Vorhersage* von Y durch X genannt, wenn

$$\mathbb{E}(\|Y - Z^\star\|^2) \leqslant \mathbb{E}(\|Y - (\mathrm{A}X + b)\|^2)$$

für alle $\mathrm{A} \in \mathbb{R}^{n\times k}$ und $b \in \mathbb{R}^n$ gilt. □

Die folgenden Ausführungen in §26.14-§26.19, die erlauben, das zentrale Resultat der Hauptkomponentenanalyse im Satz §26.20 zu beweisen, sind für diese Vorlesung nur weiterführendes Material.

§26.14 **Bemerkung**. Für eine Matrix $\mathrm{A} \in \mathbb{R}^{n\times k}$ heißt eine Matrix $\boxed{\mathrm{A}^\dagger} \in \mathbb{R}^{k\times n}$ *Moore-Penrose Inverse* von A, wenn

(MPI1) $\mathrm{A}\mathrm{A}^\dagger\mathrm{A} = \mathrm{A}$;

(MPI2) $\mathrm{A}^\dagger\mathrm{A}\mathrm{A}^\dagger = \mathrm{A}^\dagger$;

(MPI3) $\mathrm{A}\mathrm{A}^\dagger$ und $\mathrm{A}^\dagger\mathrm{A}$ sind symmetrisch.

Die Moore-Penrose Inverse ist eindeutig festgelegt. Sei $\mathrm{B} \in \mathbb{R}^{n\times n}_{\geqslant}$ positiv semidefinit, so dass $\mathrm{U}^{\mathrm{T}}\mathrm{B}\mathrm{U} = \mathrm{Diag}[\lambda_i]$ für eine orthogonale Matrix U und eine Diagonalmatrix $\mathrm{Diag}[\lambda_i]$ mit nicht-negativen Diagonaleinträgen $\lambda = (\lambda_i)_{i\in[\![n]\!]} \in \mathbb{R}^n_{\geqslant 0}$ gilt. Setzen wir

$$\lambda^\dagger = (\lambda_i^\dagger)_{i\in[\![n]\!]} \in \mathbb{R}^n_{\geqslant 0} \quad \text{mit} \quad \lambda_i^\dagger := \begin{cases} \lambda_i^{-1} & \text{, falls } \lambda_i \in \mathbb{R}_{>0}, \\ 0 & \text{, sonst.} \end{cases}$$

Dann gilt $\mathrm{B}^\dagger = \mathrm{U}\mathrm{Diag}[\lambda_i^\dagger]\mathrm{U}^{\mathrm{T}}$ für die Moore-Penrose Inverse von B und $\mathrm{A}^\dagger = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^\dagger\mathrm{A}^{\mathrm{T}} \in \mathbb{R}^{k\times n}$ für die Moore-Penrose Inverse einer beliebigen Matrix $\mathrm{A} \in \mathbb{R}^{n\times k}$. □

§26.15 **Lemma**. *Seien X und Y ein $\mathbb{R}^k$-wertiger bzw. $\mathbb{R}^n$-wertiger Zufallsvektor in $\mathcal{L}_2$. Setzten wir $\mathbb{K}\mathrm{ov}(Y,X) := \mathbb{E}((Y - \mathbb{E}(Y))(X - \mathbb{E}(X))^{\mathrm{T}})$ dann ist*

$$Z^\star = \mathbb{E}(Y) + \mathbb{K}\mathrm{ov}(Y,X)(\mathbb{K}\mathrm{ov}(X))^\dagger(X - \mathbb{E}(X))$$

eine beste lineare Vorhersage von Y durch X. Der Fehler $\varepsilon^\star := Y - Z^\star$ und $Z = \mathrm{A}X + b$ für beliebiges $\mathrm{A} \in \mathbb{R}^{n\times k}$ und $b \in \mathbb{R}^n$ sind unkorreliert, also $\mathbb{K}\mathrm{ov}(\varepsilon^\star, Z) = \mathbb{0}_{n\times n}$. Es gilt $\mathbb{E}(\varepsilon^\star) = \mathbb{0}_n$ und

$$\mathbb{K}\mathrm{ov}(\varepsilon^\star) = \mathbb{K}\mathrm{ov}(Y) - \mathbb{K}\mathrm{ov}(Y,X)(\mathbb{K}\mathrm{ov}(X))^\dagger\mathbb{K}\mathrm{ov}(X,Y).$$

§26.16 **Beweis** von Lemma §26.15. Wir setzen

$$\begin{aligned} \mathrm{A}^\star &:= \mathbb{K}\mathrm{ov}(Y,X)(\mathbb{K}\mathrm{ov}(X))^\dagger, \\ b^\star &:= \mathbb{E}(Y) - \mathbb{K}\mathrm{ov}(Y,X)(\mathbb{K}\mathrm{ov}(X))^\dagger\mathbb{E}(X), \end{aligned}$$

sodass $Z^\star = \mathrm{A}^\star X + b^\star$ und

$$\mathbb{E}(\varepsilon^\star) = \mathbb{E}(Y - Z^\star) = \mathbb{E}(Y - \mathbb{E}(Y)) - \mathrm{A}^\star\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{0}_n$$

gilt. Für $Z = \mathrm{A}X + b$ mit beliebigen $\mathrm{A} \in \mathbb{R}^{n\times k}$ und $b \in \mathbb{R}^n$ folgt damit

$$\begin{aligned} \mathbb{K}\mathrm{ov}(\varepsilon^\star, Z) &= \mathbb{E}(\varepsilon^\star X^{\mathrm{T}})\mathrm{A}^{\mathrm{T}} + \mathbb{0}_n b^{\mathrm{T}} - \mathbb{0}_n(\mathbb{E}(X^{\mathrm{T}})\mathrm{A}^{\mathrm{T}} + b^{\mathrm{T}}) \\ &= \mathbb{E}((Y - \mathbb{E}(Y))X^{\mathrm{T}})\mathrm{A}^{\mathrm{T}} - \mathrm{A}^\star\mathbb{E}((X - \mathbb{E}(X))X^{\mathrm{T}})\mathrm{A}^{\mathrm{T}} \\ &= \mathbb{K}\mathrm{ov}(Y,X)(\mathrm{I}_{k\times k} - (\mathbb{K}\mathrm{ov}(X))^\dagger\mathbb{K}\mathrm{ov}(X))\mathrm{A}^{\mathrm{T}} = \mathbb{0}_n. \end{aligned}$$

Die letzte Gleichheit ist klar, falls $\mathbb{K}\mathrm{ov}(X)$ regulär ist, ansonsten ist diese aufwendiger zu zeigen. (Wir nutzen aus, dass $\mathrm{I}_{k\times k} - (\mathbb{K}\mathrm{ov}(X))^\dagger \mathbb{K}\mathrm{ov}(X)$ eine Projektionsmatrix auf das orthogonale Komplement des Bildes von $\mathbb{K}\mathrm{ov}(X)$ ist, und zeigen dass für jedes darin enthaltene Elemente v gilt $v^{\mathrm{T}}X = v^{\mathrm{T}}\mathbb{E}(X)$ $\mathbb{P}$-f.s. und somit $\mathbb{K}\mathrm{ov}(Y,X)v = \mathbb{O}_n$.) Für jede lineare Vorhersage $Z = \mathrm{A}X + b$ von Y durch X gilt dann $\mathbb{K}\mathrm{ov}(Y - Z^\star, Z^\star - Z) = \mathbb{K}\mathrm{ov}(\varepsilon^\star, (\mathrm{A}^\star - \mathrm{A})X) = \mathbb{O}_n$ und somit

$$\begin{aligned}\mathbb{K}\mathrm{ov}(Y-Z) &= \mathbb{K}\mathrm{ov}(Y-Z^\star) + \mathbb{K}\mathrm{ov}(Z^\star - Z) + \mathbb{K}\mathrm{ov}(Y-Z^\star, Z^\star - Z) + \mathbb{K}\mathrm{ov}(Z^\star - Z, Y - Z^\star)\\ &= \mathbb{K}\mathrm{ov}(Y-Z^\star) + (\mathrm{A}^\star - \mathrm{A})\mathbb{K}\mathrm{ov}(X)(\mathrm{A}^\star - \mathrm{A})^{\mathrm{T}}.\end{aligned}$$

Damit ist die Matrix $\mathbb{K}\mathrm{ov}(Y-Z) - \mathbb{K}\mathrm{ov}(Y-Z^\star) \in \mathbb{R}^{n\times n}_{\geqslant}$ positiv semidefinit, und wir schreiben kurz $\mathbb{K}\mathrm{ov}(Y-Z) \geqslant \mathbb{K}\mathrm{ov}(Y-Z^\star)$. Wir halten weiter fest, dass dann auch gilt

$$\mathrm{Spur}(\mathbb{K}\mathrm{ov}(Y-Z)) \geqslant \mathrm{Spur}(\mathbb{K}\mathrm{ov}(Y-Z^\star)).$$

Wir benutzen weiterhin, dass $\mathbb{E}(\|Y-Z\|^2) = \mathrm{Spur}(\mathbb{K}\mathrm{ov}(Y-Z))$ gilt, und schließen

$$\mathbb{E}(\|Y-Z\|^2) = \mathrm{Spur}(\mathbb{K}\mathrm{ov}(Y-Z)) \geqslant \mathrm{Spur}(\mathbb{K}\mathrm{ov}(Y-Z^\star)) = \mathbb{E}(\|Y-Z^\star\|^2),$$

was zu zeigen ist. □

§26.17 **Korollar**. *Sei Y ein $\mathbb{R}^n$-wertiger Zufallsvektor in $\mathcal{L}_2$ und $X := \mathrm{A}Y$ für ein $\mathrm{A} \in \mathbb{R}^{k\times n}$. Dann ist*

$$Z^\star = \mathbb{E}(Y) + \mathbb{K}\mathrm{ov}(Y)\mathrm{A}^{\mathrm{T}}(\mathrm{A}\mathbb{K}\mathrm{ov}(Y)\mathrm{A}^{\mathrm{T}})^\dagger(X - \mathbb{E}(X))$$

eine beste lineare Vorhersage von Y durch $X = \mathrm{A}Y$. Der Fehler $\varepsilon^\star := Y - Z^\star$ und $Z^\star$ sind unkorreliert. Es gilt

$$\mathbb{K}\mathrm{ov}(\varepsilon^\star) = \mathbb{K}\mathrm{ov}(Y) - \mathbb{K}\mathrm{ov}(Y)\mathrm{A}^{\mathrm{T}}(\mathrm{A}\mathbb{K}\mathrm{ov}(Y)\mathrm{A}^{\mathrm{T}})^\dagger\mathrm{A}\mathbb{K}\mathrm{ov}(Y)$$

und $\mathbb{E}(\varepsilon^\star) = 0$.

§26.18 **Beweis** von Korollar §26.17. Folgt direkt aus Lemma §26.15. □

§26.19 **Bemerkung**. Seien die Spalten $u_{\bullet i} \in \mathbb{R}^n$, $i \in [\![n]\!]$, der orthogonalen Matrix U Hauptachsen der Kovarianzmatrix $\Sigma := \mathbb{K}\mathrm{ov}(Y)$ eines $\mathbb{R}^n$-wertigen Zufallsvektors Y. Für $\mu := \mathbb{E}(Y)$ gilt dann

$$Y - \mu = \mathrm{U}\mathrm{U}^{\mathrm{T}}(Y-\mu) = \sum_{i\in[\![n]\!]}\langle Y-\mu, u_{\bullet i}\rangle u_{\bullet i}.$$

Bezeichnet $(v_j)_{j\in[\![n]\!]}$ eine beliebige Orthonormalbasis des $\mathbb{R}^n$, so gilt

$$\begin{aligned}\mathbb{E}(\|Y-\mu\|^2) = \sum_{j\in[\![n]\!]}\mathbb{E}(|\langle Y-\mu, v_j\rangle|^2) = \sum_{j\in[\![n]\!]}\mathbb{V}\mathrm{ar}(\langle Y, v_j\rangle) &= \sum_{j\in[\![n]\!]}\langle \Sigma v_j, v_j\rangle\\ &= \mathrm{Spur}(\Sigma) = \sum_{j\in[\![n]\!]}\lambda_j.\end{aligned}$$

Im Folgenden nehmen wir an, dass Y zentriert ist, das heißt, $\mu = \mathbb{O}_n$ gilt. Für $k \in [\![n]\!]$ sei $\mathrm{U}_{(k)} := [u_{\bullet 1} \cdots u_{\bullet k}] \in \mathbb{R}^{n\times k}$ die Matrix mit den Spalten $(u_{\bullet j})_{j\in[\![k]\!]}$. Für

$$Z_{(k)} := \mathrm{U}_{(k)}\mathrm{U}_{(k)}^{\mathrm{T}}Y = \sum_{j\in[\![k]\!]}\langle Y, u_{\bullet j}\rangle u_{\bullet j}$$

gilt dann

$$\mathbb{E}(\|Y - Z_{(k)}\|^2) = \sum_{i\in(\!(k,n]\!]}\mathbb{E}(|\langle Y, u_{\bullet i}\rangle|^2) = \sum_{i\in(\!(k,n]\!]}\lambda_i.$$

Weiterhin sei $k^\star := \operatorname{Rang}(\Sigma) \in [\![n]\!]$ (äquivalent dazu $Y \neq \mathbb{0}_n$ $\mathbb{P}$-f.s. oder $\mathbb{E}(\|Y\|^2) \in \mathbb{R}_{>0}$) und $k \in (\!|k^\star, n]\!]$, so gilt $\lambda_l = 0$ für jedes $l \in (\!|k^\star, k]\!]$ und somit $\langle Y, u_{\bullet l}\rangle = 0$ $\mathbb{P}$-f.s. (da $\mathbb{E}(\langle Y, u_{\bullet l}\rangle) = \langle \mu, u_{\bullet l}\rangle = 0$ und $\mathbb{V}\mathrm{ar}(\langle Y, u_{\bullet l}\rangle) = \lambda_l = 0$) sowie

$$Z_{(k)} = \sum_{i \in [\![k]\!]} \langle Y, u_{\bullet i}\rangle u_{\bullet i} = \sum_{i \in [\![k^\star]\!]} \langle Y, u_{\bullet i}\rangle u_{\bullet i} = Z_{(k^\star)} \quad \mathbb{P}\text{-f.s.} \qquad \forall k \in (\!|k^\star, n]\!].$$

Sei also $k \in [\![k^\star]\!]$ und damit $\lambda_l \in \mathbb{R}_{>0}$ für jedes $l \in [\![k]\!]$. Betrachten wir den Zufallsvektor $X_{(k)} := \mathrm{U}_{(k)}^{\mathrm{T}} Y$ der ersten k Hauptkomponenten von Y, so gilt $\Sigma \mathrm{U}_{(k)} = \mathrm{U}_{(k)} \mathrm{Diag}[\lambda_\cdot]$ sowie $\mathrm{U}_{(k)}^{\mathrm{T}} \Sigma \mathrm{U}_{(k)} = \mathrm{Diag}[\lambda_\cdot]$ mit positiven Diagonaleinträgen $(\lambda_i)_{i \in [\![k]\!]} \in \mathbb{R}^k_{>0}$, also $(\mathrm{U}_{(k)}^{\mathrm{T}} \Sigma \mathrm{U}_{(k)})^\dagger = \mathrm{Diag}[\lambda_\cdot^{-1}]$. Unter Verwendung von Korollar §26.17 mit $X_{(k)} = \mathrm{A}Y$, $\mathrm{A} = \mathrm{U}_{(k)}^{\mathrm{T}}$, $\mathbb{E}(Y) = \mathbb{0}_n$ sowie $\mathbb{E}(X_{(k)}) = \mathbb{0}_k$ folgt nun, dass

$$\begin{aligned} Z^\star_{(k)} &= \mathbb{K}\mathrm{ov}(Y)\mathrm{A}^{\mathrm{T}}(\mathrm{A}\mathbb{K}\mathrm{ov}(Y)\mathrm{A}^{\mathrm{T}})^\dagger X_{(k)} \\ &= \Sigma \mathrm{U}_{(k)}(\mathrm{U}_{(k)}^{\mathrm{T}}\Sigma\mathrm{U}_{(k)})^\dagger X_{(k)} = \mathrm{U}_{(k)}\mathrm{Diag}[\lambda_\cdot]\mathrm{Diag}[\lambda_\cdot^{-1}]\mathrm{U}_{(k)}^{\mathrm{T}} Y = \mathrm{U}_{(k)}\mathrm{U}_{(k)}^{\mathrm{T}} Y = Z_{(k)} \end{aligned}$$

eine beste lineare Vorhersage von Y durch $X_{(k)}$ ist. □

§26.20 **Satz**. *Sei Y ein zentrierter $\mathbb{R}^n$-wertiger Zufallsvektor in $\mathcal{L}_2$ mit Hauptachsen $(u_{\bullet i})_{i \in [\![n]\!]}$ von $\mathbb{K}\mathrm{ov}(Y)$. Für $k \in [\![\operatorname{Rang}(\mathbb{K}\mathrm{ov}(Y))]\!]$ sei $\mathrm{U}_{(k)} := [u_{\bullet 1} \cdots u_{\bullet k}] \in \mathbb{R}^{n \times k}$ die Matrix mit den Spalten $(u_{\bullet j})_{j \in [\![k]\!]}$ und sei $X_{(k)} := \mathrm{U}_{(k)}^{\mathrm{T}} Y$ der $\mathbb{R}^k$-wertige Zufallsvektor der ersten k Hauptkomponenten $(\langle Y, u_{\bullet i}\rangle)_{i \in [\![k]\!]}$ von Y. Bezeichne mit $Z^\star_{(k)} := \mathrm{U}_{(k)}\mathrm{U}_{(k)}^{\mathrm{T}} Y = \sum_{i \in [\![k]\!]} \langle Y, u_{\bullet i}\rangle u_{\bullet i}$ eine beste lineare Vorhersage von Y durch $X_{(k)}$. Für jedes $\mathrm{A} \in \mathbb{R}^{k \times n}$ bezeichne $Z^\star_{\mathrm{A}}$ eine beste lineare Vorhersage von Y durch $\mathrm{A}Y$, dann gilt*

$$\mathbb{E}(\|Y - Z^\star_{(k)}\|^2) = \min\{\mathbb{E}(\|Y - Z^\star_{\mathrm{A}}\|^2) : \mathrm{A} \in \mathbb{R}^{k \times n}\}.$$

Damit ist $Z^\star_{(k)}$ eine beste lineare Vorhersage von Y unter allen linearen Vorhersagen von Y durch $\mathrm{A}Y$ mit $\mathrm{A} \in \mathbb{R}^{k \times n}$.

§26.21 **Beweis** von Satz §26.20. (Schritt 1) Nach Korollar §26.17 ist $Z^\star_{\mathrm{A}} = \Sigma\mathrm{A}^{\mathrm{T}}(\mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}})^\dagger \mathrm{A}Y$ eine beste lineare Vorhersage von Y durch $\mathrm{A}Y$ für $\mathrm{A} \in \mathbb{R}^{k \times n}$ mit mittlerem quadratischen Vorhersagefehler

$$\mathbb{E}(\|Y - Z^\star_{\mathrm{A}}\|^2) = \operatorname{Spur}(\Sigma - \Sigma\mathrm{A}^{\mathrm{T}}(\mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}})^\dagger\mathrm{A}\Sigma) = \operatorname{Spur}(\Sigma) - \operatorname{Spur}(\Sigma\mathrm{A}^{\mathrm{T}}(\mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}})^\dagger\mathrm{A}\Sigma)$$

Wir bestimmen für

$$\min\{\mathbb{E}(\|Y - Z^\star_{\mathrm{A}}\|^2) : \mathrm{A} \in \mathbb{R}^{k \times n}\} = \operatorname{Spur}(\Sigma) - \max\{\operatorname{Spur}(\Sigma\mathrm{A}^{\mathrm{T}}(\mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}})^\dagger\mathrm{A}\Sigma) : \mathrm{A} \in \mathbb{R}^{k \times n}\}$$

eine untere Schranke.

(Schritt 2) Für $\mathrm{B} = \Sigma^{1/2}\mathrm{A}^{\mathrm{T}} \in \mathbb{R}^{n \times k}$ ist $\mathrm{B}(\mathrm{B}^{\mathrm{T}}\mathrm{B})^\dagger\mathrm{B}^{\mathrm{T}} = \Pi_{\mathrm{Bild(B)}}$ die Darstellungsmatrix der orthogonalen Projektion auf $\mathrm{Bild}(\mathrm{B}) \subseteq \mathbb{R}^n$ und es gilt

$$\begin{aligned} \operatorname{Spur}(\Sigma\mathrm{A}^{\mathrm{T}}(\mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}})^\dagger\mathrm{A}\Sigma) &= \operatorname{Spur}(\Sigma^{1/2}\mathrm{B}(\mathrm{B}^{\mathrm{T}}\mathrm{B})^\dagger\mathrm{B}^{\mathrm{T}}\Sigma^{1/2}) \\ &= \operatorname{Spur}(\Sigma\mathrm{B}(\mathrm{B}^{\mathrm{T}}\mathrm{B})^\dagger\mathrm{B}^{\mathrm{T}}) = \operatorname{Spur}(\Pi_{\mathrm{Bild(B)}}\Sigma\Pi_{\mathrm{Bild(B)}}). \end{aligned}$$

(Schritt 3) Da $\operatorname{Rang}(\mathrm{B}) \in [\![0, k]\!]$ ist $\mathrm{Bild}(\mathrm{B}) \in \mathscr{V}_k := \{\mathcal{V} : \mathcal{V} \text{ Unterraum von } \mathbb{R}^n \text{ mit } \dim_K(\mathcal{V}) \in [\![0, k]\!]\}$ und es gilt

$$\begin{aligned} &\min\{\mathbb{E}(\|Y - Z^\star_{\mathrm{A}}\|^2) : \mathrm{A} \in \mathbb{R}^{k \times n}\} \\ &\qquad = \operatorname{Spur}(\Sigma) - \max\{\operatorname{Spur}(\Pi_{\mathrm{Bild(B)}}\Sigma\Pi_{\mathrm{Bild(B)}}) : \mathrm{B} = \Sigma^{1/2}\mathrm{A}^{\mathrm{T}} \in \mathbb{R}^{n \times k}\} \\ &\qquad \geqslant \operatorname{Spur}(\Sigma) - \max\{\operatorname{Spur}(\Pi_{\mathcal{V}}\Sigma\Pi_{\mathcal{V}}) : \mathcal{V} \in \mathscr{V}_k\} \end{aligned}$$

(Schritt 4) Für eine beliebige ONB $(v_i)_{i\in[\![l]\!]}$ von $\mathcal{V} \in \mathscr{V}_k$ gilt $\mathrm{Spur}(\Pi_\mathcal{V}\Sigma\Pi_\mathcal{V}) = \sum_{i\in[\![l]\!]} \langle \Sigma v_i, v_i\rangle$, somit entspricht dem Maximieren über alle Unterräume mit Dimension höchstens k gerade dem Maximieren über alle Orthonormalsysteme mit maximal k Elementen. Wir können nun per Induktion über k zeigen, dass

$$\max\{\mathrm{Spur}(\Pi_\mathcal{V}\Sigma\Pi_\mathcal{V}) : \mathcal{V} \in \mathscr{V}_k\} = \sum_{i\in[\![k]\!]} \langle \Sigma u_{\bullet i}, u_{\bullet i}\rangle = \sum_{i\in[\![k]\!]} \lambda_i.$$

(Schritt 5) Insgesamt also

$$\min\{\mathbb{E}(\|Y - Z^\star_{\mathrm{A}}\|^2) : \mathrm{A} \in \mathbb{R}^{k\times n}\} \geqslant \mathrm{Spur}(\Sigma) - \sum_{i\in[\![k]\!]} \lambda_i = \sum_{i\in(\![k,n]\!]} \lambda_i.$$

(Schritt 6) Die Behauptung folgt dann aus $\mathbb{E}(\|Y - Z^\star_{(k)}\|^2) = \sum_{i\in(\![k,n]\!]} \mathbb{E}(|\langle Y, u_{\bullet i}\rangle|^2) = \sum_{i\in(\![k,n]\!]} \lambda_i$. □

§26.22 **Bemerkung**. Das letzte Resultat ist die Grundlage für eine Vielzahl an Methoden zur Analyse von sehr großen Datensätzen. Betrachten wir die beste lineare Vorhersage $Z^\star_{(k)}$ von Y durch die ersten k Hauptkomponenten $(\langle Y, u_{\bullet i}\rangle)_{i\in[\![k]\!]}$ von Y, so ist der mittlere Vorhersagefehler gerade $\sum_{i\in(\![k,n]\!]} \mathbb{V}\mathrm{ar}(\langle Y, u_{\bullet i}\rangle)$. Wählen wir also die ersten k Hauptkomponenten, so können wir nur $\sum_{i\in(\![k,n]\!]} \mathbb{V}\mathrm{ar}(\langle Y, u_{\bullet i}\rangle)$ der gesamten Variabilität $\sum_{i\in[\![n]\!]} \mathbb{V}\mathrm{ar}(\langle Y, u_{\bullet i}\rangle)$ nicht erklären. Der erklärte Anteil wird typischerweise relativ angegeben, also $\frac{\sum_{i\in[\![k]\!]} \mathbb{V}\mathrm{ar}(\langle Y, u_{\bullet i}\rangle)}{\sum_{i\in[\![n]\!]} \mathbb{V}\mathrm{ar}(\langle Y, u_{\bullet i}\rangle)} \in [0,1]$. In vielen Anwendungen genügen nur wenige (zwei oder drei) Hauptkomponenten, um relativ 90% oder mehr der gesamten Variabilität zu erklären. □

§27 Statistische Inferenz: endliche Stichproben Eigenschaften

Seien $\mathbb{P}_\Theta$ eine Familie von Wahrscheinlichkeitsmaßen auf einem messbaren Raum $(\mathcal{X}, \mathscr{X})$ und $\gamma : \Theta \to \Gamma \subseteq \mathbb{R}$ ein identifizierbarer abgeleiteter Parameter. Für $n \in \mathbb{N}$ betrachten wir im Folgenden unabhängige und identisch-verteilte $\mathcal{X}$-wertige Zufallsvariablen X_i, $i \in [\![n]\!]$, mit identischer Verteilung aus $\mathbb{P}_\Theta$. Mit anderen Worten, die Zufallsvariable $X := (X_i)_{i\in[\![n]\!]}$ ist adäquat durch das statistische Produktexperiment $(\mathcal{X}^n, \mathscr{X}^n, \mathbb{P}^n_\Theta)$ beschrieben. Wir schreiben kurz $X \odot \mathbb{P}^n_\Theta$. Für jedes $\theta \in \Theta$ bezeichnet im Folgenden $\mathbb{E}^n_\theta$ stets die Erwartung bzgl. $\mathbb{P}^n_\theta$.

§27|01 Gleichmäßig bester unverfälschter Test

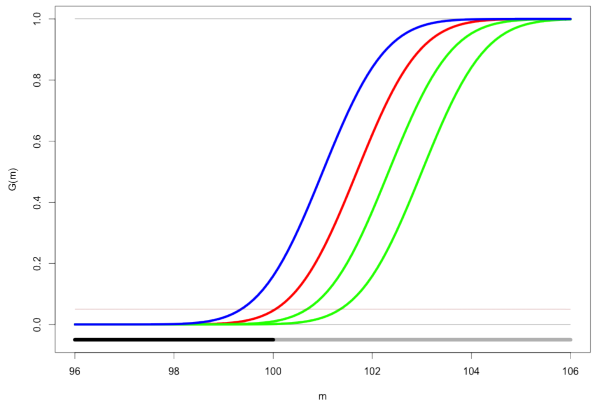
§27.01 **Definition**. Sei $\{\Gamma^0, \Gamma^1\}$ eine Partition der interessierenden Parameter und $\varphi : \mathcal{X}^n \to \{0,1\}$ ein Test, also $\varphi \in \mathcal{M}(\mathscr{X}^n)$, der Nullhypothese $H_0 : \Gamma^0$ gegen die Alternative $H_1 : \Gamma^1$. Die Abbildung

$$\boxed{\beta} : \Theta \ni \theta \mapsto \beta(\theta) := \mathbb{E}^n_\theta(\varphi) \in [0,1]$$

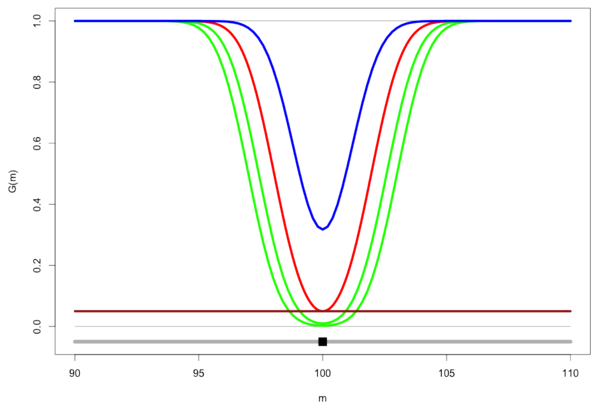
heißt *Gütefunktion* und ihre Werte $\beta(\theta)$ werden unter der Alternative, also für $\theta \in \gamma^{-1}(\Gamma^1)$, d.h. $\theta \in \Theta$ mit $\gamma(\theta) \in \Gamma^1$, *Macht* von φ genannt. Der Test φ heißt *unverfälscht* zum Niveau $\alpha \in [0,1]$, kurz *unverfälschter α-Test*, wenn φ das Niveau α einhält und seine Macht nicht kleiner als α ist, also $\beta(\theta) \geqslant \alpha$ für alle $\theta \in \gamma^{-1}(\Gamma^1)$ gilt. Ein Test φ heißt *gleichmäßig bester unverfälschter Test zum Niveau $\alpha \in [0,1]$*, falls er ein unverfälschter α-Test ist und für jedes $\theta \in \gamma^{-1}(\Gamma^1)$ die Macht $\beta_{\tilde\varphi}(\theta)$ eines jeden anderen unverfälschten α-Tests $\tilde\varphi$ nicht größer ist, d.h. $\beta_\varphi(\theta) \geqslant \beta_{\tilde\varphi}(\theta)$ gilt. □

§27.02 **Bemerkung**. Ein Test φ der Nullhypothese $H_0 : \Gamma^0$ gegen die Alternative $H_1 : \Gamma^1$ hält somit das Signifikanzniveau $\alpha \in [0,1]$ ein, falls für die assoziierte Gütefunktion $\beta(\theta) \leqslant \alpha$ für alle $\theta \in \gamma^{-1}(\Gamma^0)$ gilt. □

§27.03 **Beispiel** *(Normalverteilungsmodell §12.12).* Im stetigen statistischen Modell $X \odot (\mathrm{N}^n_{(\mu,1)})_{\mu\in\mathbb{R}}$ betrachte für beliebiges $\mu_0 \in \mathbb{R}$ das *einseitige Testproblem* der Nullhypothese $H_0 : \mu \leqslant \mu_0$, also $\Gamma^0_{\mu_0} = \mathbb{R}_{\leqslant\mu_0}$, gegen die Alternative $H_1 : \mu > \mu_0$, also $\Gamma^1_{\mu_0} = \mathbb{R}_{>\mu_0}$. Zur Erinnerung, da die Verteilungsfamilie $(\mathrm{N}^n_{(\mu,1)})_{\mu\in\mathbb{R}}$ einen monotonen Likelihood-Quotienten besitzt (Beispiel §12.12), ist für jedes $c \in \mathbb{R}$ der Neyman-Pearson-Test $\varphi^\star_c$ mit Ablehnbereich $\{\varphi^\star_c = 1\} = \{\overline{X}_n \geqslant c\}$ ein gleichmäßig bester Test zum Niveau $\beta_{\varphi^\star_c}(\mu_0) = \mathrm{N}^n_{(\mu_0,1)}(\overline{X}_n \geqslant c)$ für das Testproblem der *einfachen Nullhypothese* $H_0 : \mu = \mu_0$ gegen die *zusammengesetzte Alternative* $H_1 : \mu > \mu_0$. Da $\overline{X}_n \sim \mathrm{N}_{(\mu_0,1/n)}$ für $X \sim \mathrm{N}^n_{(\mu_0,1)}$ (vgl. Beispiel §17.13) erhalten wir $\beta_{\varphi^\star_c}(\mu_0) = \Phi(\sqrt{n}(-c+\mu_0))$ (vgl. Beispiel §10.12 (a)). Für $\alpha \in (0,1)$ und kritischen Wert $c_\alpha = \mu_0 + z_{1-\alpha}/\sqrt{n}$ mit α-Quantil z_α einer Standardnormalverteilung gilt dann $\beta_{\varphi^\star_{c_\alpha}}(\mu_0) = \alpha$ und $\beta_{\varphi^\star_{c_\alpha}}(\mu) \leqslant \alpha$ für alle $\mu \leqslant \mu_0$ (vgl. Beispiel §05.07 (c)). Somit ist für jedes $\alpha \in (0,1)$ der Neyman-Pearson-Test mit Ablehnbereich $\{\sqrt{n}(\overline{X}_n - \mu_0) \geqslant z_{1-\alpha}\}$ ein *gleichmäßig bester Test* zum Niveau α des *einseitigen Testproblems* der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$. In der folgenden Graphik sind für $n = 1$, $\mu_0 = 100$ und $\alpha = 0.05$ die Gütefunktionen des Test $\varphi^\star_c$ für die verschieden Werte $c = 1$ (blau), $c = 1.69 \approx z_{1-\alpha}$ (rot), $c = 2.33$ (grün) und $c = 3$ (grün) dargestellt.

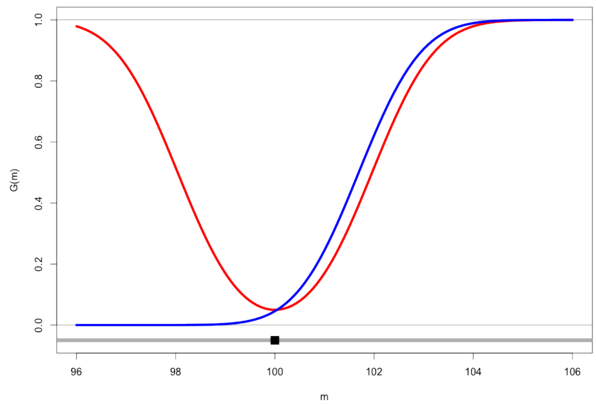


Für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$ ist jeder Test φ_c mit Ablehnbereich $\{\varphi_c = 1\} = \{\sqrt{n}|\overline{X}_n - \mu_0| \geqslant c\}$ mit $c \geqslant z_{1-\alpha/2}$ ein α-Test. In der folgenden Graphik sind für $n = 1$, $\mu_0 = 100$ und $\alpha = 0.05$ die Gütefunktionen des Test φ_c für die verschieden Werte $c = 1$ (blau), $c = 1.96 \approx z_{1-\alpha/2}$ (rot), $c = 2.58$ (grün) und $c = 3$ (grün) dargestellt.



Für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$ ist der zweiseitige Test $\varphi = \mathbb{1}_{\{\sqrt{n}|\overline{X}_n - \mu_0| \geqslant z_{1-\alpha/2}\}}$ kein gleichmäßig bester α Test, da der Test $\widetilde{\varphi} = \mathbb{1}_{\{\sqrt{n}(\overline{X}_n - \mu_0) \geqslant z_{1-\alpha}\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$ auch ein α-Test für das zweiseitige Testproblem ist und für alle $\mu > \mu_0$ eine größere Macht besitzt. In der folgenden Graphik sind für $n = 1$, $\mu_0 = 100$ und $\alpha = 0.05$ die

Gütefunktionen des zweiseitigen Tests φ (rot) und des einseitigen Tests $\widetilde{\varphi}$ (blau) dargestellt.



Offensichtlich ist die Macht des einseitigen Tests $\widetilde{\varphi}$ für alle Parameterwerte $\mu < \mu_0$ kleiner als α, so dass $\widetilde{\varphi}$ kein unverfälschter Test für das zweiseitige Testproblem ist. In der Vorlesung Statistik I zeigen wir, dass der zweiseitige Test φ optimal in der Klasse aller unverfälschten α-Test für das zweiseitige Testproblem der Nullhypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu \neq \mu_0$ ist. □

§27|02 Gleichmäßig bester unverfälschter Konfidenzbereich

§27.04 **Definition**. Seien $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ eine Familie von Partitionen in richtige und falsche interessierende Parameterwerte und B eine Bereichsschätzfunktion auf $(\mathcal{X}^n, \mathscr{X}^n)$, also $\{\gamma \in \mathrm{B}\} \in \mathscr{X}^n$ für alle $\gamma \in \Gamma$. Die Bereichsschätzfunktion B heißt *unverfälscht* zum Niveau 1-α, kurz *unverfälschter 1-α-Konfidenzbereich*, wenn B das Niveau 1-α einhält und $\mathbb{P}_\theta^n(\widetilde{\gamma} \in \mathrm{B}) \leqslant 1 - \alpha$ für alle $\widetilde{\gamma} \in \Gamma_{\gamma(\theta)}^{\mathrm{f}}$ und für alle $\theta \in \Theta$ gilt. □

§27.05 **Erinnerung**.

(a) Für $\gamma \in \Gamma$ sei $\Gamma_\gamma^{\mathrm{r}}$ und $\Gamma_\gamma^{\mathrm{f}}$ eine Partition in richtige und falsche interessierende Parameterwerte, dann bezeichnen $\Gamma_\gamma^0 = \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma_{\widetilde{\gamma}}^{\mathrm{r}}\}$ und $\Gamma_\gamma^1 = \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma_{\widetilde{\gamma}}^{\mathrm{f}}\}$ die *assoziierte Nullhypothese bzw. Alternative* der interessierenden Parameter.

(b) Zu einer Bereichsschätzfunktion B bezeichnet $(\varphi_\gamma)_{\gamma\in\Gamma}$ die *assoziierte Familie von Tests* mit Ablehnbereich $\{\varphi_\gamma = 1\} = \{\gamma \notin \mathrm{B}\}$.

(c) Zu einer $(\varphi_\gamma)_{\gamma\in\Gamma}$ Familie von Tests bezeichnet B die *assoziierte Bereichsschätzfunktion* mit $\mathrm{B}(x) := \{\gamma \in \Gamma : \varphi_\gamma(x) = 0\}$, $x \in \mathscr{X}^n$. □

§27.06 **Lemma**. *Seien $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ eine Familie von richtigen und falschen interessierenden Parameterwerte und $(\{\Gamma_\gamma^0, \Gamma_\gamma^1\})_{\gamma\in\Gamma}$ die assoziierte Familie von Nullhypothesen und Alternativen. Dann ist φ_γ ein (gleichmäßig bester) unverfälschter α-Test der Nullhypothese $H_0 : \Gamma_\gamma^0$ gegen die Alternative $H_1 : \Gamma_\gamma^1$ für jedes $\gamma \in \Gamma$, genau dann wenn die zu $(\varphi_\gamma)_{\gamma\in\Gamma}$ assoziierte Bereichsschätzfunktion B für $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ ein (gleichmäßig bester) unverfälschter 1-α-Konfidenzbereich ist.*

§27.07 **Beweis** von Lemma §27.06. In der Vorlesung. □

§27.08 **Beispiel** *(Normalverteilungsmodell §27.03)*. Im stetigen statistischen Modell $X \odot (\mathrm{N}_{(\mu,1)}^n)_{\mu\in\mathbb{R}}$ sind für beliebiges $\mu_0 \in \mathbb{R}$ und α-Quantil z_α der Standardnormalverteilung

- der rechtsseitige Test $\varphi_{z_{1-\alpha}}^r = \mathbb{1}_{\{\sqrt{n}(\overline{X}_n - \mu_0) \geqslant z_{1-\alpha}\}}$,
- der linksseitige Test $\varphi_{z_{1-\alpha}}^l = \mathbb{1}_{\{\sqrt{n}(\overline{X}_n - \mu_0) \leqslant -z_{1-\alpha}\}}$ und

- der beidseitige Test $\varphi^b_{z_{1-\alpha/2}} = \mathbb{1}_{\{\sqrt{n}|\overline{X_n}-\mu_0|\geqslant z_{1-\alpha/2}\}} = \varphi^r_{z_{1-\alpha/2}} + \varphi^l_{z_{1-\alpha/2}}$

für das entsprechende Testproblem gleichmäßig beste unverfälschte α-Tests. Damit sind für die Familie $(\{\Gamma^{\mathrm{r}}_\gamma, \Gamma^{\mathrm{f}}_\gamma\})_{\gamma\in\Gamma}$ der richtigen und falschen Parameter

(a) mit $\Gamma^{\mathrm{r}}_\mu = \mathbb{R}_{\geqslant\mu}$ und $\Gamma^{\mathrm{f}}_\mu = \mathbb{R}_{<\mu}$ der rechtsseitige KB $(\overline{X_n} - z_{1-\alpha}/\sqrt{n}, \infty)$;

(b) mit $\Gamma^{\mathrm{r}}_\mu = \mathbb{R}_{\leqslant\mu}$ und $\Gamma^{\mathrm{f}}_\mu = \mathbb{R}_{>\mu}$ der linksseitige KB $(-\infty, \overline{X_n} + z_{1-\alpha}/\sqrt{n})$;

(c) mit $\Gamma^{\mathrm{r}}_\mu = \{\mu\}$ und $\Gamma^{\mathrm{f}}_\mu = \mathbb{R}_{\neq\mu}$ der beidseitige KB $(\overline{X_n} \pm z_{1-\alpha/2}/\sqrt{n})$

gleichmäßig beste unverfälschte 1-α-Konfidenzbereiche. □

§27|03 Beste erwartungstreue Schätzfunktion

§27.09 **Definition**. Sei $\widehat{\gamma}$ eine Schätzfunktion, also $\widehat{\gamma} \in \mathcal{M}(\mathscr{X}^n)$ für den Parameter $\gamma \in \Gamma \subseteq \mathbb{R}$. Ist $\widehat{\gamma} \in \mathcal{L}_1(\mathbb{P}^n_\theta)$ für jedes $\theta \in \Theta$, so wird $\mathrm{Bias}_\theta(\widehat{\gamma}) := \mathbb{E}^n_\theta(\widehat{\gamma} - \gamma(\theta))$ *Verzerrung* oder *Bias* von $\widehat{\gamma}$ genannt. Der Schätzer heißt *erwartungstreu* oder *unverfälscht*, wenn $\mathrm{Bias}_\theta(\widehat{\gamma}) = 0$ für alle $\theta \in \Theta$ gilt. $\widehat{\gamma}$ *überschätzt* (bzw. *unterschätzt*) im Mittel γ, wenn $\mathrm{Bias}_\theta(\widehat{\gamma}) \in \mathbb{R}_{>0}$ (bzw. $\mathrm{Bias}_\theta(\widehat{\gamma}) \in \mathbb{R}_{<0}$) für alle $\theta \in \Theta$ gilt.

§27.10 **Erinnerung**. Für $k \in \mathbb{N}$ bezeichnet $\mathcal{W}_k(\mathscr{B})$ die Menge aller Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathscr{B})$ mit endlichem k-ten Moment, also für alle $\mathbb{P} \in \mathcal{W}_k(\mathscr{B})$ gilt $\mathrm{id}_\mathbb{R} \in \mathcal{L}_k(\mathbb{R}, \mathscr{B}, \mathbb{P})$ oder äquivalent dazu $\mathbb{E}(|\mathrm{id}_\mathbb{R}|^k) = \mathbb{P}(|\mathrm{id}_\mathbb{R}|^k) = \int_\mathbb{R} |x|^k \mathbb{P}(dx) \in \mathbb{R}_{\geqslant 0}$ (vgl. Definition §24.01). □

§27.11 **Definition**. Für $n \in \mathbb{N}$ sei $\mathcal{W}^n_k(\mathscr{B}) := \{\mathbb{P}^n = \bigotimes_{i\in[\![n]\!]} \mathbb{P} : \mathbb{P} \in \mathcal{W}_k(\mathscr{B})\}$ die Menge aller Produktwahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathscr{B}^n)$ mit identischer Randverteilung aus $\mathcal{W}_k(\mathscr{B})$. Für jedes $l \in [\![k]\!]$ heißen die abgeleiteten Parameter $\mathrm{m}^{(l)} : \mathcal{W}_k(\mathscr{B}) \to \mathbb{R}$ mit $\mathbb{P} \mapsto \mathrm{m}^{(l)}(\mathbb{P}) := \mathbb{P}(\mathrm{id}^l_\mathbb{R})$ und $\mathrm{M}^{(l)} : \mathcal{W}_k(\mathscr{B}) \to \mathbb{R}$ mit $\mathbb{P} \mapsto \mathrm{M}^{(l)}(\mathbb{P}) := \mathbb{P}((\mathrm{id}_\mathbb{R} - \mathrm{m}^{(1)}(\mathbb{P}))^l)$ das *l-te Moment* bzw. das *l-te zentrierte Moment*. Für jedes $\mathbb{P} \in \mathcal{W}_2(\mathscr{B})$ sind $\mu(\mathbb{P}) := \mathrm{m}^{(1)}(\mathbb{P})$ und $\sigma^2(\mathbb{P}) := \mathrm{M}^{(2)}(\mathbb{P})$ gerade der Erwartungswert und die Varianz von $\mathbb{P}$. □

§27.12 **Bemerkung**. Die (zentrierten) Momente sind identifizierbare abgeleitete Parameter und damit zum Beispiel auch $(\mathrm{m}^{(l)})_{l\in[\![k]\!]} : \mathcal{W}_k(\mathscr{B}) \to \mathbb{R}^k$, wobei die Abbildung $(\mathrm{m}^{(l)})_{l\in[\![k]\!]}$ auf $\mathcal{W}_k(\mathscr{B})$ offensichtlich nicht injektiv ist. Betrachten wir zum Beispiel ihre Einschränkung auf die Menge $\mathrm{N}_{\mathbb{R}\times\mathbb{R}_{\geqslant 0}} = \{\mathrm{N}_{(\mu,\sigma^2)} : (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_{\geqslant 0}\}$ von Normalverteilungen, so ist $\mathrm{m}^{(1)} = \mu$ und $\mathrm{M}^{(2)} = \sigma^2$. Der abgeleitete Parameter $(\mathrm{m}^{(1)}, \mathrm{M}^{(2)})$ ist auf $\mathrm{N}_{\mathbb{R}\times\mathbb{R}_{\geqslant 0}}$ somit injektiv, aber $\mathrm{m}^{(1)}$ oder $\mathrm{M}^{(2)}$ allein sind es nicht. □

§27.13 **Definition**. Für $l \in \mathbb{N}$ werden $\widehat{\mathrm{m}}^{(l)}_n \in \mathcal{M}(\mathscr{B}^n)$ mit

$$\mathbb{R}^n \ni x = (x_i)_{i\in[\![n]\!]} \mapsto \widehat{\mathrm{m}}^{(l)}_n(x) = \frac{1}{n} \sum_{i\in[\![n]\!]} x^l_i \in \mathbb{R}$$

und $\widehat{\mathrm{M}}^{(l)}_n \in \mathcal{M}(\mathscr{B}^n)$ mit

$$\mathbb{R}^n \ni x = (x_i)_{i\in[\![n]\!]} \mapsto \widehat{\mathrm{M}}^{(l)}_n(x) = \frac{1}{n} \sum_{i\in[\![n]\!]} (x_i - \mathrm{m}^{(1)})^l \in \mathbb{R}$$

empirisches Moment bzw. *zentriertes empirisches Moment* der Ordnung l genannt. □

§27.14 **Bemerkung**. Die Statistik $\widehat{\mathrm{M}}^{(l)}_n$ verwendet den abgeleiteten Parameter $\mathrm{m}^{(1)}$ und ist somit nur unter der unrealistischen Annahme, dass $\mathrm{m}^{(1)}$ bekannt ist, ein Schätzer für $\mathrm{M}^{(l)}$. Betrachten wir das statistische Produktexperiment $(\mathbb{R}^n, \mathscr{B}^n, \mathbb{W}_k^n(\mathscr{B}))$ so sind auf Grund der Linearität der Erwartung $\widehat{\mathrm{m}}_n^{(l)}$ und $\widehat{\mathrm{M}}_n^{(l)}$ *erwartungstreue Schätzer* für $\mathrm{m}^{(l)}$ bzw. $\mathrm{M}^{(l)}$. Für jedes $i \in [\![n]\!]$ erfüllt die Koordinatenprojektion $\Pi_i \in \mathcal{M}(\mathscr{B}^n)$ in der Tat $\mathrm{m}^{(l)}(\mathbb{P}) = \mathbb{P}^n(\Pi_i^l)$, so dass aus $\widehat{\mathrm{m}}_n^{(l)} = \frac{1}{n}\sum_{i\in[\![n]\!]} \Pi_i^l$ folgt

$$\mathrm{Bias}_{\mathbb{P}}(\widehat{\mathrm{m}}_n^{(l)}) = \mathbb{P}^n(\widehat{\mathrm{m}}_n^{(l)} - \mathrm{m}^{(l)}(\mathbb{P})) = \frac{1}{n}\sum_{i\in[\![n]\!]}\big(\mathbb{P}^n(\Pi_i^l) - \mathrm{m}^{(l)}(\mathbb{P})\big) = 0.$$

Weiterhin folgt aus $\mathrm{M}^{(l)}(\mathbb{P}) = \mathbb{P}^n((\Pi_i - \mathrm{m}^{(1)}(\mathbb{P}))^l)$, $i \in [\![n]\!]$, und $\widehat{\mathrm{M}}_n^{(l)} = \frac{1}{n}\sum_{i\in[\![n]\!]}(\Pi_i - \mathrm{m}^{(1)})^l$ auch

$$\mathrm{Bias}_{\mathbb{P}}(\widehat{\mathrm{M}}_n^{(l)}) = \mathbb{P}^n(\widehat{\mathrm{M}}_n^{(l)} - \mathrm{M}^{(l)}(\mathbb{P})) = \frac{1}{n}\sum_{i\in[\![n]\!]}\big(\mathbb{P}^n((\Pi_i - \mathrm{m}^{(1)}(\mathbb{P}))^l) - \mathrm{M}^{(l)}(\mathbb{P})\big) = 0.$$

Ist $\mathrm{m}^{(1)}$ nicht bekannt, und somit $\widehat{\mathrm{M}}_n^{(l)}$ als Schätzfunktion nicht mehr verwendbar, so betrachtet man üblicherweise die Statistik $\widehat{\mathrm{V}}_n^{(l)} = \frac{1}{n}\sum_{i\in[\![n]\!]}(\Pi_i - \widehat{\mathrm{m}}_n^{(1)})^l \in \mathcal{M}(\mathscr{B}^n)$ mit

$$\mathbb{R}^n \ni x = (x_i)_{i\in[\![n]\!]} \mapsto \widehat{\mathrm{V}}_n^{(l)}(x) := \frac{1}{n}\sum_{i\in[\![n]\!]}(x_i - \widehat{\mathrm{m}}_n^{(1)}(x))^l \in \mathbb{R}.$$

Die Statistik $\widehat{\mathrm{V}}_n^{(l)}$ erhalten wir aus der Statistik $\widehat{\mathrm{M}}_n^{(l)}$, in dem wir den Erwartungswert $\mathrm{m}^{(1)}(\mathbb{P}) = \mu(\mathbb{P})$ durch das empirische erste Moment $\overline{\Pi}_n := \frac{1}{n}\sum_{i\in[\![n]\!]}\Pi_i = \widehat{\mathrm{m}}_n^{(1)}$ ersetzen, diese Vorgehensweise wird häufig *Momentenmethode* genannt. Im Spezialfall $l = 2$ ist

$$\mathrm{M}^{(2)}(\mathbb{P}) = \mathbb{P}(|\mathrm{id}_{\mathbb{R}} - \mu(\mathbb{P})|^2) = \sigma^2(\mathbb{P})$$

gerade die Varianz von $\mathbb{P}$. Mit Hilfe elementarer Umformungen gilt

$$\widehat{\mathrm{V}}_n^{(2)} = \widehat{\mathrm{M}}_n^{(2)} - |\widehat{\mathrm{m}}_n^{(1)} - \mathrm{m}^{(1)}(\mathbb{P})|^2 = \frac{1}{n}\sum_{i\in[\![n]\!]}(\Pi_i - \mu(\mathbb{P}))^2 - |\overline{\Pi}_n - \mu(\mathbb{P})|^2.$$

Da $(\Pi_i)_{i\in[\![n]\!]}$ unabhängig und identisch verteilt sind, sind die Zufallsvariablen $Y_i := \frac{1}{n}(\Pi_i - \mu(\mathbb{P}))$, $i \in [\![n]\!]$, zentriert und unabhängig, sodass aus Lemma §25.03 (iii) folgt

$$\mathbb{P}^n(|\overline{\Pi}_n - \mu(\mathbb{P})|^2) = \mathbb{V}\mathrm{ar}(\sum_{i\in[\![n]\!]} Y_i) = \sum_{i\in[\![n]\!]}\mathbb{V}\mathrm{ar}(Y_i) = \frac{1}{n^2}\sum_{i\in[\![n]\!]}\mathbb{P}^n(|\Pi_i - \mu(\mathbb{P})|^2) = \frac{1}{n}\sigma^2(\mathbb{P}).$$

Damit erhalten wir $\mathrm{Bias}_{\mathbb{P}}(\widehat{\mathrm{V}}_n^{(2)}) = -\frac{1}{n}\sigma^2(\mathbb{P})$, da

$$\mathbb{P}^n(\widehat{\mathrm{V}}_n^{(2)}) = \mathbb{P}^n(\widehat{\mathrm{M}}_n^{(2)}) - \mathbb{P}^n(|\overline{\Pi}_n - \mu(\mathbb{P})|^2) = \sigma^2(\mathbb{P}) - \frac{1}{n}\sigma^2(\mathbb{P}) = \frac{n-1}{n}\sigma^2(\mathbb{P}),$$

sodass der Schätzer $\widehat{\mathrm{V}}_n^{(2)}$ im Mittel die Varianz σ^2 unterschätzt und ein erwartungstreuer Schätzer für σ^2 die Statistik $\frac{n}{n-1}\widehat{\mathrm{V}}_n^{(2)} = \frac{1}{n-1}\sum_{i\in[\![n]\!]}(\Pi_i - \overline{\Pi}_n)^2 \in \mathcal{M}(\mathscr{B}^n)$ ist. □

§27.15 **Definition**. Für $X \odot \mathbb{W}_2^n(\mathscr{B})$ heißen $\boxed{\overline{X}_n} = \widehat{\mathrm{m}}_n^{(1)}(X)$ und $\boxed{\widehat{\mathrm{S}}_n^{(2)}} := \frac{1}{n-1}\sum_{i\in[\![n]\!]}(X_i - \overline{X}_n)^2$ *empirischer Mittelwert* bzw. *empirische Varianz*. Wir schreiben abkürzend $\boxed{\widehat{\mathrm{S}}_n} := \sqrt{\widehat{\mathrm{S}}_n^{(2)}}$. □

Für jedes $\theta \in \Theta$ und jede Statistik $S \in \mathcal{L}_2(\mathbb{P}_\theta^n)$ bezeichnet $\mathbb{V}\mathrm{ar}_\theta(S) := \mathbb{E}_\theta^n(|S - \mathbb{E}_\theta^n(S)|^2)$ im Folgenden stets die Varianz von S bzgl. $\mathbb{P}_\theta^n$.

§27.16 **Definition**. Seien $\widehat{\gamma}, \widetilde{\gamma} \in \mathcal{L}_2(\mathbb{P}_\theta^n)$ für jedes $\theta \in \Theta$ zwei erwartungstreue Schätzfunktionen für γ. $\widehat{\gamma}$ heißt *(relativ) effizienter* als $\widetilde{\gamma}$, wenn $\mathbb{V}\mathrm{ar}_\theta(\widehat{\gamma}) \leqslant \mathbb{V}\mathrm{ar}_\theta(\widetilde{\gamma})$ für alle $\theta \in \Theta$ gilt und ein $\theta \in \Theta$ mit $\mathbb{V}\mathrm{ar}_\theta(\widehat{\gamma}) < \mathbb{V}\mathrm{ar}_\theta(\widetilde{\gamma})$ existiert. $\widehat{\gamma}$ wird *(absolut) effizient* genannt, falls es keinen erwartungstreuen Schätzer gibt, der (relativ) effizienter als $\widehat{\gamma}$ ist. □

§27.17 **Beispiel** *(§20.03 (d) fortgesetzt)*. Sei $X \odot (\mathrm{U}^n_{[0,\theta]})_{\theta \in \mathbb{R}_{>0}}$, so gilt

$$\mathrm{U}^n_{[0,\theta]}(\Pi_i) = \mu(\mathrm{U}_{[0,\theta]}) = \theta/2 \quad \text{und} \quad \mathbb{V}\mathrm{ar}_\theta(\Pi_i) = \sigma^2(\mathrm{U}_{[0,\theta]}) = \theta^2/12$$

für jede Koordinatenabbildung Π_i, $i \in [\![n]\!]$ (nachrechnen!). Wenden wir die Momentenmethode an, so erhalten wir den Schätzer $\widehat{\theta}_1 = 2\overline{X_n}$ für θ. Andererseits haben wir in Beispiel §12.23 (c) gezeigt, dass $\widehat{\theta}_2 = \max_{i \in [\![n]\!]} X_i$ der MLS für θ ist. Offensichtlich ist $\widehat{\theta}_1$ ein erwartungstreuer Schätzer für θ. Andererseits, nach Beispiel §20.03 (d) ist $\widehat{\theta}_2$ eine stetig-verteilte reelle Statistik mit Dichte

$$\mathbb{R} \ni x \mapsto \mathbb{f}_\theta(x) = \tfrac{n}{\theta^n} x^{n-1} \mathbb{1}_{[0,\theta]}(x) \in \mathbb{R},$$

sodass $\mathrm{U}^n_{[0,\theta]}(\widehat{\theta}_2) = \frac{n}{n+1}\theta$ und $\mathrm{Bias}_\theta(\widehat{\theta}_2) = -\theta/(n+1)$ gilt. Damit unterschätzt der MLS $\widehat{\theta}_2$ im Mittel θ. Der korrigierte MLS $\widehat{\theta}_3 := \frac{n+1}{n} \max_{i \in [\![n]\!]} X_i$ ist folglich erwartungstreu. Für $n \in \mathbb{N}_{\geqslant 2}$ ist der korrigierter MLS $\widehat{\theta}_3$ effizienter als der Momentenschätzer $\widehat{\theta}_1 = 2\overline{X_n}$, da $\mathbb{V}\mathrm{ar}_\theta(\widehat{\theta}_3) = \frac{1}{n(n+2)}\theta^2 < \frac{1}{3n}\theta^2 = \mathbb{V}\mathrm{ar}_\theta(\widehat{\theta}_1)$ für $n \in \mathbb{N}_{\geqslant 2}$ gilt (nachrechnen!). Für $n = 1$ ist $\widehat{\theta}_1 = 2X_1 = \widehat{\theta}_3$. In der Vorlesung Statistik I wird gezeigt, dass $\widehat{\theta}_3$ (absolut) effizient ist. □

§27.18 **Definition**. Für jedes $\theta \in \Theta$ und jede Statistik $S \in \mathcal{L}_2(\mathbb{P}_\theta)$ wird

$$\boxed{\mathrm{MSE}_\theta(S)} := \mathbb{E}^n_\theta(|S - \gamma(\theta)|^2)$$

mittlerer quadratischer Fehler (Mean Squared Error) von S genannt. Seien $\widehat{\gamma}, \widetilde{\gamma} \in \mathcal{L}_2(\mathbb{P}^n_\theta)$ für jedes $\theta \in \Theta$. $\widehat{\gamma}$ heißt *besser bzgl. des MSE* als $\widetilde{\gamma}$, wenn $\mathrm{MSE}_\theta(\widehat{\gamma}) \leqslant \mathrm{MSE}_\theta(\widetilde{\gamma})$ für alle $\theta \in \Theta$ gilt und ein $\theta \in \Theta$ mit $\mathrm{MSE}_\theta(\widehat{\gamma}) < \mathrm{MSE}_\theta(\widetilde{\gamma})$ existiert. □

§27.19 **Beispiel**.

(a) Für $X \odot (\mathrm{U}^n_{[0,\theta]})_{\theta \in \mathbb{R}_{>0}}$ und $n \in \mathbb{N}_{\geqslant 2}$ ist bzgl. des MSE der korrigierter Maximum-Likelihood-Schätzer (MLS) $\widehat{\theta}_3 = \frac{n+1}{n} \max\limits_{i \in [\![n]\!]} X_i$ besser als der MLS $\widehat{\theta}_2 = \max\limits_{i \in [\![n]\!]} X_i$. In der Tat es gilt

$$\mathrm{MSE}_\theta(\widehat{\theta}_3) = \frac{\theta^2}{n(n+2)} < \frac{2\theta^2}{(n+2)(n+1)} = \mathrm{MSE}_\theta(\widehat{\theta}_2).$$

Weiterhin ist der MLS $\widehat{\theta}_2 = \max_{i \in [\![n]\!]} X_i$ besser bzgl. des MSE als der Momentenschätzer $\widehat{\theta}_1 = 2\overline{X_n}$, da

$$\mathrm{MSE}_\theta(\widehat{\theta}_2) = \frac{2\theta^2}{(n+2)(n+1)} < \frac{\theta^2}{3n} = \mathrm{MSE}_\theta(\widehat{\theta}_1)$$

gilt (nachrechnen!).

(b) Für $X \odot (\mathrm{N}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}}$ und $n \in \mathbb{N}_{\geqslant 2}$ ist die empirische Varianz $\widehat{\mathrm{S}}^{(2)}_n$ ein (absolut) effizienter Schätzer für σ^2 (Vorlesung Statistik I) aber bzgl. des MSE gilt

$$\mathrm{N}^n_{(\mu,\sigma^2)}(|\widehat{\mathrm{V}}^{(2)}_n - \sigma^2|^2) = \mathrm{MSE}_{(\mu,\sigma^2)}(\widehat{\mathrm{V}}^{(2)}_n) = \tfrac{2n-1}{n^2}\sigma^4 < \tfrac{2}{n-1}\sigma^4 = \mathrm{MSE}_{(\mu,\sigma^2)}(\widehat{\mathrm{S}}^{(2)}_n),$$

sodass der Schätzer $\widehat{\mathrm{V}}^{(2)}_n$ für σ^2 besser als $\widehat{\mathrm{S}}^{(2)}_n$ bzgl. des MSE ist. □

§27|04 Lokations-Skalen-Modell

§27.20 **Definition.** Für einen $\mathbb{R}^n$-wertigen Zufallsvektor X ist ein *Lokations-Skalen-Modell* adäquat, wenn $X \odot \mathcal{W}_2^n(\mathscr{B})$ gilt, also die Koordinaten von $X = (X_i)_{i\in[\![n]\!]}$ sind unabhängige und identisch verteilte (u.i.v.) reelle Zufallsvariablen mit identischer Randverteilung $\mathbb{P} \in \mathcal{W}_2(\mathscr{B})$. Wir betrachten die identifizierbaren Parameter

$$\mu : \mathcal{W}_2(\mathscr{B}) \ni \mathbb{P} \mapsto \mu(\mathbb{P}) = \mathrm{m}^{(1)}(\mathbb{P}) = \mathbb{P}(\mathrm{id}_{\mathbb{R}}) = \int_{\mathbb{R}} x\mathbb{P}(dx) \in \mathbb{R}$$

sowie

$$\sigma^2 : \mathcal{W}_2(\mathscr{B}) \ni \mathbb{P} \mapsto \sigma^2(\mathbb{P}) = \mathrm{M}^{(2)}(\mathbb{P}) = \mathbb{P}(|\mathrm{id}_{\mathbb{R}} - \mu(\mathbb{P})|^2) = \int_{\mathbb{R}} |x - \mu(\mathbb{P})|^2\mathbb{P}(dx) \in \mathbb{R}_{\geqslant 0},$$

und schreiben abkürzend $X \odot (\mathrm{P}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}$. Wird die Varianz $\sigma_0^2 \in \mathbb{R}_{\geqslant 0}$ bzw. der Erwartungswert $\mu_0 \in \mathbb{R}$ als bekannt vorausgesetzt, so schreiben wir abkürzend $X \odot (\mathrm{P}^n_{(\mu,\sigma_0^2)})_{\mu\in\mathbb{R}}$ bzw. $X \odot (\mathrm{P}^n_{(\mu_0,\sigma^2)})_{\sigma^2\in\mathbb{R}_{>0}}$. □

§27.21 **Bemerkung.** Wir fassen unter $\mathrm{P}_{(\mu,\sigma^2)}$ alle $\mathbb{P} \in \mathcal{W}_2(\mathscr{B})$ mit $\mu = \mu(\mathbb{P})$ und $\sigma = \sigma(\mathbb{P})$ zusammen und die abkürzende Schreibweise $X \odot \mathrm{P}_{(\mu,\sigma^2)}$ meint $X \sim \mathbb{P}$ für ein $\mathbb{P} \in \mathrm{P}_{(\mu,\sigma^2)}$. □

§27.22 **Beispiel.**

(a) *Bernoulli-Schema*: $X \odot (\mathrm{B}^n_{\mathrm{p}})_{\mathrm{p}\in[0,1]}$, dann gilt $\mu(\mathrm{B}_{\mathrm{p}}) = \mathrm{p}$ und $\sigma^2(\mathrm{B}_{\mathrm{p}}) = \mathrm{p} - \mathrm{p}^2$ also auch $X \odot (\mathrm{P}^n_{(\mathrm{p},\mathrm{p}-\mathrm{p}^2)})_{\mathrm{p}\in[0,1]}$. Der MLS und Momentenschätzer $\widehat{\mathrm{p}} = \overline{X_n}$ für p ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).

(b) *Binomialverteilungsmodell*: $X \odot (\mathrm{Bin}^n_{(50,\mathrm{p})})_{\mathrm{p}\in[0,1]}$, dann gilt $\mu(\mathrm{Bin}_{(50,\mathrm{p})}) = 50\mathrm{p}$ und $\sigma^2(\mathrm{Bin}_{(50,\mathrm{p})}) = 50\mathrm{p}(1-\mathrm{p})$ also auch $X \odot (\mathrm{P}^n_{(50\mathrm{p},50(1-\mathrm{p})\mathrm{p})})_{\mathrm{p}\in[0,1]}$. Der MLS und Momentenschätzer $\widehat{\mathrm{p}} = \overline{X_n}/50$ für p ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).

(c) *Poissonverteilungsmodell*: $X \odot (\mathrm{Poi}^n_\lambda)_{\lambda\in\mathbb{R}_{>0}}$, dann gilt $\mu(\mathrm{Poi}_\lambda) = \lambda$ und $\sigma^2(\mathrm{Poi}_\lambda) = \lambda$ also auch $X \odot (\mathrm{P}^n_{(\lambda,\lambda)})_{\lambda\in\mathbb{R}_{>0}}$. Der MLS und Momentenschätzer $\widehat{\lambda} = \overline{X_n}$ für λ ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).

(d) *Exponentialverteilungsmodell*: $X \odot (\mathrm{Exp}^n_\lambda)_{\lambda\in\mathbb{R}_{>0}}$, dann gilt $\mu(\mathrm{Exp}_\lambda) = \lambda^{-1}$ und $\sigma^2(\mathrm{Exp}_\lambda) = \lambda^{-2}$ also auch $X \odot (\mathrm{P}^n_{(\lambda^{-1},\lambda^{-2})})_{\lambda\in\mathbb{R}_{>0}}$. Der MLS und Momentenschätzer $\widehat{\lambda} = (\overline{X_n})^{-1}$ überschätzt im Mittel λ mit $\mathrm{Bias}_\lambda(\widehat{\lambda}) = \lambda/(n-1)$, der korrigierter MLS $\widehat{\lambda}_2 = (\frac{n}{n-1}\overline{X_n})^{-1}$ ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I). □

§27.23 **Erinnerung.** In einem Lokations-Skalen-Modell $X \odot (\mathrm{P}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}$ ist das empirische Mittel $\overline{X_n}$ ein erwartungstreuer Schätzer für μ und die empirische Varianz $\widehat{\mathrm{S}}_n^{(2)}$ ein erwartungstreuer Schätzer für σ^2. Die folgenden Tests basieren nun auf den Statistiken $\frac{\sqrt{n}}{\widehat{\mathrm{S}}_n}(\overline{X_n} - \mu)$ bzw. $\frac{(n-1)}{\sigma^2}\widehat{\mathrm{S}}_n^{(2)}$ deren Verteilung wir unter einer zusätzlichen Normalverteilungsannahme in Abschnitt §27|06 herleiten. □

§27|04.24 **Testen des Erwartungswert.** Für einen $\mathbb{R}^n$-wertigen Zufallsvektor X sei ein Lokations-Skalen-Modell, also $X \odot (\mathrm{P}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}$ adäquat. Für $\mu_0 \in \mathbb{R}$ und $\mathrm{c} \in \mathbb{R}_{>0}$ betrachten wir

(a) den rechtsseitigen Test $\varphi^r_{\mathrm{c}} := \mathbb{1}_{\{\sqrt{n}(\overline{X_n}-\mu_0)\geqslant \mathrm{c}\widehat{\mathrm{S}}_n\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu - \mu_0 > 0$;

(b) den linksseitigen Test $\varphi^l_{\mathrm{c}} := \mathbb{1}_{\{\sqrt{n}(\overline{X_n}-\mu_0)\leqslant -\mathrm{c}\widehat{\mathrm{S}}_n\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \mu \geqslant \mu_0$ gegen die Alternative $H_1 : \mu - \mu_0 < 0$;

(c) den beidseitigen Test $\varphi^b_c := \mathbb{1}_{\{\sqrt{n}|\overline{X}_n - \mu_o| \geqslant c\widehat{S}_n\}}$ für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \mu = \mu_o$ gegen die Alternative $H_1 : \mu - \mu_o \neq 0$. □

§27|04.25 **Testen der Varianz**. Für einen $\mathbb{R}^n$-wertigen Zufallsvektor X sei ein Lokations-Skalen-Modell, also $X \circleddash (P^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}$ adäquat. Für $\sigma^2_o \in \mathbb{R}_{>0}$ und $c^r, c^l \in \mathbb{R}_{>0}$ betrachten wir

(a) den rechtsseitigen Test $\varphi^r_{c^r} := \mathbb{1}_{\{(n-1)\widehat{S}^{(2)}_n/\sigma^2_0 \geqslant c^r\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \sigma \leqslant \sigma_o$ gegen die Alternative $H_1 : \sigma/\sigma_o > 1$;

(b) den linksseitigen Test $\varphi^l_{c^l} := \mathbb{1}_{\{(n-1)\widehat{S}^{(2)}_n/\sigma^2_0 \leqslant c^l\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \sigma \geqslant \sigma_o$ gegen die Alternative $H_1 : \sigma/\sigma_o < 1$;

(c) den beidseitigen Test $\varphi^b_{c^l,c^r} := \varphi^l_{c^l} + \varphi^r_{c^r} = \mathbb{1}_{\{(n-1)\widehat{S}^{(2)}_n \leqslant c^l\sigma^2_0\}} + \mathbb{1}_{\{(n-1)\widehat{S}^{(2)}_n \geqslant c^r\sigma^2_0\}}$ für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \sigma = \sigma_o$ gegen die Alternative $H_1 : \sigma/\sigma_o \neq 1$. □

§27.26 **Bemerkung**. Möchten wir sicher stellen, dass die in den Definitionen §27|04.24 und §27|04.25 angegebenen Tests ein vorgegebenes Signifikanz-Niveau $\alpha \in (0,1)$ einhalten, so müssen wir die kritischen Werte $c, c^l, c^r \in \mathbb{R}_{>0}$ entsprechend wählen. Dies ist aber nur möglich, wenn die Verteilung der standardisierten Statistik $\frac{\sqrt{n}}{\widehat{S}_n}(\overline{X}_n - \mu_0)$ bzw. $\frac{(n-1)}{\sigma^2_0}\widehat{S}^{(2)}_n$ unter der Nullhypothese bekannt ist, wie es zum Beispiel im normalen Lokations-Skalen-Modell (Abschnitt §27|06) der Fall ist. Die angegebenen Tests können wir direkt benutzen (vgl. Satz §12.33), um einseitige bzw. zweiseitige Konfidenzbereiche für den Erwartungswert μ bzw. die Varianz σ^2) anzugeben. Zum Beispiel ist $\big(\overline{X}_n \pm c\widehat{S}_n/\sqrt{n}\big)$ bzw. $\big((n-1)\widehat{S}^{(2)}_n/c^r, (n-1)\widehat{S}^{(2)}_n/c^l\big)$) der assoziierte zweiseitige Konfidenzbereich. In Kapitel 6 untersuchen wir das asymptotische Verhalten der Verteilung von $\overline{X}_n$ sowie $\widehat{S}^{(2)}_n$, also wenn $n \to \infty$. Diese Resultate erlauben dann, die kritischen Werte $c, c^l, c^r \in \mathbb{R}_{>0}$ so zu wählen, dass zu mindesten asymptotisch, also für $n \to \infty$, das Signifikanzniveau eingehalten wird. □

§27|05 Zwei-Stichproben-Modell

§27|05.27 **Vorbemerkung**. Die zufälligen Messwerte der beiden Segmente des Fürsten G. Őmetry in Beispiel VI im Kapitel 1 Prolog können wir als Stichproben von $X \circleddash (P^{31}_{(\mu_X,\sigma^2_X)})_{(\mu_X,\sigma^2_X)\in\mathbb{R}\times\mathbb{R}_{>0}}$ und $Y \circleddash (P^{31}_{(\mu_Y,\sigma^2_Y)})_{(\mu_Y,\sigma^2_Y)\in\mathbb{R}\times\mathbb{R}_{>0}}$ auffassen, das heißt, sie können separat adäquat durch ein Lokations-Skalen-Modell beschrieben werden. Für $\mu_X = \overrightarrow{AC}$ und $\mu_Y = \overrightarrow{EB}$ ist damit die zu überprüfende Nullhypothese $H_0 : \mu_X = \mu_Y$, der interessierende Parameter ist also $\mu_X - \mu_Y$. Es erscheint realistisch, dass X und Y unabhängig sind. In diesem Fall sprechen wir von einem *Zwei-Stichproben-Modell mit unverbundenen Stichproben*. Im Folgenden nehmen wir an, dass X und Y unabhängig sind, aber dieselbe Varianz besitzen, das heißt, der Zufallsvektor (X, Y) ist adäquat durch das statistische Modell $(\mathbb{R}^{n+m}, \mathscr{B}^{n+m}, (P^n_{(\mu_X,\sigma^2)} \otimes P^m_{(\mu_Y,\sigma^2)})_{(\mu_X,\mu_Y,\sigma^2)\in\mathbb{R}^2\times\mathbb{R}_{>0}})$ beschrieben. Die Situation verschiedener Varianzen wird Behrens-Fischer Problem genannt, welches noch immer ungelöst ist. Für die interessierenden Parameter μ_X und μ_Y sind dann $\overline{X}_n$ bzw. $\overline{Y}_m$ erwartungstreue Momentenschätzer (Beispiel §27.22). Betrachten wir die Varianz σ^2, so sind $\widehat{S}^{(2)}_n := \frac{1}{n-1}\sum_{i\in[\![n]\!]}(X_i - \overline{X}_n)^2$ und $\widehat{S}^{(2)}_m := \frac{1}{m-1}\sum_{i\in[\![m]\!]}(Y_i - \overline{Y}_m)^2$ unabhängige erwartungstreue Momentenschätzer für σ^2, und somit ist der Schätzer $\boxed{\widehat{S}^{(2)}_{n,m}} := \frac{1}{n+m-2}[(n-1)\widehat{S}^{(2)}_n + (m-1)\widehat{S}^{(2)}_m]$ ebenfalls erwartungstreu. Es ist nun nahe liegend basierend auf der Statistik $\overline{X}_n - \overline{Y}_m$ für den interessierenden Parameter $\mu_X - \mu_Y$ zum Beispiel einen zweiseitigen Konfidenzbereich der Form $[\overline{X}_n - \overline{Y}_m \pm c\frac{\sqrt{n+m}}{\sqrt{nm}}\widehat{S}_{n,m}]$ mit $\boxed{\widehat{S}_{n,m}} := \sqrt{\widehat{S}^{(2)}_{n,m}}$ anzugeben und somit $\{0 \neq [\overline{X}_n - \overline{Y}_m \pm c\frac{\sqrt{n+m}}{\sqrt{nm}}\widehat{S}_{n,m}]\}$ als Ablehnbereich des assoziierten zweiseitigen Tests der Nullhypothese $H_0 : \mu_X - \mu_Y = 0$ gegen die Alternativhypothese $H_0 : \mu_X - \mu_Y \neq 0$ zu betrachten. □

§27|05.28 **Tests auf Gleichheit der Erwartungswerte.** Für zwei Zufallsvektoren X und Y mit identischer Varianz der Randverteilungen, sei ein Zwei-Stichproben-Modell mit unverbunden Stichproben, also $(X, Y) \circleddash (\mathrm{P}^n_{(\mu_X,\sigma^2)} \otimes \mathrm{P}^m_{(\mu_Y,\sigma^2)})_{(\mu_X,\mu_Y,\sigma^2)\in\mathbb{R}^2\times\mathbb{R}_{>0}}$, adäquat. Für $c \in \mathbb{R}_{>0}$ betrachten wir

(a) den rechtsseitigen Test $\varphi^r_c := \mathbb{1}_{\{\overline{X}_n - \overline{Y}_m \geqslant c\frac{\sqrt{n+m}}{\sqrt{nm}}\widehat{S}_{n,m}\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \mu_X \leqslant \mu_Y$ gegen die Alternative $H_1 : \mu_X - \mu_Y > 0$;

(b) den linksseitigen Test $\varphi^l_c := \mathbb{1}_{\{\overline{X}_n - \overline{Y}_m \leqslant -c\frac{\sqrt{n+m}}{\sqrt{nm}}\widehat{S}_{n,m}\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \mu_X \geqslant \mu_Y$ gegen die Alternative $H_1 : \mu_X - \mu_Y < 0$;

(c) den beidseitigen Test $\varphi^b_c := \varphi^l_c + \varphi^r_c = \mathbb{1}_{\{|\overline{X}_n - \overline{Y}_m| \geqslant c\frac{\sqrt{n+m}}{\sqrt{nm}}\widehat{S}_{n,m}\}}$ für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \mu_X = \mu_Y$ gegen die Alternative $H_1 : \mu_X - \mu_Y \neq 0$. □

§27.29 **Vorbemerkung.** Analog beschreiben wir die zufälligen Messwerte der abgefüllten Volumen in den 33cl bzw. 75cl Bierflaschen der Abbaye de Rochefort in Beispiel VI im Kapitel 1 Prolog als Zwei-Stichproben-Modell $(\mathbb{R}^{n+m}, \mathscr{B}^{n+m}, (\mathrm{P}^n_{(\mu_X,\sigma^2_X)} \otimes \mathrm{P}^m_{(\mu_Y,\sigma^2_Y)})_{\mu_X,\mu_Y\in\mathbb{R},\sigma^2_X,\sigma^2_Y\in\mathbb{R}_{>0}})$ mit unverbundenen Stichproben. Die Varianz σ^2_X bzw. σ^2_Y kann als Maß der Genauigkeit der Abfüllanlage interpretiert werden, so dass die Nullhypothese $H_0 : \sigma^2_X = \sigma^2_Y$ abzulehnen ist, wenn die Abfüllanlage nicht dieselbe Genauigkeit für die kleinen und großen Bierflaschen besitzt. Es ist nun nahe liegend, mit Hilfe der Statistik $\widehat{S}^{(2)}_n / \widehat{S}^{(2)}_m$ für den interessierenden Parameter σ^2_X/σ^2_Y Konfidenzbereiche bzw. Tests zu konstruieren. □

§27|05.30 **Tests auf Gleichheit der Varianzen.** Für X und Y sei ein Zwei-Stichproben-Modell mit unverbunden Stichproben, also $(X, Y) \circleddash (\mathrm{P}^n_{(\mu_X,\sigma^2_X)} \otimes \mathrm{P}^m_{(\mu_Y,\sigma^2_Y)})_{\mu_X,\mu_Y\in\mathbb{R},\sigma^2_X,\sigma^2_Y\mathbb{R}_{>0}}$, adäquat. Für $c^l, c^r \in \mathbb{R}_{>0}$ betrachten wir

(a) den rechtsseitigen Test $\varphi^r_{c^r} := \mathbb{1}_{\{\widehat{S}^{(2)}_n/\widehat{S}^{(2)}_m \geqslant c^r\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \sigma^2_X \leqslant \sigma^2_Y$ gegen die Alternative $H_1 : \sigma^2_X/\sigma^2_Y > 1$;

(b) den linksseitigen Test $\varphi^l_{c^l} := \mathbb{1}_{\{\widehat{S}^{(2)}_n/\widehat{S}^{(2)}_m \leqslant c^l\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \sigma^2_X \geqslant \sigma^2_Y$ gegen die Alternative $H_1 : \sigma^2_X/\sigma^2_Y < 1$;

(c) den beidseitigen Test $\varphi^b_{c^l,c^r} := \varphi^l_{c^l} + \varphi^r_{c^r} = \mathbb{1}_{\{\widehat{S}^{(2)}_n/\widehat{S}^{(2)}_m \leqslant c^l\}} + \mathbb{1}_{\{\widehat{S}^{(2)}_n/\widehat{S}^{(2)}_m \geqslant c^r\}}$ für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \sigma^2_X = \sigma^2_Y$ gegen die Alternative $H_1 : \sigma^2_X/\sigma^2_Y \neq 1$. □

§27.31 **Anmerkung.** Betrachten wir zum Beispiel den Verbrauch verschiedener Motorräder vor und nach einem Motortuning, so werden immer zwei Ergebnisse an einem Objekt gemessen. Beschreiben wir diese Situation durch ein Zwei-Stichproben-Modell mit $X \circleddash (\mathrm{P}^n_{(\mu_X,\sigma^2_X)})_{\mu_X\in\mathbb{R},\sigma^2_X\mathbb{R}_{>0}}$ (Verbrauch vor dem Tuning) und $Y \circleddash (\mathrm{P}^n_{(\mu_Y,\sigma^2_Y)})_{\mu_Y\in\mathbb{R},\sigma^2_Y\mathbb{R}_{>0}}$ (Verbrauch nach dem Tuning) so ist es unrealistisch anzunehmen, dass die Stichproben unabhängig sind. Ist der interessierende Parameter die Differenz des mittleren Verbrauchs $\mu_X - \mu_Y$, so betrachten wir die Differenz der Beobachtungen $X - Y := (X_i - Y_i)_{i\in[\![n]\!]}$, die dann adäquat durch ein Lokations-Skalen-Modell $X - Y \circleddash (\mathrm{P}^n_{(\mu,\sigma^2)})_{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_{>0}}$ beschrieben wird. Möchten wir Gleichheit $\mu_X = \mu_Y$ der Erwartungswerte testen, also gleicher mittlerer Verbrauch vor und nach dem Tuning, so können wir nun für die Differenz $X - Y$ direkt die Nullhypothese $H_0 : \mu = 0$ mit Hilfe der Definition §27|04.24 testen. □

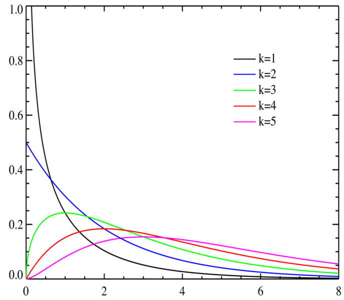
§27|06 Normales Lokations-Skalen-Modell

Im Folgenden sind $(Z_i)_{i\in[\![0,m+k]\!]}$ unabhängige und identisch standardnormalverteilte Zufallsvariablen, also $(Z_i)_{i\in[\![0,m+k]\!]} \sim \mathrm{N}^{1+m+k}_{(0,1)}$.

§27|06.32 **χ^2-Verteilung.**. Die Verteilung der Zufallsvariable

$$Q := \sum_{i \in [\![k]\!]} Z_i^2$$

heißt *(zentrale) χ^2-Verteilung* mit k *Freiheitsgraden* (vgl. Beispiel §17.12 (c)), kurz $\boxed{Q \sim \chi^2_k}$. Für $\alpha \in (0,1)$ bezeichnen wir weiterhin den Wert $\chi^2_{k,\alpha} \in \mathbb{R}_{>0}$ als α-Quantil einer (zentralen) χ^2-Verteilung mit k Freiheitsgraden, wenn $\mathrm{N}^k_{(0,1)}(Q \leqslant \chi^2_{k,\alpha}) = \alpha$ gilt. Für $\delta \in \mathbb{R}$ heißt die Verteilung der Zufallsvariable



χ^2_k-Dichtefunktionen

$$W := |Z_1 + \delta|^2 + \sum_{i \in [\![2,k]\!]} Z_i^2$$

nicht-zentrale χ^2-Verteilung mit k *Freiheitsgraden* und *Nichtzentralitätsparameter* δ^2, kurz $\boxed{W \sim \chi^2_k(\delta^2)}$. Weiterhin bezeichnet $\chi^2_{k,\alpha}(\delta^2) \in \mathbb{R}_{>0}$ das α-Quantil einer nicht-zentralen χ^2-Verteilung mit k Freiheitsgraden und Nichtzentralitätsparameter δ^2, wenn $\mathrm{N}^k_{(0,1)}(W \leqslant \chi^2_{k,\alpha}(\delta^2)) = \alpha$ gilt. □

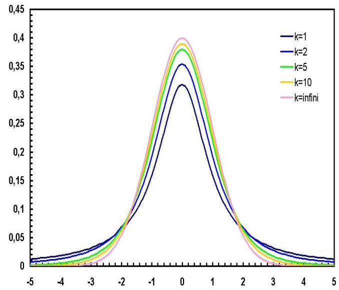
§27.33 **Korollar**. *Es gilt $\mu(\chi^2_k) = k$, $\sigma^2(\chi^2_k) = 2k$ und $\mu(\chi^2_k(\delta^2)) = \delta^2 + k$. Für $Z \sim \mathrm{N}^m_{(0,1)}$, $v \in \mathbb{R}^m$ und $\mathrm{A} \in \mathbb{R}^{m \times p}$ mit $\mathrm{Rang}(\mathrm{A}) = p$ gelten außerdem: (i) $\|\Pi_{\mathrm{Bild}(\mathrm{A})} Z\|^2 \sim \chi^2_p$ und (ii) $\|Z + v\|^2 \sim \chi^2_m(\|v\|^2)$.*

§27.34 **Beweis** von Korollar §27.33. In der Vorlesung. □

§27|06.35 **(Student-) t-Verteilung.**. Die Verteilung der Zufallsvariable

$$T := \frac{Z_0}{\sqrt{\frac{1}{k} \sum_{i \in [\![k]\!]} Z_i^2}}$$

heißt *(Student-) t-Verteilung* mit k *Freiheitsgraden*, kurz $\boxed{T \sim \mathrm{t}_k}$, und $\mathrm{t}_{k,\alpha} \in \mathbb{R}$ bezeichnet das α-Quantil einer Student-t-Verteilung mit k-Freiheitsgraden, wenn $\mathrm{N}^{k+1}_{(0,1)}(T \leqslant \mathrm{t}_{k,\alpha}) = \alpha$ gilt.



t_k-Dichtefunktionen

§27.36 **Bemerkung**. Die Student-t_1-Verteilung mit einem ($k = 1$) Freiheitsgrad entspricht gerade der Cauchy-Verteilung. Für jedes $k \in \mathbb{N}$ besitzt die t_k-Verteilung endliche Momente nur der Ordnung $s \in [\![k)$ (sie ist heavy-tailed). Insbesondere, ist $T \sim \mathrm{t}_k$ so gilt $\mu(\mathrm{t}_k) = 0$ für $k \in \mathbb{N}_{\geqslant 2}$, sowie $\sigma^2(\mathrm{t}_k) = k/(k-2)$ für $k \in \mathbb{N}_{\geqslant 3}$. □

§27.37 **Korollar**. *Für $X \sim \mathrm{N}^n_{(\mu,\sigma^2)}$ sind $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \sim \mathrm{N}_{(0,1)}$ und $\frac{(n-1)}{\sigma^2}\widehat{\mathrm{S}}^{(2)}_n \sim \chi^2_{n-1}$ unabhängig, so dass mit $\widehat{\mathrm{S}}_n = \sqrt{\widehat{\mathrm{S}}^{(2)}_n}$ gilt $\frac{\sqrt{n}}{\widehat{\mathrm{S}}_n}(\overline{X}_n - \mu) \sim \mathrm{t}_{n-1}$.*

§27.38 **Beweis** von Korollar §27.37. In der Vorlesung. □

§27.39 **t-Tests für den Erwartungswert.**. *Für einen $\mathbb{R}^n$-wertigen Zufallsvektor X sei ein normales Lokations-Skalen-Modell adäquat, also $X \odot (\mathrm{N}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}}$. Dann hält zu dem entsprechenden Testproblem in Definition §27|04.24*

(a) *der rechtsseitige t-Test $\varphi^r_{\mathrm{c}_\alpha}$ mit $\mathrm{c}_\alpha = \mathrm{t}_{(n-1),(1-\alpha)}$,*

(b) *der linksseitige t-Test $\varphi^l_{\mathrm{c}_\alpha}$ mit $\mathrm{c}_\alpha = \mathrm{t}_{(n-1),(1-\alpha)}$ sowie*

(c) *der beidseitige t-Test $\varphi^b_{\mathrm{c}_{\alpha/2}}$ mit $\mathrm{c}_{\alpha/2} = \mathrm{t}_{(n-1),(1-\alpha/2)}$*

das Signifikanz-Niveau $\alpha \in (0,1)$ ein. □

§27.40 **Bemerkung**. In der Vorlesung Statistik I zeigen wir, dass die Tests in der Aussage §27.39 auch gleichmäßig beste unverfälschte Test sind. □

§27.41 **χ^2-Tests für die Varianz.**. *Für einen $\mathbb{R}^n$-wertigen Zufallsvektor X sei ein normales Lokations-Skalen-Modell adäquat, also $X \circleddash (\mathrm{N}^n_{(\mu,\sigma^2)})_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_{>0}}$. Dann hält zu dem entsprechenden Testproblem in Definition §27|04.25*

(a) *der rechtsseitige χ^2-Test $\varphi^r_{c^r_\alpha}$ mit $c^r_\alpha = \chi^2_{(n-1),(1-\alpha)}$,*

(b) *der linksseitige χ^2-Test $\varphi^l_{c^l_\alpha}$ mit $c^l_\alpha = \chi^2_{(n-1),\alpha}$ sowie*

(c) *der beidseitige χ^2-Test $\varphi^b_{c^l_{\alpha/2},c^r_{\alpha/2}}$ mit $c^l_{\alpha/2} = \chi^2_{(n-1),\alpha/2}$ und $c^r_{\alpha/2} = \chi^2_{(n-1),(1-\alpha/2)}$*

das Signifikanz-Niveau $\alpha \in (0,1)$ ein. □

§27|07 Normales Zwei-Stichproben-Modell

§27.42 **Korollar**. *Für $(X,Y) \sim \mathrm{N}^n_{(\mu_X,\sigma^2)} \otimes \mathrm{N}^m_{(\mu_Y,\sigma^2)}$ sind*

$$\frac{\sqrt{nm}}{\sigma\sqrt{n+m}}(\overline{X_n} - \overline{Y_m} - (\mu_X - \mu_Y)) \sim \mathrm{N}_{(0,1)} \quad \textit{und} \quad \frac{(n+m-2)}{\sigma^2}\widehat{\mathrm{S}}^{(2)}_{n,m} = \frac{(n-1)}{\sigma^2}\widehat{\mathrm{S}}^{(2)}_n + \frac{(m-1)}{\sigma^2}\widehat{\mathrm{S}}^{(2)}_m \sim \chi^2_{n+m-2}$$

unabhängig, so dass mit $\widehat{\mathrm{S}}_{n,m} := \sqrt{\widehat{\mathrm{S}}^{(2)}_{n,m}}$ gilt

$$\frac{\sqrt{nm}}{\widehat{\mathrm{S}}_{n,m}\sqrt{n+m}}(\overline{X_n} - \overline{Y_m} - (\mu_X - \mu_Y)) \sim \mathrm{t}_{n+m-2}.$$

§27.43 **Beweis** von Korollar §27.42. In der Vorlesung. □

§27.44 **t-Tests auf Gleichheit der Erwartungswerte**. *Für zwei Zufallsvektoren X und Y mit identischer Varianz der Randverteilungen, sei ein normales Zwei-Stichproben-Modell mit unverbunden Stichproben, also $(X,Y) \circleddash (\mathrm{N}^n_{(\mu_X,\sigma^2)} \otimes \mathrm{N}^m_{(\mu_Y,\sigma^2)})_{\mu_X,\mu_Y\in\mathbb{R},\sigma^2\in\mathbb{R}_{>0}}$, adäquat. Dann hält zu dem entsprechenden Testproblem in Definition §27|05.28*

(a) *der rechtsseitige t-Test $\varphi^r_{c_\alpha}$ mit $c_\alpha = \mathrm{t}_{(n+m-2),(1-\alpha)}$,*

(b) *der linksseitige t-Test $\varphi^l_{c_\alpha}$ mit $c_\alpha = \mathrm{t}_{(n+m-2),(1-\alpha)}$ sowie*

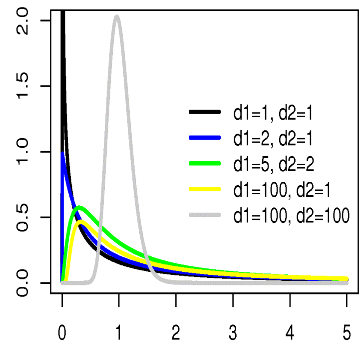
(c) *der beidseitige t-Test $\varphi^b_{c_{\alpha/2}}$ mit $c_{\alpha/2} = \mathrm{t}_{(n+m-2),(1-\alpha/2)}$*

das Signifikanz-Niveau $\alpha \in (0,1)$ ein. □

§27.45 **Beispiel**. Betrachten wir die zufälligen Messwerte der beiden Segmente des Fürsten G. Őmetry in Beispiel VI im Kapitel 1 Prolog. Dann erhalten wir Schätzwerte $\overline{x_{31}} \approx 9.51$, $\overline{y_{31}} \approx 10.49$, $\widehat{\mathrm{S}}^{(2)}_{31}(x) \approx 0.42$, $\widehat{\mathrm{S}}^{(2)}_{31}(y) \approx 1.05$, sodass $\widehat{\mathrm{s}}^{(2)}_{31,31} := \widehat{\mathrm{S}}^{(2)}_{31,31}(x,y) \approx 0.74$, $\widehat{\mathrm{s}}_{31,31} := \sqrt{\widehat{\mathrm{s}}^{(2)}_{31,31}} \approx 0.86$ und damit $\widehat{\mathrm{t}}_{31,31} := \frac{\sqrt{31}}{\widehat{\mathrm{s}}_{31,31}\sqrt{2}}|\overline{x_{31}} - \overline{y_{31}}| \approx 4.5$. Unter der Annahme, dass ein normales Zwei-Stichproben-Modell mit unverbunden Stichproben und gleichen Varianzen vorliegt, hält der beidseitige t-Test $\varphi^b_{c_{\alpha/2}}$ mit $c_{\alpha/2} = \mathrm{t}_{60,(1-\alpha/2)}$ das Niveau α ein (Aussage §27.44). Für $\alpha = 0.05$ gilt $\mathrm{t}_{60,(1-\alpha/2)} \approx 2 < 4.5 \approx \widehat{\mathrm{t}}_{31,31}$, so dass wir die Nullhypothese der Gleichheit zum Niveau $\alpha = 0.05$ ablehnen. □

§27|07.46 **(Fisher-) F-Verteilung.** Die Verteilung der Zufallsvariable

$$F := \frac{\frac{1}{m}\sum_{i\in[\![m]\!]} Z_i^2}{\frac{1}{k}\sum_{i\in[\![k]\!]} Z_{m+i}^2}$$



$\mathrm{F}_{d1,d2}$-Dichtefunktionen

heißt *zentrale (Fisher-) F-Verteilung* mit m und k *Freiheitsgraden*, kurz $\boxed{F \sim \mathrm{F}_{m,k}}$. Wir bezeichnen mit $\mathrm{F}_{m,k,\alpha} \in \mathbb{R}_{\geqslant 0}$ das α-Quantil einer zentralen Fisher-$\mathrm{F}_{m,k}$-Verteilung mit m und k Freiheitsgraden, d.h. $\mathrm{N}_{(0,1)}^{m+k}(F \leqslant \mathrm{F}_{m,k,\alpha}) = \alpha$.
Für $\delta \in \mathbb{R}$ heißt die Verteilung der Zufallsvariable

$$W := \frac{\frac{1}{m}\{(Z_1+\delta)^2 + \sum_{i\in[\![2,m]\!]} Z_i^2\}}{\frac{1}{k}\sum_{i\in[\![k]\!]} Z_{m+i}^2}$$

nicht-zentrale (Fisher-) F-Verteilung mit m und k *Freiheitsgraden* und *Nichtzentralitätsparameter* δ^2, kurz $\boxed{W \sim \mathrm{F}_{m,k}(\delta^2)}$. Weiterhin bezeichnet $\mathrm{F}_{m,k,\alpha}(\delta^2) \in \mathbb{R}_{\geqslant 0}$ das α-Quantil einer nicht-zentralen Fisher-$\mathrm{F}_{m,k}(\delta^2)$-Verteilung mit m und k Freiheitsgraden und Nichtzentralitätsparameter δ^2, wenn $\mathrm{N}_{(0,1)}^{m+k}(W \leqslant \mathrm{F}_{m,k,\alpha}(\delta^2)) = \alpha$ gilt. □

§27.47 **Bemerkung.** Sei $F \sim \mathrm{F}_{m,k}$ mit $k \in \mathbb{Z}_{\geqslant 2}$, dann ist F^{-1} eine $\mathrm{F}_{k,m}$-verteilte Zufallsvariable. Für $T \sim \mathrm{t}_k$ ist T^2 eine $\mathrm{F}_{1,k}$-verteilte Zufallsvariable. □

§27.48 **F-Tests auf Gleichheit der Varianzen.** *Für X und Y sei ein normales Zwei-Stichproben-Modell mit unverbunden Stichproben, also $(X,Y) \circleddash (\mathrm{N}^n_{(\mu_X,\sigma_X^2)} \otimes \mathrm{N}^m_{(\mu_Y,\sigma_Y^2)})_{\mu_X,\mu_Y\in\mathbb{R},\sigma_X^2,\sigma_Y^2\in\mathbb{R}_{>0}}$, adäquat. Dann hält zu dem entsprechenden Testproblem in Definition §27|05.30*

(a) *der rechtsseitige F-Test $\varphi^r_{c^r_\alpha}$ mit $c^r_\alpha = \mathrm{F}_{(n-1),(m-1),(1-\alpha)}$,*

(b) *der linksseitige F-Test $\varphi^l_{c^l_\alpha}$ mit $c^l_\alpha = \mathrm{F}_{(n-1),(m-1),\alpha}$ sowie*

(c) *der beidseitige F-Test $\varphi^b_{c^l_{\alpha/2},c^r_{\alpha/2}}$ mit $c^l_{\alpha/2} = \mathrm{F}_{(n-1),(m-1),\alpha/2}$ und $c^r_{\alpha/2} = \mathrm{F}_{(n-1),(m-1),(1-\alpha/2)}$*

das Signifikanz-Niveau $\alpha \in (0,1)$ ein. □

§27.49 **Beispiel.** Betrachten wir die zufälligen Messwerte der abgefüllten Volumen in den 33cl bzw. 75cl Bierflaschen der Abbaye de Rochefort in Beispiel VI im Kapitel 1 Prolog. Dann erhalten wir die Schätzwerte $\overline{x}_{42} \approx 32.94$, $\overline{y}_{42} \approx 74.98$, $\widehat{\mathrm{S}}^{(2)}_{42}(x) \approx 0.74$, $\widehat{\mathrm{S}}^{(2)}_{42}(y) \approx 0.81$, und damit $\widehat{\mathrm{F}}_{42,42} := \widehat{\mathrm{S}}^{(2)}_{42}(x)/\widehat{\mathrm{S}}^{(2)}_{42}(y) \approx 0.91$. Unter der Annahme, dass ein normales Zwei-Stichproben-Modell mit unverbunden Stichproben vorliegt, hält der beidseitige F-Test $\varphi^b_{c^l_{\alpha/2},c^r_{\alpha/2}}$ mit $c^l_\alpha = \mathrm{F}_{41,41,\alpha/2}$ und $c^l_\alpha = \mathrm{F}_{41,41,(1-\alpha/2)}$ das Niveau α ein (Aussage §27.48). Für $\alpha = 0.05$ gilt $\mathrm{F}_{41,41,.025} \approx 0.54$, $\mathrm{F}_{41,41,.975} \approx 1.86$ und $\widehat{\mathrm{F}}_{42,42} \approx 0.91 \in (0.54, 1.86)$, so dass wir die Nullhypothese der Gleichheit der Varianzen zum Niveau $\alpha = 0.05$ nicht ablehnen können. □

Kapitel 6

Grenzwertsätze

§28 Konvergente Folgen von Zufallsvariablen

Im Folgenden sei stets $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{L}_s := \mathcal{L}_p(\Omega, \mathscr{A}, \mathbb{P})$ mit s $\in \overline{\mathbb{R}}_{>0}$ und wir schreiben auch kurz $\overline{\mathcal{M}} := \overline{\mathcal{M}}(\mathscr{A})$, $\overline{\mathcal{M}}_{\geqslant 0} := \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, $\mathcal{M} := \mathcal{M}(\mathscr{A})$ und $\mathcal{M}_{\geqslant 0} := \mathcal{M}_{\geqslant 0}(\mathscr{A})$.

§28.01 **Erinnerung**. Für eine Folge reeller Zahlen $(x_n)_{n\in\mathbb{N}}$ existiert der Grenzwert

$$x = \lim_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n \in \mathbb{R}$$

genau dann, wenn eine der folgenden äquivalenten Bedingungen erfüllt ist:

(a) $\exists\, x \in \mathbb{R} : \limsup_{n\to\infty} |x - x_n| = 0$;

(b) $\forall\, \varepsilon \in \mathbb{R}_{>0} : \exists\, n_\varepsilon \in \mathbb{N} : \forall\, k, m \in \mathbb{N}_{\geqslant n_\varepsilon} : |x_k - x_m| \leqslant \varepsilon$;

(c) $\forall\, \varepsilon \in \mathbb{R}_{>0} : \exists\, n_\varepsilon \in \mathbb{N} : \sup_{k,m\in\mathbb{N}_{\geqslant n_\varepsilon}} |x_k - x_m| = \sup\{|x_k - x_m| : k, m \in \mathbb{N}_{\geqslant n_\varepsilon}\} \leqslant \varepsilon$;

(d) $\lim_{n\to\infty} \sup_{k,m\in\mathbb{N}_{\geqslant n}} |x_k - x_m| = 0$, also $\sup_{k,m\in\mathbb{N}_{\geqslant n}} |x_k - x_m| \downarrow 0$;

(e) $\inf_{n\in\mathbb{N}} \sup_{k,m\in\mathbb{N}_{\geqslant n}} |x_k - x_m| = 0$;

(f) $\inf_{n\in\mathbb{N}} \sup_{m\in\mathbb{N}_{\geqslant n}} |x_n - x_m| = 0$.

Eine Eigenschaft oder Beziehung wird *$\mathbb{P}$-fast sicher* ($\mathbb{P}$-f.s.) oder *$\mathbb{P}$-fast überall* ($\mathbb{P}$-f.ü.) genannt, wenn sie bis auf einer $\mathbb{P}$-Nullmenge überall gilt. Eine reelle Zufallsvariable $X \in \mathcal{M}$ auf $(\Omega, \mathscr{A}, \mathbb{P})$ besitzt den Wertebereich $\mathbb{R}$, und kann insbesondere auch als numerische Zufallsvariable, kurz $X \in \overline{\mathcal{M}}$, mit Wertebereich $\overline{\mathbb{R}}$ aufgefasst werden. □

§28.02 **Definition**. Eine Folge numerischer Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ in $\overline{\mathcal{M}}$ *konvergiert $\mathbb{P}$-fast sicher*, wenn gilt

$$X_\star = \liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n = X^\star \quad \mathbb{P}\text{-f.s.}, \qquad \text{dass heißt,} \qquad \mathbb{P}(X_\star = X^\star) = 1.$$

Wir sagen, eine Folge $(X_n)_{n\in\mathbb{N}}$ aus $\overline{\mathcal{M}}$ konvergiert $\mathbb{P}$-f.s. gegen $X \in \overline{\mathcal{M}}$, kurz $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$, wenn $X = X_\star = X^\star$ $\mathbb{P}$-f.s. gilt. □

§28.03 **Bemerkung**. Konvergiert $\mathbb{P}$-f.s. eine Folge numerischer oder reeller Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ und setzen wir $X := X_\star$, dann gilt offensichtlich $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$. Dabei legt fast sichere Konvergenz den Grenzwert bis auf Gleichheit fast überall eindeutig fest. Achtung, ist $(X_n)_{n\in\mathbb{N}}$ insbesondere eine Folge reeller Zufallsvariablen, kurz *aus* $\mathcal{M}(\mathscr{A})$, dann garantiert im Gegensatz zur nächsten Definition §28.04 die letzte Definition §28.02 nicht, dass $X_\star \in \mathbb{R}$ $\mathbb{P}$-f.s. gilt. □

§28.04 **Definition**. Eine Folge $(X_n)_{n\in\mathbb{N}}$ reeller Zufallsvariablen in $\mathcal{M}$ *konvergiert $\mathbb{P}$-fast sicher* gegen eine numerische Zufallsvariable $X \in \overline{\mathcal{M}}$, kurz $\boxed{X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X}$, wenn

$$\limsup_{n\to\infty} |X_n - X| = 0 \quad \mathbb{P}\text{-f.s.}, \qquad \text{das heißt} \qquad \mathbb{P}(\limsup_{n\to\infty} |X_n - X| = 0) = 1.$$

□

§28.05 **Cauchy-Kriterium**. *Eine Folge $(X_n)_{n\in\mathbb{N}}$ reeller Zufallsvariablen in $\mathcal{M}$ konvergiert $\mathbb{P}$-fast sicher genau dann, wenn*

$$\lim_{n\to\infty} \sup_{k,m\in\mathbb{N}_{\geqslant n}} |X_k - X_m| = \inf_{n\in\mathbb{N}} \sup_{k,m\in\mathbb{N}_{\geqslant n}} |X_k - X_m| = 0 \quad \mathbb{P}\text{-f.s.}, \tag{28.01}$$

da $(\sup_{k,m\in\mathbb{N}_{\geqslant n}} |X_k - X_m|)_{n\in\mathbb{N}}$ eine monoton fallende Folge von Zufallsvariablen ist. Weiterhin gilt (28.01) genau dann, wenn $\inf_{n\in\mathbb{N}} \sup_{m\in\mathbb{N}_{\geqslant n}} |X_n - X_m| = 0$ $\mathbb{P}$-f.s.. □

§28.06 **Lemma**. *Eine Folge $(X_n)_{n\in\mathbb{N}}$ reeller Zufallsvariablen in $\mathcal{M}$ konvergiert $\mathbb{P}$-fast sicher gegen $X \in \overline{\mathcal{M}}$, kurz $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$, genau dann, wenn $\lim_{n\to\infty} \mathbb{P}(\sup_{m\in\mathbb{N}_{\geqslant n}} |X_m - X| > \varepsilon) = 0$ für alle $\varepsilon \in \mathbb{R}_{>0}$ gilt.*

§28.07 **Beweis** von Lemma §28.06. In der Vorlesung. □

§28.08 **Lemma**. *Eine Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{M}$ konvergiert $\mathbb{P}$-fast sicher genau dann, wenn für alle $\varepsilon \in \mathbb{R}_{>0}$ gilt $\lim_{n\to\infty} \mathbb{P}(\sup_{m\in\mathbb{N}_{\geqslant n}} |X_m - X_n| > \varepsilon) = 0$.*

§28.09 **Beweis** von Lemma §28.06. In der Vorlesung. □

§28.10 **Definition**. Eine Folge $(X_n)_{n\in\mathbb{N}}$ reeller Zufallsvariablen in $\mathcal{M}$ *konvergiert $\mathbb{P}$-fast vollständig* ($\mathbb{P}$-f.v.) gegen eine numerische Zufallsvariable $X \in \overline{\mathcal{M}}$, kurz $\boxed{X_n \xrightarrow{\mathbb{P}\text{-f.v.}} X}$, wenn für alle $\varepsilon \in \mathbb{R}_{>0}$ gilt $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n - X| > \varepsilon) \in \mathbb{R}_{\geqslant 0}$. □

§28.11 **Korollar**. *Konvergiert eine Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{M}$ $\mathbb{P}$-fast vollständig gegen $X \in \overline{\mathcal{M}}$, also $X_n \xrightarrow{\mathbb{P}\text{-f.v.}} X$, dann konvergiert sie auch $\mathbb{P}$-fast sicher gegen X, also $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$.*

§28.12 **Beweis** von Korollar §28.11. In der Vorlesung. □

§28.13 **Lemma**. *Seien $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{M}$ und $(\varepsilon_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}_{\geqslant 0}$ derart, dass die folgenden beiden Bedingungen erfüllt sind:*

(B1) $\sum_{n\in\mathbb{N}} \varepsilon_n \in \mathbb{R}_{\geqslant 0}$;

(B2) $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_{n+1} - X_n| > \varepsilon_n) \in \mathbb{R}_{\geqslant 0}$.

Dann konvergiert die Folge $(X_n)_{n\in\mathbb{N}}$ $\mathbb{P}$-fast sicher und es gilt $\sum_{n\in\mathbb{N}} |X_{n+1} - X_n| \in \mathbb{R}_{\geqslant 0}$ $\mathbb{P}$-f.s..

§28.14 **Beweis** von Lemma §28.13. In der Vorlesung. □

§28.15 **Definition**. Eine Folge $(X_n)_{n\in\mathbb{N}}$ reeller Zufallsvariablen in $\mathcal{M}$ *konvergiert stochastisch* (oder auch *in $\mathbb{P}$-Wahrscheinlichkeit*), wenn für alle $\varepsilon \in \mathbb{R}_{>0}$ gilt

$$\lim_{n,m\to\infty} \mathbb{P}(|X_n - X_m| > \varepsilon) = 0.$$

Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert stochastisch gegen* eine numerische Zufallsvariable $X \in \overline{\mathcal{M}}$, kurz $\boxed{X_n \xrightarrow{\mathbb{P}} X}$, wenn für alle $\varepsilon \in \mathbb{R}_{>0}$ gilt $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. □

§28.16 **Bemerkung**. Definitionsgemäß konvergiert eine Folge $(X_n)_{n\in\mathbb{N}}$ stochastisch gegen X genau dann, wenn die Folge $(X - X_n)_{n\in\mathbb{N}}$ stochastisch gegen 0 konvergiert. Dabei legt stochastische Konvergenz den Grenzwert eindeutig fest bis auf Gleichheit fast überall. In der Tat: Sei $X_n \xrightarrow{\mathbb{P}} X$ und $X_n \xrightarrow{\mathbb{P}} Y$, dann gilt (wegen $|X - Y| \leqslant |X - X_n| + |Y - X_n|$) für jedes $\varepsilon \in \mathbb{R}_{>0}$

$$\mathbb{P}(|X - Y| > \varepsilon) \leqslant \mathbb{P}(|X - X_n| > \varepsilon/2) + \mathbb{P}(|Y - X_n| > \varepsilon/2) \xrightarrow{n\to\infty} 0.$$

Also ist $\mathbb{P}(|X - Y| > \varepsilon) = 0$ für jedes $\varepsilon \in \mathbb{R}_{>0}$ und somit $\mathbb{P}(|X - Y| = 0) = 1$. □

§28|00.17 **Vorbemerkung**. Nach Lemma §28.06 sind $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$ und $\sup_{m\in\mathbb{N}_{\geqslant n}}|X_m - X| \xrightarrow{\mathbb{P}} 0$ äquivalente Aussagen. Genauso konvergiert nach Lemma §28.08 $(X_n)_{n\in\mathbb{N}}$ $\mathbb{P}$-f.s. genau dann, wenn $\sup_{m\in\mathbb{N}_{\geqslant n}}|X_m - X_n| \xrightarrow{\mathbb{P}} 0$. Damit konvergiert $(X_n)_{n\in\mathbb{N}}$ stochastisch (gegen X), sodass aus $\mathbb{P}$-fast sicherer Konvergenz auch stochastische Konvergenz folgt. Die Umkehrung gilt nicht, aber der nächste Satz gibt eine teilweise Umkehrung. □

§28.18 **Satz**. *Fast sichere Konvergenz impliziert stochastische Konvergenz, aber nicht umgekehrt. Jede stochastisch konvergente Folge $(X_n)_{n\in\mathbb{N}}$ enthält eine fast sicher konvergente Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$. Für $X_n \xrightarrow{\mathbb{P}} X$ gilt dann $X_{n_k} \xrightarrow{\mathbb{P}\text{-f.s.}} X$.*

§28.19 **Beweis** von Satz §28.18. In der Vorlesung. □

§28.20 **Satz von der stetigen Abbildung**. *Sei $h : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion und sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{M}$, die $\mathbb{P}$-f.s. (bzw. stochastisch) gegen eine reelle Zufallsvariable $X \in \mathcal{M}$ konvergiert. Dann konvergiert $(h(X_n))_{n\in\mathbb{N}}$ auch gegen $h(X)$ $\mathbb{P}$-f.s. (bzw. stochastisch).*

§28.21 **Beweis** von Satz §28.20. In der Vorlesung. □

§28.22 **Definition**. Eine Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}_s$ für ein $s \in \overline{\mathbb{R}}_{>0}$ *konvergiert in $\mathcal{L}_s$*, wenn gilt

$$\lim_{n,m\to\infty} \|X_n - X_m\|_{\mathcal{L}_s} = 0.$$

Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert in $\mathcal{L}_s$ gegen* $X \in \mathcal{L}_s$, kurz $\boxed{X_n \xrightarrow{\mathcal{L}_s} X}$, wenn gilt:

$$\lim_{n\to\infty} \|X_n - X\|_{\mathcal{L}_s} = 0.$$

□

§28.23 **Bemerkung**. Definitionsgemäß konvergiert eine Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}_s$ gegen X genau dann, wenn die Folge $(X - X_n)_{n\in\mathbb{N}}$ gegen 0 in $\mathcal{L}_s$ konvergiert. Dabei legt $\mathcal{L}_s$-Konvergenz den Grenzwert eindeutig fest bis auf Gleichheit fast überall (Satz §28.29). Weiterhin gilt $X_n \xrightarrow{\mathcal{L}_s} X$, so gilt für $s \in \overline{\mathbb{R}}_{\geqslant 1}$ insbesondere $\|X_n\|_{\mathcal{L}_s} \xrightarrow{n\to\infty} \|X\|_{\mathcal{L}_s}$. □

§28.24 **Korollar**. *Für ein $s \in \overline{\mathbb{R}}_{>0}$ sei $(X_n)_{n\in\mathbb{N}}$ eine in $\mathcal{L}_s$ konvergente Folge. Dann ist auch für jedes $r \in (0, s]$ die Folge $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}_r$ konvergent.*

§28.25 **Beweis** von Korollar §28.24. Die Behauptung folgt direkt aus Korollar §24.07. □

§28.26 **Erinnerung**. Seien $X, \widetilde{X} \in \mathcal{L}_1(\Omega, \mathscr{A}, \mathbb{P})$ und $Y \in \overline{\mathcal{M}}(\mathscr{A})$. Dann gelten

(§22.07) $\mathbb{P}(X = 0) = 1 \quad \Leftrightarrow \quad \mathbb{E}(X\mathbb{1}_A) = 0$ für alle $A \in \mathscr{A}$.

(§22.04 (vii)) $X = Y$ $\mathbb{P}$-f.s. $\quad \Rightarrow \quad Y \in \mathcal{L}_1$ und $\mathbb{E}(X) = \mathbb{E}(Y)$.

(§21.13 (iii)) $|Y| \leqslant |X|$ $\mathbb{P}$-f.s. $\quad \Rightarrow \quad \mathbb{E}(|Y|) \leqslant \mathbb{E}(|X|)$ und $Y \in \mathcal{L}_1$.

$$\text{(§22.04 (iv))}\quad \widetilde{X} \leqslant X \ \mathbb{P}\text{-f.s.} \quad \Rightarrow \quad \mathbb{E}(\widetilde{X}) = \mathbb{E}(\widetilde{X} \wedge X) \leqslant \mathbb{E}(\widetilde{X} \vee X) = \mathbb{E}(X). \qquad \square$$

§28.27 **Lemma.**

(i) *(Monotone Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1$, derart dass:*

(mK1') $X_n \uparrow X$ $\mathbb{P}$-f.s. *(bzw.* $X_n \downarrow X$ $\mathbb{P}$-f.s.*) für ein* $X \in \overline{\mathcal{M}}$*;*

(mK2) $(|\mathbb{E}(X_n)|)_{n\in\mathbb{N}}$ *ist beschränkt, das heißt,* $\sup_{n\in\mathbb{N}}|\mathbb{E}(X_n)| \in \mathbb{R}_{\geqslant 0}$.

Dann gilt $X \in \mathcal{L}_1$ *und* $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ *(bzw.* $\mathbb{E}(X_n) \downarrow \mathbb{E}(X)$*).*

(ii) *(Dominierte Konvergenz) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1$ derart, dass gelten:*

(dK1') $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$ *für ein* $X \in \overline{\mathcal{M}}$*;*

(dK2') *Es existiert* $Y \in \mathcal{L}_1$ *mit* $\sup_{n\in\mathbb{N}}|X_n| \leqslant Y$ $\mathbb{P}$-f.s.

Dann gilt $X \in \mathcal{L}_1$, $\lim_{n\to\infty} \mathbb{E}(|X - X_n|) = 0$ *und* $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

(iii) *(Satz von Scheffé) Seien $X, X_n \in \overline{\mathcal{M}}_{\geqslant 0}$ $\mathbb{P}$-integrierbar für alle $n \in \mathbb{N}$, derart dass die Bedingungen $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$ and $\mathbb{E}(X_n) \xrightarrow{n\to\infty} \mathbb{E}(X)$ gelten. Dann gilt auch $X_n \xrightarrow{\mathcal{L}_1} X$.*

(iv) *(Satz von Riesz) Sei $s \in \mathbb{R}_{\geqslant 1}$, $X, X_n \in \mathcal{L}_s(\mathbb{P})$ für alle $n \in \mathbb{N}$ und $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$. Es gilt genau dann $\mathbb{E}(|X_n|^s) \xrightarrow{n\to\infty} \mathbb{E}(|X|^s)$, wenn $X_n \xrightarrow{\mathcal{L}_s} X$ gilt.*

§28.28 **Beweis** von Lemma §28.27. In der Vorlesung. □

§28.29 **Satz.**

(i) *Für jedes $s \in \mathbb{R}_{>0}$ ist eine $\mathcal{L}_s$-konvergente Folge $(X_n)_{n\in\mathbb{N}}$ auch stochastisch konvergent mit $\mathbb{P}$-f.s. konvergenter Teilfolge, wobei der gemeinsame Grenzwert in $\mathcal{L}_s$ ist. Für $X_n \xrightarrow{\mathcal{L}_s} X$ gilt also $X_n \xrightarrow{\mathbb{P}} X$.*

(ii) *Eine $\mathcal{L}_\infty$-konvergente Folge $(X_n)_{n\in\mathbb{N}}$ konvergiert auch $\mathbb{P}$-f.v., wobei der gemeinsame Grenzwert in $\mathcal{L}_\infty$ ist. Für $X_n \xrightarrow{\mathcal{L}_\infty} X$ gilt also $X_n \xrightarrow{\mathbb{P}\text{-f.v.}} X$.*

§28.30 **Beweis** von Satz §28.29. In der Vorlesung. □

§28.31 **Beispiel.**

(a) Betrachten wir wie in Beispiel §16.17 (b) eine Folge $(X_n)_{n\in\mathbb{N}}$ unabhängiger und identisch-verteilter (u.i.v.) reeller Zufallsvariablen mit $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. Setzen wir $S_n := \sum_{k\in[\![n]\!]} X_k \frac{1}{k}$, so gilt $\mathbb{E}(S_n) = 0$ und $\mathbb{V}\mathrm{ar}(S_n) = \sum_{k\in[\![n]\!]} \frac{1}{k^2}$. Da für alle $n \in \mathbb{N}$ and $m \in \mathbb{N}_{\geqslant n}$ gilt $\mathbb{V}\mathrm{ar}(S_m - S_n) \leqslant \sum_{k\in\mathbb{N}_{\geqslant n}} \frac{1}{k^2}$, und $\sum_{k\in\mathbb{N}_{\geqslant n}} \frac{1}{k^2} \downarrow 0$, bildet $(S_n)_{n\in\mathbb{N}}$ eine konvergente Folge in $\mathcal{L}_2$. Nach Satz §28.29 (i) existiert also $S_\infty \in \mathcal{L}_2$ mit $S_n \xrightarrow{\mathcal{L}_2} S_\infty$ und damit auch $S_n \xrightarrow{\mathbb{P}} S_\infty$.

(b) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_2$ paarweise unkorrelierter Zufallsvariablen mit beschränkter Varianz also $\mathbb{K}\mathrm{ov}(X_n, X_m) = 0$ für alle $n \in \mathbb{N}$, $m \in \mathbb{N}_{\neq n}$ sowie $\sup_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n) \in \mathbb{R}_{\geqslant 0}$. Setzen wir $X_n^\circ := X_n - \mathbb{E}(X_n)$ und $\overline{X_n^\circ} := n^{-1}\sum_{i\in[\![n]\!]} X_i^\circ$, $n \in \mathbb{N}$. Dann gilt $\mathbb{E}(\overline{X_n^\circ}) = 0$ und $\|\overline{X_n^\circ}\|_{\mathcal{L}_2}^2 \leqslant n^{-1}\sup_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n)$, sodass $\|\overline{X_n^\circ}\|_{\mathcal{L}_2}^2 \to 0$. Betrachte Teilfoge $(\overline{X_{m^2}^\circ})_{m\in\mathbb{N}}$, dann gilt für jedes

$$\forall \varepsilon \in \mathbb{R}_{>0} : \sum_{m\in\mathbb{N}} \mathbb{P}(|\overline{X_{m^2}^\circ}| > \varepsilon) \leqslant \varepsilon^{-2} \sup_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n) \sum_{m\in\mathbb{N}} m^{-2} \in \mathbb{R}_{\geqslant 0},$$

so dass $\overline{X_{m^2}^\circ} \xrightarrow{\mathbb{P}\text{-f.v.}} 0$ und damit $\overline{X_{m^2}^\circ} \xrightarrow{\mathbb{P}\text{-f.s.}} 0$ für $m \to \infty$ gilt. □

§28|00.32 **Vorbemerkung**. Unter Verwendung der *Markov-Ungleichung* §25.07 folgt aus der $\mathcal{L}_s$-Konvergenz die stochastische Konvergenz, wobei die Umkehrung nicht gilt. Unter der zusätzlichen Annahme der *gleichgradigen Integrierbarkeit* gilt dann auch die Umkehrung. □

§28.33 **Definition**. Eine Familie $(X_i)_{i\in\mathcal{I}}$ numerischer Zufallsvariablen in $\overline{\mathcal{M}}$ mit beliebiger nicht-leerer Indexmenge $\mathcal{I}$ heißt *gleichgradig $\mathbb{P}$-integrierbar*, wenn gilt

$$\lim_{n\to\infty}\sup_{i\in\mathcal{I}}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>n\}})=0.$$

□

§28.34 **Bemerkung**. Eine $\mathbb{P}$-integrierbare Zufallsvariable $X\in\mathcal{L}_1(\mathbb{P})$, das bedeutet also $\mathbb{E}(|X|)\in\mathbb{R}_{\geqslant 0}$ und somit $\mathbb{P}(|X|=\infty)=0$ (Lemma §21.13 (ii)), ist als Familie (X) gleichgradig integrierbar, da mit Hilfe des Satzes von der dominierten Konvergenz

$$\lim_{n\to\infty}\mathbb{E}(|X|\mathbb{1}_{\{|X|>n\}})=\mathbb{E}(|X|\mathbb{1}_{\{|X|=\infty\}})=0.$$

gilt. Sei $X\in\overline{\mathcal{M}}$ eine numerische Zufallsvariable. Dann sind äquivalent:

(B1) X ist $\mathbb{P}$-integrierbar, also $\mathbb{E}(|X|)\in\mathbb{R}_{\geqslant 0}$.

(B2) Für jedes $\varepsilon\in\mathbb{R}_{>0}$ existiert ein $Y\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}$ sodass $\mathbb{E}(|X|\mathbb{1}_{\{|X|\geqslant Y\}})\leqslant\varepsilon$.

(B3) $\inf\{\mathbb{E}(|X|\mathbb{1}_{\{|X|\geqslant Y\}}):\ Y\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}\}=0$.

Sei (B1) $X\in\mathcal{L}_1(\mathbb{P})$ erfüllt. Setze $Y:=2|X|\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}$, dann gilt

$$\{|X|\geqslant Y\}=\{|X|=0\}\cup\{|X|=\infty\},$$

wobei $\{|X|=\infty\}$ eine $\mathbb{P}$-Nullmenge ist und

$$\mathbb{E}(|X|\mathbb{1}_{\{|X|\geqslant Y\}})=\mathbb{E}(|X|\mathbb{1}_{\{|X|=\infty\}})=0$$

mit Lemma §21.13 (i). Somit gilt auch (B2). Sei (B2) erfüllt, also $Y\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}$ mit $\mathbb{E}(|X|\mathbb{1}_{\{|X|\geqslant Y\}})\leqslant\varepsilon\in\mathbb{R}_{>0}$ gegeben. Dann erhalten wir (B1) mit Hilfe von

$$\mathbb{E}(|X|)=\mathbb{E}(|X|\mathbb{1}_{\{|X|\geqslant Y\}})+\mathbb{E}(|X|\mathbb{1}_{\{|X|<Y\}})\leqslant\varepsilon+\mathbb{E}(Y)\in\mathbb{R}_{>0}.$$

Offensichtlich sind (B2) und (B3) äquivalent. □

§28.35 **Lemma**. *Sei $(X_i)_{i\in\mathcal{I}}$ eine Familie numerischer Zufallsvariablen in $\overline{\mathcal{M}}$ mit beliebiger nicht-leerer Indexmenge $\mathcal{I}$. Dann sind die folgenden Aussagen äquivalent:*

(A1) *$(X_i)_{i\in\mathcal{I}}$ ist gleichgradig $\mathbb{P}$-integrierbar.*

(A2) $\inf\{\sup_{i\in\mathcal{I}}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|\geqslant a\}}):\ a\in\mathbb{R}_{\geqslant 0}\}=0$.

(A3) $\inf\{\sup_{i\in\mathcal{I}}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|\geqslant Y\}}):\ Y\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}\}=0$.

§28.36 **Beweis** von Lemma §28.35. In der Vorlesung. □

§28.37 **Bemerkung**.

(a) Sei $(X_i)_{i\in\mathcal{I}}$ eine Familie in $\overline{\mathcal{M}}$ und $\varepsilon\in\mathbb{R}_{>0}$. Eine Zufallsvariable $Y\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}$ heißt *ε-Majorante* für $(X_i)_{i\in\mathcal{I}}$, wenn $\sup_{i\in\mathcal{I}}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|\geqslant Y\}})\leqslant\varepsilon$ gilt. Offensichtlich ist jede Zufallsvariable $\widetilde{Y}\in\mathcal{L}_1(\mathbb{P})\cap\overline{\mathcal{M}}_{\geqslant 0}$ mit $\widetilde{Y}\geqslant Y$ $\mathbb{P}$-f.s. ebenfalls eine ε-Majorante für $(X_i)_{i\in\mathcal{I}}$. Ist $(X_i)_{i\in\mathcal{I}}$ gleichgradig $\mathbb{P}$-integrierbar, so existiert definitionsgmäß für jedes $\varepsilon\in\mathbb{R}_{>0}$ eine ε-Majorante für $(X_i)_{i\in\mathcal{I}}$.

(b) Eine Menge $\{X_i \in \overline{\mathcal{M}} : i \in \mathcal{I}\}$ von numerischen Zufallsvariablen heißt *gleichgradig $\mathbb{P}$-integrierbar*, wenn die Familie $(X_i)_{i\in\mathcal{I}}$ in $\overline{\mathcal{M}}$ gleichgradig $\mathbb{P}$-integrierbar ist.

(c) Seien $\mathcal{X}_i \subseteq \overline{\mathcal{M}}$, $i \in [\![n]\!]$, endlich viele gleichgradig $\mathbb{P}$-integrierbare Mengen numerischer Zufallsvariablen. Dann ist ihre Vereinigung $\mathcal{X} := \cup_{i\in[\![n]\!]}\mathcal{X}_i$ ebenfalls gleichgradig $\mathbb{P}$-integrierbar. In der Tat, für jedes $\varepsilon \in \mathbb{R}_{>0}$ und ε-Majorante Y_i für $\mathcal{X}_i$, $i \in [\![n]\!]$, ist die Zufallsvariable $\max\{Y_i : i \in [\![n]\!]\}$ eine ε-Majorante für $\mathcal{X}$.

(d) Sei $(X_i)_{i\in\mathcal{I}}$ eine Familie in $\overline{\mathcal{M}}$ und für $Y \in \mathcal{L}_s(\mathbb{P}) \cap \overline{\mathcal{M}}_{\geqslant 0}$ mit $s \in \mathbb{R}_{>0}$ gelte $|X_i| \leqslant Y$ $\mathbb{P}$-f.s. für alle $i \in \mathcal{I}$. Dann ist die Familie $(|X_i|^s)_{i\in\mathcal{I}}$ gleichgradig $\mathbb{P}$-integrierbar. Betrachte die gleichgradig $\mathbb{P}$-integrierbare Familie (Y^s) und für $\varepsilon \in \mathbb{R}_{>0}$ sei $\widetilde{Y} \in \mathcal{L}_1(\mathbb{P}) \cap \overline{\mathcal{M}}_{\geqslant 0}$ eine beliebige ε-Majorante für (Y^s). Dann ist $\widetilde{Y}$ ebenfalls eine ε-Majorante für $(|X_i|^s)_{i\in\mathcal{I}}$, da $\mathbb{E}(|X_i|^s\mathbb{1}_{\{|X_i|^s>\widetilde{Y}\}}) \leqslant \mathbb{E}(Y^s\mathbb{1}_{\{Y^s>\widetilde{Y}\}}) \leqslant \varepsilon$ for all $i \in \mathcal{I}$ gilt. □

§28.38 **Satz**. *Eine Familie $(X_i)_{i\in\mathcal{I}}$ numerischer Zufallsvariablen in $\overline{\mathcal{M}}(\mathscr{A})$ mit beliebiger nicht-leerer Indexmenge $\mathcal{I}$ ist gleichgradig $\mathbb{P}$-integrierbar genau dann, wenn die folgenden zwei Bedingungen gelten:*

(gI1) *$(X_i)_{i\in\mathcal{I}}$ ist beschränkt in $\mathcal{L}_1$, dass heißt, $\sup_{i\in\mathcal{I}} \mathbb{E}(|X_i|) \in \mathbb{R}_{\geqslant 0}$;*

(gI2) *$(X_i)_{i\in\mathcal{I}}$ ist gleichgradig stetig, dass heißt,*

$$\forall\, \varepsilon \in \mathbb{R}_{>0} : \exists\, \delta_\varepsilon \in \mathbb{R}_{>0} : \forall\, A \in \mathscr{A} \text{ mit } \mathbb{P}(A) < \delta_\varepsilon : \sup_{i\in\mathcal{I}} \mathbb{E}(|X_i|\mathbb{1}_A) < \varepsilon.$$

§28.39 **Beweis** von Satz §28.38. In der Vorlesung. □

§28.40 **Bemerkung**. Die Bedingung (gI2) ist eine Verallgemeinerung von Lemma §22.09. □

§28.41 **Proposition**.

(i) *Eine Familie $(X_i)_{i\in\mathcal{I}}$ in $\mathcal{L}_1$ mit endlicher Indexmenge $\mathcal{I}$ ist gleichgradig $\mathbb{P}$-integrierbar.*

(ii) *Sind die Familien $(X_i)_{i\in\mathcal{I}}$ und $(Y_j)_{j\in\mathcal{J}}$ in $\mathcal{L}_1$ gleichgradig $\mathbb{P}$-integrierbar, dann sind auch die Familien $(X_i + Y_j)_{i\in\mathcal{I},j\in\mathcal{J}}$, $(X_i - Y_j)_{i\in\mathcal{I},j\in\mathcal{J}}$ sowie $(|X_i|)_{i\in\mathcal{I}}$ gleichgradig $\mathbb{P}$-integrierbar.*

(iii) *Ist $(X_i)_{i\in\mathcal{I}}$ gleichgradig $\mathbb{P}$-integrierbar und existiert zu jedem Y_j, $j \in \mathcal{J}$ ein $i \in \mathcal{I}$ mit $|Y_j| \leqslant |X_i|$, so ist auch $(Y_j)_{j\in\mathcal{J}}$ gleichgradig $\mathbb{P}$-integrierbar.*

(iv) *Ist $(X_i)_{i\in\mathcal{I}}$ eine Familie identisch-verteilter Zufallsvariablen in $\mathcal{L}_1$, dann ist $(X_i)_{i\in\mathcal{I}}$ gleichgradig $\mathbb{P}$-integrierbar. Im Fall $\mathcal{I} = \mathbb{N}$ ist $(\overline{X}_n)_{n\in\mathbb{N}}$ gleichgradig $\mathbb{P}$-integrierbar.*

§28.42 **Beweis** von Proposition §28.41. In der Vorlesung. □

§28.43 **Proposition**. *Für $s \in \overline{\mathbb{R}}_{>1}$ sei $(X_i)_{i\in\mathcal{I}}$ eine in $\mathcal{L}_s$ beschränkte Familie, also $\sup_{i\in\mathcal{I}} \|X_i\|_{\mathcal{L}_s} \in \mathbb{R}_{\geqslant 0}$, dann ist $(X_i)_{i\in\mathcal{I}}$ gleichgradig $\mathbb{P}$-integrierbar.*

§28.44 **Beweis** von Proposition §28.43. In der Vorlesung. □

§28.45 **Lemma**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1$. Dann gilt $X_n \xrightarrow{\mathcal{L}_1} 0$ genau dann, wenn $(X_n)_{n\in\mathbb{N}}$ gleichgradig $\mathbb{P}$-integrierbar ist und $X_n \xrightarrow{\mathbb{P}} 0$.*

§28.46 **Beweis** von Lemma §28.45. In der Vorlesung. □

§28.47 **Satz**. *Für ein $s \in \mathbb{R}_{\geqslant 1}$ sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_s$ und sei $X \in \overline{\mathcal{M}}$ eine numerische Zufallsvariable. Dann gilt $X_n \xrightarrow{\mathcal{L}_s} X$ genau dann, wenn $(|X_n|^s)_{n\in\mathbb{N}}$ gleichgradig $\mathbb{P}$-integrierbar ist und $X_n \xrightarrow{\mathbb{P}} X$.*

§28.48 **Beweis** von Satz §28.47. In der Vorlesung. □

§28.49 **Korollar** *(Dominierte Konvergenz)*. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_1$ derart, dass gelten:*

(dK1”) $X_n \xrightarrow{\mathbb{P}} X$ *für ein* $X \in \overline{\mathcal{M}}$;

(dK2”) *Es existiert eine $\mathbb{P}$-integrierbare Majorante $Y \in \mathcal{L}_1(\mathbb{P}) \cap \overline{\mathcal{M}}_{\geqslant 0}$ mit $|X_n| \leqslant Y$ $\mathbb{P}$-f.s. für jedes $n \in \mathbb{N}$.*

Dann gilt $X \in \mathcal{L}_1$, $X_n \xrightarrow{\mathcal{L}_1} X$ und $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

§28.50 **Beweis** von Korollar §28.49. Die Aussage folgt direkt aus Satz §28.47, da (dK2”) die gleichgradige $\mathbb{P}$-Integrierbarkeit der Folge $(X_n)_{n\in\mathbb{N}}$ sichert (vgl. Bemerkung §28.37 (d)). □

§28.51 **Beispiel**. Die folgenden Beispiele zeigen, dass die Umkehrungen (in grün) der direkten Implikationen (in rot) in Skizze §28.52 nicht gelten. Dazu betrachten wir Folgen $(X_n)_{n\in\mathbb{N}}$ von reellen Zufallsvariablen auf dem Wahrscheinlichkeitsraum $([0,1], \mathscr{B}_{[0,1]}, \mathrm{U}_{[0,1]})$, sowie $\mathrm{r} \in \mathbb{R}_{>0}$ und $\mathrm{s} \in \mathbb{R}_{>\mathrm{r}}$.

(a) Sei $X_n := \mathbb{1}_{[0,n^{-1}]}$, dann gilt

$(X_n \overset{\mathcal{L}_\infty}{\not\to} 0)$, da $\|X_n\|_{\mathcal{L}_\infty} = \inf\{\varepsilon \in \mathbb{R}_{\geqslant 0} : \mathbb{P}(|X_n| > \varepsilon) = 0\} = 1$;

$(X_n \xrightarrow{\mathcal{L}_\mathrm{s}} 0)$, da $\|X_n\|_{\mathcal{L}_\mathrm{s}}^{\mathrm{s}} = n^{-1}$;

$(X_n \overset{\mathbb{P}\text{-f.v.}}{\not\to} 0)$, da $\{|X_n| > \varepsilon\} = [0, n^{-1}]$ für $\varepsilon \in (0,1)$ und $\{|X_n| > \varepsilon\} = \emptyset$ für $\varepsilon \in \mathbb{R}_{\geqslant 1}$, also $\mathbb{P}(|X_n| > \varepsilon) = n^{-1}$ für $\varepsilon \in (0,1)$ und $\mathbb{P}(|X_n| > \varepsilon) = 0$ für $\varepsilon \in \mathbb{R}_{\geqslant 1}$, so dass $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) = \sum_{n\in\mathbb{N}} n^{-1} = \infty$ für alle $\varepsilon \in (0,1)$;

$(X_n \xrightarrow{\mathbb{P}\text{-f.s.}} 0)$, da $\{\limsup_{n\to\infty} |X_n| = 0\} = (0,1]$, also $\mathbb{P}(\limsup_{n\to\infty} |X_n| = 0) = 1$;

(b) Sei $X_n := 2^{n/\mathrm{s}} \mathbb{1}_{[0,2^{-n}]}$, dann gilt

$(X_n \overset{\mathcal{L}_\mathrm{s}}{\not\to} 0)$, da $\|X_n\|_{\mathcal{L}_\mathrm{s}}^{\mathrm{s}} = 1$;

$(X_n \xrightarrow{\mathbb{P}\text{-f.v.}} 0)$, da $\{|X_n| > \varepsilon\} = [0, 2^{-n}]$ für $\varepsilon \in (0, 2^{n/\mathrm{s}})$ und $\{|X_n| > \varepsilon\} = \emptyset$ für $\varepsilon \in \mathbb{R}_{\geqslant 2^{n/\mathrm{s}}}$, also $\mathbb{P}(|X_n| > \varepsilon) = 2^{-n}$ für $\varepsilon \in (0, 2^{n/\mathrm{s}})$ und $\mathbb{P}(|X_n| > \varepsilon) = 0$ für $\varepsilon \in \mathbb{R}_{\geqslant 2^{n/\mathrm{s}}}$, so dass $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) \leqslant \sum_{n\in\mathbb{N}} 2^{-n} = 1 < \infty$ für alle $\varepsilon \in \mathbb{R}_{>0}$;

(c) Sei $X_{2^k+j} := 2^{k/\mathrm{r}} \mathbb{1}_{(j2^{-k},(j+1)2^{-k}]}$ mit $n = 2^k + j$ für $j \in [\![0, 2^k - 1]\!]$, $k \in \mathbb{Z}_{\geqslant 0}$, dann gilt

$(X_n \overset{\mathcal{L}_\mathrm{s}}{\not\to} 0)$, da $\|X_n\|_{\mathcal{L}_\mathrm{s}}^{\mathrm{s}} = 2^{\mathrm{s}k/\mathrm{s}} 2^{-k} = 1$;

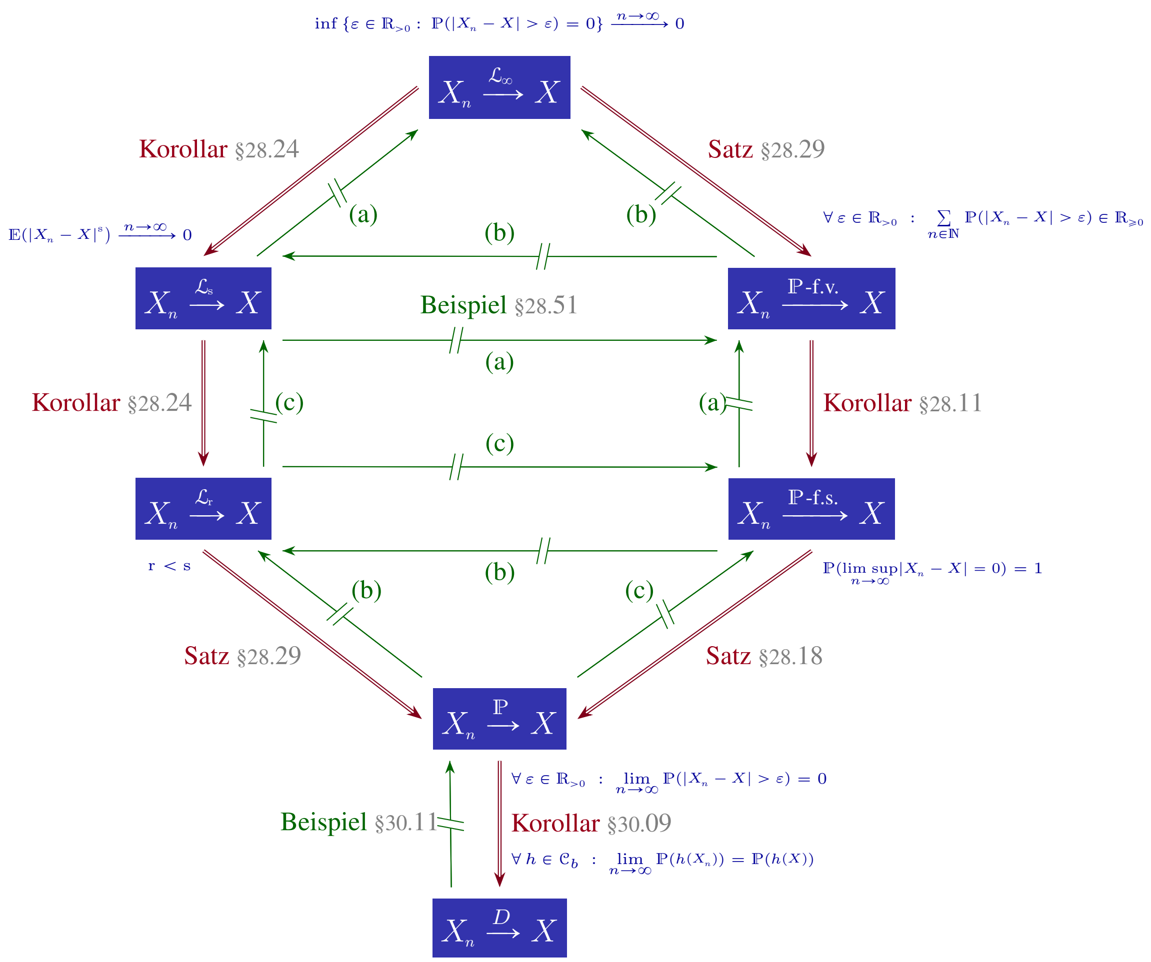
$(X_n \xrightarrow{\mathcal{L}_\mathrm{r}} 0)$, da $\|X_n\|_{\mathcal{L}_\mathrm{r}}^{\mathrm{r}} = 2^{(\mathrm{r}/\mathrm{s}-1)k}$;

$(X_n \overset{\mathbb{P}\text{-f.v.}}{\not\to} 0)$, da $\{|X_n| > \varepsilon\} = (j2^{-k}, (j+1)2^{-k}]$ für $\varepsilon \in (0, 2^{k/\mathrm{s}})$ und $\{|X_n| > \varepsilon\} = \emptyset$ für $\varepsilon \in \mathbb{R}_{\geqslant 2^{k/\mathrm{s}}}$, also $\mathbb{P}(|X_n| > \varepsilon) = 2^{-k}$ für $\varepsilon \in (0, 2^{k/\mathrm{s}})$ und $\mathbb{P}(|X_n| > \varepsilon) = 0$ für $\varepsilon \in \mathbb{R}_{\geqslant 2^{k/\mathrm{s}}}$, so dass $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) = \sum_{k\in\mathbb{N}_0} \sum_{j\in[\![0,2^k-1]\!]} 2^{-k} = \sum_{k\in\mathbb{Z}_{\geqslant 0}} 1 = \infty$ für alle $\varepsilon \in (0,1)$;

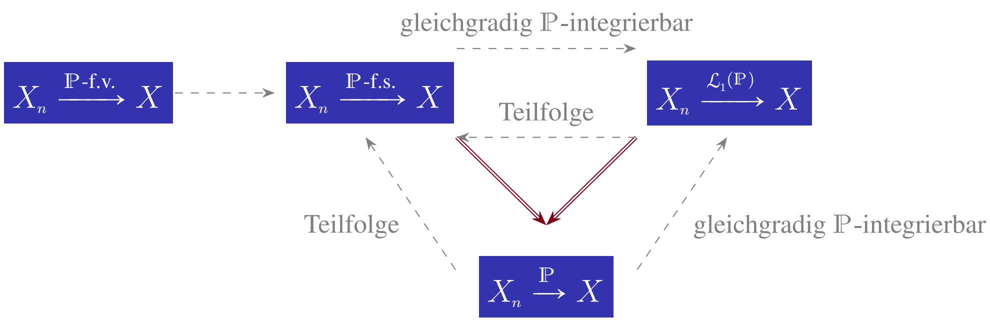
$(X_n \overset{\mathbb{P}\text{-f.s.}}{\not\to} 0)$, da $\{\limsup_{n\to\infty} |X_n| = 0\} = \{0\}$, also $\mathbb{P}(\limsup_{n\to\infty} |X_n| = 0) = 0$;

$(X_n \xrightarrow{\mathbb{P}} 0)$, da $\mathbb{P}(|X_n| > \varepsilon) \leqslant 2^{-k}$ für alle $\varepsilon \in \mathbb{R}_{>0}$. □

§28.52 **Skizze.** Die Gegenbeispiele in Beispiel §28.51 zeigen, dass die Umkehrungen (in grün) der folgenden direkten Implikationen (in rot) nicht gelten.



Die Aussagen der Implikationen zwischen den Konvergenzkonzepten der Sätze §28.18 und §28.29 fasst die nächste Darstellung zusammen, sie ist angelehnt an Klenke (2020, Abb.6.1, S.159).



□

§29 Gesetze der großen Zahlen

In diesem Abschnitt seien stets $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X_n)_{n\in\mathbb{N}}$ eine Folge reeller Zufallsvariablen in $\mathcal{L}_1 = \mathcal{L}_1(\Omega, \mathscr{A}, \mathbb{P})$. Wir definieren für jedes $n \in \mathbb{N}$ die Partialsumme und das arithmetische Mittel

$$S_n := \sum_{i\in[\![n]\!]} X_i \quad \text{und} \quad \overline{X_n} = n^{-1} S_n.$$

Wir setzen $S_0 := 0$. Da $X_i \in \mathcal{L}_1$ für jedes $i \in \mathbb{N}$ gilt, können wir die zentrierte Folge

$$(\overline{X_n} - \mathbb{E}(\overline{X_n}))_{n\in\mathbb{N}} = (n^{-1}\{S_n - \mathbb{E}(S_n)\})_{n\in\mathbb{N}}$$

betrachten, das heißt, für jedes $n \in \mathbb{N}$ gilt $\mathbb{E}(\overline{X_n} - \mathbb{E}(\overline{X_n})) = 0$. Wir sagen, die Folge $(X_n)_{n\in\mathbb{N}}$ erfüllt das *schwache* bwz. *starke Gesetz der großen Zahlen*, wenn die zentrierte Folge in $\mathbb{P}$-Wahrscheinlichkeit bzw. $\mathbb{P}$-fast sicher gegen 0 konvergiert. Damit beschäftigen wir uns in diesem Abschnitt hauptsächlich mit den asymptotischen Eigenschaften der Folge $(S_n)_{n\in\mathbb{N}}$. Wir halten fest, dass die zentrierte Folge $(n^{-1}\{S_n - \mathbb{E}(S_n)\})_{n\in\mathbb{N}}$ genau dann konvergiert, wenn für jedes feste $m \in \mathbb{Z}_{\geqslant 0}$ auch die Folge $(n^{-1}\{S_n - \mathbb{E}(S_n)\} - n^{-1}\{S_m - \mathbb{E}(S_m)\})_{n\in\mathbb{N}}$ gegen denselben Grenzwert konvergiert. Da offensichtlich

$$n^{-1}\{S_n - \mathbb{E}(S_n) - S_m + \mathbb{E}(S_m)\} = \tfrac{1}{n} \sum_{k\in(\!(m,n]\!]} (X_k - \mathbb{E}(X_k))$$

gilt, ist das Ereignis

$$\{\text{Die Folge } (n^{-1}\{S_n - \mathbb{E}(S_n)\})_{n\in\mathbb{N}} \text{ ist konvergent}\}$$

asymptotisch bzgl. $(X_n)_{n\in\mathbb{N}}$, dass heißt wie in Definition §16.13 eingeführt, ein Element der asymptotischen σ-Algebra. Falls $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsvariablen ist, folgt aus dem *0-1-Gesetz von Kolmogorov* §16.15, dass entweder mit Wahrscheinlichkeit 1 oder 0 die Folge $(n^{-1}\{S_n - \mathbb{E}(S_n)\})_{n\in\mathbb{N}}$ konvergiert.

§29|01 Gesetz der großen Zahlen in $\mathcal{L}_2$

§29.01 **Satz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_2$ paarweise unkorrelierter Zufallsvariablen mit beschränkter Varianz, das heißt $\mathbb{K}\mathrm{ov}(X_n, X_m) = 0$ für alle $n \in \mathbb{N}$, $m \in \mathbb{N}_{\neq n}$ sowie $\sup_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n) \in \mathbb{R}_{\geqslant 0}$, dann gilt*

$$\lim_{n\to\infty} \frac{S_n - \mathbb{E}(S_n)}{n} = 0 \quad \mathbb{P}\text{-f.s. } \textit{und in } \mathcal{L}_2.$$

§29.02 **Beweis** von Satz §29.01. In der Vorlesung. □

§29.03 **Korollar**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch-verteilter reeller Zufallsvariablen mit $X_1 \in \mathcal{L}_2$, dann gilt*

$$\lim_{n\to\infty} \overline{X_n} = \lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}(X_1) \quad \mathbb{P}\text{-f.s. } \textit{und in } \mathcal{L}_2.$$

§29.04 **Beweis** von Korollar §29.03. Die Aussage folgt direkt aus Satz §29.01. □

§29.05 **Bemerkung**. Wir werden zeigen, dass für fast-sichere Konvergenz Integrierbarkeit ausreicht. □

§29|02 Gesetz der großen Zahlen in $\mathcal{L}_1$

§29.06 **Lemma**. *Seien $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ zwei Folge reeller Zufallsvariablen derart, dass $\sum_{n\in\mathbb{N}} \mathbb{P}(X_n \neq Y_n) \in \mathbb{R}_{\geqslant 0}$ gilt, dann gilt auch $\sum_{n\in\mathbb{N}} |X_n - Y_n| \in \mathbb{R}_{\geqslant 0}$ $\mathbb{P}$-f.s. und somit*

$$\frac{1}{n}\left(\sum_{k\in[\![n]\!]} X_k - \sum_{k\in[\![n]\!]} Y_k\right) \xrightarrow{\mathbb{P}\text{-f.s.}} 0.$$

§29.07 **Beweis** von Lemma §29.06. In der Vorlesung. □

§29.08 **Bemerkung**. Betrachten wir Folgen $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ wie in Lemma §29.06, so erfüllt $(X_n)_{n\in\mathbb{N}}$ das starke Gesetz der großen Zahlen, wenn wir es für die Folge $(Y_n)_{n\in\mathbb{N}}$ zeigen können. □

§29.09 **Lemma**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge identisch-verteilter reeller Zufallsvariablen mit $X_1 \in \mathcal{L}_1$. Für die Folge $(Y_n)_{n\in\mathbb{N}}$ mit $Y_n := X_n \mathbb{1}_{\{|X_n|\leqslant n\}}$ gilt dann $\overline{X_n} - \overline{Y_n} \xrightarrow{\mathbb{P}\text{-f.s.}} 0$.*

§29.10 **Beweis** von Lemma §29.09. In der Vorlesung. □

§29.11 **Lemma**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch-verteilter positiver reeller Zufallsvariablen mit $\mathbb{E}(X_1) \in \mathbb{R}_{\geqslant 0}$. Für die Folge $(\overline{X_n})_{n\in\mathbb{N}}$ gilt dann $\overline{X_n} \xrightarrow{\mathbb{P}\text{-f.s.}} \mathbb{E}(X_1)$.*

§29.12 **Beweis** von Lemma §29.11. In der Vorlesung. □

§29.13 **Starkes Gesetz der großen Zahlen**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch-verteilter reeller Zufallsvariablen. Falls $X_1 \in \mathcal{L}_1$ ist, dann gilt*

$$\lim_{n\to\infty} \overline{X_n} = \mathbb{E}(X_1) \quad \mathbb{P}\text{-f.s. } \textit{und in } \mathcal{L}_1.$$

§29.14 **Beweis** von Satz §29.13. In der Vorlesung. □

§29.15 **Satz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch-verteilter reeller Zufallsvariablen.*

(i) *Falls $\mathbb{E}(X_1) = \infty$ ist, also $\mathbb{E}(X_1^-) \in \mathbb{R}_{\geqslant 0}$, gilt auch $\overline{X_n} \xrightarrow{\mathbb{P}\text{-f.s.}} \infty$.*

(ii) *Falls $\mathbb{E}(X_1) = -\infty$ ist, also $\mathbb{E}(X_1^+) \in \mathbb{R}_{\geqslant 0}$, gilt auch $\overline{X_n} \xrightarrow{\mathbb{P}\text{-f.s.}} -\infty$.*

(iii) *Falls $\mathbb{E}(|X_1|) = \infty$ ist, so gilt auch $\limsup_{n\to\infty} |\overline{X_n}| = \infty$ $\mathbb{P}$-f.s..*

§29.16 **Beweis** von Satz §29.15. In der Vorlesung. □

§29.17 **Bemerkung**. Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch-verteilter reeller Zufallsvariablen, so konvergiert also die Folge $(\overline{X_n})_{n\in\mathbb{N}}$ $\mathbb{P}$-f.s. gegen eine endlichen Grenzwert genau dann, wenn X_1 integrierbar ist. □

§29|03 Unabhängigkeit und Konvergenz in $\mathcal{L}_2$

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_2$. Wir suchen Bedingungen, sodass die zentrierte Folge $(S_n^\circ)_{n\in\mathbb{N}}$ mit

$$S_n^\circ := S_n - \mathbb{E}(S_n) = \sum_{i\in[\![n]\!]} (X_i - \mathbb{E}(X_i)) = \sum_{i\in[\![n]\!]} X_i^\circ, \quad n \in \mathbb{N},$$

in $\mathcal{L}_2$ und $\mathbb{P}$-f.s. konvergiert. Die Konvergenz in $\mathcal{L}_2$ ist einfach.

§29.18 **Satz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_2$ paarweise unkorrelierter reeller Zufallsvariablen mit summierbaren Varianzen, d.h. $\mathbb{K}\mathrm{ov}(X_n, X_m) = 0$ für alle $n \in \mathbb{N}$, $m \in \mathbb{N}_{\neq n}$ sowie $\sum_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n) \in \mathbb{R}_{\geqslant 0}$. Dann existiert $S_\infty^\circ \in \mathcal{L}_2$ mit $S_n - \mathbb{E}(S_n) \xrightarrow{\mathcal{L}_2} S_\infty^\circ$.*

§29.19 **Beweis** von Satz §29.18. In der Vorlesung. □

§29.20 **Ungleichung von Kolmogorov**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_2$ unabhängiger reeller Zufallsvariablen. Für alle $\varepsilon \in \mathbb{R}_{>0}$ und für alle $n \in \mathbb{N}$ gilt dann:*

$$\mathbb{P}(\max_{k\in[\![n]\!]} |S_k - \mathbb{E}(S_k)| > \varepsilon) \leqslant \varepsilon^{-2} \mathbb{V}\mathrm{ar}(S_n).$$

§29.21 **Beweis** von Lemma §29.20. In der Vorlesung. □

§29.22 **Bemerkung**. Für $n = 1$ erhalten wir die Tschebischeff Ungleichung §25.08. □

§29.23 **Satz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}_2$ unabhängiger reeller Zufallsvariablen mit summierbaren Varianzen, also $\sum_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n) \in \mathbb{R}_{\geqslant 0}$. Dann existiert $S_\infty^\circ \in \mathcal{L}_2$ mit*

$$\lim_{n\to\infty} \{S_n - \mathbb{E}(S_n)\} = S_\infty^\circ \quad \mathbb{P}\text{-f.s. } \textit{und in } \mathcal{L}_2.$$

Weiterhin gilt $S_\infty^\circ = \sum_{n\in\mathbb{N}}(X_n - \mathbb{E}(X_n))$ $\mathbb{P}$-f.s..

§29.24 **Beweis** von Satz §29.23. In der Vorlesung. □

§29.25 **Beispiel** *(Beispiel §28.31 fortgesetzt.)*. Sei $(Y_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger reeller Zufallsvariablen mit $\mathbb{P}(Y_n = 1/n) = \mathbb{P}(Y_n = -1/n) = 1/2$. Für jedes $n \in \mathbb{N}$ gilt dann $\mathbb{E}(Y_n) = 0$ und $\mathbb{V}\mathrm{ar}(Y_n) = n^{-2}$ und somit $\sum_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(Y_n) \in \mathbb{R}_{\geqslant 0}$. Nach Satz §29.23 konvergiert $S_n = \sum_{k\in[\![n]\!]} Y_k$ in $\mathcal{L}_2$ (wie in Beispiel §28.31 schon gezeigt) und auch $\mathbb{P}$-f.s.. □

§29.26 **Ungleichung von Lévy-Ottaviani**. *Für $n \in \mathbb{N}$ seien $(X_i)_{i\in[\![n]\!]}$ unabhängige reelle Zufallsvariablen. Für alle $a, b \in \mathbb{R}_{>0}$ gilt dann:*

$$\mathbb{P}(\max_{k\in[\![n]\!]} |S_k| \geqslant a + b) \leqslant \frac{\mathbb{P}(|S_n| \geqslant b)}{1 - \max_{k\in[\![n]\!]} \mathbb{P}(|S_n - S_k| > a)}.$$

§29.27 **Beweis** von Lemma §29.26. In der Vorlesung. □

§29.28 **Lévy's Äquivalenzsatz**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger reeller Zufallsvariablen. Dann sind äquivalent:*

(A1) *$(S_n)_{n\in\mathbb{N}}$ konvergiert fast sicher;*

(A2) *$(S_n)_{n\in\mathbb{N}}$ konvergiert stochastisch.*

Andernfalls ist $(S_n)_{n\in\mathbb{N}}$ divergent mit Wahrscheinlichkeit Eins.

§29.29 **Beweis** von Satz §29.28. In der Vorlesung. □

§29.30 **Dreireihensatz von Kolmogorov**. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger reeller Zufallsvariablen. Dann konvergiert die Folge der Partialsummen $(S_n)_{n\in\mathbb{N}}$ $\mathbb{P}$-f.s. genau dann, wenn die folgenden drei Bedingungen (für irgendein $\varepsilon \in \mathbb{R}_{>0}$ und damit für alle) gelten:*

(B1) $\sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > \varepsilon) \in \mathbb{R}_{\geqslant 0}$;

(B2) $\sum_{n\in\mathbb{N}} \mathbb{E}(X_n \mathbb{1}_{\{|X_n|\leqslant\varepsilon\}})$ *konvergiert;*

(B3) $\sum_{n\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_n \mathbb{1}_{\{|X_n|\leqslant\varepsilon\}}) \in \mathbb{R}_{\geqslant 0}$.

§29.31 **Beweis** von Satz §29.30. Klenke (2020, Satz 15.51, S.360) □

§29.32 **Beispiel** *(Beispiel §29.25 fortgesetzt.)*. Sei $(Y_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger reeller Zufallsvariablen mit $\mathbb{P}(Y_n = 1/n) = 1 - \mathbb{P}(Y_n = -1/n) = \mathrm{p} \in [0, 1]$. Für jedes $k \in \mathbb{N}$ gilt dann $\mathbb{E}(Y_k \mathbb{1}_{\{|Y_k|\leqslant 1\}}) = \frac{2\mathrm{p}-1}{k}$ und somit konvergiert $\sum_{k\in[\![n]\!]} \mathbb{E}(Y_k \mathbb{1}_{\{|Y_k|\leqslant 1\}}) = (2\mathrm{p} - 1) \sum_{k\in[\![n]\!]} \frac{1}{k}$ genau dann, wenn $\mathrm{p} = 1/2$ gilt. Nach Satz §29.30 konvergiert $S_n = \sum_{k\in[\![n]\!]} Y_k$ $\mathbb{P}$-f.s. also genau dann, wenn $\mathrm{p} = 1/2$ gilt. □

§30 Konvergenz in Verteilung

§30|00.01 **Vorbemerkung**. Für Zufallsvektoren $X, X_1, X_2, X_3, \dots$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ mit Werten in $(\mathbb{R}^k, \mathscr{B}^k)$ versehen wie bisher mit dem euklidischen Abstand $d(x, y) = \|x - y\|$ für $x, y \in \mathbb{R}^k$, kurz X und $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{M}(\mathscr{A}, \mathscr{B}^k)$, sagen wir $(X_n)_{n\in\mathbb{N}}$ konvergiert gegen X stochastisch bzw. $\mathbb{P}$-fast sicher, wenn die Folge $(\|X_n - X\|)_{n\in\mathbb{N}}$ der Abstände stochastisch bzw. $\mathbb{P}$-fast sicher gegen Null konvergiert. Wir schreiben dann ebenfalls $X_n \xrightarrow{\mathbb{P}} X$ bzw. $X_n \xrightarrow{\mathbb{P}\text{-f.s.}} X$. Wir halten fest, dass diese Konvergenzbegriffe direkt die Differenz $X_n - X$ verwenden. Häufig interessieren wir uns nur für die von X_n bzw. X induzierten Verteilungen auf $(\mathbb{R}^k, \mathscr{B}^k)$. Bemerkenswert hier bei ist, dass es für die folgende Konvergenz in Verteilung möglich ist, dass die Zufallsvariablen X_n und X nicht auf dem gleichen Wahrscheinlichkeitsraum definiert sind, da diese nicht direkt die Differenz $X_n - X$ verwendet.
Im Folgenden bezeichnen wir mit $\mathcal{C}_b := \mathcal{C}_b(\mathbb{R}^k)$ die Menge aller beschränkten, stetigen, reellen Funktionen auf $\mathbb{R}^k$. Für $h \in \mathcal{C}_b$ ist somit $\|h\|_\infty := \sup_{x\in\mathbb{R}^k} |h(x)| \in \mathbb{R}_{\geqslant 0}$, so dass für jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}^k, \mathscr{B}^k)$, kurz $\mathbb{P} \in \mathcal{W}(\mathscr{B}^k)$, gilt $\mathcal{C}_b \subseteq \mathcal{L}_\infty(\mathbb{R}^k, \mathscr{B}^k, \mathbb{P})$. □

§30.02 **Definition**.

(a) Eine Folge $(\mathbb{P}_n)_{n\in\mathbb{N}}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^k, \mathscr{B}^k)$ *konvergiert schwach* gegen ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}^k, \mathscr{B}^k)$, wenn $\lim_{n\to\infty} \mathbb{P}_n(h) = \mathbb{P}(h)$ für alle $h \in \mathcal{C}_b$ gilt. Wir schreiben kurz $\boxed{\mathbb{P}_n \xrightarrow{w} \mathbb{P}}$ oder $\mathbb{P} = \underset{n\to\infty}{w\text{-lim}}\, \mathbb{P}_n$.

(b) Für jedes $n \in \mathbb{N}$ sei $(\Omega_n, \mathscr{A}_n, \mathbb{P}_n)$ ein Wahrscheinlichkeitraum und $X_n \in \mathcal{M}(\mathscr{A}_n, \mathscr{B}^k)$ ein Zufallsvektor mit Werten in $(\mathbb{R}^k, \mathscr{B}^k)$. Die Folge $(X_n)_{n\in\mathbb{N}}$ *konvergiert in Verteilung* gegen einen $\mathbb{R}^k$-wertigen Zufallsvektor $X \in \mathcal{M}(\mathscr{A}, \mathscr{B}^k)$ auf einem Wahrscheinlichkeitraum $(\Omega, \mathscr{A}, \mathbb{P})$, kurz $\boxed{X_n \xrightarrow{D} X}$, wenn $\lim_{n\to\infty} \mathbb{P}_n(h(X_n)) = \mathbb{P}(h(X))$ für alle $h \in \mathcal{C}_b$, also $\mathbb{P}^X = \underset{n\to\infty}{w\text{-lim}}\, \mathbb{P}_n^{X_n}$, gilt.

Für eine Folge $(X_n)_{n\in\mathbb{N}}$ definieren wir *Konvergenz in Verteilung unter* $\mathbb{P}_n$ mit einem Wahrscheinlichkeitsmaß $\mathbb{P}$ als Grenzwert, kurz $X_n \xrightarrow{D} \mathbb{P}$ *unter* $\mathbb{P}_n$, allgemein durch $\mathbb{P}_n^{X_n} \xrightarrow{w} \mathbb{P}$. □

§30.03 **Beispiel**. Seien $(\mathrm{a}_n)_{n\in\mathbb{N}}$ und $(\mathrm{b}_n)_{n\in\mathbb{N}}$ konvergente Folgen in $\mathbb{R}$ bzw. $\mathbb{R}^k$ mit Grenzwerten $\mathrm{a} \in \mathbb{R}$ und $\mathrm{b} \in \mathbb{R}^k$, also $\mathrm{a}_n \xrightarrow{n\to\infty} \mathrm{a}$ und $\mathrm{b}_n \xrightarrow{n\to\infty} \mathrm{b}$. Für jeden $\mathbb{R}^k$-wertigen Zufallsvektor X gilt dann

$$\mathrm{a}_n X + \mathrm{b}_n \xrightarrow{D} \mathrm{a}X + \mathrm{b}.$$

In der Tat für $h \in \mathcal{C}_b$ impliziert die Stetigkeit von h, dass

$$h(\mathrm{a}_n X(\omega) + \mathrm{b}_n) \xrightarrow{n\to\infty} h(\mathrm{a}X(\omega) + \mathrm{b}) \quad \forall \omega \in \Omega.$$

Da $\|h\|_\infty \in \mathbb{R}_{\geqslant 0}$ eine integrierbare Majorante ist, folgt mit dem Satz von der dominierten Konvergenz

$$\mathbb{E}(h(\mathrm{a}_n X + \mathrm{b}_n)) \xrightarrow{n\to\infty} \mathbb{E}(h(\mathrm{a}X + \mathrm{b})).$$

Achtung: $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ impliziert *nicht* $\mathbb{P}_n(B) \xrightarrow{n\to\infty} \mathbb{P}(B)$ für alle $B \in \mathscr{B}^k$. In der Tat: setzen wir $\mathrm{a}_n = a = 0$, so gilt $\delta_{\mathrm{b}_n} =: \mathbb{P}_n \xrightarrow{w} \mathbb{P} := \delta_{\mathrm{b}}$, während für $\mathrm{b}_n \neq b$ für alle $n \in \mathbb{N}$ gilt $\mathbb{P}_n(\{\mathrm{b}\}) = 0 \neq 1 = \mathbb{P}(\{\mathrm{b}\})$. □

§30.04 **Anmerkung**. Es folgen Charakterisierungen der schwachen Konvergenz. Wir erinnern daran, dass wir kurz $h \in \mathcal{M}(\mathscr{B}^k) := \mathcal{M}(\mathscr{B}^k, \mathscr{B})$ für eine Borel-messbare Funktion $h : \mathbb{R}^k \to \mathbb{R}$ schreiben.

Mit $\mathcal{U}_h := \{x \in \mathbb{R}^k : h \text{ ist nicht stetig in } x\}$ bezeichnen wir die Borel-messbare Menge aller *Unstetigkeitsstellen* von h, sodass $\mathcal{U}_h^c = \mathbb{R} \setminus \mathcal{U}_h$ gerade die Menge der Stetigkeitspunkte von h ist. Weiterhin heißt eine Funktion $h : \mathbb{R}^k \to \mathbb{R}$ *Lipschitz-stetig*, wenn ein $L \in \mathbb{R}_{>0}$ existiert mit $|h(x) - h(y)| \leqslant L\|x - y\|$ für alle $x, y \in \mathbb{R}^k$. Ist $B \in \mathscr{B}^k$, so bezeichnen wir mit $\overline{B}$ den *Abschluss* von B, mit B° das *Innere* und mit $\partial B = \overline{B} \setminus B^\circ$ den *Rand* von B. □

§30.05 **Lemma** *(Portmanteau)*. *Für Wahrscheinlichkeitmaße $\mathbb{P}, \mathbb{P}_n \in \mathcal{W}(\mathscr{B}^k)$, $n \in \mathbb{N}$, sind die folgenden Aussagen äquivalent:*

(A1) $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

(A2) *Für jede beschränkte Lipschitz-stetige Funktion $h : \mathbb{R}^k \to \mathbb{R}$ gilt $\lim_{n\to\infty} \mathbb{P}_n(h) = \mathbb{P}(h)$.*

(A3) *Für jede beschränkte Funktion $h \in \mathcal{M}(\mathscr{B}^k)$ mit $\mathbb{P}(\mathcal{U}_h) = 0$ gilt $\lim_{n\to\infty} \mathbb{P}_n(h) = \mathbb{P}(h)$.*

(A4) *Für jedes abgeschlossene Ereignis $A \in \mathscr{B}^k$ gilt $\lim\sup_{n\to\infty} \mathbb{P}_n(A) \leqslant \mathbb{P}(A)d$.*

(A5) *Für jedes offene Ereignis $O \in \mathscr{B}^k$ gilt $\lim\inf_{n\to\infty} \mathbb{P}_n(O) \geqslant \mathbb{P}(O)$.*

(A6) *Für jedes Ereignis $B \in \mathscr{B}^k$ mit $\mathbb{P}(\partial B) = 0$ gilt $\lim_{n\to\infty} \mathbb{P}_n(B) = \mathbb{P}(B)$.*

§30.06 **Beweis** von Lemma §30.05. In der Vorlesung. □

§30.07 **Satz von Slutzky**. *Seien X, X_n und Y_n, $n \in \mathbb{N}$, Zufallsvektoren. Konvergiert $(Y_n)_{n\in\mathbb{N}}$ in Verteilung gegen X, also $Y_n \xrightarrow{D} X$, und konvergiert $(X_n - Y_n)_{n\in\mathbb{N}}$ stochastisch gegen Null, also $\|X_n - Y_n\| \xrightarrow{\mathbb{P}} 0$, dann konvergiert $(X_n)_{n\in\mathbb{N}}$ in Verteilung auch gegen X, also $X_n \xrightarrow{D} X$.*

§30.08 **Beweis** von Satz §30.07. In der Vorlesung. □

§30.09 **Korollar**. *Konvergiert eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvektoren gegen einen Zufallsvektor X stochastisch, also $X_n \xrightarrow{\mathbb{P}} X$, so auch in Verteilung, also $X_n \xrightarrow{D} X$, aber die Umkehrung gilt nicht.*

§30.10 **Beweis** von Korollar §30.09. Die Aussage folgt (mit $Y_n := X$) direkt aus Satz §30.07. □

§30.11 **Beispiel**. Sei $Z \sim \mathrm{N}_{(0,1)}$. Setzen wir $X_n := Z$ für $n \in \mathbb{N}$ und $X := -Z$, dann gilt $X_n \xrightarrow{D} X$, während für alle $\varepsilon \in \mathbb{R}_{>0}$ die Wahrscheinlichkeit

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(2|Z| > \varepsilon) = 2\Phi(-\varepsilon/2) \in \mathbb{R}_{>0}$$

für $n \to \infty$ nicht gegen Null konvergiert. □

§30.12 **Korollar**. *Eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvektoren konvergiert in Verteilung gegen eine Konstante $\mathrm{a} \in \mathbb{R}^k$, also $X_n \xrightarrow{D} \mathrm{a}$, genau dann, wenn $\lim_{n\to\infty} \mathbb{P}(\|X_n - \mathrm{a}\| > \varepsilon) = 0$ für alle $\varepsilon \in \mathbb{R}_{>0}$ gilt.*

§30.13 **Beweis** von Korollar §30.12. In der Vorlesung. □

§30.14 **Bemerkung**. In der Situation von Korollar §30.12, sagen wir, dass $(X_n)_{n\in\mathbb{N}}$ gegen a stochastisch konvergiert und wir schreiben auch kurz $X_n \xrightarrow{\mathbb{P}} \mathrm{a}$, selbst wenn die Zufallsvariablen nicht über dem gleichen Wahrscheinlichkeitsraum definiert sind. □

§30.15 **Satz von der stetigen Abbildung**. *Sei $h \in \mathcal{M}(\mathscr{B}^k, \mathscr{B}^m)$ mit Unstetigkeitsmenge $\mathcal{U}_h \in \mathscr{B}^k$.*

(i) *Seien $\mathbb{P}, \mathbb{P}_n \in \mathcal{W}(\mathscr{B}^k)$, $n \in \mathbb{N}$, Wahrscheinlichkeitmaße mit $\mathbb{P}(\mathcal{U}_h) = 0$ und $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$. Dann gilt $\mathbb{P}_n \circ h^{-1} = \mathbb{P}_n^h \xrightarrow{w} \mathbb{P}^h = \mathbb{P} \circ h^{-1}$.*

(ii) *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvektoren, die in Verteilung gegen einen Zufallsvektor X konvergiert, also $X_n \xrightarrow{D} X$. Ist $\mathbb{P}(X \in \mathcal{U}_h) = 0$ so konvergiert $(h(X_n))_{n\in\mathbb{N}}$ auch gegen $h(X)$ in Verteilung, also $h(X_n) \xrightarrow{D} h(X)$.*

§30.16 **Beweis** von Satz §30.15. Der Beweis von (i) ist eine Übungsaufgabe, und (ii) ist nur eine Umformulierung von (i). □

§30.17 **Korollar**. *Sei $((X_n, Y_n))_{n\in\mathbb{N}}$ eine Folge von $\mathbb{R}^{k+m}$-wertigen Zufallsvektoren, X ein $\mathbb{R}^k$-wertiger Zufallsvektor und $\mathrm{a} \in \mathbb{R}^m$ eine Konstante. Falls $X_n \xrightarrow{D} X$ (in $\mathbb{R}^k$) und $Y_n \xrightarrow{\mathbb{P}} a$ (in $\mathbb{R}^m$), so gilt $(X_n, Y_n) \xrightarrow{D} (X, \mathrm{a})$, im Fall $k = m$, $X_n + Y_n \xrightarrow{D} X + \mathrm{a}$ sowie im Fall $m = 1$, $Y_n X_n \xrightarrow{D} \mathrm{a}X$.*

§30.18 **Beweis** von Korollar §30.17. In der Vorlesung. □

§30.19 **Satz** *(Delta Methode)*. *Seien $(X_n)_{n\in\mathbb{N}}$ eine Folge reeller Zufallsvariablen, Y eine reelle Zufallsvariable, $x \in \mathbb{R}$ eine Konstante und $(\mathrm{a}_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}_{>0}$ mit $\mathrm{a}_n \xrightarrow{n\to\infty} \infty$, derart dass $\mathrm{a}_n(X_n - x) \xrightarrow{D} Y$. Dann gilt $X_n \xrightarrow{\mathbb{P}} x$. Ist weiterhin $f \in \mathcal{M}(\mathscr{B})$ differenzierbar in x, so gilt*

$$\mathrm{a}_n(f(X_n) - f(x)) - \mathrm{a}_n(X_n - x)f'(x) \xrightarrow{\mathbb{P}} 0$$

und somit auch

$$\mathrm{a}_n(f(X_n) - f(x)) \xrightarrow{D} f'(x)Y.$$

§30.20 **Beweis** von Satz §30.19. In der Vorlesung. □

§30.21 **Definition**. Eine Folge $(\mathbb{F}_n)_{n\in\mathbb{N}}$ von Verteilungsfunktionen auf $\mathbb{R}$ *konvergiert schwach* gegen eine Verteilungsfunktion $\mathbb{F}$ auf $\mathbb{R}$, kurz $\mathbb{F}_n \xrightarrow{w} \mathbb{F}$, wenn $\mathbb{F}_n(x) \xrightarrow{n\to\infty} \mathbb{F}(x)$ für alle $x \in \mathcal{U}_{\mathbb{F}}^c = \mathbb{R}\setminus\mathcal{U}_{\mathbb{F}}$, an denen $\mathbb{F}$ also stetig ist (Stetigkeitspunkte von $\mathbb{F}$), gilt. □

§30.22 **Satz**. *Für reelle Zufallsvariablen sind äquivalent:*

(i) *$X_n \xrightarrow{D} X$;*

(ii) *Die Verteilungsfunktionen erfüllen $\mathbb{F}^{X_n} \xrightarrow{w} \mathbb{F}^X$.*

§30.23 **Beweis** von Satz §30.22. In der Vorlesung. □

§30.24 **Beispiel**.

(a) Sei $(\sigma_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}_{>0}$ mit $\sigma_n \xrightarrow{n\to\infty} 0$ und $X_n \sim \mathrm{N}_{(0,\sigma_n^2)}$ für $n \in \mathbb{N}$. Dann gilt $X_n \xrightarrow{D} \delta_0 = \mathrm{N}_{(0,0)}$.

(b) Seien X und X_n für $n \in \mathbb{N}$ diskrete $\mathbb{Z}$-wertige Zufallsvariablen, so gilt $X_n \xrightarrow{D} X$ genau dann, wenn für die Zähldichten $\mathbb{p}^{X_n}(z) \xrightarrow{n\to\infty} \mathbb{p}^X(z)$ für alle $z \in \mathbb{Z}$ gilt. Somit besagt der *Poissonsche Grenzwertsatz* §04.09 $\mathrm{Bin}_{(n,\mathrm{p}_n)} \xrightarrow{w} \mathrm{Poi}_\lambda$ für $np_n \xrightarrow{n\to\infty} \lambda \in \mathbb{R}_{>0}$.

(c) Sei $(U_i)_{i\in\mathbb{N}}$ eine Folge unabhängiger und identisch $\mathrm{U}_{[0,1]}$-verteilter Zufallsvariablen. Für jedes $n \in \mathbb{N}$ besitzt dann die reelle Zufallsvariable $X_n := n \min_{i\in[\![n]\!]} U_i$ die Verteilungsfunktion

$$\mathbb{R}_{\geqslant 0} \ni x \mapsto \mathbb{F}^{X_n}(x) = 1 - ((1 - \tfrac{x}{n}) \vee 0)^n \in [0,1].$$

Damit folgt $n \min_{i\in[\![n]\!]} U_i \xrightarrow{D} \mathrm{Exp}_1$. Da $1 - U_1 \sim \mathrm{U}_{[0,1]}$ gilt auch $n(1 - \max_{i\in[\![n]\!]} U_i) \xrightarrow{D} \mathrm{Exp}_1$.

(d) Für $\theta \in \mathbb{R}_{>0}$ sei $(X_i)_{i\in\mathbb{N}}$ eine Folge unabhängiger und identisch $\mathrm{U}_{[0,\theta]}$-verteilter Zufallsvariablen. Da $U_i := 1 - \frac{1}{\theta}X_i \sim \mathrm{U}_{[0,1]}$ gilt $\frac{n}{\theta}(\theta - \max_{i\in[\![n]\!]} X_i) \xrightarrow{D} \mathrm{Exp}_1$. □

§30.25 **Auswahlsatz von Helly**. *Sei $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine Folge von Verteilungen auf $(\mathbb{R},\mathscr{B})$ mit entsprechenden Verteilungsfunktionen $(\mathbb{F}_n)_{n\in\mathbb{N}}$. Dann existiert eine Teilfolge $(\mathbb{F}_{n_k})_{k\in\mathbb{N}}$ und eine monoton wachsende und rechtsstetige Funktion $\mathbb{F} : \mathbb{R} \to [0,1]$ mit $\mathbb{F}_{n_k}(x) \xrightarrow{k\to\infty} \mathbb{F}(x)$ für alle $x \in \mathcal{U}_{\mathbb{F}}^c = \mathbb{R}\backslash\mathcal{U}_{\mathbb{F}}$.*

§30.26 **Beweis** von Satz §30.25. In der Vorlesung. □

§30.27 **Beispiel**.

(a) Sind $(\mathbb{P}_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R},\mathscr{B})$ mit Verteilungsfunktionen $(\mathbb{F}_n)_{n\in\mathbb{N}}$, so folgt aus $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ nach Satz §30.22, dass $\mathbb{F}_n(x) \to \mathbb{F}(x)$ an den Stetigkeitspunkten x von $\mathbb{F}$ gerade für die Verteilungsfunktion $\mathbb{F}$ von $\mathbb{P}$ gilt.

(b) Für $n \in \mathbb{N}$ sei $\mathbb{P}_n = \mathrm{U}_{[n,n+1]}$ mit Verteilungsfunktion $\mathbb{F}_n$, so gilt $\mathbb{F}_n(x) \xrightarrow{n\to\infty} 0$ für alle $x \in \mathbb{R}$. Für $\mathbb{P}_n = \mathrm{N}_{(0,n)}$ gilt $\mathbb{F}_n(x) \xrightarrow{n\to\infty} 1/2$ für alle $x \in \mathbb{R}$. Die Funktion $\mathbb{F}$ im Auswahlsatz von Helly ist hier jeweils keine Verteilungsfunktion. □

§30.28 **Definition**. Eine Familie $\mathcal{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathscr{B})$ heißt *gleichgradig straff*, wenn für jedes $\varepsilon \in \mathbb{R}_{>0}$ ein $\mathrm{K}_\varepsilon \in \mathbb{R}_{>0}$ existiert mit $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}([-\mathrm{K}_\varepsilon, \mathrm{K}_\varepsilon]^c) \leqslant \varepsilon$. □

§30.29 **Bemerkung**.

(a) Eine Familie $\mathcal{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathscr{B})$ mit entsprechender Familie $\mathcal{F}$ von Verteilungsfunktionen ist gleichgradig straff genau dann, wenn

$$\lim_{x\to-\infty} \sup_{\mathbb{F}\in\mathcal{F}} \mathbb{F}(x) = 0 \quad \text{und} \quad \lim_{x\to\infty} \inf_{\mathbb{F}\in\mathcal{F}} \mathbb{F}(x) = 1$$

gilt. In der Tat, es genügt zu bemerken, dass für $x \in \mathbb{R}_{>0}$, $y \in \mathbb{R}_{>x}$ gilt

$$\mathbb{F}(-y) + 1 - \mathbb{F}(x) \leqslant \mathbb{P}([-x,x]^c) \leqslant \mathbb{F}(-x) + 1 - \mathbb{F}(x).$$

(b) Jede endliche Familie $\mathcal{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathscr{B})$ ist gleichgradig straff. □

§30.30 **Beispiel**. Sei $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine schwach konvergente Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathscr{B})$, etwa $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, und bezeichne $\mathbb{F}$ die Verteilungsfunktion von $\mathbb{P}$. Dann gibt es für jedes $\varepsilon \in \mathbb{R}_{>0}$ ein $x \in \mathbb{R}_{>0}$, derart dass $\mathbb{F}(-x) \leqslant \varepsilon/4$, $\mathbb{F}(x) \geqslant 1 - \varepsilon/4$ und $\mathbb{F}$ ist stetig in x und $-x$. Dann folgt

$$\mathbb{P}_n([-x,x]) \xrightarrow{n\to\infty} \mathbb{F}(x) - \mathbb{F}{-x} \geqslant 1 - \varepsilon/2.$$

Damit existiert ein $N \in \mathbb{N}$ mit $\mathbb{P}_n([-x,x]) \geqslant 1-\varepsilon$ für alle $n \in \mathbb{N}_{\geqslant N}$. Weiterhin existiert auf Grund der σ-Stetigkeit ein $y \in \mathbb{R}_{\geqslant x}$ mit $\mathbb{P}_n([-y,y]) \geqslant 1-\varepsilon$ für alle $n \in [\![1,N)\!)$. Somit gilt $\mathbb{P}_n([-y,y]^c) \leqslant \varepsilon$ für alle $n \in \mathbb{N}$ und $(\mathbb{P}_n)_{n\in\mathbb{N}}$ ist also straff. □

§30.31 **Korollar**. *Sei $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine gleichgradig straffe Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathscr{B})$. Dann existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R},\mathscr{B})$ derart, dass $\mathbb{P}_{n_k} \xrightarrow{w} \mathbb{P}$ gilt.*

§30.32 **Beweis** von Korollar §30.31. In der Vorlesung. □

§30.33 **Korollar**. *Für Wahrscheinlichkleitsmaße $\mathbb{P}, \mathbb{P}_n \in \mathcal{W}(\mathscr{B})$, $n \in \mathbb{N}$ gilt $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ genau dann, wenn jede Teilfolge $(\mathbb{P}_{n_k})_{k\in\mathbb{N}}$ eine schwach gegen $\mathbb{P}$ konvergierende Teilfolge $(\mathbb{P}_{n_{k_j}})_{j\in\mathbb{N}}$, also $\mathbb{P}_{n_{k_j}} \xrightarrow{w} \mathbb{P}$, enthält.*

§30.34 **Beweis** von Korollar §30.33. In der Vorlesung. □

§31 Charakteristische Funktionen

§31|00.01 **Vorbemerkung**. Im Folgenden bezeichnet $\mathbb{C} = \{a + \imath b \mid a, b \in \mathbb{R}\}$ den Körper der komplexen Zahlen. Für eine komplexe Zahl $z = a + \imath b \in \mathbb{C}$ bezeichnen $\mathrm{Re}(z) = a$ und $\mathrm{Im}(z) = b$ den Realteil bzw. den Imaginärteil von z, $\overline{z} = a - \imath b$ die zu z komplex konjugierte Zahl, und $|z| = \sqrt{a^2 + b^2}$ den Betrag von z. Die komplexe Exponentialfunktion ist definiert durch

$$\exp : \mathbb{C} \ni z \mapsto e^z = \exp(z) := \exp(x)(\cos(y) + \imath \sin(y)) \in \mathbb{C}$$

oder äquivalent durch die Potenzreihe $\exp(z) = \sum_{n\in\mathbb{Z}_{\geqslant 0}} z^n/n!$. Wir bezeichnen mit $\mathscr{B}_{\mathbb{C}}$ die Borel-σ-Algebra über $\mathbb{C}$, wobei wir wie üblich jede komplexe Zahl $z = a + \imath b$ mit $(a, b) \in \mathbb{R}^2$ identifizieren. Sei $(\Omega, \mathscr{B}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Abbildung $Z : \Omega \to \mathbb{C}$ heißt komplexe Zufallsvariable, wenn sie $\mathscr{A}$-$\mathscr{B}_{\mathbb{C}}$ messbar ist. Eine notwendige und hinreichende Bedingung ist, dass $X = \mathrm{Re}(Z)$ und $Y = \mathrm{Im}(Z)$ reelle Zufallsvariablen sind. Eine komplexe Zufallsvariable $Z \in \mathcal{M}(\mathscr{A}, \mathscr{B}_{\mathbb{C}})$ heißt $\mathbb{P}$-integrierbar, wenn $|Z| \in \mathcal{L}_1(\mathbb{P})$ gilt. Wir bezeichnen die Menge aller $\mathbb{P}$-integrierbaren komplexen Zufallsvariablen ebenfalls mit $\mathcal{L}_1 := \mathcal{L}_1(\mathbb{P})$. Offensichtlich gilt $Z = X + \imath Y \in \mathcal{L}_1$ genau dann, wenn $X, Y \in \mathcal{L}_1$ gilt. In diesem Fall definieren wir $\mathbb{E}(Z) := \mathbb{E}(X) + \imath\mathbb{E}(Y)$, wobei insbesondere $|\mathbb{E}(Z)| \leqslant \mathbb{E}(|Z|)$ gilt und auch die Erwartung $\mathbb{E} : \mathcal{L}_1 \to \mathbb{C}$ ist linear. Weiterhin, gilt für die komplex konjugierte $\overline{Z}$ von $Z \in \mathbb{C}$, dass $\overline{Z} \in \mathcal{L}_1$ genau dann gilt, wenn $Z \in \mathcal{L}_1$ ist, und in diesem Fall gilt $\mathbb{E}(\overline{Z}) = \overline{\mathbb{E}(Z)}$. Wenn weiterhin Z_1, Z_2 unabhängige komplexe Zufallsvariablen in $\mathcal{L}_1$ sind, dann gilt $Z_1Z_2 \in \mathcal{L}_1$ und $\mathbb{E}(Z_1Z_2) = \mathbb{E}(Z_1)\mathbb{E}(Z_2)$. □

§31.02 **Definition**.

(a) Für eine reelle Zufallsvariable X bezeichnet $\boxed{\varphi_X} : \mathbb{R} \to \mathbb{C}$ mit

$$u \mapsto \varphi(u) := \mathbb{E}(e^{\imath uX}) = \mathbb{E}(\cos(uX)) + \imath\mathbb{E}(\sin(uX))$$

die *charakteristische Funktion* von X.

(b) Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathscr{B})$ bezeichnet $\boxed{\varphi_{\mathbb{P}}} : \mathbb{R} \to \mathbb{C}$ mit

$$u \mapsto \varphi_{\mathbb{P}}(u) := \int_{\mathbb{R}} e^{\imath ux}\mathbb{P}(dx) = \int_{\mathbb{R}} \cos(ux)\mathbb{P}dx + \imath \int_{\mathbb{R}} \sin(ux)\mathbb{P}(dx)$$

die *charakteristische Funktion* von $\mathbb{P}$. □

§31.03 **Lemma**. *Eine charakteristische Funktion φ_X ist gleichmäßig stetig auf $\mathbb{R}$ und erfüllt:*

(i) $|\varphi(u)| \leqslant 1$ *für jedes* $u \in \mathbb{R}$ *und* $\varphi(0) = 1$;

(ii) $\varphi_{aX+b}(u) = e^{\imath bu}\varphi_X(au)$ *für alle* $u, a, b \in \mathbb{R}$;

(iii) $\varphi_{-X}(u) = \overline{\varphi(u)}$ *für alle* $u \in \mathbb{R}$, *sodass* $\mathbb{P}^X = \mathbb{P}^{-X}$ *genau dann gilt, wenn* φ_X *reell ist.*

§31.04 **Beweis** von Lemma §31.03. In der Vorlesung. □

§31.05 **Lemma**. *Es gilt* $\varphi_{\mathbb{P}\star\widetilde{\mathbb{P}}}(u) = \varphi_{\mathbb{P}}(u)\varphi_{\widetilde{\mathbb{P}}}(u)$ *für alle* $u \in \mathbb{R}$, *somit für unabhängige reelle Zufallsvariablen* X, Y *also* $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$ *für alle* $u \in \mathbb{R}$.

§31.06 **Beweis** von Lemma §31.05. In der Vorlesung. □

§31.07 **Beispiel**.

(a) $\varphi_{B_p}(u) = (pe^{\imath u} + 1 - p)$ sowie $\varphi_{\mathrm{Bin}_{(n,p)}}(u) = (pe^{\imath u} + 1 - p)^n$;

(b) $\varphi_{\mathrm{Poi}_\lambda}(u) = \exp(\lambda(e^{\imath u} - 1))$;

(c) $\varphi_{\mathrm{U}_{[-1,1]}}(u) = \sin(u)/u$;

(d) $\varphi_{\mathrm{N}_{(0,1)}}(u) = e^{-u^2/2}$ und $\varphi_{\mathrm{N}_{(\mu,\sigma^2)}}(u) = e^{\imath u\mu}e^{-\sigma^2u^2/2}$. □

§31.08 **Lemma**. *Für alle $t \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt*

$$\left| e^{\imath t} - \sum_{k \in [\![0,n)\!)} \frac{(\imath t)^k}{k!} \right| \leqslant \frac{|t|^n}{n!}.$$

§31.09 **Beweis** von Lemma §31.08. In der Vorlesung. □

§31.10 **Lemma**. *Sei X eine reelle Zufallsvariable in $\mathcal{L}_m$ für ein $m \in \mathbb{N}$. Dann ist φ_X m-mal stetig differenzierbar und für alle $k \in [\![0,m]\!]$ und $u \in \mathbb{R}$ gilt $\varphi^{(k)}(u) = \mathbb{E}((\imath X)^k e^{\imath uX})$.*

§31.11 **Beweis** von Lemma §31.10. In der Vorlesung. □

§31.12 **Eindeutigkeitssatz**. *Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B})$. Für $\mathrm{a} \in \mathbb{R}$ mit $\mathrm{b} \in \mathbb{R}_{>\mathrm{a}}$ gilt die Inversionsformel*

$$\mathbb{P}((\mathrm{a},\mathrm{b})) + \tfrac{1}{2}\mathbb{P}(\{\mathrm{a}\}) + \tfrac{1}{2}\mathbb{P}(\{\mathrm{b}\}) = \tfrac{1}{2\pi} \lim_{t\to\infty} \int_{-t}^{t} \frac{e^{-\imath u\mathrm{a}} - e^{-\imath u\mathrm{b}}}{iu} \varphi_{\mathbb{P}}(u) du.$$

Insbesondere sind zwei Wahrscheinlichkeitsmaße mit derselben charakteristischen Funktion identisch.

§31.13 **Beweis** von Satz §31.12. In der Vorlesung. □

§31.14 **Bemerkung**. Seien X und $(X_n)_{n\in\mathbb{N}}$ reelle Zufallsvariablen mit entsprechenden charakteristischen Funktion φ_X und $(\varphi_{X_n})_{n\in\mathbb{N}}$. Da für jedes $u \in \mathbb{R}$ die Abbildung $x \mapsto e^{\imath ux}$ stetig und beschränkt ist, folgt aus $X_n \xrightarrow{D} X$ auch $\varphi_{X_n}(u) = \mathbb{E}(e^{\imath uX_n}) \xrightarrow{n\to\infty} \mathbb{E}(e^{\imath uX}) = \varphi_X(u)$ für alle $u \in \mathbb{R}$. □

§31.15 **Stetigkeitssatz von Lévy**. *Sei $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathscr{B})$ mit entsprechenden charakteristischen Funktionen $(\varphi_n)_{n\in\mathbb{N}}$. Existiert eine Funktion $h : \mathbb{R} \to \mathbb{C}$, die stetig in Null ist derart, dass $\varphi_n(u) \xrightarrow{n\to\infty} h(u)$ für alle $u \in \mathbb{R}$ (punktweise) gilt, dann ist $h = \varphi_{\mathbb{P}}$, die charakteristische Funktion eines Wahrscheinlichkeitsmaßes $\mathbb{P}$ auf $(\mathbb{R}, \mathscr{B})$, und es gilt $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.*

§31.16 **Beweis** von Satz §31.15. In der Vorlesung. □

§31.17 **Beispiel**.

(a) Sei $(\mathrm{p}_n)_{n\in\mathbb{N}}$ eine Folge in $[0,1]$ mit $n\mathrm{p}_n \xrightarrow{n\to\infty} \lambda \in \mathbb{R}_{>0}$. Mit Beispiel §31.07 (a) und (b) gilt dann punktweise $\varphi_{\mathrm{Bin}_{(n,\mathrm{p}_n)}} \xrightarrow{n\to\infty} \varphi_{\mathrm{Poi}_\lambda}$. Der *Stetigkeitssatz von Lévy* §31.15 impliziert daher den *Poissonschen Grenzwertsatz* §04.09 $\mathrm{Bin}_{(n,\mathrm{p}_n)} \xrightarrow{w} \mathrm{Poi}_\lambda$.

(b) Für $n \in \mathbb{N}$ sei $S_n \sim \mathrm{Bin}_{(n,\mathrm{p})}$ mit $\mathrm{p} \in (0,1)$. Dann ist $S_n^* := \frac{S_n - n\mathrm{p}}{\sqrt{n\mathrm{p}(1-\mathrm{p})}}$ standardisiert, dass heißt, zentriert mit Varianz Eins. Für $\mathrm{q} := 1 - \mathrm{p}$ gilt

$$\varphi_{S_n^*}(u) = \left(\mathrm{p}e^{\imath u\mathrm{q}/\sqrt{n\mathrm{pq}}} + \mathrm{q}e^{-\imath u\mathrm{p}/\sqrt{n\mathrm{pq}}}\right)^n = \left(1 - u^2/(2n) + r_n/n\right)^n,$$

mit $r_n \xrightarrow{n\to\infty} 0$, sodass punktweise $\varphi_{S_n^*} \xrightarrow{n\to\infty} \varphi_{\mathrm{N}_{(0,1)}}$. Mit Satz §31.15 folgt so $S_n^* \xrightarrow{D} \mathrm{N}_{(0,1)}$.

(c) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch $\mathrm{U}_{[-1,1]}$-verteilter Zufallsvariablen. Dann gilt $\mathbb{E}(X_1) = 0$ und $\mathbb{V}\mathrm{ar}(X_1) = 1/3$. Die standardisierte Summe $S_n^* = \sqrt{\frac{3}{n}} \sum_{k\in[\![n]\!]} X_k$ hat die charakteristische Funktion

$$\varphi_{S_n^*}(u) = \left(\frac{\sin(\sqrt{3}u/\sqrt{n})}{\sqrt{3}u/\sqrt{n}}\right)^n = \left(1 - u^2/(2n) + r_n/n\right)^n$$

mit $r_n \xrightarrow{n\to\infty} 0$. Es folgt wiederum $S_n^* \xrightarrow{D} \mathrm{N}_{(0,1)}$. □

§32 Zentrale Grenzwertsätze

§32.01 **Erinnerung.** Wie in Beispiel §04.08 (c) bezeichnet Φ die Verteilungsfunktion einer Standarnormalverteilung $\mathrm{N}_{(0,1)}$. □

§32.02 **Zentraler Grenzwertsatz.** *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter (u.i.v.) reeller Zufallsvariablen in $\mathcal{L}_2$ mit $\mu := \mathbb{E}(X_1)$ und $\sigma^2 := \mathbb{V}\mathrm{ar}(X_1) \in \mathbb{R}_{>0}$. Dann erfüllt die standardisierte Summe*

$$S_n^* := \tfrac{1}{\sqrt{n}} \sum_{i\in[\![n]\!]} \frac{X_i - \mu}{\sigma} \xrightarrow{D} \mathrm{N}_{(0,1)}.$$

Insbesondere gilt für $\mathrm{a} \in \mathbb{R}$ *mit* $\mathrm{b} \in \mathbb{R}_{>\mathrm{a}}$ *also* $\mathbb{P}(\mathrm{a} < S_n^* \leqslant \mathrm{b}) \xrightarrow{n\to\infty} \Phi(\mathrm{b}) - \Phi(\mathrm{a})$.

§32.03 **Beweis** von Satz §32.02. In der Vorlesung. □

§32.04 **Bemerkung.** Mit Hilfe von Korollar §30.17 gilt unter den Voraussetzungen des zentralen Grenzwertsatz $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} \mathrm{N}_{(0,\sigma^2)}$. □

§32.05 **Definition.**

(a) Für jedes $n \in \mathbb{N}$ seien $(X_{n,j})_{j\in[\![n]\!]}$ reelle Zufallsvariablen in $\mathcal{L}_2$. Wir nennen die Familie $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ ein *standardisiertes Dreiecksschema*, wenn folgende Bedingungen für jedes $n \in \mathbb{N}$ erfüllt sind:

(sD1) Die Familie $(X_{n,j})_{j\in[\![n]\!]}$ ist unabhängig;

(sD2) $\mathbb{E}(X_{n,j}) = 0$ für jedes $j \in [\![n]\!]$ und $\sum_{j\in[\![n]\!]} \mathbb{V}\mathrm{ar}(X_{n,j}) = 1$.

(b) Ein standardisiertes Dreiecksschema $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ erfüllt die *Lindeberg-Bedingung*, wenn

$$\lim_{n\to\infty} \sum_{j\in[\![n]\!]} \mathbb{E}(X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}|\geqslant\delta\}}) = 0$$

für jedes $\delta \in \mathbb{R}_{>0}$ gilt.

(c) Ein standardisiertes Dreiecksschema $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ erfüllt die *Lyapunov-Bedingung*, wenn

$$\lim_{n\to\infty} \sum_{j\in[\![n]\!]} \mathbb{E}(|X_{n,j}|^{2+\delta}) = 0.$$

für ein $\delta \in \mathbb{R}_{>0}$ gilt. □

§32.06 **Beispiel.**

(a) Sei $(Y_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter (u.i.v.) reeller Zufallsvariablen in $\mathcal{L}_2$. Setzen wir $\mu := \mathbb{E}(Y_1)$, $\sigma^2 := \mathbb{V}\mathrm{ar}(Y_1) \in \mathbb{R}_{>0}$ und $X_{n,j} := (Y_j - \mu)/(\sigma\sqrt{n})$ für $n \in \mathbb{N}$ und $j \in [\![n]\!]$, so ist $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung genügt, da für alle $\delta \in \mathbb{R}_{>0}$

$$\sum_{j\in[\![n]\!]} \mathbb{E}(X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}|\geqslant\delta\}}) = \sigma^{-2}\mathbb{E}((Y_1-\mu)^2\mathbb{1}_{\{|Y_1-\mu|\geqslant\delta\sigma\sqrt{n}\}}) \xrightarrow{n\to\infty} 0.$$

(b) Sei $(Y_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und standardisierter reeller Zufallsvariablen, die für ein $\delta \in \mathbb{R}_{>0}$ in $\mathcal{L}_{2+\delta}$ beschränkt ist, d.h. $\sup_{n\in\mathbb{N}} \mathbb{E}(|Y_n|^{2+\delta}) \in \mathbb{R}_{\geqslant 0}$. Setzen wir $X_{n,j} := Y_j/\sqrt{n}$ für $n \in \mathbb{N}$ und $j \in [\![n]\!]$, so ist $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ ein standardisiertes Dreiecksschema, das der Lyapunov-Bedingung genügt:

$$\sum_{j\in[\![n]\!]} \mathbb{E}(|X_{n,j}|^{2+\delta}) = n^{-1-\delta/2}\sum_{j\in[\![n]\!]} \mathbb{E}(|Y_j|^{2+\delta}) \leqslant n^{-\delta/2}\sup_{j\in\mathbb{N}}\mathbb{E}(|Y_j|^{2+\delta}) \xrightarrow{n\to\infty} 0.$$

(c) Wie in Beispiel §29.25 sei $(Y_n)_{n\in\mathbb{N}}$ mit $\mathbb{P}(Y_n = -1/n) = 1/2 = \mathbb{P}(Y_n = 1/n)$ für $n \in \mathbb{N}$ ein Folge unabhängiger Zufallsvariablen. Wir setzen $\sigma_n^2 := \sum_{k\in[\![n]\!]} \mathbb{V}\mathrm{ar}(Y_k)$ und $X_{n,j} := Y_j/\sigma_n$ für $j \in [\![n]\!]$ und $n \in \mathbb{N}$. Dann gilt $\sigma_n^2 = \sum_{k\in[\![n]\!]} k^{-2} \uparrow \sigma_\infty^2 = \pi^2/6$ und $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ ist ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung nicht genügt, da für alle $\delta \in (0, \sigma_\infty^{-1})$, also $\delta\sigma_n \leqslant \sigma_n/\sigma_\infty \leqslant 1$, gilt:

$$\sum_{j\in[\![n]\!]} \mathbb{E}(X_{n,j}^2\mathbb{1}_{\{|X_{n,j}|\geqslant\delta\}}) = \sum_{j\in[\![n]\!]} \frac{1}{j^2\sigma_n^2}\mathbb{1}_{\{1\geqslant j\sigma_n\delta\}} \geqslant \frac{1}{\sigma_n^2} \geqslant \frac{1}{\sigma_\infty^2} > 0$$

□

§32.07 **Lemma**. *Ein standardisiertes Dreiecksschema, das der Lyapunov-Bedingung genügt, erfüllt auch die Lindeberg-Bedingung.*

§32.08 **Beweis** von Lemma §32.07. In der Vorlesung. □

§32.09 **Zentraler Grenzwertsatz nach Lindeberg (1922)**. *Sei $(X_{n,j})_{j\in[\![n]\!],n\in\mathbb{N}}$ ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung genügt, so gilt für (die Zeilensumme)* $S_n^* = \sum_{j\in[\![n]\!]} X_{n,j} \xrightarrow{D} \mathrm{N}_{(0,1)}$.

§32.10 **Beweis** von Satz §32.09. In der Vorlesung. □

§32.11 **Definition**. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter reeller Zufallsvariablen. Für jedes $n \in \mathbb{N}$ heißt das Wahrscheinlichkeitsmaß

$$\boxed{\widehat{\mathbb{P}}_n} := \tfrac{1}{n}\sum_{i\in[\![n]\!]} \delta_{X_i}$$

empirische Verteilung oder *empirisches Wahrscheinlichkeitsmaß,* sowie die entsprechende Verteilungsfunktion

$$\mathbb{R} \ni x \mapsto \boxed{\widehat{\mathbb{F}}_n(x)} := \widehat{\mathbb{P}}_n(\mathbb{R}_{\leqslant x}) = \tfrac{1}{n}\sum_{i\in[\![n]\!]} \delta_{X_i}(\mathbb{R}_{\leqslant x}) = \tfrac{1}{n}\sum_{i\in[\![n]\!]} \mathbb{1}_{\mathbb{R}_{\leqslant x}}(X_i) \in [0,1],$$

empirische Verteilungsfunktion. □

§32.12 **Satz**. *Für alle $x \in \mathbb{R}$ gilt* $\lim_{n\to\infty} \widehat{\mathbb{F}}_n(x) = \mathbb{F}(x) = \mathbb{P}(X_1 \leqslant x)$ $\mathbb{P}$-f.s. *Für alle $x \in \mathbb{R}$ mit $\mathbb{F}(x) \in (0,1)$ gilt*

$$\sqrt{n}(\widehat{\mathbb{F}}_n(x) - \mathbb{F}(x)) \xrightarrow{D} \mathrm{N}_{(0,\mathbb{F}(x)(1-\mathbb{F}(x)))}.$$

§32.13 **Beweis** von Satz §32.12. Übungsaufgabe. □

§32.14 **Satz von Glivenko-Cantelli**. *Die empirische Verteilungsfunktion konvergiert gleichmäßig fast sicher gegen die wahre Verteilungsfunktion:*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\widehat{\mathbb{F}}_n(x) - \mathbb{F}(x)| = 0 \quad \mathbb{P}\text{-f.s.}$$

§32.15 **Beweis** von Satz §32.14. In der Vorlesung. □

§33 Statistische Inferenz: asymptotische Eigenschaften

Sei $\mathbb{P}_\Theta$ eine Familie von Wahrscheinlichkeitsmaßen auf einem messbaren Raum $(\mathcal{X}, \mathscr{X})$. Im Folgenden betrachten wir eine Folge $(X_i)_{i\in\mathbb{N}}$ unabhängiger und identisch-verteilter $\mathcal{X}$-wertiger Zufallsvariablen mit identischer Randverteilung aus $\mathbb{P}_\Theta$. Für jedes $n \in \mathbb{N}$ ist die Zufallsvariable $X^{(n)} := (X_i)_{i\in[\![n]\!]}$ also adäquat durch das statistische Produktexperiment $(\mathcal{X}^n, \mathscr{X}^n, \mathbb{P}^n_\Theta)$ beschrieben. Wir schreiben kurz $(X_i)_{i\in\mathbb{N}} \circledsim \mathbb{P}^{\mathbb{N}}_\Theta$. Für jedes $\theta \in \Theta$ bezeichnet im Folgenden $\mathbb{E}^n_\theta$ stets die Erwartung bzgl. $\mathbb{P}^n_\theta$. Weiterhin sei $\gamma : \Theta \to \Gamma \subseteq \mathbb{R}$ ein identifizierbarer abgeleiteter Parameter und für jedes $n \in \mathbb{N}$ sei $\widehat{\gamma}_n : \mathcal{X}^n \to \Gamma$ ein Schätzer für γ, also $\widehat{\gamma}_n \in \mathcal{M}(\mathscr{X}^n)$.

§33|01 Konsistenz und asymptotische Verteilung eines Schätzers

§33.01 **Definition**. Die Folge von Schätzern $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ für γ heißt

(a) *(schwach) konsistent*, wenn $\widehat{\gamma}_n \xrightarrow{\mathbb{P}_\theta} \gamma(\theta)$ für jedes $\theta \in \Theta$ gilt;

(b) *stark konsistent*, wenn $\widehat{\gamma}_n \xrightarrow{\mathbb{P}_\theta\text{-f.s.}} \gamma(\theta)$ für jedes $\theta \in \Theta$ gilt;

(c) *$\mathcal{L}_k$-konsistent*, wenn $\widehat{\gamma}_n \xrightarrow{\mathcal{L}_k(\mathbb{P}_\theta)} \gamma(\theta)$ für jedes $\theta \in \Theta$ gilt.

Ist $\widehat{\gamma}_n \in \mathcal{L}_1(\mathbb{P}^n_\theta)$ für alle $\theta \in \Theta$ und $n \in \mathbb{N}$, so heißt $\widehat{\gamma}_n$ *asymptotisch erwartungstreu (unverfälscht)* wenn $\mathrm{Bias}_\theta(\widehat{\gamma}_n) = \mathbb{E}^n_\theta(\widehat{\gamma}_n) - \gamma(\theta) \xrightarrow{n\to\infty} 0$ für jedes $\theta \in \Theta$ gilt. □

§33.02 **Sprechweise**. Zum Beispiel meint *„$\widehat{\gamma}_n$ ist konsistent“* stets, dass die Folge von Schätzern $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ konsistent ist. □

§33.03 **Beispiel**. Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge unabhängiger und identisch verteilter reeller Zufallsvariablen mit endlichem k-ten Moment, also $(X_i)_{i\in\mathbb{N}} \circledsim \mathcal{W}^{\mathbb{N}}_k(\mathscr{B})$. Für $l \in [\![k]\!]$ sind dann das *empirische l-te Moment*

$$\boxed{\widehat{\mathrm{m}}^{(l)}_n} = \tfrac{1}{n} \sum_{i\in[\![n]\!]} X_i^l$$

und das *zentrierte empirische l-te Moment*

$$\boxed{\widehat{\mathrm{M}}^{(l)}_n} = \tfrac{1}{n} \sum_{i\in[\![n]\!]} (X_i - \mathrm{m}^{(1)})^l,$$

unter der unrealistischen Anahme, dass $\mathrm{m}^{(1)}$ bekannt ist, *stark konsistente Schätzer* für das l-te Moment $\mathrm{m}^{(l)}$ bzw. das zentrierte l-te Moment $\mathrm{M}^{(l)}$. In der Tat, für $(X_i)_{i\in\mathbb{N}} \sim \mathbb{P}^{\mathbb{N}}$ mit $\mathbb{P} \in \mathcal{W}_k(\mathscr{B})$ sind $(X_i^l)_{i\in\mathbb{N}}$ und $(Y_i^l)_{i\in\mathbb{N}}$ mit $Y_i := X_i - \mathrm{m}^{(1)}(\mathbb{P})$ Folgen unabhängiger und identisch verteilter reeller

Zufallsvariablen in $\mathcal{L}_1(\mathbb{P})$, so dass mit dem *starken Gesetz der großen Zahlen* §29.11 gilt $\widehat{\mathrm{m}}_n^{(l)} = \overline{X_n^l} \xrightarrow{\mathbb{P}\text{-f.s.}} \mathrm{m}^{(l)}(\mathbb{P})$ bzw. $\widehat{\mathrm{M}}_n^{(l)} = \overline{Y_n^l} \xrightarrow{\mathbb{P}\text{-f.s.}} \mathrm{M}^{(l)}(\mathbb{P})$. Ist $\mathrm{m}^{(1)}$ nicht bekannt, so ist der *Momentenschätzer*

$$\boxed{\widehat{\mathrm{V}}_n^{(l)}} = \tfrac{1}{n} \sum_{i \in [\![n]\!]} (X_i - \widehat{\mathrm{m}}_n^{(1)})^l$$

ein *stark konsistenter Schätzer* für das zentrierte l-te Moment $\mathrm{M}^{(l)}$. In der Tat, für $(X_i)_{i\in\mathbb{N}} \sim \mathbb{P}^{\mathbb{N}}$ mit $\mathbb{P} \in \mathcal{W}_k(\mathscr{B})$ gilt

$$\widehat{\mathrm{V}}_n^{(l)} = \widehat{\mathrm{M}}_n^{(l)} + \sum_{j\in[\![l]\!]} \binom{l}{j} (\mathrm{m}^{(1)}(\mathbb{P}) - \widehat{\mathrm{m}}_n^{(1)})^j \tfrac{1}{n} \sum_{i\in[\![n]\!]} Y_i^{l-j},$$

wobei für jedes $j \in [\![l]\!]$ gilt $(\mathrm{m}^{(1)}(\mathbb{P}) - \widehat{\mathrm{m}}_n^{(1)})^j \xrightarrow{\mathbb{P}\text{-f.s.}} 0$ und $\frac{1}{n}\sum_{i\in[\![n]\!]} Y_i^{l-j} \xrightarrow{\mathbb{P}\text{-f.s.}} \mathrm{M}^{(l-j)}(\mathbb{P})$. Mit dem *Satz von der Stetigen Abbildung* §28.20 folgt damit auch $\widehat{\mathrm{V}}_n^{(l)} \xrightarrow{\mathbb{P}\text{-f.s.}} \mathrm{M}^{(l)}(\mathbb{P})$. □

§33.04 **Beispiel**. Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{U}_{[0,\theta]}^{\mathbb{N}})_{\theta\in\mathbb{R}_{>0}}$. Der *MLS* $\widehat{\theta}_n = \max_{i\in[\![n]\!]} X_i$ unterschätzt im Mittel θ mit $\mathrm{Bias}_\theta(\widehat{\theta}_n) = -\theta/(n+1)$ (vgl. Beispiel §27.17). Da $\mathrm{Bias}_\theta(\widehat{\theta}_n) \xrightarrow{n\to\infty} 0$ ist $\widehat{\theta}_n$ *asymptotisch erwartungstreu*. Da für den mittleren quadratischen Fehler $\mathrm{MSE}_\theta(\widehat{\theta}_n) = \frac{2\theta^2}{(n+2)(n+1)} \xrightarrow{n\to\infty} 0$ für jedes $\theta \in \Theta$ (vgl. Beispiel §27.19 (a)) gilt, ist der MLS $\widehat{\theta}_n$ $\mathcal{L}_2$-konsistent und somit auch (schwach) konsistent ist. □

§33.05 **Bemerkung**. Für eine Folge von Schätzern $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ für γ in $\mathcal{L}_2(\mathbb{P}_\theta)$ gilt mit Hilfe der Bias^2-Varianz-Zerlegung (vgl. Satz §25.06)

$$\mathrm{MSE}_\theta(\widehat{\gamma}_n) = \mathbb{E}_\theta^n((\widehat{\gamma}_n - \gamma(\theta))^2) = \mathbb{V}\mathrm{ar}_\theta(\widehat{\gamma}_n) + (\mathrm{Bias}_\theta(\widehat{\gamma}_n))^2.$$

Damit ist $\widehat{\gamma}_n$ genau dann *$\mathcal{L}_2$-konsistent*, wenn $\mathbb{V}\mathrm{ar}_\theta(\widehat{\gamma}_n) \xrightarrow{n\to\infty} 0$ gilt und $\widehat{\gamma}_n$ *asymptotisch unverfälscht* ist. □

Zur Erinnerung, für jedes $n \in \mathbb{N}$ und $\theta \in \Theta$ ist $\widehat{\gamma}_n \in \mathcal{M}(\mathscr{X}^n)$ eine reelle Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\mathcal{X}^n, \mathscr{X}^n, \mathbb{P}_\theta^n)$. Eine Konsistenzaussage für einen Schätzer $\widehat{\gamma}_n$ quantifiziert nicht, wie schnell der Schätzer konvergiert. In der nächsten Definition wird dies mit Hilfe von Definition §30.02 formalisiert.

§33.06 **Definition**. Sei $(\mathrm{d}_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{>0}$ mit $\mathrm{d}_n \uparrow \infty$. Die Folge von Schätzern $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ für γ heißt *d_n-konsistent*, wenn für jedes $\theta \in \Theta$ die Folge $(\mathrm{d}_n(\widehat{\gamma}_n - \gamma(\theta)))_{n\in\mathbb{N}}$ in Verteilung unter $\mathbb{P}_\theta^n$ konvergiert. Existiert weiterhin ein Wahrscheinlichkeitsmaß $\mathbb{P} \in \mathcal{W}(\mathscr{B})$ und ein identifizierbarer abgeleiteter Parameter $\tau : \Theta \to \mathbb{R}_{\neq 0}$, genannt *Störparameter*, derart, dass für jedes $\theta \in \Theta$ gilt $\frac{\mathrm{d}_n}{\tau(\theta)}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{D} \mathbb{P}$ *unter* $\mathbb{P}_\theta^n$, so nennen wir $\mathbb{P}$ *asymptotische Verteilung* von $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$. □

§33.07 **Bemerkung**. Ist $\widehat{\gamma}_n$ ein d_n-konsistenter Schätzer für γ, dann gilt für jede Folge $(\mathrm{c}_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{>0}$ mit $\mathrm{c}_n/\mathrm{d}_n \xrightarrow{n\to\infty} 0$, dass $\mathrm{c}_n(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{D} 0$ und nach Korollar §30.12 somit $\mathrm{c}_n(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{\mathbb{P}_\theta} 0$. Insbesondere ist $\widehat{\gamma}_n$ somit schwach konsistent für γ. □

§33.08 **Beispiel**. Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{U}_{[0,\theta]}^{\mathbb{N}})_{\theta\in\mathbb{R}_{>0}}$. Der MLS $\widehat{\theta}_n = \max_{i\in[\![n]\!]} X_i$ ist *n-konsistent* mit Störparameter θ und asymptotischer Exp_1-Verteilung, da $\frac{n}{\theta}(\theta - \widehat{\theta}_n) \xrightarrow{D} \mathrm{Exp}_1$ unter $\mathrm{U}_{[0,\theta]}^n$ für jedes $\theta \in \mathbb{R}_{>0}$ gilt (vgl. Beispiel §30.24). □

§33.09 **Beispiel**. Sei $(X_i)_{i\in\mathbb{N}} \odot \mathcal{W}_{2k}^{\mathbb{N}}(\mathscr{B})$. Für $l \in [\![k]\!]$ ist unter Verwendung des *zentralen Grenzwertsatzes* §32.02 das *empirische l-te Moment* $\boxed{\widehat{\mathrm{m}}_n^{(l)}} = \frac{1}{n}\sum_{i\in[\![n]\!]} X_i^l$ ein *$\sqrt{n}$-konsistenter* Schätzer für das l-te Moment $\mathrm{m}^{(l)}$ mit Störparameter

$$\sigma_l := \sqrt{\mathrm{m}^{(2l)} - (\mathrm{m}^{(l)})^2} : \mathcal{W}_{2k}(\mathscr{B}) \ni \mathbb{P} \mapsto \sigma_l(\mathbb{P}) := \sqrt{\mathrm{m}^{(2l)}(\mathbb{P}) - (\mathrm{m}^{(l)}(\mathbb{P}))^2} \in \mathbb{R}_{\geqslant 0}$$

und asymptotischer Standardnormalverteilung. Ist $\mathrm{m}^{(1)}$ bekannt, so folgt aus dem zentralen Grenzwertsatz, dass das *zentrierte empirische l-te Moment*

$$\boxed{\widehat{\mathrm{M}}_n^{(l)}} = \frac{1}{n}\sum_{i\in[\![n]\!]} (X_i - \mathrm{m}^{(1)})^l$$

ein *$\sqrt{n}$-konsistenter* Schätzer für das l-te zentrierte Moment $\mathrm{M}^{(l)}$ mit Störparameter

$$\sigma_l := \sqrt{\mathrm{M}^{(2l)} - (\mathrm{M}^{(l)})^2} : \mathcal{W}_{2k}(\mathscr{B}) \ni \mathbb{P} \mapsto \sigma_l(\mathbb{P}) := \sqrt{\mathrm{M}^{(2l)}(\mathbb{P}) - (\mathrm{M}^{(l)}(\mathbb{P}))^2} \in \mathbb{R}_{\geqslant 0}$$

und asymptotischer Standardnormalverteilung. Im realistischen Fall ist $\mathrm{m}^{(1)}$ unbekannt und wir betrachten

$$\boxed{\widehat{\mathrm{V}}_n^{(l)}} = \frac{1}{n}\sum_{i\in[\![n]\!]} (X_i - \widehat{\mathrm{m}}_n^{(1)})^l.$$

Im Spezialfall $l = 2$ gilt $\widehat{\mathrm{V}}_n^{(2)} = \widehat{\mathrm{M}}_n^{(2)} - (\widehat{\mathrm{m}}_n^{(1)} - \mathrm{m}^{(1)})^2$. Da $\sqrt{n}(\widehat{\mathrm{m}}_n^{(1)} - \mathrm{m}^{(1)})$ in Verteilung konvergiert, folgt mit dem *Satz von der stetigen Abbildung* §30.15, dass $\sqrt{n}(\widehat{\mathrm{m}}_n^{(1)} - \mathrm{m}^{(1)})^2 \xrightarrow{D} 0$ und somit auch stochastisch gegen Null konvergiert (vgl. Korollar §30.12). Mit Hilfe des *Satzes von Slutzky* §30.07 schließen wir, dass $\widehat{\mathrm{V}}_n^{(2)}$, wie $\widehat{\mathrm{M}}_n^{(2)}$, ein *$\sqrt{n}$-konsistenter* Schätzer für das 2-te zentrierte Moment $\mathrm{M}^{(2)}$, die Varianz, mit Störparameter $\sigma_2 := \sqrt{\mathrm{M}^{(4)} - (\mathrm{M}^{(2)})^2}$ und asymptotischer Standardnormalverteilung ist. Abschliessend halten wir fest, dass damit auch die *empirische Varianz* $\boxed{\widehat{\mathrm{S}}_n^{(2)}} = \frac{n}{n-1}\widehat{\mathrm{V}}_n^{(2)}$ ein *$\sqrt{n}$-konsistenter* Schätzer für die Varianz, mit Störparameter $\sigma_2 := \sqrt{\mathrm{M}^{(4)} - (\mathrm{M}^{(2)})^2}$ und asymptotischer Standardnormalverteilung ist. □

§33|02 Asymptotischer 1-α-Konfidenzbereich

§33.10 **Definition**. Seien $\alpha \in (0,1)$, $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ eine Familie von Partitionen der Menge der interessierenden Parameterwerte und B^n eine Bereichsschätzfunktion auf $(\mathcal{X}^n, \mathscr{X}^n)$ für jedes $n \in \mathbb{N}$. Die Folge von Bereichsfunktionen $(\mathrm{B}^n)_{n\in\mathbb{N}}$ hält für $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ das *Niveau 1-α asymptotisch* ein, wenn $\liminf_{n\to\infty} \mathbb{P}_\theta^n(\widetilde{\gamma} \in \mathrm{B}^n) \geqslant 1-\alpha$ für alle $\widetilde{\gamma} \in \Gamma_{\gamma(\theta)}^{\mathrm{r}}$ und für alle $\theta \in \Theta$ gilt. □

§33.11 **Bemerkung**. Ist B^n für jedes $n \in \mathbb{N}$ ein 1-α-Konfidenzbereich, so hält die Folge $(\mathrm{B}^n)_{n\in\mathbb{N}}$ auch asymptotisch das Niveau 1-α ein. Wir sagen B^n ist ein *asymptotischer 1-α-Konfidenzbereich*, wenn die Folge $(\mathrm{B}^n)_{n\in\mathbb{N}}$ asymptotisch das Niveau 1-α einhält. □

§33.12 **Beispiel** *(Beispiel §33.08 fortgesetzt)*. Für $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{U}_{[0,\theta]}^{\mathbb{N}})_{\theta\in\mathbb{R}_{>0}}$ ist der MLS $\widehat{\theta}_n = \max_{i\in[\![n]\!]} X_i$ *n-konsistent* mit Störparameter θ und stetiger asymptotischer Verteilung Exp_1. Da $\widehat{\theta}_n$ schwach konsistent für θ ist, gilt mit dem *Satz von Slutzky* §30.07 auch $\frac{n}{\widehat{\theta}_n}(\theta - \widehat{\theta}_n) \xrightarrow{D} \mathrm{Exp}_1$ *unter* $\mathrm{U}_{[0,\theta]}^n$. Sei $\alpha \in (0,1)$ und $\mathrm{q}_\alpha \in \mathbb{R}_{>0}$ das α-Quantil der stetigen Verteilung Exp_1. Für jedes $\widetilde{\theta} \in \mathbb{R}_{\geqslant\theta}$, also $\widetilde{\theta} - \theta \in \mathbb{R}_{\geqslant 0}$, gilt dann

$$\begin{aligned}\mathrm{U}_{[0,\theta]}^n(\widetilde{\theta} \in [\widehat{\theta}_n + \tfrac{\widehat{\theta}_n}{n}\mathrm{q}_\alpha, \infty)) &= \mathrm{U}_{[0,\theta]}^n(\widetilde{\theta} - \theta - \tfrac{\widehat{\theta}_n}{n}\mathrm{q}_\alpha \geqslant \widehat{\theta}_n - \theta) \\ &\geqslant \mathrm{U}_{[0,\theta]}^n(-\mathrm{q}_\alpha \geqslant \tfrac{n}{\widehat{\theta}_n}(\widehat{\theta}_n - \theta)) \xrightarrow{n\to\infty} \mathrm{Exp}_1(\mathbb{R}_{\geqslant \mathrm{q}_\alpha}) = 1-\alpha,\end{aligned}$$

so dass $[\widehat{\theta}_n + \frac{\widehat{\theta}_n}{n}\mathrm{q}_\alpha, \infty)$ ein asymptotischer 1-α-Konfidenzbereich für die Mengen der richtigen und falschen Parameter $\Gamma_\theta^{\mathrm{r}} = \mathbb{R}_{\geqslant\theta}$ bzw. $\Gamma_\theta^{\mathrm{f}} = (0,\theta)$ ist. Analog zeigen wir, dass $(0, \widehat{\theta}_n + \frac{\widehat{\theta}_n}{n}\mathrm{q}_{1-\alpha}]$ und $[\widehat{\theta}_n + \frac{\widehat{\theta}_n}{n}\mathrm{q}_{\alpha/2}, \widehat{\theta}_n + \frac{\widehat{\theta}_n}{n}\mathrm{q}_{1-\alpha/2}]$ asymptotische 1-α-Konfidenzbereiche für die Mengen der richtigen Parameter $(0,\theta]$ bzw. $\{\theta\}$ mit $\theta \in \mathbb{R}_{>0}$ sind. □

§33.13 **Lemma**. *Sei $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ eine d_n-konsistente Folge von Schätzern für den abgeleiteten Parameter $\gamma : \Theta \to \mathbb{R}$ mit Störparameter $\tau : \Theta \to \mathbb{R}_{\neq 0}$ und stetiger asymptotischer Verteilung $\mathbb{P} \in \mathcal{W}(\mathscr{B})$, das heißt, $\frac{\mathrm{d}_n}{\tau(\theta)}(\widehat{\gamma}_n - \gamma(\theta)) \xrightarrow{D} \mathbb{P}$ unter $\mathbb{P}_\theta^n$. Für $\alpha \in (0,1)$ bezeichne $\mathrm{q}_\alpha \in \mathbb{R}$ das α-Quantil der stetigen Verteilung $\mathbb{P}$. Ist $\widehat{\tau}_n$ ein schwach konsister Schätzer für τ, dann sind*

$$[\widehat{\gamma}_n - \tfrac{\widehat{\tau}_n}{\mathrm{d}_n}\mathrm{q}_{1-\alpha}, \infty), \quad (-\infty, \widehat{\gamma}_n - \tfrac{\widehat{\tau}_n}{\mathrm{d}_n}\mathrm{q}_\alpha] \quad \textit{und} \quad [\widehat{\gamma}_n - \tfrac{\widehat{\tau}_n}{\mathrm{d}_n}\mathrm{q}_{1-\alpha/2}, \widehat{\gamma}_n - \tfrac{\widehat{\tau}_n}{\mathrm{d}_n}\mathrm{q}_{\alpha/2}]$$

asymptotische 1-α-Konfidenzbereiche für die Mengen der richtigen Parameter $\mathbb{R}_{\geqslant\gamma}$, $\mathbb{R}_{\leqslant\gamma}$ bzw. $\{\gamma\}$ mit $\gamma \in \Gamma$.

§33.14 **Beweis** von Lemma §33.13. In der Vorlesung. □

§33.15 **Beispiel** *(Beispiel §33.09 fortgesetzt)*. Sei $(X_i)_{i\in\mathbb{N}} \circleddash \mathcal{W}_4^{\mathbb{N}}(\mathscr{B})$ und die Varianz

$$\sigma^2 : \mathcal{W}_4(\mathscr{B}) \ni \mathbb{P} \mapsto \sigma^2(\mathbb{P}) = \mathrm{M}^{(2)}(\mathbb{P}) \in \mathbb{R}_{\geqslant 0}$$

der interessierende Parameter. Dann ist die empirische Varianz $\widehat{\mathrm{S}}_n^{(2)}$ ein *$\sqrt{n}$-konsistenter* Schätzer für σ^2 mit Störparameter $\tau := \sqrt{\mathrm{M}^{(4)} - (\mathrm{M}^{(2)})^2}$ und asymptotischer Standardnormalverteilung (vgl. Beispiel §33.09). In Beispiel §33.03 haben wir gezeigt, dass $\widehat{\mathrm{V}}_n^{(4)}$ und $\widehat{\mathrm{S}}_n^{(2)}$ stark konsistente Schätzer für $\mathrm{M}^{(4)}$ bzw. $\mathrm{M}^{(2)}$ sind. Mit Hilfe des *Satzes von der Stetigen Abbildung* §28.20 ist damit $\widehat{\tau} := \sqrt{\widehat{\mathrm{V}}_n^{(4)} - (\widehat{\mathrm{S}}_n^{(2)})^2}$ ein (stark) konsistenter Schätzer für den Störparameter τ. Für $\alpha \in (0,1)$ und α-Quantil $z_\alpha \in \mathbb{R}$ der Standardnormalverteilung sind nach Lemma §33.13

$$[\widehat{\mathrm{S}}_n^{(2)} - \tfrac{\widehat{\tau}_n}{\sqrt{n}}z_{1-\alpha}, \infty), \quad (-\infty, \widehat{\mathrm{S}}_n^{(2)} + \tfrac{\widehat{\tau}_n}{\sqrt{n}}z_{1-\alpha}] \quad \text{und} \quad [\widehat{\mathrm{S}}_n^{(2)} \pm \tfrac{\widehat{\tau}_n}{\sqrt{n}}z_{1-\alpha/2}]$$

somit *asymptotische 1-α-Konfidenzbereiche* für die Mengen der richtigen Parameter $[\sigma^2, \infty)$, $(0,\sigma^2]$ bzw. $\{\sigma^2\}$ mit $\sigma^2 \in \mathbb{R}_{\geqslant 0}$. □

§33|03 Asymptotischer α-Test

§33.16 **Definition**. Sei $\{\Gamma^0, \Gamma^1\}$ eine Partition der Menge der interessierenden Parameterwerte Γ, und $\varphi^n : \mathcal{X} \to \{0,1\}$ für jedes $n \in \mathbb{N}$ ein Test mit Gütefunktion $\beta_{\varphi^n}(\theta) := \mathbb{P}_\theta^n(\varphi^n)$, $\theta \in \Theta$. Für das Testproblem der Nullhypothese $H_0 : \Gamma^0$ gegen die Alternativhypothese $H_1 : \Gamma^1$ hält die Folge von Tests $(\varphi^n)_{n\in\mathbb{N}}$ *asymptotisch* das *Signifikanz-Niveau* $\alpha \in (0,1)$ ein, wenn $\limsup_{n\to\infty} \beta_{\varphi^n}(\theta) \leqslant \alpha$ für alle $\theta \in \Theta$ mit $\gamma(\theta) \in \Gamma^0$ gilt. □

§33.17 **Bemerkung**. Ist φ^n für jedes $n \in \mathbb{N}$ ein α-Test, so hält die Folge von Test $(\varphi^n)_{n\in\mathbb{N}}$ auch asymptotisch das Signifikanz-Niveau α ein. Wir sagen φ^n ist ein *asymptotischer α-Test*, wenn die Folge $(\varphi^n)_{n\in\mathbb{N}}$ asymptotisch das Niveau α einhält. □

§33.18 **Erinnerung**.

(a) Für $\gamma \in \Gamma$ sei $\Gamma_\gamma^{\mathrm{r}}$ und $\Gamma_\gamma^{\mathrm{f}}$ eine Partition in richtige und falsche interessierende Parameterwerte, dann bezeichnen $\Gamma_\gamma^0 = \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma_{\widetilde{\gamma}}^{\mathrm{r}}\}$ und $\Gamma_\gamma^1 = \{\widetilde{\gamma} \in \Gamma : \gamma \in \Gamma_{\widetilde{\gamma}}^{\mathrm{f}}\}$ die *assozierte Nullhypothese bzw. Alternative* der interessierenden Parameter.

(b) Für $n \in \mathbb{N}$ sei B^n eine Bereichsschätzfunktion, dann bezeichnet $(\varphi_\gamma^n)_{\gamma\in\Gamma}$ die *assoziierte Familie von Tests* mit Ablehnbereich $\{\varphi_\gamma^n = 1\} = \{\gamma \notin \mathrm{B}^n(X^{(n)})\}$.

(c) Für $n \in \mathbb{N}$ sei $(\varphi_\gamma^n)_{\gamma\in\Gamma}$ einer Familie von Tests, dann bezeichnet B^n die *assoziierte Bereichsschätzfunktion* mit $\mathrm{B}^n(x) := \{\gamma \in \Gamma : \varphi_\gamma^n(x) = 0\}$. □

§33.19 **Lemma**. *Seien $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$ eine Familie von richtigen und falschen interessierenden Parameterwerte und $(\{\Gamma_\gamma^0, \Gamma_\gamma^1\})_{\gamma\in\Gamma}$ die assozierte Familie von Nullhypothesen und Alternativen. Dann gilt für eine Familie $((\varphi_\gamma^n)_{n\in\mathbb{N}})_{\gamma\in\Gamma}$ von Testfolgen, dass φ_γ^n ein ein asymptotischer α-Test der Nullhypothese $H_0 : \Gamma_\gamma^0$ gegen die Alternative $H_1 : \Gamma_\gamma^1$ für jedes $\gamma \in \Gamma$ genau dann ist, wenn die assozierte Folge von Bereichsfunktionen $(\mathrm{B}^n)_{n\in\mathbb{N}}$ für $(\{\Gamma_\gamma^{\mathrm{r}}, \Gamma_\gamma^{\mathrm{f}}\})_{\gamma\in\Gamma}$) ein asymptotischer 1-α-Konfidenzbereich ist.*

§33.20 **Beweis** von Lemma §33.19. In der Vorlesung. □

§33.21 **Beispiel** *(Beispiel §33.12 fortgesetzt).* Sei $(X_i)_{i\in\mathbb{N}} \circledcirc (\mathrm{U}_{[0,\theta]}^{\mathbb{N}})_{\theta\in\mathbb{R}_{>0}}$ und $\alpha \in (0,1)$. Dann sind

$$[\widehat{\theta}_n + \tfrac{\widehat{\theta}_n}{n}\mathrm{q}_\alpha, \infty), \quad (0, \widehat{\theta}_n + \tfrac{\widehat{\theta}_n}{n}\mathrm{q}_{1-\alpha}] \quad \text{und} \quad [\widehat{\theta}_n + \tfrac{\widehat{\theta}_n}{n}\mathrm{q}_{\alpha/2}, \widehat{\theta}_n + \tfrac{\widehat{\theta}_n}{n}\mathrm{q}_{1-\alpha/2}],$$

wobei $\mathrm{q}_\alpha \in \mathbb{R}_{>0}$ das α-Quantil der stetigen Verteilung Exp_1 bezeichnet, asymptotische 1-α-Konfidenzbereiche für die Mengen der richtigen Parameter $[\theta, \infty)$, $(0, \theta]$ bzw. $\{\theta\}$ mit $\theta \in \mathbb{R}_{>0}$. Für $\theta_{\rm o} \in \mathbb{R}_{>0}$ mit $T_n^* := \frac{n}{\widehat{\theta}_n}(\theta_{\rm o} - \widehat{\theta}_n)$ sind damit

(a) der rechtsseitige Test $\varphi_{\mathrm{q}_{1-\alpha}}^r := \mathbb{1}_{\{T_n^* > \mathrm{q}_{1-\alpha}\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \theta \geqslant \theta_{\rm o}$ gegen die Alternative $H_1 : \theta_{\rm o} - \theta > 0$;

(b) der linksseitige Test $\varphi_{\mathrm{q}_\alpha}^l := \mathbb{1}_{\{T_n^* < \mathrm{q}_\alpha\}}$ für das *einseitige Testproblem* der Nullhypothese $H_0 : \theta \leqslant \theta_{\rm o}$ gegen die Alternative $H_1 : \theta_{\rm o} - \theta < 0$;

(c) der beidseitige Test $\varphi_{\mathrm{q}_{\alpha/2},\mathrm{q}_{1-\alpha/2}}^b := \mathbb{1}_{\{T_n^* < \mathrm{q}_{\alpha/2}\}} + \mathbb{1}_{\{T_n^* > \mathrm{q}_{1-\alpha/2}\}} = \varphi_{\mathrm{q}_{\alpha/2}}^l + \varphi_{\mathrm{q}_{1-\alpha/2}}^r$ für das *zweiseitige Testproblem* der Nullhypothese $H_0 : \theta = \theta_{\rm o}$ gegen die Alternative $H_1 : \theta_{\rm o} - \theta \neq 0$

asymptotische α-Tests. □

§33.22 **Lemma**. *Sei $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ eine d_n-konsistente Folge von Schätzern für den abgeleiteten Parameter $\gamma : \Theta \to \mathbb{R}$ mit Störparameter $\tau : \Theta \to \mathbb{R}_{\neq 0}$ und stetiger asymptotischer Verteilung $\mathbb{P} \in \mathcal{W}(\mathscr{B})$, d.h. $\frac{\mathrm{d}_n}{\tau(\theta)}(\widehat{\gamma}_n - \gamma\theta) \xrightarrow{D} \mathbb{P}$ unter $\mathbb{P}_\theta^n$. Für $\alpha \in (0,1)$ bezeichne $\mathrm{q}_\alpha \in \mathbb{R}$ das α-Quantil der stetigen Verteilung $\mathbb{P}$. Ist $\widehat{\tau}_n$ ein schwach konsister Schätzer für τ, dann sind für $\gamma_{\rm o} \in \mathbb{R}$ mit $T_n^* := \frac{\mathrm{d}_n}{\widehat{\tau}_n}(\widehat{\gamma}_n - \gamma_{\rm o})$*

(i) *der rechtsseitige Test $\varphi_{\mathrm{q}_{1-\alpha}}^r := \mathbb{1}_{\{T_n^* > \mathrm{q}_{1-\alpha}\}}$ für das einseitige Testproblem der Nullhypothese $H_0 : \gamma \leqslant \gamma_{\rm o}$ gegen die Alternative $H_1 : \gamma - \gamma_{\rm o} > 0$;*

(ii) *der linksseitige Test $\varphi_{\mathrm{q}_\alpha}^l := \mathbb{1}_{\{T_n^* < \mathrm{q}_\alpha\}}$ für das einseitige Testproblem der Nullhypothese $H_0 : \gamma \geqslant \gamma_{\rm o}$ gegen die Alternative $H_1 : \gamma - \gamma_{\rm o} < 0$;*

(iii) *der beidseitige Test $\varphi_{\mathrm{q}_{\alpha/2},\mathrm{q}_{1-\alpha/2}}^b := \mathbb{1}_{\{T_n^* < \mathrm{q}_{\alpha/2}\}} + \mathbb{1}_{\{T_n^* > \mathrm{q}_{1-\alpha/2}\}} = \varphi_{\mathrm{q}_{\alpha/2}}^l + \varphi_{\mathrm{q}_{1-\alpha/2}}^r$ für das zweiseitige Testproblem der Nullhypothese $H_0 : \gamma = \gamma_{\rm o}$ gegen die Alternative $H_1 : \gamma - \gamma_{\rm o} \neq 0$*

asymptotische α-Tests.

§33.23 **Beweis** von Lemma §33.22. In der Vorlesung. □

§33.24 **Beispiel** *(Beispiel §33.15 fortgesetzt).* Sei $(X_i)_{i\in\mathbb{N}} \circledcirc \mathcal{W}_4^{\mathbb{N}}(\mathscr{B})$ und die Varianz

$$\sigma^2 : \mathcal{W}_4(\mathscr{B}) \ni \mathbb{P} \mapsto \sigma^2(\mathbb{P}) = \mathrm{M}^{(2)}(\mathbb{P}) \in \mathbb{R}_{\geqslant 0}$$

der interessierende Parameter. Für $\sigma_0^2 \in \mathbb{R}_{>0}$ mit $T_n^* := \frac{\sqrt{n}}{\widehat{\tau}_n}(\widehat{S}_n^{(2)} - \sigma_0^2)$, $\alpha \in (0,1)$ und α-Quantil z_α der Standardnormalverteilung sind

- der rechtsseitige Test $\varphi^r_{z_{1-\alpha}} := \mathbb{1}_{\{T_n^* > z_{1-\alpha}\}}$,
- der linksseitige Test $\varphi^l_{z_\alpha} := \mathbb{1}_{\{T_n^* < z_\alpha\}}$ und
- der beidseitige Test $\varphi^b_{z_{1-\alpha/2}} := \mathbb{1}_{\{|T_n^*| > z_{1-\alpha/2}\}}$

für das entsprechnende Testproblem asymptotische α-Tests. □

§33|04 Lokations-Skalen-Modell

§33.25 **Korollar.** *Für eine Folge $(X_i)_{i\in\mathbb{N}}$ unabhängiger und identisch verteilter reeller Zufallsvariablen sei ein Lokations-Skalen-Modell adäquat, also $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu,\sigma^2)})_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_{>0}}$. Dann sind der empirische Mittelwert $\overline{X}_n$ und die empirische Varianz $\widehat{S}_n^{(2)}$ stark konsistente Schätzer für die abgeleiteten Parameter μ bzw. σ^2.*

§33.26 **Beweis** von Korollar §33.25. Folgt aus Beispiel §33.03 mit $\overline{X}_n = \widehat{m}_n^{(1)}$ und $\widehat{S}_n^{(2)} = \frac{n}{n-1}\widehat{V}_n^{(2)}$. □

§33.27 **Beispiel.**

(a) Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{B}^{\mathbb{N}}_{\mathrm{p}})_{\mathrm{p}\in[0,1]}$. Der MLS $\widehat{\mathrm{p}}_n = \overline{X}_n$ für p ist stark konsistent.

(b) Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{Poi}^{\mathbb{N}}_{\lambda})_{\lambda\in\mathbb{R}_{>0}}$. Der MLS $\widehat{\lambda}_n = \overline{X}_n$ ist stark konsistent für λ.

(c) Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{Exp}^{\mathbb{N}}_{\lambda})_{\lambda\in\mathbb{R}_{>0}}$. Der MLS $\widehat{\lambda}_n = (\overline{X}_n)^{-1}$ und der korrigierter MLS $\widetilde{\lambda}_n = (\frac{n}{n-1}\overline{X}_n)^{-1}$ für λ sind stark konsistent. □

§33.28 **Korollar.** *Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu,\sigma^2)})_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_{>0}}$. Dann ist der empirische Mittelwert $\overline{X}_n$ ein $\sqrt{n}$-konsistenter Schätzer für μ mit Störparameter σ und asymptotischer Standardnormalverteilung.*

§33.29 **Beweis** von Korollar §33.28. Die Behauptung folgt direkt aus dem zentralen Grenzwertsatz §32.02. □

§33.30 **Beispiel.**

(a) Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{B}^{\mathbb{N}}_{\mathrm{p}})_{\mathrm{p}\in[0,1]}$. Der MLS $\widehat{\mathrm{p}}_n = \overline{X}_n$ für $\mathrm{p} \in (0,1)$ ist $\sqrt{n}$-konsistent, mit Störparameter $\sqrt{\mathrm{p}(1-\mathrm{p})}$ und asymptotischer Standardnormalverteilung.

(b) Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{Poi}^{\mathbb{N}}_{\lambda})_{\lambda\in\mathbb{R}_{>0}}$. Der MLS $\widehat{\lambda}_n = \overline{X}_n$ ist $\sqrt{n}$-konsistent für λ, mit Störparameter $\sqrt{\lambda}$ und asymptotischer Standardnormalverteilung. □

§33.31 **Ausblick.** Wir haben in verschiedenen statistischen Modellen die asymptotische Verteilung des Maximum-Likelihood-Schätzers hergeleitet. In der Vorlesung Statistik I bestimmen wir Bedingungen an das statistische Modell, sodass der MLS asymptotisch normalverteilt ist. □

§33.32 **Korollar.** *Sei $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu,\sigma^2)})_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_{>0}}$, $\alpha \in (0,1)$ und $z_\alpha \in \mathbb{R}$ das α-Quantil der Standardnormalverteilung. Dann sind*

$$[\overline{X}_n - \tfrac{\widehat{S}_n}{\sqrt{n}} z_{1-\alpha}, \infty), \quad (-\infty, \overline{X}_n + \tfrac{\widehat{S}_n}{\sqrt{n}} z_{1-\alpha}] \quad \textit{und} \quad [\overline{X}_n \pm \tfrac{\widehat{S}_n}{\sqrt{n}} z_{1-\alpha/2}]$$

asymptotische 1-α-Konfidenzbereiche für die Mengen der richtigen Parameter $\mathbb{R}_{\geqslant\mu}$, $\mathbb{R}_{\leqslant\mu}$ bzw. $\{\mu\}$ mit $\mu \in \mathbb{R}$.

§33.33 **Beweis** von Korollar §33.32. Die Behauptung folgt direkt aus Lemma §33.13. □

§33.34 **Beispiel**. Sei $\alpha \in (0,1)$ und $z_\alpha \in \mathbb{R}$ das α-Quantil der Standardnormalverteilung.

(a) Für $(X_i)_{i\in\mathbb{N}} \circleddash (\mathrm{B}_\mathrm{p}^{\mathbb{N}})_{\mathrm{p}\in(0,1)}$ sind

$$[\overline{X}_n - \tfrac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}} z_{1-\alpha}, 1), \quad (0, \overline{X}_n + \tfrac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}} z_{1-\alpha}] \quad \text{und} \quad [\overline{X}_n \pm \tfrac{\sqrt{\overline{X}_n(1-\overline{X}_n)}}{\sqrt{n}} z_{1-\alpha/2}]$$

asymptotische 1-α-Konfidenzbereich für die Mengen der richtigen Parameter $[\mathrm{p}, 1)$, $(0, \mathrm{p}]$ bzw. $\{\mathrm{p}\}$ mit $\mathrm{p} \in (0,1)$.

(b) Für $(X_i)_{i\in\mathbb{N}} \circleddash (\mathrm{Poi}_\lambda^{\mathbb{N}})_{\lambda\in\mathbb{R}_{>0}}$ sind

$$[\overline{X}_n - \tfrac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha}, \infty), \quad (0, \overline{X}_n + \tfrac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha}] \quad \text{und} \quad [\overline{X}_n \pm \tfrac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha/2}]$$

asymptotische 1-α-Konfidenzbereich für die Mengen der richtigen Parameter $[\lambda, \infty)$, $(0, \lambda]$ bzw. $\{\lambda\}$ mit $\lambda \in \mathbb{R}_{>0}$. □

§33.35 **Asymptotische Tests für den Erwartungswert**. *Sei $(X_i)_{i\in\mathbb{N}} \circleddash (\mathrm{P}_{(\mu,\sigma^2)}^{\mathbb{N}})_{(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}_{>0}}$. Für $\mu_0 \in \mathbb{R}$ mit $T_n^* := \frac{\sqrt{n}}{\widehat{S}_n}(\overline{X}_n - \mu_0)$, $\alpha \in (0,1)$ und α-Quantil z_α der Standardnormalverteilung sind*

(i) *der rechtsseitige Test $\varphi_{z_{1-\alpha}}^r := \mathbb{1}_{\{T_n^* > z_{1-\alpha}\}}$ für das einseitige Testproblem der Nullhypothese $H_0 : \mu \leqslant \mu_0$ gegen die Alternative $H_1 : \mu - \mu_0 > 0$;*

(ii) *der linksseitige Test $\varphi_{z_\alpha}^l := \mathbb{1}_{\{T_n^* < z_\alpha\}}$ für das einseitige Testproblem der Nullhypothese $H_0 : \mu \geqslant \mu_0$ gegen die Alternative $H_1 : \mu - \mu_0 < 0$;*

(iii) *der beidseitige Test $\varphi_{z_{1-\alpha/2}}^b := \mathbb{1}_{\{|T_n^*| > z_{1-\alpha/2}\}}$ für das zweiseitige Testproblem der Nullhypothese $H_0 : \mu = \mu_0$ gegen die Alternative $H_1 : \mu - \mu_0 \neq 0$*

asymptotische α-Tests.

§33.36 **Beweis** von Korollar §33.35. Die Behauptung folgt direkt aus Lemma §33.22. □

§33.37 **Beispiel**. Sei $\alpha \in (0,1)$ und $z_\alpha \in \mathbb{R}$ das α-Quantil der Standardnormalverteilung.

(a) *Bernoulli-Schema*: Für $\mathrm{p}_0 \in (0,1)$ mit $T_n^* := \frac{\sqrt{n}}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}(\overline{X}_n - \mathrm{p}_0)$ sind

- der rechtsseitige Test $\varphi_{z_{1-\alpha}}^r := \mathbb{1}_{\{T_n^* > z_{1-\alpha}\}}$,
- der linksseitige Test $\varphi_{z_\alpha}^l := \mathbb{1}_{\{T_n^* < z_\alpha\}}$ und
- der beidseitige Test $\varphi_{z_{1-\alpha/2}}^b := \mathbb{1}_{\{|T_n^*| > z_{1-\alpha/2}\}}$

für das entsprechende Testproblem asymptotische α-Tests.

(b) *Poissonverteilungsmodell*: Für $\lambda_0 \in \mathbb{R}_{>0}$ und $T_n^* := \frac{\sqrt{n}}{\sqrt{\overline{X}_n}}(\overline{X}_n - \lambda_0)$ sind

- der rechtsseitige Test $\varphi_{z_{1-\alpha}}^r := \mathbb{1}_{\{T_n^* > z_{1-\alpha}\}}$,
- der linksseitige Test $\varphi_{z_\alpha}^l := \mathbb{1}_{\{T_n^* < z_\alpha\}}$ und
- der beidseitige Test $\varphi_{z_{1-\alpha/2}}^b := \mathbb{1}_{\{|T_n^*| > z_{1-\alpha/2}\}}$

für das entsprechnende Testproblem asymptotische α-Tests. □

§33|05 Zwei-Stichproben-Modell

§33.38 **Vorbemerkung**. Die zufälligen Geschlechter der Konsumierenden der beiden Erhebungen können wir als Stichproben $X \odot (\mathrm{P}^{1000}_{(\mathrm{p}_X,\mathrm{p}_X(1-\mathrm{p}_X))})_{\mathrm{p}_X\in(0,1)}$ und $Y \odot (\mathrm{P}^{1000}_{(\mathrm{p}_Y,\mathrm{p}_Y(1-\mathrm{p}_Y))})_{\mathrm{p}_Y\in(0,1)}$ auffassen, dass heißt, sie können separat adäquat durch ein Lokations-Skalen-Modell beschrieben werden. Da p_X und p_Y der Anteil der Konsumentinnen in der ersten bzw. zweiten Erhebung ist, ist damit die zu überprüfende Nullhypothese $H_0 : \mathrm{p}_X = \mathrm{p}_Y$, der interessierende Parameter ist also $\mathrm{p}_X - \mathrm{p}_Y$. Es erscheint realistisch, dass X und Y unabhängig sind, sodass X und Y einem *Zwei-Stichproben-Modell mit unverbundenen Sichproben* folgt. Im Folgenden nehmen wir an, dass $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu_X,\sigma_X^2)})_{(\mu_X,\sigma_X)\in\mathbb{R}\times\mathbb{R}_{>0}}$ und $(Y_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu_Y,\sigma_Y^2)})_{(\mu_Y,\sigma_Y)\in\mathbb{R}\times\mathbb{R}_{>0}}$ unabhängig sind. Für die interessierenden Parameter μ_X und μ_Y sind dann $\overline{X}_n$ und $\overline{Y}_n$ *$\sqrt{n}$-konsistente* Schätzer mit Störparameter σ_X bzw. σ_Y und asymptotischer Standardnormalverteilung. Aufgrund der Unabhängigkeit ist $\overline{X}_n - \overline{Y}_n$ ein *$\sqrt{n}$-konsistenter* Schätzer für $\mu_X - \mu_Y$ mit Störparameter $\tau = \sqrt{\sigma_X^2 + \sigma_Y^2}$ und asymptotischer Standardnormalverteilung. Betrachten wir die Varianzen, so sind

$$\widehat{\mathrm{S}}^{(2)}_{n,X} := \frac{1}{n-1} \sum_{i\in[\![n]\!]} (X_i - \overline{X}_n)^2 \quad \text{und} \quad \widehat{\mathrm{S}}^{(2)}_{n,Y} := \frac{1}{n-1} \sum_{i\in[\![n]\!]} (Y_i - \overline{Y}_n)^2$$

konsistente Momentenschätzer für σ_X^2 bzw. σ_Y^2, und somit ist

$$\widehat{\tau}_n := \sqrt{\widehat{\mathrm{S}}^{(2)}_{n,X} + \widehat{\mathrm{S}}^{(2)}_{n,Y}}$$

ein konsistenter Schätzer für den Störparameter τ. □

§33.39 **Asymptotische Tests auf Gleichheit der Erwartungswerte**.
Seien $(X_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu_X,\sigma_X^2)})_{(\mu_X,\sigma_X)\in\mathbb{R}\times\mathbb{R}_{>0}}$ und $(Y_i)_{i\in\mathbb{N}} \odot (\mathrm{P}^{\mathbb{N}}_{(\mu_Y,\sigma_Y^2)})_{(\mu_Y,\sigma_Y)\in\mathbb{R}\times\mathbb{R}_{>0}}$ unabhängig. Für $\alpha \in (0,1)$, α-Quantil $z_\alpha \in \mathbb{R}$ der Standardnormalverteilung und $T_n^\star := \frac{\sqrt{n}}{\widehat{\tau}_n}(\overline{X}_n - \overline{Y}_n)$ sind

(i) *der rechtsseitige Test $\varphi^r_{z_{1-\alpha}} := \mathbb{1}_{\{T_n^* > z_{1-\alpha}\}}$ für das einseitige Testproblem der Nullhypothese $H_0 : \mu_X \leqslant \mu_Y$ gegen die Alternative $H_1 : \mu_X - \mu_Y > 0$;*

(ii) *der linksseitige Test $\varphi^l_{z_\alpha} := \mathbb{1}_{\{T_n^* < z_\alpha\}}$ für das einseitige Testproblem der Nullhypothese $H_0 : \mu_X \geqslant \mu_Y$ gegen die Alternative $H_1 : \mu_X - \mu_Y < 0$;*

(iii) *der beidseitige Test $\varphi^b_{z_{1-\alpha/2}} := \mathbb{1}_{\{|T_n^*| > z_{1-\alpha/2}\}}$ für das zweiseitige Testproblem der Nullhypothese $H_0 : \mu_X = \mu_Y$ gegen die Alternative $H_1 : \mu_X - \mu_Y \neq 0$*

asymptotische α-Tests.

§33.40 **Beweis** von Korollar §33.39. In der Vorlesung. □

§33.41 **Beispiel**. Fasse die zufälligen Gesschlechter der Konsumierenden in Beispiel I im Kapitel 1 Prolog als zwei unabhängige *Bernoulli-Schema* auf. Dann gilt

$$\frac{\sqrt{n}}{\tau}(\overline{X}_n - \overline{Y}_n) \xrightarrow{D} \mathrm{N}_{(0,1)} \quad \text{mit} \quad \tau = \sqrt{\mathrm{p}_X(1-\mathrm{p}_X) + \mathrm{p}_Y(1-\mathrm{p}_Y)}.$$

Für $T_n^\star := \frac{\sqrt{n}}{\widehat{\tau}_n}(\overline{X}_n - \overline{Y}_n)$ mit $\widehat{\tau}_n = \sqrt{\overline{X}_n(1-\overline{X}_n) + \overline{Y}_n(1-\overline{Y}_n)}$ sind

- der rechtsseitige Test $\varphi^r_{z_{1-\alpha}} := \mathbb{1}_{\{T_n^\star > z_{1-\alpha}\}}$,
- der linksseitige Test $\varphi^l_{z_\alpha} := \mathbb{1}_{\{T_n^\star < z_\alpha\}}$ und
- der beidseitige Test $\varphi^b_{z_{1-\alpha/2}} := \mathbb{1}_{\{|T_n^\star| > z_{1-\alpha/2}\}}$

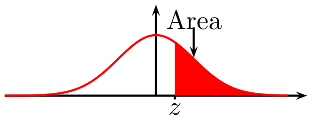
asymptotische α-Test für die entprechenden Testprobleme in Korollar §33.39. Dann erhalten wir die Schätzwerte $\overline{x_{1000}} = 0.699$ und $\overline{y_{1000}} = 0.389$, und damit $\widehat{\tau} \approx 0.669$, so dass $\mathrm{t}^{\star}_{1000} \approx 14.64$. Für $\alpha = 0.05$ erhalten wir $z_{1-\alpha/2} = 1.96$. Da $14.64 > 1.96$ können wir die Hypothese der Gleichheit der Erwartungswerte zum asymptotischen Niveau $\alpha = 0.05$ ablehnen. □

Anhang A

Anhang

A.1 Normalverteilung

Figure 1: Normal Curve Areas. Standard normal probability in right-hand tail. For negative values of z, areas are found by symmetry.



	Second decimal place of z									
z	0	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

Literatur

Bauer, H. (1992). *Maß- und Integrationstheorie*. 2., überarbeitete Auflage: Berlin etc.: Walter de Gruyter.

Elstrodt, J. (2011). *Maß- und Integrationstheorie.* 7., überarbeitete und ergänzte Auflage: Berlin: Springer.

Georgii, H.-O. (2015). *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik.* 5., überarbeitete und ergänzte Auflage: Berlin: De Gruyter.

Klenke, A. (2012). *Wahrscheinlichkeitstheorie*. 3., überarbeitete und ergänzte Auflage: Springer Spektrum.

— (2020). *Wahrscheinlichkeitstheorie*. 4., überarbeitete und ergänzte Auflage: Springer Spektrum.

Krengel, U. (2005). *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* 8., erweiterte Auflage: Braunschweig: Vieweg.

von Bortkewitsch, L. (1898). *Das Gesetz der kleinen Zahlen.* Leipzig: B. G. Teubner.