



*Skript zur Vorlesung*

# EINFÜHRUNG IN DIE WAHRSCHEINLICHKEITSTHEORIE UND STATISTIK

*Wintersemester 2020/21*

*Fassung Stand 16. Dezember 2020*

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

MATHEMATIKON, Im Neuenheimer Feld 205, 69120 Heidelberg

Telefon: +49 6221 54.14.190 – Fax: +49 6221 54.14.101

eMail: [johannes@math.uni-heidelberg.de](mailto:johannes@math.uni-heidelberg.de)

Webseite: [sip.math.uni-heidelberg.de](http://sip.math.uni-heidelberg.de)



# Inhaltsverzeichnis

|   |           |
|---|-----------|
| <b>1 Prolog</b>   | <b>1</b>  |
| Beispiel I: Geschlecht von Konsumierenden . . . . .                     | 1         |
| Beispiel II: Qualitätskontrolle . . . . .                               | 2         |
| Beispiel III: Anzahl . . . . .  | 3         |
| Beispiel IV: Wartezeit . . . . .  | 3         |
| Beispiel V: Lebensdauer . . . . .                                       | 3         |
| Beispiel VI: Messfehler . . . . .                                       | 4         |
| <b>2 Wahrscheinlichkeitsraum</b>  | <b>5</b>  |
| §01 Stichprobenraum . . . . .   | 5         |
| §02 Wahrscheinlichkeit . . . . .  | 8         |
| §03 Dynkin'scher $\pi$ - $\lambda$ -Satz . . . . .                      | 10        |
| §04 Diskreter Wahrscheinlichkeitsraum . . . . .                         | 12        |
| §05 Stetiger Wahrscheinlichkeitsraum . . . . .                          | 16        |
| §06 Statistisches Modell . . . . .                                      | 19        |
| <b>3 Zufallsvariablen</b>   | <b>25</b> |
| §07 Zufallsvariable . . . . .   | 25        |
| §08 Numerische und reelle Zufallsvariablen . . . . .                    | 26        |
| §09 Einfache Zufallsvariable . . . . .                                  | 27        |
| §10 Verteilung einer Zufallsvariablen . . . . .                         | 28        |
| §11 Verteilung einer Familie von Zufallsvariablen . . . . .             | 32        |
| §12 Statistische Inferenz . . . . .                                     | 36        |
| <b>4 Bedingte Wahrscheinlichkeit und Unabhängigkeit</b>                 | <b>45</b> |
| §13 Bedingte Wahrscheinlichkeiten und Bayes-Formel . . . . .            | 45        |
| §14 Unabhängige Ereignisse . . . . .                                    | 46        |
| §15 Unabhängige $\sigma$ -Algebren . . . . .                            | 47        |
| §16 Unabhängige Zufallsvariablen . . . . .                              | 48        |
| §17 Faltung . . . . .   | 51        |
| §18 Multivariate Normalverteilung . . . . .                             | 53        |
| §19 Beispiele statistischer Modelle . . . . .                           | 55        |
| <b>5 Erwartungswert</b>   | <b>57</b> |
| §20 Positive numerische Zufallsvariablen . . . . .                      | 57        |
| §21 Integrierbare Zufallsvariablen . . . . .                            | 59        |
| §22 Variablentransformation . . . . .                                   | 62        |
| §23 $\mathcal{L}_p$ -integrierbare Zufallsvariablen . . . . .           | 63        |
| §24 Varianz, Kovarianz und Korrelation . . . . .                        | 65        |
| §25 Hauptkomponentenanalyse . . . . .                                   | 68        |
| §26 Statistische Inferenz: endliche Stichproben Eigenschaften . . . . . | 72        |
| <b>Anhang</b>   | <b>85</b> |



# Kapitel 1

## Prolog

### Beispiel I: Geschlecht von Konsumierenden

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

W W W W M W W W M W M W W M W W W M W M W W M W M M W W W W W W W W W W M  
W W M W W W M W W W W W M W W M W W M W W W W M M M W W W W W W M W M M W  
W W M W W W M W W W M W W M W W W W W M W M W W W M M W W W W W W W M W W W  
M W W W W W W W M W W W M W W M M W W W M W M M W W W W W M W W M W M W W W  
W W W W W M W W W W W M W W M M W W M W W M M M W W W W W M W M W W W M M M  
W W M W W W W W W W W W W W M W M W W W W W W W W M M W W M W W W W W M W W  
W W M M W M W M W M W M W W W W M M W W W W M M W M W W W W W M W W W M W M  
W W W W W W W W M W W W M W M W W W W W M M M M W M M W W W M W W W W W W M  
W W W W W W W W M W W W W W M W M W W W W M W W W M M W W M W M W M W W W M  
W W M W M W W W W M M W M W M W W W W W W W M W W W M W W M W W W M W W W M W  
M M M W W W W W W W M W W W W W M M M W W W W W W W W W W W W W W W W W W W  
M W W W W W M W W W W W M M W W M W W W W W M M M W W W M M W W W W W W M W W  
M W M M W W W W W W M W W W W W M W W W W W M M M M W W M W W W W W W M  
M W  
W W M W  
W W W M M M W M W  
W M W M M W W M M M W W W W M W W W W W M W W W W W W W W W W W W W W W W W  
W W W W W M M W  
W W W W W W M W  
M W W W W W M W M 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 Verbraucherser-

vice wieder das Geschlecht von 1000 Konsumierenden des Getränkes erhoben. Mit folgendem Resultat:

M M M M M M M M W M W M W W M W M M M M M W M M W W W M M M M W W M M M M M  
M M M W M W W M M M M M W W M M W M W W W W W M M W W M M W W W W M M M M  
M M M M M M M M M M M W M M M M M M W M M W M M W W W M W M W W M M W  
M W W W W M W M M M M W M M W W W M M M M M M W M M W M W M M M W W W M W M  
W W W W M W W M M W W M M W M W M M M M M M W M M M M M M M M W M M W W W  
M W M M M W M M M W W M M M W M M M M W M M W M M M M M M M W M M M W M M M  
W M M M M W M M M W W W W M M W M M W M M M W W M W M M M M M M W W M W M M  
W M W W M M M W W M W W M M M W M M M M M M W M M W M W W M M W M W M M W M  
W W W W M M W M M W W M W M W M M W W M M M W W M M M M W M M W M W M M W M  
M M W W M M M W M W M M M M M M M W M M M M M M M M W M M W W M M M M M W W  
W M M M M M M W W M W M M M W W W M M M W W W M M M W W M M M M W M M M M W  
M M M W M W M M M M W M W W M W W M M M W W M W W W M M M W W M W M M M M M  
W W W W M M M M W M M W M W M W M M W M M M M M W W W M M W M M M M M W W M  
W M M M M W W M M M M W M M M M W M M W M M M W W M M M W M M M M W M W W M  
M M M W M M W W M M W M M W M M W W W W M M W W M M M M W M W W M M M M M  
W M W M W M M M W M W W W W M W M M M W M W W W W M M M M M W M W M M M W M  
M M M W W W M M M M W W W M W M W W M M M M M M W M W M W M M W W W W M W W  
M W W M M M M W W M M W W M W M M M M W M M M M M W M M W W M M M W M M M M  
M M M W M M W W M M M W W M M W W M M M M M M M M M W M M W M W M W W M M M  
M W W W M M W M M M M W M W M M M M W M M M M W M W M M M W W W W W W W W W  
M M W M M W M M M W M W M W M M M M M M M M W M W M W W M W W W W M M M  
M M M M M M M M M W W W M W W W M M W M M M W M W W M M M W M M M M M W W  
W W M M M M W W W M W W W M M W M M W W W W W M W W W M M W W M W M W M W W  
M M W W W M M W W M M M W M W M M W M M M M M W W M W M W M W M M M W W M M  
W W W M M W M W W M M M

(Anzahl M=611)

## Beispiel II: Qualitätskontrolle

Eine Schraubenherstellerin hat einen Vertrag mit einem Klienten abgeschlossen, in dem sie sich verpflichtet, dass in der nächsten Lieferung *höchstens 1 von 100 Schrauben beschädigt* ist. Der Klient 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 0 0 1 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)

### 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

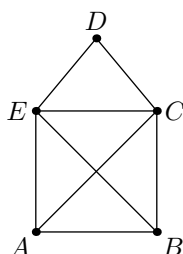
Eine Stadtverwaltung benutzt in ihren Gebäuden 35.000 Glühlampen. Der Hersteller der Glühlampen garantiert, dass *die Lebensdauer der Glühlampe* mindestens 1000 Stunden beträgt. Um diese Aussage zu überprüfen, führt die Einkaufsabteilung eine Studie mit 100 Glühlampen durch. Mit dem folgenden Resultat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

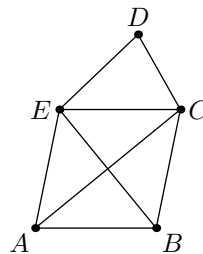
(Mittelwert=1026.37)

## Beispiel VI: Messfehler

Nach einem starken Sturm fürchtet der alte Fürst G. Ömetry 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. Ömetry 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.

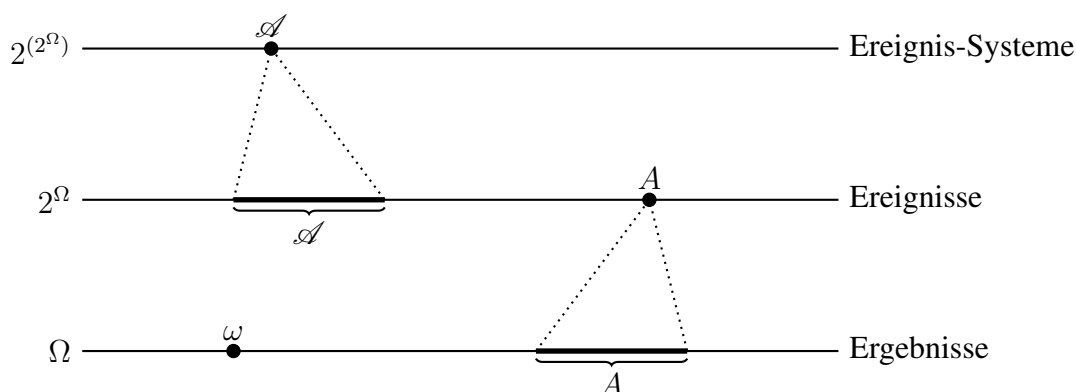
§01.01 **Definition.** Eine nicht-leere Menge  $\Omega$  aller möglichen Ausgänge eines zufälligen Experiments wird *Ergebnismenge* (*Grundmenge* oder *Stichprobenraum*) genannt. Ein möglicher Versuchsausgang  $\omega$  des zufälligen Experiments, also ein Element von  $\Omega$ , kurz  $\omega \in \Omega$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

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  $\Omega$ , also ein Element der Potenzmenge  $2^\Omega$  von  $\Omega$ . Für einen Versuchsausgang  $\omega \in \Omega$  ist ein Ereignis  $A \in 2^\Omega$  eingetreten, wenn  $\omega \in A$  gilt. Wir bezeichnen mit  $A^c := \Omega \setminus A \in 2^\Omega$  das Komplement von  $A$  in  $\Omega$ .  $\square$

§01.03 **Sprechweise.** Die Mengen  $\emptyset$ ,  $\Omega$  und  $\{\omega\}$ ,  $\omega \in \Omega$ , heißen das (absolut) *unmögliche Ereignis*, das (absolut) *sichere Ereignis* bzw. ein *Elementarereignis*; Ereignisse  $A$  und  $B$  mit  $A \cap B = \emptyset$  nennt man *unvereinbar* oder *unverträglich*.  $\square$

§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.  $\square$

§01.05 **Definition.** Ein Teilmengensystem  $\mathcal{A} \subseteq 2^\Omega$  heißt  *$\sigma$ -Algebra* über  $\Omega$ , falls gilt: (a)  $\Omega \in \mathcal{A}$ ; (b)  $A^c \in \mathcal{A}$  für alle  $A \in \mathcal{A}$ ; (c)  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$  für alle  $A_n \in \mathcal{A}$ ,  $n \in \mathbb{N}$ . Das Paar  $(\Omega, \mathcal{A})$  wird *messbarer Raum* genannt und  $\mathcal{A}$  wird als *Menge der interessierenden Ereignisse* bezeichnet.  $\square$

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

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

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

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

§01.08 **Beispiel.**

- (a) Auf jeder nicht-leeren Ergebnismenge  $\Omega$  existieren die **triviale  $\sigma$ -Algebra**  $\{\emptyset, \Omega\}$  sowie die Potenzmenge  $2^\Omega$  als  $\sigma$ -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  $\mathcal{E} = \left\{ A_i \in 2^\Omega, i \in \mathcal{I}, \text{ paarweise disjunkt und } \biguplus_{i \in \mathcal{I}} A_i = \Omega \right\}$  eine Partition von  $\Omega$ . Dann ist  $\sigma(\mathcal{E}) = \left\{ \biguplus_{j \in \mathcal{J}} A_j \mid \mathcal{J} \subseteq \mathcal{I} \right\}$ . □

§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  $\sigma$ -Algebra  $\mathcal{A}$  über  $\Omega$  gilt:

- (i)  $\emptyset \in \mathcal{A}$ ;
- (ii) Für  $A, B \in \mathcal{A}$  sind  $A \cup B, A \cap B, A \setminus B := A \cap B^c$  und  $A \Delta B := (A \setminus B) \cup (B \setminus A)$  Elemente von  $\mathcal{A}$ ;
- (iii) Für eine abzählbare Indexmenge  $\mathcal{I}$  und  $\{A_i : i \in \mathcal{I}\} \subseteq \mathcal{A}$  gilt  $\bigcap_{i \in \mathcal{I}} A_i \in \mathcal{A}$ .

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

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

§01.13 **Definition.** Sei  $A_n \subseteq \Omega$  für  $n \in \mathbb{N}$ . Die Folge  $(A_n)_{n \in \mathbb{N}}$  heißt **monoton wachsend** (bzw. **fallend**), wenn  $A_n \subseteq A_{n+1}$  (bzw.  $A_{n+1} \subseteq A_n$ ) für alle  $n \in \mathbb{N}$  gilt. Weiterhin heißen

$$A_\star := \liminf_{n \rightarrow \infty} A_n := \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m := \bigcup \left\{ \bigcap \{ A_m \mid m \in \mathbb{N} \wedge m \geq n \} \mid n \in \mathbb{N} \right\} \quad \text{und}$$

$$A^\star := \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m.$$

**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  $\lim_{n \rightarrow \infty} A_n = A$ . □

§01.14 **Bemerkung.**

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

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

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

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

In anderen Worten,  $A_\star$  ist also das Ereignis, dass *schließlich alle der  $A_n$  eintreten*.  $A^\star$  ist hingegen das Ereignis, dass *unendlich viele der  $A_n$  eintreten*. Insbesondere gilt  $A_\star = \liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n = A^\star$ .  $\square$

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

- (i)  $\liminf_{n \rightarrow \infty} A_n \in \mathcal{A}$  und  $\limsup_{n \rightarrow \infty} A_n \in \mathcal{A}$ ;
- (ii) Falls  $A := \lim_{n \rightarrow \infty} A_n$  existiert, dann ist  $A \in \mathcal{A}$ .

§01.16 **Beweis von Lemma §01.15.** Übungsaufgabe.  $\square$

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

§01.18 **Bemerkung.** Häufig sind wir an der Borel- $\sigma$ -Algebra  $\mathcal{B}^n := \mathcal{B}_{\mathbb{R}^n}$  über  $\mathbb{R}^n$  interessiert, wobei  $\mathbb{R}^n$  versehen ist mit dem Euklidischen Abstand  $d(x, y) = \|x - y\| = \sqrt{\sum_{i \in \llbracket n \rrbracket} (x_i - y_i)^2}$  für  $x = (x_i)_{i \in \llbracket n \rrbracket}, y = (y_i)_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n$  mit  $\llbracket n \rrbracket := [1, n] \cap \mathbb{N}$ .  $\square$

§01.19 **Schreibweise.** Wir schreiben  $\llbracket a, b \rrbracket := [a, b] \cap \mathbb{Z}$  für  $a, b \in \mathbb{R}$  mit  $a \leq b$ . Wir setzen  $\mathbb{R}^+ := [0, \infty)$ ,  $\mathbb{Q}^+ := [0, \infty) \cap \mathbb{Q}$ ,  $\mathbb{R}_0^+ := (0, \infty)$ ,  $\mathbb{Q}_0^+ := (0, \infty) \cap \mathbb{Q}$ ,  $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$ , sowie  $\mathbb{Q}_0 := \mathbb{Q} \setminus \{0\}$ . Für  $a, b \in \mathbb{R}^n$  schreiben wir  $a < b$ , wenn  $a_i < b_i$  für alle  $i \in \llbracket n \rrbracket$  gilt. Für  $a < b$ , definieren wir den offenen *Quader* als das Kartesische Produkt  $(a, b) := \prod_{i \in \llbracket n \rrbracket} (a_i, b_i) := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$ . Analog, sind  $[a, b]$ ,  $(a, b]$  sowie  $[a, b)$  definiert. Weiterhin, sei  $(-\infty, b) := \prod_{i \in \llbracket n \rrbracket} (-\infty, b_i)$  und analog  $(-\infty, b]$  definiert.  $\square$

§01.20 **Bemerkung.** Wie in *Schreibweise §01.19* bezeichnet  $\prod_{i \in \mathcal{I}} S_i$  das Kartesische Produkt der Mengen  $S_i$ ,  $i \in \mathcal{I}$ , und für  $s \in \prod_{i \in \mathcal{I}} S_i$  bezeichnen  $s_i$ ,  $i \in \mathcal{I}$ , stets die Komponenten von  $s = (s_i)_{i \in \mathcal{I}}$ .  $\square$

§01.21 **Satz.** Die Borel- $\sigma$ -Algebra  $\mathcal{B}^n$  über  $\mathbb{R}^n$  wird auch erzeugt von folgenden Mengensystemen:

- (i)  $\mathcal{E}_1 := \{A \subseteq \mathbb{R}^n \mid A \text{ ist abgeschlossen}\}$ ; (ii)  $\mathcal{E}_2 := \{A \subseteq \mathbb{R}^n \mid A \text{ ist kompakt}\}$ ;
- (iii)  $\mathcal{E}_3 := \{(a, b) \mid a, b \in \mathbb{Q}^n, a < b\}$ ; (iv)  $\mathcal{E}_4 := \{[a, b] \mid a, b \in \mathbb{Q}^n, a < b\}$ ;
- (v)  $\mathcal{E}_5 := \{(a, b) \mid a, b \in \mathbb{Q}^n, a < b\}$ ; (vi)  $\mathcal{E}_6 := \{[a, b] \mid a, b \in \mathbb{Q}^n, a < b\}$ ;
- (vii)  $\mathcal{E}_7 := \{(-\infty, b] \mid b \in \mathbb{Q}^n\}$ ; (viii)  $\mathcal{E}_8 := \{(-\infty, b) \mid b \in \mathbb{Q}^n\}$ ; (ix)  $\mathcal{E}_9 := \{(a, \infty) \mid a \in \mathbb{Q}^n\}$  und (x)  $\mathcal{E}_{10} := \{[a, \infty) \mid a \in \mathbb{Q}^n\}$ .

§01.22 **Beweis von Satz §01.21.** In der Vorlesung Wahrscheinlichkeitstheorie 1.  $\square$

§01.23 **Lemma.** Seien  $(\Omega, \mathcal{A})$  ein messbarer Raum und  $B \subseteq \Omega$  eine nicht-leere Teilmenge. Dann ist  $\mathcal{A}|_B := \mathcal{A} \cap B := \{A \cap B \mid A \in \mathcal{A}\}$  eine  $\sigma$ -Algebra über  $B$ , die sogenannte *Spur* oder *Einschränkung* von  $\mathcal{A}$  auf  $B$ . Des Weiteren, für  $\mathcal{E} \in 2^\Omega$  gilt  $\sigma(\mathcal{E})|_B = \sigma(\mathcal{E}|_B)$ .

§01.24 **Beweis** von Lemma §01.23. Übungsaufgabe.  $\square$

§01.25 **Schreibweisen.** Wir bezeichnen mit  $\overline{\mathcal{B}} := \mathcal{B}_{\overline{\mathbb{R}}}$  die Borel- $\sigma$ -Algebra über der kompaktifizierten Zahlengerade  $\overline{\mathbb{R}} := [-\infty, \infty]$ , wobei die Mengen  $\{-\infty\}$ ,  $\{\infty\}$  und  $\mathbb{R}$  in  $\overline{\mathbb{R}}$  abgeschlossen bzw. offen, also Borel-Mengen sind. Insbesondere, ist  $\mathcal{B} := \mathcal{B}_{\mathbb{R}} = \overline{\mathcal{B}} \cap \mathbb{R}$  die Borel- $\sigma$ -Algebra über  $\mathbb{R}$ . Weiterhin schreiben wir  $\overline{\mathcal{B}}^+ := \overline{\mathcal{B}} \cap \mathbb{R}^+$ ,  $\mathcal{B}^+ := \mathcal{B} \cap \mathbb{R}^+$  und  $\mathcal{B}_{\geq 0}^+ := \mathcal{B} \cap \mathbb{R}_{\geq 0}^+$ .  $\square$

## §02 Wahrscheinlichkeit

In einem zufälligem Experiment, das durch einen messbaren Raum  $(\Omega, \mathcal{A})$  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  $(\Omega, \mathcal{A})$  ein messbarer Raum. Eine Abbildung  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  heißt *Wahrscheinlichkeitsmaß* (*Wahrscheinlichkeitsverteilung* oder kurz *Verteilung*) auf  $\mathcal{A}$ , wenn sie folgenden Bedingungen genügt:

- (a)  $\mathbb{P}(\Omega) = 1$ ; (normiert)
- (b)  $\mathbb{P}(\biguplus_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$  für jede Folge  $(A_n)_{n \in \mathbb{N}}$  paarweise disjunkter Ereignisse aus  $\mathcal{A}$ , d.h.  $A_n \cap A_m = \emptyset$  für alle  $n, m \in \mathbb{N}$  mit  $n \neq m$ . ( $\sigma$ -additiv)

Wir bezeichnen mit  $\mathcal{W} := \mathcal{W}(\mathcal{A})$  die Menge aller Wahrscheinlichkeitsmaße auf  $\mathcal{A}$ . Ein Tripel  $(\Omega, \mathcal{A}, \mathbb{P})$  bestehend aus einer Ergebnismenge  $\Omega$ , einer  $\sigma$ -Algebra  $\mathcal{A}$  interessierender Ereignisse sowie einem Wahrscheinlichkeitsmaß  $\mathbb{P} \in \mathcal{W}$  wird *Wahrscheinlichkeitsraum* genannt.  $\square$

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

§02.03 **Beispiel.** Sei  $(\Omega, \mathcal{A})$  ein messbarer Raum. Für  $A \subseteq \Omega$  bezeichnet  $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$  mit  $\mathbb{1}_A^{-1}(\{1\}) = A$  und  $\mathbb{1}_A^{-1}(\{0\}) = A^c$  die *Indikatorfunktion* auf  $A$ . Für jedes  $\omega \in \Omega$  ist das *Einpunkt-* oder *Diracmaß*  $\delta_\omega : \mathcal{A} \rightarrow \{0, 1\}$  mit  $\delta_\omega(A) := \mathbb{1}_A(\omega)$  für alle  $A \in \mathcal{A}$  ein Wahrscheinlichkeitsmaß aus  $\mathcal{W}(\mathcal{A})$ . Ist  $(\mathbb{P}_n)_{n \in \mathbb{N}}$  eine Folge von Wahrscheinlichkeitsmaße in  $\mathcal{W}$ , so ist auch jede Konvexkombination  $\sum_{n \in \mathbb{N}} w_n \mathbb{P}_n$  mit  $w_n \geq 0$ ,  $n \in \mathbb{N}$ , und  $\sum_{n \in \mathbb{N}} w_n = 1$  in  $\mathcal{W}$ . Die Diracmaße bilden Extrempunkte der konvexen Menge aller Wahrscheinlichkeitsmaße.  $\square$

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

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

§02.06 **Beweis** von Lemma §02.05. Da es höchstens  $n$  Atome mit Masse gleich oder größer  $1/n$  gibt, kann die Anzahl der Atome höchstens abzählbar unendlich sein.  $\square$

§02.07 **Schreibweise.** Für  $a, b \in \mathbb{R}$  schreiben wir kurz  $a \vee b := \max\{a, b\}$  sowie  $a \wedge b := \min\{a, b\}$ .  $\square$

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

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

- (ii) Für jedes  $A \in \mathcal{A}$  gilt  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ ;
- (iii) Für alle  $A, B \in \mathcal{A}$  gilt (Boole)  
 $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ ,  
 $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ ,  
 $\mathbb{P}(A \Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$ ;  
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  $A_1, \dots, A_n \in \mathcal{A}$  gilt (Poincaré und Sylvester)  
 $\mathbb{P}(\bigcup_{i \in [n]} A_i) = \sum_{\emptyset \neq I \subseteq [n]} (-1)^{|I|-1} \mathbb{P}(\bigcap_{i \in I} A_i)$ ;
- (v) Für alle  $A, B \in \mathcal{A}$  mit  $A \subseteq B$  gilt  $0 \leq \mathbb{P}(A) \leq \mathbb{P}(B) \leq 1$  (Isotonie);
- (vi) Für  $A, B \in \mathcal{A}$  gilt: (Bonferroni)  
 $\mathbb{P}(A) \vee \mathbb{P}(B) \leq \mathbb{P}(A \cup B) \leq \{\mathbb{P}(A) + \mathbb{P}(B)\} \wedge 1$ ;  
 $\{\mathbb{P}(A) + \mathbb{P}(B) - 1\} \vee 0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(A) \wedge \mathbb{P}(B)$ ;  
 $|\mathbb{P}(A) - \mathbb{P}(B)| \leq \mathbb{P}(A \Delta B)$ .

§02.09 Beweis von Lemma §02.08. Übungsaufgabe. □

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

- (i)  $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$ ; (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  $\lim_{n \rightarrow \infty} A_n = A$ , dann gilt  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n \Delta A) = 0$  und somit  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ .

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

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

§02.13 Beweis von Lemma §02.12. In der Vorlesung. □

§02.14 **Bemerkung.** Die letzte Aussage in Lemma §02.12 erlaubt eine alternative aber zu Definition §02.01 äquivalente Definition eines Wahrscheinlichkeitsmaßes, d.h. eine Mengenfunktion  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  ist ein Wahrscheinlichkeitsmaß wenn es (a) normiert, (b) additiv und (c)  $\sigma$ -stetig ist. □

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

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

$$\mathbb{P}(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_{n \rightarrow \infty} A_n).$$

§02.16 Beweis von Lemma §02.12. Übungsaufgabe. □

§02.17 **Definition.** Für ein Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $(\mathbb{R}, \mathcal{B})$  wird  $\mathbb{F} : \mathbb{R} \rightarrow [0, 1]$  mit  $x \mapsto \mathbb{F}(x) := \mathbb{P}((-\infty, x])$  die zugehörige Verteilungsfunktion genannt. □

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

- (i)  $F$  ist monoton wachsend, d.h., für alle  $x, y \in \mathbb{R}$  mit  $x \leq y$  gilt  $F(x) \leq F(y)$ .
- (ii)  $F$  ist rechtsstetig, d.h., für alle  $x_n \downarrow x$  in  $\mathbb{R}$  gilt  $F(x_n) \downarrow F(x)$ .
- (iii)  $F$  besitzt einen linksseitigen Grenzwert, das heißt für alle  $x_n \uparrow x$  in  $\mathbb{R}$  gilt  $F(x_n) \uparrow F(x-) := \mathbb{P}((-\infty, x))$ .
- (iv) Die Sprünge von  $F$  korrespondieren mit den Atomen von  $\mathbb{P}$ , d.h.  $F(x) - F(x-) = \mathbb{P}(\{x\})$  für alle  $x \in \mathbb{R}$ .
- (v) Es gilt  $\lim_{x \rightarrow -\infty} F(x) = 0$  und  $\lim_{x \rightarrow \infty} F(x) = 1$ .

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

§02.20 **Bemerkung.** Aufgrund von Lemma §02.05 und Lemma §02.18 (iv) ist die Anzahl der Unstetigkeitsstellen einer Verteilungsfunktion  $F$  abzählbar. □

### §03 Dynkin'scher $\pi$ - $\lambda$ -Satz

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

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

- (a)  $\cap$ -stabil (sprich: *schnittstabil*) oder ein  $\pi$ -System, falls für je zwei Mengen  $A, B \in \mathcal{E}$  gilt, dass auch  $A \cap B \in \mathcal{E}$ ;
- (b)  $\sigma$ - $\cap$ -stabil (sprich: *sigma-schnittstabil*), falls für jede Folge  $(A_n)_{n \in \mathbb{N}}$  von Teilmengen aus  $\mathcal{E}$  gilt, dass auch  $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{E}$ ;
- (c)  $\cup$ -stabil (sprich: *vereinigungsstabil*), falls für je zwei Mengen  $A, B \in \mathcal{E}$  gilt, dass auch  $A \cup B \in \mathcal{E}$ ;
- (d)  $\sigma$ - $\cup$ -stabil (sprich: *sigma-vereinigungsstabil*), falls für jede Folge  $(A_n)_{n \in \mathbb{N}}$  von Teilmengen aus  $\mathcal{E}$  gilt, dass auch  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$ ;
- (e)  $\setminus$ -stabil (sprich: *differenzmengenstabil*), falls je zwei Mengen  $A, B \in \mathcal{E}$  gilt, dass auch  $A \setminus B \in \mathcal{E}$ ;
- (f) *komplementstabil*, falls mit jeder Menge  $A \in \mathcal{E}$  auch  $A^c \in \mathcal{E}$  gilt. □

§03.02 **Bemerkung.**

- (i) Ein Teilmengensystem  $\mathcal{E}$  aus  $2^\Omega$  mit  $\Omega \in \mathcal{E}$ , das komplementstabil und  $\sigma$ - $\cup$ -stabil ist, ist somit eine  $\sigma$ -Algebra.
- (ii) Für ein komplementstabiles Teilmengensystem  $\mathcal{E}$  aus  $2^\Omega$  folgen aus den de Morgan'schen Regeln die Äquivalenzen von  $\cup$ -stabil und  $\cap$ -stabil als auch von  $\sigma$ - $\cup$ -stabil und  $\sigma$ - $\cap$ -stabil. □

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

- ( $\lambda_1$ )  $\Omega \in \mathcal{D}$ ;



$(\lambda_2)$   $\mathcal{D}$  ist komplementstabil;

$(\lambda_3)$  für je abzählbar viele, paarweise disjunkte Menge  $A_1, A_2, \dots$  in  $\mathcal{D}$  gilt  $\biguplus_{n=1}^{\infty} A_n \in \mathcal{D}$ .  $\square$

§03.04 **Bemerkung.** Die Bedingung  $(\lambda_2)$ , das  $\mathcal{D}$  komplementstabil ist; kann äquivalent ersetzt werden durch die scheinbar stärkere Bedingung

$(\lambda'_2)$  für alle  $A, B \in \mathcal{D}$  mit  $A \subseteq B$  gilt  $B \setminus A \in \mathcal{D}$ .

Da jedes Dynkin-System auch  $(\lambda'_2)$  erfüllt. Denn für  $A, B \in \mathcal{D}$  mit  $A \subseteq B$  sind  $A$  und  $B^c$  disjunkt und es gilt  $B \setminus A = (A \biguplus B^c)^c \in \mathcal{D}$ .  $\square$

§03.5 **Anmerkung.** Jede  $\sigma$ -Algebra ist ein Dynkin-System. Die Umkehrung gilt nicht, da  $(\lambda_3)$  nur für Folgen paarweise disjunkter Ereignisse gefordert ist. Zum **Beispiel** für  $\Omega = \{1, 2, 3, 4\}$  ist  $\mathcal{D} = \{\emptyset, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \Omega\}$  ein Dynkin-System aber keine  $\sigma$ -Algebra. Der Unterschied ist allerdings nicht sehr groß.  $\square$

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

§03.07 **Beweis** von **Lemma** §03.06. In der Vorlesung.  $\square$

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

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

das kleinste Dynkin-System auf  $\Omega$ , das  $\mathcal{E}$  enthält.  $\mathcal{E}$  heißt **Erzeuger** von  $\delta(\mathcal{E})$  und  $\delta(\mathcal{E})$  das von  $\mathcal{E}$  erzeugte Dynkin-System auf  $\Omega$ .

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

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

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

§03.12 **Beweis** von **Satz** §03.11. In der Vorlesung.  $\square$

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

§03.14 **Beweis** von **Korollar** §03.13. „ $\supseteq$ “ Dies ist klar nach **Bemerkung** §03.10.

„ $\subseteq$ “ Nach **Satz** §03.11 ist das Dynkin-System  $\delta(\mathcal{E})$   $\cap$ -stabil, und somit nach **Lemma** §03.06 auch eine  $\sigma$ -Algebra, die definitionsgemäß  $\mathcal{E}$  enthält.  $\square$

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

§03.16 **Beweis** von **Korollar** §03.15. Für das erzeugte Dynkin-System  $\delta(\mathcal{E})$  gilt definitionsgemäß  $\delta(\mathcal{E}) \subseteq \mathcal{D}$ . Da  $\mathcal{E}$   $\cap$ -stabil ist, folgt  $\sigma(\mathcal{E}) = \delta(\mathcal{E})$  aus **Korollar** §03.13, also auch  $\sigma(\mathcal{E}) \subseteq \mathcal{D}$ .  $\square$

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

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

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

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

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

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

§03.19 **Beweis** von **Satz** §03.18. Übungsaufgabe unter Verwendung der **Beweisstrategie** §03.17 mit  $(R) = „\mathbb{P}(A) = \tilde{\mathbb{P}}(A)“$ .  $\square$

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

§03.21 **Beweis** von **Korollar** §03.20. Die **Existenz** wird in der Vorlesung Wahrscheinlichkeitstheorie 1 gezeigt, die **Eindeutigkeit** folgt direkt aus **Satz** §03.18 mit  $\cap$ -stabilen  $\mathcal{E} = \{(-\infty, x] \mid x \in \mathbb{R}\}$ .  $\square$

## §04 Diskreter Wahrscheinlichkeitsraum

§04.01 **Definition**. Ist  $\Omega$  eine abzählbare Menge, so wird ein Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $\mathcal{A} = 2^\Omega$  **diskret** genannt und  $(\Omega, \mathcal{A}, \mathbb{P})$  heißt **diskreter Wahrscheinlichkeitsraum**. Die Abbildung  $\mathbb{p} : \Omega \rightarrow [0, 1]$  mit  $\omega \mapsto \mathbb{p}(\omega) := \mathbb{P}(\{\omega\})$  wird **Zähldichte** des Wahrscheinlichkeitsmaßes  $\mathbb{P}$  genannt.  $\square$

§04.02 **Lemma**.

- (i) Ist  $(\Omega, \mathcal{A}, \mathbb{P})$  ein diskreter Wahrscheinlichkeitsraum, so ist  $\mathbb{P}$  eindeutig durch seine Zähldichte  $\mathbb{p}$  festgelegt.
- (ii) Ist andererseits  $\Omega$  eine abzählbare Menge und besitzt  $\mathbb{p} : \Omega \rightarrow [0, 1]$  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  $\mathbb{P} : 2^\Omega \rightarrow [0, 1]$  definiert, dessen Zähldichte gerade  $\mathbb{p}$  ist.

§04.03 **Beweis** von **Lemma** §04.02. Übungsaufgabe.  $\square$

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

*Ziehen mit Zurücklegen, mit Betrachtung der Reihenfolge:*

$$\Omega_1 = \{(k_i)_{i \in [n]} \mid k_i \in [N], \forall i \in [n]\} = [N]^n, |\Omega_1| = N^n;$$

*Ziehen ohne Zurücklegen, mit Betrachtung der Reihenfolge für  $n \leq N$ :*

$$\Omega_2 = \{(k_i)_{i \in [n]} \in [N]^n \mid k_1, \dots, k_n \text{ paarweise verschieden}\}, |\Omega_2| = \frac{N!}{(N-n)!};$$

*Ziehen ohne Zurücklegen, ohne Betrachtung der Reihenfolge für  $n \leq N$ :*

$$\Omega_3 = \{A \subseteq [N] \mid |A| = n\}, |\Omega_3| = \binom{N}{n};$$

*Ziehen mit Zurücklegen, ohne Betrachtung der Reihenfolge:*

$$\Omega_4 = \{(k_i)_{i \in [n]} \in [N]^n \mid k_1 \leq k_2 \leq \dots \leq k_n\}, |\Omega_4| = \binom{N+n-1}{n}.$$

§04.05 **Beweis** von **Lemma** §04.04. In der Vorlesung.  $\square$



§04.06 **Beispiel.**

- (a) Die Menge aller Ergebnisse im Lotto „6 aus 49“ ist  $\Omega_3$  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  $\Omega_1 = \llbracket N \rrbracket^n$  mit  $N = 365$ . Das Ereignis „keine zwei Personen haben am selben Tag Geburtstag“ entspricht  $\Omega_2 = \{(k_i)_{i \in \llbracket n \rrbracket} \in \llbracket N \rrbracket^n \mid k_1, \dots, k_n \text{ paarweise verschieden}\}$ . 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  $3.1 \cdot 10^{-7}$ .  $\square$

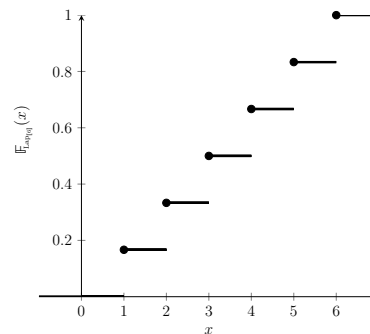
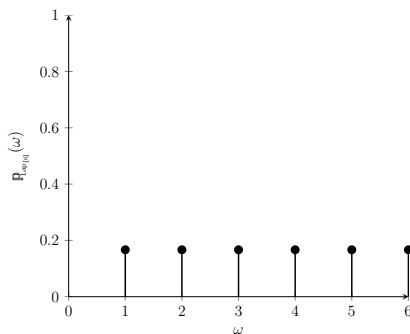
§04.7 **Anmerkung.** Sei  $\emptyset \neq \Omega \subseteq \mathbb{R}$  abzählbar und  $\mathbb{p}$  eine Zähldichte auf  $\Omega$ . Das Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $(\mathbb{R}, \mathcal{B})$  mit  $\mathbb{P}(B) := \sum_{\omega \in \Omega} \mathbb{p}(\omega) \delta_\omega$  für  $B \in \mathcal{B}$  und die zugehörige Verteilungsfunktion  $\mathbb{F} : \mathbb{R} \rightarrow [0, 1]$  mit  $\mathbb{F}(x) = \mathbb{P}((-\infty, x]) = \sum_{\omega \in \Omega} \mathbb{p}(\omega) \delta_\omega((-\infty, x]) = \sum_{\omega \in \Omega} \mathbb{1}_{(-\infty, x]}(\omega) \mathbb{p}(\omega) =: \sum_{\omega \leq x} \mathbb{p}(\omega)$  für  $x \in \mathbb{R}$  werden **diskret** genannt.  $\square$

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

- (a) **Laplace-/Gleich-Verteilung**, kurz **Lap $_\Omega$** , mit  $|\Omega| < \infty$ :

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

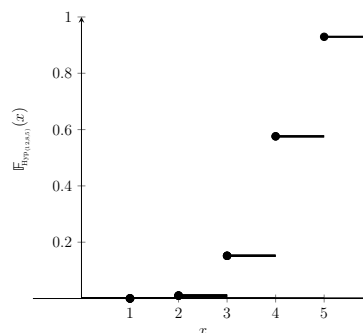
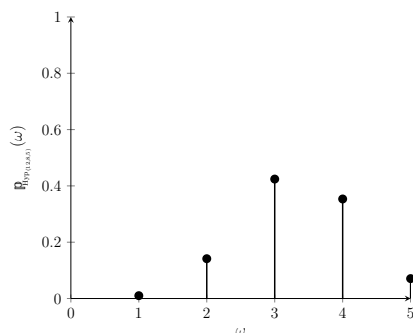
Betrachten wir den **Wurf eines Würfels**, so ist  $\Omega = \llbracket 6 \rrbracket$ . Ist der Würfel fair, so erhalten wir die Zähldichte  $\mathbb{p}_{\text{Lap}_{\llbracket 6 \rrbracket}}(\omega) = 1/6$ ,  $\omega \in \llbracket 1, 6 \rrbracket$ , mit der zugehörigen Verteilungsfunktion  $\mathbb{F}_{\text{Lap}_{\llbracket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



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

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

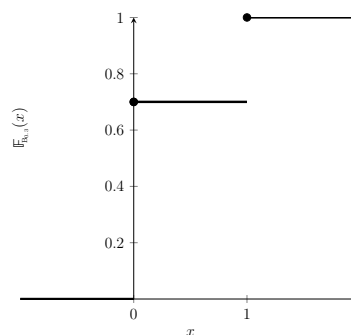
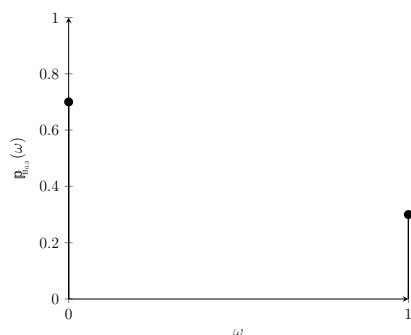
Ziehen wir  $n$  Kugeln aus einer Urne mit  $K$  schwarzen und  $N - K$  weißen Kugeln, so ist die Wahrscheinlichkeit, dass davon  $\omega$  schwarz sind, gerade durch  $\mathbb{p}_{\text{Hyp}_{(N,K,n)}}(\omega)$  gegeben. Für  $N = 12$ ,  $K = 8$  und  $n = 5$  erhalten wir die Zähldichte und die zugehörigen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



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

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

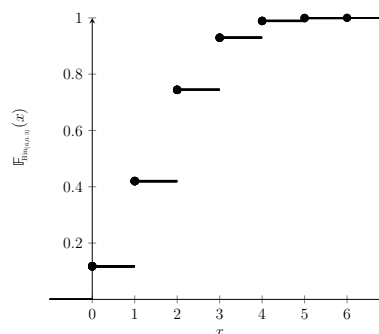
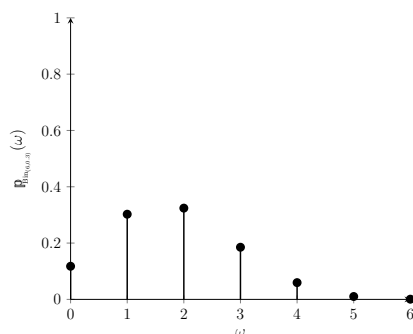
Im Fall  $n = 1$ , beschreibt eine **Bernoulli-Verteilung**, kurz  $B_p$ , also  $\mathbb{P}_{B_p}(\omega) = p^\omega (1-p)^{1-\omega}$  für  $\omega \in \{0, 1\}$ , gerade den Wurf einer Münze, wobei wir die Ergebnisse 1=„Kopf“ als Erfolg und 0=„Zahl“ als Misserfolg auffassen. Für  $B_{0.3}$  (die Münze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



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

$$\mathbb{P}_{\text{Bin}_{(n,p)}}(\omega) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega} \text{ für } \omega \in \llbracket 0, n \rrbracket.$$

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



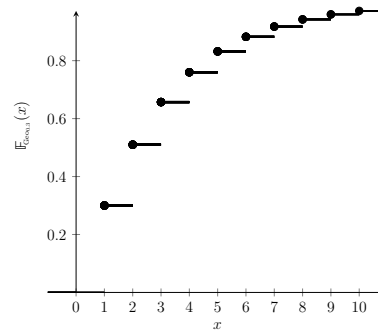
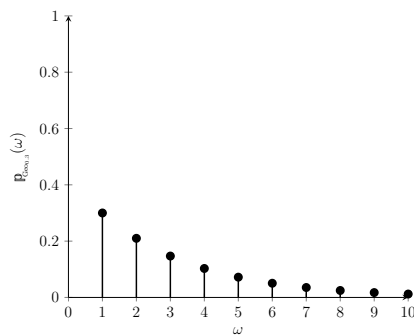
Ziehen wir wie in (b) Kugeln aus einer Urne aber mit Zurücklegen, so erhalten wir für die

Anzahl der gezogenen schwarzen Kugeln eine  $\text{Bin}_{(n,K/N)}$ -Verteilung. Die Binomialapproximation einer Hypergeometrischen Verteilung ist eine Übungsaufgabe.

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

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

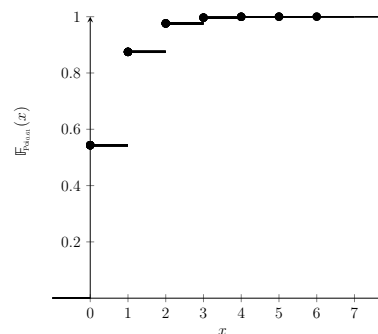
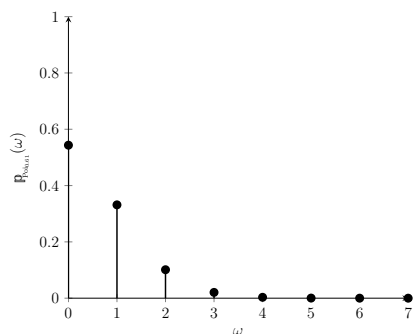
Werfen wir eine Münze wie in (c) solange bis das erste Mal „Kopf“ fällt, so kann die Anzahl der benötigten Würfe durch eine **Geo<sub>0.3</sub>**-Verteilung beschrieben werden, für die wir folgende Zähldichte und zugehörige Verteilungsfunktion erhalten:



(f) *Poissonverteilung*, kurz **Poi<sub>λ</sub>**, mit Parameter  $\lambda \in \mathbb{R}_0^+$  und  $\Omega = \mathbb{N}_0$ :

$$\mathbb{P}_{\text{Poi}_\lambda}(\omega) = \exp(-\lambda) \frac{\lambda^\omega}{\omega!} \text{ für } \omega \in \mathbb{N}_0.$$

Ladislau von Bortkewitsch [1898] beschreibt die Anzahl an Todesfällen in der preußischen Kavallerie durch den Schlag eines Pferdes (vgl. **Beispiel §04.12**) durch eine **Poi<sub>0.61</sub>**-Verteilung, für die wir folgende Zähldichte und zugehörigen Verteilungsfunktion erhalten:



□

§04.09 **Poissonscher Grenzwertsatz**. Es sei eine Folge von reellen Zahlen  $(p_n)_{n \in \mathbb{N}}$  aus dem Intervall  $[0, 1]$  mit  $\lambda := \lim_{n \rightarrow \infty} np_n \in \mathbb{R}_0^+$  gegeben. Dann gilt für alle  $\omega \in \mathbb{N}_0$ :

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\text{Bin}_{(n,p_n)}}(\omega) = \mathbb{P}_{\text{Poi}_\lambda}(\omega).$$

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

§04.11 **Bemerkung**. Damit kann man für hinreichend großes  $n$  und  $np$  „mittelgroß“ die  $\text{Bin}_{(n,p)}$ -Verteilung durch eine  $\text{Poi}_\lambda$ -Verteilung approximieren (zur Güte der Approximation vgl. Satz 5.9 in Krenzel [2005]). Das heißt insbesondere, dass die Erfolgswahrscheinlichkeit  $p$  klein sein muss. Die Abhängigkeit von  $p = p_n$  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 $\omega$                        | 0   | 1  | 2  | 3 | 4 | $\geq 5$ |
|--|-----|----|----|---|---|----------|
| Anzahl Korpsjahre mit $\omega$ Todesfällen | 109 | 65 | 22 | 3 | 1 | 0        |
| $200 p_{\text{Poi}_{0.61}}(\omega)$        | 109 | 66 | 20 | 4 | 1 | 0        |

Der Parameter  $\lambda$  der Poissonverteilung  $\text{Poi}_\lambda$  wurde als mittlere Anzahl an Todesfällen pro Korpsjahr gewählt  $\lambda = (1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 1 \cdot 4)/200 = 0.61$ .  $\square$

§04.13 **Definition.** Sind  $p_1, \dots, p_n$  Zähldichten auf  $\Omega$ , so heißt

$$p(\omega) := \prod_{i \in [n]} p_i(\omega_i) \quad \text{für } \omega = (\omega_i)_{i \in [n]} \in \Omega^n$$

die *Produktzähldichte* der  $(p_i)_{i \in [n]}$  auf  $\Omega^n$ .  $\square$

§04.14 **Beispiel.** Die Zähldichte eines Bernoulli-Schema wie **Beispiele** §04.08 (d) ist die Produktzähldichte von  $n$  Zähldichten zu identischen Bernoulli-Verteilungen  $B_p$ .  $\square$

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

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

die Abbildung, 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  $\mathbb{P}$  auf der Potenzmenge  $2^\Omega$ , das folgender Invarianzeigenschaft genügt: Für alle  $A \in 2^\Omega$  und  $n \in \mathbb{N}$  gilt  $\mathbb{P}(T_n A) = \mathbb{P}(A)$ . (Dies drückt die Fairness der Münze und die Unabhängigkeit der Würfe aus.)

§04.16 **Beweis von Satz §04.15.** In der Vorlesung.  $\square$

## §05 Stetiger Wahrscheinlichkeitsraum

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

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

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

ein eindeutiges Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf  $(\mathbb{R}^n, \mathcal{B}^n)$  und es gilt  $\mathbb{P}(B) = \int_B f(x) dx$  für alle  $B \in \mathcal{B}^n$ .

§05.03 **Beweis von Satz §05.02.** In der Vorlesung Wahrscheinlichkeitstheorie 1.  $\square$

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

§05.05 **Lemma.**

- (i) Ist  $f$  eine Dichte eines Wahrscheinlichkeitsmaßes  $\mathbb{P}$  auf  $\mathcal{B}$  mit Verteilungsfunktion  $F$ , so gilt  $F(x) = \int_{-\infty}^x f(y)dy$  für alle  $x \in \mathbb{R}$ .
- (ii) Ist die Verteilungsfunktion  $F$  eines Wahrscheinlichkeitsmaßes  $\mathbb{P}$  auf  $\mathcal{B}$  (schwach) differenzierbar, so ist  $f := F'$  die zugehörige Wahrscheinlichkeitsdichte.

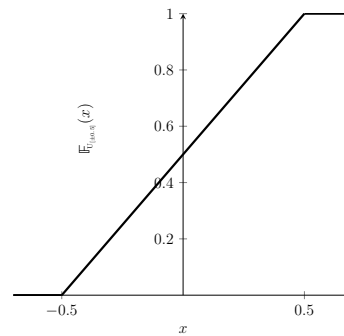
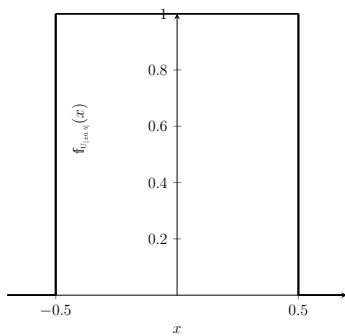
§05.06 **Beweis** von Lemma §05.05. In der Vorlesung Analysis 3.  $\square$

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

- (a) *Gleich-/Uniformverteilung*, kurz  $U_G$ , auf  $G \in \mathcal{B}$  mit Lebesgue-Maß  $\lambda(G) \in \mathbb{R}_0^+$ :

$$f_{U_G}(x) = \frac{1}{\lambda(G)} \mathbb{1}_G(x) \text{ für } x \in \mathbb{R}.$$

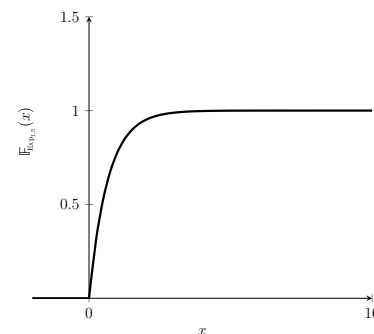
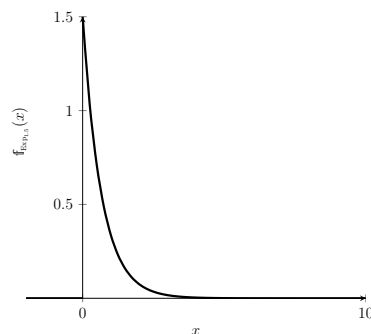
Runden wir Messungen reelle Größen auf die jeweils nächstgelegene ganze Zahl hin auf bzw. ab, so kann der Rundungsfehler durch eine Gleichverteilung  $U_{[\pm 0.5]}$  auf dem Intervall  $[\pm 0.5] := [-0.5, 0.5]$  mit Dichte  $f_{U_{[\pm 0.5]}}(x) = \mathbb{1}_{[\pm 0.5]}(x)$  und Verteilungsfunktion  $F_{U_{[\pm 0.5]}}(x) = (x + 0.5)\mathbb{1}_{[\pm 0.5]}(x) + \mathbb{1}_{\mathbb{R}_0^+}(x - 0.5)$  für  $x \in \mathbb{R}$  beschrieben werden:



- (b) *Exponentialverteilung*, kurz  $\text{Exp}_\lambda$ , mit Parametern  $\lambda \in \mathbb{R}_0^+$ :

$$f_{\text{Exp}_\lambda}(x) = \lambda \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x) \text{ für } x \in \mathbb{R}.$$

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



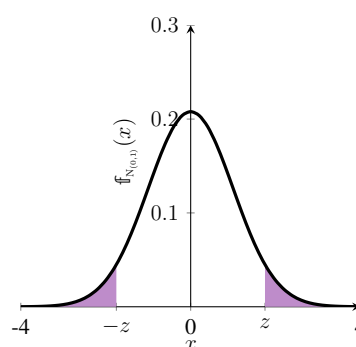
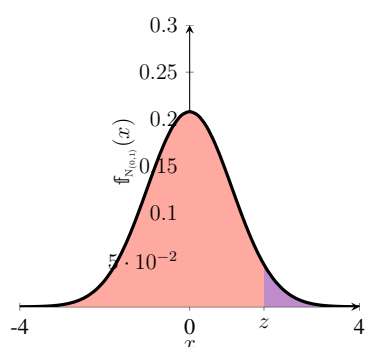
(c) *Normalverteilung*, kurz  $N_{(\mu, \sigma^2)}$ , mit Parametern  $\mu \in \mathbb{R}$  und  $\sigma^2 \in \mathbb{R}_0^+$ :

$$f_{N_{(\mu, \sigma^2)}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) \text{ für } x \in \mathbb{R}.$$

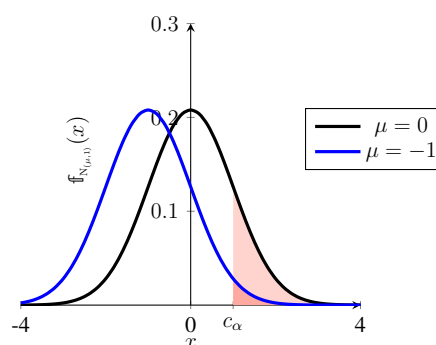
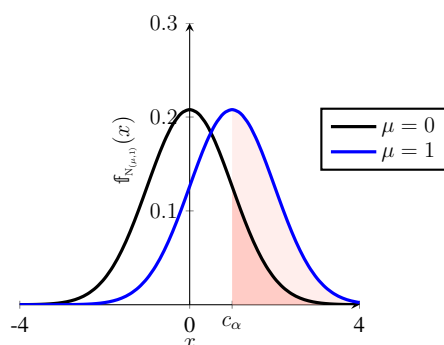
Die Verteilung gemittelter zufälliger Größen (z. Bsp. der jährliche Wasserverbrauch eines Haushaltes) kann häufig durch eine Normalverteilung gut approximiert werden (vgl. ??). Eine  $N_{(0,1)}$ -Verteilung heißt *Standardnormalverteilung*. Wir bezeichnen mit  $\Phi$  die Verteilungsfunktion einer Standardnormalverteilung,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

$$\Phi(z) = [N_{(0,1)}]((-\infty, z]) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx.$$

Für alle  $z \in \mathbb{R}^+$  gilt dann  $\Phi(-z) = 1 - \Phi(z)$ :



Weiterhin heißt für  $\alpha \in (0, 1)$  der eindeutig bestimmte Wert  $z_\alpha$  mit  $\alpha = \Phi(z_\alpha)$  das  $\alpha$ -Quantil einer Standardnormalverteilung. Typische Werte für Quantile der Standardnormalverteilung sind tabellarisiert, siehe Seite 85 im Anhang. Wir halten weiterhin fest, dass für  $\alpha \in (0, 1)$  und  $c_\alpha \in \mathbb{R}$  mit  $N_{(\mu, \sigma^2)}([c_\alpha, \infty)) = \alpha$  gilt  $N_{(\mu+c, \sigma^2)}([c_\alpha, \infty)) > \alpha$  für  $c > 0$  sowie  $N_{(\mu+c, \sigma^2)}([c_\alpha, \infty)) < \alpha$  für  $c < 0$ :



□

§05.08 **Definition.** Sind  $f_1, \dots, f_n$  Wahrscheinlichkeitsdichten auf  $\mathbb{R}$ , so heißt

$$f(x) := \prod_{i \in [n]} f_i(x_i) \text{ für } x = (x_i)_{i \in [n]} \in \mathbb{R}^n$$

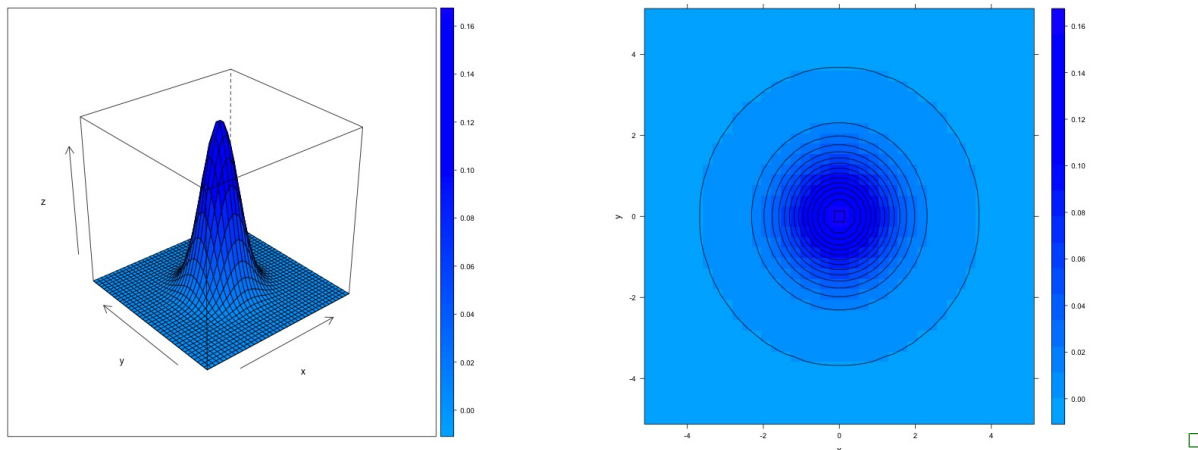
die Produktdichte der  $(f_i)_{i \in [n]}$  auf  $\mathbb{R}^n$ .

□

§05.09 **Beispiel.** Die *n-dimensionale Standard-Normalverteilung*  $N_{(0, E_n)}$  im  $\mathbb{R}^n$ , wobei  $E_n$  die  $n$ -dimensionale Einheitsmatrix und  $0$  den  $n$ -dimensionalen Nullvektor bezeichnet, ist definiert durch die Dichte

$$f_{N_{(0, E_n)}}(x) := \prod_{i \in \llbracket n \rrbracket} f_{N_{(0, 1)}}(x_i) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i \in \llbracket n \rrbracket} x_i^2\right) \text{ für } x \in \mathbb{R}^n;$$

Für  $n = 2$  erhalten wir die Dichte  $f_{N_{(0, E_2)}}(x_1, x_2) = (2\pi)^{-1} \exp(-\frac{1}{2}(x_1^2 + x_2^2))$  für  $x_1, x_2 \in \mathbb{R}$ :



## §06 Statistisches Modell

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

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

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

*diskret*, wenn der Stichprobenraum  $\mathcal{X}$  abzählbar und  $\mathcal{F} = 2^{\mathcal{X}}$  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  $(\mathbb{R}^n, \mathcal{B}^n)$  der Stichprobenraum ist und für jeden Parameter  $\theta \in \Theta$  ist  $\mathbb{P}_\theta$  ein stetiges Wahrscheinlichkeitsmaß auf  $(\mathbb{R}^n, \mathcal{B}^n)$ . In diesem Fall bezeichnet  $f_{\mathbb{P}_\Theta} := (f_{\mathbb{P}_\theta})_{\theta \in \Theta}$  die zur Familie  $\mathbb{P}_\Theta$  von stetigen Wahrscheinlichkeitsmaßen gehörige Familie von Wahrscheinlichkeitsdichten.  $\square$

§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 das männliche Geschlecht eines Konsumierenden, so beschreiben wir das zufällige Geschlecht eines Konsumierenden durch ein *Bernoulliverteilungsmodell*  $(\{0, 1\}, 2^{\{0,1\}}, (B_p)_{p \in [0,1]})$ , vgl. **Beispiele** §04.08 (c).
- Beispiel II: Die zufällige Anzahl beschädigter Schrauben in einem Beutel mit 50 Schrauben beschreiben wir durch ein *Binomialverteilungsmodell*  $([0, 50], 2^{[0,50]}, (\text{Bin}_{(50,p)})_{p \in [0,1]})$ , vgl. **Beispiele** §04.08 (d).
- Beispiel III Die zufällige Anzahl der eingegangenen Anrufe innerhalb einer Viertelstunde beschreiben wir durch ein *Poissonverteilungsmodell*  $(\mathbb{N}_0, 2^{\mathbb{N}_0}, (\text{Poi}_\lambda)_{\lambda \in \mathbb{R}_0^+})$ , vgl. **Beispiele** §04.08 (f).
- Beispiel IV Die zufällige Wartezeit an der U-Bahn Haltestelle an einem Tag (in Minuten) beschreiben wir durch ein *Uniformverteilungsmodell*  $(\mathbb{R}, \mathcal{B}, (U_{[0,\theta]})_{\theta \in \mathbb{R}_0^+})$ , vgl. **Beispiele** §05.07 (a).
- Beispiel V Die zufällige Lebensdauer einer Glühlampe (in Stunden) beschreiben wir durch ein *Exponentialverteilungsmodell*  $(\mathbb{R}, \mathcal{B}, (\text{Exp}_\lambda)_{\lambda \in \mathbb{R}_0^+})$ , vgl. **Beispiele** §05.07 (b).
- Beispiel VI Die zufälligen Fehler in gemessenen Werten beschreiben wir durch ein *Normalverteilungsmodell*  $(\mathbb{R}, \mathcal{B}, (N_{(\mu,\sigma^2)})_{(\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+})$ , vgl. **Beispiele** §05.07 (c).  $\square$

### §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*  $([0, 1000], 2^{[0,1000]}, (\text{Bin}_{(1000,p)})_{p \in [0,1]})$  beschreiben (formale Begründung in **Beispiel** §19.03, (a)). Zur (unrealistischen) Vereinfachung geht der Verbraucherservice zunächst davon aus, dass  $\{0.5, 0.7\}$  die Parametermenge ist, das heißt, das statistische Modell  $([0, 1000], 2^{[0,1000]}, (\text{Bin}_{(1000,0.5)}, \text{Bin}_{(1000,0.7)}))$  beschreibt adäquat das zufällige Experiment. In anderen Worten, entweder ist  $\mathbb{P}_0 = \text{Bin}_{(1000,0.5)}$  oder  $\mathbb{P}_1 = \text{Bin}_{(1000,0.7)}$  die wahre Verteilung.  $\square$
- §06.06 **Definition.** Sei  $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_0, \mathbb{P}_1))$  ein (binäres) statistisches Experiment. Die Entscheidung, ob die *Nullhypothese*  $H_0 : \{\mathbb{P}_0\}$  oder die *Alternativhypothese*  $H_1 : \{\mathbb{P}_1\}$  vorliegt, wird (*statistisches*) *Testproblem mit einfachen Hypothesen* genannt. Eine Entscheidungsfunktion  $\varphi : \mathcal{X} \rightarrow \{0, 1\}$ , also Entscheidungen nur anhand einer Stichprobe  $x$  aus dem Stichprobenraum  $\mathcal{X}$ , mit dem Ereignis  $\varphi^{-1}(\{1\}) \in \mathcal{F}$  aller Stichproben, die zu einer Entscheidung gegen die Nullhypothese  $H_0$ , also für die Alternativhypothese  $H_1$  führen, sowie dem Ereignis  $\varphi^{-1}(\{0\}) \in \mathcal{F}$  aller Stichproben, die zu einer Entscheidung für die Nullhypothese  $H_0$  führen, heißt (*statistischer*) *Hypothesentest*, kurz *Test*.  $\square$
- §06.07 **Sprechweise.** Die Sprechweise Testproblem mit *einfachen* Hypothesen bedeutet in diesem Zusammenhang, dass sowohl die Nullhypothese als auch die Alternativhypothese jeweils einelementig ist. Üblicherweise bezeichnen wir eine Entscheidung für die Alternativhypothese  $H_1$  als *Ablehnen der Nullhypothese*, wogegen eine Entscheidung für die Nullhypothese  $H_0$  *nicht Ablehnen der Nullhypothese* genannt wird. Für einen Test  $\varphi$  bezeichnen wir das Ereignis  $\mathcal{A} := \varphi^{-1}(\{1\}) \in \mathcal{F}$  aller Stichproben, die zum Ablehnen der Nullhypothese  $H_0$  führen, als *Ablehnbereich* des Tests  $\varphi$ . Da wir hier nur die zwei Möglichkeiten Ablehnen oder nicht Ablehnen für einen Test zulassen, ist  $\varphi = \mathbb{1}_{\mathcal{A}}$  eine Indikatorfunktion und  $\mathcal{A}^c = \varphi^{-1}(0) \in \mathcal{F}$



das Ereignis aller Stichproben, die nicht zu einem Ablehnen der Nullhypothese führen, genannt *Annahmebereich*, wobei definitionsgemäß  $\mathcal{A} \cup \mathcal{A}^c = \mathcal{X}$  gilt.  $\square$

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

§06.09 **Bemerkung**. Durch Angabe des Ablehnbereiches  $\mathcal{A}$  ist ein Test  $\varphi = \mathbb{1}_{\mathcal{A}}$  eindeutig festgelegt. Offensichtlich können in einem statistischen 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 Alternativhypothese  $H_1 : \{\mathbb{P}_1\}$  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  $\mathcal{A} \in \mathcal{F}$  gilt. Damit können wir den Ereignissen  $\mathcal{A}$  und  $\mathcal{A}^c$  Wahrscheinlichkeiten zuordnen.  $\square$

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

§06.11 **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  $\mathbb{P}_1(\mathcal{A}^c) = \mathbb{P}_1(\emptyset) = 0$ . Damit können wir im Allgemeinen 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.  $\square$

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

§06.13 **Bemerkung**. In der Definition eines besten  $\alpha$ -Tests werden die Fehler 1. und 2. Art nicht sym-

metrisch 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 Alternativhypothese dies widerspiegeln. Vereinfacht gesprochen, das Ziel ist es, die Nullhypothese abzulehnen, da nur dann die Wahrscheinlichkeit sich zu irren, immer klein ist.  $\square$

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

$$\mathcal{A}_k := \{x \in \mathcal{X} \mid \mathbb{P}_1(x) \geq k \mathbb{P}_0(x)\}$$

für einen *kritischen Wert*  $k \in \mathbb{R}^+$  heißt *Neyman-Pearson-Test*.  $\square$

§06.15 **Beispiel (Binomialverteilungsmodell §06.08 fortgesetzt).** Wir betrachten das Testproblem der einfachen Nullhypothese  $H_0 : \{\text{Bin}_{(1000, 0.5)}\}$  gegen die einfache Alternativhypothese  $H_1 : \{\text{Bin}_{(1000, 0.7)}\}$  im binären diskreten statistischen Experiment  $(\llbracket 0, 1000 \rrbracket, 2^{\llbracket 0, 1000 \rrbracket}, (\text{Bin}_{(1000, p)})_{p \in \{0.5, 0.7\}})$ . Setzen wir  $p = 0.5$  und  $q = 0.7$ , so ist für einen kritischem Wert  $k \in \mathbb{R}^+$  ein Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\mathcal{A}_k}$  gegeben durch den Ablehnbereich

$$\begin{aligned} \mathcal{A}_k &= \left\{ x \in \llbracket 0, 1000 \rrbracket \mid \binom{1000}{x} q^x (1-q)^{1000-x} \geq k \binom{1000}{x} p^x (1-p)^{1000-x} \right\} \\ &= \left\{ x \in \llbracket 0, 1000 \rrbracket \mid \left( \frac{q/(1-q)}{p/(1-p)} \right)^x \left( \frac{1-q}{1-p} \right)^{1000} \geq k \right\} \end{aligned}$$

Da  $q > p$  gilt, folgt  $\frac{q/(1-q)}{p/(1-p)} > 1$  und die Funktion  $L_{q,p}(x) := \left( \frac{q/(1-q)}{p/(1-p)} \right)^x \left( \frac{1-q}{1-p} \right)^{1000}$ ,  $x \in \mathbb{R}^+$  ist somit streng monoton wachsend. Damit gibt es zu jedem kritischen Wert  $k \in \mathbb{R}^+$  einen Wert  $x^* \in \llbracket 0, 1000 \rrbracket$ , derart dass  $\mathcal{A}_k = \llbracket x^*, 1000 \rrbracket$  gilt, so dass jeder Neyman-Pearson-Test durch einen Ablehnbereich dieser Form gegeben ist.  $\square$

§06.16 **Neyman-Pearson Lemma.** Sei  $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_0, \mathbb{P}_1))$  ein (binäres) diskretes statistisches Experiment. Für das Testproblem der einfachen Nullhypothese  $H_0 : \{\mathbb{P}_0\}$  gegen die einfache Alternativhypothese  $H_1 : \{\mathbb{P}_1\}$  ist jeder *Neyman-Pearson-Test*  $\varphi = \mathbb{1}_{\mathcal{A}_k}$  mit kritischem Wert  $k \in \mathbb{R}^+$  und Ablehnbereich  $\mathcal{A}_k \in \mathcal{F}$  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.  $\square$

§06.18 **Bemerkung.** Variieren wir den kritischen Wert  $k$  im Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\mathcal{A}_k}$ , so ändert sich auch der entsprechende Fehler 1. Art  $\mathbb{P}_0(\mathcal{A}_k)$ . Insbesondere gilt  $\mathbb{P}_0(\mathcal{A}_k) \downarrow \mathbb{P}_0(\mathbb{P}_0 = 0) = 0$  für  $k \rightarrow \infty$  und  $\mathbb{P}_0(\mathcal{A}_k) \uparrow \mathbb{P}_0(\{\mathbb{P}_1 > 0\} \cup \{\mathbb{P}_0 = 0\}) \in [0, 1]$  für  $k \rightarrow 0$ . Diese Änderung ist aber im Allgemeinen nicht stetig, so dass zu einem vorgegebenen Niveau  $\alpha \in (0, 1)$  nicht immer ein kritischer Wert  $k_\alpha$  und entsprechender Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\mathcal{A}_{k_\alpha}}$  mit  $\mathbb{P}_0(\mathcal{A}_{k_\alpha}) = \alpha$  gewählt werden kann. In diesem Fall wählen wir zu einem vorgegebenen Niveau  $\alpha \in (0, 1)$  den kritischen Wert

$$k_\alpha := \arg \max \{k \in \mathbb{R}^+ \mid \alpha \geq \mathbb{P}_0(\mathcal{A}_k)\}.$$

Offensichtlich gilt dann auch  $\mathbb{P}_0(\mathcal{A}_{k_\alpha}) \leq \alpha$  und der entsprechende Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\mathcal{A}_{k_\alpha}}$  mit Ablehnbereich  $\mathcal{A}_{k_\alpha}$  ist ein  $\alpha$ -Test, der aber im Allgemeinen nicht mehr ein bester  $\alpha$ -Test ist, wie wir in der Vorlesung Statistik I sehen werden.  $\square$

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

$$\text{Bin}_{(1000,0.5)}(\llbracket x^*, 1000 \rrbracket) = \sum_{x \in \llbracket x^*, 1000 \rrbracket} \binom{1000}{x} 0.5^x (1-0.5)^{1000-x} = 0.5^{1000} \sum_{x \in \llbracket x^*, 1000 \rrbracket} \binom{1000}{x}$$

wobei der entsprechende Fehler 2. Art

$$\text{Bin}_{(1000,0.7)}(\llbracket 0, x^* - 1 \rrbracket) = \sum_{x \in \llbracket 0, x^* - 1 \rrbracket} \binom{1000}{x} 0.7^x (1-0.7)^{1000-x}$$

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

$$x_\alpha := \arg \min \left\{ x^* \in \llbracket 0, 1000 \rrbracket \mid \alpha \geq \text{Bin}_{(1000,0.5)}(\llbracket x^*, 1000 \rrbracket) \right\}.$$

Dann ist der entsprechende Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\llbracket x_\alpha, 1000 \rrbracket}$  mit Ablehnbereich  $\llbracket x_\alpha, 1000 \rrbracket$  ein  $\alpha$ -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  $\varphi = \mathbb{1}_{\llbracket 538, 1000 \rrbracket}$  den Ablehnbereich  $\llbracket 538, 1000 \rrbracket$  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 Alternativhypothese, der Anteil der Konsumentinnen beträgt 70%, zum Niveau  $\alpha = 0.01$  abgelehnt werden kann. Offensichtlich hängt ein Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\llbracket x^*, 1000 \rrbracket}$  mit Ablehnbereich  $\llbracket x^*, 1000 \rrbracket$  nicht von dem Parameter  $q \in (p, 1]$  der Alternativhypothese ab. Damit ist der Neyman-Pearson-Test mit Ablehnbereich  $\llbracket 538, 1000 \rrbracket$  auch bester Test zum Niveau  $\alpha = \text{Bin}_{(1000,0.5)}(\llbracket 538, 1000 \rrbracket)$  der einfachen Nullhypothese  $H_0 : \{\text{Bin}_{(1000,0.5)}\}$  gegen die einfache Alternativhypothese  $H_1 : \{\text{Bin}_{(1000,0.55)}\}$  im binären diskreten statistischen Experiment  $(\llbracket 0, 1000 \rrbracket, 2^{\llbracket 0, 1000 \rrbracket}, (\text{Bin}_{(1000,p)})_{p \in \{0.5, 0.55\}})$ . Man beachte, dass der minimale Fehler 2. Art dann  $\text{Bin}_{(1000,0.55)}(\llbracket 0, 537 \rrbracket) \approx 0.213$  ist, und dieser hängt natürlich von der Alternativhypothese 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  $(\llbracket 0, n \rrbracket, 2^{\llbracket 0, n \rrbracket}, (\text{Bin}_{(n,p)})_{p \in \{0.5, 0.55\}})$  für das Testproblem der einfachen Nullhypothese  $H_0 : \{\text{Bin}_{(n,0.5)}\}$  gegen die einfache Alternativhypothese  $H_1 : \{\text{Bin}_{(n,0.55)}\}$  den Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\llbracket x_{\alpha,n}, n \rrbracket}$  mit Ablehnbereich  $\llbracket x_{\alpha,n}, n \rrbracket$ , derart dass

$$x_{\alpha,n} := \arg \min \{ x^* \in \llbracket 0, n \rrbracket \mid \alpha \geq \text{Bin}_{(n,0.5)}(\llbracket x^*, n \rrbracket) \},$$

wählen. Der Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\llbracket x_{\alpha,n}, n \rrbracket}$  ist dann ein  $\alpha$ -Test, sowie der beste Test zum Niveau  $\text{Bin}_{(n,0.5)}(\llbracket x_{\alpha,n}, n \rrbracket) \leq \alpha$ . 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  $\text{Bin}_{(500,0.5)}(\llbracket 277, 500 \rrbracket) \approx 0.0088$ ,  $\text{Bin}_{(1000,0.5)}(\llbracket 538, 1000 \rrbracket) \approx 0.0088$  und  $\text{Bin}_{(2000,0.5)}(\llbracket 1053, 2000 \rrbracket) \approx 0.0094$ , sowie Fehler 2. Art  $\text{Bin}_{(500,0.55)}(\llbracket 0, 276 \rrbracket) \approx 0.553$ ,  $\text{Bin}_{(1000,0.55)}(\llbracket 0, 537 \rrbracket) \approx 0.213$  und  $\text{Bin}_{(2000,0.55)}(\llbracket 0, 1052 \rrbracket) \approx 0.016$ , 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  $\varphi = \mathbb{1}_{\llbracket x_{\alpha,n}, n \rrbracket}$  mit Ablehnbereich  $\llbracket x_{\alpha,n}, n \rrbracket$  diese einhält, das heißt,

$$n_{\alpha,\beta} := \arg \min \{ n \in \mathbb{N} \mid \beta \geq \text{Bin}_{(n,0.55)}(\llbracket 0, x_{\alpha,n} \rrbracket) \}.$$

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  $\text{Bin}_{(2170,0.5)}(\llbracket 1140, 2170 \rrbracket) \approx 0.0096$  und der Fehler 2. Art  $\text{Bin}_{(2170,0.55)}(\llbracket 0, 1139 \rrbracket) \approx 0.00997$  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.  $\square$

§06.20 **Ausblick.** Betrachten wir für ein beliebiges  $p \in (0, 1)$ ,  $q \in (p, 1]$  und  $n \in \mathbb{N}$  das Testproblem der einfachen Nullhypothese  $H_0 : \{\text{Bin}_{(n,p)}\}$  gegen die einfache Alternativhypothese  $H_1 : \{\text{Bin}_{(n,q)}\}$  im diskreten statistischen Experiment  $(\llbracket 0, n \rrbracket, 2^{\llbracket 0, n \rrbracket}, (\text{Bin}_{(n,p)}, \text{Bin}_{(n,q)}))$ , so hängt der Ablehnbereich  $\llbracket x^*, n \rrbracket$  mit  $x^* \in \llbracket 0, n \rrbracket$  eines Neyman-Pearson-Tests  $\varphi = \mathbb{1}_{\llbracket x^*, n \rrbracket}$  nicht von dem Wert  $q \in (p, 1]$  der Alternativhypothese ab. Damit ist für jedes  $q \in (p, 1]$  ein Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\llbracket x^*, n \rrbracket}$  mit Ablehnbereich  $\llbracket x^*, n \rrbracket$  bester Test zum Niveau  $\text{Bin}_{(n,p)}(\llbracket x^*, n \rrbracket)$  der *einfachen Nullhypothese*  $H_0 : \{\text{Bin}_{(n,p)}\}$  gegen die *einfache Alternativhypothese*  $H_1 : \{\text{Bin}_{(n,q)}\}$ . Dies erlaubt uns, das diskrete statistische Experiment  $(\llbracket 0, n \rrbracket, 2^{\llbracket 0, n \rrbracket}, (\text{Bin}_{(n,q)})_{q \in [p, 1]})$  zu betrachten. Für jedes  $x^* \in \llbracket 0, n \rrbracket$  ist der Neyman-Pearson Test  $\varphi = \mathbb{1}_{\llbracket x^*, n \rrbracket}$  mit Ablehnbereich  $\llbracket x^*, n \rrbracket$  dann *gleichmäßig bester Test* zum Niveau  $\alpha = \text{Bin}_{(n,p)}(\llbracket x^*, n \rrbracket)$  der *einfachen Nullhypothese*  $H_0 : \{\text{Bin}_{(n,p)}\}$  gegen die *zusammengesetzte Alternativhypothese*  $H_1 : \{\text{Bin}_{(n,q)}, q \in (p, 1]\}$ , da der Fehler 2. Art für jeden anderen Test  $\varphi = \mathbb{1}_{\mathcal{A}}$  zum Niveau  $\alpha$  mit Ablehnbereich  $\mathcal{A}$  gleichmäßig nicht kleiner ist, das heißt,  $\text{Bin}_{(n,q)}(\llbracket x^*, n \rrbracket^c) \leq \text{Bin}_{(n,q)}(\mathcal{A}^c)$  für alle  $q \in (p, 1]$  gilt. Wir halten fest, dass diese Schlussfolgerung möglich ist, da für jedes  $p \in [0, 1]$  und  $q \in (p, 1]$  die Funktion  $L_{p,q}(x) := \left(\frac{q/(1-q)}{p/(1-p)}\right)^x \left(\frac{1-q}{1-p}\right)^n$ ,  $x \in \mathbb{R}^+$ , streng monoton wachsend ist. Bezeichnet  $\mathbb{p}_p$  die Zähldichte der Verteilung  $\text{Bin}_{(n,p)}$  für  $p \in [0, 1]$ , so gilt offensichtlich  $L_{p,q}(x) = \frac{\mathbb{p}_q(x)}{\mathbb{p}_p(x)}$ ,  $x \in \llbracket 0, n \rrbracket$  und  $L_{p,q}$  wird *Likelihood-Quotient* (oder *Dichtequotient*) genannt. Die Verteilungsfamilie  $(\text{Bin}_{(n,p)})_{p \in [0, 1]}$  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 Alternativhypothese, der Anteil der Konsumentinnen beträgt mehr als 50%, zum Niveau  $\alpha = 0.01$  ablehnen.  $\square$

# Kapitel 3

## Zufallsvariablen

### §07 Zufallsvariable

§07.01 **Definition.** Es seien  $(\Omega, \mathcal{A})$  und  $(\mathcal{S}, \mathcal{S})$  zwei messbare Räume. Eine Funktion  $X : \Omega \rightarrow \mathcal{S}$  heißt  $\mathcal{A}$ - $\mathcal{S}$ -messbar (kurz *messbar*), falls  $\sigma(X) := X^{-1}(\mathcal{S}) := \{X^{-1}(S) \mid S \in \mathcal{S}\} \subseteq \mathcal{A}$  gilt. Jede solche messbare Funktion wird  $((\mathcal{S}, \mathcal{S})$ -wertige) *Zufallsvariable* genannt.  $\square$

§07.02 **Schreibweisen.**  $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$ ,  $\{X \in S\} := X^{-1}(S) = \{\omega \in \Omega \mid X(\omega) \in S\}$  und  $\{X = x\} := X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}$ .  $\square$

§07.03 **Lemma.** Es seien  $X : \Omega \rightarrow \mathcal{S}$  eine Abbildung und  $\mathcal{S}$  eine  $\sigma$ -Algebra auf  $\mathcal{S}$ . Dann ist das Teilmengensystem  $\sigma(X) = X^{-1}(\mathcal{S})$  eine  $\sigma$ -Algebra auf  $\Omega$ , die sogenannte von  $X$  erzeugte  $\sigma$ -Algebra.

§07.04 **Beweis** von Lemma §07.03. Übungsaufgabe.  $\square$

§07.05 **Bemerkung.** Nach Lemma §07.03 ist  $X^{-1}(\mathcal{S})$  die kleinste  $\sigma$ -Algebra auf  $\Omega$ , sodass die Funktion  $X : \Omega \rightarrow \mathcal{S}$  eine Zufallsvariable ist.  $\square$

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

§07.07 **Beweis** von Lemma §07.06. In der Vorlesung.  $\square$

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

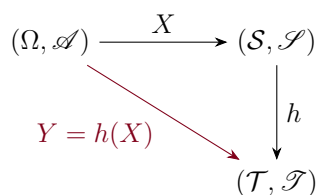
§07.09 **Beweis** von Lemma §07.08. Übungsaufgabe.  $\square$

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

§07.11 **Beweis** von Korollar §07.10. In der Vorlesung.  $\square$

§07.12 **Proposition.** Seien  $(\Omega, \mathcal{A})$ ,  $(\mathcal{S}, \mathcal{S})$  und  $(\mathcal{T}, \mathcal{T})$  drei messbare Räume. Sind  $X : \Omega \rightarrow \mathcal{S}$   $\mathcal{A}$ - $\mathcal{S}$ -messbar und  $h : \mathcal{S} \rightarrow \mathcal{T}$   $\mathcal{S}$ - $\mathcal{T}$ -messbar, so ist die Hintereinanderausführung  $h \circ X$   $\mathcal{A}$ - $\mathcal{T}$ -messbar, also  $Y = h(X)$  eine  $(\mathcal{T}, \mathcal{T})$ -wertige Zufallsvariable.

§07.13 **Beweis** von Proposition §07.12. In der Vorlesung.  $\square$

§07.14 **Skizze.**

□

## §08 Numerische und reelle Zufallsvariablen

§08.01 **Definition.** Sei  $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$  eine Zufallsvariable.

- (a) Falls  $(\mathcal{S}, \mathcal{S}) = (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ , so heißt  $X$  *numerische Zufallsvariable*, kurz  $X \in \overline{\mathcal{A}}$ .  
 Falls  $(\mathcal{S}, \mathcal{S}) = (\overline{\mathbb{R}^+}, \overline{\mathcal{B}^+})$ , so heißt  $X$  *positive numerische Zufallsvariable*, kurz  $X \in \overline{\mathcal{A}^+}$ .
- (b) Falls  $(\mathcal{S}, \mathcal{S}) = (\mathbb{R}, \mathcal{B})$ , so heißt  $X$  *reelle Zufallsvariable*, kurz  $X \in \mathcal{A}$ .  
 Falls  $(\mathcal{S}, \mathcal{S}) = (\mathbb{R}^+, \mathcal{B}^+)$ , so heißt  $X$  *positive reelle Zufallsvariable*, kurz  $X \in \mathcal{A}^+$ .
- (c) Falls  $(\mathcal{S}, \mathcal{S}) = (\mathbb{R}^n, \mathcal{B}^n)$ , so heißt  $X$  *Zufallsvektor*, kurz  $X \in \mathcal{A}^n$ . □

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

§08.03 **Beispiel.**

- (a) Seien  $(\Omega, \mathcal{A})$  ein messbarer Raum und  $1_A$  mit  $A \subseteq \Omega$  die Indikatorfunktion. Dann gilt  $1_A \in \mathcal{A}$ , d.h.  $1_A$  ist eine reelle Zufallsvariable, genau dann, wenn  $A \in \mathcal{A}$  gilt.
- (b) Sei  $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P})$  ein stetiger Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsdichte  $f : \mathbb{R}^n \rightarrow \mathbb{R}^+$ , dann ist  $f$  eine positive reelle Zufallsvariable. □

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

- (i) Eine Funktion  $X : \Omega \rightarrow \overline{\mathbb{R}}$  ist eine numerische Zufallsvariable genau dann, wenn für alle  $x \in \mathbb{R}$  gilt  $\{X \leq x\} = X^{-1}([-\infty, x]) \in \mathcal{A}$ .
- (ii) Eine Funktion  $X : \Omega \rightarrow \mathbb{R}^n$  ist ein Zufallsvektor 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 Zufallsvariablen  $X, Y \in \overline{\mathcal{A}}$  gilt:

- (i) Für alle  $a \in \mathbb{R}$  gilt  $aX \in \overline{\mathcal{A}}$  (mit der Konvention  $0 \times \infty = 0$ );
- (ii)  $X \vee Y, X \wedge Y \in \overline{\mathcal{A}}$  und insbesondere  $X^+ := X \vee 0, X^- := (-X)^+ \in \overline{\mathcal{A}^+}$  und  $|X| \in \overline{\mathcal{A}^+}$ ;
- (iii)  $\{X < Y\}, \{X \leq Y\}, \{X = Y\} \in \mathcal{A}$

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

§08.08 **Lemma.** Es seien  $X_1, \dots, X_n \in \mathcal{A}$  und  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  Borel-messbar. Dann ist  $h(X_1, \dots, X_n) \in \mathcal{A}^m$ . Insbesondere gilt also  $(X_1, \dots, X_n) \in \mathcal{A}^n$  und  $X_1 + X_2, X_1 - X_2, X_1 X_2 \in \mathcal{A}$  sowie, falls überall wohldefiniert,  $X_1/X_2 \in \mathcal{A}$ .



§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  $(\Omega, \mathcal{A})$ . Die Folge  $(X_n)_{n \in \mathbb{N}}$  heißt *(punktweise) monoton wachsend* (bzw. *fallend*), wenn  $X_n(\omega) \leq X_{n+1}(\omega)$  (bzw.  $X_{n+1}(\omega) \leq X_n(\omega)$ ) für alle  $\omega \in \Omega$  und für alle  $n \in \mathbb{N}$  gilt. Für jedes  $\omega \in \Omega$  definiere

$$[\liminf_{n \rightarrow \infty} X_n](\omega) := \sup_{n \geq 1} \inf_{m \geq n} X_m(\omega) := \sup \{ \inf \{ X_m(\omega) \mid m \in \mathbb{N} \wedge m \geq n \} \mid n \in \mathbb{N} \};$$

$$[\limsup_{n \rightarrow \infty} X_n](\omega) := \inf_{n \geq 1} \sup_{m \geq n} X_m(\omega) := \inf \{ \sup \{ X_m(\omega) \mid m \in \mathbb{N} \wedge m \geq n \} \mid n \in \mathbb{N} \}.$$

Dann heißen  $X_* := \liminf_{n \rightarrow \infty} X_n$  und  $X^* := \limsup_{n \rightarrow \infty} X_n$  *Limes inferior* bzw. *Limes superior*

der Folge  $(X_n)_{n \in \mathbb{N}}$ . Die Folge  $(X_n)_{n \in \mathbb{N}}$  heißt *konvergent*, wenn  $X_* = X^* =: X$  gilt, d.h. der punktweise Grenzwert existiert überall. In diesem Fall schreiben wir kurz  $\lim_{n \rightarrow \infty} X_n := X$ .

§08.11 **Bemerkung.**

- (i) Jede *monoton wachsende* (bzw. *fallende*) Folge  $(X_n)_{n \in \mathbb{N}}$  von numerischen oder reellen Zufallsvariablen ist konvergent mit  $X := \lim_{n \rightarrow \infty} X_n = \sup_{n \in \mathbb{N}} X_n$  (bzw.  $X := \lim_{n \rightarrow \infty} X_n = \inf_{n \in \mathbb{N}} X_n$ ). In diesem Fall schreiben wir kurz  $X_n \uparrow X$  (bzw.  $X_n \downarrow X$ ).
- (ii) Sei  $(A_n)_{n \in \mathbb{N}}$  Folge von Teilmengen aus  $\Omega$ . Für  $A_*$  und  $A^*$  wie in **Definition** §01.13 gilt dann  $\mathbb{1}_{A_*} = (\mathbb{1}_{A_n})_* = \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n}$  und  $\mathbb{1}_{A^*} = (\mathbb{1}_{A_n})^* = \limsup_{n \rightarrow \infty} \mathbb{1}_{A_n}$ . □

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

- (i)  $\sup_{n \in \mathbb{N}} X_n \in \overline{\mathcal{A}}$ ,  $\inf_{n \in \mathbb{N}} X_n \in \overline{\mathcal{A}}$ ,  $X_* = \liminf_{n \rightarrow \infty} X_n \in \overline{\mathcal{A}}$  und  $X^* = \limsup_{n \rightarrow \infty} X_n \in \overline{\mathcal{A}}$ ;
- (ii) Falls  $X := \lim_{n \rightarrow \infty} X_n$  existiert, dann ist  $X \in \overline{\mathcal{A}}$ .

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

## §09 Einfache Zufallsvariable

Eine reelle Zufallsvariable  $Y$ , die nur die Werte 0 oder 1 annimmt, wird **Beispiele** §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  $\mathbb{1}_A \leq \mathbb{1}_B$ .
- (ii) Für  $X : \Omega \rightarrow \mathcal{S}$  und  $S \subseteq \mathcal{S}$  gilt  $\mathbb{1}_{\{X \in S\}} = \mathbb{1}_{X^{-1}(S)} = \mathbb{1}_S \circ X = \mathbb{1}_S(X)$ .

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

§09.03 **Definition.** Eine numerische (bzw. reelle) Zufallsvariable  $X$  auf  $(\Omega, \mathcal{A})$  heißt *einfach (elementar)*, falls sie nur endlich viele verschiedene numerische (bzw. reelle) Werte annimmt, d.h. für  $X(\Omega) = \{x_i, i \in \mathcal{I}\}$  besitzt  $X$  eine Darstellung der Form

$$X = \sum_{i \in \mathcal{I}} x_i \mathbb{1}_{A_i} \quad \text{mit} \quad A_i := X^{-1}(\{x_i\}) = \{X = x_i\} \in \mathcal{A},$$

wobei  $\mathcal{I}$  endlich ist und alle  $x_i \in \overline{\mathbb{R}}$  (bzw.  $x_i \in \mathbb{R}$ ) verschieden sind.  $\square$

Einfache Zufallsvariablen erlauben numerische (bzw. reelle) Zufallsvariablen zu approximieren. Im Folgenden bezeichnen wir für  $a \in \overline{\mathbb{R}}^+$  mit  $\lfloor a \rfloor := \sup\{z \in \mathbb{Z} \mid z \leq a\}$ , d.h. für  $a < \infty$  ist  $\lfloor a \rfloor$  die nächstliegende nicht größere ganze Zahl.

§09.04 **Lemma.** Für jedes  $X \in \overline{\mathcal{A}}^+$  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  $\overline{\mathcal{A}}^+$ , derart dass gilt

- (i)  $X_n \uparrow X$ ;
- (ii)  $X_n \leq X \wedge n$ , d.h.  $X_n(\omega) \leq X(\omega) \wedge n$  für alle  $\omega \in \Omega$ ;
- (iii) Für jedes  $c \in \mathbb{R}^+$  gilt  $\lim_{n \rightarrow \infty} X_n = X$  gleichmäßig auf  $\{X \leq c\}$ , d.h.  $\lim_{n \rightarrow \infty} \sup_{\{\omega: X(\omega) \leq c\}} |X(\omega) - X_n(\omega)| = 0$ .

§09.05 **Beweis** von Lemma §09.04. In der Vorlesung.  $\square$

§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  $Y = Y^+ - Y^-$ .  $\square$

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

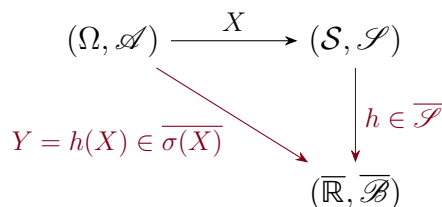
§09.07 **Proposition.** Seien  $(\Omega, \mathcal{A})$ ,  $(\mathcal{S}, \mathcal{S})$  zwei messbare Räume,  $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{S}, \mathcal{S})$  messbar und  $Y : \Omega \rightarrow \overline{\mathbb{R}}$ . Dann sind die folgenden Aussagen äquivalent:

- (i)  $Y$  ist messbar bzgl.  $\sigma(X) = X^{-1}(\mathcal{S})$ , kurz  $Y \in \overline{\sigma(X)}$ ;
- (ii) Es existiert eine messbare Funktion  $h : (\mathcal{S}, \mathcal{S}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ , kurz  $h \in \overline{\mathcal{S}}$ , derart dass  $Y = h(X)$  gilt.

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

§09.08 **Beweis** von Proposition §09.07. In der Vorlesung.  $\square$

§09.09 **Skizze.**



$\square$

## §10 Verteilung einer Zufallsvariablen

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



§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.  $\square$

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

§10.04 **Beweis von Lemma §10.03.** Übungsaufgabe.  $\square$

§10.05 **Definition.** Die Verteilung  $\mathbb{P}^X$  von  $X$ , kurz  $X \sim \mathbb{P}^X$ , wird auch Bildmaß von  $\mathbb{P}$  unter  $X$  genannt. Mit der Verteilungsfunktion  $\mathbb{F}^X$  (Dichte  $\mathbb{f}^X$ , Zähldichte  $\mathbb{p}^X$ ) 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  $(\mathcal{S}, \mathcal{S})$  ist. Eine reelle Zufallsvariable (Zufallsvektor)  $X$  mit stetigem Bildmaß  $\mathbb{P}^X$  wird *stetig-verteilt* genannt.  $\square$

§10.06 **Schreibweise.**

- (i)  $\mathbb{F}^X(x) = \mathbb{P}(X \leq x)$
- (ii) Bei Zufallsvariablen spielt der Ausgangsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  häufig keine Rolle und wird daher dann auch nicht angegeben.
- (iii) Ist die Verteilung einer reellen Zufallsvariable  $X$  zum Beispiel eine Normalverteilung  $N_{(\mu, \sigma^2)}$  (vgl. **Beispiele §04.08**), so schreiben wir kurz  $X \sim N_{(\mu, \sigma^2)}$ .  $\square$

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

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

§10.09 **Beweis von Satz §10.08.** In der Vorlesung Analysis 3.  $\square$

§10.10 **Korollar.**

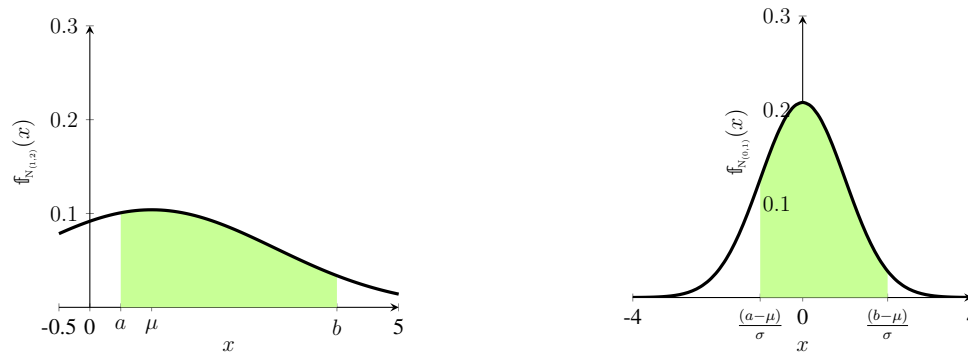
- (i) Ist  $X$  eine reelle Zufallsvariable mit Dichte  $\mathbb{f}^X$ , so besitzt  $Y = aX + b$  für  $a \in \mathbb{R}_{\setminus 0}$ ,  $b \in \mathbb{R}$  die Dichte  $\mathbb{f}^Y(y) = \frac{1}{|a|} \mathbb{f}^X(a^{-1}(y - b))$ .
- (ii) Ist  $X$  aufgefasst als Spaltenvektor ein  $n$ -dimensionaler Zufallsvektor mit Dichte  $\mathbb{f}^X$ , so besitzt  $Y = AX + b$  für reguläres  $A \in \mathbb{R}^{(n,n)}$  und  $b \in \mathbb{R}^{(n,1)}$  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.  $\square$

§10.12 **Beispiel.**

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

Für  $a, b \in \mathbb{R}$ ,  $a < b$ , gilt  $[N_{(\mu, \sigma^2)}](a, b) = \mathbb{P}(X \in (a, b)) = \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \mathbb{P}\left(Z \in \left[\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right)\right) = [N_{(0,1)}]\left(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$  (vgl. §§04.08 (c)):



- (b) Für  $X \sim N_{(0,1)}$  ist  $S = X^2$  eine  $\chi_1^2$ -verteilte Zufallsvariable, wobei die  $\chi^2$ -Verteilung mit einem Freiheitsgrad gegeben ist durch die Dichte  $f_{\chi_1^2}(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \mathbb{1}_{\mathbb{R}^+}(y)$ .
- (c) Ist  $X$  aufgefasst als Spaltenvektor ein  $n$ -dimensionaler standard-normalverteilter Zufallsvektor wie in Beispiel §05.09, so ist  $Y = \mu + AX$  mit  $\mu \in \mathbb{R}^n$  und regulärem  $A \in \mathbb{R}^{(n,n)}$  ein  $n$ -dimensionaler Zufallsvektor mit Dichte

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

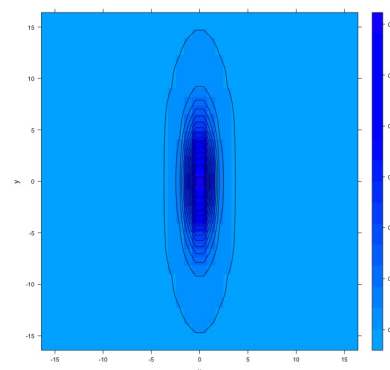
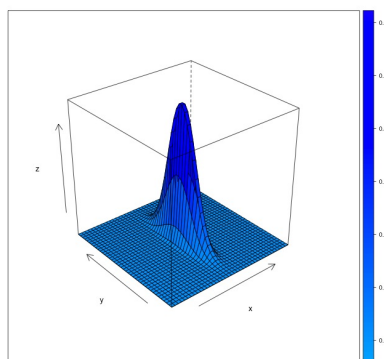
wobei  $\Sigma = AA^t$  eine symmetrische positive definite Matrix ist.  $Y$  heißt *normalverteilt* mit Vektor  $\mu \in \mathbb{R}^n$  und Matrix  $\Sigma \in \mathbb{R}^{(n,n)}$ , kurz  $Y \sim N_{(\mu, \Sigma)}$ .

- (d) Der stetig-verteilte Zufallsvektor  $(X, Y)$  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 die gemeinsame Dichte durch

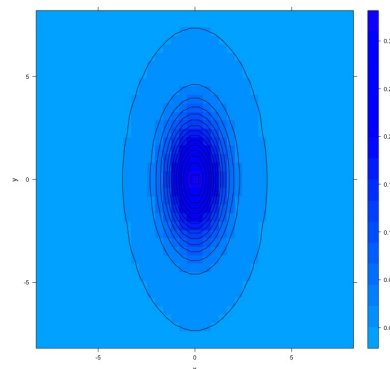
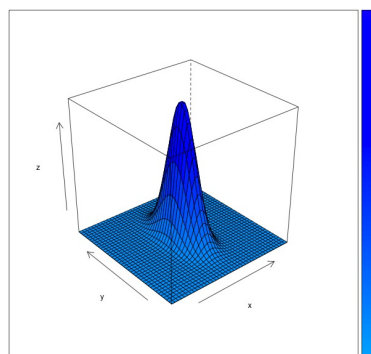
$$f^{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{(x-\mu_1)^2}{2(1-\rho^2)\sigma_1^2}\right) \exp\left(\frac{2\rho(x-\mu_1)(y-\mu_2)}{2(1-\rho^2)\sigma_1\sigma_2}\right) \exp\left(-\frac{(y-\mu_2)^2}{2(1-\rho^2)\sigma_2^2}\right)$$

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

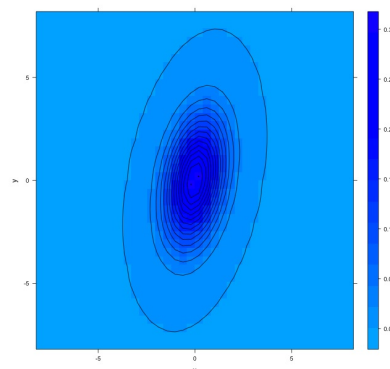
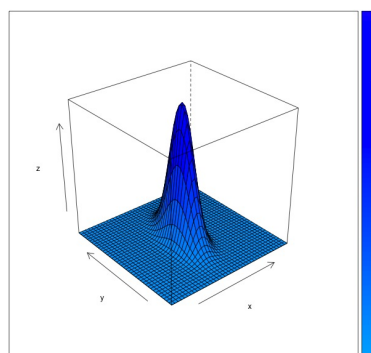
Für  $\mu_1 = 0 = \mu_2$ ,  $\sigma_1 = 1$ ,  $\sigma_2 = 4$  und  $\rho = 0$ :



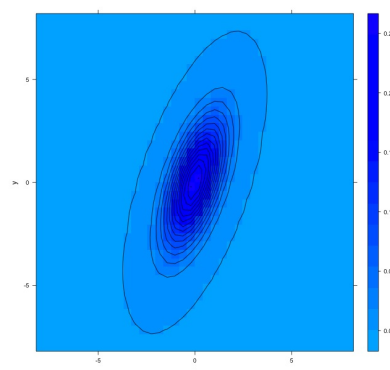
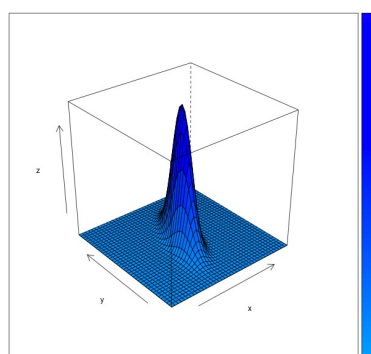
Für  $\mu_1 = 0 = \mu_2$ ,  $\sigma_1 = 1$ ,  $\sigma_2 = 2$  und  $\rho = 0$ :



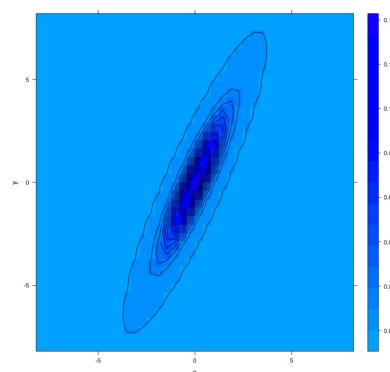
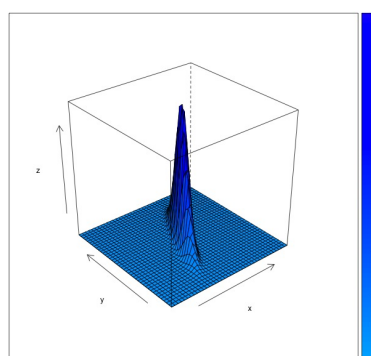
Für  $\mu_1 = 0 = \mu_2$ ,  $\sigma_1 = 1$ ,  $\sigma_2 = 2$  und  $\rho = 0,3$ :



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  $\rho = 0,9$ :



□

## §11 Verteilung einer Familie von Zufallsvariablen

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

§11.01 **Definition.** Für eine Familie  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  von Teil- $\sigma$ -Algebren aus  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$  bezeichnet

$$\bigwedge_{i \in \mathcal{I}} \mathcal{A}_i := \bigcap_{i \in \mathcal{I}} \mathcal{A}_i \quad \text{und} \quad \bigvee_{i \in \mathcal{I}} \mathcal{A}_i := \sigma\left(\bigcup_{i \in \mathcal{I}} \mathcal{A}_i\right)$$

die *größte  $\sigma$ -Algebra*, die in allen  $\mathcal{A}_i$ ,  $i \in \mathcal{I}$ , enthalten ist, bzw. die *kleinste  $\sigma$ -Algebra*, die alle  $\mathcal{A}_i$ ,  $i \in \mathcal{I}$  enthält.  $\square$

§11.02 **Erinnerung..** Die Menge  $\mathcal{S}_{\mathcal{I}} := \mathbf{X}_{i \in \mathcal{I}} \mathcal{S}_i$  aller Abbildungen  $(s_i)_{i \in \mathcal{I}} : \mathcal{I} \rightarrow \bigcup_{i \in \mathcal{I}} \mathcal{S}_i$  sodass  $s_i \in \mathcal{S}_i$  für alle  $i \in \mathcal{I}$  ist, heißt *Produktraum* oder kartesisches Produkt. Sind alle  $\mathcal{S}_i$  gleich, etwa  $\mathcal{S}_i = \mathcal{S}$ , dann schreiben wir  $\mathcal{S}^{\mathcal{I}} := \mathcal{S}_{\mathcal{I}}$ , im Fall  $n := |\mathcal{I}| < \infty$ , auch nur kurz  $\mathcal{S}^n := \mathcal{S}^{\mathcal{I}}$ .

§11.03 **Definition.** Für jedes  $j \in \mathcal{I}$  bezeichne  $\Pi_j : \mathcal{S}_{\mathcal{I}} \rightarrow \mathcal{S}_j$  mit  $(s_i)_{i \in \mathcal{I}} \mapsto s_j$  die Koordinatenabbildung. Die Produkt- $\sigma$ -Algebra  $\mathcal{I}_{\mathcal{I}} := \bigotimes_{i \in \mathcal{I}} \mathcal{I}_i$  auf dem Produktraum  $\mathcal{S}_{\mathcal{I}}$  ist die kleinste  $\sigma$ -Algebra, sodass für jedes  $j \in \mathcal{I}$  die Koordinatenabbildung  $\Pi_j : \mathcal{S}_{\mathcal{I}} \rightarrow \mathcal{S}_j$   $\mathcal{I}_{\mathcal{I}}$ - $\mathcal{I}_j$ -messbar ist, d.h.  $\mathcal{I}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathcal{I}_i := \bigvee_{j \in \mathcal{I}} \sigma(\Pi_j) = \bigvee_{j \in \mathcal{I}} \Pi_j^{-1}(\mathcal{I}_j)$ . Sind alle  $(\mathcal{S}_i, \mathcal{I}_i)$  gleich, etwa  $(\mathcal{S}, \mathcal{I}) = (\mathcal{S}, \mathcal{I})$ , dann schreiben wir  $\mathcal{I}^{\mathcal{I}} := \mathcal{I}_{\mathcal{I}}$ , im Fall  $n := |\mathcal{I}| < \infty$ , auch nur kurz  $\mathcal{I}^n := \mathcal{I}^{\mathcal{I}}$ . Ist für jedes  $i \in \mathcal{I}$  weiterhin  $\mathbb{P}_i$  ein Wahrscheinlichkeitsmaß auf  $(\mathcal{S}_i, \mathcal{I}_i)$ , dann heißt ein Wahrscheinlichkeitsmaß  $\mathbb{P}_{\mathcal{I}}$  auf  $\bigotimes_{i \in \mathcal{I}} \mathcal{I}_i$  *Produktmaß*, wenn für alle endlichen  $\mathcal{J} \subseteq \mathcal{I}$  und  $S_j \in \mathcal{I}_j$ ,  $j \in \mathcal{J}$  gilt:

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

In dem Fall schreiben wir  $\mathbb{P}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$ . Sind alle  $\mathbb{P}_i$  gleich, etwa  $\mathbb{P}_i = \mathbb{P}$ , dann schreiben wir kurz  $\mathbb{P}_{\mathcal{I}} = \mathbb{P}^{\mathcal{I}}$  und im Fall  $n := |\mathcal{I}| < \infty$  auch  $\mathbb{P}_{\mathcal{I}} = \mathbb{P}^n$ .  $\square$

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

§11.05 **Beweis von Lemma §11.04.** In der Vorlesung.  $\square$

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

§11.07 **Satz.** Für jede Familie  $(\mathcal{S}_i, \mathcal{I}_i, \mathbb{P}_i)_{i \in \mathcal{I}}$  von Wahrscheinlichkeitsräumen mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$  existiert stets ein eindeutiges Produktmaß  $\mathbb{P} = \bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$  auf  $(\mathcal{S}_{\mathcal{I}}, \mathcal{I}_{\mathcal{I}})$ .

§11.08 **Beweis von Satz §11.07.** In der Vorlesung Wahrscheinlichkeitstheorie I.  $\square$

§11.09 **Definition.** Das Bildmaß  $\mathbb{P}^X := \mathbb{P} \circ X^{-1}$  unter  $X := (X_i)_{i \in \mathcal{I}}$  auf  $(\mathcal{S}_{\mathcal{I}}, \mathcal{S}_{\mathcal{I}})$  heißt **gemeinsame Verteilung** der Familie  $(X_i)_{i \in \mathcal{I}}$ . Für jedes  $i \in \mathcal{I}$  wird das Bildmaß  $\mathbb{P}^{X_i} := \mathbb{P} \circ X_i^{-1} = \mathbb{P} \circ (\Pi_i \circ X)^{-1} = \mathbb{P}^X \circ \Pi_i^{-1}$  **Randverteilung (marginale Verteilung)** von  $X_i$  bzgl.  $\mathbb{P}^X$  genannt.  $\square$

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

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

$$\mathbb{F}^X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \quad \text{für alle } x \in \overline{\mathbb{R}}^n,$$

dann ist für jedes  $X_i$  die **Randverteilungsfunktion**

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

(ii) Ist  $X$  eine diskret-verteilte Zufallsvariable mit **gemeinsamer Zähldichte**  $\mathbb{p}^X : \mathbf{X}_{i \in \llbracket n \rrbracket} \mathcal{S}_i \rightarrow [0, 1]$ , dann ist jedes  $X_i$  diskret-verteilt mit **Randzähldichte**

$$\mathbb{p}^{X_i}(s_i) = \sum_{s_1 \in \mathcal{S}_1} \cdots \sum_{s_{i-1} \in \mathcal{S}_{i-1}} \sum_{s_{i+1} \in \mathcal{S}_{i+1}} \cdots \sum_{s_n \in \mathcal{S}_n} \mathbb{p}^X(s_1, \dots, s_n) \quad \text{für alle } s_i \in \mathcal{S}_i.$$

(iii) Ist  $X$  ein stetig-verteilter Zufallsvektor mit **gemeinsamer Dichte**  $\mathbb{f}^X : \mathbb{R}^n \rightarrow \mathbb{R}^+$ , dann ist jedes  $X_i$  stetig-verteilt mit **Randdichte**

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

§11.11 **Beweis** von **Satz** §11.10. In der Vorlesung.  $\square$

§11.12 **Beispiel.**

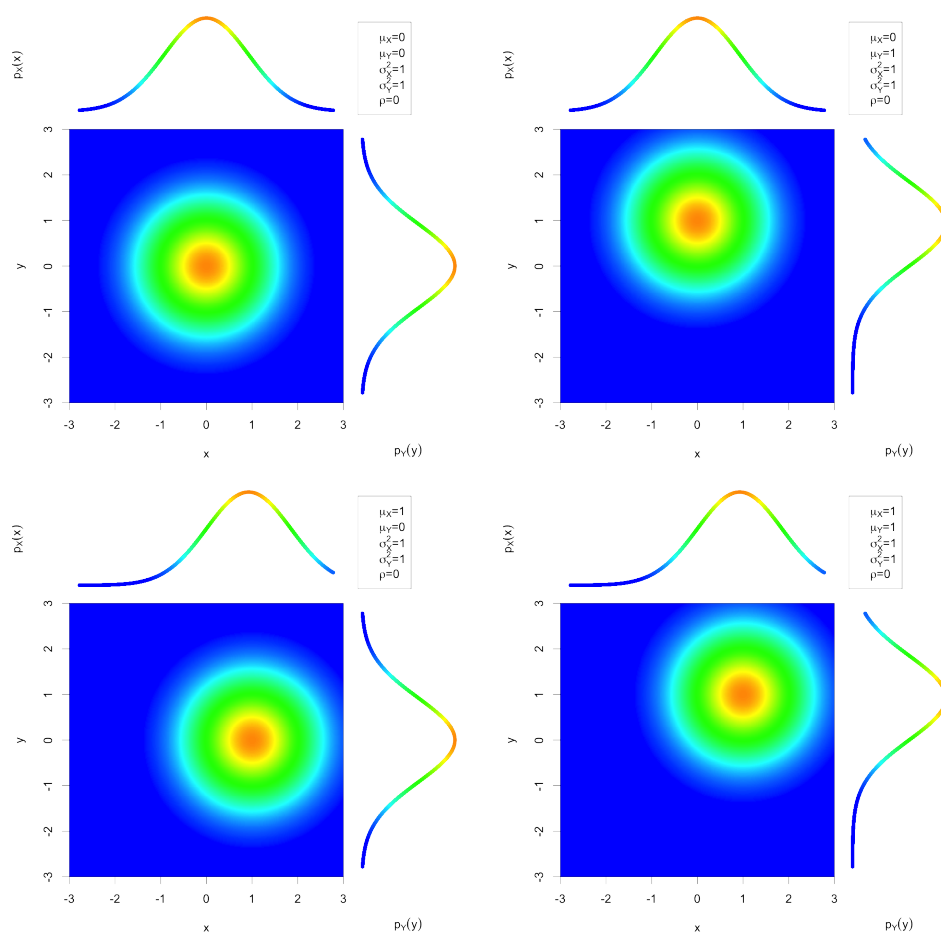
(a) Betrachten wir den Wurf von zwei fairen Würfeln (mit Gleichverteilung). Dann sei  $X$  der Absolutbetrag der Differenz der Augenzahlen und  $Y$  die Summe der Augenzahlen. Dann ist  $Z = (X, Y)$  diskret-verteilt mit gemeinsamer Zähldichte und Randzähldichten:

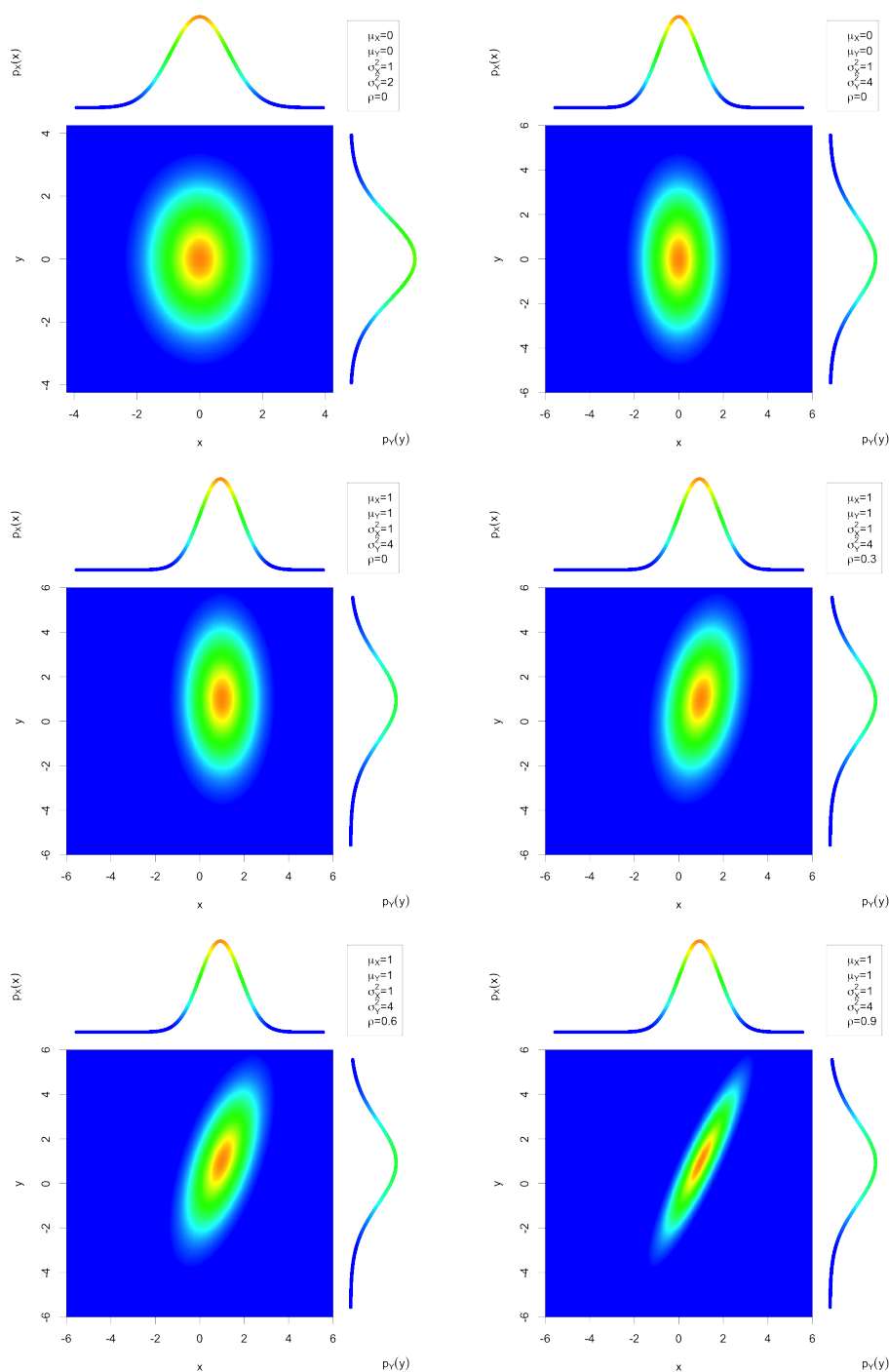
|                     |   | $y$            |                |                |                |                |                |                |                |                |                |                |                   |  |
|---------------------|---|----------------|----------------|----------------|----------------|----------------|--------------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

|                      |                | $y$            |                |                |                |                 |                 |                 |                 |                 |                 |                 |                 |                   |  |
|----------------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------------------|--|
| $\mathbb{F}^Z(x, y)$ |                | $(-\infty, 2)$ | $[2, 3)$       | $[3, 4)$       | $[4, 5)$       | $[5, 6)$        | $[6, 7)$        | $[7, 8)$        | $[8, 9)$        | $[9, 10)$       | $[10, 11)$      | $[11, 12)$      | $[12, \infty)$  | $\mathbb{F}^X(x)$ |  |
| $x$                  | $(-\infty, 0)$ | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0                 |  |
|                      | $[0, 1)$       | 0              | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{2}{36}$  | $\frac{3}{36}$  | $\frac{3}{36}$  | $\frac{4}{36}$  | $\frac{4}{36}$  | $\frac{5}{36}$  | $\frac{5}{36}$  | $\frac{6}{36}$  | $\frac{6}{36}$    |  |
|                      | $[1, 2)$       | 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  $\mu_1, \mu_2 \in \mathbb{R}$ ,  $\sigma_1, \sigma_2 \in \mathbb{R}_0^+$  und  $\rho \in (-1, 1)$  dann sind die Randverteilungen  $X \sim N(\mu_1, \sigma_1^2)$  und  $Y \sim N(\mu_2, \sigma_2^2)$ . Die nächsten Graphiken stellen die gemeinsamen sowie die marginalen Dichten für verschiedene Werte der Parameter dar.





(c) Für  $\mu \in \mathbb{R}$  und  $\sigma \in \mathbb{R}_0^+$  besitzt das Produktmaß  $N_{(\mu, \sigma^2)}^n$  auf  $(\mathbb{R}^n, \mathcal{B}^n)$  die Dichte

$$\prod_{i \in \llbracket n \rrbracket} \mathbb{f}_{N_{(\mu, \sigma^2)}}(x_i) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i \in \llbracket n \rrbracket} (x_i - \mu)^2\right) \text{ für } x \in \mathbb{R}^n.$$

Andererseits, seien  $\mathbf{1}_n := (1)_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n$  und  $E_n \in \mathbb{R}^{(n, n)}$  die Einheitsmatrix, sodass  $\mu \mathbf{1}_n = (\mu)_{i \in \llbracket n \rrbracket}$  und  $\sigma^2 E_n = \text{Diag}(\sigma^2 \mathbf{1}_n)$  die Diagonalmatrix mit Diagonaleinträgen  $\sigma^2 \mathbf{1}_n$  ist. Die Normalverteilung  $N_{(\mu \mathbf{1}_n, \sigma^2 E_n)}$  auf  $(\mathbb{R}^n, \mathcal{B}^n)$  besitzt dann die Dichte (vgl. [Beispiel §10.12](#))



(c))

$$f_{N_{(\mu \mathbb{1}_n, \sigma^2 E_n)}}(x) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i \in [n]} (x_i - \mu)^2\right) \text{ für } x \in \mathbb{R}^n,$$

sodass  $N_{(\mu \mathbb{1}_n, \sigma^2 E_n)} = N_{(\mu, \sigma^2)}^n$ , da sie dieselbe Dichte besitzen.  $\square$

## §12 Statistische Inferenz

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

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

§12.03 **Beispiel (Normalverteilungsmodell).** Wir betrachten wie in **Beispiel §11.12 (c)** eine Familie von Normalverteilungen  $(N_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$  auf  $(\mathbb{R}^n, \mathcal{B}^n)$ . Wir sagen ein  $\mathbb{R}^n$ -wertiger Spaltenvektor  $X$  ist *normalverteilt mit unbekannten Parametern  $\mu$  und  $\sigma$* , wenn  $X \odot (N_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$ . In dieser Situation sind typischerweise  $\mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}$  mit  $(\mu, \sigma) \mapsto \mu$  oder  $\mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$  mit  $(\mu, \sigma) \mapsto \sigma$  interessierende Parameter. Andererseits sagen wir,  $X$  ist *normalverteilt mit unbekanntem Parameter  $\mu$  und bekanntem Parameter  $\sigma$* , wenn  $X \odot (N_{(\mu, \sigma^2)}^n)_{\mu \in \mathbb{R}}$ , oder auch *normalverteilt mit unbekanntem Parameter  $\sigma$  und bekanntem Parameter  $\mu$* , wenn  $X \odot (N_{(\mu, \sigma^2)}^n)_{\sigma \in \mathbb{R}_0^+}$ . In jedem dieser Fälle ist das arithmetische Mittel  $\bar{X}_n := \frac{1}{n} \sum_{i \in [n]} X_i$  eine reelle Statistik, und für  $X \sim N_{(\mu, \sigma^2)}^n$  bezeichnen wir mit  $\mathbb{P}_{\mu, \sigma}^{\bar{X}_n}$  die von  $\bar{X}_n$  auf  $(\mathbb{R}, \mathcal{B})$  induzierte Verteilung.  $\square$

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

### §12.01 Hypothesentest

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



- (a)  $\mathcal{H}_{\gamma_o}^0 = (a, \gamma_o]$  und  $\mathcal{H}_{\gamma_o}^1 = (\gamma_o, b)$  (bzw.  $\mathcal{H}_{\gamma_o}^0 = [\gamma_o, b)$  und  $\mathcal{H}_{\gamma_o}^1 = (a, \gamma_o)$ ) für ein  $\gamma_o \in \Gamma$  wird das Testproblem *einseitig* genannt und wir schreiben es auch in der Form: Nullhypothese  $H_0 : \gamma \leq \gamma_o$  gegen Alternativhypothese  $H_1 : \gamma > \gamma_o$  (bzw.  $H_0 : \gamma \geq \gamma_o$  gegen  $H_1 : \gamma < \gamma_o$ ).
- (b)  $\mathcal{H}_{\gamma_o}^0 = \{\gamma_o\}$  und  $\mathcal{H}_{\gamma_o}^1 = (a, \gamma_o) \cup (\gamma_o, b)$  für ein  $\gamma_o \in \Gamma$  wird das Testproblem *zweiseitig* genannt und wir schreiben es auch in der Form: Nullhypothese  $H_0 : \gamma = \gamma_o$  gegen Alternativhypothese  $H_1 : \gamma \neq \gamma_o$ .  $\square$

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

§12.06 **Beispiel (Normalverteilungsmodell §12.03 fortgesetzt).** Sei  $X \odot (N_{(\mu, \sigma^2)}^n)_{\mu \in \mathbb{R}}$ , also ein normalverteilter  $\mathbb{R}^n$ -wertiger Spaltenvektor mit unbekanntem Parameter  $\mu$  und bekanntem Parameter  $\sigma$ . Für  $\mu_o \in \mathbb{R}$  betrachten wir das einseitige Testproblem der Nullhypothese  $H_0 : \mu \leq \mu_o$  gegen die Alternativhypothese  $H_1 : \mu > \mu_o$ .  $\square$

§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  $H_0 : \mathcal{H}^0$  wird abgelehnt, also der Ablehnbereich  $\mathcal{A}$  tritt ein, obwohl das statistische Experiment  $(\mathcal{X}, \mathcal{F}, \mathbb{P}_{\gamma^{-1}(\mathcal{H}^0)})$  adäquat ist, oder die Nullhypothese wird nicht gegen die Alternativhypothese  $H_1 : \mathcal{H}^1$  abgelehnt, also der Annahmehereich  $\mathcal{A}^c$  tritt ein, obwohl  $(\mathcal{X}, \mathcal{F}, \mathbb{P}_{\gamma^{-1}(\mathcal{H}^1)})$  adäquat ist.  $\square$

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

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

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

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

§12.09 **Beispiel (Binomialverteilungsmodell §06.19 fortgesetzt).** Betrachte für beliebige  $p_o \in (0, 1)$  und  $n \in \mathbb{N}$  erneut das *einseitigen Testproblem* der Nullhypothese  $H_0 : p \leq p_o$  gegen die Alternativhypothese  $H_1 : p > p_o$  im Binomialverteilungsmodell  $(\llbracket 0, n \rrbracket, 2^{\llbracket 0, n \rrbracket}, (\text{Bin}_{(n,p)})_{p \in (0,1)})$ , so ist  $\llbracket x^*, n \rrbracket$  mit  $x^* \in \llbracket 0, n \rrbracket$  der Ablehnbereich eines Neyman-Pearson Tests  $\varphi = \mathbb{1}_{\llbracket x^*, n \rrbracket}$ , der für das Testproblem der *einfachen Nullhypothese*  $H_0 : p = p_o$  gegen die *zusammengesetzte Alternativhypothese*  $H_1 : p > p_o$  ein gleichmäßig bester Test zum Niveau  $\alpha = \text{Bin}_{(n,p_o)}(\llbracket x^*, n \rrbracket)$  ist (vgl. *Ausblick* §06.20). Wir zeigen in der Vorlesung Statistik I, dass für alle  $p < p_o$  auch gilt

$\text{Bin}_{(n,p)}(\llbracket x^*, n \rrbracket) \leq \alpha$ . Somit ist ein Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\llbracket x^*, n \rrbracket}$  mit Ablehnbereich  $\llbracket x^*, n \rrbracket$  auch ein *gleichmäßig bester Test* zum Niveau  $\alpha = \text{Bin}_{(n,p_0)}(\llbracket x^*, n \rrbracket)$  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 Alternativhypothese, der Anteil der Konsumentinnen beträgt mehr als 50%, zum Niveau  $\alpha = 0.01$  ablehnen.  $\square$

§12.10 **Definition.** Sei  $(\mathbb{R}^n, \mathcal{B}^n, (\mathbb{P}_0, \mathbb{P}_1))$  ein (binäres) stetiges statistisches Experiment mit entsprechenden Wahrscheinlichkeitsdichten  $f_0$  und  $f_1$ . Jeder Test  $\varphi = \mathbb{1}_{\mathcal{A}_k}$  mit Ablehnbereich der Form

$$\mathcal{A}_k := \{x \in \mathbb{R}^n \mid f_1(x) \geq k f_0(x)\} = \{f_1 \geq k f_0\}$$

für einen *kritischen Wert*  $k \in \mathbb{R}^+$  heißt *Neyman-Pearson-Test*.  $\square$

§12.11 **Bemerkung.** Da  $f_0$  und  $f_1$  reelle Zufallsvariablen sind, gilt  $\mathcal{A}_k = \{f_1 \geq k f_0\} \in \mathcal{B}^n$  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.  $\square$

§12.12 **Beispiel (Normalverteilungsmodell §12.06 fortgesetzt).** Für  $\sigma \in \mathbb{R}_0^+$  und  $\mu_0, \mu_1 \in \mathbb{R}$  mit  $\mu_0 < \mu_1$  betrachte im binären stetigen statistischen Experiment  $(\mathbb{R}^n, \mathcal{B}^n, (\mathbb{N}_{(\mu, \sigma^2)}^n)_{\mu \in \{\mu_0, \mu_1\}})$  zunächst das Testproblem der einfachen Nullhypothese  $H_0 : \{\mathbb{N}_{(\mu_0, \sigma^2)}^n\}$  gegen die einfache Alternativhypothese  $H_1 : \{\mathbb{N}_{(\mu_1, \sigma^2)}^n\}$ . Setzen wir  $\bar{x}_n := \frac{1}{n} \sum_{i \in [n]} x_i$  für  $x \in \mathbb{R}^n$ , so ist ein Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\mathcal{A}_k}$  für einen kritischem Wert  $k \in \mathbb{R}^+$  gegeben durch den Ablehnbereich

$$\begin{aligned} \mathcal{A}_k &= \left\{x \in \mathbb{R}^n \mid f_{\mathbb{N}_{(\mu_1, \sigma^2)}^n}(x) \geq k f_{\mathbb{N}_{(\mu_0, \sigma^2)}^n}(x)\right\} \\ &= \left\{x \in \mathbb{R}^n \mid \exp\left(\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2\right) \geq k\right\} \\ &= \left\{x \in \mathbb{R}^n \mid \exp\left(\frac{n(\mu_1 - \mu_0)}{\sigma^2} \bar{x}_n + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2}\right) \geq k\right\}. \end{aligned}$$

Für  $\mu_1 > \mu_0$  ist die Funktion  $L_{\mu_1, \mu_0}(x) := \exp\left(\frac{n(\mu_1 - \mu_0)}{\sigma^2} \bar{x}_n + \frac{n(\mu_0^2 - \mu_1^2)}{2\sigma^2}\right)$ ,  $x \in \mathbb{R}^n$ , streng monoton wachsend in  $\bar{x}_n$ . Damit gibt es zu jedem kritischen Wert  $k \in \mathbb{R}^+$  einen Wert  $c^* \in \mathbb{R}$  derart, dass  $\mathcal{A}_k = \{x \in \mathbb{R}^n \mid \bar{x}_n \geq c^*\} = \{\bar{X}_n \geq c^*\}$  für die reelle Statistik  $\bar{X}_n : \mathbb{R}^n \rightarrow \mathbb{R}$  mit  $x \mapsto \bar{x}_n$  gilt. Wir halten fest, dass in einem Normalverteilungsmodell ein Neyman-Pearson-Test  $\varphi = \mathbb{1}_{\{\bar{X}_n \geq c^*\}}$  durch einen Ablehnbereich der Form  $\{\varphi = 1\} = \{\bar{X}_n \geq c^*\}$  für ein  $c^* \in \mathbb{R}$  gegeben ist. Insbesondere, gilt also  $\mathbb{N}_{(\mu_0, \sigma^2)}^n(\varphi = 1) = \mathbb{P}_{\mu_0, \sigma^2}^{\bar{X}_n}([c^*, \infty))$ .  $\square$

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

§12.14 **Beweis von Satz §12.13.** Übungsaufgabe.  $\square$

§12.15 **Beispiel (Normalverteilungsmodell §12.12 fortgesetzt).** Betrachten wir für  $\sigma \in \mathbb{R}_0^+$  (und  $n = 1$ ) im stetigen statistischen Experiment  $X \odot (\mathbb{N}_{(\mu, \sigma^2)})_{\mu \in \mathbb{R}}$  mit beliebigem  $\mu_0 \in \mathbb{R}$  das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu \leq \mu_0$  gegen die Alternativhypothese  $H_1 : \mu > \mu_0$ , so hängt der Ablehnbereich  $\{\varphi_{c^*} = 1\} = \{X \geq c^*\}$  mit  $c^* \in \mathbb{R}$  eines Neyman-Pearson-Tests

$\varphi_{c^*} = \mathbb{1}_{\{X \geq c^*\}}$  nicht von der Alternativhypothese ab (vgl. **Beispiel §12.12**), da die Verteilungsfamilie  $(N_{(\mu, \sigma^2)})_{\mu \in \mathbb{R}}$  einen *monotonen Likelihood-Quotienten* besitzt. Damit ist der Neyman-Pearson-Test  $\varphi_{c^*}$  für das Testproblem der *einfachen Nullhypothese*  $H_0 : \mu = \mu_o$  gegen die *zusammengesetzte Alternativhypothese*  $H_1 : \mu > \mu_o$  ein gleichmäßig bester Test zum Niveau  $N_{(\mu_o, \sigma^2)}(\varphi_{c^*} = 1) = N_{(\mu_o, \sigma^2)}([c^*, \infty)) = N_{(0,1)}([\frac{c^* - \mu_o}{\sigma}, \infty)) = \Phi(\frac{c^* - \mu_o}{\sigma})$  (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  $\alpha = \Phi(\frac{-c_\alpha + \mu_o}{\sigma})$  gilt. In der Tat bezeichnet  $z_\alpha$  das  $\alpha$ -Quantil einer Standardnormalverteilung (vgl. **Beispiele §05.07 (c)**), also  $\alpha = \Phi(z_\alpha)$ , so ist  $c_\alpha = \mu_o - \sigma z_\alpha$  der gesuchte kritische Wert. Weiterhin gilt  $N_{(\mu, \sigma^2)}([c_\alpha, \infty)) < \alpha$  für alle  $\mu < \mu_o$  (vgl. **Beispiele §05.07 (c)**). Somit ist  $\varphi_{c_\alpha} = \mathbb{1}_{\{X \geq \mu_o - \sigma z_\alpha\}}$  für jedes  $\alpha \in (0, 1)$  ein Neyman-Pearson-Test mit Ablehnbereich  $\{X \geq \mu_o - \sigma z_\alpha\}$ , der ein *gleichmäßig bester Test* zum Niveau  $\alpha$  des *einseitigen Testproblems* der Nullhypothese  $H_0 : \mu \leq \mu_o$  gegen die Alternativhypothese  $H_1 : \mu > \mu_o$  ist.  $\square$

## §12.02 Schätzfunktion

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

§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  $(\mathbb{N}, 2^{\mathbb{N}}, (\text{Lap}_{[1,N]})_{N \in \mathbb{N}})$  mit zugehöriger Familie von Zähldichten  $(p_N)_{N \in \mathbb{N}}$  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  $\hat{N}(x)$  denjenigen Parameter  $N \in \mathbb{N}$  zu wählen, für den die Wahrscheinlichkeit  $p_N(x)$  des Eintreten von  $x$  maximiert wird, d.h.

$$\begin{aligned} \hat{N}(x) &\in \arg \max_{N \in \mathbb{N}} \{p_N(x)\} := \left\{ N \in \mathbb{N} \mid p_N(x) \geq p_{\tilde{N}}(x) \text{ für alle } \tilde{N} \in \mathbb{N} \right\} \\ &= \arg \max_{N \in \mathbb{N}} \left\{ \frac{1}{N} \mathbb{1}_{[1,N]}(x) \right\} = \{x\}. \end{aligned}$$

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

§12.18 **Erinnerung.** Ist  $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\theta)$  ein diskretes bzw. stetiges statistisches Experiment mit zugehöriger Familie von Zähldichten bzw. Wahrscheinlichkeitsdichten  $\mathbb{P}_\theta := (p_\theta)_{\theta \in \Theta}$ , so ist für jedes  $\theta \in \Theta$  die Funktion  $p_\theta : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$  messbar, so dass wir  $p_\theta$  im Folgenden als reelle Statistik auffassen.  $\square$

§12.19 **Definition.** Sei  $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\theta)$  ein diskretes bzw. stetiges statistisches Experiment mit zugehöriger Familie von Zähldichten bzw. Wahrscheinlichkeitsdichten  $(p_\theta)_{\theta \in \Theta}$ . Für jedes  $\theta \in \Theta$  betrachten wir die reelle Statistik  $L(\theta) : \mathcal{X} \rightarrow \mathbb{R}^+$  mit  $x \mapsto [L(\theta)](x) := L(\theta, x) := p_\theta(x)$ . Die (zufällige) Funktion  $\theta \mapsto L(\theta)$  auf  $\Theta$  wird *Likelihood-Funktion* genannt.  $\square$

§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*.  $\square$

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

§12.22 **Bemerkung.** Man beachte, dass  $L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$  meint  $[L(\hat{\theta}(x))](x) = \sup_{\theta \in \Theta} [L(\theta)](x)$  für alle  $x \in \mathcal{X}$ . Weiterhin sind weder Existenz noch Eindeutigkeit eines MLS ohne Weiteres garantiert. Bei Mehrdeutigkeit wählt man üblicherweise einen maximierenden Parameter  $\theta$  nach Belieben.  $\square$

§12.23 **Beispiel.**

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

$$\arg \max_{p \in [0,1]} \left\{ \mathbb{P}_{\text{Bin}_{(n,p)}}(x) \right\} = \arg \max_{p \in [0,1]} \left\{ \binom{n}{x} p^x (1-p)^{n-x} \right\} = \left\{ \frac{x}{n} \right\}$$

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

(b) *Poissonverteilungsmodell*,  $(\mathbb{N}_0^n, 2^{\mathbb{N}_0^n}, (\text{Poi}_\lambda^n)_{\lambda \in \mathbb{R}_0^+})$  mit Produktzähldichten  $(\mathbb{P}_{\text{Poi}_\lambda^n})_{\lambda \in \mathbb{R}_0^+}$ :

$$\arg \max_{\lambda \in \mathbb{R}_0^+} \left\{ \mathbb{P}_{\text{Poi}_\lambda^n}(x) \right\} = \arg \max_{\lambda \in \mathbb{R}_0^+} \left\{ \frac{\lambda^{\sum_{i \in [n]} x_i}}{\prod_{i \in [n]} (x_i!)} e^{-n\lambda} \right\} = \left\{ \frac{1}{n} \sum_{i \in [n]} x_i \right\}.$$

Die Statistik  $\hat{\lambda} : \mathbb{N}_0^n \rightarrow \mathbb{R}^+$  mit  $x \mapsto \hat{\lambda}(x) = \frac{1}{n} \sum_{i \in [n]} x_i =: \bar{x}_n$  ist der MLS für  $\lambda$ .

(c) *Uniformverteilungsmodell*,  $(\mathbb{R}^n, \mathcal{B}^n, (\text{U}_{[0,\theta]}^n)_{\theta \in \mathbb{R}_0^+})$  mit Produktdichten  $(\mathbb{f}_{\text{U}_{[0,\theta]}^n}^n)_{\theta \in \mathbb{R}_0^+}$ :

$$\arg \max_{\theta \in \mathbb{R}_0^+} \left\{ \mathbb{f}_{\text{U}_{[0,\theta]}^n}^n(x) \right\} = \arg \max_{\theta \in \mathbb{R}_0^+} \left\{ \frac{1}{\theta^n} \prod_{i \in [n]} \mathbb{1}_{[0,\theta]}(x_i) \right\} = \left\{ \max_{i \in [n]} x_i \right\}.$$

Die Statistik  $\hat{\theta} : \mathbb{R}^n \rightarrow \mathbb{R}^+$  mit  $x \mapsto \hat{\theta}(x) = \max_{i \in [n]} x_i$  ist der MLS für  $\theta$ .

(d) *Exponentialverteilungsmodell*,  $(\mathbb{R}^n, \mathcal{B}^n, (\text{Exp}_\lambda^n)_{\lambda \in \mathbb{R}_0^+})$  mit Produktdichten  $(\mathbb{f}_{\text{Exp}_\lambda^n}^n)_{\lambda \in \mathbb{R}_0^+}$ :

$$\arg \max_{\lambda \in \mathbb{R}_0^+} \left\{ \mathbb{f}_{\text{Exp}_\lambda^n}^n(x) \right\} = \arg \max_{\lambda \in \mathbb{R}_0^+} \left\{ \lambda^n \exp(-\lambda \sum_{i \in [n]} x_i) \right\} = \left\{ \frac{1}{\bar{x}_n} \right\}.$$

Die Statistik  $\hat{\lambda} : \mathbb{R}^n \rightarrow \mathbb{R}^+$  mit  $x \mapsto \hat{\lambda}(x) = (\bar{x}_n)^{-1}$  ist der MLS für  $\lambda$ .

(e) *Normalverteilungsmodell*,  $(\mathbb{R}^n, \mathcal{B}^n, (\text{N}_{(\mu, \sigma^2)}^n)_{\mu \in \mathbb{R}})$ ,  $\sigma \in \mathbb{R}_0^+$ , mit Produktdichten  $(\mathbb{f}_{\text{N}_{(\mu, \sigma^2)}^n}^n)_{\mu \in \mathbb{R}}$ :

$$\arg \max_{\mu \in \mathbb{R}} \left\{ \mathbb{f}_{\text{N}_{(\mu, \sigma^2)}^n}^n(x) \right\} = \arg \max_{\mu \in \mathbb{R}} \left\{ \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i \in [n]} (x_i - \mu)^2\right) \right\} = \{\bar{x}_n\}.$$

Die Statistik  $\hat{\mu} : \mathbb{R}^n \rightarrow \mathbb{R}$  mit  $x \mapsto \hat{\mu}(x) = \bar{x}_n$  ist der MLS für  $\mu$ .  $\square$

§12.24 **Bemerkung.** Im Binomialverteilungsmodell  $X \odot (\text{Bin}_{(n,p)})_{p \in [0,1]}$  nimmt der MLS  $\hat{p} = X/n$  für  $p$  (vgl. **Beispiel** §12.23 (a)) nur Werte in  $\hat{p}([0, n]) = \{\frac{0}{n}, \dots, \frac{n}{n}\}$  an. Für jeden Parameter  $p \in [0, 1] \setminus \hat{p}([0, n])$  gilt somit  $\text{Bin}_{(n,p)}(\hat{p} = p) = 0$ . Analog, für  $\sigma \in \mathbb{R}_0^+$  und  $n = 1$  im Normalverteilungsmodell  $X \odot (\text{N}_{(\mu, \sigma^2)})_{\mu \in \mathbb{R}}$  ist  $\hat{\mu} = X$  der MLS für  $\mu$  (vgl. **Beispiel** §12.23 (e)), sodass für jeden Parameterwert  $\mu \in \mathbb{R}$  gilt  $\text{N}_{(\mu, \sigma^2)}(\hat{\mu} = \mu) = \text{N}_{(\mu, \sigma^2)}(\{\mu\}) = 0$ . Im Allgemeinen können wir nicht davon ausgehen, dass der Schätzwert des MLS gleich dem wahren Wert ist. Andererseits im Normalverteilungsmodell, gilt  $\{\mu \in [\hat{\mu} - c, \hat{\mu} + c]\} = \{x \in \mathbb{R} \mid \mu \in [\mu(x) - c, \mu(x) + c]\} = \{\hat{\mu} \in [\mu - c, \mu + c]\}$  für jedes  $c \in \mathbb{R}_0^+$ , so dass dass  $\text{N}_{(\mu, \sigma^2)}(\mu \in [\hat{\mu} - c, \hat{\mu} + c]) = \text{N}_{(\mu, \sigma^2)}(\hat{\mu} \in [\mu - c, \mu + c]) = \text{N}_{(\mu, \sigma^2)}([\mu - c, \mu + c]) > 0$ . Für eine Stichprobe  $x$  ist es somit sinnvoller einen Bereich  $[\hat{\mu}(x) - c, \hat{\mu}(x) + c]$  anzugeben als nur den Schätzwert  $\hat{\mu}(x)$ .  $\square$

### §12.03 Konfidenzbereich

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

§12.26 **Bemerkung.** Für jedes  $\gamma \in \Gamma$  sei  $\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\}$  eine Partition der Menge der interessierenden Parameterwerte  $\Gamma$ , also  $\mathcal{R}_\gamma \uplus \mathcal{F}_\gamma = \Gamma$ . Für jedes  $\theta \in \Theta$  fassen wir  $\mathcal{R}_{\gamma(\theta)}$  und  $\mathcal{F}_{\gamma(\theta)}$  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  $\mathbb{P}_\theta(\tilde{\gamma} \in B)$  für jeden richtigen Parameterwert  $\tilde{\gamma} \in \mathcal{R}_{\gamma(\theta)}$  besitzt, und gleichzeitig aber eine möglichst präzise Überdeckung besitzt, das heißt, für jeden falschen Parameterwert  $\tilde{\gamma} \in \mathcal{F}_{\gamma(\theta)}$  soll die Überdeckungswahrscheinlichkeit  $\mathbb{P}_\theta(\tilde{\gamma} \in B)$  möglichst klein ist. Wie im Fall des Hypothesentest ist aber ein gleichzeitiges Maximieren und Minimieren dieser Überdeckungswahrscheinlichkeit nicht möglich.  $\square$

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

- (a)  $\mathcal{R}_\gamma = \{\gamma\}$  und  $\mathcal{F}_\gamma = \{\gamma\}^c = (-\infty, \gamma) \cup (\gamma, \infty) = \mathbb{R} \setminus \{\gamma\} =: \mathbb{R}_\gamma$ ;
- (b)  $\mathcal{R}_\gamma = (-\infty, \gamma]$  und  $\mathcal{F}_\gamma = (-\infty, \gamma]^c = (\gamma, \infty)$ ;
- (c)  $\mathcal{R}_\gamma = [\gamma, \infty)$  und  $\mathcal{F}_\gamma = [\gamma, \infty)^c = (-\infty, \gamma)$ .  $\square$

§12.28 **Definition.** Sei  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  eine Familie von Partitionen der Menge der interessierenden Parameterwerte  $\Gamma$ . Für ein  $\alpha \in (0, 1)$  heißt eine Bereichsschätzfunktion  $B$  **Konfidenzbereich zum Niveau  $1 - \alpha$  für  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$** , wenn  $\mathbb{P}_\theta(\tilde{\gamma} \in B) \geq 1 - \alpha$  für alle  $\tilde{\gamma} \in \mathcal{R}_{\gamma(\theta)}$  und für alle  $\theta \in \Theta$  gilt. In diesem Fall wird  $B$  kurz  **$(1 - \alpha)$ -Konfidenzbereich** (für  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ ) genannt. Eine Bereichsschätzfunktion  $B^*$  für  $\Gamma$  heißt **gleichmäßig bester Konfidenzbereich zum Niveau  $1 - \alpha$  für  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$** , falls sie ein  $(1 - \alpha)$ -Konfidenzbereich ist, und jeder andere  $(1 - \alpha)$ -Konfidenzbereich  $B$  (für  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$ ) keine kleinere Überdeckungswahrscheinlichkeit der falschen abgeleiteten Parameterwerte besitzt, das heißt, für alle  $\theta \in \Theta$  und  $\tilde{\gamma} \in \mathcal{F}_{\gamma(\theta)}$  gilt  $\mathbb{P}_\theta(\tilde{\gamma} \in B^*) \leq \mathbb{P}_\theta(\tilde{\gamma} \in B)$ .  $\square$

§12.29 **Beispiel (Normalverteilungsmodell §12.15 fortgesetzt).** Betrachte für  $\sigma \in \mathbb{R}_0^+$  das stetige statistische Experiment  $X \odot (\text{N}_{(\theta, \sigma^2)})_{\theta \in \mathbb{R}}$ , wobei die Menge der interessierenden Parameter  $\Gamma = \Theta = \mathbb{R}$  ist. Für jedes  $c \in \mathbb{R}_0^+$  ist  $B : \mathbb{R} \rightarrow 2^\mathbb{R}$  mit  $B(x) := (x \pm c) := (x - c, x + c)$ , kurz  $B = (X \pm c)$ , eine BSF, da  $\{\tilde{\theta} \in B\} = (\tilde{\theta} \pm c) \in \mathcal{B}$  für jedes  $\tilde{\theta} \in \mathbb{R}$  gilt, mit



Überdeckungswahrscheinlichkeit  $N_{(\theta, \sigma^2)}(\tilde{\theta} \in B) = N_{(\theta, \sigma^2)}((\tilde{\theta} \pm c)) = \Phi(\frac{\tilde{\theta} - \theta + c}{\sigma}) - \Phi(\frac{\tilde{\theta} - \theta - c}{\sigma})$  für alle  $\tilde{\theta}, \theta \in \mathbb{R}$  (vgl. **Beispiel §10.12 (a)**). Interessieren wir uns nur für den wahren Parameter, also  $\mathcal{R}_\theta = \{\theta\}$  und dementsprechend  $\mathcal{F}_\theta = \mathbb{R}_\theta$  als Menge der falschen Parameterwerte.  $B$  ist dann ein  $(1-\alpha)$ -Konfidenzbereich für  $(\{\{\theta\}, \mathbb{R}_\theta\})_{\theta \in \mathbb{R}}$  wenn für alle  $\theta \in \mathbb{R}$  gilt  $N_{(\theta, \sigma^2)}(\theta \in B) = \Phi(\frac{c}{\sigma}) - \Phi(\frac{-c}{\sigma}) \geq 1 - \alpha$ . Wählen wir  $B = (X \pm \sigma z_{1-\alpha/2})$  mit  $1-\alpha/2$ -Quantil  $z_{1-\alpha/2}$  der Standardnormalverteilung, das heißt  $1 - \alpha/2 = \Phi(z_{1-\alpha/2}) = 1 - \Phi(-z_{1-\alpha/2})$ , so gilt  $N_{(\theta, \sigma^2)}(\theta \in B) = \Phi(z_{1-\alpha/2}) - \Phi(-z_{1-\alpha/2}) = 1 - \alpha/2 - (1 - (1 - \alpha/2)) = 1 - \alpha$  für alle  $\theta \in \mathbb{R}$ . Damit ist  $(X \pm \sigma z_{1-\alpha/2})$  ein  $(1-\alpha)$ -Konfidenzbereich für den wahren Parameter  $\theta$  und die falschen Parameter  $\mathbb{R} \setminus \{\theta\}, \theta \in \mathbb{R}$ .  $\square$

### §12.30 Definition.

- (a) Ist  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  eine Familie von Partitionen der Menge der interessierenden Parameterwerte  $\Gamma$ , so definiert  $\mathcal{H}_\gamma^0 := \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{R}_{\tilde{\gamma}}\}$  und  $\mathcal{H}_\gamma^1 := \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{F}_{\tilde{\gamma}}\}$  für  $\gamma \in \Gamma$  eine *assoziierte Familie*  $(\{\mathcal{H}_\gamma^0, \mathcal{H}_\gamma^1\})_{\gamma \in \Gamma}$  *von Partitionen* der Parametermenge  $\Gamma$ , dabei fassen wir  $\mathcal{H}_\gamma^0$  als Nullhypothese und  $\mathcal{H}_\gamma^1$  als Alternativhypothese eines Testproblems auffassen.
- (b) Für eine Bereichsschätzfunktion  $B$  definiert  $(\varphi_\gamma = \mathbb{1}_{\mathcal{A}_\gamma})_{\gamma \in \Gamma}$  eine *assoziierte Familie von Tests* mit Ablehnbereich  $\{\varphi_\gamma = 1\} = \mathcal{A}_\gamma$ , da definitionsgemäß  $\mathcal{A}_\gamma := \{\gamma \notin B\} \in \mathcal{F}$  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  $\mathcal{A}_\gamma = \{\varphi_\gamma = 1\} \in \mathcal{F}$  definiert  $B : \mathcal{X} \rightarrow 2^\Gamma$  mit  $B(x) := \{\gamma \in \Gamma \mid x \notin \mathcal{A}_\gamma\}$  eine *assoziierte Bereichsschätzfunktion*, da in der Tat  $\{\gamma \in B\} = \{x \in \mathcal{X} \mid \gamma \in B(x)\} = \{x \in \mathcal{X} \mid x \notin \mathcal{A}_\gamma\} = \mathcal{A}_\gamma^c \in \mathcal{F}$  für jeden abgeleiteten Parameterwert  $\gamma \in \Gamma$  gilt.  $\square$

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

- (a) Für  $\mathcal{R}_\gamma = \{\gamma\}$  und  $\mathcal{F}_\gamma = \mathbb{R}_\gamma$  sind  $\mathcal{H}_\gamma^0 = \{\gamma\}$  und  $\mathcal{H}_\gamma^1 = \mathbb{R}_\gamma$ ;
- (b) Für  $\mathcal{R}_\gamma = (-\infty, \gamma]$  und  $\mathcal{F}_\gamma = (\gamma, \infty)$  sind  $\mathcal{H}_\gamma^0 = \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{R}_{\tilde{\gamma}}\} = \{\tilde{\gamma} \in \Gamma \mid \gamma \leq \tilde{\gamma}\} = [\gamma, \infty)$  und  $\mathcal{H}_\gamma^1 = (-\infty, \gamma)$ ;
- (c) Für  $\mathcal{R}_\gamma = [\gamma, \infty)$  und  $\mathcal{F}_\gamma = (-\infty, \gamma)$  sind  $\mathcal{H}_\gamma^0 = (-\infty, \gamma]$  und  $\mathcal{H}_\gamma^1 = (\gamma, \infty)$ .  $\square$

§12.32 **Beispiel** (*Normalverteilungsmodell §12.15 fortgesetzt*). Betrachten wir für  $\sigma \in \mathbb{R}_0^+$  im stetigen statistischen Experiment  $X \odot (N_{(\mu, \sigma^2)})_{\mu \in \mathbb{R}}$  mit beliebigem  $\mu_o \in \mathbb{R}$  das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu \leq \mu_o$ , also  $\mathcal{H}_{\mu_o}^0 = (-\infty, \mu_o]$ , gegen die Alternativhypothese  $H_1 : \mu > \mu_o$ , also  $\mathcal{H}_{\mu_o}^1 = (\mu_o, \infty)$ , so ist für jedes  $\alpha \in (0, 1)$  ein Neyman-Pearson-Test  $\varphi_{\mu_o} = \mathbb{1}_{\{X \geq \mu_o - \sigma z_\alpha\}}$  mit Ablehnbereich  $\{\varphi_{\mu_o} = 1\} = \{X \geq \mu_o - \sigma z_\alpha\}$  ein *gleichmäßig bester Test* zum Niveau  $\alpha$ . Die zu der Familie  $(\varphi_{\mu_o})_{\mu_o \in \mathbb{R}}$  *assoziierte Bereichsschätzfunktion*  $B$  erfüllt dann  $B(x) = \{\mu_o \in \mathbb{R} \mid x \in \{\varphi_{\mu_o} = 0\}\} = \{\mu_o \in \mathbb{R} \mid x < \mu_o - \sigma z_\alpha\} = (x + \sigma z_\alpha, \infty)$ .  $\square$

§12.33 **Satz**. Seien  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  eine Familie von Partitionen in richtige und falsche interessierenden Parameterwerte und  $(\{\mathcal{H}_\gamma^0, \mathcal{H}_\gamma^1\})_{\gamma \in \Gamma}$  die assoziierte Familie von Partitionen in Null- und Alternativhypothesen. Dann gilt für eine Familie  $(\varphi_\gamma)_{\gamma \in \Gamma}$  von Tests, dass  $\varphi_\gamma$  ein (gleichmäßig bester)  $\alpha$ -Test der Nullhypothese  $H_0 : \mathcal{H}_\gamma^0$  gegen die Alternativhypothese  $H_1 : \mathcal{H}_\gamma^1$  für jedes  $\gamma \in \Gamma$  ist, genau dann wenn die assoziierte Bereichsschätzfunktion  $B$  für  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  ein (gleichmäßig bester)  $(1-\alpha)$ -Konfidenzbereich ist.

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

§12.35 **Beispiel** (*Normalverteilungsmodell §12.32 fortgesetzt*). Für  $X \sim (N_{(\mu, \sigma^2)})_{\mu \in \mathbb{R}}$  mit  $\sigma \in \mathbb{R}_0^+$  ist  $(X + \sigma z_\alpha, \infty)$  der assoziierte Konfidenzbereich zu einer Familie gleichmäßig bester  $\alpha$ -Tests (vgl. **Beispiel** §12.32). Nach **Satz** §12.33 ist  $(X + \sigma z_\alpha, \infty)$  damit auch ein gleichmäßig bester  $1-\alpha$ -Konfidenzbereich für die Mengen der richtigen Parameter  $\mathcal{R}_{\mu_o} = [\mu_o, \infty)$  und der falschen Parameter  $\mathcal{F}_{\mu_o} = (-\infty, \mu_o)$  mit  $\mu_o \in \mathbb{R}$ . Andererseits, ist  $(X \pm \sigma z_{1-\alpha/2})$  ein  $(1-\alpha)$ -Konfidenzbereich für die Menge der richtigen Parameter  $\mathcal{R}_{\mu_o} = \{\mu_o\}$  und der falschen Parameter  $\mathcal{F}_{\mu_o} = \mathbb{R}_{\setminus \mu_o}$ ,  $\mu_o \in \mathbb{R}$  (vgl. **Beispiel** §12.29). Nach **Satz** §12.33 ist damit der assoziierte Test  $\phi_{\mu_o}$  mit Ablehnbereich  $\{\phi_{\mu_o} = 1\} = \{\mu_o \notin (X \pm \sigma z_{1-\alpha/2})\} = \{|X - \mu_o| \geq \sigma z_{1-\alpha/2}\}$  ein  $\alpha$ -Test für das *zweiseitige Testproblem* der Nullhypothese  $H_0 : \mu = \mu_o$  gegen die Alternativhypothese  $H_1 : \mu \neq \mu_o$ . □





## Kapitel 4

# Bedingte Wahrscheinlichkeit und Unabhängigkeit

## §13 Bedingte Wahrscheinlichkeiten und Bayes-Formel

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

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

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

die *bedingte Wahrscheinlichkeit* von  $A$  gegeben (oder unter)  $B$  bezeichnet.  $\square$

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

- (i)  $\tilde{\mathbb{P}} : \mathcal{A} \rightarrow [0, 1]$  mit  $A \mapsto \tilde{\mathbb{P}}(A) := \mathbb{P}(A|B)$  ist ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ . Der Wahrscheinlichkeitsraum  $(B, \mathcal{A}|_B, \tilde{\mathbb{P}}|_B)$  wird die *Spur* von  $(\Omega, \mathcal{A}, \mathbb{P})$  über  $B$  genannt.
- (ii) Es sei  $\{B_i, i \in \mathcal{I}\}$  eine messbare Partition von  $B$  mit  $\mathbb{P}(B_i) > 0$  für alle  $i \in \mathcal{I}$ . Für alle  $A \in \mathcal{A}$  gilt dann die *Formel von der totalen Wahrscheinlichkeit*:

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

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

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

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

§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:  $\frac{0,002 \cdot 0,999}{0,001 \cdot 0,999 + 0,999 \cdot 0,002} = 2/3$ , d.h. über 66%.  $\square$

§13.05 **Lemma.** Für  $A_i \in \mathcal{A}$ ,  $i \in \llbracket n+1 \rrbracket$  mit  $\mathbb{P}(\bigcap_{i \in \llbracket n \rrbracket} A_i) > 0$  gilt

$$\mathbb{P}\left(\bigcap_{i \in \llbracket n+1 \rrbracket} A_i\right) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_{n+1} | \bigcap_{i \in \llbracket n \rrbracket} A_i).$$

§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  $\frac{S}{N} \cdot \frac{S-1}{N-1} \cdot \frac{W}{N-2}$ . 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  $(\Omega, \mathcal{A}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $\mathcal{I}$  eine beliebige nicht-leere Indexmenge.

§14.01 **Definition.**

- (a) Zwei Ereignisse  $A, B \in \mathcal{A}$  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  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$  heißt *(stochastisch) unabhängig* (unter  $\mathbb{P}$ ), kurz  $\perp\!\!\!\perp_{i \in \mathcal{I}} A_i$ , wenn für jede endliche Teilmenge  $\mathcal{J} \subseteq \mathcal{I}$  gilt  $\mathbb{P}(\bigcap_{j \in \mathcal{J}} A_j) = \prod_{j \in \mathcal{J}} \mathbb{P}(A_j)$ . □

§14.02 **Bemerkung.**

- (i) Es gilt  $\perp\!\!\!\perp_{i \in \mathcal{I}} A_i$  genau dann, wenn  $\perp\!\!\!\perp_{i \in \mathcal{J}} A_i$  für jede endliche Teilmenge  $\mathcal{J} \subseteq \mathcal{I}$  gilt.
- (ii) Für  $|\mathcal{I}| \geq 3$  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  $i, j \in \mathcal{I}$  mit  $i \neq j$ , 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.** Sei  $(A_i)_{i \in \mathcal{I}}$  eine nicht-leere Familie von unabhängigen Ereignissen aus  $\mathcal{A}$ . Für  $A'_i \in \sigma(\{A_i\}) = \{\emptyset, A_i, A_i^c, \Omega\}$ ,  $i \in \mathcal{I}$ , ist die Familie  $(A'_i)_{i \in \mathcal{I}}$  von Ereignissen aus  $\mathcal{A}$  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  $\mathcal{A}$  gilt:

- (i) Aus  $\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty$  folgt  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0$ ;
- (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  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$ .

§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  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}(A) < 1$ . Damit kann also auf die Unabhängigkeit in Teil (ii) von **Satz** §14.06 nicht verzichtet werden.
- (b) Im unendlichen Münzwurfexperiment bezeichne  $A_n^M := \{\omega \in \Omega \mid \omega_n = \dots = \omega_{n+M-1} = 1\}$  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  $\mathbb{P}(\limsup_{k \rightarrow \infty} A_{kM}^M) = 1$  für jedes  $M \in \mathbb{N}$ . Dies impliziert  $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n^M) = 1$ . Es gilt sogar  $\mathbb{P}(\bigcap_{M \in \mathbb{N}} \limsup_{n \rightarrow \infty} A_n^M) = 1$ , da der abzählbare Schnitt von Einsmengen wieder eine Einsmenge ist. □

## §15 Unabhängige $\sigma$ -Algebren

Für jedes  $i \in \mathcal{I}$  einer nicht-leeren Indexmenge  $\mathcal{I}$  sei  $\mathcal{A}_i \subseteq \mathcal{A}$  eine  $\sigma$ -Algebra. Die Definition der Unabhängigkeit der Familie  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  von Teil- $\sigma$ -Algebren von  $\mathcal{A}$  ist ein Spezialfall der folgenden Definition.

§15.01 **Definition**. Eine Familie  $(\mathcal{E}_i)_{i \in \mathcal{I}}$  von Teilmengensystemen aus  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$ , d.h.  $\mathcal{E}_i \subseteq \mathcal{A}$  für alle  $i \in \mathcal{I}$ , heißt (*stochastisch*) *unabhängig* (unter  $\mathbb{P}$ ), kurz  $\perp\!\!\!\perp_{i \in \mathcal{I}} \mathcal{E}_i$ , wenn  $\perp\!\!\!\perp_{i \in \mathcal{I}} A_i$  für alle  $A_i \in \mathcal{E}_i$  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  $\pi$ -System, also ein schnittstabiles Teilmengensystem, aus  $\mathcal{A}$  ist, ist dies ein Spezialfall der nächsten Aussage, welches wir mit Hilfe des  $\pi$ - $\lambda$  **Satz** §03.11 zeigen. □

§15.03 **Satz**. Sei  $(\mathcal{E}_i)_{i \in \mathcal{I}}$  eine unabhängige Familie von  $\pi$ -Systemen aus  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge, also  $\perp\!\!\!\perp_{i \in \mathcal{I}} \mathcal{E}_i$ , dann gilt  $\perp\!\!\!\perp_{i \in \mathcal{I}} \sigma(\mathcal{E}_i)$ .

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

§15.05 **Lemma**. Für eine Familie  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  von Teil- $\sigma$ -Algebren aus  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$  sei

$$\mathcal{E} = \left\{ \bigcap_{i \in \mathcal{J}} A_i \mid A_i \in \mathcal{A}_i, i \in \mathcal{J}, \mathcal{J} \subseteq \mathcal{I}, |\mathcal{J}| < \infty \right\}.$$

Dann ist  $\mathcal{E}$  ein  $\pi$ -System und  $\bigvee_{i \in \mathcal{I}} \mathcal{A}_i = \sigma(\mathcal{E})$ , also  $\mathcal{E}$  ist ein schnittstabiler Erzeuger von  $\bigvee_{i \in \mathcal{I}} \mathcal{A}_i$ .

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

§15.07 **Satz**. Für jede unabhängige Familie  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  von Teil- $\sigma$ -Algebren aus  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$ , also  $\perp\!\!\!\perp_{i \in \mathcal{I}} \mathcal{A}_i$ , und jede Partition  $\{\mathcal{J}_k : k \in \mathcal{K}\}$  von  $\mathcal{I}$  ist die Familie  $(\bigvee_{i \in \mathcal{J}_k} \mathcal{A}_i)_{k \in \mathcal{K}}$  von Teil- $\sigma$ -Algebren aus  $\mathcal{A}$  unabhängig, also  $\perp\!\!\!\perp_{k \in \mathcal{K}} \bigvee_{i \in \mathcal{J}_k} \mathcal{A}_i$ .

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

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

$$\mathcal{A}_\infty := \bigcap_{n \geq 1} \bigvee_{m \geq n} \mathcal{A}_m = \bigcap_{n \geq 1} \bigvee_{m \geq n} \mathcal{A}_m = \bigcap_{n \in \mathbb{N}} \sigma(\bigcup \{\mathcal{A}_m \mid m \in \mathbb{N} \wedge m \geq n\})$$

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

§15.10 **Beispiel.** Sei  $(A_n)_{n \in \mathbb{N}}$  eine Folge von Ereignissen aus  $\mathcal{A}$ , dann sind  $A_\star = \liminf_{n \rightarrow \infty} A_n$  und  $A^\star = \limsup_{n \rightarrow \infty} A_n$  (vgl. Definition §01.13) asymptotische Ereignisse bzgl.  $(\sigma(\{A_n\}))_{n \in \mathbb{N}}$ . Setzen wir nämlich  $B_n := \bigcap_{m=n}^\infty A_m$  für  $n \in \mathbb{N}$ , dann gilt  $B_n \uparrow A_\star$  und  $B_n \in \bigvee_{m \geq N} \sigma(\{A_n\})$  für jedes  $n \geq N$ , sodass  $A_\star \in \bigvee_{m \geq N} \sigma(\{A_n\})$  für alle  $N \in \mathbb{N}$  und damit auch  $A_\star \in \bigwedge_{N \in \mathbb{N}} \bigvee_{m \geq N} \sigma(\{A_n\})$ . Für  $A^\star$  geht dies analog. Insbesondere ist die asymptotische  $\sigma$ -Algebra  $\bigwedge_{N \in \mathbb{N}} \bigvee_{m \geq N} \sigma(\{A_n\})$  nicht leer.  $\square$

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  $A^\star = \limsup_{n \rightarrow \infty} A_n$  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.** Sei  $(\mathcal{A}_n)_{n \in \mathbb{N}}$  eine Folge von unabhängigen Teil- $\sigma$ -Algebren aus  $\mathcal{A}$ . Dann ist die Wahrscheinlichkeit für jedes bzgl.  $(\mathcal{A}_n)_{n \in \mathbb{N}}$  asymptotischen Ereignisses entweder 0 oder 1, also

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

§15.12 **Beweis** von Satz §15.11. In der Vorlesung.  $\square$

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

§15.14 **Beispiel (Beispiel §15.10 fortgesetzt).** Sei  $(A_n)_{n \in \mathbb{N}}$  eine Folge von *unabhängigen* Ereignissen aus  $\mathcal{A}$ . Da  $A_\star = \liminf_{n \rightarrow \infty} A_n$  und  $A^\star = \limsup_{n \rightarrow \infty} A_n$  (vgl. Beispiel §15.10) asymptotische Ereignisse bzgl. der Folge unabhängiger  $\sigma$ -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  $\mathbb{P}(A^\star) \in \{0, 1\}$ .  $\square$

## §16 Unabhängige Zufallsvariablen

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

§16.01 **Erinnerung..** b Für jedes  $i \in \mathcal{I}$  ist die von  $X_i$  erzeugte  $\sigma$ -Algebra  $\sigma(X_i) = X_i^{-1}(\mathcal{I}_i) = \{X_i(B) \mid B \in \mathcal{I}_i\}$  eine Teil- $\sigma$ -Algebra von  $\mathcal{A}$ .  $\square$

§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- $\sigma$ -Algebren von  $\mathcal{A}$  unabhängig ist.  $\square$

§16.03 **Bemerkung.** Definitionsgemäß gilt  $\perp\!\!\!\perp_{i \in \mathcal{I}} X_i$  genau dann, wenn  $\perp\!\!\!\perp_{i \in \mathcal{J}} X_i$  für jede endliche Teilmenge  $\mathcal{J} \subseteq \mathcal{I}$  gilt. Weiterhin für jede endliche Familie  $(B_j)_{j \in \mathcal{J}}$  von Ereignissen also mit  $B_j \in \mathcal{I}_j, j \in \mathcal{J} \subseteq \mathcal{I}, |\mathcal{J}| < \infty$  gilt  $\mathbb{P}(\bigcap_{j \in \mathcal{J}} \{X_j \in B_j\}) = \prod_{j \in \mathcal{J}} \mathbb{P}(X_j \in B_j)$ .  $\square$

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

§16.05 **Lemma.** Seien  $(\mathcal{T}_i, \mathcal{F}_i)_{i \in \mathcal{I}}$  eine Familie messbarer Räume und für jedes  $i \in \mathcal{I}$ ,  $h_i : \mathcal{S}_i \rightarrow \mathcal{T}_i$  eine  $\mathcal{S}_i$ - $\mathcal{T}_i$ -messbare Funktion. Ist  $\perp\!\!\!\perp_{i \in \mathcal{I}} X_i$ , so gilt auch  $\perp\!\!\!\perp_{i \in \mathcal{I}} h_i(X_i)$ . Für jede Partition  $\{\mathcal{J}_k : k \in \mathcal{K}\}$  von  $\mathcal{I}$  ist  $((h_i(X_i))_{i \in \mathcal{J}_k})_{k \in \mathcal{K}}$  eine Familie von unabhängigen Zufallsvariablen, also  $\perp\!\!\!\perp_{k \in \mathcal{K}} (h_i(X_i))_{i \in \mathcal{J}_k}$ .

§16.06 **Beweis** von **Lemma** §16.05. In der Vorlesung.  $\square$

§16.07 **Korollar.**

- (i) Eine Familie  $(X_i)_{i \in \mathcal{I}}$  numerischer Zufallsvariablen aus  $\overline{\mathcal{A}}$  ist unabhängig genau dann, wenn für jede endliche Teilmenge  $\mathcal{J} \subseteq \mathcal{I}$  und für alle  $x_i \in \mathbb{R}$ ,  $i \in \mathcal{J}$ , gilt:

$$\mathbb{P}\left(\bigcap_{i \in \mathcal{J}} \{X_i \leq x_i\}\right) = \prod_{i \in \mathcal{J}} \mathbb{P}(X_i \leq x_i).$$

Für  $\mathcal{I} = \llbracket n \rrbracket$  sind also  $X_1, \dots, X_n$  genau dann unabhängig, wenn für alle  $x \in \mathbb{R}^n$  gilt

$$\mathbb{F}^{(X_1, \dots, X_n)}(x) = \prod_{i \in \llbracket n \rrbracket} \mathbb{F}^{X_i}(x_i).$$

- (ii) Diskret-verteilte Zufallsvariablen  $X_1, \dots, X_n$  sind genau dann unabhängig, wenn für alle  $s \in \times_{i \in \llbracket n \rrbracket} \mathcal{S}_i$  gilt

$$\mathbb{p}^{(X_1, \dots, X_n)}(s) = \prod_{i \in \llbracket n \rrbracket} \mathbb{p}^{X_i}(s_i).$$

- (iii) Ist  $X = (X_1, \dots, X_n)$  ein stetig-verteilter Zufallsvektor, dann sind  $X_1, \dots, X_n$  genau dann unabhängig, wenn für Lebesgue-fast alle  $x \in \mathbb{R}^n$  gilt

$$\mathbb{f}^X(x) = \prod_{i \in \llbracket n \rrbracket} \mathbb{f}^{X_i}(x_i).$$

§16.08 **Beweis** von **Korollar** §16.07. In der Vorlesung.  $\square$

§16.09 **Beispiel.**

- (a) Beim Würfelwurf mit zwei Würfeln sind  $W_i : \Omega \rightarrow \llbracket 6 \rrbracket$  mit  $W_i(\omega_1, \omega_2) = \omega_i$  für  $i \in \llbracket 2 \rrbracket$  unabhängige Zufallsvariablen (unter Gleichverteilung). Dazu genügt es, für  $s_1, s_2 \in \llbracket 6 \rrbracket$  die entsprechenden Zähldichten  $\mathbb{p}^{W_1}(s_1) = \mathbb{p}^{W_2}(s_2) = 1/6$  sowie  $\mathbb{p}^W(s_1, s_2) = 1/36$  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  $p \in [0, 1]$  (vgl. **Beispiele** §04.08 (c)). Dann sind  $(X_i)_{i \in [n]}$  unabhängige und identisch  $B_p$ -verteilte Bernoulli-Zufallsvariablen mit identischer Randzähldichte  $\mathbb{P}(x) = p^x(1-p)^{1-x}$ ,  $x \in \{0, 1\}$ , da für die gemeinsame Zähldichte  $p^{\sum_{i \in [n]} x_i} (1-p)^{n - \sum_{i \in [n]} x_i} = \prod_{i \in [n]} \mathbb{P}(x_i)$  für alle  $(x_i)_{i \in [n]} \in \{0, 1\}^n$  gilt.
- (d) Sind  $\mathbb{f}_1, \dots, \mathbb{f}_n$  Dichten auf  $\mathbb{R}$ , so erzeugt die Produktdichte  $\mathbb{f}(x) = \prod_{i \in [n]} \mathbb{f}_i(x_i)$  ein Produktmaß auf  $(\mathbb{R}^n, \mathcal{B}^n)$  (vgl. **Definition** §11.03). Die Koordinatenabbildungen  $X_i : \mathbb{R}^n \rightarrow \mathbb{R}$  mit  $x \mapsto X_i(x) := x_i$ ,  $i \in [n]$  sind Borel-messbar, also Zufallsvariablen. 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.  $\square$

§16.10 **Erinnerung..** Nach **Satz** §11.07 existiert zu jeder Familie  $(\mathcal{S}_i, \mathcal{I}_i, \mathbb{P}_i)_{i \in \mathcal{I}}$  von Wahrscheinlichkeitsräumen ein eindeutig bestimmtes Produktmaß  $\bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$  auf dem zugehörigen Produktraum  $\mathcal{S}_{\mathcal{I}} = \prod_{i \in \mathcal{I}} \mathcal{S}_i$  versehen mit der Produkt- $\sigma$ -Algebra  $\mathcal{S}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathcal{S}_i$ .  $\square$

§16.11 **Korollar.** Für jede Familie  $(\mathcal{S}_i, \mathcal{I}_i, \mathbb{P}_i)_{i \in \mathcal{I}}$  von Wahrscheinlichkeitsräumen mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$  existiert eine Familie  $(X_i)_{i \in \mathcal{I}}$  unabhängiger  $(\mathcal{S}_i, \mathcal{I}_i)$ -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  $(\mathcal{S}_{\mathcal{I}}, \mathcal{S}_{\mathcal{I}})$ .

§16.12 **Beweis von Korollar** §16.11. In der Vorlesung.  $\square$

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

$$\mathcal{A}_X := \bigcap_{n \geq 1} \bigcup_{m \geq n} \sigma(X_m) = \bigcap_{n \geq 1} \bigcup_{m \geq n} \sigma(X_m) = \bigcap_{n \in \mathbb{N}} \sigma(\bigcup \{\sigma(X_m) \mid m \in \mathbb{N} \wedge m \geq n\})$$

die *asymptotische  $\sigma$ -Algebra*. Ein Ereignis  $A \in \mathcal{A}_X$  wird *asymptotisch* bzgl.  $(X_n)_{n \in \mathbb{N}}$  genannt, d.h.  $A$  hängt für alle  $n \geq 1$  nur von  $(X_m)_{m \geq n}$  ab.  $\square$

§16.14 **Beispiel.**

- (a) Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge numerischer Zufallsvariablen, so sind die Abbildungen  $X_* := \liminf_{n \rightarrow \infty} X_n$  und  $X^* := \limsup_{n \rightarrow \infty} X_n$  (vgl. **Definition** §08.10) messbar bzgl.  $\mathcal{A}_X$ . In der Tat: setzen wir  $Y_n := \sup_{m \geq n} X_m$  mit  $Y_n \downarrow X^*$ , so ist für jedes  $N \in \mathbb{N}$  die Zufallsvariable  $X^* = \inf_{n \geq 1} Y_n = \inf_{n \geq N} Y_n$  messbar bzgl.  $\bigcap_{n \geq N} \sigma(X_n)$ , also auch bzgl.  $\bigcap_{n \geq 1} \bigcup_{m \geq n} \sigma(X_m)$ . Für  $X_*$  geht dies analog.
- (b) Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge reeller Zufallsvariablen sowie  $A$  das Ereignis, dass der Grenzwert  $\lim_{n \rightarrow \infty} \sum_{k \in [n]} \frac{X_k}{k}$  existiert. Setzen wir  $B_{N,\varepsilon} := \bigcap_{m > n \geq N} \{\sum_{k \in [n,m]} \frac{X_k}{k} \in (-\varepsilon, \varepsilon)\}$  für  $\varepsilon \in \mathbb{Q}_0^+$  und  $N \in \mathbb{N}$ , 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,  $B_{N,\varepsilon} \in \bigvee_{k \geq N} \sigma(X_k)$  für alle  $N \in \mathbb{N}$  und für jedes  $n \in \mathbb{N}$  und alle  $m \geq n$  gilt  $\bigvee_{k \geq m} \sigma(X_k) \subseteq \bigvee_{k \geq n} \sigma(X_k)$ , so dass  $\bigcup_{N=1}^m B_{N,\varepsilon} = B_{m,\varepsilon} \in \bigvee_{k \geq n} \sigma(X_k)$  für alle  $m \geq n$ . Folglich gilt  $\bigcup_{N=1}^{\infty} B_{N,\varepsilon} \in \bigvee_{k \geq n} \sigma(X_k)$  für alle  $n \in \mathbb{N}$  und damit  $C_\varepsilon := \bigcup_{N=1}^{\infty} B_{N,\varepsilon} \in \mathcal{A}_X$ . Da  $\mathcal{A}_X$  als  $\sigma$ -Algebra  $\sigma$ - $\cap$ -stabil ist, folgt auch  $A = \bigcap_{\varepsilon \in \mathbb{Q}_0^+} C_\varepsilon \in \mathcal{A}_X$ .  $\square$

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

$$\mathcal{A}_X \subseteq \overline{\mathcal{F}} := \{A \in \mathcal{A} \mid \mathbb{P}(A) \in \{0, 1\}\}.$$

§16.16 **Beweis von Korollar §16.15.** Direktes Anwenden von Satz §15.11.  $\square$

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

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

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

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

$$\mathbb{P}(X_* < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}(X_* \leq n) = 0,$$

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

$$\mathbb{P}(X_* \leq x_*) = \lim_{n \rightarrow \infty} \mathbb{P}(X_* \leq x_* + \frac{1}{n}) = 1 \text{ und}$$

$$\mathbb{P}(X_* < x_*) = \lim_{n \rightarrow \infty} \mathbb{P}(X_* \leq x_* - \frac{1}{n}) = 0,$$

also  $\mathbb{P}(X_* = x_*) = 1$ . Ist  $x_* = -\infty$ , so gilt  $\mathbb{P}(X_* > -\infty) = \lim_{n \rightarrow \infty} \mathbb{P}(X_* > n) = 0$ , also auch  $\mathbb{P}(X_* = -\infty) = 1$ . 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 jede beliebige Randverteilung  $\mathbb{P}_{X_k}$  der Zufallsvariablen  $X_k, k \in \mathbb{N}$ .  $\square$

## §17 Faltung

§17.01 **Vorbermerkung.** 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  $\mathbb{R}$ . Gemäß Satz §11.07 hat dann das Paar  $(X, Y)$  die Verteilung  $\mathbb{P}^X \otimes \mathbb{P}^Y$  auf  $\mathbb{R}^2$ . Andererseits ist  $X + Y$  eine reelle Zufallsvariable, und für die Borel-messbare Additionsabbildung  $\oplus : \mathbb{R}^2 \rightarrow \mathbb{R}$  mit  $(x, y) \mapsto \oplus((x, y)) := x + y$  gilt  $X + Y = \oplus \circ (X, Y)$ . Damit hat  $X + Y$  die Verteilung  $(\mathbb{P}^X \otimes \mathbb{P}^Y) \circ \oplus^{-1}$  auf  $\mathbb{R}$ . Aufgrund der Kommutativität der Addition ist  $(\mathbb{P}^Y \otimes \mathbb{P}^X) \circ \oplus^{-1}$  auch die Verteilung von  $Y + X$ .  $\square$

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

§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** §§17.01.  $\square$

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

§17.06 **Beweis** von **Korollar** §17.05. Die Kommutativität folgt aus der **Vorbemerkung** §§17.01. Die Assoziativität der Addition vererbt sich analog auf die Faltung.  $\square$

§17.07 **Lemma (Diskreter Fall).**

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

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

(ii) Besitzen  $\mathbb{P}$  und  $\tilde{\mathbb{P}}$  Zähldichten  $\mathbb{p}$  und  $\tilde{\mathbb{p}}$  auf  $\mathbb{N}_0$ , so besitzt  $\mathbb{P} \star \tilde{\mathbb{P}}$  die Zähldichte

$$[\mathbb{p} \star \tilde{\mathbb{p}}](n) := \sum_{k \in [0, n]} \mathbb{p}(n - k) \tilde{\mathbb{p}}(k), \quad n \in \mathbb{N}_0.$$

§17.08 **Beweis** von **Lemma** §17.07. In der Vorlesung.  $\square$

§17.09 **Beispiel.**

(a) Für  $n \in \mathbb{N}$  sei die gemeinsame Verteilung von  $(X_i)_{i \in [n]}$  ein Bernoulli-Schema mit Parameter  $p \in [0, 1]$  (vgl. **Beispiele** §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  $\text{Bin}_{(2,p)}$ -verteilt (vgl. **Beispiele** §04.08 (d)), da  $B_p \star B_p = \text{Bin}_{(2,p)}$ :

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

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

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

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

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

§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

$$f_{\Gamma_{(\lambda,p)}}(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x), \quad x \in \mathbb{R},$$

mit der Gamma-Funktion  $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt$ , sodass  $\Gamma(n+1) = n!$  für  $n \in \mathbb{N}_0$  und  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

- (a) Spezialfälle sind  $\Gamma_{(\lambda,1)} = \text{Exp}_\lambda$  und  $\Gamma_{(1/2,1/2)} = \chi_1^2$ .
- (b) Setzt man zusätzlich  $\Gamma_{(\lambda,0)} := \delta_0$ , so bildet  $(\Gamma_{(\lambda,p)})_{p \in \mathbb{R}^+}$  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  $\chi_1^2$ -verteilten Zufallsvariablen* (äquivalent dazu die Summe der Quadrate,  $\sum_{i \in [k]} Z_i^2$ , von  $k$  unabhängigen standard-normalverteilten Zufallsvariablen  $Z_1, \dots, Z_k$ , oder äquivalent  $\|Z\|^2$  eines standard-normalverteilten Zufallsvektors  $Z = (Z_i)_{i \in [k]} \sim N_{(0,E_k)}$  im  $\mathbb{R}^k$ ) gemäß  $\chi_k^2 := \Gamma_{(1/2,k/2)}$ -verteilt (Sprechweise:  *$\chi^2$ -Verteilung mit  $k$  Freiheitsgraden*). □

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

## §18 Multivariate Normalverteilung

§18.01 **Vorbemerkung.** Im Folgenden fassen wir Vektoren als Spaltenvektoren  $a = (a_1 \cdots a_n)^t \in \mathbb{R}^n$  auf. Wir bezeichnen mit  $\langle \cdot, \cdot \rangle$  das Standardskalarprodukt auf  $\mathbb{R}^n$ , das heißt,  $\langle a, b \rangle = b^t a = \sum_{i \in [n]} a_i b_i$  für alle  $a, b \in \mathbb{R}^n$ .

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

Zufallsvektor  $Y = AX + b$  die Dichte  $f^Y(y) = \frac{f^X(A^{-1}(y-b))}{|\det(A)|}$ . Sind insbesondere die Komponenten  $X_1, \dots, X_n$  von  $X$  unabhängig und ist  $A$  eine Diagonalmatrix mit Diagonaleinträgen  $c = (c_i)_{i \in [n]}$ , kurz  $A = \text{Diag}(c)$ , mit  $c_i \in \mathbb{R}_{>0}$ ,  $i \in [n]$ , dann gilt  $\det(A) = \prod_{i \in [n]} c_i \neq 0$

und  $Y = Ax + b = \begin{pmatrix} c_1 X_1 + b_1 \\ \vdots \\ c_n X_n + b_n \end{pmatrix}$  besitzt die Dichte  $f^Y(y) = \prod_{i \in [n]} \frac{1}{|c_i|} f^{X_i}\left(\frac{y_i - b_i}{c_i}\right)$ . In Bei-

spiel §10.12 (c) haben wir weiterhin auf  $(\mathbb{R}^n, \mathcal{B}^n)$  die multivariate Normalverteilung  $N_{(\mu, \Sigma)}$  mit Vektor  $\mu \in \mathbb{R}^n$  und positiv definiten, also symmetrischen und regulären, Matrix  $\Sigma \in \mathbb{R}^{(n,n)}$  über ihre Dichte  $f_{N_{(\mu, \Sigma)}}$  eingeführt. Im Fall einer nicht regulären Matrix  $\Sigma$  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  $\mathbb{R}^n$  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  $\langle X, c \rangle$  für alle  $c \in \mathbb{R}^n$ .

§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.  $\square$

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

$$f^{\langle X, c \rangle}(y) = f_{N_{(0, \langle c, c \rangle)}}(y) = \frac{1}{\sqrt{2\pi \langle c, c \rangle}} \exp\left(-\frac{y^2}{2\langle c, c \rangle}\right), \quad y \in \mathbb{R}.$$

§18.05 **Beweis** von Korollar §18.04. In der Vorlesung.  $\square$

§18.06 **Definition.** Ein  $\mathbb{R}^n$ -wertiger Zufallsvektor  $X$  besitzt eine *multivariate Normalverteilung*  $N_{(\mu, \Sigma)}$  mit  $\mu \in \mathbb{R}^n$  und positiv semi-definiter Matrix  $\Sigma \in \mathbb{R}^{(n,n)}$ , falls für alle  $c \in \mathbb{R}^n$  die reelle Zufallsvariable  $\langle X, c \rangle$  eine  $N_{(\langle \mu, c \rangle, \langle \Sigma c, c \rangle)}$ -Verteilung besitzt. Das Produktmaß  $N_{(0, E_n)} = \bigotimes_{i \in \llbracket n \rrbracket} N_{(0,1)} = N_{(0,1)}^n$  heißt insbesondere (*n-dimensionale*) *Standardnormalverteilung*, wobei  $E_n$  die  $n$ -dimensionale Einheitsmatrix ist.  $\square$

§18.07 **Vorbemerkung.** Für eine Matrix  $A \in \mathbb{R}^{(n,m)}$  mit Spaltenvektoren  $(a_{\bullet i})_{i \in \llbracket n \rrbracket}$  bezeichnet  $\text{Bild}(A) = \langle (a_{\bullet i})_{i \in \llbracket n \rrbracket} \rangle \subseteq \mathbb{R}^n$  die lineare Hülle der Spaltenvektoren, also das Bild der linearen Abbildung  $\mathbb{R}^m \rightarrow \mathbb{R}^n$  mit  $x \mapsto Ax$ . 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  $v \in \mathcal{V}$  gilt  $\langle u, v \rangle = 0$ , und jedes Element  $x \in \mathbb{R}^n$  hat eine eindeutige Darstellung  $x = u + v$  mit  $u \in \mathcal{U}$  und  $v \in \mathcal{U}^\perp$ . Wir bezeichnen mit  $\Pi_{\mathcal{U}}$  die Darstellungsmatrix der orthogonalen Projektion von  $\mathbb{R}^n$  in  $\mathcal{U}$ , also  $\mathcal{U} \oplus \mathcal{U}^\perp \rightarrow \mathcal{U}$  mit  $x = u + v \mapsto u = \Pi_{\mathcal{U}}x$ . Eine Matrix  $U \in \mathbb{R}^{(n,m)}$  heißt partielle Isometrie, falls  $UU^t = \Pi_{\text{Bild}(U)}$  und  $U^tU = \Pi_{\text{Bild}(U^t)}$ .  $\square$

§18.08 **Lemma.** Seien  $X \sim N_{(0, E_m)}$  und  $Y \sim N_{(0, E_k)}$ , dann gelten die folgenden Aussagen

- (i) Falls  $A \in \mathbb{R}^{(n,m)}$  und  $B \in \mathbb{R}^{(n,k)}$  mit  $AA^t = BB^t$  gilt, dann sind die  $\mathbb{R}^n$ -wertigen Zufallsvektoren  $AX$  und  $BY$  identisch verteilt.
- (ii) Falls  $U \in \mathbb{R}^{(n,m)}$  eine partielle Isometrie ist, dann gilt  $UX \sim N_{(0, \Pi_{\text{Bild}(U)})}$ .
- (iii) Falls  $A \in \mathbb{R}^{(m,n)}$  und  $B \in \mathbb{R}^{(m,k)}$  mit  $A^tB = 0$ . Dann sind  $\Pi_{\text{Bild}(A)}X \sim N_{(0, \Pi_{\text{Bild}(A)})}$  und  $\Pi_{\text{Bild}(B)}X \sim N_{(0, \Pi_{\text{Bild}(B)})}$  unabhängig.

§18.09 **Beweis** von Lemma §18.08. In der Vorlesung.  $\square$

§18.10 **Korollar.** Sei  $X \sim N_{(\mu, \Sigma)}$  mit  $\mu = (\mu_i)_{i \in \llbracket n \rrbracket} \in \mathbb{R}^n$  und  $\Sigma = (\Sigma_{ij})_{i,j \in \llbracket n \rrbracket} \in \mathbb{R}^{(n,n)}$  positiv semi-definit, dann gelten die folgenden Aussagen:

- (i) Für alle  $i \in \llbracket n \rrbracket$  gilt  $X_i \sim N_{(\mu_i, \Sigma_{ii})}$ .
- (ii) Für alle  $i, j \in \llbracket n \rrbracket$  mit  $i \neq j$  sind die Koordinaten  $X_i$  und  $X_j$  von  $X$  genau dann unabhängig, wenn  $\Sigma_{ij} = 0$  gilt.
- (iii) Für  $A \in \mathbb{R}^{(m,n)}$  und  $b \in \mathbb{R}^m$  gilt  $Y = AX + b \sim N_{(A\mu + b, A\Sigma A^t)}$ .
- (iv) Ist  $\Sigma$  positiv definit, dann ist  $X$  stetig verteilt mit Lebesgue-Dichte

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

§18.11 **Beweis** von Korollar §18.10. Übungsaufgabe.  $\square$

## §19 Beispiele statistischer Modelle

§19.01 **Definition.** Sei  $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\theta)$  ein adäquates statistisches Experiment für eine  $(\mathcal{X}, \mathcal{F})$ -wertige Zufallsvariable  $X$  und sei  $(X_i)_{i \in \mathcal{I}}$  eine Familie von  $(\mathcal{X}, \mathcal{F})$ -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  $(\mathcal{X}^{\mathcal{I}}, \mathcal{F}^{\mathcal{I}})$  die gemeinsame Verteilung von  $(X_i)_{i \in \mathcal{I}}$  ist. In diesem Fall ist das statistische Produktexperiment  $(\mathcal{X}^{\mathcal{I}}, \mathcal{F}^{\mathcal{I}}, \mathbb{P}_\theta^{\mathcal{I}})$  also adäquat für  $(X_i)_{i \in \mathcal{I}}$ .  $(\mathcal{X}^{\mathcal{I}}, \mathcal{F}^{\mathcal{I}}, \mathbb{P}_\theta^{\mathcal{I}})$  wird auch adäquates *statistisches u.i.v. Modell* für  $(X_i)_{i \in \mathcal{I}}$  genannt.  $\square$

§19.02 **Schreibweise.**  $(X_i)_{i \in \mathcal{I}} \odot \mathbb{P}_\theta^{\mathcal{I}}$  und im Fall  $\mathcal{I} = \llbracket n \rrbracket$  auch  $(X_i)_{i \in \llbracket n \rrbracket} \odot \mathbb{P}_\theta^n$ .  $\square$

§19.03 **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 männliche Geschlecht eines Konsumierenden, so beschreiben wir das Geschlecht der  $n = 1000$  befragten Konsumierenden als Stichprobe eines *Bernoulli-Schemas*  $(\{0, 1\}^n, 2^{\{0, 1\}^n}, (\mathbb{B}_p^n)_{p \in [0, 1]})$ , vgl. **Beispiele** §04.08 (c). Sei  $(X_i)_{i \in \llbracket n \rrbracket} \odot (\mathbb{B}_p^n)_{p \in [0, 1]}$ . Nach **Beispiel** §17.09 (a) ist dann ein *Binomialverteilungsmodell*  $(\llbracket 0, n \rrbracket, 2^{\llbracket 0, n \rrbracket}, (\text{Bin}_{(n, p)}^n)_{p \in [0, 1]})$  ein adäquates statistisches Experiment für die Anzahl  $\sum_{i \in \llbracket n \rrbracket} X_i$  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 **Beispiele** §04.08 (d) durch ein *Binomialverteilungsmodell*  $(\llbracket 0, 50 \rrbracket^n, 2^{\llbracket 0, 50 \rrbracket^n}, (\text{Bin}_{(50, p)}^n)_{p \in [0, 1]})$ . Sei  $(X_i)_{i \in \llbracket n \rrbracket} \odot (\text{Bin}_{(50, p)}^n)_{p \in [0, 1]}$ . Nach §17.09 (a) ist dann ein *Binomialverteilungsmodell*  $(\llbracket 0, 50n \rrbracket, 2^{\llbracket 0, 50n \rrbracket}, (\text{Bin}_{(50n, p)}^n)_{p \in [0, 1]})$  ein adäquates statistisches Experiment für die Gesamtanzahl  $\sum_{i \in \llbracket n \rrbracket} X_i$  an beschädigten Schrauben.
- (c) **Beispiel III:** Die  $n = 280$  zufälligen Anzahlen eingegangener Anrufe innerhalb einer Woche beschreiben wir durch ein *Poissonverteilungsmodell*  $(\mathbb{N}_0^n, 2^{\mathbb{N}_0^n}, (\text{Poi}_\lambda^n)_{\lambda \in \mathbb{R}_0^+})$ , vgl. **Beispiele** §04.08 (f). Sei  $(X_i)_{i \in \llbracket n \rrbracket} \odot (\text{Poi}_\lambda^n)_{\lambda \in \mathbb{R}_0^+}$ . Nach **Beispiel** §17.09 (a) ist dann ein *Poissonverteilungsmodell*  $(\mathbb{N}_0, 2^{\mathbb{N}_0}, (\text{Poi}_{n\lambda})_{\lambda \in \mathbb{R}_0^+})$  ein adäquates statistisches Experiment für die Gesamtanzahl  $\sum_{i \in \llbracket n \rrbracket} X_i$  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*  $(\mathbb{R}^n, \mathcal{B}^n, (\text{U}_{[0, \theta]}^n)_{\theta \in \mathbb{R}_0^+})$ , vgl. **Beispiele** §05.07 (a). Sei  $(X_i)_{i \in \llbracket n \rrbracket} \odot (\text{U}_{[0, \theta]}^n)_{\theta \in \mathbb{R}_0^+}$ . Bezeichne für  $\theta \in \mathbb{R}_0^+$  mit  $\mathbb{P}_\theta$  das stetige Wahrscheinlichkeitsmaß auf  $(\mathbb{R}, \mathcal{B})$  mit Wahrscheinlichkeitsdichte  $f_\theta(x) = \frac{nx^{n-1}}{\theta^n} \mathbb{1}_{[0, \theta]}(x)$ ,  $x \in \mathbb{R}$ . Das stetige statistische Experiment  $(\mathbb{R}, \mathcal{B}, (\mathbb{P}_\theta)_{\theta \in \mathbb{R}_0^+})$  ist dann adäquat für die maximale Wartezeit  $\max_{i \in \llbracket n \rrbracket} X_i$ .
- (e) **Beispiel V:** Die zufällige Lebensdauer der  $n = 100$  Glühlampen (in Stunden) beschreiben wir durch ein *Exponentialverteilungsmodell*  $(\mathbb{R}^n, \mathcal{B}^n, (\text{Exp}_\lambda^n)_{\lambda \in \mathbb{R}_0^+})$ , vgl. **Beispiele** §05.07 (b). Sei  $(X_i)_{i \in \llbracket n \rrbracket} \odot (\text{Exp}_\lambda^n)_{\lambda \in \mathbb{R}_0^+}$ . Nach **Beispiel** §17.12 ist dann ein *Gammaverteilungsmodell*  $(\mathbb{R}, \mathcal{B}, (\Gamma_{(\lambda, n)})_{\lambda \in \mathbb{R}_0^+})$  ein adäquates statistisches Experiment für die kumulierte Lebensdauer  $\sum_{i \in \llbracket n \rrbracket} X_i$  der Glühlampen.
- (f) **Beispiel VI:** Die zufälligen Messwerte beschreiben wir durch ein *Normalverteilungsmodell*  $(\mathbb{R}^n, \mathcal{B}^n, (\text{N}_{(\mu, \sigma^2)}^n)_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+})$ , vgl. **Beispiele** §05.07 (c). Ein *Normalverteilungsmodell*

*dell*  $(\mathbb{R}, \mathcal{B}, (N_{(n\mu, n\sigma^2)}))_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+}$  ist dann nach **Beispiel §17.13** ein adäquates statistisches Experiment für den kumulierten Messwert  $\sum_{i=1}^n X_i$  und nach **Beispiel §10.12 (a)** ist  $(\mathbb{R}, \mathcal{B}, (N_{(\mu, \sigma^2/n)}))_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+}$  adäquat für den mittleren Messwert  $\frac{1}{n} \sum_{i \in \llbracket n \rrbracket} X_i = \bar{X}_n$ .  $\square$ .



# Kapitel 5

## Erwartungswert

### §20 Positive numerische Zufallsvariablen

§20.01 **Satz.** Für jedes Wahrscheinlichkeitsmaß  $\mathbb{P}$  auf einem messbarem Raum  $(\Omega, \mathcal{A})$  existiert ein eindeutig bestimmtes Funktional  $\mathbb{E} : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$  derart, dass die folgenden Bedingungen erfüllt sind:

- (i) Für alle  $X, Y \in \overline{\mathcal{A}}^+$  und  $a, b \in \overline{\mathbb{R}}^+$  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  $\overline{\mathcal{A}}^+$  gilt  $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$ ; ( $\sigma$ -stetig)
- (iii) Für jedes  $A \in \mathcal{A}$  gilt  $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$ . (normiert)

Das Funktional  $\mathbb{E}$  wird **Erwartung** bzgl.  $\mathbb{P}$  genannt und für jedes  $X \in \overline{\mathcal{A}}^+$  heißt  $\mathbb{E}(X)$  der **Erwartungswert** von  $X$ .

§20.02 **Beweis** von **Satz** §20.01. Der Satz fasst die Hauptaussage dieses Abschnittes zusammen, der Beweis der Aussage erfolgt in mehreren Schritten. Wir zeigen zuerst in **Satz** §20.05 die Eindeutigkeitsaussage, und geben dann in **Definition** §20.07 ein Funktional  $\tilde{\mathbb{E}} : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$  explizit an, für das wir in mehreren Schritten die Bedingungen (i)-(iii) nachweisen. Zusammenfassend zeigen wir damit dann in **Satz** §20.11 auch die Existenzaussage.  $\square$

§20.03 **Schreibweise.** Wenn es uns wichtig ist, den Stichprobenraum und das Wahrscheinlichkeitsmaß zu betonen, so schreiben wir auch

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}_{\mathbb{P}}(X) = \mathbb{P}(X) = \int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) \mathbb{P}(d\omega) \quad \text{sowie} \\ \mathbb{E}(X \mathbb{1}_A) &= \mathbb{P}(X \mathbb{1}_A) = \int_{\Omega} \mathbb{1}_A(\omega) X(\omega) \mathbb{P}(d\omega) = \int_A X d\mathbb{P} \quad \text{für } A \in \mathcal{A}.\end{aligned}$$

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.  $\square$

§20.04 **Bemerkung.** Der Beweis der nächsten Aussage wendet die **Beweisstrategie** §09.06 an.  $\square$

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

§20.06 **Beweis** von **Satz** §20.05. In der Vorlesung.  $\square$

Eine Partition  $\mathcal{P} := \{A_i \mid i \in \mathcal{I}\} \subseteq \mathcal{A}$  von  $\Omega$  nennen wir endlich, wenn die nicht-leere Indexmenge  $\mathcal{I}$  endlich ist. Für jedes  $A \in \mathcal{P}$  gilt also  $\emptyset \neq A \in \mathcal{A}$ . Für den weiteren Verlauf sei  $\mathcal{P} := \{\mathcal{P} \subseteq \mathcal{A} \mid \mathcal{P} \text{ endliche Partition von } \Omega\}$ .



§20.07 **Definition.** Sei  $\tilde{\mathbb{E}} : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$  mit

$$X \mapsto \tilde{\mathbb{E}}(X) := \sup_{\mathcal{P} \in \mathcal{P}} \left\{ \sum_{\substack{A \in \mathcal{P}: \\ \mathbb{P}(A) > 0}} \left( \inf_{\omega \in A} X(\omega) \right) \mathbb{P}(A) \right\}.$$

□

§20.08 **Erinnerung.** Für  $X, Y \in \overline{\mathcal{A}}^+$  mit  $X(\omega) \leq Y(\omega)$  für alle  $\omega \in \Omega$  schreiben wir  $X \leq Y$ . □

§20.09 **Lemma.**

- (i) Sei  $\{B_j, | j \in \mathcal{J}\} \in \mathcal{P}$  eine endliche, messbare Partition von  $\Omega$  und sei  $X = \sum_{j \in \mathcal{J}} x_j \mathbb{1}_{B_j}$  eine einfache Zufallsvariable aus  $\overline{\mathcal{A}}^+$ . Dann gilt

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

Insbesondere, erfüllt  $\tilde{\mathbb{E}}$  also die Bedingung **Satz §20.01 (iii)**. (normiert)

- (ii) Für alle  $X, Y \in \overline{\mathcal{A}}^+$  mit  $X \leq Y$  gilt  $\tilde{\mathbb{E}}(X) \leq \tilde{\mathbb{E}}(Y)$ . (monoton)

- (iii) Für alle  $(X_n)_{n \in \mathbb{N}} \uparrow X$  in  $\overline{\mathcal{A}}^+$  gilt  $\tilde{\mathbb{E}}(X_n) \uparrow \tilde{\mathbb{E}}(X)$ . (monotone Konvergenz,  $\sigma$ -stetig)  
 $\tilde{\mathbb{E}}$  erfüllt also die Bedingung **Satz §20.01 (ii)**.

- (iv) Für alle  $X, Y \in \overline{\mathcal{A}}^+$  und  $a, b \in \overline{\mathbb{R}}^+$  gilt  $\tilde{\mathbb{E}}(aX + bY) = a\tilde{\mathbb{E}}(X) + b\tilde{\mathbb{E}}(Y)$ . (linear)  
 $\tilde{\mathbb{E}}$  erfüllt die Bedingung **Satz §20.01 (i)**.

§20.10 **Beweis** von **Lemma §20.09**. In der Vorlesung. □

§20.11 **Existenzsatz.** Das Funktional  $\tilde{\mathbb{E}} : \overline{\mathcal{A}}^+ \rightarrow \overline{\mathbb{R}}^+$  wie in **Definition §20.07** ist eine Erwartung bzgl.  $\mathbb{P}$ , das heißt, es erfüllt die Bedingungen (i)-(iii) aus **Satz §20.01**.

§20.12 **Beweis** von **Satz §20.11**. Die Aussage folgt direkt aus **Lemma §20.09 (i), (iii) und (iv)**. □

Damit haben wir **Satz §20.01** nachgewiesen, so dass die Erwartung  $\mathbb{E}$  eindeutig ist und durch die explizite Form in **Definition §20.07** gegeben ist.

## Weitere Eigenschaften

§20.13 **Lemma.**

- (i) Für  $X \in \overline{\mathcal{A}}^+$  ist  $\mathbb{P}(X = 0) = 1$  genau dann wenn  $\mathbb{E}(X) = 0$ .  
(ii) Sei  $X \in \overline{\mathcal{A}}^+$  mit  $\mathbb{E}(X) < \infty$ . Dann gilt  $\mathbb{P}(X = +\infty) = 0$ . (endlich)  
(iii) Für  $X, Y \in \overline{\mathcal{A}}^+$  gilt  $\mathbb{P}(X \leq Y) = 1$  genau dann, wenn  $\mathbb{E}(X \mathbb{1}_A) \leq \mathbb{E}(Y \mathbb{1}_A)$  für alle  $A \in \mathcal{A}$  gilt. Insbesondere, aus  $\mathbb{P}(X \leq Y) = 1$  folgt  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .  
(iv) Für  $X, Y \in \overline{\mathcal{A}}^+$  gilt  $\mathbb{P}(X = Y) = 1$  genau dann, wenn  $\mathbb{E}(X \mathbb{1}_A) = \mathbb{E}(Y \mathbb{1}_A)$  für alle  $A \in \mathcal{A}$  gilt.

§20.14 **Beweis** von **Lemma §20.13**. In der Vorlesung. □

§20.15 **Schreibweise.** 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  $X < \infty$   $\mathbb{P}$ -f.s.. Falls  $\mathbb{P}$  aus dem Kontext klar ist, wird dies häufig auch einfach weggelassen. □

§20.16 **Lemma von Fatou.** Für jede Folge  $(X_n)_{n \in \mathbb{N}}$  aus  $\overline{\mathcal{A}}^+$  gilt

$$\mathbb{E}\left(\liminf_{n \rightarrow \infty} X_n\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n).$$

§20.17 **Beweis von Lemma §20.16.** In der Vorlesung. □

§20.18 **Beispiel.** Auf dem Wahrscheinlichkeitsraum  $(\mathbb{R}, \mathcal{B}, U_{0,1})$  betrachte für  $n \in \mathbb{N}$  die reelle Zufallsvariable  $X_n := n \mathbb{1}_{(0, 1/n]}$ . Für jedes  $n \in \mathbb{N}$  gilt dann  $\mathbb{E}(X_n) = n U_{0,1}\left((0, \frac{1}{n}]\right) = 1$  und somit  $\liminf_{n \rightarrow \infty} \mathbb{E}(X_n) = 1$ . Andererseits, für jedes  $\omega \in \mathbb{R}$  ist  $\lim_{n \rightarrow \infty} X_n(\omega) = 0$ , sodass  $\mathbb{E}(\liminf_{n \rightarrow \infty} X_n) = 0 < 1 \liminf_{n \rightarrow \infty} \mathbb{E}(X_n)$  gilt. □

§20.19 **Lemma.** Für jede Folge  $(X_n)_{n \in \mathbb{N}}$  aus  $\overline{\mathcal{A}}^+$  gilt  $\sum_{n \in \mathbb{N}} X_n \in \overline{\mathcal{A}}^+$  und

$$\mathbb{E}\left(\sum_{n \in \mathbb{N}} X_n\right) = \sum_{n \in \mathbb{N}} \mathbb{E}(X_n).$$

§20.20 **Beweis von Lemma §20.19.** Übungsaufgabe unter Verwendung von Satz §20.01 (i) (linear) und (ii) (monotone Konvergenz). □

§20.21 **Lemma.** Eine Familie  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  von Teil- $\sigma$ -Algebren aus  $\mathcal{A}$  mit beliebiger nicht-leerer Indexmenge  $\mathcal{I}$  ist unabhängig, also  $\prod_{i \in \mathcal{I}} \mathcal{A}_i$ , genau dann, wenn für jede Familie  $(X_i)_{i \in \mathcal{I}}$  positiver numerischer Zufallsvariablen mit  $X_i \in \overline{\mathcal{A}}_i^+$ ,  $i \in \mathcal{I}$ , und jede endliche nicht-leere Teilmenge  $\mathcal{J} \subseteq \mathcal{I}$  gilt  $\mathbb{E}(\prod_{j \in \mathcal{J}} X_j) = \prod_{j \in \mathcal{J}} \mathbb{E}(X_j)$ .

§20.22 **Beweis von Lemma §20.21.** Übungsaufgabe unter Verwendung der Beweisstrategie §09.06. □

## §21 Integrierbare Zufallsvariablen

§21.01 **Vorbemerkung.** Sei  $(\Omega, \mathcal{A}, \mathbb{P})$  ein Wahrscheinlichkeitsraum und  $X \in \overline{\mathcal{A}}$  eine numerische Zufallsvariable. Dann definieren wir den Erwartungswert von  $X$  mit Hilfe der Zerlegung von  $X = X^+ - X^-$  in die positiven numerischen Zufallsvariablen  $X^+ = X \vee 0$  und  $X^- = (-X) \vee 0$  aus  $\overline{\mathcal{A}}^+$ . Zur Erinnerung, es gilt:  $0 \leq X^+$ ,  $0 \leq X^-$ ,  $|X| = X^+ + X^-$  sowie  $X^+ X^- = 0$ . □

§21.02 **Definition.** Sei  $X \in \overline{\mathcal{A}}$  eine numerische Zufallsvariable.

- (a) Ist höchstens einer der beiden Erwartungswerte  $\mathbb{E}(X^+)$  und  $\mathbb{E}(X^-)$  nicht endlich, das heißt,  $\mathbb{E}(X^+) \wedge \mathbb{E}(X^-) < \infty$ , so definiert  $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$  den *Erwartungswert* von  $X$  mit den üblichen Konventionen  $\infty + x = \infty$  und  $-\infty + x = -\infty$  für alle  $x \in \mathbb{R}$ . Der Erwartungswert von  $X$  ist nicht definiert, wenn  $\mathbb{E}(X^+) = \mathbb{E}(X^-) = \infty$  gilt.
- (b) Falls  $\mathbb{E}(|X|) < \infty$ , also falls  $\mathbb{E}(X^+) < \infty$  und  $\mathbb{E}(X^-) < \infty$ , gilt, dann heißt  $X$  *integrierbar*. Die Menge aller integrierbaren numerischen Zufallsvariablen bezeichnen wir mit  $\mathcal{L}_1 := \mathcal{L}_1(\mathbb{P}) := \mathcal{L}_1(\Omega, \mathcal{A}, \mathbb{P}) := \{X \in \overline{\mathcal{A}} \mid \mathbb{E}(|X|) < \infty\}$ . □

§21.03 **Bemerkung.** Sei  $X \in \mathcal{L}_1$ . Auf Grund der Endlichkeit in Lemma §20.13 (ii) gilt somit auch  $\mathbb{P}(|X| < \infty) = 1$ , das heißt,  $X$  ist fast sicher gleich einer reellen Zufallsvariablen. Wie in Schreibweise §20.03 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:

- (i) Sei  $\mathbb{P}$  ein *stetiges* Wahrscheinlichkeitsmaß auf  $\mathbb{R}^n$  mit Dichte  $f$ . Dann gilt  $X \in \mathcal{L}_1(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P})$  genau dann, wenn  $\int_{\mathbb{R}^n} |X(\omega)| f(\omega) d\omega < \infty$ . In diesem Fall ist

$$\mathbb{E}(X) = \mathbb{P}(X) = \int_{\mathbb{R}^n} X(\omega) f(\omega) d\omega.$$

- (ii) Sei  $\mathbb{P}$  ein *diskretes* Wahrscheinlichkeitsmaß mit Zähldichte  $p$ . Dann gilt  $X \in \mathcal{L}_1$  genau dann wenn  $\sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty$ . In diesem Fall ist

$$\mathbb{E}(X) = \mathbb{P}(X) = \sum_{\omega \in \Omega} X(\omega) p(\omega).$$

□

§21.04 **Proposition.**  $\mathcal{L}_1$  ist ein Vektorraum. Für das Funktional  $\mathbb{E} : \mathcal{L}_1 \rightarrow \mathbb{R}$  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  $X \geq 0$  gilt  $\mathbb{E}(X) \geq 0$ ; (positiv)
- (iii) Für alle  $X \in \mathcal{L}_1$  gilt  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ ;
- (iv) Für alle  $X, Y \in \mathcal{L}_1$  mit  $X \leq Y$  gilt  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ ; (monoton)
- (v) Für alle  $X \in \mathcal{A}$  mit  $\sup_{\omega \in \Omega} |X(\omega)| < \infty$  gilt  $X \in \mathcal{L}_1$ . In diesem Fall heißt  $X$  **beschränkt**.
- (vi) Seien  $X \in \mathcal{L}_1$  und  $Y \in \overline{\mathcal{A}}$  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  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

§21.05 **Beweis von Proposition §21.04.** In der Vorlesung sowie (ii)-(vi) Übungsaufgabe, wobei für (vi) zuerst gezeigt werden sollte, dass  $\mathbb{E}(|X - Y|) = 0$  gilt. □

§21.06 **Beispiel.**

- (a) Sei  $(X_i)_{i \in \llbracket n \rrbracket} \sim B_p^n$ ,  $n \in \mathbb{N}$ ,  $p \in [0, 1]$ , ein Bernoulli-Schema. Für jedes  $i \in \llbracket n \rrbracket$  ist  $X_i \sim B_p$  mit  $\mathbb{P}(X_i = 1) = p$  und  $\mathbb{P}(X_i = 0) = 1 - p$ , sodass  $X_i \in \mathcal{L}_1$  mit  $\mathbb{E}(X_i) = p$ . Da  $X = \sum_{i \in \llbracket n \rrbracket} X_i \sim \text{Bin}_{(n,p)}$  (vgl. **Beispiel §17.09 (a)**) ist auch  $X \in \mathcal{L}_1$  mit  $\mathbb{E}(X) = \sum_{i \in \llbracket n \rrbracket} \mathbb{E}(X_i) = np$ . Da der Erwartungswert nur von der Verteilung abhängt, sagt man zum Beispiel, dass die  $\text{Bin}_{(n,p)}$ -Verteilung den Erwartungswert  $np$  besitzt.
- (b) Sei  $X \sim \text{Poi}_\lambda$  mit  $\lambda \in \mathbb{R}_0^+$  und Zähldichte  $p_{\text{Poi}_\lambda}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ,  $k \in \mathbb{N}_0$  (vgl. **Beispiele §04.08 (f)**). Dann ist  $X \in \mathcal{L}_1$  mit

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k p_{\text{Poi}_\lambda}(k) = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\infty} p_{\text{Poi}_\lambda}(k) = \lambda.$$

- (c) Für  $Z \sim N_{(0,1)}$  gilt  $Z \in \mathcal{L}_1$  und auf Grund der symmetrischen Dichte  $\mathbb{E}(Z) = 0$ . □

## Weitere Eigenschaften

§21.07 **Lemma.** Für  $X \in \mathcal{L}_1$  ist  $\mathbb{P}(X = 0) = 1$  genau dann wenn  $\mathbb{E}(X \mathbb{1}_A) = 0$  für alle  $A \in \mathcal{A}$  gilt.

§21.08 **Beweis von Lemma §21.07.** In der Vorlesung. □

§21.09 **Lemma (Stetigkeit).** Sei  $X \in \mathcal{L}_1$ . Dann gilt

$$\forall \varepsilon \in \mathbb{R}_0^+ : \exists \delta_\varepsilon \in \mathbb{R}_0^+ : \forall A \in \mathcal{A} : \mathbb{P}(A) < \delta_\varepsilon \Rightarrow \mathbb{E}(|X| \mathbf{1}_A) \leq \varepsilon.$$

Insbesondere, für alle  $(A_n)_{n \in \mathbb{N}}$  aus  $\mathcal{A}$  mit  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0$  gilt  $\lim_{n \rightarrow \infty} \mathbb{E}(|X| \mathbf{1}_{A_n}) = 0$ .

§21.10 **Beweis** von **Lemma** §21.09. In der Vorlesung. □

§21.11 **Erinnerung.** Eine Funktion  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  heißt *konvex*, wenn

$$\phi(\lambda x + (1 - \lambda)y) \leq \lambda \phi(x) + (1 - \lambda)\phi(y)$$

für alle  $x, y \in \mathbb{R}$  und für alle  $\lambda \in [0, 1]$  gilt. □

§21.12 **Ungleichung von Jensen.** Seien  $X \in \mathcal{L}_1$  und  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  eine konvexe Funktion mit  $\phi(X) \in \mathcal{L}_1$ . Dann gilt  $\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$ .

§21.13 **Beweis** von **Satz** §21.12. In der Vorlesung. □

§21.14 **Beispiel.** Für  $p, q \in \mathbb{R}_0^+$  mit  $p < q$  seien  $|X|^p, |X|^q \in \mathcal{L}_1$ . Da  $\phi(x) = |x|^{q/p}$  konvex ist, folgt aus §§21.12 die *Ungleichung von Lyapounov*  $\mathbb{E}(|X|^p) \leq (\mathbb{E}(|X|^q))^{p/q}$ . □

## Konvergenzsätze

§21.15 **Satz von der monotonen Konvergenz.** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge aus  $\mathcal{L}$ , derart dass:

(M1)  $(X_n)_{n \in \mathbb{N}}$  ist monoton wachsend (fallend), also  $X_n \uparrow X$  (bzw.  $X_n \downarrow X$ ) für ein  $X \in \overline{\mathcal{A}}$ ;

(M2)  $(|\mathbb{E}(X_n)|)_{n \in \mathbb{N}}$  ist beschränkt, das heißt,  $\sup_{n \in \mathbb{N}} |\mathbb{E}(X_n)| < \infty$ .

Dann gilt  $X \in \mathcal{L}$  und  $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$  (bzw.  $\mathbb{E}(X_n) \downarrow \mathbb{E}(X)$ ).

§21.16 **Beweis** von **Satz** §21.15. In der Vorlesung. □

§21.17 **Satz von der dominierten Konvergenz.** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge aus  $\mathcal{L}$  derart, dass gelten:

(D1)  $(X_n)_{n \in \mathbb{N}}$  ist (punktweise) konvergent, also  $\lim_{n \rightarrow \infty} X_n = X$  für ein  $X \in \overline{\mathcal{A}}$ ;

(D2) Es existiert  $Y \in \mathcal{L}_1$  mit  $\sup_{n \in \mathbb{N}} |X_n| \leq Y$ , also  $\sup_{n \in \mathbb{N}} |X_n| \in \mathcal{L}_1$ .

Dann gilt  $X \in \mathcal{L}$ ,  $\lim_{n \rightarrow \infty} \mathbb{E}(|X - X_n|) = 0$  und  $\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$ .

§21.18 **Beweis** von **Satz** §21.17. In der Vorlesung. □

§21.19 **Satz von der konvergenten Reihe.** Sei  $(X_n)_{n \in \mathbb{N}}$  eine Folge aus  $\mathcal{L}_1$  mit  $\sum_{n \in \mathbb{N}} \mathbb{E}(|X_n|) < \infty$ . Dann gilt:

(i) Die Reihe  $\sum_{n \in \mathbb{N}} X_n$  konvergiert absolut fast sicher, also  $\mathbb{P}(\sum_{n \in \mathbb{N}} |X_n| < \infty) = 1$ ;

(ii)  $\sum_{n \in \mathbb{N}} X_n \in \mathcal{L}_1$  und

(iii)  $\mathbb{E}(\sum_{n \in \mathbb{N}} X_n) = \sum_{n \in \mathbb{N}} \mathbb{E}(X_n)$ .

§21.20 **Beweis** von **Satz** §21.19. Übungsaufgabe. Hinweis: Benutzen Sie **Lemma** §20.19 und die Endlichkeitsaussage aus **Lemma** §20.13 (ii) um zu zeigen, dass  $Y := \sum_{n \in \mathbb{N}} |X_n| < \infty$   $\mathbb{P}$ -f.s., und dass  $Y \in \mathcal{L}_1$ . Anschließend wenden Sie den **Beweis** §21.18 von der dominierten Konvergenz auf die Folge der Partialsummen an. □

§21.21 **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** §21.04 (vi)). □

## §22 Variablentransformation

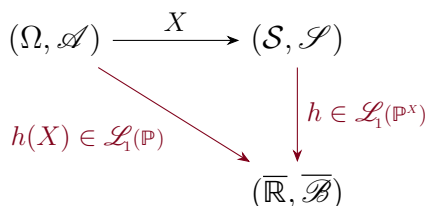
§22.01 **Satz.** Sei  $X$  eine  $(\mathcal{S}, \mathcal{S})$ -wertige Zufallsvariable auf  $(\Omega, \mathcal{A}, \mathbb{P})$ , sei  $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$  die Verteilung von  $X$  auf  $(\mathcal{S}, \mathcal{S})$  und sei  $h \in \overline{\mathcal{S}}$  eine numerische Zufallsvariable auf  $(\mathcal{S}, \mathcal{S})$ .

(i) Für  $h \geq 0$ , also  $h \in \overline{\mathcal{S}}^+$ , gilt

$$\mathbb{P}(h(X)) = \int_{\Omega} h(X) d\mathbb{P} = \mathbb{E}_{\mathbb{P}}(h(X)) = \mathbb{E}_{\mathbb{P}^X}(h) = \int_{\mathcal{S}} h d\mathbb{P}^X = \mathbb{P}^X(h). \quad (22.1)$$

(ii)  $h(X) \in \mathcal{L}_1(\Omega, \mathcal{A}, \mathbb{P})$  gilt genau dann, wenn  $h \in \mathcal{L}_1(\mathcal{S}, \mathcal{S}, \mathbb{P}^X)$ . In diesem Fall ist (22.1) ebenfalls erfüllt.

§22.02 **Skizze.**



□

§22.03 **Bemerkung.** Analog zu **Bemerkung** §21.03 werden in der Maßtheorie weiterhin die folgenden zwei Darstellungen gezeigt:

(i) Sei  $X$  ein *stetig-verteilter* Zufallsvektor im  $\mathbb{R}^n$  mit Dichte  $f^X$  und sei  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  Borel-messbar. Dann gilt  $h \in \mathcal{L}_1(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}^X)$  genau dann wenn  $\int_{\mathbb{R}^n} |h(x)| f^X(x) dx < \infty$ . In diesem Fall ist

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}^n} h(x) f^X(x) dx = \mathbb{P}^X(h).$$

(ii) Seien  $X$  eine *diskret-verteilte*  $\mathcal{X} := X(\Omega)$ -wertige Zufallsvariable mit Zähldichte  $\mathbb{p}^X$  und  $h : \mathcal{X} \rightarrow \mathbb{R}$ , 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

$$\mathbb{E}(h(X)) = \sum_{x \in \mathcal{X}} h(x) \mathbb{p}^X(x) = \mathbb{P}^X(h).$$

□

§22.04 **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  $(W_1, W_2) \sim \text{Lap}_{[1,6]}^2$  (vgl. **Beispiel** §11.12 (a)). Dann gilt

$$\mathbb{E}(|W_1 - W_2|) = \sum_{(i,j) \in [6]^2} |i - j| \frac{1}{36} = \sum_{i \in [6]} \sum_{j \in [6]} |i - j| \frac{1}{36} = \frac{35}{18}$$

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  $X \sim \text{Poi}_{\lambda}$  mit  $\lambda \in \mathbb{R}_{>0}^+$  und Zähldichte  $\mathbb{p}_{\text{Poi}_{\lambda}}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ,  $k \in \mathbb{N}_0$  (vgl. **Beispiel** §21.06 (b)). Dann gilt

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) \mathbb{p}_{\text{Poi}_{\lambda}}(k) = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2 \sum_{k=0}^{\infty} \mathbb{p}_{\text{Poi}_{\lambda}}(k) = \lambda^2$$

und somit  $\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \lambda^2 + \lambda$ .

- (c) Für  $Z \sim N_{(0,1)}$  gilt mit partieller Integration unter Benutzung von  $f'_{N_{(0,1)}}(z) = -zf_{N_{(0,1)}}(z)$  sowie  $0 = \mathbb{E}(Z) = \int_{\mathbb{R}} zf_{N_{(0,1)}}(z)dz$

$$\mathbb{E}(Z^2) = \int_{\mathbb{R}} z^2 f_{N_{(0,1)}}(z) = - \int_{\mathbb{R}} z f'_{N_{(0,1)}}(z) dz = \int_{\mathbb{R}} f_{N_{(0,1)}}(z) dz = 1.$$

Alternativ, da  $Y = Z^2 \sim \chi_1^2$  gilt auch  $\mathbb{E}(Z^2) = \int_0^\infty y(2\pi y)^{-1/2} e^{-y/2} dy = 1$ .

- (d) 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  $Z_1 \sim N_{(0,1)}$  und  $Z_2 \sim N_{(0,1)}$  (vgl. **Beispiel §11.12 (b)**). Weiterhin für  $Z \sim N_{(0,1)}$  ist  $V = \sqrt{1-\rho^2}Z \sim N_{(0,1-\rho^2)}$  (vgl. **Beispiel §10.12 (a)**) mit Dichte  $f^V(v) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp(-\frac{v^2}{2(1-\rho^2)})$  und

$$0 = \sqrt{1-\rho^2}\mathbb{E}(Z) = \mathbb{E}(V) = \int_{\mathbb{R}} v f^V(v) dv = \int_{\mathbb{R}} \frac{v}{\sqrt{2\pi(1-\rho^2)}} \exp(-\frac{v^2}{2(1-\rho^2)}) dv.$$

Somit gilt, unter Verwendung der gemeinsamen Dichte  $f^{(Z_1, Z_2)}$  von  $(Z_1, Z_2)$  und der Substitution von  $z_2 - \rho z_1$  durch  $v$ ,

$$\begin{aligned} \mathbb{E}(Z_1(Z_2 - \rho Z_1)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} z_1(z_2 - \rho z_1) f^{(Z_1, Z_2)}(z_1, z_2) dz_1 dz_2 \\ &= \int_{\mathbb{R}} \frac{z_1}{\sqrt{2\pi}} \exp(-\frac{(1-\rho^2)z_1^2}{2(1-\rho^2)}) \int_{\mathbb{R}} \frac{(z_2 - \rho z_1)}{\sqrt{2\pi(1-\rho^2)}} \exp(-\frac{(z_2 - \rho z_1)^2}{2(1-\rho^2)}) dz_2 dz_1 \\ &= \int_{\mathbb{R}} z_1 f^{Z_1}(z_1) \int_{\mathbb{R}} v f^V(v) dv dz_1 = \mathbb{E}(Z_1)\mathbb{E}(V) = 0. \end{aligned}$$

Folglich gilt  $\mathbb{E}(Z_1 Z_2) = \mathbb{E}(Z_1(Z_2 - \rho Z_1)) + \rho \mathbb{E}(Z_1^2) = \rho$ , da  $\mathbb{E}(Z_1^2) = 1$  nach (b).  $\square$

§22.05 **Proposition.** Sei  $X \in \overline{\mathcal{A}}^+$ , dann gilt  $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > y) dy$ .

§22.06 **Beweis** von **Proposition §22.05**. Übungsaufgabe.  $\square$

## §23 $\mathcal{L}_p$ -integrierbare Zufallsvariablen

Sei  $(\Omega, \mathcal{A}, \mathbb{P})$  ein Wahrscheinlichkeitsraum.

§23.01 **Definition.** Sei  $X \in \overline{\mathcal{A}}$  eine numerische Zufallsvariable.

- (a) Für  $p \in \mathbb{R}_{>0}^+$  definiere

$$\|X\|_{\mathcal{L}_p} := \begin{cases} (\mathbb{E}(|X|^p))^{1/p} & , \quad \text{für } p \in \mathbb{R}_{>0}^+, \\ \inf \{x \in \mathbb{R}^+ : \mathbb{P}(|X| > x) = 0\} & , \quad \text{für } p = \infty. \end{cases}$$

$X$  heißt  $\mathcal{L}_p$ -*integrierbar*, wenn  $\|X\|_{\mathcal{L}_p} < \infty$ . Die Menge aller  $\mathcal{L}_p$ -integrierbaren Zufallsvariablen bezeichnen wir mit  $\mathcal{L}_p := \mathcal{L}_p(\mathbb{P}) := \mathcal{L}_p(\Omega, \mathcal{A}, \mathbb{P}) := \{X \in \overline{\mathcal{A}} \mid \|X\|_{\mathcal{L}_p} < \infty\}$ .

- (b) Für  $X \in \mathcal{L}_p$  und  $p \in \mathbb{N}$  heißt  $\mathbb{E}(X^p)$  das  $p$ -te *Moment* von  $X$ ; für  $X \in \mathcal{L}_p$  und  $p \in \mathbb{R}_{>0}^+$  heißt  $\mathbb{E}(|X|^p)$  das  $p$ -te *absolute Moment* von  $X$ .  $\square$

§23.02 **Bemerkung.**

- (i) Für  $p = 1$  stimmt die letzte Definition mit **Definition** §21.02 (b) überein.
- (ii) Für  $p \in \mathbb{R}_0^+$  gilt  $X \in \mathcal{L}_p$  genau dann, wenn  $|X|^p \in \mathcal{L}_1$  gilt.
- (iii) Für  $p = \infty$  gilt  $\mathbb{P}(|X| > \|X\|_{\mathcal{L}_\infty}) = 0$ , da die Funktion  $x \mapsto \mathbb{P}(|X| > x)$  rechtsstetig ist. □

§23.03 **Korollar.** Für  $p, q \in \mathbb{R}_0^+$  mit  $p \leq q$  und  $X \in \mathcal{L}_q$  gilt  $\|X\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_q}$  und somit  $\mathcal{L}_q \subseteq \mathcal{L}_p$ .

§23.04 **Beweis** von **Korollar** §23.03. In der Vorlesung. □

§23.05 **Lemma.** Seien  $X \in \overline{\mathcal{A}}$  und  $p \in \mathbb{R}_0^+$ .

- (i)  $\|X\|_{\mathcal{L}_p} = 0$  gilt genau dann, wenn  $\mathbb{P}(X = 0) = 1$  gilt;
- (ii) Für  $a \in \mathbb{R}$  gilt  $\|aX\|_{\mathcal{L}_p} = |a|\|X\|_{\mathcal{L}_p}$ ;
- (iii)  $\|1\|_{\mathcal{L}_p} = 1$ ;
- (iv) Für  $X \in \mathcal{L}_p$  gilt  $\mathbb{P}(|X| < \infty) = 1$ ;
- (v) Für  $Y \in \overline{\mathcal{A}}$  mit  $\mathbb{P}(X = Y) = 1$  gilt  $\|X\|_{\mathcal{L}_p} = \|Y\|_{\mathcal{L}_p}$ .

§23.06 **Beweis** von **Lemma** §23.05. Übungsaufgabe. □

§23.07 **Hölder Ungleichung.** Seien  $X, Y \in \overline{\mathcal{A}}$  und  $p, q \in [1, \infty]$  mit  $\frac{1}{p} + \frac{1}{q} = 1$ . Dann gilt:

$$\mathbb{E}(|XY|) \leq \|X\|_{\mathcal{L}_p} \|Y\|_{\mathcal{L}_q}.$$

§23.08 **Beweis** von **Satz** §23.07. In der Vorlesung. □

§23.09 **Cauchy-Schwarz Ungleichung.** Für  $X, Y \in \mathcal{L}_2$  gilt  $XY \in \mathcal{L}_1$  und

$$|\mathbb{E}(XY)|^2 \leq \|X\|_{\mathcal{L}_2}^2 \|Y\|_{\mathcal{L}_2}^2 = \mathbb{E}(|X|^2) \mathbb{E}(|Y|^2).$$

§23.10 **Beweis** von **Satz** §23.09. Die Aussage folgt direkt aus **Satz** §23.07. □

§23.11 **Minkowski Ungleichung.** Für  $X, Y \in \mathcal{L}_p$  mit  $p \in [1, \infty]$  gilt  $X + Y \in \mathcal{L}_p$  und

$$\|X + Y\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_p} + \|Y\|_{\mathcal{L}_p}.$$

§23.12 **Beweis** von **Satz** §23.11. In der Vorlesung. □

§23.13 **Bemerkung.** Für  $p \in [1, \infty]$  ist  $\|\cdot\|_{\mathcal{L}_p}$  eine Pseudonorm auf  $\mathcal{L}_p$ , das heißt für  $X, Y \in \mathcal{L}_p$  und  $a \in \mathbb{R}$  gilt:

- (i)  $\|aX\|_{\mathcal{L}_p} = |a|\|X\|_{\mathcal{L}_p}$  (**Lemma** §23.05 (ii))
- (ii)  $\|X + Y\|_{\mathcal{L}_p} \leq \|X\|_{\mathcal{L}_p} + \|Y\|_{\mathcal{L}_p}$  (Minkowski-Ungleichung §§23.11)
- (iii)  $\|X\|_{\mathcal{L}_p} \geq 0$  für alle  $X$  und  $\|X\|_{\mathcal{L}_p} = 0$ , falls  $X = 0$   $\mathbb{P}$ -f.s.. (**Lemma** §23.05 (i)) □



## §24 Varianz, Kovarianz und Korrelation

§24.01 **Erinnerung..** Für  $X, Y \in \mathcal{L}_2$  folgt aus §§23.07, dass  $X, Y, XY \in \mathcal{L}_1$  sind.  $\square$

§24.02 **Definition.** Für Zufallsvariablen  $X, Y \in \mathcal{L}_2$  bezeichnet

$$\text{Cov}(X, Y) := \mathbb{E}(\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

die *Kovarianz* zwischen  $X$  und  $Y$ .

$$\text{Var}(X) := \text{Cov}(X, X) = \mathbb{E}(\{X - \mathbb{E}(X)\}^2) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 \quad \text{und} \quad \text{std}(X) := \sqrt{\text{Var}(X)}$$

heißt die *Varianz* bzw. *Standardabweichung* (standard deviation) von  $X$ .

§24.03 **Lemma.** Seien  $X, Y, Z \in \mathcal{L}_2$  und  $a, b \in \mathbb{R}$ .

- (i) Es gilt  $\text{Var}(X) = 0$  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)  $\text{Cov} : \mathcal{L}_2 \times \mathcal{L}_2 \rightarrow \mathbb{R}$  ist eine positiv semi-definite, symmetrische Bilinearform, das heißt
  - (a)  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$  (symmetrisch)
  - (b)  $\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$  (linear)
  - (c)  $\text{Cov}(X, X) \geq 0$  (positiv semi-definit)

und es gilt weiterhin  $\text{Cov}(a, X) = 0$ .

- (iii)  $\text{Var} : \mathcal{L}_2 \rightarrow \mathbb{R}^+$  ist die von  $\text{Cov}$  induzierte quadratische Form, sodass

- (a)  $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- (b)  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$

gelten.

§24.04 **Beweis** von Lemma §24.03. Übung  $\square$

§24.05 **Beispiel** (Beispiel §22.04 fortgesetzt.).

- (a) Sei  $X \sim \text{Poi}_\lambda$  mit  $\lambda \in \mathbb{R}_0^+$ . Dann gilt  $\mathbb{E}(X) = \lambda$  und  $\mathbb{E}(X^2) = \lambda^2 + \lambda$  (vgl. Beispiel §22.04 (b)), sodass  $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$  gilt.
- (b) Für  $Z \sim N_{(0,1)}$  gilt  $\mathbb{E}(Z) = 0$  und  $\mathbb{E}(Z^2) = 1$  (vgl. Beispiel §22.04 (c)), so dass  $\text{Var}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2 = 1$ . Für  $\mu \in \mathbb{R}$  und  $\sigma \in \mathbb{R}_0^+$  ist  $X = \mu + \sigma Z \sim N_{(\mu, \sigma^2)}$  (vgl. Beispiel §10.12 (a)), sodass  $\mathbb{E}(X) = \sigma \mathbb{E}(Z) + \mu = \mu$  und  $\text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2$  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  $X \sim N_{(\mu_1, \sigma_1^2)}$  und  $Y \sim N_{(\mu_2, \sigma_2^2)}$  (vgl. Beispiel §11.12 (b)), sowie  $Z_1 := \sigma_1^{-1}(X - \mu_1) \sim N_{(0,1)}$  und  $Z_2 := \sigma_2^{-1}(Y - \mu_2) \sim N_{(0,1)}$  (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 §22.04 (d)). Insbesondere, gilt damit  $\text{Cov}(X, Y) = \mathbb{E}(\{X - \mu_1\}\{Y - \mu_2\}) = \mathbb{E}(\{\sigma_1 Z_1\}\{\sigma_2 Z_2\}) = \sigma_1 \sigma_2 \rho$ .  $\square$

§24.06 **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<sup>2</sup>-Varianz-Zerlegung*

$$\mathbb{E}(|Y - a|^2) = |\mathbb{E}(Y) - a|^2 + \mathbb{V}\text{ar}(Y).$$

Die Abbildung  $a \mapsto \mathbb{E}(|Y - a|^2)$  nimmt ihr Minimum auf  $\mathbb{R}$  genau bei  $a^* = \mathbb{E}(Y)$  an, sodass

$$\min_{a \in \mathbb{R}} \mathbb{E}(|Y - a|^2) = \mathbb{V}\text{ar}(Y) \quad \text{und} \quad \{\mathbb{E}(Y)\} = \arg \min_{a \in \mathbb{R}} \mathbb{E}(|Y - a|^2)$$

gilt und  $a^* = \mathbb{E}(Y)$  *beste konstante Vorhersage* von  $Y$  genannt wird. □

§24.07 **Markov Ungleichung.** Für  $Y \in \overline{\mathcal{A}}^+$  und  $p, \varepsilon \in \mathbb{R}_0^+$  gilt  $\varepsilon^p \mathbb{1}_{\{Y > \varepsilon\}} \leq Y^p$ , sodass *Proposition §21.04 (iv)* impliziert  $\mathbb{P}(Y > \varepsilon) \leq \varepsilon^{-p} \mathbb{E}(Y^p)$ . □

§24.08 **Tschebischeff Ungleichung.** Für  $X \in \mathcal{L}_2$  und  $\varepsilon \in \mathbb{R}_0^+$  impliziert das Anwenden von *Satz §24.07* auf  $Y = |X - \mathbb{E}(X)|$ , dass  $\mathbb{P}(|X - \mathbb{E}(X)| > \varepsilon) \leq \varepsilon^{-2} \mathbb{V}\text{ar}(X)$ . □

§24.09 **Bemerkung.** Der Erwartungswert  $\mu := \mathbb{E}(X)$  und die Varianz  $\sigma^2 := \mathbb{V}\text{ar}(X)$  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  $(\Omega, \mathcal{A}, \mathbb{P})$  häufig nicht angegeben wird, da die Beobachtung von  $X$  nur Rückschlüsse auf die Verteilung von  $X$  erlaubt. Betrachten wir die Varianz, so ist diese eine Maß der Streuung von  $X$  um seinen Erwartungswert, da nach §§24.08  $\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$  für alle  $k \in \mathbb{N}$  gilt. Damit erhalten wir zum Beispiel  $\mathbb{P}(|X - \mu| \geq \sigma) \leq 1$ ,  $\mathbb{P}(|X - \mu| \geq 2\sigma) \leq 0,25$  sowie  $\mathbb{P}(|X - \mu| \geq 3\sigma) \leq 0,12$ . Die Abschätzung §§24.08 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  $X \sim N(\mu, \sigma^2)$ , gilt  $\mathbb{P}(|X - \mu| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$  (vgl. *Beispiel §10.12 (a)*), sodass  $\mathbb{P}(|X - \mu| \geq k\sigma) = 2 - 2\Phi(k)$  und insbesondere  $\mathbb{P}(|X - \mu| \geq \sigma) \approx 0,32$ ,  $\mathbb{P}(|X - \mu| \geq 2\sigma) \approx 0,04$  sowie  $\mathbb{P}(|X - \mu| \geq 3\sigma) \approx 0,0026$ . □

§24.10 **Erinnerung..** Für  $X, Y \in \mathcal{L}_2$  gilt  $|\text{Cov}(X, Y)| \leq \text{std}(X) \text{std}(Y)$  durch Anwenden der Cauchy-Schwarz Ungleichung §§23.09. □

§24.11 **Definition.** Für  $X, Y \in \mathcal{L}_2$  mit  $\text{std}(X) > 0$  und  $\text{std}(Y) > 0$  bezeichnet

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\text{std}(X) \text{std}(Y)} \in [-1, 1]$$

die *Korrelation* zwischen  $X$  und  $Y$ . Falls  $\text{Cov}(X, Y) = 0$  gilt, heißen  $X$  und  $Y$  *unkorreliert*.

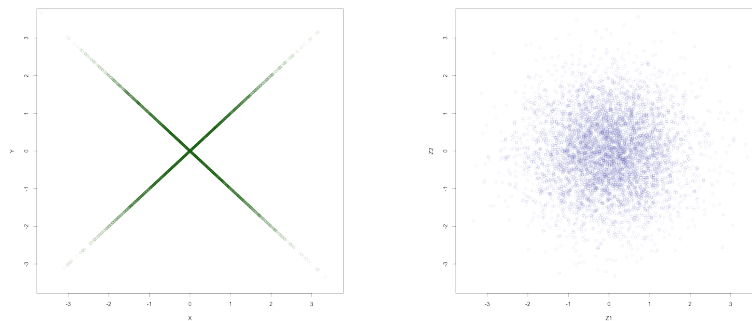
§24.12 **Korollar.**

- (i) Für unkorrelierte Zufallsvariablen  $X, Y \in \mathcal{L}_2$  gilt  $\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$ .
- (ii) Unabhängige Zufallsvariablen  $X, Y \in \mathcal{L}_2$  sind unkorreliert. Die Umkehrung gilt nicht.

§24.13 **Beweis von Korollar §24.12.** Die Aussage (i) folgt aus *Lemma §24.03 (iii) (b)* und die Aussage (ii) direkt aus *Proposition §21.04 (vii)* und *Beispiel §24.14*. □

§24.14 **Beispiel.**

- (a) Sei  $(X_i)_{i \in \llbracket n \rrbracket} \sim B_p^n$ ,  $n \in \mathbb{N}$ ,  $p \in [0, 1]$ , ein Bernoulli-Schema. Für jedes  $i \in \llbracket n \rrbracket$  ist  $X_i \sim B_p$  mit  $\mathbb{P}(X_i = 1) = p$  und  $\mathbb{P}(X_i = 0) = 1 - p$ , sodass  $X_i \in \mathcal{L}_2$  mit  $\mathbb{E}(X_i^2) = p$  und  $\text{Var}(X_i) = p(1 - p)$ . Da  $X = \sum_{i \in \llbracket n \rrbracket} X_i \sim \text{Bin}(n, p)$  (vgl. **Beispiel §17.09 (a)**) und  $\perp_{i \in \llbracket n \rrbracket} X_i$  ist auch  $X \in \mathcal{L}_2$  mit  $\text{Var}(X) = \sum_{i \in \llbracket n \rrbracket} \text{Var}(X_i) = np(1 - p)$ . Im Fall  $p \in \{0, 1\}$  gilt also  $\text{Var}(X) = 0$  sowie stets  $\text{Var}(X) \leq n/4$ . Die relative Häufigkeit von Erfolgen  $\hat{p} := X/n$  erfüllt  $\mathbb{E}(\hat{p}) = p$  und  $\text{Var}(\hat{p}) = \frac{p(1-p)}{n} \leq \frac{1}{4n}$ .
- (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  $(W_1, W_2) \sim \text{Lap}_{[6]}^2$  (vgl. **Beispiel §11.12 (a)**). Dann sind  $X$  und  $Y$  nicht unabhängig (vgl. **Beispiel §16.09 (b)**), aber  $X$  und  $Y$  sind unkorreliert, da  $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \frac{245}{18} - \frac{35}{18} \cdot 7 = 0$  (vgl. **Beispiel §22.04 (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  $\text{std}(X) = \sigma_1$  und  $\text{std}(Y) = \sigma_2$  sowie  $\text{Cov}(X, Y) = \sigma_1\sigma_2\rho$  (vgl. **Beispiel §24.05 (b)** bzw. (c)) und somit  $\rho(X, Y) = \rho$ . 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  $X \sim N_{(\mu_1, \sigma_1^2)}$  und  $Y \sim N_{(\mu_2, \sigma_2^2)}$  unkorreliert sind, aber der Vektor  $(X, Y)$  nicht bivariat normalverteilt ist. Betrachte dazu zwei unabhängige Zufallsvariablen  $X$  und  $V$ , wobei  $X \sim N_{(0,1)}$  und  $V$  ist eine Rademacher-Zufallsvariable, d.h.  $V \in \{-1, 1\}$  mit  $P(V = -1) = 1/2 = P(V = 1)$ . Es ist nun leicht zu zeigen, dass die Zufallsvariablen  $Y := VX$  und  $X$  unkorreliert sind und dass  $Y \sim N_{(0,1)}$  (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.



□

§24.15 **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  $\text{Var}(X) = 0$ , so befinden wir uns im Fall der *besten konstanten Vorhersage* §§24.06, sodass wir im Folgenden  $\text{Var}(X) > 0$  annehmen und mit

$$L_X := \{aX + b \mid a, b \in \mathbb{R}\} \subseteq \mathcal{L}_2$$

die Menge der Zufallsvariablen, die affine Funktionen von  $X$  sind, bezeichnen. □

§24.16 **Beste lineare Vorhersage.** Seien  $X, Y \in \mathcal{L}_2$  mit  $\mathbb{V}\text{ar}(X) > 0$ . Für jedes  $Z = aX + b \in L_X$  gilt die **Bias<sup>2</sup>+Varianz-Zerlegung**

$$\mathbb{E}(|Y - Z|^2) = |\mathbb{E}(Y) - \mathbb{E}(Z)|^2 + \mathbb{V}\text{ar}(Y - Z) = |\mathbb{E}(Y) - a\mathbb{E}(X) - b|^2 + \mathbb{V}\text{ar}(Y - aX).$$

Da für  $a^* := \frac{\text{Cov}(X, Y)}{\mathbb{V}\text{ar}(X)}$  gilt  $\text{Cov}(Y - a^*X, X) = \text{Cov}(Y, X) - a^*\mathbb{V}\text{ar}(X) = 0$ , folgt

$$\begin{aligned}\mathbb{V}\text{ar}(Y - aX) &= \mathbb{V}\text{ar}(Y - a^*X + (a^* - a)X) = \mathbb{V}\text{ar}(Y - a^*X) + (a^* - a)^2 \mathbb{V}\text{ar}(X); \\ \mathbb{V}\text{ar}(Y - a^*X) &= \mathbb{V}\text{ar}(Y) + (a^*)^2 \mathbb{V}\text{ar}(X) - 2\text{Cov}(Y, a^*X) \\ &= \mathbb{V}\text{ar}(Y) - \frac{\text{Cov}(X, Y)^2}{\mathbb{V}\text{ar}(X)} = \mathbb{V}\text{ar}(Y)(1 - \rho^2(X, Y)).\end{aligned}$$

Für  $b^* := \mathbb{E}(Y) - a^*\mathbb{E}(X)$  und  $Z^* = a^*X + b^*$  gilt somit

$$\mathbb{E}(|Y - Z|^2) \geq \mathbb{V}\text{ar}(Y - aX) \geq \mathbb{V}\text{ar}(Y - a^*X) = \mathbb{E}(|Y - Z^*|^2)$$

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

$$\begin{aligned}\min_{Z \in L_X} \mathbb{E}(|Y - Z|^2) &= \mathbb{E}(|Y - Z^*|^2) = \mathbb{V}\text{ar}(Y)(1 - \rho^2(X, Y)) \\ \text{und} \quad \left\{ \mathbb{E}(Y) + \frac{\text{Cov}(X, Y)}{\mathbb{V}\text{ar}(X)}(X - \mathbb{E}(X)) \right\} &= \{Z^*\} = \arg \min_{Z \in L_X} \mathbb{E}(|Y - Z|^2)\end{aligned}$$

gilt.  $Z^* = a^*X + b^*$  wird **beste lineare Vorhersage** von  $Y$  durch  $X$  genannt.  $\square$

§24.17 **Bemerkung.** Da  $\min_{Z \in L_X} \mathbb{E}(|Y - Z|^2) = \mathbb{V}\text{ar}(Y)(1 - \rho^2(X, Y))$  gilt, interpretiert man  $\rho(X, Y)$  auch als Maß für die Stärke der **linearen** Abhängigkeit zwischen  $X$  und  $Y$ . Zur Erinnerung  $\rho(X, Y) \in [-1, 1]$  und im Extremfall  $|\rho(X, Y)| = 1$  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 §20.13 (i)**, sodass bis auf einer Nullmenge  $Y$  gerade gleich einer affinen Funktion von  $X$  ist.  $\square$

## §25 Hauptkomponentenanalyse

§25.01 **Erinnerung.** Im Folgenden fassen wir wieder Vektoren als Spaltenvektoren auf, das heißt  $a = (a_1 \cdots a_n)^t \in \mathbb{R}^n$ . Wir bezeichnen mit  $\|\cdot\|$  die vom Standardskalarprodukt  $\langle \cdot, \cdot \rangle$  auf  $\mathbb{R}^n$  induzierte Norm, das heißt,  $\|a\| = \sqrt{\langle a, a \rangle} = \sqrt{a^t a} = (\sum_{i \in \llbracket n \rrbracket} a_i^2)^{1/2}$  für alle  $a \in \mathbb{R}^n$ . Weiterhin setzen wir  $|a| := \sum_{i \in \{n, j \in \mathbb{N}\}} |a_i|$ .  $\square$

§25.02 **Definition.** Sei  $X = (X_1 \cdots X_n)^t$  ein Zufallsvektor (aufgefasst als Spaltenvektor).

(a) Falls  $|X| = \sum_{i \in \llbracket n \rrbracket} |X_i| \in \mathcal{L}_1$ , also  $X_i \in \mathcal{L}_1$ ,  $i \in \llbracket n \rrbracket$ , kurz  $X \in \mathcal{L}_1$ , dann heißt

$$\mathbb{E}(X) = (\mathbb{E}(X_1) \cdots \mathbb{E}(X_n))^t = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix} \in \mathbb{R}^n$$

**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

$$\begin{aligned} \text{Cov}(X) &= (\text{Cov}(X_i, X_j))_{i,j \in [n]} = \left( \mathbb{E}(\{X_i - \mathbb{E}(X_i)\}\{X_j - \mathbb{E}(X_j)\}) \right)_{i,j \in [n]} \\ &= \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^t) = \mathbb{E}(XX^t) - \mathbb{E}(X)(\mathbb{E}(X))^t \in \mathbb{R}^{(n,n)} \end{aligned}$$

*Kovarianzmatrix* von  $X$ . □

§25.03 **Bemerkung.** Für die Kovarianzmatrix  $\Sigma := \text{Cov}(X)$  eines  $\mathbb{R}^n$ -wertigen Zufallsvektors  $X$ , falls sie existiert, gilt  $\text{Cov}(\langle a, X \rangle, \langle X, b \rangle) = \sum_{i \in [n]} \sum_{j \in [n]} a_i b_j \text{Cov}(X_i, X_j) = b^t \Sigma a = \langle \Sigma a, b \rangle$  für alle  $a, b \in \mathbb{R}^n$ . Insbesondere ist  $\Sigma$  *symmetrisch*, da  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$  für alle  $i, j \in [n]$ , und *positiv semidefinit*, da für alle  $c \in \mathbb{R}^n$  gilt  $\langle \Sigma c, c \rangle = \text{Var}(\langle X, c \rangle) \geq 0$ . □

§25.04 **Lemma.** Sei  $X$  ein  $\mathbb{R}^m$ -wertiger Zufallsvektor in  $\mathcal{L}_2$ . Für alle  $b \in \mathbb{R}^n$  und  $A \in \mathbb{R}^{(n,m)}$  ist dann  $Y = AX + b$  ein  $\mathbb{R}^n$ -wertiger Zufallsvektor in  $\mathcal{L}_2$ . Bezeichnen wir weiterhin mit  $\mu := \mathbb{E}(X) \in \mathbb{R}^m$  und  $\Sigma := \text{Cov}(X) \in \mathbb{R}^{(m,m)}$  den Erwartungswertvektor und die Kovarianzmatrix von  $X$ , dann gilt  $\mathbb{E}(Y) = A\mu + b$  und  $\text{Cov}(Y) = A\Sigma A^t$ .

§25.05 **Beweis** von Lemma §25.04. In der Vorlesung. □

§25.06 **Bemerkung.** Ist  $X$  ein  $\mathbb{R}^m$ -wertiger Zufallsvektor in  $\mathcal{L}_2$  mit  $\mu := \mathbb{E}(X) \in \mathbb{R}^m$  und  $\Sigma := \text{Cov}(X) \in \mathbb{R}^{(m,m)}$ . Dann gilt  $\mathbb{E}(\langle X, a \rangle) = \langle \mu, a \rangle$  und  $\text{Cov}(\langle a, X \rangle, \langle X, b \rangle) = \langle \Sigma a, b \rangle$  für alle  $a, b \in \mathbb{R}^n$ . □

§25.07 **Beispiel.** Für  $X \sim N_{(\mu, \Sigma)}$  ist  $\mu = \mathbb{E}(X) \in \mathbb{R}^n$  der Erwartungswertvektor und  $\Sigma = \text{Cov}(X) \in \mathbb{R}^{(n,n)}$  die Kovarianzmatrix von  $X$ . □

§25.08 **Erinnerung.** Sei  $\Sigma$  eine positiv semidefinite Matrix. Da  $\Sigma$  symmetrisch ist, ist  $\Sigma$  insbesondere diagonalisierbar, also, ähnlich zu einer Diagonalmatrix  $\text{Diag}(\lambda)$  mit reellen Diagonaleinträgen  $\lambda = (\lambda_i)_{i \in [n]}$ , genannt *Eigenwerte* von  $\Sigma$ . Es existiert also eine orthogonale Matrix  $U$ , also  $U^t U = E_n = U U^t$ , deren Spalten  $(u_i)_{i \in [n]}$ , genannt *Eigenvektoren*, eine Orthonormalbasis des  $\mathbb{R}^n$  bilden, derart dass  $U^t \Sigma U = \text{Diag}(\lambda)$  gilt. Da  $\Sigma$  positiv semidefinit ist, sind alle Eigenwerte positiv, also  $\lambda_i \in \mathbb{R}^+$ ,  $i \in [n]$ . Im Folgenden nehmen wir an, dass die orthogonale Matrix  $U$  so gewählt ist, dass die Eigenwerte der Größe nach angeordnet sind, das heißt  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$  gilt. □

§25.09 **Definition.** Sei  $X$  ein  $\mathbb{R}^n$ -wertiger Zufallsvektor in  $\mathcal{L}_2$  mit positiv semidefiniter Kovarianzmatrix  $\text{Cov}(X) = \Sigma \in \mathbb{R}^{(n,n)}$ . Weiterhin seien  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$  die der Größe nach geordneten Eigenwerte der Matrix  $\Sigma$  und  $u_1, \dots, u_n \in \mathbb{R}^n$  die zugehörigen Eigenvektoren. Dann heißen  $u_1, \dots, u_n$  *Hauptachsen* von  $\Sigma$  und die reellen Zufallsvariablen  $\langle X, u_1 \rangle, \dots, \langle X, u_n \rangle$  *Hauptkomponenten* von  $X$ . □

§25.10 **Lemma.** Seien  $u_1, \dots, u_n \in \mathbb{R}^n$  Hauptachsen zu den der Größe nach angeordneten Eigenwerte  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$  der Kovarianzmatrix  $\Sigma = \text{Cov}(X)$  eines  $\mathbb{R}^n$ -wertigen Zufallsvektors  $X$  in  $\mathcal{L}_2$ . Dann sind die Hauptkomponenten  $(\langle X, u_i \rangle)_{i \in [n]}$  unkorreliert. Es gilt  $\text{Cov}(\langle X, u_i \rangle, \langle X, u_j \rangle) = \langle \Sigma u_i, u_j \rangle = \lambda_i \mathbb{1}_{\{i=j\}}$ , also insbesondere  $\text{Var}(\langle X, u_i \rangle) = \lambda_i$ , für alle  $i, j \in [n]$ , und somit  $\text{Var}(\langle X, u_1 \rangle) \geq \text{Var}(\langle X, u_2 \rangle) \geq \dots \geq \text{Var}(\langle X, u_n \rangle)$ .

§25.11 **Beweis** von Lemma §25.10. Direktes Anwenden der Definition. □

§25.12 **Beispiel.** Seien  $X \sim N_{(\mu, \Sigma)}$ ,  $\lambda = (\lambda_i)_{i \in \llbracket n \rrbracket}$  die Eigenwerte von  $\Sigma$  und die Hauptachsen  $(u_i)_{i \in \llbracket n \rrbracket}$  von  $\Sigma$  die Spalten der orthogonalen Matrix  $U$ . Dann gilt  $U^t X \sim N_{(U^t \mu, \text{Diag}(\lambda))}$  und die Hauptkomponenten  $(\langle X, u_i \rangle)_{i \in \llbracket n \rrbracket}$  sind somit unabhängig.  $\square$

§25.13 **Definition.** Seien  $X$  und  $Y$  ein  $\mathbb{R}^k$ -wertiger bzw.  $\mathbb{R}^n$ -wertiger Zufallsvektor in  $\mathcal{L}_2$ . Für  $A^* \in \mathbb{R}^{(n, k)}$  und  $b^* \in \mathbb{R}^n$  heißt  $Z^* = A^* X + b^*$  eine *lineare Vorhersage* von  $Y$  durch  $X$ .  $Z^*$  wird *beste lineare Vorhersage* von  $Y$  durch  $X$  genannt, wenn für alle  $A \in \mathbb{R}^{(n, k)}$  und  $b \in \mathbb{R}^n$  gilt  $\mathbb{E} \|Y - Z^*\|^2 \leq \mathbb{E} \|Y - (AX + b)\|^2$ .  $\square$

*Die folgenden Ausführungen in §§25.14–§§25.19 erlauben das zentrale Resultat der Hauptkomponentenanalyse im Satz §25.20 zu beweisen, sind für diese Vorlesung aber nur weiterführendes Material.*

§25.14 **Bemerkung.** Für eine Matrix  $A \in \mathbb{R}^{(n, k)}$  heißt eine Matrix  $A^+ \in \mathbb{R}^{(k, n)}$  *Moore-Penrose Inverse* von  $A$ , wenn  $AA^+A = A$ ,  $A^+AA^+ = A^+$  und  $AA^+$  sowie  $A^+A$  symmetrisch sind. Die Moore-Penrose Inverse ist eindeutig festgelegt. Ist  $B \in \mathbb{R}^n$  positiv semidefinit, so dass  $U^t B U = \text{Diag}(\lambda)$  für eine orthogonale Matrix  $U$  und eine Diagonalmatrix  $\text{Diag}(\lambda)$  mit reellen Diagonaleinträgen  $\lambda = (\lambda_i)_{i \in \llbracket n \rrbracket}$ . Setzen wir  $\lambda^+ = (\lambda_i^+)_{i \in \llbracket n \rrbracket}$  mit  $\lambda_i^+ = \lambda_i^{-1} \mathbb{1}_{\mathbb{R}_0^+}(\lambda_i)$ , das heißt für  $\lambda_i \in \mathbb{R}_0^+$  ist  $\lambda_i^+ = \lambda_i^{-1}$  und ansonsten  $\lambda_i^+ = 0$ . Dann ist  $B^+ = U \text{Diag}(\lambda^+) U^t$  die Moore-Penrose Inverse von  $B$ . Allgemein ist  $A^+ = (A^t A)^+ A^t \in \mathbb{R}^{(k, n)}$  die Moore-Penrose Inverse von  $A \in \mathbb{R}^{(n, k)}$ .  $\square$

§25.15 **Lemma.** Seien  $X$  und  $Y$  ein  $\mathbb{R}^k$ -wertiger bzw.  $\mathbb{R}^n$ -wertiger Zufallsvektor in  $\mathcal{L}_2$ . Setzen wir  $\text{Cov}(Y, X) := \mathbb{E}((Y - \mathbb{E}(Y))(X - \mathbb{E}(X))^t)$  dann ist  $Z^* = \mathbb{E}(Y) + \text{Cov}(Y, X) \text{Cov}(X)^+(X - \mathbb{E}(X))$  die *beste lineare Vorhersage* von  $Y$  durch  $X$ . Der Fehler  $\varepsilon := Y - Z^*$  und  $AX$  für beliebiges  $A \in \mathbb{R}^{(n, k)}$  sind unkorreliert, also  $\text{Cov}(\varepsilon, AX) = 0$ . Es gilt  $\text{Cov}(\varepsilon) = \text{Cov}(Y) - \text{Cov}(Y, X) \text{Cov}(X)^+ \text{Cov}(X, Y)$  und  $\mathbb{E}(\varepsilon) = 0$ .

§25.16 **Beweis von Lemma §25.15.** Sei  $Z^* = A^* X + b^*$ , also  $A^* = \text{Cov}(Y, X) \text{Cov}(X)^+$  und  $b^* = \mathbb{E}(Y) - \text{Cov}(Y, X) \text{Cov}(X)^+ \mathbb{E}(X)$ . Dann gilt

$$\begin{aligned} \text{Cov}(Y - Z^*, AX) &= \mathbb{E}((Y - Z^*)X^t)A^t \\ &= \mathbb{E}((Y - \mathbb{E}(Y))X^t)A^t - \text{Cov}(Y, X) \text{Cov}(X)^+ \mathbb{E}((X - \mathbb{E}(X))X^t)A^t \\ &= \text{Cov}(Y, X)(E_n - \text{Cov}(X)^+ \text{Cov}(X))A^t = 0. \end{aligned}$$

Die letzte Gleichheit ist klar, falls  $\text{Cov}(X)$  regulär ist, ansonsten ist diese aufwendiger zu zeigen. (Wir nutzen aus, dass  $E_n - \text{Cov}(X)^+ \text{Cov}(X)$  eine Projektionsmatrix auf das orthogonale Komplement des Bildes von  $\text{Cov}(X)$ , und zeigen dass für jedes darin enthaltene Element  $v$  gilt  $v^t X = v^t \mathbb{E}(X)$  und somit  $\text{Cov}(Y, X)v = 0$ .) Für jede lineare Vorhersage  $Z = AX + b$  von  $Y$  durch  $X$  gilt dann  $\text{Cov}(Y - Z^*, Z^* - Z) = \text{Cov}(Y - Z^*, (A^* - A)X) = 0$  und somit

$$\begin{aligned} \text{Cov}(Y - Z) &= \text{Cov}(Y - Z^*) + \text{Cov}(Z^* - Z) + \text{Cov}(Y - Z^*, Z^* - Z) + \text{Cov}(Z^* - Z, Y - Z^*) \\ &= \text{Cov}(Y - Z^*) + (A^* - A) \text{Cov}(X) (A^* - A) \end{aligned}$$

damit ist die Matrix  $\text{Cov}(Y - Z) - \text{Cov}(Y - Z^*)$  positiv semidefinit, und wir schreiben kurz  $\text{Cov}(Y - Z) \geq \text{Cov}(Y - Z^*)$  und halten fest, dass  $\text{Spur}(\text{Cov}(Y - Z)) \geq \text{Spur}(\text{Cov}(Y - Z^*))$ . Wir benutzen weiterhin  $\mathbb{E} \|Y - Z\|^2 = \text{Spur}(\text{Cov}(Y - Z))$  und schließen  $\mathbb{E} \|Y - Z\|^2 = \text{Spur}(\text{Cov}(Y - Z)) \geq \text{Spur}(\text{Cov}(Y - Z^*)) = \mathbb{E} \|Y - Z^*\|^2$ , was zu zeigen war.  $\square$



§25.17 **Korollar.** Sei  $Y$  ein  $\mathbb{R}^n$ -wertiger Zufallsvektor in  $\mathcal{L}_2$  und  $X = AY$  für ein  $A \in \mathbb{R}^{(k,n)}$ . Dann ist  $Z^* = \mathbb{E}(Y) + \text{Cov}(Y)A^t(A\text{Cov}(Y)A^t)^+(X - \mathbb{E}(X))$  eine **beste lineare Vorhersage** von  $Y$  durch  $X = AY$ . Der Fehler  $\varepsilon := Y - Z^*$  und  $Z^*$  sind unkorreliert. Es gilt  $\mathbb{E}(\varepsilon) = 0$  und  $\text{Cov}(\varepsilon) = \text{Cov}(Y) - \text{Cov}(Y)A^t(A\text{Cov}(Y)A^t)^+A\text{Cov}(Y)$ .

§25.18 **Beweis von Korollar §25.17.** Folgt direkt aus Lemma §25.15.  $\square$

§25.19 **Bemerkung.** Seien die Spalten  $u_1, \dots, u_n \in \mathbb{R}^n$  der orthogonalen Matrix  $U$  Hauptachsen der Kovarianzmatrix  $\Sigma = \text{Cov}(Y)$  eines  $\mathbb{R}^n$ -wertigen Zufallsvektors  $Y$ . Für  $\mu := \mathbb{E}(Y)$  gilt dann  $Y - \mu = UU^t(Y - \mu) = \sum_{i \in [n]} \langle Y - \mu, u_i \rangle u_i$ . Bezeichnet  $(v_j)_{j \in [n]}$  eine beliebige Orthonormalbasis des  $\mathbb{R}^n$ , so gilt

$$\begin{aligned} \mathbb{E}\|Y - \mu\|^2 &= \sum_{j \in [n]} \mathbb{E}(\langle Y - \mu, v_j \rangle^2) = \sum_{j \in [n]} \text{Var}(\langle Y, v_j \rangle) = \sum_{j \in [n]} \langle \Sigma v_j, v_j \rangle \\ &= \text{Spur}(\Sigma) = \sum_{j \in [n]} \lambda_j. \end{aligned}$$

Im Folgenden nehmen wir an, dass  $Y$  zentriert ist, das heißt,  $\mu = 0$ . Für  $k \in [n]$  sei  $U_{(k)} \in \mathbb{R}^{(n,k)}$  die Matrix mit den Spalten  $(u_j)_{j \in [k]}$ . Für  $Z_{(k)} := U_{(k)}U_{(k)}^t Y = \sum_{i \in [k]} \langle Y, u_i \rangle u_i$  gilt dann  $\mathbb{E}\|Y - Z_{(k)}\|^2 = \sum_{j \in [k+1, n]} \mathbb{E}(\langle Y, u_j \rangle^2) = \sum_{j \in [k+1, n]} \lambda_j$ . Weiterhin ist  $k > \text{rg}(\text{Cov}(Y)) =: k^*$ , so gilt  $\lambda_l = 0$  für jedes  $l \in [k^* + 1, k]$  und somit  $\langle Y, u_l \rangle = 0$   $\mathbb{P}$ -f.s. (da  $\text{Var}(\langle Y, u_l \rangle) = \lambda_l$ ), so dass in diesem Fall  $Z_{(k)} = \sum_{i \in [k]} \langle Y, u_i \rangle u_i = \sum_{i \in [k^*]} \langle Y, u_i \rangle u_i = Z_{(k^*)}$   $\mathbb{P}$ -f.s. gilt. Sei also  $k \in [\text{rg}(\text{Cov}(Y))]$ . Betrachten wir den  $\mathbb{R}^k$ -wertigen Zufallsvektor  $X_{(k)} = U_{(k)}^t Y$  der ersten  $k$  Hauptkomponenten von  $Y$ , so gilt  $U_{(k)}^t \text{Cov}(Y) U_{(k)} = \text{Diag}(\lambda_{(k)})$  mit  $\lambda_{(k)} := (\lambda_i)_{i \in [k]}$ , also  $(U_{(k)}^t \text{Cov}(Y) U_{(k)})^+ = \text{Diag}(\lambda_{(k)}^{-1})$  und

$$\begin{aligned} Z_{(k)}^* &= \text{Cov}(Y)U_{(k)}^t(U_{(k)}\text{Cov}(Y)U_{(k)}^t)^+X_{(k)} = \text{Cov}(Y)U_{(k)}^t \text{Diag}(\lambda_{(k)}^{-1})U_{(k)}^t Y \\ &= U_{(k)}U_{(k)}^t Y = Z_{(k)} \end{aligned}$$

ist die beste lineare Vorhersage von  $Y$  durch  $X_{(k)}$ .  $\square$

§25.20 **Satz.** Sei  $Y$  ein zentrierter  $\mathbb{R}^n$ -wertiger Zufallsvektor in  $\mathcal{L}_2$  mit Hauptachsen  $(u_i)_{i \in [n]}$  von  $\Sigma = \text{Cov}(Y)$ . Für  $k \in [\text{rg}(\text{Cov}(Y))]$  sei  $U_{(k)} \in \mathbb{R}^{(n,k)}$  die Matrix mit den Spalten  $(u_i)_{i \in [k]}$  und sei  $X_{(k)} = U_{(k)}^t Y$  der  $\mathbb{R}^k$ -wertige Zufallsvektor der ersten  $k$  Hauptkomponenten  $(\langle Y, u_i \rangle)_{i \in [k]}$  von  $Y$ . Bezeichne mit  $Z_{(k)}^* := U_{(k)}U_{(k)}^t Y = \sum_{i \in [k]} \langle Y, u_i \rangle u_i$  die beste lineare Vorhersage von  $Y$  durch  $X_{(k)}$ . Für jedes  $A \in \mathbb{R}^{k,n}$  bezeichne  $Z_A^*$  die beste lineare Vorhersage von  $Y$  durch  $AY$ , dann gilt  $\mathbb{E}\|Y - Z_{(k)}^*\|^2 = \min_{A \in \mathbb{R}^{k,n}} \mathbb{E}\|Y - Z_A^*\|^2$ . Damit ist  $Z_{(k)}^*$  die beste lineare Vorhersage von  $Y$  unter allen linearen Vorhersagen von  $Y$  durch  $AY$  mit  $A \in \mathbb{R}^{(k,n)}$ .

§25.21 **Beweis von Satz §25.20.** (i) Nach Korollar §25.17 ist  $Z_A^* = \Sigma A^t(A\Sigma A^t)^+AY$  die beste lineare Vorhersage von  $Y$  durch  $AY$  für  $A \in \mathbb{R}^{(k,n)}$  mit mittlerem quadratischen Vorhersagefehler

$$\mathbb{E}\|Y - Z_A^*\|^2 = \text{Spur}(\Sigma - \Sigma A^t(A\Sigma A^t)^+A\Sigma)$$

Wir bestimmen für  $\min_{A \in \mathbb{R}^{k,n}} \mathbb{E}\|Y - Z_A^*\|^2 = \text{Spur}(\Sigma) - \max_A \text{Spur}(\Sigma A^t(A\Sigma A^t)^+A\Sigma)$  eine untere Schranke.



- (ii) Für  $B = \Sigma^{1/2} A^t$  ist  $B(B^t B)^+ B^t = \Pi_{\text{Bild}(B)}$  die Darstellungsmatrix der orthogonalen Projektion auf  $\mathcal{V} := \text{Bild}(B)$  und es gilt

$$\begin{aligned} \text{Spur}(\Sigma A^t (A \Sigma A^t)^+ A \Sigma) &= \text{Spur}(\Sigma^{1/2} B (B^t B)^+ B^t \Sigma^{1/2}) \\ &= \text{Spur}(\Sigma B (B^t B)^+ B^t) = \text{Spur}(\Pi_{\text{Bild}(B)} \Sigma \Pi_{\text{Bild}(B)}). \end{aligned}$$

- (iii) Da  $\text{rg}(B) \leq k$  ist  $\text{Bild}(B) \in \mathcal{V}_k := \{\mathcal{V} \mid \mathcal{V} \text{ Unterraum von } \mathbb{R}^n \text{ mit } \dim(\mathcal{V}) \leq k\}$ , so dass

$$\begin{aligned} \min_{A \in \mathbb{R}^{k,n}} \mathbb{E} \|Y - Z_A^*\|^2 &= \text{Spur}(\Sigma) - \max_{A \in \mathbb{R}^{k,n}} \text{Spur}(\Pi_{\text{Bild}(B)} \Sigma \Pi_{\text{Bild}(B)}) \\ &\geq \text{Spur}(\Sigma) - \max_{\mathcal{V} \in \mathcal{V}_k} \text{Spur}(\Pi_{\mathcal{V}} \Sigma \Pi_{\mathcal{V}}) \end{aligned}$$

- (iv) Für beliebige ONB  $(v_i)_{i=1}^l$  von  $\mathcal{V} \in \mathcal{V}_K$  gilt  $\text{Spur}(\Pi_{\mathcal{V}} \Sigma \Pi_{\mathcal{V}}) = \sum_{i \in [l]} \langle \Sigma v_i, v_i \rangle$ , somit entspricht dem Maximieren über alle Unterraum mit Dimension höchstens  $k$  gerade dem Maximieren über alle Orthonormalsystemen mit maximal  $k$  Elementen. Wir können nun per Induktion über  $k$  zeigen, dass  $\max_{\mathcal{V}} \text{Spur}(\Pi_{\mathcal{V}} \Sigma \Pi_{\mathcal{V}}) = \sum_{i=1}^k \langle \Sigma u_i, u_i \rangle = \sum_{i \in [k]} \lambda_i$ .

- (v) Insgesamt also

$$\min_{A \in \mathbb{R}^{k,n}} \mathbb{E} \|Y - Z_A^*\|^2 \geq \text{Spur}(\Sigma) - \sum_{i \in [k]} \lambda_i = \sum_{i \in [k+1,n]} \lambda_i$$

- (vi) Die Behauptung folgt dann aus  $\mathbb{E} \|Y - Z_{(k)}^*\|^2 = \sum_{i \in [k+1,n]} \mathbb{E} (\langle Y, u_i \rangle^2) = \sum_{i \in [k+1,n]} \lambda_i$ .  $\square$

§25.22 **Bemerkung.** Das letzte Resultat ist die Grundlage für eine Vielzahl an Methoden zur Analyse von sehr großen Datensätzen. Betrachten wir die beste lineare Vorhersage  $Z_{(k)}^*$  von  $Y$  durch die ersten  $k$  Hauptkomponenten  $(\langle Y, u_i \rangle)_{i \in [k]}$  von  $Y$ , so ist der mittlere Vorhersagefehler gerade  $\sum_{i \in [k+1,n]} \text{Var}(\langle Y, u_i \rangle)$ . Wählen wir also die ersten  $k$  Hauptkomponenten so können wir nur  $\sum_{i \in [k+1,n]} \text{Var}(\langle Y, u_i \rangle)$  der gesamten Variabilität  $\sum_{i \in [n]} \text{Var}(\langle Y, u_i \rangle)$  nicht erklären. Der erklärte Anteil wird typischerweise relativ angegeben, also  $\frac{\sum_{i \in [k]} \text{Var}(\langle Y, u_i \rangle)}{\sum_{i \in [n]} \text{Var}(\langle Y, u_i \rangle)} \in [0, 1]$ . In vielen Anwendungen genügen nur wenige (zwei oder drei) Hauptkomponenten, um relativ 90% oder mehr der gesamten Variabilität zu erklären.  $\square$

## §26 Statistische Inferenz: endliche Stichproben Eigenschaften

Seien  $\mathbb{P}_\theta$  eine Familie von Wahrscheinlichkeitsmaßen auf einem messbaren Raum  $(\mathcal{X}, \mathcal{F})$  und  $\gamma : \Theta \rightarrow \Gamma \subseteq \mathbb{R}$  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  $(\mathcal{X}^n, \mathcal{F}^n, \mathbb{P}_\theta^n)$  beschrieben. Wir schreiben kurz  $X \sim \mathbb{P}_\theta^n$ . Für jedes  $\theta \in \Theta$  bezeichnet im Folgenden  $\mathbb{E}_\theta^n$  stets die Erwartung bzgl.  $\mathbb{P}_\theta^n$ .

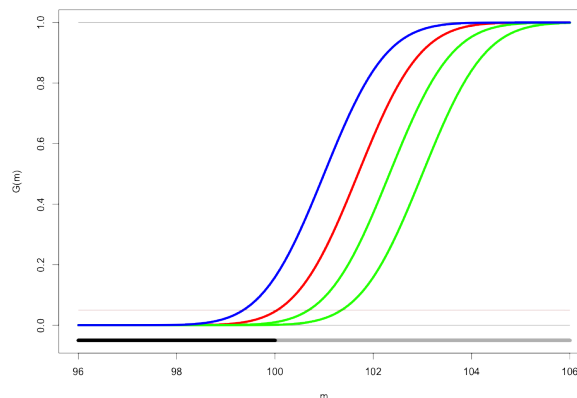
### §26.01 Gleichmäßig bester unverfälschter Test

§26.01 **Definition.** Sei  $\{\mathcal{H}^0, \mathcal{H}^1\}$  eine Partition der interessierenden Parameter und  $\varphi : \mathcal{X}^n \rightarrow \{0, 1\}$  ein Test, also  $\varphi \in \mathcal{F}^n$ , der Nullhypothese  $H_0 : \mathcal{H}^0$  gegen die Alternativhypothese  $H_1 : \mathcal{H}^1$ .

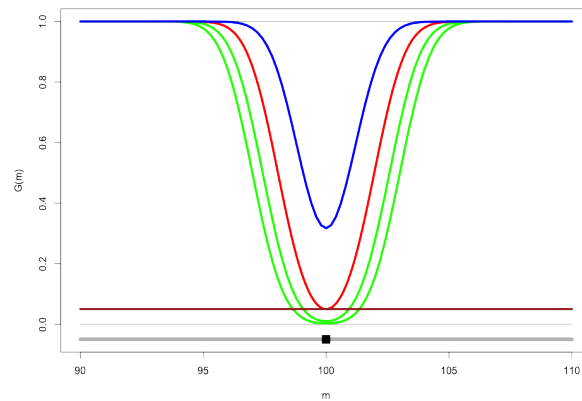
Die Abbildung  $\beta_\varphi : \Theta \rightarrow [0, 1]$  mit  $\theta \mapsto \beta_\varphi(\theta) := \mathbb{E}_\theta^n(\varphi)$  heißt *Gütefunktion* und ihre Werte  $\beta_\varphi(\theta)$  werden unter der Alternativhypothese, also  $\theta \in \Theta$  mit  $\gamma(\theta) \in \mathcal{H}^1$ , kurz  $\theta \in \gamma^{-1}(\mathcal{H}^1)$ , *Macht* von  $\varphi$  genannt. Der Test  $\varphi$  heißt *unverfälscht* zum Niveau  $\alpha \in [0, 1]$ , kurz *unverfälschter  $\alpha$ -Test*, wenn  $\varphi$  das Niveau  $\alpha$  einhält und seine Macht nicht kleiner als  $\alpha$  ist, also  $\beta_\varphi(\theta) \geq \alpha$  für alle  $\theta \in \gamma^{-1}(\mathcal{H}^1)$ . Ein Test  $\varphi$  heißt *gleichmäßig bester unverfälschter Test zum Niveau  $\alpha \in [0, 1]$* , falls er ein unverfälschter  $\alpha$ -Test ist und für jedes  $\theta \in \gamma^{-1}(\mathcal{H}^1)$  die Macht  $\beta_\varphi(\theta)$  eines jeden anderen unverfälschten  $\alpha$ -Tests  $\tilde{\varphi}$  nicht größer ist, das heißt,  $\beta_\varphi(\theta) \geq \beta_{\tilde{\varphi}}(\theta)$  gilt.  $\square$

§26.02 **Bemerkung.** Ein Test  $\varphi$  der Nullhypothese  $H_0 : \mathcal{H}^0$  gegen die Alternativhypothese  $H_1 : \mathcal{H}^1$  hält somit das Signifikanzniveau  $\alpha \in [0, 1]$  ein, falls für die assoziierte Gütefunktion  $\beta_\varphi(\theta) \leq \alpha$  für alle  $\theta \in \gamma^{-1}(\mathcal{H}^0)$  gilt.  $\square$

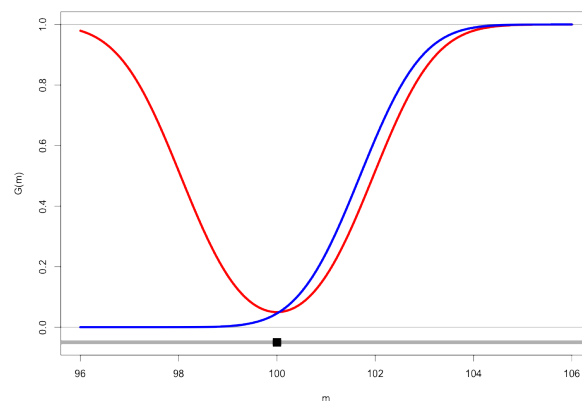
§26.03 **Beispiel (Normalverteilungsmodell §§12.12).** Im stetigen statistischen Modell  $X \odot (N_{(\mu,1)}^n)_{\mu \in \mathbb{R}}$  betrachte für beliebiges  $\mu_o \in \mathbb{R}$  das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu \leq \mu_o$ , also  $\mathcal{H}_{\mu_o}^0 = (-\infty, \mu_o]$ , gegen die Alternativhypothese  $H_1 : \mu > \mu_o$ , also  $\mathcal{H}_{\mu_o}^1 = (\mu_o, \infty)$ . Zur Erinnerung, da die Verteilungsfamilie  $(N_{(\mu,1)}^n)_{\mu \in \mathbb{R}}$  einen monotonen Likelihood-Quotienten besitzt (Beispiel §12.12), ist für jedes  $c \in \mathbb{R}$  der Neyman-Pearson-Test  $\varphi_c$  mit Ablehnbereich  $\{\varphi_c = 1\} = \{\bar{X}_n \geq c\}$  ein gleichmäßig bester Test zum Niveau  $\beta_{\varphi_c}(\mu_o) = N_{(\mu_o,1)}^n(\bar{X}_n \geq c)$  für das Testproblem der *einfachen Nullhypothese*  $H_0 : \mu = \mu_o$  gegen die *zusammengesetzte Alternativhypothese*  $H_1 : \mu > \mu_o$ . Da  $\bar{X}_n \sim N_{(\mu_o,1/n)}$  für  $X \sim N_{(\mu_o,1)}^n$  (vgl. Beispiel §17.13) erhalten wir  $\beta_{\varphi_c}(\mu_o) = \Phi(\sqrt{n}(-c + \mu_o))$  (vgl. Beispiel §10.12 (a)). Für  $\alpha \in (0, 1)$  und kritischen Wert  $c_\alpha = \mu_o + z_{1-\alpha}/\sqrt{n}$  mit  $\alpha$ -Quantil  $z_\alpha$  einer Standardnormalverteilung gilt dann  $\beta_{\varphi_{c_\alpha}}(\mu_o) = \alpha$  und  $\beta_{\varphi_{c_\alpha}}(\mu) \leq \alpha$  für alle  $\mu < \mu_o$  (vgl. §§05.07 (c)). Somit ist für jedes  $\alpha \in (0, 1)$  der Neyman-Pearson-Test mit Ablehnbereich  $\{\sqrt{n}(\bar{X}_n - \mu_o) \geq z_{1-\alpha}\}$  ein *gleichmäßig bester Test* zum Niveau  $\alpha$  des *einseitigen Testproblems* der Nullhypothese  $H_0 : \mu \leq \mu_o$  gegen die Alternativhypothese  $H_1 : \mu > \mu_o$ . In der folgenden Graphik sind für  $n = 1$ ,  $\mu_o = 100$  und  $\alpha = 0.05$  die Gütefunktionen des Test  $\varphi_c$  für die verschiedenen 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  $H_0 : \mu = \mu_o$  gegen die Alternativhypothese  $H_1 : \mu \neq \mu_o$  ist jeder Test  $\varphi_c$  mit Ablehnbereich  $\{\varphi_c = 1\} = \{\sqrt{n}|\bar{X}_n - \mu_o| \geq c\}$  mit  $c \geq z_{1-\alpha/2}$  ein  $\alpha$ -Test. In der folgenden Graphik sind für  $n = 1$ ,  $\mu_o = 100$  und  $\alpha = 0.05$  die Gütefunktionen des Test  $\varphi_c$  für die verschiedenen 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  $H_0 : \mu = \mu_o$  gegen die Alternativhypothese  $H_1 : \mu \neq \mu_o$  ist der zweiseitige Test  $\varphi = \mathbb{1}_{\{\sqrt{n}|X-\mu_o| \geq z_{1-\alpha/2}\}}$  kein gleichmäßig bester  $\alpha$  Test, da der Test  $\tilde{\varphi} = \mathbb{1}_{\{\sqrt{n}(X-\mu_o) \geq z_{1-\alpha}\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu \leq \mu_o$  gegen die Alternativhypothese  $H_1 : \mu > \mu_o$  auch ein  $\alpha$ -Test für das zweiseitige Testproblem ist und für alle  $\mu > \mu_o$  eine größere Macht besitzt. In der folgenden Graphik sind für  $n = 1$ ,  $\mu_o = 100$  und  $\alpha = 0.05$  die Gütefunktionen des zweiseitigen Tests  $\varphi$  (rot) und des einseitigen Tests  $\tilde{\varphi}$  (blau) dargestellt.



Offensichtlich ist die Macht des einseitigen Tests  $\tilde{\varphi}$  für alle Parameterwerte  $\mu < \mu_o$  kleiner als  $\alpha$ , so dass  $\tilde{\varphi}$  kein unverfälschter Test für das zweiseitige Testproblem ist. In der Vorlesung Statistik I zeigen wir, dass der zweiseitige Test  $\varphi$  optimal in der Klasse aller unverfälschten  $\alpha$ -Test für das zweiseitige Testproblem der Nullhypothese  $H_0 : \mu = \mu_o$  gegen die Alternativhypothese  $H_1 : \mu \neq \mu_o$  ist.  $\square$

## §26.02 Gleichmäßig bester unverfälschter Konfidenzbereich

§26.04 **Definition.** Seien  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  eine Familie von Partitionen in richtige und falsche interessierende Parameterwerte und  $B$  eine Bereichsschätzfunktion auf  $(\mathcal{X}^n, \mathcal{F}^n)$ , also  $\{\gamma \in B\} \in \mathcal{F}^n$  für alle  $\gamma \in \Gamma$ . Die Bereichsschätzfunktion  $B$  heißt *unverfälscht* zum Niveau  $1-\alpha$ , kurz *unverfälschter  $1-\alpha$ -Konfidenzbereich*, wenn  $B$  das Niveau  $1-\alpha$  einhält und  $\mathbb{P}_\theta^n(\tilde{\gamma} \in B) \leq 1 - \alpha$  für alle  $\tilde{\gamma} \in \mathcal{F}_{\gamma(\theta)}$  und für alle  $\theta \in \Theta$  gilt.  $\square$

§26.05 **Erinnerung.**

- (i) Für  $\gamma \in \Gamma$  sei  $\mathcal{R}_\gamma$  und  $\mathcal{F}_\gamma$  eine Partition in richtige und falsche interessierende Parameter-

werte, dann bezeichnen  $\mathcal{H}_\gamma^0 = \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{R}_{\tilde{\gamma}}\}$  und  $\mathcal{H}_\gamma^1 = \{\tilde{\gamma} \in \Gamma \mid \gamma \in \mathcal{F}_{\tilde{\gamma}}\}$  die *assoziierte Null- bzw. Alternativhypothese* der interessierende Parameter.

- (ii) Zu einer Bereichsschätzfunktion  $B$  bezeichnet  $(\varphi_\gamma)_{\gamma \in \Gamma}$  die *assoziierte Familie von Tests* mit Ablehnbereich  $\{\varphi_\gamma = 1\} = \{\gamma \notin B\}$ .
- (iii) Zu einer  $(\varphi_\gamma)_{\gamma \in \Gamma}$  Familie von Tests bezeichnet  $B$  die *assoziierte Bereichsschätzfunktion* mit  $B(x) := \{\gamma \in \Gamma \mid \varphi_\gamma(x) = 0\}$ .  $\square$

§26.06 **Lemma.** Seien  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  eine Familie von richtigen und falschen interessierenden Parameterwerte und  $(\{\mathcal{H}_\gamma^0, \mathcal{H}_\gamma^1\})_{\gamma \in \Gamma}$  die assoziierte Familie von Null- und Alternativhypothesen. Dann ist  $\varphi_\gamma$  ein *(gleichmäßig bester) unverfälschter  $\alpha$ -Test* der Nullhypothese  $H_0 : \mathcal{H}_\gamma^0$  gegen die Alternativhypothese  $H_1 : \mathcal{H}_\gamma^1$  für jedes  $\gamma \in \Gamma$ , genau dann wenn die zu  $(\varphi_\gamma)_{\gamma \in \Gamma}$  assoziierte Bereichsschätzfunktion  $B$  für  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  ein *(gleichmäßig bester) unverfälschter  $1-\alpha$ -Konfidenzbereich* ist.

§26.07 **Beweis** von Lemma §26.06. In der Vorlesung.  $\square$

§26.08 **Beispiel** (Normalverteilungsmodell §§26.03). Im stetigen statistischen Modell  $X \odot (N_{(\mu,1)}^n)_{\mu \in \mathbb{R}}$  sind für beliebiges  $\mu_o \in \mathbb{R}$  und  $\alpha$ -Quantil  $z_\alpha$  der Standardnormalverteilung der rechtsseitige Test  $\varphi_{z_{1-\alpha}}^r = \mathbb{1}_{\{\sqrt{n}(\bar{X}_n - \mu_o) \geq z_{1-\alpha}\}}$ , der linksseitige Test  $\varphi_{z_{1-\alpha}}^l = \mathbb{1}_{\{\sqrt{n}(\bar{X}_n - \mu_o) \leq -z_{1-\alpha}\}}$  und der beidseitige Test  $\varphi_{z_{1-\alpha/2}}^b = \mathbb{1}_{\{\sqrt{n}|\bar{X}_n - \mu_o| \geq z_{1-\alpha/2}\}} = \varphi_{z_{1-\alpha/2}}^r + \varphi_{z_{1-\alpha/2}}^l$  für das entsprechende Testproblem gleichmäßig beste unverfälschte  $\alpha$ -Tests. Damit sind für die Familie  $(\{\mathcal{R}_\gamma, \mathcal{F}_\gamma\})_{\gamma \in \Gamma}$  der richtigen und falschen Parameter

- (a) mit  $\mathcal{R}_\mu = [\mu, \infty)$  und  $\mathcal{F}_\mu = (-\infty, \mu)$  der rechtsseitige KB  $(\bar{X}_n - z_{1-\alpha}/\sqrt{n}, \infty)$ ;
- (b) mit  $\mathcal{R}_\mu = (-\infty, \mu]$  und  $\mathcal{F}_\mu = (\mu, \infty)$  der linksseitige KB  $(-\infty, \bar{X}_n + z_{1-\alpha}/\sqrt{n})$ ;
- (c) mit  $\mathcal{R}_\mu = \{\mu\}$  und  $\mathcal{F}_\mu = \mathbb{R}_\mu$  der beidseitige KB  $(\bar{X}_n \pm z_{1-\alpha/2}/\sqrt{n})$

gleichmäßig beste unverfälschte  $1-\alpha$ -Konfidenzbereiche.  $\square$

### §26.03 Beste erwartungstreue Schätzfunktion

§26.09 **Definition.** Sei  $\hat{\gamma}$  eine Schätzfunktion, also  $\hat{\gamma} \in \mathcal{F}^n$ , für den Parameter  $\gamma$ . Ist  $\hat{\gamma} \in \mathcal{L}_1(\mathbb{P}^n)$  für jedes  $\theta \in \Theta$ , so wird  $\text{Bias}_\theta(\hat{\gamma}) := \mathbb{E}_\theta(\hat{\gamma} - \gamma(\theta))$  *Verzerrung* oder *Bias* von  $\hat{\gamma}$  genannt. Der Schätzer heißt *erwartungstreu* oder *unverfälscht*, wenn  $\text{Bias}_\theta(\hat{\gamma}) = 0$  für alle  $\theta \in \Theta$  gilt.  $\hat{\gamma}$  *überschätzt* (bzw. *unterschätzt*) im Mittel  $\gamma$ , wenn  $\text{Bias}_\theta(\hat{\gamma}) > 0$  (bzw.  $\text{Bias}_\theta(\hat{\gamma}) < 0$ ) für alle  $\theta \in \Theta$  gilt.

§26.10 **Definition.** Für  $k \in \mathbb{N}$  sei  $\mathcal{W}_k$  die Menge aller Wahrscheinlichkeitsmaße auf  $(\mathbb{R}, \mathcal{B})$  mit endlichem  $k$ -ten absolutem Moment, also  $\mathbb{E}_\mathbb{P}(|X|^k) = \mathbb{P}(|X|^k) < \infty$  für alle  $\mathbb{P} \in \mathcal{W}_k$ , und für  $n \in \mathbb{N}$  sei  $\mathcal{W}_k^n := \{\mathbb{P}^n = \bigotimes_{i \in [n]} \mathbb{P} \mid \mathbb{P} \in \mathcal{W}_k\}$  die Menge aller Produktmaße  $\mathbb{P}^n$  auf  $(\mathbb{R}^n, \mathcal{B}^n)$  mit identischen Randverteilungen aus  $\mathcal{W}_k$ . Für  $l \in [k]$  heißen die abgeleiteten Parameter  $m^{(l)} : \mathcal{W}_k \rightarrow \mathbb{R}$  mit  $\mathbb{P} \mapsto m^{(l)}(\mathbb{P}) := \mathbb{P}(X^l) = \mathbb{E}_\mathbb{P}(X^l)$  und  $M^{(l)} : \mathcal{W}_k \rightarrow \mathbb{R}$  mit  $\mathbb{P} \mapsto M^{(l)}(\mathbb{P}) := \mathbb{E}_\mathbb{P}((X - m^{(1)}(\mathbb{P}))^l)$  das  *$l$ -te Moment* bzw. das  *$l$ -te zentrierte Moment* von  $\mathbb{P}$ .  $\square$

§26.11 **Bemerkung.** Die Momente sind identifizierbare abgeleitete Parameter und damit zum Beispiel auch  $(m^{(l)})_{l \in [k]} : \mathcal{W}_k \rightarrow \mathbb{R}^k$ , aber die Abbildung  $(m^{(l)})_{l \in [k]}$  ist nicht injektiv. Nehmen wir zum Beispiel zusätzlich an, dass  $X$  normalverteilt ist, also  $X \odot (N_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+}$ , so ist  $m^{(1)} = \mu$

und  $M^{(2)} = \sigma^2$ . Die Parametrisierung  $(m^{(1)}, M^{(2)})$  ist somit bijektiv, aber  $m^{(1)}$  oder  $M^{(2)}$  allein sind nicht bijektiv.  $\square$

§26.12 **Definition.** Ist  $(\mathbb{R}^n, \mathcal{B}^n, \mathcal{W}_k^n)$  ein adäquates statistisches Experiment für  $X = (X_i)_{i \in [n]}$ , so werden für  $l \in [1, k]$  die reellen Statistiken  $\widehat{m}_n^{(l)}$  und  $\widehat{M}_n^{(l)}$  auf  $(\mathbb{R}^n, \mathcal{B}^n)$  mit  $\widehat{m}_n^{(l)} := \widehat{m}_n^{(l)}(X) = \frac{1}{n} \sum_{i \in [n]} X_i^l$  und  $\widehat{M}_n^{(l)} := \widehat{M}_n^{(l)}(X) = \frac{1}{n} \sum_{i \in [n]} (X_i - m^{(1)})^l$  *empirisches Moment* bzw. *zentriertes empirisches Moment* der Ordnung  $l$  genannt.  $\square$

§26.13 **Bemerkung.** Die Statistik  $\widehat{M}_n^{(l)}$  verwendet den wahren Wert  $m^{(1)}$  und ist somit nur unter der unrealistischen Annahme, dass  $m^{(1)}$  bekannt ist, ein Schätzer für  $M^{(l)}$ . Auf Grund der Linearität der Erwartung sind  $\widehat{m}_n^{(l)}$  und  $\widehat{M}_n^{(l)}$  *erwartungstreue Schätzer* für  $m^{(l)}$  bzw.  $M^{(l)}$ . In der Tat, für  $X \sim \mathbb{P}^n$  mit  $m^{(l)}(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(X_1^l)$  sowie  $M^{(l)}(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}((X_1 - m^{(1)}(\mathbb{P}))^l)$  gilt  $\text{Bias}_{\mathbb{P}}(\widehat{m}_n^{(l)}) = \mathbb{E}_{\mathbb{P}}(\widehat{m}_n^{(l)} - m^{(l)}(\mathbb{P})) = \frac{1}{n} \sum_{i \in [n]} \{\mathbb{E}_{\mathbb{P}}(X_i^l) - m^{(l)}(\mathbb{P})\} = 0$  sowie  $\text{Bias}_{\mathbb{P}}(\widehat{M}_n^{(l)}) = \mathbb{E}_{\mathbb{P}}(\widehat{M}_n^{(l)} - M^{(l)}(\mathbb{P})) = \frac{1}{n} \sum_{i \in [n]} \{\mathbb{E}_{\mathbb{P}}(X_i - m^{(1)}(\mathbb{P}))^l - M^{(l)}(\mathbb{P})\} = 0$ . Ist  $m^{(1)}$  nicht bekannt, und somit  $\widehat{M}_n^{(l)}$  als Schätzfunktion nicht mehr verwendbar, so betrachtet man üblicherweise die Statistik  $\widehat{V}_n^{(l)} := \frac{1}{n} \sum_{i \in [n]} (X_i - \widehat{m}_n^{(1)})^l$ . Die Statistik  $\widehat{V}_n^{(l)}$  erhalten wir durch das Ersetzen des Erwartungswertes  $m^{(1)}(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(X) =: \mu$  durch das empirischen Moment  $\widehat{m}_n^{(1)} = \bar{X}_n$ , diese Vorgehensweise wird häufig *Momentenmethode* genannt. Im Spezialfall  $l = 2$  ist  $M^{(2)}(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(X - \mu)^2 =: \sigma^2$  gerade die Varianz von  $X$ . Mit Hilfe elementarer Umformungen gilt  $\widehat{V}_n^{(2)} = \widehat{M}_n^{(2)} - (\widehat{m}_n^{(1)} - m^{(1)})^2 = \frac{1}{n} \sum_{i \in [n]} (X_i - \mu)^2 - (\bar{X}_n - \mu)^2$ . Da  $X = (X_i)_{i \in [1, n]}$  unabhängig und identisch verteilt sind, sind die Zufallsvariablen  $Y_i := \frac{1}{n}(X_i - \mu)$ ,  $i \in [n]$ , zentriert und unabhängig, sodass aus **Lemma** §24.03 (iii) folgt  $\mathbb{E}_{\mathbb{P}}((\bar{X}_n - \mu)^2) = \mathbb{V}_{\mathbb{P}}(\sum_{i=1}^n Y_i) = \sum_{i \in [n]} \mathbb{V}_{\mathbb{P}}(Y_i) = \frac{1}{n^2} \sum_{i \in [n]} \mathbb{V}_{\mathbb{P}}(X_i) = \frac{\sigma^2}{n}$ . Damit erhalten wir  $\mathbb{E}_{\mathbb{P}}(\widehat{V}_n^{(2)}) = \mathbb{E}_{\mathbb{P}}(\widehat{M}_n^{(2)}) - \mathbb{E}_{\mathbb{P}}((\bar{X}_n - \mu)^2) = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2$ , also  $\text{Bias}_{\mathbb{P}}(\widehat{V}_n^{(2)} - \sigma^2) = -\frac{1}{n}\sigma^2$ , sodass der Schätzer  $\widehat{V}_n^{(2)}$  im Mittel die Varianz  $\sigma^2$  unterschätzt. Folglich ist  $\widehat{S}_n^{(2)} := \frac{n}{n-1} \widehat{V}_n^{(2)} = \frac{1}{n-1} \sum_{i \in [n]} (X_i - \bar{X}_n)^2$  ein erwartungstreuer Schätzer für  $\sigma^2$ .  $\square$

§26.14 **Definition.** Für  $X \odot \mathcal{W}_2^n$  heißen  $\widehat{m}_n^{(1)} = \bar{X}_n$  und  $\widehat{S}_n^{(2)} := \frac{1}{n-1} \sum_{i \in [n]} (X_i - \bar{X}_n)^2$  *empirischer Mittelwert* bzw. *empirische Varianz*. Wir schreiben abkürzend  $\widehat{S}_n := \sqrt{\widehat{S}_n^{(2)}}$ .  $\square$

§26.15 **Definition.** Seien  $\widehat{\gamma}, \widetilde{\gamma} \in \mathcal{L}_2(\mathbb{P}^n)$  für jedes  $\theta \in \Theta$  zwei erwartungstreue Schätzfunktionen für  $\gamma$ .  $\widehat{\gamma}$  heißt *(relativ) effizienter* als  $\widetilde{\gamma}$ , wenn  $\mathbb{V}_{\mathbb{P}}(\widehat{\gamma}) \leq \mathbb{V}_{\mathbb{P}}(\widetilde{\gamma})$  für alle  $\theta \in \Theta$  gilt und ein  $\theta \in \Theta$  mit  $\mathbb{V}_{\mathbb{P}}(\widehat{\gamma}) < \mathbb{V}_{\mathbb{P}}(\widetilde{\gamma})$  existiert.  $\widehat{\gamma}$  wird *(absolut) effizient* genannt, falls es keinen erwartungstreuen Schätzer gibt, der (relativ) effizienter als  $\widehat{\gamma}$  ist.  $\square$

§26.16 **Beispiel** (§§19.03 (d) fortgesetzt). Sei  $X \odot (\mathcal{U}_{[0, \theta]}^n)_{\theta \in \mathbb{R}_0^+}$ , so gilt  $\mathbb{E}_{\theta}(X_1) = \theta/2$  und  $\mathbb{V}_{\theta}(X_1) = \theta^2/12$  (nachrechnen!). Wenden wir die Momentenmethode an, so erhalten wir den Schätzer  $\widehat{\theta}_1 = 2\bar{X}_n$  für  $\theta$ . Andererseits haben wir in **Beispiel** §12.23 (c) gezeigt, dass  $\widehat{\theta}_2 = \max_{i \in [n]} X_i$  der MLS für  $\theta$  ist. Offensichtlich ist  $\widehat{\theta}_1$  ein erwartungstreuer Schätzer für  $\theta$ . Andererseits, nach **Beispiel** §19.03 (d) ist  $\widehat{\theta}_2$  eine stetig-verteilte Zufallsvariable mit Dichte  $f_{\theta}(x) = \frac{nx^{n-1}}{\theta^n} \mathbb{1}_{[0, \theta]}(x)$ ,  $x \in \mathbb{R}$ , sodass  $\mathbb{E}_{\theta}(\widehat{\theta}_2) = \frac{n}{n+1}\theta$  und  $\text{Bias}_{\theta}(\widehat{\theta}_2) = -\theta/(n+1)$  gilt. Damit unterschätzt der MLS  $\widehat{\theta}_2$  im Mittel  $\theta$ . Der korrigierte MLS  $\widehat{\theta}_3 := \frac{n+1}{n} \max_{i \in [n]} X_i$  ist folglich erwartungstreu. Für  $n > 1$  ist der korrigierte MLS  $\widehat{\theta}_3$  effizienter als der Momentenschätzer  $\widehat{\theta}_1 = 2\bar{X}_n$ , da  $\mathbb{V}_{\theta}(\widehat{\theta}_3) =$



$\frac{1}{n(n+2)}\theta^2 < \frac{1}{3n}\theta^2 = \text{Var}_\theta(\hat{\theta}_1)$  für  $n > 1$  gilt (nachrechnen!). Für  $n = 1$  ist  $\hat{\theta}_1 = 2X_1 = \hat{\theta}_3$ . In der Vorlesung Statistik I wird gezeigt, dass  $\hat{\theta}_3$  (absolut) effizient ist.  $\square$

§26.17 **Definition.** Seien  $\hat{\gamma}, \tilde{\gamma}$  Schätzfunktionen für  $\gamma$ . Ist  $\hat{\gamma} \in \mathcal{L}_2(\mathbb{P})$  für jedes  $\theta \in \Theta$ , so wird  $\text{MSE}_\theta(\hat{\gamma}) := \mathbb{E}_\theta(|\hat{\gamma} - \gamma(\theta)|^2)$  *mittlerer quadratischer Fehler* (Mean Squared Error) von  $\hat{\gamma}$  genannt.  $\hat{\gamma}$  heißt *besser bzgl. des MSE* als  $\tilde{\gamma}$ , wenn  $\text{MSE}_\theta(\hat{\gamma}) \leq \text{MSE}_\theta(\tilde{\gamma})$  für alle  $\theta \in \Theta$  gilt und ein  $\theta \in \Theta$  mit  $\text{MSE}_\theta(\hat{\gamma}) < \text{MSE}_\theta(\tilde{\gamma})$  existiert.  $\square$

§26.18 **Beispiel.**

- (a) Für  $X \odot (\mathcal{U}_{[0,\theta]}^n)_{\theta \in \mathbb{R}_0^+}$  und  $n > 1$  ist bzgl. des MSE der korrigierter Maximum-Likelihood-Schätzer (MLS)  $\hat{\theta}_3 = \frac{n+1}{n} \max_{i \in [n]} X_i$  besser als der MLS  $\hat{\theta}_2 = \max_{i \in [n]} X_i$ . In der Tat es gilt  $\text{MSE}_\theta(\hat{\theta}_3) = \frac{\theta^2}{n(n+2)} < \frac{2\theta^2}{(n+2)(n+1)} = \text{MSE}_\theta(\hat{\theta}_2)$ . Weiterhin ist der MLS  $\hat{\theta}_2 = \max_{i \in [n]} X_i$  besser bzgl. des MSE als der Momentenschätzer  $\hat{\theta}_1 = 2\bar{X}_n$ , da  $\text{MSE}_\theta(\hat{\theta}_2) = \frac{2\theta^2}{(n+2)(n+1)} < \frac{\theta^2}{3n} = \text{MSE}_\theta(\hat{\theta}_1)$  gilt (nachrechnen!).
- (b) Für  $X \odot (\mathcal{N}_{(\mu,\sigma^2)}^n)_{(\mu,\sigma^2) \in \mathbb{R} \times \mathbb{R}_0^+}$  und  $n > 1$  ist die empirische Varianz  $\hat{S}_n^{(2)}$  ein (absolut) effizienter Schätzer für  $\sigma^2$  (Vorlesung Statistik I) aber bzgl. des MSE ist der Schätzer  $\hat{V}_n^{(2)}$  für  $\sigma^2$  besser als  $\hat{S}_n^{(2)}$ , da  $\text{MSE}_{\mu,\sigma}(\hat{V}_n^{(2)}) = \frac{2n-1}{n^2}\sigma^4 < \frac{2}{n-1}\sigma^4 = \text{MSE}_{\mu,\sigma}(\hat{S}_n^{(2)})$ .  $\square$

## §26.04 Lokations-Skalen-Modell

§26.19 **Definition.** Für einen  $\mathbb{R}^n$ -wertigen Zufallsvektor  $X$  ist ein *Lokations-Skalen-Modell* adäquat, wenn  $X \odot \mathcal{W}_2^n$  gilt, also die Koordinaten von  $X$  unabhängige und identisch verteilte (u.i.v.) reelle Zufallsvariablen mit identischer Randverteilung aus  $\mathcal{W}_2$  sind. Wir betrachten die identifizierbaren Parameter  $\mu : \mathcal{W}_2 \rightarrow \mathbb{R}$  mit  $\mathbb{P} \mapsto \mu(\mathbb{P}) := \mathbb{P}(X) = \mathbb{E}_\mathbb{P}(X)$  und  $\sigma^2 : \mathcal{W}_2 \rightarrow \mathbb{R}^+$  mit  $\mathbb{P} \mapsto \sigma^2(\mathbb{P}) := \text{Var}_\mathbb{P}(X)$ , und schreiben abkürzend  $X \odot (\mathcal{P}_{(\mu,\sigma^2)}^n)_{(\mu,\sigma) \in \mathbb{R} \times \mathbb{R}^+}$ . Wird die Varianz  $\sigma_o^2 \in \mathbb{R}^+$  bzw. der Erwartungswert  $\mu_o \in \mathbb{R}$  als bekannt vorausgesetzt, so schreiben wir abkürzend  $X \odot (\mathcal{P}_{(\mu,\sigma_o^2)}^n)_{\mu \in \mathbb{R}}$  bzw.  $X \odot (\mathcal{P}_{(\mu_o,\sigma^2)}^n)_{\sigma^2 \in \mathbb{R}^+}$ .  $\square$

§26.20 **Bemerkung.** Wir fassen unter  $\mathcal{P}_{(\mu,\sigma^2)}^n$  alle  $\mathbb{P}^n \in \mathcal{W}_2^n$  mit  $\mu = \mu(\mathbb{P})$  und  $\sigma = \sigma(\mathbb{P})$  zusammen und die abkürzende Schreibweise  $X \odot \mathcal{P}_{(\mu,\sigma^2)}^n$  ist so zu verstehen, dass es ein  $\mathbb{P} \in \mathcal{P}_{(\mu,\sigma^2)}^n$  gibt mit  $X \sim \mathbb{P}$ .  $\square$

§26.21 **Beispiel.**

- (a) *Bernoulli-Schema:*  $X \odot (\mathcal{B}_p^n)_{p \in [0,1]}$ , dann gilt  $\mathbb{E}_p(X_1) = p$  und  $\text{Var}_p(X_1) = p - p^2$ , also gilt auch  $X \odot (\mathcal{P}_{(p,p-p^2)}^n)_{p \in [0,1]}$ . Der MLS und Momentenschätzer  $\hat{p} = \bar{X}_n$  für  $p$  ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).
- (b) *Binomialverteilungsmodell:*  $X \odot (\text{Bin}_{(50,p)}^n)_{p \in [0,1]}$ , dann gilt  $\mathbb{E}_p(X_1) = 50p$  und  $\text{Var}_p(X_1) = 50p(1-p)$ , also gilt auch  $X \odot (\mathcal{P}_{(50p,50p(1-p))}^n)_{p \in [0,1]}$ . Der MLS und Momentenschätzer  $\hat{p} = \bar{X}_n/50$  für  $p$  ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).
- (c) *Poissonverteilungsmodell:*  $X \odot (\text{Poi}_\lambda^n)_{\lambda \in \mathbb{R}_0^+}$ , dann gilt  $\mathbb{E}_\lambda(X_1) = \lambda$  und  $\text{Var}_\lambda(X_1) = \lambda$ , also gilt auch  $X \odot (\mathcal{P}_{(\lambda,\lambda)}^n)_{\lambda \in \mathbb{R}_0^+}$ . Der MLS und Momentenschätzer  $\hat{\lambda} = \bar{X}_n$  für  $\lambda$  ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).

(d) **Exponentialverteilungsmodell:**  $X \odot (\text{Exp}_\lambda^n)_{\lambda \in \mathbb{R}_0^+}$ , dann gilt  $\mathbb{E}_\lambda(X_1) = 1/\lambda$  und  $\text{Var}_\lambda(X_1) = 1/\lambda^2$ , also  $X \odot (\mathcal{P}_{(1/\lambda, 1/\lambda^2)}^n)_{\lambda \in \mathbb{R}_0^+}$ . Der MLS und Momentenschätzer  $\hat{\lambda} = (\bar{X}_n)^{-1}$  überschätzt im Mittel  $\lambda$  mit  $\text{Bias}_\lambda(\hat{\lambda}) = \lambda/(n-1)$ , der korrigierter MLS  $\hat{\lambda}_2 = (\frac{n}{n-1}\bar{X}_n)^{-1}$  ist erwartungstreu und (absolut) effizient (Vorlesung Statistik I).  $\square$

§26.22 **Erinnerung.** In einem Lokations-Skalen-Modell  $X \odot (\mathcal{P}_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+}$  ist das empirische Mittel  $\bar{X}_n$  ein erwartungstreuer Schätzer für  $\mu$  und die empirische Varianz  $\hat{S}_n^{(2)}$  ein erwartungstreuer Schätzer für  $\sigma^2$ . Die folgenden Tests basieren nun auf den Statistiken  $\frac{\sqrt{n}}{\hat{S}_n}(\bar{X}_n - \mu)$  mit  $\hat{S}_n := \sqrt{\hat{S}_n^{(2)}}$  bzw.  $\frac{(n-1)}{\sigma^2} \hat{S}_n^{(2)}$  deren Verteilung wir unter einer zusätzlichen Normalverteilungsannahme in Abschnitt §26.05 herleiten.  $\square$

§26.23 **Testen des Erwartungswerts.** Für einen  $\mathbb{R}^n$ -wertigen Zufallsvektor  $X$  sei ein Lokations-Skalen-Modell, also  $X \odot (\mathcal{P}_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+}$  adäquat. Für  $\mu_o \in \mathbb{R}$  und  $c \in \mathbb{R}_0^+$  betrachten wir

- (a) den rechtsseitigen Test  $\varphi_c^r := \mathbb{1}_{\{\sqrt{n}(\bar{X}_n - \mu_o) \geq c\hat{S}_n\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu \leq \mu_o$  gegen die Alternativhypothese  $H_1 : \mu - \mu_o > 0$ ;
- (b) den linksseitigen Test  $\varphi_c^l := \mathbb{1}_{\{\sqrt{n}(\bar{X}_n - \mu_o) \leq -c\hat{S}_n\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu \geq \mu_o$  gegen die Alternativhypothese  $H_1 : \mu - \mu_o < 0$ ;
- (c) den beidseitigen Test  $\varphi_c^b := \mathbb{1}_{\{\sqrt{n}|\bar{X}_n - \mu_o| \geq c\hat{S}_n\}}$  für das *zweiseitige Testproblem* der Nullhypothese  $H_0 : \mu = \mu_o$  gegen die Alternativhypothese  $H_1 : \mu - \mu_o \neq 0$ .  $\square$

§26.24 **Testen der Varianz.** Für einen  $\mathbb{R}^n$ -wertigen Zufallsvektor  $X$  sei ein Lokations-Skalen-Modell, also  $X \odot (\mathcal{P}_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$  adäquat. Für  $\sigma_o^2 \in \mathbb{R}_0^+$  und  $c^o, c_u \in \mathbb{R}_0^+$  betrachten wir

- (a) den rechtsseitigen Test  $\varphi_{c_u}^r := \mathbb{1}_{\{(n-1)\hat{S}_n^{(2)}/\sigma_o^2 \geq c_u\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \sigma \leq \sigma_o$  gegen die Alternativhypothese  $H_1 : \sigma/\sigma_o > 1$ ;
- (b) den linksseitigen Test  $\varphi_{c_o}^l := \mathbb{1}_{\{(n-1)\hat{S}_n^{(2)}/\sigma_o^2 \leq c_o\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \sigma \geq \sigma_o$  gegen die Alternativhypothese  $H_1 : \sigma/\sigma_o < 1$ ;
- (c) den beidseitigen Test  $\varphi_{c^o, c_u}^b := \mathbb{1}_{\{(n-1)\hat{S}_n^{(2)} \leq c_o\sigma_o^2\}} + \mathbb{1}_{\{(n-1)\hat{S}_n^{(2)} \geq c_u\sigma_o^2\}}$  für das *zweiseitige Testproblem* der Nullhypothese  $H_0 : \sigma = \sigma_o$  gegen die Alternativhypothese  $H_1 : \sigma/\sigma_o \neq 1$ .  $\square$

§26.25 **Bemerkung.** Möchten wir sicher stellen, dass die in Definition §26.23 (bzw. Definition §26.24) angegebenen Tests ein vorgegebenes Signifikanz-Niveau  $\alpha \in (0, 1)$  einhalten, so müssen wir den kritischen Wert  $c \in \mathbb{R}_0^+$  (bzw.  $c^o, c_u \in \mathbb{R}_0^+$ ) entsprechend wählen. Dies ist aber nur möglich, wenn die Verteilung der standardisierten Statistik  $\frac{\sqrt{n}}{\hat{S}_n}(\bar{X}_n - \mu_o)$  (bzw.  $\frac{(n-1)}{\sigma_o^2} \hat{S}_n^{(2)}$ ) unter der Nullhypothese bekannt ist, wie es zum Beispiel im normalen Lokations-Skalen-Modell (Abschnitt §26.05) der Fall ist. Die angegebenen Tests können wir direkt benutzen (vgl. Satz §12.33) um einseitige bzw. zweiseitige Konfidenzbereiche für den Erwartungswert  $\mu$  (bzw.  $\sigma^2$ ) anzugeben. Zum Beispiel ist  $(\bar{X}_n \pm c\hat{S}_n/\sqrt{n})$  (bzw.  $((n-1)\hat{S}_n^{(2)}/c_u, (n-1)\hat{S}_n^{(2)}/c_o)$ ) der assoziierte zweiseitige Konfidenzbereich. In ?? untersuchen wir das asymptotische Verhalten der Verteilung von  $\bar{X}_n$  sowie  $\hat{S}_n^{(2)}$ , also wenn  $n \rightarrow \infty$ . Diese Resultate erlauben dann, den kritischen Wert  $c$  so zu wählen, dass zu mindesten asymptotisch, also für  $n \rightarrow \infty$ , dass das Signifikanzniveau eingehalten wird.  $\square$

## Zwei-Stichproben-Modell



§26.26 **Vorbemerkung.** Die zufälligen Messwerte der beiden Segmente des Fürsten G. Ömetry in Beispiel VI im Kapitel 1 Prolog können wir als Stichproben  $X \odot (\mathcal{P}_{(\mu_X, \sigma_X^2)}^{31})_{(\mu_X, \sigma_X) \in \mathbb{R} \times \mathbb{R}_0^+}$  und  $Y \odot (\mathcal{P}_{(\mu_Y, \sigma_Y^2)}^{31})_{(\mu_Y, \sigma_Y) \in \mathbb{R} \times \mathbb{R}_0^+}$  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 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  $(\mathbb{R}^{n+m}, \mathcal{B}^{n+m}, (\mathcal{P}_{(\mu_X, \sigma^2)}^n \otimes \mathcal{P}_{(\mu_Y, \sigma^2)}^m)_{(\mu_X, \mu_Y, \sigma) \in \mathbb{R}^2 \times \mathbb{R}_0^+})$  beschrieben. Die Situation verschiedener Varianzen wird Behrens-Fischer Problem genannt, welches noch immer ungelöst ist. Für die interessierenden Parameter  $\mu_X$  und  $\mu_Y$  sind dann  $\bar{X}_n$  bzw.  $\bar{Y}_m$  erwartungstreue Momentenschätzer (Beispiel §26.21). Betrachten wir die Varianz  $\sigma^2$ , so sind  $\hat{S}_X^{(2)} := \frac{1}{n-1} \sum_{i \in [n]} (X_i - \bar{X}_n)^2$  und  $\hat{S}_Y^{(2)} := \frac{1}{m-1} \sum_{i \in [m]} (Y_i - \bar{Y}_m)^2$  unabhängige erwartungstreue Momentenschätzer für  $\sigma^2$ , und somit ist der Schätzer  $\hat{S}_{n+m}^{(2)} := \frac{1}{n+m-2} [(n-1)\hat{S}_X^{(2)} + (m-1)\hat{S}_Y^{(2)}]$  ebenfalls erwartungstreu. Es ist nun nahe liegend basierend auf der Statistik  $\bar{X}_n - \bar{Y}_m$  für den interessierenden Parameter  $\mu_X - \mu_Y$  zum Beispiel einen zweiseitigen Konfidenzbereich der Form  $[\bar{X}_n - \bar{Y}_m \pm c \frac{\sqrt{n+m}}{\sqrt{nm}} \hat{S}_{n+m}]$  mit  $\hat{S}_{n+m} := \sqrt{\hat{S}_{n+m}^{(2)}}$  anzugeben und somit  $\{0 \neq [\bar{X}_n - \bar{Y}_m \pm c \frac{\sqrt{n+m}}{\sqrt{nm}} \hat{S}_{n+m}]\}$  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.  $\square$

§26.27 **Tests auf Gleichheit der Erwartungswerte.** Für zwei Zufallsvektoren  $X$  und  $Y$  mit derselben Varianz der Randverteilungen, sei ein Zwei-Stichproben-Modell mit unverbundenen Stichproben, also  $(X, Y) \odot (\mathcal{P}_{(\mu_X, \sigma^2)}^n \otimes \mathcal{P}_{(\mu_Y, \sigma^2)}^m)_{(\mu_X, \mu_Y, \sigma) \in \mathbb{R}^2 \times \mathbb{R}_0^+}$ , adäquat. Für  $c \in \mathbb{R}_0^+$  betrachten wir

- (a) den rechtsseitigen Test  $\varphi_c^r := \mathbb{1}_{\{\bar{X}_n - \bar{Y}_m \geq c \frac{\sqrt{n+m}}{\sqrt{nm}} \hat{S}_{n+m}\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu_X \leq \mu_Y$  gegen die Alternativhypothese  $H_1 : \mu_X - \mu_Y > 0$ ;
- (b) den linksseitigen Test  $\varphi_c^l := \mathbb{1}_{\{\bar{X}_n - \bar{Y}_m \leq -c \frac{\sqrt{n+m}}{\sqrt{nm}} \hat{S}_{n+m}\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \mu_X \geq \mu_Y$  gegen die Alternativhypothese  $H_1 : \mu_X - \mu_Y < 0$ ;
- (c) den beidseitigen Test  $\varphi_c^b := \mathbb{1}_{\{|\bar{X}_n - \bar{Y}_m| \geq c \frac{\sqrt{n+m}}{\sqrt{nm}} \hat{S}_{n+m}\}}$  für das *zweiseitige Testproblem* der Nullhypothese  $H_0 : \mu_X = \mu_Y$  gegen die Alternativhypothese  $H_1 : \mu_X - \mu_Y \neq 0$ .  $\square$

§26.28 **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 mit unverbundenen Stichproben, also  $(\mathbb{R}^{n+m}, \mathcal{B}^{n+m}, (\mathcal{P}_{(\mu_X, \sigma_X^2)}^n \otimes \mathcal{P}_{(\mu_Y, \sigma_Y^2)}^m)_{\mu_X, \mu_Y \in \mathbb{R}, \sigma_X, \sigma_Y \in \mathbb{R}_0^+})$ . Die Varianz  $\sigma_X^2$  bzw.  $\sigma_Y^2$  kann als Maß der Genauigkeit der Abfüllanlage interpretiert werden, so dass die Nullhypothese  $H_0 : \sigma_X^2 = \sigma_Y^2$  abzulehnen ist, wenn die Abfüllanlage nicht dieselbe Genauigkeit für die kleinen und großen Bierflaschen besitzt. Es ist nun nahe liegend, mit Hilfe der Statistik  $\hat{S}_X^{(2)}/\hat{S}_Y^{(2)}$  für den interessierenden Parameter  $\sigma_X/\sigma_Y$  Konfidenzbereiche bzw. Tests zu konstruieren.  $\square$

§26.29 **Tests auf Gleichheit der Varianzen.** Für  $X$  und  $Y$  sei ein Zwei-Stichproben-Modell mit unverbundenen Stichproben, also  $(X, Y) \odot (\mathcal{P}_{(\mu_X, \sigma_X^2)}^n \otimes \mathcal{P}_{(\mu_Y, \sigma_Y^2)}^m)_{\mu_X, \mu_Y \in \mathbb{R}, \sigma_X, \sigma_Y \in \mathbb{R}_0^+}$ , adäquat. Für  $c^o, c_u \in$

$\mathbb{R}_0^+$  betrachten wir

- (a) den rechtsseitigen Test  $\varphi_{c_u}^r := \mathbb{1}_{\{\hat{S}_X^{(2)}/\hat{S}_Y^{(2)} \geq c_u\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \sigma_X \leq \sigma_Y$  gegen die Alternativhypothese  $H_1 : \sigma_X/\sigma_Y > 1$ ;
- (b) den linksseitigen Test  $\varphi_{c_o}^l := \mathbb{1}_{\{\hat{S}_X^{(2)}/\hat{S}_Y^{(2)} \leq c_o\}}$  für das *einseitige Testproblem* der Nullhypothese  $H_0 : \sigma_X \geq \sigma_Y$  gegen die Alternativhypothese  $H_1 : \sigma_X/\sigma_Y < 1$ ;
- (c) den beidseitigen Test  $\varphi_{c_o, c_u}^b := \mathbb{1}_{\{\hat{S}_X^{(2)}/\hat{S}_Y^{(2)} \leq c_o\}} + \mathbb{1}_{\{\hat{S}_X^{(2)}/\hat{S}_Y^{(2)} \geq c_u\}}$  für das *zweiseitige Testproblem* der Nullhypothese  $H_0 : \sigma_X = \sigma_Y$  gegen die Alternativhypothese  $H_1 : \sigma_X/\sigma_Y \neq 1$ .  $\square$

§26.30 **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  $X \odot (\mathcal{P}_{(\mu_X, \sigma_X^2)}^n)_{\mu_X \in \mathbb{R}, \sigma_X \in \mathbb{R}_0^+}$  (Verbrauch vor dem Tuning) und  $Y \odot (\mathcal{P}_{(\mu_Y, \sigma_Y^2)}^n)_{\mu_Y \in \mathbb{R}, \sigma_Y \in \mathbb{R}_0^+}$  (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  $X - Y \odot (\mathcal{P}_{(\mu, \sigma^2)}^n)_{\mu \in \mathbb{R}, \sigma \in \mathbb{R}_0^+}$  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 §26.23](#) testen.  $\square$

## §26.05 Normales Lokations-Skalen-Modell

Im Folgenden sind  $(Z_i)_{i \in [0, m+k]}$  unabhängige und identisch standardnormalverteilte Zufallsvariablen, also  $(Z_i)_{i \in [0, m+k]} \sim N_{(0,1)}^{1+m+k}$ .

§26.31  **$\chi^2$ -Verteilung.** Die Verteilung der Zufallsvariable

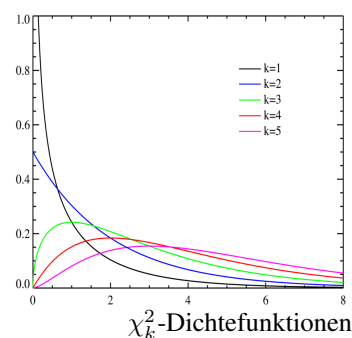
$$Q := \sum_{i \in [k]} Z_i^2$$

heißt (*zentrale*)  $\chi^2$ -Verteilung mit  $k$  *Freiheitsgraden* (vgl. [Beispiel §17.12 \(c\)](#)), kurz  $Q \sim \chi_k^2$ . Für  $\alpha \in (0, 1)$  bezeichnen wir weiterhin den Wert  $\chi_{k, \alpha}^2 \in \mathbb{R}_0^+$  als  $\alpha$ -Quantil einer (zentralen)  $\chi^2$ -Verteilung mit  $k$  Freiheitsgraden, falls  $\mathbb{P}(Q \leq \chi_{k, \alpha}^2) = \alpha$  gilt.

Für  $\delta \in \mathbb{R}$  heißt die Verteilung der Zufallsvariable

$$Q := (Z_1 + \delta)^2 + \sum_{i \in [2, k]} Z_i^2$$

*nicht-zentrale*  $\chi_k^2$ -Verteilung mit  $k$  *Freiheitsgraden* und *Nichtzentralitätsparameter*  $\delta^2$ , kurz  $Q \sim \chi_k^2(\delta^2)$  sowie  $\chi_{k, \alpha}^2(\delta^2) \in \mathbb{R}_0^+$  für das  $\alpha$ -Quantil einer nicht-zentralen  $\chi^2$ -Verteilung mit  $k$  Freiheitsgraden und Nichtzentralitätsparameter  $\delta^2$ , d.h.  $\mathbb{P}(Q \leq \chi_{k, \alpha}^2(\delta^2)) = \alpha$ .  $\square$



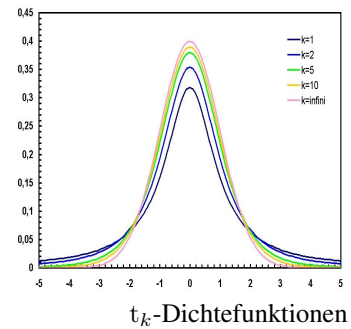
§26.32 **Korollar.** Sei  $Q \sim \chi_k^2$  und  $W \sim \chi_k^2(\delta^2)$ , dann gilt  $\mathbb{E}(Q) = k$ ,  $\mathbb{V}\text{ar}(Q) = 2k$  und  $\mathbb{E}(W) = \delta^2 + k$ . Für  $Z \sim N_{(0,1)}^m$ ,  $v \in \mathbb{R}^m$  und  $A \in \mathbb{R}^{m \times p}$  mit  $\text{rg}(A) = p$  gelten außerdem: (i)  $\|\Pi_{\text{Bild}(A)} Z\|^2 \sim \chi_p^2$  und (ii)  $\|Z + v\|^2 \sim \chi_m^2(\|v\|^2)$ .

§26.33 **Beweis** von **Korollar** §26.32. In der Vorlesung. □

§26.34 **(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  $T \sim t_k$ , und  $t_{k,\alpha} \in \mathbb{R}$  bezeichnet das  $\alpha$ -Quantil einer Student-t-Verteilung mit  $k$ -Freiheitsgraden, d.h.  $\mathbb{P}(T \leq t_{k,\alpha}) = \alpha$ .



§26.35 **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 bis zur Ordnung  $p < k$  (sie ist heavy-tailed). Insbesondere, ist  $T \sim t_k$  so gilt  $\mathbb{E}(T) = 0$  für  $k > 1$ , sowie  $\text{Var}(T) = k/(k-2)$  für  $k > 2$ . □

§26.36 **Korollar.** Für  $X \sim N_{(\mu, \sigma^2)}^n$  sind  $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \sim N_{(0,1)}$  und  $\frac{(n-1)}{\sigma^2} \hat{S}_n^{(2)} \sim \chi_{n-1}^2$  unabhängig, so dass  $\hat{T}_n = \frac{\sqrt{n}}{\hat{S}_n}(\bar{X}_n - \mu) \sim t_{n-1}$  mit  $\hat{S}_n := \sqrt{\hat{S}_n^{(2)}}$  gilt.

§26.37 **Beweis** von **Korollar** §26.36. In der Vorlesung. □

§26.38 **t-Tests für den Erwartungswert.** Für einen  $\mathbb{R}^n$ -wertigen Zufallsvektor  $X$  sei ein normales Lokations-Skalen-Modell adäquat, also  $X \odot (N_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$ . Dann hält der rechtsseitige Test  $\varphi_c^r$  mit  $c = t_{(n-1), (1-\alpha)}$ , der linksseitige Test  $\varphi_c^l$  mit  $c = t_{(n-1), (1-\alpha)}$  und der beidseitige  $\varphi_c^b$  mit  $c = t_{(n-1), (1-\alpha/2)}$  zu dem entsprechenden Testproblem in **Definition** §26.23 das Signifikanz-Niveau  $\alpha \in (0, 1)$  ein.

§26.39 **Bemerkung.** In der Vorlesung Statistik I zeigen wir, dass diese Tests auch gleichmäßig beste unverfälschte Test sind. □

§26.40  **$\chi^2$ -Tests für die Varianz.** Für einen  $\mathbb{R}^n$ -wertigen Zufallsvektor  $X$  sei ein normales Lokations-Skalen-Modell adäquat, also  $X \odot (N_{(\mu, \sigma^2)}^n)_{(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+}$ . Dann hält der rechtsseitige Test  $\varphi_{c_u}^r$  mit  $c_u = \chi_{(n-1), (1-\alpha)}^2$ , der linksseitige Test  $\varphi_{c^o}^l$  mit  $c^o = \chi_{(n-1), \alpha}^2$  und der beidseitige Test  $\varphi_{c^o, c_u}^b$  mit  $c^o = \chi_{(n-1), (1-\alpha/2)}^2$  und  $c_u = \chi_{(n-1), \alpha/2}^2$  zu dem entsprechenden Testproblem in **Definition** §26.24 das Signifikanz-Niveau  $\alpha \in (0, 1)$  ein.

### Zweistichproben Problem

§26.41 **Korollar.** Für  $(X, Y) \sim N_{(\mu_X, \sigma^2)}^n \otimes N_{(\mu_Y, \sigma^2)}^m$  sind  $\frac{\sqrt{nm}}{\sigma \sqrt{n+m}}(\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)) \sim N_{(0,1)}$  und  $(n+m-2)\hat{S}_{n+m}^{(2)} = \frac{(n-1)}{\sigma^2} \hat{S}_X^{(2)} + \frac{(m-1)}{\sigma^2} \hat{S}_Y^{(2)} \sim \chi_{n+m-2}^2$  unabhängig, so dass  $\hat{T}_{n,m} = \frac{\sqrt{nm}}{\hat{S}_{n+m} \sqrt{n+m}}(\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)) \sim t_{n+m-2}$  mit  $\hat{S}_{n+m} := \sqrt{\hat{S}_{n+m}^{(2)}}$  gilt.

§26.42 **Beweis** von **Korollar** §26.41. In der Vorlesung. □

§26.43 **t-Tests auf Gleichheit der Erwartungswerte.** Für zwei Zufallsvektoren  $X$  und  $Y$  mit derselben Varianz der Randverteilungen, sei ein normales Zwei-Stichproben-Modell mit unverbundenen Stichproben, also  $(X, Y) \odot (N_{(\mu_X, \sigma^2)}^n \otimes N_{(\mu_Y, \sigma^2)}^m)_{(\mu_X, \mu_Y, \sigma) \in \mathbb{R}^2 \times \mathbb{R}_0^+}$ , adäquat. Dann hält der rechtsseitige Test  $\varphi_c^r$  mit  $c = t_{(n+m-2), (1-\alpha)}$ , der linksseitige Test  $\varphi_c^l$  mit  $c = t_{(n+m-2), (1-\alpha)}$  und der beidseitige  $\varphi_c^b$  mit  $c = t_{(n+m-2), (1-\alpha/2)}$  zu dem entsprechenden Testproblem in Definition §26.27 das Signifikanz-Niveau  $\alpha \in (0, 1)$  ein.

§26.44 **Beispiel.** Betrachten wir die zufälligen Messwerte der beiden Segmente des Fürsten G. Ömetry in Beispiel VI im Kapitel 1 Prolog. Dann erhalten wir die Schätzwerte  $\bar{x}_{31} \approx 9.51$ ,  $\bar{y}_{31} \approx 10.49$ ,  $\hat{s}_x^{(2)} \approx 0.42$ ,  $\hat{s}_y^{(2)} \approx 1.05$ , sodass  $\hat{s}_{60}^2 \approx 0.74$ ,  $\hat{s}_{60} = \sqrt{\hat{s}_{60}^2} \approx 0.86$  und damit  $\frac{\sqrt{31}}{\hat{s}_{60}\sqrt{2}}|\bar{x}_{31} - \bar{y}_{31}| \approx 4.5$ . Unter der Annahme, dass ein normales Zwei-Stichproben-Modell mit unverbundenen Stichproben, und gleichen Varianzen vorliegt, hält der beidseitige  $\varphi_c^b$  mit  $c = t_{(60), (1-\alpha/2)}$  das Niveau  $\alpha$  ein. Für  $\alpha = 0.05$  erhalten wir  $t_{(60), (1-\alpha/2)} = 2.00$ . Da  $4.5 > 2$  können wir die Hypothese der Gleichheit zum Niveau  $\alpha = 0.05$  ablehnen.  $\square$

§26.45 **(Fisher-) F-Verteilung.** Die Verteilung der Zufallsvariable

$$F := \frac{\frac{1}{m} \sum_{i \in \llbracket m \rrbracket} Z_i^2}{\frac{1}{k} \sum_{i \in \llbracket m+1, m+k \rrbracket} Z_i^2}$$

heißt **zentrale (Fisher-) F-Verteilung** mit  $m$  und  $k$  **Freiheitsgraden**, kurz  $F \sim F_{n,k}$  und bezeichnen mit  $F_{m,k,\alpha}$  das  $\alpha$ -Quantil einer zentralen Fisher-F $_{n,k}$ -Verteilung mit  $m$  und  $k$  Freiheitsgraden, d.h.  $P(F \leq F_{m,k,\alpha}) = \alpha$ .

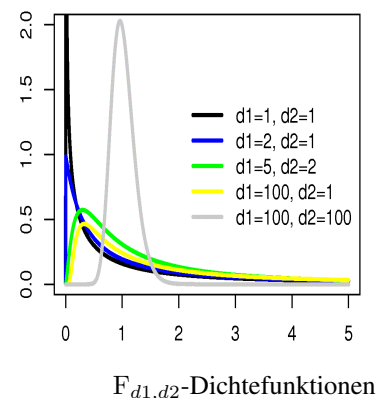
Für  $\delta \in \mathbb{R}$  heißt die Verteilung der Zufallsvariable

$$F := \frac{\frac{1}{m} \{(Z_1 + \delta)^2 + \sum_{i \in \llbracket 2, m \rrbracket} Z_i^2\}}{\frac{1}{k} \sum_{i \in \llbracket m+1, m+k \rrbracket} Z_i^2}$$

**nicht-zentrale (Fisher-) F-Verteilung** mit  $m$  und  $k$  **Freiheitsgraden** und **Nichtzentralitätsparameter**  $\delta^2$ , kurz  $F \sim F_{m,k}^{\delta^2}$  sowie  $F_{m,k,\alpha}^{\delta^2} \in \mathbb{R}$  für das  $\alpha$ -Quantil einer nicht-zentralen  $F_{n,k}^{\delta^2}$ -Verteilung mit  $m$  und  $k$  Freiheitsgraden und Nichtzentralitätsparameter  $\delta^2$ , d.h.  $P(F \leq F_{m,k,\alpha}^{\delta^2}) = \alpha$ .  $\square$

§26.46 **Bemerkung.** Sei  $F \sim F_{n,k}$  mit  $k > 1$ , dann ist  $F^{-1}$  eine  $F_{k,m}$ -verteilte Zufallsvariable. Für  $T \sim t_k$  ist  $T^2$  eine  $F_{1,k}$ -verteilte Zufallsvariable.  $\square$

§26.47 **F $_{n,k}$ -Tests auf Gleichheit der Varianzen.** Für  $X$  und  $Y$  sei ein normales Zwei-Stichproben-Modell mit unverbundenen Stichproben, also  $(X, Y) \odot (N_{(\mu_X, \sigma_X^2)}^n \otimes N_{(\mu_Y, \sigma_Y^2)}^m)_{\mu_X, \mu_Y \in \mathbb{R}, \sigma_X, \sigma_Y \in \mathbb{R}_0^+}$ , adäquat. Dann hält der rechtsseitige Test  $\varphi_{c_u}^r$  mit  $c_u = F_{(n-1), (m-1), (1-\alpha)}$ , der linksseitige Test  $\varphi_{c_o}^l$  mit  $c_o = F_{(n-1), (m-1), \alpha}$  und der beidseitige t-Test  $\varphi_{c_o, c_u}^b$  mit  $c_o = F_{(n-1), (m-1), (1-\alpha/2)}$  und  $c_u = F_{(n-1), (m-1), \alpha/2}$  zu dem entsprechenden Testproblem in Definition §26.29 das Signifikanz-Niveau  $\alpha \in (0, 1)$  ein.



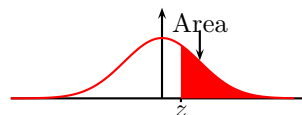
§26.48 **Beispiel.** Betrachten wir die zufälligen Messwerte der abgefüllten Volumen in den 33cl bzw. 75cl Bierflaschen der Abbaye de Rochefort in Beispiel VI im Kapitel 1 Prolog. Dann erhalten wir die Schätzwerte  $\bar{x}_{42} \approx 32.94$ ,  $\bar{y}_{42} \approx 74.98$ ,  $s_X^{(2)} \approx 0.74$ ,  $s_Y^{(2)} \approx 0.81$ , und damit  $\frac{s_X^{(2)}}{s_Y^{(2)}} \approx 0.91$ . Unter der Annahme, dass ein normales Zwei-Stichproben-Modell mit unverbunden Stichproben vorliegt, hält der beidseitige  $F_{n,k}$ -Test  $\varphi_{c^o, c_u}^b$  mit  $c^o = F_{41,41,(1-\alpha/2)}$  und  $c_u = F_{41,41,\alpha/2}$  das Niveau  $\alpha$  ein. Für  $\alpha = 0.05$  erhalten wir  $F_{41,41,0.975} = 1.86$  und  $F_{41,41,0.025} \approx 0.54$ . Da  $0.91 \in (0.54, 1.86)$  können wir die Hypothese der Gleichheit der Varianzen zum Niveau  $\alpha = 0.05$  nicht ablehnen.  $\square$



# 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.



| $z$ | Second decimal place of $z$ |       |       |       |       |       |       |       |       |       |
|-----|-----------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     | 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                |       |       |       |       |       |       |       |       |       |





## Literaturverzeichnis

- H. Bauer. *Maß- und Integrationstheorie*. Berlin etc.: Walter de Gruyter, 2., überarbeitete Auflage, 1992.
- J. Elstrodt. *Maß- und Integrationstheorie*. Berlin: Springer, 7., überarbeitete und ergänzte Auflage, 2011.
- H.-O. Georgii. *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Berlin: De Gruyter, 5., überarbeitete und ergänzte Auflage, 2015.
- A. Klenke. *Wahrscheinlichkeitstheorie*. Springer Spektrum, 3., überarbeitete und ergänzte Auflage, 2012.
- U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Braunschweig: Vieweg, 8., erweiterte Auflage, 2005.
- L. von Bortkewitsch. *Das Gesetz der kleinen Zahlen*. Leipzig: B. G. Teubner., 1898.

