

Wahrscheinlichkeitsrechnung I & II

Ein vorläufiges Skript zu den (2+1)-Vorlesungen im WS 2023/24 bzw. SS 2024

WS 2023/24: Vorlesung (Mattner): Di 14:15–15:45 HS 9,
3 Ü-Gruppen (Lea Willems lea.willems@uni-trier.de, Lena Jonas s4lejona@uni-trier.de):
Do 9:00–9:45 E 45, Fr 14:00–14:45 digital
<https://uni-trier.zoom-x.de/j/61635484751?pwd=L1g0RTQ4V1ZEQWdEeE5TaHpiMTVDdz09>,
Meeting-ID: 616 3548 4751, Kenncode: WR123456, und Mo 12:15–13:00 HS 8,
Korrekteure: Hannah Gentes s3hagent@uni-trier.de, Lena Jonas.

SS 2024: Vorlesung (Mattner): Di 12:15–13:45, HS 9.
Übungsgruppen (Lea Willems und Lena Jonas): Fr 11:00–11:45 und Fr 12:15–13:00, E 44.

Prof. Dr. Lutz Mattner,

Universität Trier, Fachbereich IV - Mathematik, mattner@uni-trier.de

2024-02-27

Inhaltsverzeichnis

Formalia	ii
Übungs- und Vorlesungsübersicht WS 2023/24	iii
Übungs- und Vorlesungsübersicht SS 2024	iv
Index	v
Literatur	ix
Allgemeine Notationsvereinbarungen	xiv
Einleitung	1
1 Aussagenlogik und Boole-Algebren	4
2 Mathematisierung des empirischen Wahrscheinlichkeitsbegriffes	15
3 Elementare Wahrscheinlichkeitsräume	30
4 Elemente der Kombinatorik	48
5 Zufallsgrößen und ihre Verteilungen	56
6 Bedingte Wahrscheinlichkeiten	64
7 Kopplungsmodelle	70
8 Stochastische Unabhängigkeit	81
9 Bernoulli-Ketten, Binomial-, Negativbinomial- und Multinomialverteilungen, Faltungen	91
10 Verteilungsfunktionen, Quantile, Erwartungswerte, Gesetz großer Zahlen	103
11 Poisson-Approximation	119
12 Ein Zentraler Grenzwertsatz	124
13 Statistische Modelle, Schätzprobleme, Tests	140
A Boole-Algebren	155
B Summen	169
Klausurvorbereitungsaufgaben zur WR I	177
Hinweise zu den mündlichen Prüfungen der WR I & II	185

Formalia

Dieses Dokument wird vorlesungsbegleitend aktualisiert und möglichst verbessert, und etwa wöchentlich auf Stud.IP neu hochgeladen. Es wird in der Vorlesung und den Übungen in der aktuellen Fassung bis zum jeweiligen Vorlesungsstand als vorliegend vorausgesetzt. Es soll am Ende mindestens ein Inhaltsverzeichnis der Vorlesungen, praktisch alles in der Vorlesung angeschriebene (Ausnahmen: alle Bilder), vieles aber sicher nicht alles des gesagten, wenigstens nicht vorgetragenes (vgl. die Vorlesungsablaufpläne), alle Übungsaufgaben, und einige prüfungsrelevante Informationen enthalten. Dabei wird versucht, folgenden Regeln zu genügen:

1. Schreibe nur Wahres.
2. Formuliere so eindeutig und klar wie möglich, solange dies nicht Regel 1 verletzt.
3. Formuliere kurz, einfach, und elegant, solange dies nicht Regel 1 oder 2 verletzt.
4. Halte Dich an etablierte Konventionen, wie Sprach-, Rechtschreib-, und Kommaregeln, solange dies nicht eine der Regeln 1 bis 3 verletzt.

Verbesserungsvorschläge, insbesondere die Regeln 1 und 2 betreffend, sind stets sehr willkommen.

Unabhängig von diesem Dokument bilden die Vorlesungen und Übungen eine *Präsenzveranstaltung*; alle dort gegebenen Informationen sind also zu beachten.

Adressaten der **WR I** im Wintersemester sind Bachelor-Studenten der Mathematik (1-Fach), Wirtschaftsmathematik, oder des Lehramts (jeweils als Pflichtteilmodul eines Moduls “Stochastik”), des Nebenfachs Mathematik oder der Informatik (als Pflichtmodul “Wahrscheinlichkeitsrechnung”), sowie des freien Wahlbereiches nichtmathematischer Bachelorstudiengänge (Modul “Wahrscheinlichkeitsrechnung I”).

Adressaten der **WR II** im Sommersemester als zweites Teilmodul der WR sind die Bachelor-Studenten des Lehramts (Pflichtmodul “Modul 7: Stochastik”).

Weitere Hörer wären natürlich auch herzlich willkommen.

Prüfungskrams: Zur WR I ist eine **Klausur** (105 Minuten) am Dienstag 2024-02-27 ab 09:00 Uhr im Audimax vorgesehen, dazu ein Wiederholungstermin am Dienstag 2024-03-19 ab 10:00 Uhr auch im Audimax. Ein weiterer Wiederholungstermin im akademischen Jahr 2023/24 ist nicht vorgesehen. Klausureinsicht jeweils am Mittwoch danach von 11:00 bis 12:00 in E 50 bzw. E 51.

Zulassungsvoraussetzung zur Klausur sind 40 % der erreichbaren Übungspunkte der WR I.

Für Studenten der Informatik, des Nebenfachs Mathematik, oder eines freien Wahlbereichs ist die Klausur die Modulprüfung.

Für Studenten der Mathematik (1-Fach) oder Wirtschaftsmathematik ist das Bestehen der Klausur eine Prüfungszulassungsvoraussetzung für die mündliche Prüfung des Moduls “Stochastik”, dort bestehend aus WR I und Wahrscheinlichkeitstheorie.

Für Studenten des Lehramts ist das Bestehen der Klausur zur WR I und eine erfolgreiche Teilnahme an den Übungen zur WR II Prüfungszulassungsvoraussetzung für die mündliche Modulprüfung über die WR; dabei bedeutet “erfolgreich”, dass mindestens 25 % der Aufgaben in Einzelabgabe bearbeitet wurden und aus diesen nach Wahl des Übungsleiters mindestens zwei in den Übungen richtig und didaktisch akzeptabel vorgetragen wurden, inklusive der Beantwortung von Fragen der Kommilitonen und des Übungsleiters. Bei längeren, etwa vierteiligen Aufgaben können nach Maßgabe des Übungsleiters auch Teillösungen reichen. Für die sich dann über den Stoff der WR I und II erstreckende mündliche Prüfung werden Prüfungstermine wie folgt angeboten: 1. Unmittelbar gegen Vorlesungsende (im Regelfall nicht empfohlen, gedacht z.B. für ins Ausland gehende). 2. Mit einigen Wochen Abstand zum Vorlesungsende, voraussichtlich in der zweiten Septemberhälfte, aber jedenfalls nicht in den Kalenderwochen 33–37. Genaueres nach Absprache in der Vorlesung im Sommersemester. 3. Ein Wiederholungstermin nach der Vorlesungszeit des Wintersemesters 2024/25, voraussichtlich in der zweiten Märzhälfte 2025.

Ein Bestehen der Klausur lediglich als Prüfungsvorleistung kann nicht als Modulprüfung in einem anderen Studiengang anerkannt werden.

Bei allem sind Änderungen vorbehalten.

Vorlesungs- und Übungsübersicht WS 2023/24

(15 Wochen, geplant 15 V- und knapp 15 Ü-Termine)

<i>Vorlesungstermin:</i>	<i>Behandelt bzw. geplant (mit Skript bis dahin Ihnen vorliegend vorausgesetzt!):</i>
1. 2023-10-24	Einleitung – Satz 1.7 (Bew. Satz 1.4 nur knapp skizziert)
2. 2023-10-31	Rest Bew. 1.7(b), Bem. hinter Aufg. 1.10, Anhang A – Satz A.8 (ohne Bew. (6,7))
3. 2023-11-07	Satz A.11, Def. A.14, Aufg. A.15(a), Satz 1.11 – dessen Beweisskizze (ohne 3.)
4. 2023-11-14	Def. 1.19 – Bsp. 1.22, A.16, A.17, Bsp. 2.1 – Def. 2.2, Satz 2.6 knapp angedeutet
5. 2023-11-21	Def. 2.3 (W-Inhalte), Bem. 2.4, Bew. Satz 2.6 – Text vor Def. 2.11 (Auswahlregeln)
6. 2023-11-28	Def. 2.11 – Satz 2.18
7. 2023-12-05	Ende Kap. 2, Quasipartitionen p. xiv, Kap. 3 (Elementare W-Räume) bis Satz 3.3
8. 2023-12-12	Bem. 3.4 (ohne Details) – Bem. 3.18
9. 2023-12-19	Beispiel 3.19 – Satz 3.25 (Siebformel)
10. 2024-01-09	Kap. 4 (Kombinatorik) – Def. A_k in Bsp. 4.11 (Rencontre-Problem)
11. 2024-01-16	Rest Bsp. 4.11, Anfang Kap. 5 (Zufallsgrößen) – Anfang Bsp. 5.6
12. 2024-01-23	Rest Bsp. 5.6 – Ende Kap. 5, Anfang Kap. 6 (Bedingte W.keiten) – Bsp. 6.3
13. 2024-01-30	Satz 6.4 – Behauptung Satz 7.2(a), Beispiel 7.4
14. 2024-02-06	Satz 7.2 mit Bew. Teil (a) – Bemerkungen 7.9
15. 2024-02-13	Bsp. 7.10 – Def. 7.14. Grob angedeutet: Def 8.1 mittels 7.14 wegen 8.21, 8.18(b,c).

<i>Aufgaben (eine Woche vor Abgabe spezifiziert)</i>	<i>Abg. bis 9:00 Uhr in Kasten E 10, oder in Ü in E 45, am Donnerstag (mit Ausnahmen wie angegeben)</i>
0.1, 1.1, 1.3, 1.8(a)	Präsenzaufgaben 2023-10-26, -27, -30, keine Abgabe
0.2, 1.8(b-d), 1.9**, 1.10	2023-11-02
A.4, A.7, A.10	2023-11-09
A.12*, A.13(a,d), 1.15, 1.17(3 der 5 ad lib.), A.21(a,b,c)*	2023-11-16
A.15(3 der 8 ad lib.), 2.5, 2.7, 1.18*, A.22*	2023-11-23
2.10, A.18, A.19*, A.20, A.23*	2023-11-30
2.14, 2.15, 2.17, A.24*	2023-12-07
3.3(7,8,13), 3.5, 3.14	2023-12-14
3.15, 3.27, 3.28	2023-12-21
3.24(c)*, 3.29*, 3.26*, 3.30*, 4.14*	2024-01-11
4.6, 4.7, 4.8*, 4.12	2024-01-18
4.13, 5.3, 5.7, 4.9*	2024-01-25
5.14, 5.16, 6.4(6,2,7), 6.10*	2024-02-01
6.5, 6.9, 7.5, 7.6*	2024-02-08
7.7*, 7.12*, 7.13*	2024-02-15
4.15*, 4.16*, 5.17*, 8.3*	Aufg. zur ersten Woche der WR II im SS 2024

Es ist sowohl Einzel- als auch Zweierabgabe erlaubt. Es gibt jede Woche in der Regel drei normale Aufgaben mit je gleichvielen Punkten; schwerere oder längere Aufgaben werden großzügiger bewertet. Mit einem Sternchen versehene Aufgaben sind Zusatzaufgaben, deren Punkte nur in den Zähler der Übungspunktprozentzahl des jeweiligen Bearbeiters eingehen. Manche Aufgaben haben im Skript mehr Teile als hier, zwecks Erleichterung des Übens des *sorgfältigen* Aufschreibens, zur Abgabe gefordert werden; in solchen Fällen lohnt sich natürlich auch eine nicht notwendig schriftliche Beschäftigung mit den anderen Teilen. Ein **Problem**, hier mit zwei Sternchen gekennzeichnet, ist eine vermutlich lösbare Aufgabe, zu der Ihr Dozent jedoch keine Lösung kennt. Es bringt analog doppelte Punktzahl im Zähler. Die Aufgaben stehen illustrierend oder ergänzend im Haupttext des Skriptes, oder aber am Ende des jeweiligen Kapitels falls sie in diesem ein Nebenthema betreffen (z.B. A.21, 3.27) oder mehr oder weniger austauschbare Trainingsaufgaben sind (z.B. A.22); nicht immer ist diese Klassifikation eindeutig.

Vorlesungs- und Übungsübersicht SS 2024

(14 Wochen, geplant 14 V- und 14 Ü-Termine)

<i>Vorlesungstermin:</i>	<i>Behandelt bzw. geplant (mit Skript bis dahin Ihnen vorliegend vorausgesetzt!):</i>
1. 2024-04-16	Bsple. 7.15, Bem. 7.16, Kapitel 8 (Stochastische Unabhängigkeit) – ...
2. 2024-04-23	
3. 2024-04-30	
4. 2024-05-07	
5. 2024-05-14	
6. 2024-05-21	
7. 2024-05-28	
8. 2024-06-04	
9. 2024-06-11	
10. 2024-06-18	
11. 2024-06-25	
12. 2024-07-02	
13. 2024-07-09	
14. 2022-07-16	Ende Kapitel 13

<i>Aufgaben</i>	<i>Abgabe bis 13:00 Uhr am Donnerstag</i> (evtl. mit Ausnahmen wg. Himmelfahrt und Fronleichnam)
4.15*, 4.16*, 5.17*, 8.3*	2024-04-18
8.4, ...	2024-04-25
	2024-05-02
	2024-05-09
	2024-05-16
	2024-05-23
	2024-05-30
	2024-06-06
	2024-06-13
	2024-06-20
	2024-06-27
	2024-07-04
	2024-07-11
	2024-07-18

Bis zur Vorwoche in beiden Gruppen noch nicht vorgerechnete Aufgaben können auch später noch abgegeben und vorgerechnet werden.

Index

- absorbierend, 157
- Absorbtiionsgesetz, 159
- Addition, 166
- Additivität der Summenbildung, 169
- Additivität des Erwartungswertes, 107
- Additivität eines Inhalts, 16
- Alphabet, 4
- also, xiv
- Alternative, 146
- äquivalent, 7
- auf, xv
- aus M , 49
- Aussagenkalkül, 7
- Austauschbarkeit, 77
- Auswahlregel, 23
- Automorphismus, 163

- Bayes-Formel, 66
- bedingendes Ereignis, 27, 65
- bedingt auf, 65
- bedingte Verteilung einer Zufallsgröße, 80
- bedingte Verteilung eines Quasikollektivs, 27
- bedingte Wahrscheinlichkeit, 27, 65
- Beobachtungsgröße, 140
- Bernoulli-Kette, 91
- Bernoulli-Kettenverteilung, 91
- Bernoulli-Verteilung, 37, 60
- Bijektionsmethode, 50
- Bild, xiv
- Bildmaß, 56
- Bildraum, 56
- Bilinearität, 112
- bimodal, 37
- binäre Relation, 8
- Binomialkonfidenzintervalle, 143
- Binomialmodell, 141
- Binomialverteilung, 37, 38, 78
- Bonferroni-Ungleichungen, 44
- Boole-, 155
- Boole-Algebra, 155
- Boole-Algebra-Homomorphismus, 163
- Boole-Welt, 13
- boolescher Verband, 162
- boolescher Ring, 166
- Borel- σ -Algebra, 156
- Buchstabe, 4

- Cantelli-Ungleichung, 114
- Chebychev-Ungleichung, 113
- Clopper-Pearson-Intervalle, 143

- d.h., xiv
- das heißt, xiv
- de Finetti, Indikatorklammernotation, xv
- De Morgansche Regeln, 157
- Dichte, 34
- Dichte der Standardnormalverteilung, 125
- Dirac-Maß, 39
- Dirac-Verteilung, 37, 39
- disjunkt, 164
- Disjunktifizierung, 33
- diskrete Arkussinusverteilung, 47
- diskrete Dreiecksverteilung, 59
- diskreter Wahrscheinlichkeitsraum, 30
- Distributivität, 172
- Doppelsummensatz, 169
- Dreiecksungleichung für Erwartungswerte, 107
- Dualität, 156

- eigentlicher Filter, 168
- einfache Zufallsstichprobe, 62
- Einpunktverteilung, 39
- Eins einer Boole-Algebra, 155
- einseitige Abweichungswahrscheinlichkeiten, 114
- Einsmenge, 35
- Einzelwahrscheinlichkeiten, 30
- Elementanzahl, xiv
- Elementaraussage, 4
- elementare (Wahrscheinlichkeits)zähldichte, 34
- elementare bedingte Verteilung, 79
- elementare Funktion, 126
- elementarer Desintegrationssatz, 71
- elementarer Grundraum, 30
- elementarer Wahrscheinlichkeitsraum, 30
- Elementarereignis, 30
- elementares Wahrscheinlichkeitsmaß, 30
- Empiristische Wahrscheinlichkeitstheorie, 1
- endliche Additivität, 16
- Endlichkeitskorrektur, 112
- entweder oder, xiv
- Ereignis, 30
- Ergebnis, 30
- erste Pfadregel, 73
- erstes Argument eines Markov-Kerns, 70
- Erwartungswert, 106
- erweiterte reelle Zahlen, 170
- erzeugte Unteralgebra, 165
- EWT, 1
- Experiment, 141
- Extrempunkt, 47

- faktorierte bedingte Verteilung einer Zufallsgröße, 79
- Faltung, 94
- fast sicher, 58
- Fehlerniveau (eines Tests), 146
- Filter, 168
- Fishers exakter Test, 147
- Fixmenge, xiv
- Fixpunkt, xiv
- Flanken einer Verteilung, 125
- folglich, xiv
- Formel von der totalen Wahrscheinlichkeit, 65
- frei, 10

- gemeinsame Verteilung, 61
- geometrische Verteilung, 37, 99
- gerichtet, 103
- Gleichverteilung, 39
- Grundmenge, 156

- Häufigkeitsinterpretation, 116
- halbsummierbar, 173
- Hochrechnung, 141
- Homogenität, 172
- Homogenität des Erwartungswertes, 107
- hypergeometrische Verteilung, 37, 62
- hypergeometrisches Modell, 141
- Hypothese, 146

- idempotent, 157, 166
- identisch verteilt, 57
- Indikator, xv
- Indikatorfunktion, xv
- Indikatorgröße, 60
- indizierte Partition, xiv
- indizierte Quasipartition, xiv
- Inhalt, 16
- Inklusionsindikation, 23
- Inklusionsindikator, 23
- intention to treat, 153
- interessierender Parameter, 141
- Involution, 7
- irreales Weltereignis, 13
- Isomorphismus, 163
- Isotonie, 16, 31, 172
- Isotonie des Erwartungswertes, 107
- Iverson, Indikatorklammernotation, xv

- Johnson-Identität, 7

- k -Teilmenge, 49
- kanonische Zufallsgröße, 58
- kanonisches statistisches Modell, 140
- Kette, 32
- Kollektiv, 24
- Kombinationen mit Wiederholung, 49
- Kombinationen ohne Wiederholung, 49
- kommutativ, 166
- komplementär, 164
- komplementäre Verteilungsfunktion, 104
- Komplementarität, 31
- Komplementierung, 155
- komplementstabil, 156
- Komplexprodukt, 95
- Konfidenzbereich, 141
- Konfidenzniveau, 141
- konvex, 47
- Konvexkombination, 17
- Koordinatengröße, 75
- Kopplung, 71, 73
- Kopplungsmodell, 73
- Korrelationskoeffizient, 117
- Kovarianz, 112
- kritischer Bereich eines Tests, 146

- Länge, 91
- Laplace-Verteilung, 39
- Leibnizreihen, 53

- Markov-Kern, 70
- Markovsche Kopplung, 74
- maximale Kette, 32
- Median, 105
- Mengen- σ -Algebra, 156
- Mengen-Algebra, 156
- Merkmalalgebra, 16
- Minkowski-Summe, 95
- Modalstellenmenge, 43
- Modelle, 141
- Modellereignis, 16, 30
- Moment, 110
- Mononumeralist, 142
- Morphismus von Boole-Algebren, 163
- Multiindex, 99
- Multinomialkoeffizient, 99
- Multinomialverteilung, 100
- multinomische Formel, 99
- Multiplikation, 166

- Negation, 4, 166
- Negativbinomialverteilung, 37, 99
- Nichtabbruchsbedingung, 24
- nichtdegeneriert, 16, 155
- Niveau eines Konfidenzbereichs (= Konfidenzniveau), 141
- Niveau eines Tests (= Fehlerniveau), 146

- Normalverteilung, 139
- Normiertheit, 16, 31, 34
- Null einer Boole-Algebra, 155
- Nullkonvention, 171
- Nullmenge, 35
- Nulltreue, 16, 31

- obere Faktorielle, 48
- obere Quantilfunktion, 105
- offenbar, 165

- P von A , 31
- p.d., xiv, 17, 168
- paarweise Additivität eines Wahrscheinlichkeitsmaßes, 31
- paarweise disjunkt, xiv, 17, 168
- paarweise Subadditivität, 31
- paarweise Unabhängigkeit, 88
- Parameterraum, 140
- Partition, xiv
- Periodensterbetafel, 2
- Permutationen mit Wiederholung, 49
- Permutationen ohne Wiederholung, 49
- Permutationsinvarianz, 77, 169
- Poisson-Verteilung, 37, 54
- polnische Notation, 5
- positiv abhängig, 87
- Positivität, 16, 30, 139
- Potenzmengenalgebra, 156
- $\text{Prob}(\Omega)$, $\text{prob}(\Omega)$, 34
- Problem, iii
- Produkt, 75
- Produkt(wahrscheinlichkeits)maß, 74
- Produktverteilung, 74
- punktweise fest, xiv

- Quantil, 105
- Quantilfunktion (= untere Quantilfunktion), 105
- Quantilfunktion der Standardnormalverteilung, 125
- Quasikollektiv, 16
- Quasipartition, xiv

- Randverteilung, 61
- reales Weltereignis, 13
- Realisierung, 57
- Realität, 13
- rechtszensierte Negativbinomialverteilung, 98
- Rencontre-Verteilung, 54
- Ring, 166
- Rng, 166

- S -Aussage, 4
- S -Ereignis, 7
- Schäferprinzip, 50
- Schätzer, 141
- Schätzproblem, 141
- Schnittbildung, 155
- sicheres Ereignis, 30
- Siebformel, 44
- Siebformel für $n = 2$, 31
- Simpson-Paradoxon, 66
- simultane Konfidenzbereiche, 145
- Stabdiagramm, 37
- Standardabweichung, 110
- standardisierte Zufallsgröße, 113
- Standardnormalverteilung, 125
- starke Nullkonvention, 172
- Startverteilung, 71
- statistische Inferenz, 146
- statistisches Schließen, 146
- Stichprobenraum, 140
- stochastisch unabhängig, 81
- streng wachsendes k -Tupel, 49
- Substitutionsregel, 169
- Subtraktivität, 31
- Sukzessives Ziehen, 76
- Summe, 30, 172, 173
- summierbar, 173
- Supremumsabstand, 121
- symmetrische Differenz in einer Boole-Algebra, 166

- Test, 146
- Testproblem, 146
- totale Additivität, 31
- totale Subadditivität, 31
- Totalvariationsabstand, 121
- Träger (im engeren oder weiteren Sinn), 35
- Trägermenge einer algebraischen Struktur, 155
- Trefferwahrscheinlichkeit, 91
- Trefferwahrscheinlichkeit, 38, 60, 78
- Tupel aus M (mit Wiederholung), 49
- Tupel aus M ohne Wiederholung, 49

- Übergangskern, 70
- Übergangs(zähl)dichte, 70
- Ultrafilter, 168
- umhüllen, 44
- Umordnungssatz, 30, 169
- unäre Verknüpfung, 7
- unabhängig, 81
- unabhängige Kopplung, 74
- und so weiter, 44
- uniforme Verteilung, 39
- unimodal (auf \mathbb{Z}), 43

unkanonisches statistisches Modell, 140
unkorreliert, 118
unmögliches Ereignis, 30
Unter-(Boole-)Algebra, 165
untere Faktorielle, 48
Urbild, xiv
Ursache, 87

Vandermonde'sche Faltungsformel, 51
Varianz, 110
Verband, 162
Verbundverteilung, 61
Vereinigungsbildung, 155
 \cup -stabil (lies: vereinigungsstabil), 156
Verschiebungssatz), 111
Vertauschbarkeit, 77
Verteilung, 16, 31, 56, 141
Verteilungsannahme, 141
Verteilungsfunktion, 103
Verteilungsfunktion der Standardnormalverteilung, 125
Verundung, 4
Vorstellung (einen Weltausschnitt betreffend), 13

wachsendes k -Tupel, 49
Wahrscheinlichkeit, 16
Wahrscheinlichkeit von A , 31
Wahrscheinlichkeitsfunktion, 34
Wahrscheinlichkeitsinhalt, 16
Wahrscheinlichkeitstheorie, 1
Wahrscheinlichkeitsverteilung, 31
Wahrscheinlichkeitszähldichte, 17
Weltereignis, 13
WT, 1

Zählgröße, 61
Zentraler Grenzwertsatz, 126
zentrierte Zufallsgröße, 113
ZGWS, 126
Ziegenproblem, 78
Zufallsgröße, 56
Zufallsvariable, 56
zugehörigen, 140
Zwei-Jungen-Problem, 72
zweiseitige Abweichungswahrscheinlichkeiten, 114
zweites Argument eines Markov-Kerns, 70
zweites Moment, 110

Literatur

Einführungen in die Stochastik

Unsere Vorlesungen orientieren sich ab Kapitel 3 überwiegend an langbewährtem Lehrbuchmaterial, wie etwa Henze (2021) oder Krenzel (2005); eine Ausnahme ist ein technisches Detail in der Definition des Begriffs “elementarer Wahrscheinlichkeitsraum”, hier im wesentlichen Storch/Wiebe (2003, Satz 7.A.9) folgend. In den Kapiteln 1 und 2 dagegen gehen wir viel genauer als üblich auf das Verhältnis zwischen Wahrscheinlichkeitsrechnung und Erfahrungswelt ein, hier von Mises (1951) folgend und präzisierend.

Jedes der in der folgenden Liste genannten Bücher, und manches nichtgenannte, bringt gegenüber allen anderen einen Mehrwert. Das eine oder andere ältere mag stellenweise mathematisch unpräziser¹ und daher zunächst schwerer lesbar sein als neuere, enthält aber interessantes und anderswo fehlendes Zusatzmaterial. An freien Internet-Materialien sind sicher die *stochastikclips* von Norbert Henze, https://www.youtube.com/channel/UCbFqBeoq75hF0xQT9Q1bDaA/feed?activity_view=3, für den Anfänger empfehlenswert, so manch anderes (zum Beispiel die konfuse YouTube-Beiträge eines gewissen Studienabbrechers) dagegen nicht.

Die hier nicht weiter aufgeführten üblichen Einführungen in die allgemeine maßtheoretische Wahrscheinlichkeitstheorie, siehe etwa Billingsley (1995) auf p. xi, setzen meist eine “elementare” Wahrscheinlichkeitsrechnung als bekannt voraus.

- BANDELOW, C. (1989). *Einführung in die Wahrscheinlichkeitstheorie*. 2., vollständig überarbeitete Auflage, BI Wissenschaftsverlag.
- BEHNEN, K. und NEUHAUS, G. (2003). *Grundkurs Stochastik*. 4. neu bearbeitete und erweiterte Auflage, PD-Verlag Heidenau. Korrekturen: <https://www.math.uni-hamburg.de/personen/neuhaus/febn01.pdf>
- DÜMBGEN, L. (2003). *Stochastik für Informatiker*. Springer.
- FELLER, W. (1970). *An Introduction to Probability Theory and Its Applications, Volume I*. Third edition (copyright 1968), revised printing, Wiley.
- FREEDMAN, D., PISANI, R. und PURVES, R. (2007). *Statistics*. Fourth Edition, Norton.
- GEORGH, H.-O. (2007). *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 3. Auflage, de Gruyter.
- HENZE, N. (1997, . . . , 2017, 2018, 2021). *Stochastik für Einsteiger. Eine Einführung in die faszinierende Welt des Zufalls*. 13 Auflagen, Springer Spektrum.
- HESSE, C. (2009). *Wahrscheinlichkeitstheorie. Eine Einführung mit Beispielen und Anwendungen*. 2., überarbeitete Auflage, Vieweg+Teubner.
- HODGES, J.L. und LEHMANN, E.L. (1964). *Basic Concepts of Probability and Statistics*. Holden-Day.
- JACOBS, K. (1992). *Discrete Stochastics*. Birkhäuser.
- JORDAN, K. (1972). *Chapters on the Classical Calculus of Probability*. Akadémiai Kiadó. [Übersetzung des 1956 erschienen ungarischen Originals.]

¹Mathematische Präzision ist ein relativer Begriff, dessen Anforderungen sich jedenfalls bisher im Laufe der Zeit immer weiter verschärft haben, siehe Klein (1926, p. 53) und ähnlich von Mises (1990, p. 191) in der auf p. x beginnenden zweiten Literaturliste.

- KERRICH, J.E. (1946). *An Experimental Introduction to the Theory of Probability*. Einar Munksgaard.
- KOLMOGOROFF, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer.
- KRENGEL, U. (2005). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 8., erweiterte Auflage, Vieweg.
- MILBRODT, H. (2010). *Wahrscheinlichkeitstheorie. Ein Einführung mit Anwendungen und Beispielen aus der Versicherungs- und Finanzmathematik*. DGVMF.
- MORGENSTERN, D. (1968). *Einführung in die Wahrscheinlichkeitstheorie und mathematische Statistik*. Zweite, verbesserte Auflage, Springer.
- MOSTELLER, F. (1965). *Fifty Challenging Problems in Probability with Solutions*. Addison-Wesley.
- MÜLLER, C. und DENECKE, L. (2013). *Stochastik in den Ingenieurwissenschaften. Eine Einführung mit R*. Springer.
- PFANZAGL, J. (1991). *Elementare Wahrscheinlichkeitsrechnung*. 2. Auflage, de Gruyter.
- PLACHKY, D., BARINGHAUS, L. und SCHMITZ, N. (1978). *Stochastik I. Eine elementare Einführung in die Grundbegriffe der Wahrscheinlichkeitsrechnung und Statistik*. Akademische Verlagsgesellschaft.
- RÉNYI, A. (1962, 1966, . . . , 1979). *Wahrscheinlichkeitstheorie. Mit einem Anhang über Informationstheorie*. Sechs ab 1966 (Ausgabe letzter Hand) wohl kaum veränderte Auflagen, DVW.
- (1970 a). *Probability theory*. North-Holland. [Etwas erweiterte Übersetzung von Rényi (1966).]
- (1970 b). *Foundations of Probability*. Holden-Day. [Ein zu obigen beiden weitgehend komplementäres Buch.]
- STORCH, U. und WIEBE, H. (2003). *Lehrbuch der Mathematik. Band 1: Analysis einer Veränderlichen*. 3. Auflage, Spektrum Akademischer Verlag.
- USPENSKY, J.V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill.
- VON MISES, R. (1928, 1936, 1951, 1972). *Wahrscheinlichkeit, Statistik und Wahrheit*. Vier Auflagen, Springer. [Die dritte Auflage ist letzter Hand, die vierte “durchgesehen von Hilda Geiringer”.]
- (1931). *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Deuticke.
- (1964). *Mathematical Theory of Probability and Statistics*. Edited and complemented by Hilda Geiringer. Academic Press.

Weitere im folgenden gelegentlich genannte Literatur

Der Anfang von Kapitel 1 folgt zum Teil den ersten 11 Seiten von Halmos (1956); für den Rest des Kapitels und für Anhang A kann man etwa die ersten Seiten von Halmos (1963) oder Givant/Halmos (2009) konsultieren. Das Sternchen * an einer Literaturangabe zeigt an, dass diese hier aus einer Sekundärquelle übernommen wurde, ohne das Original gesehen zu haben.

- ARBUTHNOTT, J. (1712). An argument for divine providence, taken from the constant regularity observ'd in the births of both sexes. *Philosophical Transactions* **27**, 186–190.
- BARBOUR, A.D., HOLST, L. und JANSON, S. (1992). *Poisson Approximation*. Oxford University Press.
- BEHREND, E. (2007). *Analysis, Band 2*. 2., aktualisierte Auflage, Vieweg.
- BELLHOUSE, D.R. (2000). De Vetula: a medieval manuscript containing probability calculations. *International Statistical Review* **68**, 123–136.
- BICKEL, P.J., HAMMEL, E.A. und O'CONNELL, J.W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science* **187**, 398–404.

- BILLINGSLEY, P. (1995). *Probability and Measure*. Third Edition, Wiley.
- BRANDENBURG, M. (2017). *Einführung in die Kategorientheorie*. 2. Auflage, Springer.
- CARATHEODORY, C. (1956). *Maß und Integral und ihre Algebralisierung*. Birkhäuser.
- CARTAN, H. (1937). Théorie des filtres. Filtres et ultrafiltres. *Comptes Rendus de l'Académie des Sciences de Paris* **205**, 595–598, 777–779. Auch in Cartan (1979, pp. 953–959).
- (1979). *Œuvres. Collected Works, Volume III*. Edited by R. Remmert and J-P. Serre, Springer.
- CHOW, Y.S. und TEICHER, H. (1997). *Probability Theory. Independence, Interchangeability, Martingales*. Third Edition, Springer.
- CORONARY DRUG PROJECT RESEARCH GROUP (1980). Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *New England Journal of Medicine* **303**, 1038-1041, <https://pubmed.ncbi.nlm.nih.gov/6999345/>.
- DEDEKIND, R. (1888, 1893, 1911, 1918). *Was sind und was sollen die Zahlen?* Vier Aufl., Vieweg.
- *DE FINETTI, B. (1967). Sur quelques conventions qui semblent utiles. *Revue Roumaine de Mathématiques Pures e Appliquées* **12**, 1227-1233. [Englische Übersetzung in de Finetti (1972, pp. xviii–xxiv).]
- (1972). *Probability, Induction and Statistics. The Art of Guessing*. Wiley.
- DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics.
- DIAMOND, A.H. und MCKINSEY, J.C.C. (1947). Algebras and their sub-algebras. *Bulletin of the American Mathematical Society* **53**, 959–962.
- DINEV, T. und MATTNER, L. (2012). The asymptotic Berry-Esseen constant for intervals. *Teoriya Veroyatnostei i ee Primeneniya* **57**, 381-384. Auch in *Theory of Probability and Its Applications* **57**, 323-325 (2013).
- DUBINS, L.E. und SAVAGE, L.J. (1965). *How to Gamble If You Must: Inequalities for Stochastic Processes*. McGraw-Hill. [1976 “unabridged, corrected republication” bei Dover mit dem auf Druck des neuen Verlages geänderten Titel *Inequalities for Stochastic Processes (How to Gamble If You Must)*. 2014 dann beim selben Verlag wieder unter dem alten Titel, “edited and updated by” W.D. Sudderth und D. Gilat.]
- DUDLEY, R.M. (2002). *Real Analysis and Probability*. 2nd Ed., Cambridge University Press.
- DUGUNDJI, J. (1966). *Topology*. Allyn and Bacon.
- EHM, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statistics & Probability Letters* **11**, 7–16.
- ELSTRODT, J. (2018). *Maß- und Integrationstheorie*. 8., erw. u. aktualisierte Auflage, Springer.
- FEICHTINGER, G. (1973) *Bevölkerungststatistik*. de Gruyter.
- FELLER, W. (1939). Über die Existenz von sogenannten Kollektiven. *Fundamenta Mathematicae* **32**, 87–96.
- (1971). *An Introduction to Probability Theory and Its Applications, Vol. II*. Second edition, Wiley.
- FISHER, R.A. (1958 a). Lung cancer and cigarettes? *Nature* **182**, 108. <https://www.nature.com/articles/182108a0>
- (1958 b). Cancer and smoking. *Nature* **182**, 596. <https://www.nature.com/articles/182596a0>
- FREEDMAN, D. (1977). A remark on the difference between sampling with and without replacement. *Journal of the American Statistical Association* **72**, 681.
- GÄNSSLER, P. und STUTE, W. (1977). *Wahrscheinlichkeitstheorie*. Springer.
- GAWRONSKI, W. (1996). *Grundlagen der Linearen Algebra*. Aula-Verlag.
- GIVANT, S. und HALMOS, P. (2009). *Introduction to Boolean Algebras*. Springer.
- GRÄTZER, G. (2011). *Lattice Theory: Foundation*. Birkhäuser.
- HAILPERIN, T. (1965). Best possible inequalities for the probability of a logical function of events. *American Mathematical Monthly* **72**, 343–359.

- HALD, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. Wiley.
 (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley.
- HALMOS, P.R. (1956). The basic concepts of algebraic logic. *American Mathematical Monthly* **53**, 363–387. [Auch in Halmos (1962), dort pp. 9–33.]
 (1960). *Naive Set Theory*. Van Nostrand.
 (1962). *Algebraic Logic*. Chelsea.
 (1963). *Lectures on Boolean Algebras*. Van Nostrand. [Nachdruck 1974, Springer.]
- HENZE, N. (2013 b). *Irrfahrten und verwandte Zufälle. Ein elementarer Einstieg in die stochastischen Prozesse*. Springer Spektrum.
- HERRLICH, H. und STRECKER, G.E. (2007). *Category Theory*. 3rd edition, Heldermann.
- HEUSER, H. (2009). *Lehrbuch der Analysis, Teil 1*. 17. Auflage, Vieweg+Teubner.
- HILBERT, D. (1899, 1903, 1909, . . . , 1977). *Grundlagen der Geometrie*. 12 Auflagen, Teubner.
 (1900). Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker-Vereinigung* **8**, 180–184.
 Erhältlich (2021-07-14) über <https://www.mathematik.de/jahresberichte-archiv>.
- HILBERT, D. und BERNAYS, P. (1964). *Grundlagen der Mathematik I*. Springer.
- HUNTINGTON, E.V. (1904). Postulates for the algebra of logic. *Transactions of the American Mathematical Society* **5**, 288–309, 552.
 (1933). New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell’s Principia Mathematica. *Transactions of the American Mathematical Society* **35**, 274–304, 557–558, 971.
- IVERSON, K.E. (1962). *A Programming Language*. Wiley.
- JOHNSON, N.L., KEMP, A.W. und KOTZ, S. (2005). *Univariate Discrete Distributions*. Third Edition, Wiley.
- JOHNSON, W.E. (1892). The logical calculus. I. General principles, II, III. *Mind* **1**, 3–30, 235–250, 340–357.
- KELLEY, J.L. (1961). *General Topology*. Corrected reprint, van Nostrand.
- KLEIN, F. (1926). *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Teil I*. Springer. [Nachdrucke in einem Band mit Teil II (1927) bei Chelsea 1967, Springer 1979.]
- LANG, S. (2002). *Algebra*. Revised Third Edition, Springer.
- LAPLACE, P.S. (1812, 1814, 1820). *Théorie analytique des probabilités*. Drei Auflagen, Courcier.
 Nachdruck der ersten Auflage 1967, Culture et Civilisation. Nachdruck der dritten Auflage in Laplace (1847), *Œuvres complètes*, vol. 7, Gauthier-Villars, diese nachgedruckt 1966, Olms.
- LEHMANN, E.L. und ROMANO, J.P. (2005). *Testing Statistical Hypotheses*. Third Ed., Springer.
- LORENZEN, S. (1998). BSE ist keine Seuche. Die Massentötung von Naturrindern in Deutschland ist illegal. *Die Zeit*, http://www.zeit.de/1998/09/BSE_ist_keine_Seuche
- MATTNER, L. (2014). *Also existiert Gott*. https://www.math.uni-trier.de/~mattner/2014-11-16_Also_existiert_Gott.pdf. Als Leserbrief gekürzt in *Forschung & Lehre*, Heft 10/14, p. 990, <http://www.forschung-und-lehre.de/>.
- MATTNER, L und SCHULZ, J. (2018). On normal approximations to symmetric hypergeometric laws. *Transactions of the American Mathematical Society* **370**, 727–748.
- MEREDITH, C.A. und PRIOR, A.N. (1968). Equational logic. *Notre Dame Journal of Formal Logic* **9**, 212–226.
- NOLL, W. (1997). *The role of the professor*. 3 pp., <https://www.math.cmu.edu/~wn0g/RP.pdf>.
- PADMANABHAN, R. und RUDEANU, S. (2008). *Axioms for Lattices and Boolean Algebras*. World Scientific.
- PEARSON, K. (1978). *The History of Statistics in the 17th and 18th Centuries, against the Changing Background of Intellectual, Scientific and Religious Thought*. Lectures by Karl Pearson given at the University College London during the academic sessions 1921–1933, edited by E.S. Pearson. Griffin.

- PÓLYA, G. und SZEGÖ, G. (1970). *Aufgaben und Lehrsätze der Analysis I*. 4. Auflage, Springer.
- RICKEY, V.F. (1985). Mathematics of the Gregorian Calendar. *Mathematical Intelligencer* **7**, 53–56.
- ROOS, B. (1996). *Metrische Poisson-Approximation*. Dissertation, Universität Oldenburg.
- ROSENBLICHT, M. (1972). Integration in finite terms. *American Mathematical Monthly* **79**, 963–972.
- RUDIN, W. (1987). *Real and Complex Analysis*. 3rd edition, McGraw-Hill.
- SCHNEIDER, I. (HRSG.) (1988). *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933. Einführungen und Texte*. Wissenschaftliche Buchgesellschaft.
- SCHULZ, J. (2016). *The Optimal Berry-Esseen Constant in the Binomial Case*. Dissertation, Universität Trier, <http://ubt.opus.hbz-nrw.de/volltexte/2016/1007/>
- SHAFER, G. und VOVK, V. (2006). The sources of Kolmogorov’s Grundbegriffe. *Statistical Science* **21**, 70–98.
- SHEFFER, H. (1913). A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society* **14**, 481–488.
- SHEVTSOVA, I.G. (2013). On the absolute constants in the Berry-Esseen inequality and its structural and nonuniform improvements. (Russisch, mit englischer Zusammenfassung.) *Informatika i Ee Primeneniya* **7**, 124–125.
- STATISTISCHES BUNDESAMT (2021). *Sterbetafeln. Ergebnisse aus der laufenden Berechnung von Periodensterbetafeln für Deutschland und die Bundesländer 2018/2020*, https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00144056. [Link Stand 2023-10-13. Leider konnte man sich in den letzten Jahren weder auf Permanenz noch auf Benutzerfreundlichkeit verlassen. Sterbetafeln auch früherer Jahre bis 2011/13 unter https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00003028. Aktuelle Sterbetafel 2020/2022 erhältlich über https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/Tabellen/_tabellen-innen-lebenserwartung-sterbetafel.html.]
- STONE, M.H. (1936). The theory of representations for Boolean algebras. *Transactions of the American Mathematical Society* **40**, 37–111.
- TARSKI, A. (1935). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* **1**, 261–405.
- TVERSKY, A. und KAHNEMAN, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* **90**, 293–315.
- VAN NERVEN, P. (2023). *Optimal Error Bounds in Normal and Edgeworth Approximation of Symmetric Binomial and Related Laws*. Eingereichte Dissertation (Oktober 2023), Universität Trier.
- VATUTIN, V.A. und MIKHAILOV, V.G. (1983). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory of Probability and Its Applications* **27**, 734–743. Russisches Original in *Teoriya Veroyatnostei i ee Primeneniya* **27**, 684–692 (1982).
- VON MISES, R. (1990). *Kleines Lehrbuch des Positivismus. Einführung in die empiristische Wissenschaftsauffassung*. Herausgegeben und eingeleitet von Friedrich Stadler. Suhrkamp. [Originalausgabe 1939, Van Stockum, Den Haag. Übersetzt 1951 als: *Positivism. A Study in Human Understanding*. Harvard University Press.]
- WALD, A. (1937). Die Widerspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung. *Ergebnisse eine Mathematischen Kolloquiums* **8**, 38–72.
- WITTING, H. (1985). *Mathematische Statistik I*. Teubner.
- ZOLOTUKHIN, A., NAGAEV, S. und CHEBOTAREV, V. (2018). On a bound of the absolute constant in the Berry-Esseen inequality for i.i.d. Bernoulli random variables. *Modern Stochastics: Theory and Applications* **5**, 385–410.

Allgemeine Notationsvereinbarungen

Wir verwenden “**das heißt**”, abgekürzt “**d.h.**”, nur zur Formulierung einer behaupteten oder definitorischen Äquivalenz; im Gegensatz dazu behauptet “**also**” dasselbe wie “**folglich**”, nämlich nur eine Implikation. Also (!) ist zum Beispiel “Es sei $x \in \mathbb{R}$. Ist $x > 0$, d.h. $x^2 > 0 \dots$ ” nach unserem Sprachgebrauch falsch, dagegen “Es sei $x \in \mathbb{R}$. Ist $x > 0$, d.h. $x^3 > 0 \dots$ ” richtig.

“**Entweder ... oder**” ist stets ausschließend gemeint. Zum Beispiel sind zwei Äquivalenzklassen zur selben Äquivalenzrelation stets entweder gleich oder disjunkt, während irgend zwei Mengen durchaus sowohl gleich als auch disjunkt sein können.

Bei uns ist $\mathbb{N} = \{1, 2, \dots\}$ und $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

Es sei M eine Menge. Dann bezeichnet $\#M \in \mathbb{N}_0 \cup \{\infty\}$, oder auch $\#(M)$, die **Elementanzahl** von M . Für $k \in \mathbb{N}_0$ sei

$$M_{\neq}^k := \{x \in M^k : x_i \neq x_j \text{ für } i, j \in \{1, \dots, k\} \text{ mit } i \neq j\}.$$

Bilder und **Urbilder** von Mengen schreiben wir mit *eckigen* Klammern, also $f[A] := \{f(x) : x \in A\}$ und $f^{-1}[B] := \{x \in X : f(x) \in B\}$ falls $f : X$ eine Funktion ist und A, B Mengen sind mit $A \subseteq X$.

Ist $f : X \rightarrow X$, also eine Funktion von X nach X , so heißt ein $x \in X$ mit $f(x) = x$ **Fixpunkt** von f , ein $A \subseteq X$ mit $f[A] = A$ **Fixmenge** von f , und ein $A \subseteq X$ mit $f(x) = x$ für $x \in A$ **punktweise fest** unter f .

Ist A eine einelementige Menge, so bezeichne $\{\cdot\}^{-1}(A)$ ihr Element; für andere A bleibe $\{\cdot\}^{-1}(A)$ undefiniert². Ist nun A eine zunächst beliebige Menge, so ist damit etwa “ $x := \{\cdot\}^{-1}(A)$ ” zugleich eine Behauptung, nämlich die der Einelementigkeit von A , und eine Definition.

Eine Familie $(A_i : i \in I)$ von Mengen heißt **paarweise disjunkt**, kurz **p.d.**, falls $A_i \cap A_j = \emptyset$ für $(i, j) \in I_{\neq}^2$ gilt. Eine Menge \mathcal{A} von Mengen heißt p.d. falls $\text{id}_{\mathcal{A}} = (A : A \in \mathcal{A})$ dies ist.

Es sei \mathcal{X} eine Menge. Eine Familie $(A_i : i \in I)$ von Teilmengen von \mathcal{X} heißt **indizierte Quasipartition** von \mathcal{X} falls sie p.d. ist und $\bigcup_{i \in I} A_i = \mathcal{X}$ gilt; sie heißt **indizierte Partition** von \mathcal{X} falls zusätzlich $A_i \neq \emptyset$ für $i \in I$ gilt. Statt “indiziert” sagt man auch “parametrisiert” oder, falls auf I eine Ordnung gegeben ist, “geordnet”. Eine Teilmenge \mathcal{A} von $2^{\mathcal{X}}$ heißt **(nichtindizierte) Partition** von \mathcal{X} falls $\text{id}_{\mathcal{A}} = (A : A \in \mathcal{A})$ eine indizierte Partition von \mathcal{X} ist; analog mit “Quasi-”.³ Es ist

²Im Unterschied zur üblichen allgemeineren Notation $\bigcup A := \{\xi : \exists a \in A \text{ mit } \xi \in a\}$, die für jede Menge A (allgemeiner: für jede Klasse A) definiert ist, und die ja speziell für $A = \{x\}$ gerade $\bigcup \{x\} = x = \{\cdot\}^{-1}(\{x\})$ liefert.

Zum Verständnis des vorstehenden Satzes ist zu beachten, dass in den heutzutage üblichen Formalisierungen der Mathematik jedes Element einer Menge (oder einer Klasse) eine Menge ist. Siehe dazu etwa Halmos (1960) im Fall der Zermelo-Fraenkel-Mengenlehre ohne Klassen, und Kelley (1961), dort der Anhang “Elementary Set Theory”, für die Morse-Kelley-Version einer Mengen- und Klassenlehre. Die Behauptung der Identität $\bigcup \{x\} = x$ ist also nur dann sinnvoll, wenn x eine Menge ist, und wird dann durch Nachrechnen von $\xi \in \bigcup \{x\} \Leftrightarrow \xi \in x$ verifiziert.

³Häufig wird schon Partition genannt was bei uns Quasipartition heißt. Ein in mehreren sehr

also \mathcal{A} eine Partition von \mathcal{X} genau dann, wenn gilt: $\emptyset \notin \mathcal{A}$, $\bigcup \mathcal{A} = \mathcal{X}$, und wenn $A, B \in \mathcal{A}$, dann $A = B$ oder $A \cap B = \emptyset$.

Es sei zum Beispiel $\mathcal{X} := \{1, 2, 3\}$. Dann ist $\{\{1\}, \{2, 3\}\}$ eine Partition (von \mathcal{X}), während $\{\{1\}, \{2, 3\}, \emptyset\}$ lediglich eine Quasipartition ist. $(\{1\}, \emptyset, \{2, 3\}, \emptyset)$ ist eine geordnete Quasipartition. Das Quadrupel $(A_i : i \in I) := (\{1\}, \{2, 3\}, \emptyset, \{1\})$ ist keine indizierte Quasipartition, aber sein Bild $\{A_i : i \in I\} = \{\{1\}, \{2, 3\}, \emptyset\}$ ist eine Quasipartition.

Sind I und Ω Mengen und ist X_i für $i \in I$ eine Funktion mit Definitionsbereich Ω , so bezeichnen wir die “familienwertige” Funktion $\Omega \ni \omega \mapsto (X_i(\omega) : i \in I)$ mit

$$(X_i : i \in I);$$

dagegen bezeichnet $(X_i : i \in I)$ ja die Funktion $I \ni i \mapsto X_i$ mit Definitionsbereich I , deren Werte Funktionen mit Definitionsbereich Ω sind. Im zweiten Fall könnten die Funktionen X_i auch verschiedene Definitionsbereiche haben. Für “Funktionen von Zufallsgrößen” benutzen wir manchmal auch übliche aber eigentlich illegale Notationen wie $T(X)$ statt $T \circ X$ und $T(X_1, \dots, X_n)$ statt $T \circ (X_1, \dots, X_n)$.

Wir definieren den **Indikator** einer Aussage durch⁴

$$(\text{Aussage}) := \begin{cases} 1 & \text{wenn Aussage wahr,} \\ 0 & \text{wenn Aussage falsch oder sinnlos.} \end{cases}$$

also zum Beispiel $(3^2 = 8) = 0$, $(3^2 = 9) = 1$, $(\infty - \infty = 0) = 0$. Ist $A \subseteq \mathcal{X}$ eine Menge, so schreibt man

$$\mathbf{1}_A(x) := (x \in A) = \begin{cases} 1 & \text{für } x \in A, \\ 0 & \text{für } x \in \mathcal{X} \setminus A \end{cases}$$

und nennt die dadurch definierte Funktion $\mathbf{1}_A : \mathcal{X} \rightarrow \{0, 1\}$ den **Indikator** oder die **Indikatorfunktion** von A .

Bei Jahresangaben wie zum Beispiel in Fußnote 5 auf Seite 1 steht “v.d.Z.” für “vor der Zeitrechnungswende”, also im dortigen Beispiel 43 v.d.Z. = im Jahre -42 .

guten Büchern begangener Standardfehler besteht dann darin, nicht nötigenfalls die Betrachtung auf die dann “Partitionen bestehend aus nichtleeren Mengen” zu nennenden unserigen Partitionen einzuschränken. Siehe zum Beispiel Kelley (1961, p. 96; die “projection of X onto \mathcal{D} ” ist gar nicht *onto* (= **auf** := surjektiv) falls $\emptyset \in \mathcal{D}$), Dugundji (1966, p. 16, Remark 7.5; unter den “sets A_α ” könnte \emptyset sein, und dies ist keine Äquivalenzklasse), und Gawronski (1996, p. 3, Satz 0.1.5).

⁴Notation von **Iverson** (1962, p. 11) - **de Finetti** (1967, pp. xx–xxi in der englischen Übersetzung 1972).

Einleitung

Die **Wahrscheinlichkeitstheorie** (kurz: **WT**) ist ein Teilgebiet der Mathematik mit besonders vielfältigen Anwendungsmöglichkeiten, von der Bestimmung optimaler Strategien bei Glücksspielen⁵ über die Berechnung fairer Versicherungsprämien bis hin zur Bewertung medizinischer Therapien mittels der auf der WT aufbauenden Mathematischen Statistik. In ihrer seit 1933 üblichen Axiomatisierung durch⁶ Andrej Nikolaevič Kolmogorov (1903–1987) besteht die WT, wie jedes entwickelte Teilgebiet der Mathematik, aus recht wohlangeordneten Definitionen, Sätzen, Beweisen und innermathematischen Beispielen, und ist insofern leicht zu erlernen. Jedoch erscheint die Verbindung zwischen der WT und der Erfahrungswelt oft weniger klar als bei anderen anwendungsnahen Teilen der Mathematik, etwa der Geometrie, was nicht nur Anfängern dann große Schwierigkeiten bereitet.

Was zum Beispiel bedeutet es für einen männlichen, gerade 25-jährigen und den Abschluss einer Risikolebensversicherung ins Auge fassenden Familiengründer, dass seine (?) Sterbewahrscheinlichkeit für den Zeitraum bis zur 50-Jährigkeit, berechnet aus einer Sterbetafel des Statistischen Bundesamtes (2021, pp. 4-5), rund 0.03 sei? Während etwa der Strahlensatz der ebenen Geometrie die Berechnung einer gesuchten Länge in unserer Umwelt aus geeigneten anderen gemessenen Längen gestattet, und das Ergebnis der Rechnung auch prinzipiell ausführbaren oder doch zumindest idealisierend denkbaren Messungen bis auf kleine Fehler entspricht, welche Messung oder Beobachtung soll denn für unseren Familiengründer der Sterbewahrscheinlichkeit 0.03 entsprechen?

Ihr Dozent glaubt mit Richard von Mises (1883–1953), dass dessen ab 1919 in mehreren Publikationen begründete und 1937 von Abraham Wald (1902–1950) in einem wesentlichen Punkt präzisierende **Empiristische Wahrscheinlichkeitstheorie** (kurz: **EWT**) - in welcher genauen mathematischen Formulierung oder künftigen Weiterentwicklung auch immer - hier Klarheit schafft.⁷ Wesentliche Punkte der

⁵Wir nennen dieses vermeintlich belanglose oder gar unsittliche Beispiel als erstes, da ihm die WT einerseits wohl ihre Existenz, und andererseits sicherlich immer wieder wichtige Anregungen verdankt: Der erste uns bekannte Text zur WT ist das Gedicht *De vetula*, vorgeblich eine Autobiographie von Ovid (43 v.d.Z.–17?), aber geschrieben vermutlich um 1250; dort werden auf vier Seiten im Prinzip alle Einzelwahrscheinlichkeiten für die Augensumme beim dreifachen Würfelwurf angegeben, siehe Schneider (1988, pp. 1, 5-8) oder Bellhouse (2000). “How to Gamble If You Must” ist der (ursprüngliche und neuerliche) Haupttitel des wichtigen Buches von Dubins/Savage (1965) zur Theorie des optimalen Stoppens, thematisch in die in Trier meist im Wintersemester angebotene Vorlesung “Stochastische Prozesse” gehörend.

⁶Wir verwenden bei nichtlateinisch geschriebenen Eigennamen die wissenschaftlich übliche Transliteration. Bei Literaturangaben wird dagegen der Autorenname wie in der jeweiligen Publikation verwendet, was etwa im Fall von Kolmogoroff (1933) zu scheinbaren Inkonsistenzen führt.

⁷Die meisten Stochastik-Dozenten scheinen diesen Glauben nicht zu teilen, oder das alles für so trivial zu halten, dass sich eine Ausformulierung nicht lohne. Dass eine Begründung des Wahrscheinlichkeitsbegriffs - wenn sie hinausgehen soll über rein formale mathematische Definitionen wie “Anzahl der günstigen Fälle durch Anzahl der möglichen” in der kombinatorischen Wahrscheinlichkeitsrechnung, oder Definition 3.2, oder die maßtheoretische Definition der allgemeinen WT - nicht trivial ist, belegen jedoch die ernsthaften aber meist wortreichen und unbefriedigenden sowie ein-

EWT sind:

- In ihr sind Wahrscheinlichkeiten Grenzwerte relativer Häufigkeiten.
- Wahrscheinlichkeiten kommen nicht “Weltereignissen” zu, sondern “Kollektiven” von solchen, und dann auch zugehörigen “Modellereignissen” in einem “Merkmalraum”.
- Die übliche WT behandelt das Rechnen mit Wahrscheinlichkeiten⁸ von Modellereignissen, die EWT klärt den Anwendungsbezug.

Dementsprechend, und damit mehr als üblich auf die “Voraussetzungen für die Anwendbarkeit der Wahrscheinlichkeitsrechnung auf die Welt der reellen Geschehnisse”⁹ eingehend, werden in diesen Vorlesungen Wahrscheinlichkeiten in Kapitel 2 empiristisch eingeführt, wobei gegenüber von Mises eine höhere mathematische Präzision und damit auch mehr Verständlichkeit angestrebt wird. Da dort wie oben von verschiedenen Arten von “Ereignissen” die Rede ist, wird zuvor die “Algebra der Ereignisse”, nämlich die Aussagenlogik in Form der Booleschen Algebra, in Kapitel 1 behandelt. Ab Kapitel 3, und damit in ihrem Großteil, behandelt diese Vorlesung dann allgemein übliche WT, genauer deren elementaren, nicht die Maß- und Integrationstheorie voraussetzenden Teil.

Der Name “WR” dieser Vorlesungen, statt etwa “Einführung in die WT” oder “Einführung die Stochastik”, dient lediglich der bequemen Abgrenzung zur Sommersemestervorlesung WT, welche den Stoff der Wintersemestervorlesungen WR I und MIT als bekannt voraussetzt.

0.1 Aufgabe. Erklären Sie, wie man auf das obige Ergebnis “rund 0.03” an Hand der angegebenen Sterbetafel kommt. Zum Verständnis der Sterbetafel könnte Pfanzagl (1991, p. 239), Hesse (2009, p. 286), oder Feichtinger (1973, Abschnitt 3.1) nützlich sein. Was ergibt sich analog für weibliche Familiengründer?¹⁰ □

ander zum Teil widersprechenden Versuche vieler namhafter Mathematiker, Naturwissenschaftler, und Philosophen, von denen Jordan (1972, Chapter 1) einige zusammenfassend wiedergibt.

Der hier gewählte Name “EWT” ist zwar nicht allgemein üblich, erscheint jedoch später als fast selbsterklärend, und ist jedenfalls zur Abgrenzung gegenüber der üblichen WT gut geeignet.

⁸“Ce calcul délicat s’étend aux questions les plus importantes de la vie, qui ne sont en effet, pour la plupart, que des problèmes de probabilité” schreibt Laplace (1812, p. 0). Übertreibung oder nicht?

⁹Formulierung von Kolmogoroff (1933, p. 3), mit zustimmendem Verweis auf von Mises.

¹⁰Die in dieser und der nächsten Aufgabe zu verwendenden Tafeln sind sogenannte **Periodensterbetafeln**. Eine dort “Sterbewahrscheinlichkeit vom Alter x bis $x+1$ ” genannte und traditionell mit q_x bezeichnete Zahl bezieht sich auf eine Population (z.B. die männlichen Einwohner Deutschlands) in einem Beobachtungszeitraum (z.B. 2018–2020). Naiv denken wir uns hier zunächst q_x definiert als den Quotienten $\frac{D_x}{P_x}$ mit $P_x :=$ Anzahl der im Beobachtungszeitraum x -jährigen und $D_x :=$ Anzahl der im Beobachtungszeitraum x -jährig verstorbenen Populationsmitglieder.

Aus Periodensterbetafeln abgeleitete Vorhersagen können naturgemäß nur insoweit gültig sein, wie sich die Sterberaten in den betrachteten Alterstufen nicht ändern.

Die obige naive Definition ist unklar (wann im Beobachtungszeitraum x -jährig?) und würde bei

0.2 Aufgabe. (a) Bestimmen Sie analog zu Aufgabe 0.1 jeweils die zwei Sterbewahrscheinlichkeiten für 25-jährige bis zur 50-Jährigkeit mit den neun Sterbetafeln 2020/2022, 2018/2020 (bekannt aus Aufgabe 0.1), 2017/2019, 2016/2015, ..., 2011/2013.

(b) Für jedes der beiden Geschlechter: Betrachten Sie die maximale Differenz der Sterbewahrscheinlichkeiten aus Teil (a). Rund welcher Differenz von Sterbefallzahlen eines fixierten Geburtsjahrgangs, sagen wir 2002 in Deutschland, entspricht diese? \square

Mitzzählung im Nenner P_x von spät (z.B. Weihnachten 2020) x -jährig werdenden nicht berücksichtigen, dass diese nur wenig Zeit hätten, zum Zähler D_x beizutragen. Die tatsächliche Definition von q_x ist daher etwas komplizierter und - für Ihren Dozenten - mit der angegebenen Quelle nicht nachvollziehbar.

1 Aussagenlogik und Boole-Algebren

1.1 Aufgabe. Die folgende Liste enthält 5 Aussagen. Wie viele der Aussagen und welche sind wahr?

- Genau eine Aussage auf dieser Liste ist falsch.
- Genau zwei Aussagen auf dieser Liste sind falsch.
- Genau drei Aussagen auf dieser Liste sind falsch.
- Genau vier Aussagen auf dieser Liste sind falsch.
- Genau fünf Aussagen auf dieser Liste sind falsch. □

In WT-Texten ist ständig von “Ereignissen” die Rede. Rein mathematisch gesehen sind dies dort beliebige Elemente einer jeweils spezifizierten Boole-Algebra gemäß Definition A.1, wobei letztere meist speziell als Mengen- σ -Algebra gemäß Bemerkung A.5 gewählt wird, und in elementaren Vorlesungen wie dieser meist noch spezieller als die Potenzmengenalgebra einer spezifizierten Grundmenge, siehe unten Beispiel A.3(b) und Definition 3.2.

Aus Anwendungssicht dagegen soll ein “Ereignis” einer “Aussage” über einen betrachteten Weltausschnitt entsprechen. Dabei sollen “äquivalente” Aussagen “identifiziert” werden, und bequemerweise sollen auch “offensichtlich unzutreffende” Aussagen zugelassen sein. Dies wird mathematisch präzisiert durch die im folgenden kurz dargestellte Aussagenlogik.¹

1.2 Definition. Es sei Σ eine nichtleere Menge, es seien $\&, N \notin \Sigma$ mit $\& \neq N$, und es sei² $T := \Sigma \cup \{\&, N\}$. Mit

$$\begin{aligned} Ns &:= (N, s_1, \dots, s_m) && \text{für } m \in \mathbb{N} \text{ und } s \in T^m, \\ \&st &:= (\&, s_1, \dots, s_\ell, t_1, \dots, t_m) && \text{für } \ell, m \in \mathbb{N}, s \in T^\ell, t \in T^m, \end{aligned}$$

und $S := \Sigma^1 = \{(\sigma) : \sigma \in \Sigma\}$ sei $S_1 := S$ und induktiv

$$S_{n+1} := S_n \cup \{Ns : s \in S_n\} \cup \{\&st : s, t \in S_n\} \quad \text{für } n \in \mathbb{N}.$$

Damit heißt

$$S^* := \bigcup_{n \in \mathbb{N}} S_n$$

Menge aller S -Aussagen. Die Elemente des **Alphabets** Σ heißen **Buchstaben**, die von S **Elementaraussagen**. Für $s, t \in S^*$ heißt Ns die **Negation** von s , und $\&st$ die **Verundung** von s und t . □

¹Wir folgen hier im wesentlichen der von Halmos (1956, insbesondere Abschnitt 3) gegebenen Skizze. Ziel ist zunächst nur eine Klärung der Begriffe “Aussage” und “Ereignis” := “Äquivalenzklasse von Aussagen” unter der Annahme, dass wir schon Mengen, Relationen, Verknüpfungen, und die natürlichen Zahlen kennen. Die im “logischen” Aufbau der Mathematik *vor* der Mengenlehre benötigte Logik - siehe dazu etwa Halmos (1960, pp. v, 5-6) - kann damit natürlich nicht begründet, sondern höchstens nachträglich beschrieben werden.

²Es folgt ein griechisches “Tau”, d.h. der im griechischen Alphabet auf Σ folgende Großbuchstabe, und kein lateinisches “Te”; letzteres wäre in Formeln kursiv: T .

1.3 Aufgabe. (a) Schreiben Sie im Fall von $\Sigma = \{\sigma, \tau, \omega\}$ dreielementig die Mengen S_1 und S_2 durch explizite Auflistung der in ihnen enthaltenen Tupel hin.

(b) Rekursionsformel für $\#S_n$, Wertetabelle für kleine n und $\#\Sigma$. □

Interpretation: Die Elemente von S^* sind die mit “nicht” und “und” aus den Elementaraussagen in endlich vielen Schritten verknüpften Aussagen. Zum Beispiel wird für S -Aussagen s und t (elementar oder nicht) die S -Aussage

$$(1) \quad \&N\&stN\&sNt$$

sukzessive interpretiert als “ $N\&st$ und $N\&sNt$ ”, “Nicht $\&st$ und nicht $\&sNt$ ”,

$$(2) \quad \text{“Weder } s \text{ und } t, \text{ noch } s \text{ und nicht } t\text{”}.$$

Dagegen kann (1) zum Beispiel nicht als “ $N\&stN$ und $\&sNt$ ” gelesen werden, da $N\&stN$ gar nicht definiert ist, da der letzte Eintrag dieses Tupels N ist, was nach Definition 1.2 (oder auch nach 1.4(1) unten mit $k = m - 1$ und $k = m$) für kein Element von S^* gilt.

Die hier exemplarisch bemerkte eindeutige Lesbarkeit der auf Klammern verzichtenden sogenannten **polnischen Notation**³ in (1) wird unten in Satz 1.4(c) präzisiert. Zum Vergleich sei bemerkt, dass das Komma in (2) eine dort nötige Klammerung andeutet.

1.4 Satz. (a) **Abgeschlossenheit von S^* .** *Es ist $S \subseteq S^*$ und es gelten die beiden Implikationen*

$$s \in S^* \Rightarrow Ns \in S^*, \quad s, t \in S^* \Rightarrow \&st \in S^*.$$

(b) **Kleinstheit von S^* .** *Gilt (a) mit einer Menge U statt S^* , so gilt $S^* \subseteq U$.*

(c) **Eindeutige Lesbarkeit.** *Für $s \in S^*$ gilt genau eine der folgenden Aussagen:*

- (i) $s \in S$.
- (ii) *Es gibt ein $t \in S^*$ mit $s = Nt$.*
- (iii) *Es gibt ein Paar $(t, u) \in S^* \times S^*$ mit $s = \&tu$.*

Dabei ist t in (ii), beziehungsweise (t, u) in (iii), gegebenenfalls eindeutig bestimmt.

Beweis. (a) Sind $s, t \in S^*$, so gibt es wegen der aus $S_n \subseteq S_{n+1}$ für $n \in \mathbb{N}$, also der Isotonie der Mengenfolge $(S_n)_{n \in \mathbb{N}}$, ein $n \in \mathbb{N}$ mit $s, t \in S_n$, also $\&st \in S_{n+1}$ und damit $\&st \in S^*$. Die anderen beiden Teilaussagen sind offensichtlicher.

(b) Es sei U eine Menge mit $S \subseteq U$ und den beiden Implikationen $s \in U \Rightarrow Ns \in U$, $s, t \in U \Rightarrow \&st \in U$. Dann gilt $S_1 \subseteq U$, und für $n \in \mathbb{N}$ folgt aus der Annahme $S_n \subseteq U$ offenbar $S_{n+1} \subseteq U$. Also gilt $S_n \subseteq U$ für jedes n , also $S^* \subseteq U$.

(c) Vorüberlegung: Für $s = (s_1, \dots, s_m) \in S^*$ gilt für $k \in \{1, \dots, m\}$

$$(1) \quad \Delta(s, k) := \#\{i \in \{1, \dots, k\} : s_i \in \Sigma\} - \#\{i \in \{1, \dots, k\} : s_i = \&\}$$

$$\begin{cases} = 1 & \text{falls } k = m, \\ \leq 0 & \text{falls } k < m, \end{cases}$$

³Laut Tarski (1935, p. 283) von Łukasiewicz eingeführt.

wie man für $s \in S_n$ durch Induktion über n zeigt: Im Fall von $s = (s_1, \dots, s_m) \in S_1$ ist $m = 1$, also $k = 1$ in (1) und in der Tat $\Delta(s, k) = 1$. Gilt andererseits (1) für ein $n \in \mathbb{N}$ und jedes $s = (s_1, \dots, s_m) \in S_n$, so folgt für jedes solche s für $k \in \{1, \dots, m + 1\}$ dann

$$\Delta(Ns, k) = \left\{ \begin{array}{ll} 0 & \text{falls } k = 1, \\ \Delta(s, k - 1) & \text{falls } k > 1 \end{array} \right\} \left\{ \begin{array}{ll} = 1 & \text{falls } k = m + 1, \\ \leq 0 & \text{falls } k < m + 1 \end{array} \right.$$

wegen (1) für $k - 1$ statt k im zweiten Schritt, und mit $t = (t_1, \dots, t_\ell) \in S_n$ auch

$$\Delta(\&st, k) \left\{ \begin{array}{ll} = 1 & \text{falls } k = m + \ell + 1, \\ \leq 0 & \text{falls } k < m + \ell + 1 \end{array} \right.$$

trivialerweise im Fall $k = 1$, wegen (1) für $k - 1$ statt k im zweiten Schritt falls $k \in \{2, \dots, m + 1\}$, und wegen (1) für $(t, \ell, k - m - 1)$ statt (s, m, k) falls $k \in \{m + 2, \dots, m + \ell + 1\}$; also gilt dann auch (1) für $s \in S_{n+1}$. Aus (1) folgt, dass kein echtes Anfangsstück eines $s \in S^*$ ebenfalls in S^* liegt.

Es sei nun $s \in S^*$. Offenbar schließen sich (i),(ii),(iii) paarweise aus, im Fall von (ii) und (iii) wegen $N \neq \&$. Andererseits ist wiederum nach der induktiven Definition der S_n aber $s \in S$, oder $s = Nt$ mit einem $t \in S^*$, oder $s = \&tu$ mit geeigneten $t, u \in S^*$, d.h. einer der drei Fälle (i),(ii),(iii) tritt ein. Wäre in Fall (ii) auch $Nt = Nu$ mit einem $u \in S^*$, so folgte direkt $t = u$. Wäre im Fall (iii) auch $(v, w) \in S^* \times S^*$ mit $s = \&vw$ und $(v, w) \neq (t, u)$, so folgte mit $\&tu = \&vw$ dass t ein echtes Anfangsstück von v wäre, oder umgekehrt, im Widerspruch zur Vorüberlegung. Also gelten in den Fällen (ii) und (iii) die jeweiligen Eindeutigkeitsaussagen. \square

Unser Sprachgebrauch suggeriert nun, dass gewisse Elemente von S^* in einem noch zu präzisierenden Sinne als äquivalent zueinander anzusehen sind.

1.5 Beispiel. In einem Zweizeiler des römischen Dichters Martial ergibt die zweite Zeile, “nec tecum possum vivere nec sine te”, auf deutsch:

(1) “Weder kann ich mit dir leben, noch ohne dich”.

Rein inhaltlich gesehen hätte Martial kürzer “non possum vivere”, auf deutsch also

(2) “Ich kann nicht leben”

schreiben können. Mit

$$s := \text{“Ich kann leben”}, \quad t := \text{“Ich kann mit dir sein”}$$

kann nun einerseits (1) zu 1.3(2) und damit zu 1.3(1) umgeschrieben werden, andererseits (2) zu “Nicht s ” und damit zu Ns . Die hier empfundene “Äquivalenz” von (1) mit (2) suggeriert als eine Forderung an eine noch zu definierende Äquivalenzrelation \equiv auf S^* also die unten angegebene Gleichung 1.6(4). \square

Formal definiert wird nun \equiv als die kleinste Äquivalenzrelation auf S^* welche 1.6(4) und die weiteren, näherliegenden, Forderungen 1.6(1,2,3,5,6) erfüllt. Dabei bezieht sich die Kleinstheit von \equiv , also von einer Menge geordneter Paare (siehe dazu auch den Beweis von Satz 1.7), auf die Mengeninklusion als Ordnung: Jede andere Äquivalenzrelation E auf S^* welche 1.6(1–6) für $s, t, u \in S^*$ mit E statt \equiv erfüllt ist eine echte Obermenge von \equiv , d.h. für $s, t \in S^*$ gilt sEt nicht nur wenn $s \equiv t$, sondern auch in gewissen weiteren Fällen.

1.6 Definition. In der Situation aus Definition 1.2 bezeichne \mathcal{E} die Menge aller Äquivalenzrelationen \equiv auf S^* derart, dass für $s, t, u \in S^*$ gilt:

- (1) $\&sttu \equiv \&\&stu,$
- (2) $\&st \equiv \&ts,$
- (3) $NNs \equiv s,$
- (4) $\&N\&stN\&sNt \equiv Ns,$
- (5) $s \equiv t \Rightarrow Ns \equiv Nt,$
- (6) $s \equiv t \Rightarrow \&su \equiv \&tu.$

Es bezeichne nun \equiv den Schnitt $\bigcap \mathcal{E}$ all dieser Äquivalenzrelationen. Dann heißt (S^*, \equiv) **Aussagenkalkül** zur Grundmenge S , und zwei S -Aussagen $s, t \in S^*$ heißen **äquivalent** falls $s \equiv t$ gilt. □

1.7 Satz. Gegeben sei die Situation aus Definition 1.6.

(a) \equiv ist eine Äquivalenzrelation auf S^* mit 1.6(1–6) für $s, t, u \in S^*$.

(b) Für $s \in S^*$ bezeichne $[s]$ die zugehörige Äquivalenzklasse. Damit werden auf dem Quotienten

$$\mathcal{A} := S^*/\equiv = \{[s] : s \in S^*\}$$

durch

$$\begin{aligned} [s] \wedge [t] &:= [\&st] \quad \text{für } s, t \in S^*, \\ [s]' &:= [Ns] \quad \text{für } s \in S^* \end{aligned}$$

zwei Verknüpfungen wohldefiniert. Die binäre Verknüpfung \wedge ist assoziativ und kommutativ, die unäre Verknüpfung $'$ ist involutiv⁴, und es gilt die **Johnson-Identität**⁵

$$(1) \quad (A \wedge B)' \wedge (A \wedge B)' = A' \quad \text{für } A, B \in \mathcal{A}.$$

Die Elemente von \mathcal{A} heißen **S-Ereignisse**. □

⁴Eine **unäre Verknüpfung** auf eine Menge X ist praktisch nichts anderes als eine Funktion $f : X \rightarrow X$; diese heißt **Involution** von X falls noch $f \circ f = \text{id}_X$ gilt. Hier wird also $A'' = A$ für $A \in \mathcal{A}$ behauptet.

⁵Von uns wegen Satz A.11 so genannt. In (1) und auch sonst wird in Ausdrücken wie $A \wedge B'$ die üblichen Konvention “Unär geht vor binär” angenommen.

Der Terminus “Ereignis” ist hier entsprechend dem etwas unglücklichen Sprachgebrauch der WT gewählt⁶; besser wäre etwa “denkbares Ereignis” oder auch “Vorstellung”. Anschaulich denken wir bei einem $A \in S^*/\equiv$ an eine durch viele $s \in S^*$ beschreibbare Vorstellung von der Welt, welche zutreffend sein mag oder auch nicht. Zum Beispiel könnte A das Nichtlebenkönnen Martials repräsentieren, oder ein denkbares Ergebnis eines morgigen Fußballspiels, oder auch eine Präzisierung eines Details der antiken Geschichte.

Beweis Satz 1.7. Siehe Aufgabe 1.8 für einige hier nur angedeutete Details.

(a) Eine (**binäre**) **Relation** R auf einer Menge X ist definitionsgemäß nichts anderes als eine Teilmenge von X^2 , nur dass man statt $(x, y) \in R$ gerne xRy schreibt. Ein beliebiger Schnitt von Relationen auf X ist wieder eine solche: Ist nämlich \mathcal{R} eine Menge von Relationen auf X , d.h. $\mathcal{R} \subseteq 2^{X^2}$, so ist entweder $\mathcal{R} = \emptyset$ und dann definitionsgemäß der “leere Schnitt” $\bigcap \mathcal{R}$ gleich der Grundmenge, hier also X^2 , und dies ist eine Relation auf X , oder es ist $\mathcal{R} \neq \emptyset$ und damit für jedes $z \in \bigcap \mathcal{R}$ jedenfalls $z \in R$ für ein $R \in \mathcal{R}$ und damit $z \in X^2$, also wieder $\bigcap \mathcal{R} \subseteq X^2$. Ist speziell \mathcal{R} eine Menge von Äquivalenzrelationen auf X , so ist auch $\bigcap \mathcal{R}$ eine Äquivalenzrelation.

Also ist hier $\equiv = \bigcap \mathcal{E}$ eine Äquivalenzrelation auf S^* . Es seien nun $s, t, u \in S^*$. Für jedes $E \in \mathcal{E}$ gilt dann 1.6(1) mit E statt \equiv , d.h. $(\&s\&tu, \&\&stu) \in E$, und es folgt $(\&s\&tu, \&\&stu) \in \bigcap \mathcal{E} = \equiv$, d.h. 1.6(1). Analog bzw. ähnlich folgen 1.6(2–6).

(b) Wohldefiniertheit: Für $s, t, u, v \in S^*$ mit $[s] = [u]$ und $[t] = [v]$ gilt $s \equiv u$ und $t \equiv v$, und mit alternierender Anwendung von 1.6(6) und 1.6(2) folgt

$$\&st \equiv \&ut \equiv \&tu \equiv \&vu \equiv \&uv,$$

also $\&st \equiv \&uv$ wegen der Transitivität von \equiv , also $[\&st] = [\&uv]$, und mit 1.6(5) direkt $Ns \equiv Nu$, also $[Ns] = [Nu]$.

Die behaupteten Eigenschaften von \wedge und $'$ ergeben sich nun direkt aus 1.6(1–4): Für $A, B, C \in \mathcal{A}$, also $A = [s], B = [t], C = [u]$ mit $s, t, u \in S^*$, gilt

$$\begin{aligned} A \wedge (B \wedge C) &= [s] \wedge [\&tu] = [\&s\&tu] = [\&\&stu] = [\&st] \wedge [u] \\ &= (A \wedge B) \wedge C \end{aligned}$$

usw. □

1.8 Aufgabe. Vervollständigen Sie den Beweis von Satz 1.7 bezüglich

(a) der Äquivalenzrelationsbehauptung,

(b) der Eigenschaft 1.6(4),

(c) der Eigenschaft 1.6(6),

(d) der Involutionseigenschaft von $'$. □

⁶Halmos (1956), dort nicht in Hinblick auf die WT schreibend, nennt die S -Aussagen “sentences” und die Elemente von S^*/\equiv “propositions”. Normale Übersetzungen ins Deutsche wären “Sätze” und “Aussagen”, aber leider ist ersteres auch Übersetzung von “theorems”.

1.9 Problem. In der Situation von Definition 1.6 ergibt sich später aus Satz 1.11 leicht in Korollar 1.13 die Gültigkeit von

$$(1) \quad \&ss \equiv s \quad \text{für } s \in S^*.$$

Mit anderen Worten: Jede Äquivalenzrelation \equiv auf S^* mit 1.6(1-6) für $s, t, u \in S^*$ erfüllt auch (1). Können Sie dies hier einfach und direkt beweisen? \square

1.10 Aufgabe. Zeigen Sie in der Situation von Aufgabe 1.3(a), dass die Menge $\{[s] : s \in S_2\}$ höchstens 9 Elemente enthält. Dabei dürfen Sie die erst später zu beweisende Aussage 1.9(1) als bekannt ansehen. \square

Neben der Negation und der Verundung in Definition 1.2 hätten wir noch weitere Verknüpfungen in S^* und damit in $\mathcal{A} = S^*/\equiv$ einführen können, etwa eine “Veroderung”, und wir hätten neben 1.6(1–4) noch weitere Forderungen, etwa 1.9(1), durch Beispiele wie 1.5 motivieren können. Die nun folgenden Überlegungen zeigen jedoch, dass dies überflüssig wäre, da, im durch Satz 1.11(a) präzisierten Sinne, $(\mathcal{A}, \wedge, ')$ praktisch schon eine Boole-Algebra ist, woraus sich dann zum Beispiel 1.9(1) ergibt.

Wir setzen ab jetzt A.1–A.11 als bekannt voraus.

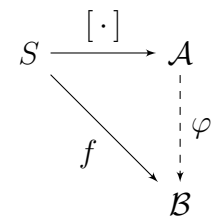
Um nun den Satz A.11 von Johnson–Huntington auf $(\mathcal{A}, \wedge, ')$ aus Satz 1.7 anwenden zu können, und damit unsere boolesche Analyse des Aussagenkalküls aus Definition 1.6 zu vollenden, muss noch $\#\mathcal{A} \geq 2$ gezeigt werden. Daher ist der folgende Teilsatz 1.11(a) noch nicht trivial, und es bereitet nur wenig Mehraufwand, gleich 1.11(b) mitzubeweisen.

Wir setzen daher ab jetzt auch A.14 und A.15(a) als bekannt voraus.

1.11 Satz. Gegeben sei die Situation aus Satz 1.7.

(a) Es ist $\mathcal{A} = S^*/\equiv$ mit \wedge und $'$ aus 1.7(b) und \vee, O, I aus A.11(1–2) eine Boole-Algebra.

(b) Es sei \mathcal{B} eine Boole-Algebra und $f : S \rightarrow \mathcal{B}$ eine Funktion. Dann gibt es genau einen Boole-Algebra-Morphismus $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ mit $\varphi([s]) = f(s)$ für $s \in S$.



Der Beweis wird im Anschluss an Bemerkung 1.16 geführt.

1.12 Korollar 1 (Konsistenz des Aussagenkalküls). In der Situation aus Definition 1.6 gilt $s \not\equiv Ns$ für jedes $s \in S^*$.

Beweis. Für $s \in S^*$ ist, mit $[\cdot]$ und $'$ wie in Satz 1.7, $[s] \neq [s]' = [Ns]$ wegen Satz 1.11(a) und A.8(7) im ersten Schritt, also $s \not\equiv Ns$. \square

1.13 Korollar 2. Im Aussagenkalkül gelten über 1.6(1–4) hinaus weitere Äquivalenzen, die sich aus den nicht O oder I enthaltenden Gleichungen in A.1(2), A.8(1,5), A.10(2) ergeben, darunter 1.9(1).

Beweis von 1.9(1). Für $s \in S^*$ ist, mit $[\cdot]$ und \wedge wie in Satz 1.7, $[\&ss] = [s] \wedge [s] = [s]$ wegen 1.11(a) und A.8(1) im zweiten Schritt, also $\&ss \equiv s$. \square

In den obigen beiden Korollaren wurde von Satz 1.11 nur dessen Teil (a) benutzt; ein Beispiel einer Anwendung von Teil (b) ist:

1.14 Korollar 3 (Inäquivalenz von Elementaraussagen). *In der Situation aus Definition 1.6 gilt $s \not\equiv t$ für $s, t \in S$ mit $s \neq t$.*

Beweis. Für gegebene verschiedene $s, t \in S$ liefert eine Anwendung von Satz 1.11(b) auf die zweielementige Boole-Algebra $\mathcal{B} := \{O, I\}$ gemäß Beispiel A.3(a), und auf $f : S \rightarrow \mathcal{B}$ definiert durch $f(s) = O$ und $f(u) := I$ für $u \in S \setminus \{s\}$, eine Funktion $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ mit

$$\varphi([s]) = f(s) \neq f(t) = \varphi([t]),$$

also $[s] \neq [t]$, also $s \not\equiv t$. □

1.15 Aufgabe. Der vollständige Zweizeiler aus Beispiel 1.5 lautet im Original:

“Difficilis, facilis, iucundus, acerbus es idem,
nec tecum possum vivere nec sine te”.

Verschaffen Sie sich eine Übersetzung ins Deutsche, formalisieren Sie diese analog zu unserem Vorgehen in 1.5, und geben Sie das zugehörige Element von S^*/\equiv mittels 1.11(a) in möglichst einfacher Form an. □

1.16 Bemerkung. In der Terminologie der etwas abstrakteren Algebra sagt die Existenzaussage in Satz 1.11(b), dass die Familie $([s] : s \in S)$ **frei** in der Boole-Algebra \mathcal{A} ist. (Die Eindeutigkeitsaussage ergibt sich dann daraus, dass die Menge $\{[s] : s \in S\}$ ein Erzeuger von \mathcal{A} ist.) Freie Familien sind notwendig injektiv, siehe Korollar 1.14.

Anschaulich bedeutet die Freiheit, dass für die Elemente dieser Familie keine besonderen Gleichungen gelten, d.h. keine die sich nicht schon, wie etwa die De Morganschen Regeln, aus den Boole-Algebra-Axiomen ergeben. Siehe dazu neben Korollar 1.14 auch die folgende Aufgabe 1.17. □

1.17 Aufgabe. In der Situation von Satz 1.11(a) seien $s, t, u \in S$. Dann gilt

- (1) $[s] \neq O,$
- (2) $[s] \wedge [t] \wedge [u] \neq O,$
- (3) $[s] \not\leq [t]$ falls $s \neq t,$
- (4) $s \neq Nt,$
- (5) $s \equiv \&tu$ genau dann, wenn $s = t = u.$

□

1.18 Aufgabe. Es seien S und T zwei gleichmächtige nichtleere Mengen von 1-Tupeln und es seien (S^*, \equiv) und (T^*, \equiv) die zugehörigen Aussagenkalküle gemäß Definition 1.6. Dann sind die Boole-Algebren $\mathcal{A} := S^*/\equiv$ und $\mathcal{B} := T^*/\equiv$ isomorph.

(*Hinweis:* Wenden Sie Satz 1.11(b) insgesamt vier mal an, zunächst auf zwei geeignete Funktionen von S nach \mathcal{B} beziehungsweise von T nach \mathcal{A} .) □

Beweis von Satz 1.11. 1. Es gelte die Prämisse von (b).

1.α Nach Definition 1.2 gilt

$$(1) \quad S = S_1 \subseteq S_n \subseteq S_{n+1} \quad \text{für } n \in \mathbb{N}$$

und $S^* = \bigcup_{n \in \mathbb{N}} S_n$.

Wegen der Gleichung und der ersten Inklusion in (1) und wegen Satz 1.4(c) können wir induktiv Funktionen $f_n : S_n \rightarrow \mathcal{B}$ für $n \in \mathbb{N}$ durch $f_1 := f$ und

$$f_{n+1}(s) := \left. \begin{array}{ll} f_n(s) & \text{falls } s \in S_n, \\ f_n(t)' & \text{falls } S_n \not\ni s = Nt \text{ mit } t \in S_n \\ f_n(t) \wedge f_n(u) & \text{falls } S_n \not\ni s = \&tu \text{ mit } t, u \in S_n \end{array} \right\} \text{ für } s \in S_{n+1}$$

definieren, und mit der zweiten Inklusion in (1) folgt $f_{n+1} \supseteq f_n$ für $n \in \mathbb{N}$. Daher ist

$$g := \bigcup_{n \in \mathbb{N}} f_n : S^* \rightarrow \mathcal{B}$$

eine Funktion, und es gilt

$$(2) \quad g(s) = f(s) \quad \text{für } s \in S,$$

$$(3) \quad g(Nt) = g(t)' \quad \text{und} \quad g(\&tu) = g(t) \wedge g(u) \quad \text{für } t, u \in S^*.$$

1.β Es sei \simeq die von g auf S^* erzeugte Äquivalenzrelation, also

$$(4) \quad s \simeq t \Leftrightarrow g(s) = g(t) \quad \text{für } s, t \in S^*.$$

Dann ist $\simeq \in \mathcal{E}$, mit \mathcal{E} definiert in 1.6, denn:

1.6(1) gilt mit \simeq statt \equiv , denn für $s, t, u \in S^*$ gilt

$$\begin{aligned} g(\&s\&tu) &= g(s) \wedge g(\&tu) = g(s) \wedge (g(t) \wedge g(u)) \\ &= (g(s) \wedge g(t)) \wedge g(u) = g(\&st) \wedge g(u) = g(\&\&stu) \end{aligned}$$

wegen der Assoziativität von \wedge in der Boole-Algebra \mathcal{B} im zentralen Schritt und wegen (3) in den übrigen Schritten, und mit (4) folgt $\&s\&tu \simeq \&\&stu$.

Analog zeigt man, dass 1.6(2-4) mit \simeq statt \equiv gelten.

Zu 1.6(5): Gilt $s \simeq t$, also $g(s) = g(t)$, so folgt mit (3) im ersten und im dritten Schritt

$$g(Ns) = g(s)' = g(t)' = g(Nt),$$

also $Ns \simeq Nt$.

Analog behandelt man 1.6(6).

1.γ Wegen $\simeq \in \mathcal{E}$ ist für $A \in \mathcal{A}$ die Menge $g[A]$, also das g -Bild einer \equiv -Äquivalenzklasse, einelementig, denn für $s, t \in A$ gilt $s \equiv t$, also $s \simeq t$ wegen $\simeq \in \mathcal{E}$, also $g(s) = g(t)$. Also definiert⁷

$$\varphi(A) := \{\cdot\}^{-1}(g[A]) \quad \text{für } A \in \mathcal{A}$$

⁷Zur Schreibweise $\{\cdot\}^{-1}$ siehe Seite xiv.

eine Funktion $\varphi : \mathcal{A} \rightarrow \mathcal{B}$, und es gilt

$$(5) \quad \varphi([s]) = g(s) \quad \text{für } s \in S^*,$$

wegen (2) also

$$(6) \quad \varphi([s]) = f(s) \quad \text{für } s \in S.$$

2. Eine Anwendung von 1. auf zum Beispiel $\mathcal{B} := \{O, I\}$ und $f(s) := 0$ für $s \in S$ liefert für ein (wegen der Voraussetzung $\Sigma \neq \emptyset$ aus Definition 1.2 existierendes) $s \in S$ mit (3) im ersten Schritt und mit A.8(7) im zweiten

$$g(Ns) = g(s)' \neq g(s),$$

also $Ns \not\sim s$, also $Ns \neq s$, also $\#\mathcal{A} \geq 2$, und damit (a) nach Satz 1.7 und Satz A.11.

3. Es gelte nun wieder die Prämisse von (b). Mit 1. und 2. und Aufgabe A.15(a) ist dann φ ein Morphismus von \mathcal{A} nach \mathcal{B} , denn für $s, t \in S^*$ gilt, jeweils nach der Definition in 1.7(b) im ersten Schritt, (5) im zweiten und im vierten, und (3) im dritten,

$$\varphi([s] \wedge [t]) = \varphi([\&st]) = g(\&st) = g(s) \wedge g(t) = \varphi([s]) \wedge \varphi([t])$$

und

$$\varphi([s]') = \varphi([Ns]) = g(Ns) = g(s)' = \varphi([s])'.$$

Zum Nachweis der Eindeutigkeit von φ sei schließlich $\psi : \mathcal{A} \rightarrow \mathcal{B}$ ein Morphismus mit $\psi([s]) = f(s)$ für $s \in S$. Wegen (6) und $S_1 = S$ gilt dann

$$(7) \quad \psi([s]) = \varphi([s])$$

für $s \in S_1$. Ist weiter $n \in \mathbb{N}$ und gilt (7) für jedes $s \in S_n$, so gilt für $s \in S_{n+}$ im Fall von $s \in S_n$ trivialerweise (7), im Fall von $s = Nt$ mit $t \in S_n$ unter Verwendung der Induktionsannahme im zentralen Schritt

$$\begin{aligned} \psi([s]) &= \psi([Nt]) = \psi([t]') = \psi([t])' \\ &= \varphi([t])' = \varphi([t]') = \varphi([Nt]) = \varphi([s]), \end{aligned}$$

und auch im verbleibenden Fall von $s = \&tu$ mit $t, u \in S_n$ analog wieder (7). Also gilt $\psi = \varphi$ auf $\{[s] : s \in S^*\} = \mathcal{A}$. □

Wegen Satz 1.11(a) definieren wir im Anschluss an die Diskussion hinter Satz 1.7 nun, von der konkreten Konstruktion von \mathcal{A} abstrahierend, und \mathcal{C} statt \mathcal{A} schreibend:

1.19 Definition. Eine **Vorstellung (einen Weltausschnitt betreffend)** ist eine Boole-Algebra⁸ \mathcal{C} . Die Elemente von \mathcal{C} heißen **(denkbare) Weltereignisse**. \square

Von einem “Weltereignis” der Erfahrungswelt⁹ denken wir uns, dass entweder es selbst oder sein (Boole-)Komplement “real eintritt”, auch wenn wir diese beiden Möglichkeiten mangels Beobachtungen bei weitem nicht in jedem Einzelfall unterscheiden können, und auch wenn nach Beobachtungen die Unterscheidung willkürliche Elemente enthält, etwa bei einer Mehrheits- oder Autoritätsentscheidung.¹⁰ Weiter denken wir uns mit zwei real eintretenden Ereignissen auch deren (Boole-) Schnitt als real eintretend, und umgekehrt. Dementsprechend definieren wir:

1.20 Definition. Es sei \mathcal{C} eine Vorstellung und es sei $\mathcal{R} \subseteq \mathcal{C}$ mit den Eigenschaften

- (1) $A, B \in \mathcal{R} \Leftrightarrow A \wedge B \in \mathcal{R}$ für $A, B \in \mathcal{C}$,
- (2) $A \in \mathcal{C} \Rightarrow$ entweder $A \in \mathcal{R}$ oder $A' \in \mathcal{R}$.

Dann heißt \mathcal{R} **Realität**, und $(\mathcal{C}, \mathcal{R})$ heißt **Boole-Welt**. Die Elemente von \mathcal{R} heißen **real**, die von $\mathcal{C} \setminus \mathcal{R}$ **irreal**. \square

Die Konklusion rechts in (2) kann offenbar auch geschrieben werden als

- (3) $A \in \mathcal{R} \Leftrightarrow A' \notin \mathcal{R}$.

1.21 Bemerkung. Es sei $(\mathcal{C}, \mathcal{R})$ eine Boole-Welt.

(a) Es ist O irreal und I real, denn nach 1.20(2) angewandt auf $A := I$ gilt

- (1) entweder $I \in \mathcal{R}$ oder $O \in \mathcal{R}$,

und im Fall $O \in \mathcal{R}$ ergäbe sich $O \wedge I = O \in \mathcal{R}$, also auch $I \in \mathcal{R}$ nach 1.20(1), im Widerspruch zu (1) (siehe p. xiv!); aus (1) ergibt sich also $I \in \mathcal{R}$ und $O \notin \mathcal{R}$.

(b) Für gemäß A.15(1) disjunkte $A, B \in \mathcal{C}$ gilt die Äquivalenz

$$A \vee B \in \mathcal{R} \Leftrightarrow \text{entweder } A \in \mathcal{R} \text{ oder } B \in \mathcal{R},$$

denn für beliebige $A, B \in \mathcal{C}$ gilt wegen 1.20(2,3) und de Morgan A.8(5) im ersten Schritt, 1.20(1) im zweiten, und 1.20(2) im dritten, die Äquivalenzkette

$$A \vee B \in \mathcal{R} \Leftrightarrow A' \wedge B' \notin \mathcal{R} \Leftrightarrow A' \notin \mathcal{R} \text{ oder } B' \notin \mathcal{R} \Leftrightarrow A \in \mathcal{R} \text{ oder } B \in \mathcal{R},$$

und im Fall disjunkter A, B würde “ $A \in \mathcal{R}$ und $B \in \mathcal{R}$ ” mit 1.20(1) auf $O = A \wedge B \in \mathcal{R}$ führen, im Widerspruch zu (a). \square

⁸ \mathcal{C} wie lateinisch *conceptio*, englisch *conception*, deutsch *Konzeption*, *Vorstellung*, *Auffassung*. Nicht \mathcal{A} , damit \mathcal{A} auch von uns später wie üblich verwendet werden kann.

⁹Die Wörter “Vorstellung”, “Weltereignis”, “Realität”, “real”, “irreal” werden einerseits in den Definitionen 1.19 und 1.20 als Namen mathematischer Begriffe eingeführt, sollen aber andererseits weiter beim Reden über diese Begriffe zu modellierende außermathematische Begriffe verwendet werden können; wir fügen daher nötigenfalls zur Unterscheidung einerseits “im mathematischen Sinne” beziehungsweise andererseits “in der Erfahrungswelt”, oder ähnliches, bei.

¹⁰Man denke etwa an eine Aussage beim Fußball wie: “Das war Abseits”.

1.22 Beispiel. (a) Es sei \mathcal{C} eine Mengenalgebra auf einer Grundmenge \mathcal{Z} , und es sei $r \in \mathcal{Z}$. Dann ist

$$\mathcal{R} := \{C \in \mathcal{C} : C \ni r\}$$

eine Realität, denn für $A, B \in \mathcal{C}$ gilt die Äquivalenzkette

$$A, B \in \mathcal{R} \Leftrightarrow r \in A, r \in B \Leftrightarrow r \in A \cap B \Leftrightarrow A \cap B \in \mathcal{R},$$

und für $A \in \mathcal{C}$ gilt auch entweder $r \in A$ oder $r \in A^c$, also entweder $A \in \mathcal{R}$ oder $A^c \in \mathcal{R}$.

(b) In Teil (a) könnte \mathcal{C} zum Beispiel die Potenzmengenalgebra auf

$$\mathcal{Z} := \mathcal{X}^{\mathbb{N}} \quad \text{mit } \mathcal{X} := \{25, \dots, 150\}$$

sein, mit der an das einleitende Beispiel von Seite 1 anknüpfenden Interpretation: Für $z \in \mathcal{Z}$ und $\nu \in \mathbb{N}$ ist z_ν das denkbare Todesalter in Jahren des ν -ten ab dem Jahre 2000 25-jährig gewordenen bzw. noch werdenden männlichen Deutschen, und ein gewisses $r \in \mathcal{Z}$ entspricht der Realität. Diese Realität ist etwa für einen Lebensversicherer von Interesse. Der fragt sich etwa, “wie viele” der denkbaren Weltereignisse $C_\nu := \{z \in \mathcal{Z} : z_\nu < 50\}$ real sind. Präziser formuliert, bei Beschränkung auf die ersten n der betrachteten Personen, interessiert der (ungefähre) Wert von

$$\sum_{\nu=1}^n (C_\nu \in \mathcal{R}) = \sum_{\nu=1}^n (r \in C_\nu) = \sum_{\nu=1}^n (r_\nu < 50),$$

wobei wir die auf p. xv erklärte Indikatorklammernotation von Iverson (1962) und de Finetti (1967) verwendet haben. \square

Die hier in Hinblick auf die von uns beabsichtigte Interpretation als “Realitäten” bezeichneten mathematischen Objekte heißen anderswo in der Mathematik “Ultrafilter”. “Filter” und Ultrafilter, siehe die Aufgaben A.23 und A.24, wurden von Henri Cartan (1937) im Spezialfall von Potenzmengenalgebren als Grundbegriffe einer allgemeinen Konvergenztheorie eingeführt, was ihre im hiesigen Kontext wenig intuitiv anmutenden Namen erklären mag.

Dass Beispiele wie 1.22(b), und praktisch alle im weiteren Verlauf konkret betrachteten, sich mit Mengen-Algebren hinschreiben lassen, erklärt sich aus den Darstellungssätzen A.16 und A.17, die jedoch hier nicht wirklich benutzt werden.

2 Mathematisierung des empirischen Wahrscheinlichkeitsbegriffes

2.1 Beispiel (Freitag, der Dreizehnte). Wie “wahrscheinlich” ist es, dass ein Monatsdreizehnter ein Freitag ist?

Zur Mathematisierung dieser Frage wählen wir als Vorstellung \mathcal{C} die Potenzmengenalgebra auf $\mathcal{Z} := \{1, \dots, 7\}^{\mathbb{N}}$. Dabei stehen die Zahlen von 1 bis 7 für die Wochentage von Montag bis Sonntag, und wir denken für ein $z \in \mathcal{Z}$ zum Beispiel an die Interpretation

$$z_\nu = 5 \hat{=} \text{“Der } \nu\text{-te Dreizehnte ist ein Freitag”},$$

ausgehend von einem fixierten Startmonat, etwa¹ 1883-04. Die Realität sollte dann $\mathcal{R} = \{C \in \mathcal{C} : C \ni r\}$ mit einem $r \in \mathcal{Z}$ definiert durch unseren Gregorianischen Kalender sein. Als Ad-hoc-Definition bietet sich nun²

$$(1) \quad \text{Gesuchte Wahrscheinlichkeit} := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\nu=1}^n (r_\nu = 5)$$

an. Diese untersuchen wir nun, vor der numerischen Beantwortung der Eingangsfrage, um zunächst zu einer allgemeinen Definition zu gelangen:

Mit

$$\begin{aligned} \mathcal{X} &:= \{1, \dots, 7\}, \\ \mathcal{A} &:= \text{Potenzmengenalgebra auf } \mathcal{X}, \\ \pi_\nu &:= \nu\text{-te Koordinatenprojektion von } \mathcal{Z} \text{ auf } \mathcal{X} \end{aligned}$$

ist

$$\varphi_\nu := \pi_\nu^{-1}[\cdot],$$

also die Urbildabbildung der ν -ten Koordinatenprojektion, ein Boole-Algebra-Morphismus von \mathcal{A} nach \mathcal{C} , siehe Aufgabe A.15(c), und mit $A := \{5\} \in \mathcal{A}$ gilt für $\nu \in \mathbb{N}$ die Äquivalenzkette

$$r_\nu = 5 \Leftrightarrow \pi_\nu(r) \in A \Leftrightarrow r \in \pi_\nu^{-1}[A] \Leftrightarrow \varphi_\nu(A) \in \mathcal{R};$$

damit gilt also

$$\text{R.S.}(1) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\nu=1}^n (\varphi_\nu(A) \in \mathcal{R}).$$

→ 2.7

¹Der Geburtsmonat von Richard von Mises.

²Wir erinnern an die auf p. xv eingeführte Indikatorklammernotation.

2.2 Definition. Es sei $(\mathcal{C}, \mathcal{R})$ eine Boole-Welt, \mathcal{A} eine Boole-Algebra, und $\varphi_\bullet = (\varphi_\nu)_{\nu \in \mathbb{N}}$ eine Folge von Boole-Morphismen $\varphi_\nu : \mathcal{A} \rightarrow \mathcal{C}$ derart, dass

$$(1) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\nu=1}^n (\varphi_\nu(A) \in \mathcal{R}) =: P(A) \quad \text{für } A \in \mathcal{A}$$

existiert. Dann heißt φ_\bullet ein **Quasikollektiv** in $(\mathcal{C}, \mathcal{R})$, mit der **Verteilung** P auf der **Merkmalsalgebra** \mathcal{A} . Die Elemente von \mathcal{A} heißen **Modellereignisse**. \square

Zur Formulierung allgemeiner Eigenschaften von Verteilungen von Quasikollektiven in Satz 2.6 definieren wir:

2.3 Definition. Es sei \mathcal{A} eine Boole-Algebra und es sei $P : \mathcal{A} \rightarrow [0, \infty]$ eine Funktion mit

- (1) $P(A) \in [0, \infty]$ für $A \in \mathcal{A}$ (**Positivität**),
- (2) $P(O) = 0$ (**Nulltreue**),
- (3) $P(A \vee B) = P(A) + P(B)$ für disjunkte $A, B \in \mathcal{A}$ (**(paarweise) Additivität**).

Dann heißt P **Inhalt**³ auf \mathcal{A} . Gilt noch

$$(4) \quad P(I) = 1 \quad \text{(Normiertheit),}$$

so heißt P **Wahrscheinlichkeitsinhalt** auf \mathcal{A} ; dieser heißt **nichtdegeneriert** falls es ein $A \in \mathcal{A}$ mit $0 < P(A) < 1$ gibt. \square

Bei gegebenem Wahrscheinlichkeitsinhalt P lesen wir einen Wert $P(A)$ auch als **Wahrscheinlichkeit** von A .

2.4 Bemerkung. Es sei \mathcal{A} eine Boole-Algebra und es sei $P : \mathcal{A} \rightarrow [0, \infty]$ eine Funktion.

(a) P sei ein Inhalt. Dann gilt

$$(1) \quad P(A) \leq P(B) \quad \text{für } A, B \in \mathcal{A} \text{ mit } A \leq B \quad \text{(Isotonie),}$$

wegen $B = BA \vee BA' = A \vee BA'$ und damit $P(B) = P(A \vee BA') = P(A) + P(BA') \geq P(A)$, sowie die **(endliche) Additivität**

$$(2) \quad P\left(\bigvee_{j \in J} A_j\right) = \sum_{j \in J} P(A_j) \quad \text{für } J \text{ endlich und } A_\bullet \text{ paarweise disjunkt}^4 \text{ in } \mathcal{A}$$

per Induktion über $\#J$.

³In einer systematischeren Darstellung des Inhaltsbegriffes - für die hier kein Platz ist - würde man die Voraussetzung " \mathcal{A} Boole-Algebra" so verallgemeinern, dass zum Beispiel auch der gewöhnliche Flächeninhalt von Vereinigungen endlich vieler Dreiecke in der Ebene ein Inhalt wäre. Idee: Ersetze die Komplementierung $'$ und die Eins I durch eine (Boole-)Differenz genannte binäre Verknüpfung \setminus mit Axiomen die im Mengenfall $\mathcal{A} \subseteq 2^\Omega$ von $(\mathcal{A}, \cap, \cup, \setminus, \emptyset)$ analog zu Beispiel A.3(b) erfüllt werden. zwei mögliche Präzisierungen zum Beispiel bei Grätzer (2011, pp. 143–144, "generalized Boolean lattice").

Dass die in der Maß- und Integrationstheorie übliche Betrachtung lediglich des Mengenfalls (dort "Ring" genannt, besser wäre "(Mengen)-Rng"), elementargeometrisch unbefriedigend ist, zeigt die Frage: Ist die Differenz zweier Dreiecke stets eine endliche Vereinigung von Dreiecken? Genaueres dazu auf den ersten Seite von Caratheodory (1956); dort heißen die verallgemeinerten Boole-Algebren "Somen-Ringe".

(b) P ist schon dann ein Wahrscheinlichkeitsinhalt wenn die Bedingungen 2.3(1,3,4) gelten, denn dann folgt $1 = P(I) = P(I \vee O) = P(I) + P(O) = 1 + P(O)$ und damit 2.3(2). Ist P ein Wahrscheinlichkeitsinhalt, so gilt

$$(3) \quad P(A) \in [0, 1] \quad \text{für } A \in \mathcal{A},$$

wegen (1) mit $B := I$. □

Die folgenden einfachen Konstruktionen 2.5 von Wahrscheinlichkeitsinhalten werden unten im Beweis von Satz 2.6 benutzt.

2.5 Aufgabe. Es sei jeweils \mathcal{A} eine Boole-Algebra.

(a) Es sei \mathcal{R} eine Realität auf \mathcal{A} . Dann definiert

$$P(A) := (A \in \mathcal{R}) \quad \text{für } A \in \mathcal{A}$$

einen (degenerierten) Wahrscheinlichkeitsinhalt P auf \mathcal{A} .

(b) Es sei $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ ein Boole-Algebra-Morphismus, und Q ein Wahrscheinlichkeitsinhalt auf \mathcal{B} . Dann ist $Q \circ \varphi$ ein Wahrscheinlichkeitsinhalt auf \mathcal{A} .

(c) Es sei J eine Menge⁵, $\lambda_\bullet \in [0, 1]^J$ eine **Wahrscheinlichkeitszähldichte** auf J (d.h. es gelte noch $\sum_{j \in J} \lambda_j = 1$), und $P_\bullet = (P_j : j \in J)$ eine Familie von Wahrscheinlichkeitsinhalten auf \mathcal{A} . Dann ist die **Konvexkombination** $\sum_{j \in J} \lambda_j P_j$, definiert durch

$$\left(\sum_{j \in J} \lambda_j P_j \right)(A) := \sum_{j \in J} \lambda_j P_j(A) \quad \text{für } A \in \mathcal{A},$$

ein Wahrscheinlichkeitsinhalt auf \mathcal{A} .

(d) Es sei $(P_n)_{n \in \mathbb{N}}$ eine Folge von Wahrscheinlichkeitsinhalten auf \mathcal{A} derart, dass $P := \lim_{n \rightarrow \infty} P_n$ punktweise existiert, d.h. $P(A) = \lim_{n \rightarrow \infty} P_n(A)$ für $A \in \mathcal{A}$. Dann ist P ein Wahrscheinlichkeitsinhalt auf \mathcal{A} . □

2.6 Satz. *In der Situation von Definition 2.2 ist P ein Wahrscheinlichkeitsinhalt auf \mathcal{A} .*

Beweis. Für $\nu \in \mathbb{N}$ definiert

$$Q_\nu(A) := (\varphi_\nu(A) \in \mathcal{R}) = \left((\cdot \in \mathcal{R}) \circ \varphi_\nu \right)(A) \quad \text{für } A \in \mathcal{A}$$

⁴Eine Familie $A_\bullet = (A_j)_{j \in J}$ in einer Boole-Algebra heißt nach Aufgabe A.15(h) **paarweise disjunkt**, abgekürzt **p.d.**, wenn $A_j \wedge A_k = O$ für $(j, k) \in J_\neq^2$ gilt. Zur Notation $\bigvee_{j \in J} A_j$ siehe A.9(c).

⁵Diese sei hier noch als abzählbar vorausgesetzt. Im Beweis von Satz 2.6 wird nur der Fall einer endlichen Menge J benutzt. Dass die Aussage tatsächlich bei beliebiger Mächtigkeit von J gilt, sehen wir später nach Definition B.9 aus Anhang B mittels Satz B.10.

einen Wahrscheinlichkeitsinhalt Q_ν auf \mathcal{A} , wegen 2.5(a) mit \mathcal{C} statt \mathcal{A} , und 2.5(b) mit \mathcal{C} statt \mathcal{B} . Damit ist aber auch $P_n := \frac{1}{n} \sum_{\nu=1}^n Q_\nu$ für jedes $n \in \mathbb{N}$ ein Wahrscheinlichkeitsinhalt auf \mathcal{A} , wegen 2.5(c) mit $I = \{1, \dots, n\}$ und $\lambda_\nu = \frac{1}{n}$, und damit ist dies nach der Grenzwertexistenzannahme in Definition 2.2 und nach 2.5(d) auch $P = \lim_{n \rightarrow \infty} P_n$. \square

Nun zurück zu unserem Beispiel 2.1. Die Berechnung der Zahlenwerte überlassen wir Ihnen mit der folgenden **R**-Aufgabe.

2.7 Aufgabe (2.1 ← Freitag der Dreizehnte). Unser Kalender hat eine Periodizität von vierhundert Jahren. Mithilfe der Programmiersprache **R** lässt sich die Anzahl der Freitage unter den Monatsdreizehnten eines Zyklus im Prinzip sehr leicht berechnen, und dies sollen Sie hier tun.

- (a) Wie wahrscheinlich ist es, dass ein Monatsdreizehnter ein Freitag ist?
 (b) Welcher Wochentag ist als Monatsdreizehnter am wahrscheinlichsten?

Anleitung:

- **R** ist frei verfügbar und an allen Universitätsrechnern vorinstalliert. Bei den Rechnern der Mathematik-Cip-Pools können Sie einfach **R** in die Konsole eingeben und diese dann als **R**-Konsole verwenden. Zum Beispiel ist dann die Verwendung als Taschenrechner möglich. Mit `?` oder `help` können sie Hilfe zu den meisten **R** Funktionen erhalten. Etwa mit `?Date` oder `help(ops.Date)`.
- Zum Datum 2018-12-24 erzeugt die Eingabe von `x<-as.Date("2018-12-24")` ein Objekt `x` der Klasse `Date`, welches nette Eigenschaften hat. So lässt sich durch `weekdays(x)` der Wochentag von `x` erfragen. Letzteres funktioniert auch wenn `x` ein Vektor von Daten ist.
- Mit der Funktion `seq` (beschrieben in `?seq.Date`) lassen sich ganze Vektoren von Daten erzeugen, wie zum Beispiel ein `y` der Form `(2018-12-24, 2019-01-24, ..., 2118-12-24)`.
- Durch `sum(weekdays(y)=="Montag")` lässt sich die Anzahl der Montage in `y` bestimmen. Dabei liefert `y=="Montag"` einen Vektor gleicher Länge wie `y`, mit *i*-tem Eintrag `TRUE` wenn der *i*-te Eintrag von `y` ein Montag ist, und `FALSE` sonst. `sum` summiert dann über diesen Vektor, was gleichbedeutend dazu ist die `TRUE`-Einträge zu zählen. Vorsicht: `Montag` muss in der von **R** auf Ihrem System verwendeten Sprache (wohl Deutsch oder Englisch) geschrieben werden.
- Die Aufgabe lässt sich mit den hier genannten Funktion in sehr wenigen Programmierzeilen lösen (nur ein bis zwei für einen festen Wochentag).
- Schreiben Sie Ihre funktionierenden Kommandos in einen Texteditor. Für die Abgabe können Sie die Ergebnisse dazu kopieren, das ganze (höchstens eine Seite!) ausdrucken, und eventuell noch Erläuterungen anbringen. → 2.8

2.8 Beispiel (2.7 ← Freitag der Dreizehnte). In der Situation aus Beispiel 2.1 liefern Definition 2.2 und Aufgabe 2.7

k	1	2	3	4	5	6	7
$4800P(\{k\})$	685	685	687	684	688	684	687

(im endlichen exakt für $n \in 4800\mathbb{N}$), also kurioserweise $P(\{k\})$ größt für $k = 5$, mit

$$p := P(\{5\}) = \frac{688}{4800} = \frac{43}{300} = \frac{1}{7} \cdot \frac{301}{300} = \frac{1}{7} \cdot \left(1 + \frac{1}{3}\%\right) > \frac{1}{7}.$$

Die empirische Bedeutung von p ist nun:

- (1) In jeder langen Folge konsekutiver Monate ist der Anteil der Freitage unter den Monatsdreizehnten ungefähr p , also ein drittel Prozent mehr als ein Siebtel.

Diese kann jedoch bei Abweichungen von der oben geforderten Auswahl *aller* Monate verloren gehen; wir geben dafür zwei Beispiele:

(a) Mit “Januare” statt “Monate” ergibt sich statt p ein q , für welches mit einem $\ell \in \mathbb{N}$

$$q = \frac{\ell}{4800/12} \neq \frac{688/12}{4800/12} = p$$

gilt, wegen $688/12 \notin \mathbb{N}$ im zentralen Schritt, also $q \neq p$.

(b) Mit “Monate, die auf zwei Monate mit je einem Freitag den Dreizehnten folgen,” statt “Monate” ergibt sich 0 statt p . □

Für mehr Mathematik und zur Geschichte unseres Kalenders siehe Rickey (1985).

Obiges Beispiel ist im Grunde rein innermathematisch, da eine mathematisch wohldefinierte unendliche Folge $r \in \{1, \dots, 7\}^{\mathbb{N}}$ betrachtet wurde. In Beispielen aus der Erfahrungswelt kommen dagegen nur endliche Folgen vor, die man sich allenfalls als unbegrenzt fortsetzbar denken mag, und den präzisen Grenzwertbildungen 2.2(1) entspricht die Berechnung relativer Häufigkeiten mit einem festen, vorzugsweise großen, aber eigentlich immer recht willkürlich gewählten n :

2.9 Beispiel (Geschlechterverhältnis Neugeborener). Dass ungefähr gleich viele Jungen wie Mädchen geboren werden, muss seit Urzeiten in vielen Kulturen als offensichtliche Erfahrungstatsache gegolten haben. Die frühesten uns heute noch bekannten Daten zur Untersuchung dieser Frage stammen jedoch erst aus den Jahren 1336–1338 vom damaligen Priester der Taufkirche San Giovanni des Doms von Florenz, laut Pearson (1978, pp. 7–8): In den genannten drei Jahren fanden jährlich rund 5750 Taufen pro Jahr statt, mit rund 3075 männlichen und 2675 weiblichen Täuflingen; der Anteil der Jungen betrug also mit rund $\frac{3075}{5750} = 0.535$ doch etwas mehr als $\frac{1}{2}$.

Genauere und umfangreicherer Daten finden sich bei Graunt (1662), welcher die jährlichen Taufzahlen von Jungen und von Mädchen in London von 1629 bis 1660 und in Romsey von 1569 bis 1658 angibt - und ebenfalls einen leichten Jungenüberschuss feststellt, siehe Hald (1990, p. 93).

Arbuthnott (1712) präsentiert und untersucht noch umfangreichere Daten: Er betrachtet für jedes der 82 Jahre von 1629 bis 1710 die Anzahlen der in London getauften Jungen und Mädchen. Mit $i \hat{=} \text{Jahr} - 1628 \in \{1, \dots, 82\}$ gibt in der Tabelle auf der nächsten Seite n_i die Gesamttaufzahl des i -ten Jahres an, x_i die Jungentaufzahl, also $n_i - x_i$ die Mädchentaufzahl und x_i/n_i den Anteil der Jungen, gerundet auf drei Nachkommastellen. Da dieser Anteil in jedem der 82 Jahre größer als $\frac{1}{2}$ ist, schließt Arbuthnott auf göttliche Vorsehung zwecks Ausgleichs der “external Accidents to which are Males subject (who must seek their Food with danger)”.

Als Anteil der Jungen über den Gesamtzeitraum erhalten wir

$$(1) \quad p := \frac{\sum_{i=1}^{82} x_i}{\sum_{i=1}^{82} n_i} = 0.516$$

Mit $n := \sum_{i=1}^{82} n_i = 938\,223$ ist

$$(2) \quad p = \frac{1}{n} \sum_{\nu=1}^n (r_\nu = 1)$$

mit $r \in \{0, 1\}^n$,

$$r_\nu := \begin{cases} 1 \\ 0 \end{cases} \text{ falls } \nu\text{-ter Täufling } \begin{cases} \text{männlich} \\ \text{weiblich} \end{cases},$$

wobei von r leider nur die Summen

$$x_i = \sum_{\nu=1+\sum_{j=1}^{i-1} n_j}^{\sum_{j=1}^i n_j} r_\nu$$

überliefert sind.

Definition (1) ist wie 2.1(1), nur ohne Limesbildung, welche in der Erfahrungswelt nun mal

i	x_i	$n_i - x_i$	n_i	x_i/n_i
1	5218	4683	9901	0.527
2	4858	4457	9315	0.522
3	4422	4102	8524	0.519
4	4994	4590	9584	0.521
5	5158	4839	9997	0.516
6	5035	4820	9855	0.511
7	5106	4928	10034	0.509
8	4917	4605	9522	0.516
9	4703	4457	9160	0.513
10	5359	4952	10311	0.520
11	5366	4784	10150	0.529
12	5518	5332	10850	0.509
13	5470	5200	10670	0.513
14	5460	4910	10370	0.527
15	4793	4617	9410	0.509
16	4107	3997	8104	0.507
17	4047	3919	7966	0.508
18	3768	3395	7163	0.526
19	3796	3536	7332	0.518
20	3363	3181	6544	0.514
21	3079	2746	5825	0.529
22	2890	2722	5612	0.515
23	3231	2840	6071	0.532
24	3220	2908	6128	0.525
25	3196	2959	6155	0.519
26	3441	3179	6620	0.520
27	3655	3349	7004	0.522
28	3668	3382	7050	0.520
29	3396	3289	6685	0.508
30	3157	3013	6170	0.512
31	3209	2781	5990	0.536
32	3724	3247	6971	0.534
33	4748	4107	8855	0.536
34	5216	4803	10019	0.521
35	5411	4881	10292	0.526
36	6041	5681	11722	0.515
37	5114	4858	9972	0.513
38	4678	4319	8997	0.520
39	5616	5322	10938	0.513
40	6073	5560	11633	0.522
41	6506	5829	12335	0.527
42	6278	5719	11997	0.523
43	6449	6061	12510	0.516
44	6443	6120	12563	0.513
45	6073	5822	11895	0.511
46	6113	5738	11851	0.516
47	6058	5717	11775	0.514
48	6552	5847	12399	0.528
49	6423	6203	12626	0.509
50	6568	6033	12601	0.521
51	6247	6041	12288	0.508
52	6548	6299	12847	0.510
53	6822	6533	13355	0.511
54	6909	6744	13653	0.506
55	7577	7158	14735	0.514
56	7575	7127	14702	0.515
57	7484	7246	14730	0.508
58	7575	7119	14694	0.516
59	7737	7214	14951	0.517
60	7487	7101	14588	0.513
61	7604	7167	14771	0.515
62	7909	7302	15211	0.520
63	7662	7392	15054	0.509
64	7602	7316	14918	0.510
65	7676	7483	15159	0.506
66	6985	6647	13632	0.512
67	7263	6713	13976	0.520
68	7632	7229	14861	0.514
69	8062	7767	15829	0.509
70	8426	7626	16052	0.525
71	7911	7452	15363	0.515
72	7578	7061	14639	0.518
73	8102	7514	15616	0.519
74	8031	7656	15687	0.512
75	7765	7683	15448	0.503
76	6113	5738	11851	0.516
77	8366	7779	16145	0.518
78	7952	7417	15369	0.517
79	8379	7687	16066	0.522
80	8239	7623	15862	0.519
81	7840	7380	15220	0.515
82	7640	7288	14928	0.512

nicht vorkommt⁶. Letzteres ignorierend erhalten wir, ansonsten völlig analog zu Beispiel 2.1, mit $\mathcal{X} := \{0, 1\}$, $\mathcal{Z} := \mathcal{X}^n$, $\mathcal{C} := 2^{\mathcal{Z}}$, $\mathcal{R} := \{C \in \mathcal{C} : C \ni r\}$, $\pi_\nu := \nu$ -te Koordinatenprojektion von \mathcal{Z} auf \mathcal{X} , $\varphi_\nu := \pi_\nu^{-1}[\cdot]$ für $A := \{1\}$ dann

$$\text{R.S.}(2) = \frac{1}{n} \sum_{\nu=1}^n (\varphi_\nu(A) \in \mathcal{R}).$$

Interpretation von p :

- (3) In jeder langen Folge konsekutiver Taufen ist (zumindest in London um das Jahr 1670) der Anteil der Jungen unter den Täuflingen ungefähr 0.516.

Bleibt nun hier vielleicht, anders als in 2.8(a,b), diese Interpretation gültig wenn die Konsekutivitätsforderung weggelassen wird? Wenn wir etwa aus Arbuthnots Daten nur die Jahre auswählen, welche auf ein Jahr mit einem überdurchschnittlich hohen Jungenanteil, sagen wir ≥ 0.520 , folgen, erhalten wir dann einen Jungenanteil q von wieder ungefähr $p = 0.516$, oder müssten wir vielleicht zum Ausgleich einen spürbar kleineren Anteil erhalten? Die Rechnung ergibt

$$(4) \quad q = \frac{\sum_{i=2}^{82} x_i \cdot \left(\frac{x_{i-1}}{n_{i-1}} \geq 0.520\right)}{\sum_{i=2}^{82} n_i \cdot \left(\frac{x_{i-1}}{n_{i-1}} \geq 0.520\right)} = 0.518,$$

also tatsächlich ungefähr p .

(Die Daten und die bei der Berechnung von p und q verwendeten R-Kommandos stehen in der auf StudIP hochgeladenen Datei `Arbuthnott-Daten.txt`.) → 2.20

2.10 Aufgabe (Kreiselnde Münzen). Wie wahrscheinlich ist es, dass eine kreiselnde Münze mit “Zahl” nach oben zur Ruhe kommt?

Fixieren Sie ein Eurostück und lassen Sie es auf einen glatten Untergrund, z.B. einer Küchentischplatte, 100 mal (häufiger wäre eigentlich noch besser) hintereinander auf dem Münzrand schnell und ungestört kreiseln. In aller Regel wird die Münze bei jedem der 100 Versuche nach einer Weile auf einer Seite zur Ruhe kommen. In diesem Fall notieren Sie eine 1, falls “Zahl” oben liegt, und eine 0, falls “Zahl” unten liegt.⁷ Sollte ausnahmsweise die Münze auf dem Rand stehen bleiben oder vom Tisch fallen, werde der jeweilige Versuch solange wiederholt, bis die Münze das erste Mal auf der Seite liegen bleibt. Wie groß ist der Anteil der Einsen

- (a) bei allen Versuchen,
- (b) bei den Versuchen mit ungerader Nummer,
- (c) bei den Versuchen mit gerader Nummer,

⁶Analog ist eine Geschwindigkeit in der Mathematischen Physik eine Ableitung, also ein Limes von Differenzenquotienten, einer Bahnkurve, in der Experimentalphysik dagegen ein Differenzenquotient von Orts- und Zeitmessungen, mit vorzugsweise kleinem Zeitinkrement.

⁷Liste aufheben und immer mal wieder anschauen!

(d) bei den Versuchen, welche auf ein Ergebnis “Eins” folgen?

Bei dieser Aufgabe ist wie üblich eine Gruppenabgabe basierend auf einer Ergebnisfolge erlaubt, jedoch muss für volle Punktzahl jedes Gruppenmitglied seine eigene Ergebnisfolge in einer txt-Datei wie folgt hochladen:

- Schreiben Sie die 100 Nullen und Einsen untereinander in die txt-Datei (also 100 Zeilen mit nur je einem Zeichen); sonst darf die Datei nichts enthalten.
- Nennen Sie die Datei `KM2024_1234567.txt`, wobei `1234567` durch Ihre (in der Regel siebenstellige) Matrikelnummer zu ersetzen ist.
- Die Datei ist in den Ordner “Hausarbeiten” zur Vorlesung hochzuladen.

Können Sie die leicht von Hand ausführbaren Rechnungen auch mit **R** erledigen?

Hinweise:

- Eingabe von `getwd()` in R sagt Ihnen, wo R Ihre Ergebnisdatei sucht. (Dies könnten Sie auch mit `setwd(...)` ändern.)
- Mit `KM <- read.table("KM2024_1234567.txt")` haben Sie die Zahlen dann in R als sogenanntes data frame `KM` zur Verfügung.
- Data frames sind zur Bearbeitung komplizierterer, mehr als einspaltiger Tabellen gedacht. Um nun hier die einzige Spalte als Vektor `r` zur weiteren Berechnung zur Verfügung zu haben, kann man `r <- KM$V1` verwenden, dabei ist `V1` der Default-Name der ersten Spalte (wobei das `V` für “Variable” steht).
- Im Detail sind viele Lösungen möglich. Was liefern z.B. sukzessive die Kommandos `x<c(0,1,0,1,1,0,1,0,0,1)`, `x`, `sum(x)`, `length(x)`, `mean(x)`, `y<-x[2:10]`, `g<-2*(1:5)`, `x[g-1]`, `x*y`, `z<-c(y,0)` ? □

Die Verteilung eines Quasikollektivs gemäß Definition 2.2 liefert Vorhersagen über den durch die betrachtete Boole-Welt modellierten Ausschnitt der Erfahrungswelt, wie

- in Beispiel 2.8 die Aussage 2.8(1),
- in Beispiel 2.9 die Aussage 2.9(3),
- in Aufgabe 2.10 vielleicht: In jeder langen Folge konsekutiver Münzkreiselungen (mit dem von Herrn Mattner verwendeten irischen Eurostück) ist der Anteil von “Zahl oben” ungefähr 0.4.

Jedoch unterscheiden sich diese Beispiele wohl bezüglich der Bedeutung der jeweiligen Konsekutivitätsvoraussetzung: In Beispiel 2.8 ist sie wesentlich, wie 2.8(b) besonders drastisch zeigt, in Beispiel 2.9 und Aufgabe 2.10 vielleicht nicht. Die nun angesteuerte Verschärfung 2.12 der Definition 2.2 eliminiert Effekte wie in Beispiel

2.8(a,b) und soll das vielleicht von Ihnen in Aufgabe 2.10 beobachtete Verhalten widerspiegeln.

Grob gesprochen fordert Definition 2.12 über 2.2 hinausgehend, dass sich die Limeses in 2.2(1) nicht ändern, wenn nach einer “nichthellseherischen Regel” nur manche, aber immer noch unendlich viele, der dortigen Indikatoren zur Berechnung der relativen Häufigkeiten verwendet werden, d.h. wenn aus 2.12(1) auch 2.12(2) für nach einer besagten Regel berechnete $s \in \{0, 1\}^{\mathbb{N}}$ folgt. Dabei wird “nichthellseherisch” dahingehend präzisiert, dass s_ν als Funktion von ν und den Indikatoren $(\varphi_i(B) \in \mathcal{R})$ mit $i \in \{1, \dots, \nu - 1\}$ und $B \in \mathcal{A}$ berechnet werden können muss. Zwei triviale Beispiele, gar nicht von den genannten Indikatoren abhängig:

$$s_\nu := 1 \quad \text{für jedes } \nu \in \mathbb{N},$$

$$s_\nu := \begin{cases} 1 \\ 0 \end{cases} \quad \text{falls } \nu \begin{cases} \text{ungerade} \\ \text{gerade} \end{cases}.$$

Ein einfaches nichttriviales Beispiel, 2.10(d) idealisierend: $\mathcal{A} :=$ Potenzmengenalgebra auf $\{0, 1\}$, $B_0 := \{1\}$, $s_1 := 0$, und

$$s_\nu := (\varphi_{\nu-1}(B_0) \in \mathcal{R}) \quad \text{für } \nu \geq 2.$$

Die allgemeine Formalisierung dieser Idee ist naheliegend, aber leider nicht ganz einfach lesbar:

2.11 Definition. Es sei \mathcal{A} eine Boole-Algebra, und es sei

$$S : \bigcup_{\nu \in \mathbb{N}} \{0, 1\}^{\{1, \dots, \nu-1\} \times \mathcal{A}} \rightarrow \{0, 1\}$$

eine Funktion. Dann heißt S **\mathcal{A} -Auswahlregel**, und für $\nu \in \mathbb{N}$ wird ihre Restriktion auf $\{0, 1\}^{\{1, \dots, \nu-1\} \times \mathcal{A}}$ mit S_ν bezeichnet und heißt ν -ter **Inklusionsindikator** von S .

Ist weiter $(\mathcal{C}, \mathcal{R})$ eine Boole-Welt und $\varphi_\bullet = (\varphi_\nu)_{\nu \in \mathbb{N}}$ eine Folge von Boole-Morphismen $\varphi_\nu : \mathcal{A} \rightarrow \mathcal{C}$, so heißt für $\nu \in \mathbb{N}$

$$(1) \quad s_\nu := s_{\nu, \mathcal{A}, \mathcal{C}, \mathcal{R}, \varphi_\bullet, S} := S_\nu \left((\varphi_i(B) \in \mathcal{R}) : i \in \{1, \dots, \nu - 1\}, B \in \mathcal{A} \right)$$

die ν -te **Inklusionsindikation** von S für φ_\bullet und \mathcal{R} . □

In der obigen Definition ist S_1 konstant, denn der Definitionsbereich von S_1 , also

$$\{0, 1\}^{\{1, \dots, 0\} \times \mathcal{A}} = \{0, 1\}^{\emptyset \times \mathcal{A}} = \{0, 1\}^\emptyset = \{\emptyset\},$$

ist einelementig. Weiter sind die Definitionsbereiche der S_ν paarweise disjunkt, so dass eine Auswahlregel S also nach Angabe beliebiger $S_\nu : \{0, 1\}^{\{1, \dots, \nu-1\} \times \mathcal{A}} \rightarrow \{0, 1\}$ dann durch $S := \bigcup_{\nu \in \mathbb{N}} S_\nu$ definiert ist.

2.12 Definition. Es sei φ_\bullet ein Quasikollektiv in $(\mathcal{C}, \mathcal{R})$ mit der Verteilung P auf \mathcal{A} , und es sei \mathcal{S} eine Menge von \mathcal{A} -Auswahlregeln. Für jedes $S \in \mathcal{S}$ gelte mit s_ν gemäß 2.11(1) für $\nu \in \mathbb{N}$ unter der **Nichtabbruchsbedingung**

$$(1) \quad \lim_{n \rightarrow \infty} \sum_{\nu=1}^n s_\nu = \infty$$

auch

$$(2) \quad \lim_{n \rightarrow \infty} \frac{\sum_{\nu=1}^n s_\nu \cdot (\varphi_\nu(A) \in \mathcal{R})}{\sum_{\nu=1}^n s_\nu} = P(A) \quad \text{für } A \in \mathcal{A}.$$

Dann heißt φ_\bullet ein **\mathcal{S} -Kollektiv** in $(\mathcal{C}, \mathcal{R})$, mit der Verteilung P . □

2.13 Beispiele. (a) Quasikollektive. Es sei φ_\bullet ein Quasikollektiv in $(\mathcal{C}, \mathcal{R})$, mit der Verteilung P auf \mathcal{A} . Dann ist die konstante Funktion 1 auf $\bigcup_{\nu \in \mathbb{N}} \{0, 1\}^{\{1, \dots, \nu-1\} \times \mathcal{A}}$ eine Auswahlregel, und mit $\mathcal{S} := \{1\}$ ist φ_\bullet ein \mathcal{S} -Kollektiv, da für $S = 1$ die Limesbedingung 2.12(2) schon durch 2.2(1) gegeben ist.

(b) Kreiselmünzen idealisiert. Analog zur Formalisierung in Beispiel 2.1 sei $\mathcal{X} := \{0, 1\}$, $\mathcal{A} :=$ Potenzmengenalgebra auf \mathcal{X} , $\mathcal{C} :=$ Potenzmengenalgebra auf $\mathcal{X}^{\mathbb{N}}$,

$$(1) \quad r \text{ “} := \text{Ergebnis einer unendlichen Folge von Kreiselmünzen”} \in \mathcal{X}^{\mathbb{N}},$$

$\mathcal{R} := \{C \in \mathcal{C} : C \ni r\}$, und $\varphi_\nu := \pi_\nu^{-1}[\cdot] : \mathcal{A} \rightarrow \mathcal{C}$ die Urbildbildung der ν -ten Koordinatenprojektion für $\nu \in \mathbb{N}$. Dann ist “nach Aufgabe 2.10(a)” vielleicht⁸ φ_\bullet ein Quasikollektiv in $(\mathcal{C}, \mathcal{R})$ mit einer Verteilung P auf \mathcal{A} , bestimmt durch ein $p \in [0, 1]$ mit $P(\{1\}) = p$.

Weiter ist mit den Auswahlregeln S, T, U , definiert durch

$$\begin{aligned} S_\nu(f) &:= (\nu \text{ ungerade}), \\ T_\nu(f) &:= (\nu \text{ gerade}), \\ U_\nu(f) &:= \begin{cases} 0 & \text{falls } \nu = 1, \\ f(\nu - 1, \{1\}) & \text{falls } \nu \geq 2 \end{cases} \end{aligned}$$

für $\nu \in \mathbb{N}$ und $f \in \{0, 1\}^{\{1, \dots, \nu-1\} \times \mathcal{A}}$ dann φ_\bullet “nach 2.10(b-d)” vielleicht sogar ein $\{S, T, U\}$ -Kollektiv; hier ist gemäß 2.11(1) für $\nu \in \mathbb{N}$

$$(2) \quad \begin{aligned} s_\nu &= (\nu \text{ ungerade}), \\ t_\nu &= (\nu \text{ gerade}), \\ u_\nu &= \begin{cases} 0 & \text{falls } \nu = 1, \\ (\varphi_{\nu-1}(\{1\}) \in \mathcal{R}) = (r_{\nu-1} = 1) & \text{falls } \nu \geq 2; \end{cases} \end{aligned}$$

⁸Rein mathematisch gesehen ist von einem un spezifizierten $r \in \{0, 1\}^{\mathbb{N}}$ die Rede, welches die im folgenden betrachteten Eigenschaften haben könnte oder auch nicht; der in Anführungsstriche gesetzte Teil von (1) beschreibt lediglich unsere außermathematische Interpretation.

dabei ist die Rechnung zur letzten Gleichung (2) im Fall $\nu \geq 2$ ausführlicher

$$\begin{aligned} u_\nu &= U_\nu\left((\varphi_i(B) \in \mathcal{R}) : i \in \{1, \dots, \nu - 1\}, B \in \mathcal{A}\right) \\ &= (\varphi_{\nu-1}(\{1\}) \in \mathcal{R}) \\ &= (\pi_{\nu-1}^{-1}[\{1\}] \ni r) = (\pi_{\nu-1}(r) \in \{1\}) = (r_{\nu-1} = 1). \quad \square \end{aligned}$$

Allgemeine Interpretation von 2.11(1) und 2.12(2): $s_\nu = 1$ bedeutet, aufgrund des Eintretens oder Nichteintretens gewisser mittels der $\varphi_1, \dots, \varphi_{\nu-1}$ beschreibbarer Weltereignisse zu entscheiden, dass die “ φ_ν -Ereignisse”⁹ in der Mittelbildung in 2.12(2) berücksichtigt werden. Dabei hängt im Beispiel 2.13(b) im Fall der Auswahlregeln S und T die Entscheidung von keinem der mittels $\varphi_1, \dots, \varphi_{\nu-1}$ beschreibbaren Weltereignisse ab, und im Fall von U nur von $\varphi_{\nu-1}(\{1\})$.

2.14 Aufgabe (Nullen und Einsen von RANDOM.ORG). Verschaffen Sie sich eine Folge r von $N := 10\,000$ “zufälligen” Nullen und Einsen, indem Sie entweder eine Münze N mal werfen oder den auf <https://www.random.org/integers/> zur Verfügung stehenden Random Integer Generator nutzen. Verhält sich Ihre Folge wie ein Anfangsstück eines Kollektivs?

Berechnen Sie dazu mit **R** näherungsweise die entsprechenden Grenzwerte relativer Häufigkeiten für die Auswahlregeln “Alle”, “Jede zehnte Zahl” (10 Startwerte!), “Jede mit Quadratzahlindex”, sowie “Jede nach (mindestens) k konsekutiven Nullen” (für $k \in \{1, 2, 3, 5\}$).

Anleitung: Um die vom Random Integer Generator produzierten Zahlen nach **R** zu transportieren, kann man diese etwa in einer langen Spalte in einen Textfile `Richard` kopieren, und hat die Zahlen dann mit `R <- read.table("Richard")` in **R** als sogenanntes data frame **R** zur Verfügung. Wie schon in Aufgabe 2.10 bemerkt, liefert nun `r <- R$V1` die einzige Spalte von **R** als Vektor **r**. □

2.15 Aufgabe. Formalisieren Sie die Situation aus Aufgabe 2.14 analog zu Beispiel 2.13(b). □

Verteilungen von \mathcal{S} -Kollektiven haben eine umso größere empirische Bedeutung, je größer \mathcal{S} ist. Siehe etwa die recht eingeschränkte empirische Bedeutung des Quasikollektivs, also $\{1\}$ -Kollektivs, aus Beispiel 2.8. Für die WT, im Sinne der Einleitung dieses Skriptums, ist zunächst jedoch nur die in Satz 2.6 festgestellte Wahrscheinlichkeitsinhaltseigenschaft einer Verteilung eines Quasikollektivs φ_\bullet relevant.

Dennoch stellt sich zu einem gegebenen Wahrscheinlichkeitsinhalt P die Frage: Gibt es \mathcal{S} -Kollektive mit der Verteilung P und großen \mathcal{S} ? Und gibt es überhaupt Quasikollektive mit der Verteilung P ?

⁹Wie liegen diese, also die $\varphi_\nu(A)$ mit $A \in \mathcal{A}$, in \mathcal{C} ? Siehe dazu Aufgabe A.20(c).

2.16 Satz (Widerspruchsfreiheitssatz von Feller¹⁰, 1939). *Es sei P ein Wahrscheinlichkeitsinhalt auf einer Boole-Algebra \mathcal{A} . Es gebe eine abzählbare Menge $\mathcal{A}_0 \subseteq \mathcal{A}$ mit*

$$(1) \quad P(A) = \sup_{B \in \mathcal{A}_0, B \leq A} P(B) \quad \text{für } A \in \mathcal{A}.$$

Dann gibt es zu jeder abzählbaren Menge \mathcal{S} von \mathcal{A} -Auswahlregeln ein \mathcal{S} -Kollektiv mit der Verteilung P .

Beweis. Siehe Feller (1939).¹¹ □

In der Regularitätsvoraussetzung (1) bezeichnet natürlich “ \leq ” die in Aufgabe A.13 eingeführte Ordnungsrelation auf der Boole-Algebra \mathcal{A} , im Spezialfall einer Mengenalgebra also die Teilmengenrelation. Diese Voraussetzung ist kaum einschränkend: Sie ist einerseits für die ab Definition 3.2 in diesen Vorlesungen fast ausschließlich betrachteten elementaren Wahrscheinlichkeitsmaße, und diese sind nach Satz 3.3 auch Wahrscheinlichkeitsinhalte auf einer Potenzmengenalgebra gemäß Definition 2.3, stets erfüllt, siehe Satz 3.16, also beispielsweise für die speziellen Verteilungen aus 3.12. Andererseits ist sie auch für die allgemeineren Wahrscheinlichkeitsmaße der MIT und WT gemäß Bemerkung 12.21 bei Beschränkung auf geeignete Definitionsalgebren erfüllt, wie in der folgenden Aufgabe exemplarisch präzisiert:

2.17 Aufgabe. Analog zu Beispiel A.3(c) ist offenbar

$$\mathcal{A} := \{A \subseteq [0, 1] : A \text{ endliche Vereinigung von Intervallen}\}$$

eine Mengenalgebra. Für ein Intervall $J \subseteq [0, 1]$, also $J \in \{]a, b[,]a, b], [a, b[, [a, b]\}$ mit $0 \leq a \leq b \leq 1$, sei $P_0(J)$ dessen Länge, also $b - a$. In der MIT wird gezeigt, dass durch

$$(1) \quad P(A) := \sum_{k=1}^n P_0(J_k) \quad \text{falls } n \in \mathbb{N}_0 \text{ und } A = \bigcup_{k=1}^n J_k \text{ mit p.d. Intervallen } J_k$$

ein Wahrscheinlichkeitsinhalt P auf \mathcal{A} wohldefiniert wird.¹² Zeigen Sie, dass es dann eine abzählbare Menge $\mathcal{A}_0 \subseteq \mathcal{A}$ mit 2.16(1) gibt.¹³ □

¹⁰Hier verallgemeinert durch eine einfache Anwendung des Satzes von Stone A.17; Fellers Formulierung entspricht dem Spezialfall wo \mathcal{A} als Mengenalgebra vorausgesetzt ist.

¹¹Eine hoffentlich leichter lesbare Fassung des Beweises steht in einer demnächst fertigzustellenden Arbeit Ihres Dozenten.

¹²Zum Beispiel so: Wenn man das Lebesgue-Maß auf der Borel- σ -Algebra des Intervalls $[0, 1]$ wie auch immer konstruiert hat, so ist P dessen Restriktion auf \mathcal{A} , und die Wohldefiniertheit in (1) ergibt sich aus der Additivität.

¹³Bemerkungen für Kenner der MIT: 1. Ersetzt man das obige \mathcal{A} durch $\mathcal{A} :=$ Borel- σ -Algebra auf $[0, 1]$, und dementsprechend P durch das Lebesgue-Maß auf dieser, so gibt es keine abzählbare Menge $\mathcal{A}_0 \subseteq \mathcal{A}$ mit 2.16(1), denn für jede abzählbare Menge $\mathcal{A}_0 \subseteq \mathcal{A}$ kann man für jedes nichtleere $B \in \mathcal{A}_0$ ein $y_B \in B$ wählen und hat dann mit $A := [0, 1] \setminus \{y_B : B \in \mathcal{A}_0, P(B) > 0\}$ ein Element von \mathcal{A} mit $P(A) = 1$, aber $P(B) = 0$ für jedes $B \in \mathcal{A}_0$ mit $B \subseteq A$.

2. Die in der Aufgabe zu zeigende Aussage gilt allgemeiner für jede Restriktion, auf die Intervallalgebra \mathcal{A} , eines Wahrscheinlichkeitsmaßes auf der Borel- σ -Algebra, mit nur unwesentlich komplizierterem Beweis.

Die Definitionen 2.11 und 2.12 präzisieren die Vorstellung der Unempfindlichkeit der Limites in 2.2(1) gegen eine Auswahl nur mancher (aber unendlich vieler) der dort vorkommenden Indikatoren nach “nichthellseherischen” Regeln, und Satz 2.16 zeigt, dass diese Präzisierung widerspruchsfrei ist. Dabei stellt die Abzählbarkeitsvoraussetzung an \mathcal{S} keine praktische Einschränkung dar, da überhaupt nur abzählbar viele “Regeln” hinschreibbar sind. Die in den Definitionen präzierte Vorstellung der “Nichthellseherei”, also die “Berechenbarkeit von s_ν aus den φ_i -Ereignissen mit $i < \nu$ ”, ist dabei eine nicht nur anschaulich naheliegende, sondern auch mathematisch nicht überflüssige Bedingung:

Setzt man, bei gegebenem Quasikollektiv φ_\bullet in $(\mathcal{C}, \mathcal{R})$ mit der Verteilung P auf \mathcal{A} , im Gegensatz zu 2.11(1) speziell

$$(2) \quad s_\nu := (\varphi_\nu(B_0) \in \mathcal{R}) \quad \text{für } \nu \in \mathbb{N}$$

mit einem festen $B_0 \in \mathcal{A}$ mit $P(B_0) > 0$, also s_ν fast wie in 2.11(1) oder wie spezieller u_ν in 2.13(2), aber eben gerade nicht mehr aus den $\varphi_i(B)$ mit $B \in \mathcal{A}$ und $i < \nu$ berechenbar, so existiert zwar stets der Limes links in 2.12(2), ist aber oft von $P(A)$ verschieden. Dies führt hier in Satz 2.18 auf den Begriff der bedingten Wahrscheinlichkeit, und in einem extremen Spezialfall in Korollar 2.19 zu einem **Scheingegenbeispiel**¹⁴ zu Satz 2.16. Zum Vergleich der Limites in 2.12(2) unter Verwendung der s_ν aus (2) mit den Limites aus 2.18(1) beachten wir

$$(3) \quad s_\nu \cdot (\varphi_\nu(A) \in \mathcal{R}) = (\varphi_\nu(A) \in \mathcal{R}, \varphi_\nu(B_0) \in \mathcal{R}) \quad \text{für } A \in \mathcal{A} \text{ und } \nu \in \mathbb{N}.$$

2.18 Satz. *Es sei φ_\bullet ein Quasikollektiv in der Boole-Welt $(\mathcal{C}, \mathcal{R})$ mit der Verteilung P auf der Merkmalsalgebra \mathcal{A} , und es sei $B \in \mathcal{A}$ mit $P(B) > 0$. Dann existiert¹⁵*

$$(1) \quad P(A|B) := \lim_{n \rightarrow \infty} \frac{\sum_{\nu=1}^n (\varphi_\nu(A) \in \mathcal{R}, \varphi_\nu(B) \in \mathcal{R})}{\sum_{\nu=1}^n (\varphi_\nu(B) \in \mathcal{R})} = \frac{P(A \wedge B)}{P(B)} \quad \text{für } A \in \mathcal{A}$$

und heißt die **bedingte Wahrscheinlichkeit** von A gegeben das **bedingende Ereignis** B . Die damit auf \mathcal{A} definierte Funktion $P(\cdot | B)$ heißt die **bedingte Verteilung** von φ_\bullet gegeben B . □

Beweis. Wegen der definierenden Eigenschaft 1.20(1) der Realität \mathcal{R} im ersten Schritt und der Morphismuseigenschaft der φ_ν im zweiten gilt

$$(\varphi_\nu(A) \in \mathcal{R}, \varphi_\nu(B) \in \mathcal{R}) = (\varphi_\nu(A) \wedge \varphi_\nu(B) \in \mathcal{R}) = (\varphi_\nu(A \wedge B) \in \mathcal{R}),$$

¹⁴Ein Beispiel, welches einer betrachteten Aussage zwar nicht widerspricht, wohl aber einer irgendwie nicht völlig fernliegenden Abwandlung der Aussage.

¹⁵Wir setzen $0/0 := 0$; damit ist der in (1) rechts auftretende Bruch stets definiert, was er aber, wegen der Voraussetzung $P(B) > 0$, ab einem n_0 auch ohne diese Konvention ist.

und damit ist der in (1) zu limitierende Bruch gleich

$$\frac{\frac{1}{n} \sum_{\nu=1}^n (\varphi_\nu(A \wedge B) \in \mathcal{R})}{\frac{1}{n} \sum_{\nu=1}^n (\varphi_\nu(B) \in \mathcal{R})}$$

und konvergiert also gegen R.S.(1). □

Interpretationsbeispiele zu Satz 2.18 bringt Kapitel 6, unter anderem:

- (2) Wer häufig Doppelkopf spielt, muss bei ungefähr 11 von 35 Blättern ohne eigene Kreuz-Dame mit einer Hochzeit rechnen.

2.19 Korollar (Scheingegenbeispiel zu Satz 2.16). *Es sei φ_\bullet ein Quasikollektiv in $(\mathcal{C}, \mathcal{R})$ mit der nichtdegenerierten Verteilung P auf der Merkmalalgebra \mathcal{A} . Dann gibt es eine \mathcal{A} -Auswahlregel S für die mit der Notation 2.11(1) zwar 2.12(1), aber nicht 2.12(2) gilt.*

Beweis. Mit einem aufgrund der vorausgesetzten Nichtdegeneriertheit existierenden $B_0 \in \mathcal{A}$ mit $P(B_0) \in]0, 1[$ definieren wir $s_\bullet \in \{0, 1\}^{\mathbb{N}}$ durch 2.17(2), und damit $S := \bigcup_{\nu \in \mathbb{N}} S_\nu$ durch

$$S_\nu(f) := s_\nu \quad \text{für } \nu \in \mathbb{N} \text{ und } f \in \{0, 1\}^{\{1, \dots, \nu-1\} \times \mathcal{A}}.$$

Dann gilt $\lim_{n \rightarrow \infty} \sum_{\nu=1}^n s_\nu = \infty$ wegen $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\nu=1}^n s_\nu = P(B_0) > 0$, es ist S eine \mathcal{A} -Auswahlregel, die hiesigen s_ν aus 2.17(2) sind gleich den s_ν aus 2.11(1) mit dem hiesigen S , und damit gilt zwar 2.12(1), aber speziell für $A := B_0$

$$\text{L.S. 2.12(2)} = \frac{P(B_0 \wedge B_0)}{P(B_0)} = 1 \neq \text{R.S. 2.12(2)}$$

nach 2.17(2,3) und 2.18(1) angewandt auf $A := B := B_0$. □

Dies steht in keinem Widerspruch zu Satz 2.16.

2.20 Bemerkung (2.9←). Ist $q = 0.518$ in 2.9(4) mittels einer Auswahlregel entstanden (bei idealisierender Betrachtung)? Nun, Ihr Dozent hatte erst auf die Arbutnott-Daten auf p. 20 geschaut um sicherzugehen, dass “ ≥ 0.520 ” mindestens ein paar mal vorkommt, und sich erst dann endgültig entschieden, q zu berechnen. Diese Art der “Auswahl” passt jedoch nicht zur Definition 2.11. Zwei passendere Beispiele wären

$$\frac{\sum_{i=3}^{82} x_i \cdot \left(\frac{x_{i-2}}{n_{i-2}} < \frac{x_{i-1}}{n_{i-1}} \right)}{\sum_{i=3}^{82} n_i \cdot \left(\frac{x_{i-2}}{n_{i-2}} < \frac{x_{i-1}}{n_{i-1}} \right)} \quad \text{und} \quad \frac{\sum_{i=2}^{82} x_i \cdot \left(\frac{x_{i-1}}{n_{i-1}} = \max_{j=1}^{i-1} \frac{x_j}{n_j} \right)}{\sum_{i=2}^{82} n_i \cdot \left(\frac{x_{i-1}}{n_{i-1}} = \max_{j=1}^{i-1} \frac{x_j}{n_j} \right)},$$

also “nur Jahre nach einem Anstieg” und “nur Jahre nach einem Rekord”. □

Mit Definition 2.12, ihrer Illustration durch Beispiele wie etwa in den Aufgaben 2.14 und 2.15, der Feststellung ihrer Widerspruchsfreiheit in Satz 2.16, und mit der Herleitung des Begriffs der bedingten Wahrscheinlichkeit in Satz 2.18, ist unsere Einführung des (empiristischen, von Mises folgenden) Wahrscheinlichkeitsbegriffs abgeschlossen, und wir wenden uns ab Kapitel 3 der *Wahrscheinlichkeitsrechnung* im engeren Sinne, also dem Rechnen mit Wahrscheinlichkeitsinhalten, hier spezialisiert auf elementare Wahrscheinlichkeitsmaße, zu.

3 Elementare Wahrscheinlichkeitsräume

Wir beschränken uns ab jetzt der Einfachheit halber bis auf weiteres auf elementare Wahrscheinlichkeitsräume. Diese erscheinen einerseits - weiter spezialisiert auf abzählbare oder sogar endliche Grundräume – sowohl für die Schulstochastik als auch für die meisten Anwendungen in der Informatik ausreichend, und ihr Studium ist andererseits eine gute Vorbereitung für die allgemeine, auf der Maß- und Integrationstheorie aufbauende Wahrscheinlichkeitstheorie.

Die elementare Wahrscheinlichkeitstheorie baut auf der Summentheorie auf.

3.1 Bemerkung (Summen in $[0, \infty]$ mit beliebig vielen Summanden). Für eine Familie $x = (x_i : i \in I)$ in $[0, \infty]$, also $x \in [0, \infty]^I$ mit einer beliebigen Menge I , heißt

$$(1) \quad \sum x := \sum_{i \in I} x_i := \sup \left\{ \sum_{i \in I_0} x_i : I_0 \subseteq I \text{ endlich} \right\} \in [0, \infty]$$

ihre **Summe**. Für $x \in [0, \infty]^I$ gelten damit Rechenregeln wie für Summen mit endlichen Indexmengen, wie zum Beispiel

$$(2) \quad I = \bigcup_{\alpha \in A} I_\alpha, \quad I \bullet \text{ p.d.} \Rightarrow \sum x = \sum_{\alpha \in A} \sum_{i \in I_\alpha} x_i \quad (\text{Umordnungssatz}),$$

sowie die Implikation

$$(3) \quad \sum x < \infty \Rightarrow I_0 := \{i \in I : x_i > 0\} \text{ abzählbar, } \sum x = \sum_{i \in I_0} x_i.$$

Für mehr dazu, wie später in Kapitel 10 bei der Einführung von Erwartungswerten benötigt, siehe Anhang B. \square

Siehe p. xiv für den ab jetzt verwendeten Begriff “Quasipartition”.

3.2 Definition. Es sei Ω eine Menge und es sei $P : 2^\Omega$ eine Funktion mit den Eigenschaften

- (1) $P(A) \in [0, \infty]$ für $A \subseteq \Omega$ (**Positivität**),
- (2) $\sum_{A \in \mathcal{A}} P(A) = 1$ für jede Quasipartition \mathcal{A} von Ω (**totale Additivität und Normiertheit**).

Dann heißt (Ω, P) **elementarer¹ Wahrscheinlichkeitsraum**, mit dem (**elementarem**) **Grundraum** Ω und dem (**elementaren**) **Wahrscheinlichkeitsmaß** P auf Ω . Die Elemente von 2^Ω heißen **Ereignisse**², die von Ω **Ergebnisse**. Die ein-elementigen Ereignisse heißen **Elementarereignisse**, deren Wahrscheinlichkeiten **Einzelwahrscheinlichkeiten**. \emptyset heißt **unmögliches Ereignis**, Ω **sicheres Ereignis**. \square

¹In der Literatur findet man statt “elementar” oft “**diskret**”, was jedoch in der allgemeineren maßtheoretischen WT nicht ganz dasselbe bezeichnet.

²Im Sinne der EWT aus Kapitel 2 genauer: **Modellereignisse**.

Die kurz und bündig die gewünschten Eigenschaften von P postulierende Definition 3.2 mag als etwas abstrakt empfunden werden. Sie kann äquivalent “konkreter”, aber dann umständlicher, gemäß Bemerkung 3.9 unten formuliert werden, wie zum Beispiel bei Dümbgen (2003, p. 10, Definition 2.8), Storch/Wiebe (2003, Definitionen 7.A.6 und 7.A.8), oder Henze (2021, die “Reparatur” 22.2 der Definition 22.1).

Vereinbarung: Wir lassen ab hier bis auf weiteres zur Abkürzung das Adjektiv “elementar” meist weg.

Statt “Wahrscheinlichkeitsmaß” sagt man auch **Wahrscheinlichkeitsverteilung** oder kurz **Verteilung**. Statt “Wahrscheinlichkeitsmaß auf Ω ” sagt man auch, richtiger aber etwas pedantisch, “Wahrscheinlichkeitsmaß auf 2^Ω ”.

Wir lesen $P(A)$ als “ P von A ” oder auch als “**Wahrscheinlichkeit von A** ”. Neben P werden natürlich auch andere Buchstaben zur Bezeichnung von Wahrscheinlichkeitsmaßen verwendet, zum Beispiel \mathbb{P} , Q und R .

Für Ereignisse $A \subseteq \Omega$ sei im folgenden $A^c := \Omega \setminus A$ das Komplement von A in Ω .

3.3 Satz (Rechenregeln). *Es sei (Ω, P) ein Wahrscheinlichkeitsraum. Dann ist Ω nichtleer, und für $A, B \subseteq \Omega$ und für jede Familie $A_\bullet = (A_i)_{i \in I}$ von Ereignissen $A_i \subseteq \Omega$ gelten die Rechenregeln*

- (1) $P(\emptyset) = 0$, (Nulltreue),
- (2) $P(\Omega) = 1$, (Normiertheit),
- (3) $P(A) \in [0, 1]$,
- (4) $P(A) = 1 - P(A^c)$ (Komplementarität),
- (5) A, B disjunkt $\Rightarrow P(A \cup B) = P(A) + P(B)$ ((paarw.) Additivität),
- (6) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Siebformel, $n = 2$),
- (7) $P(A \cup B) \leq P(A) + P(B)$ (paarweise Subadditivität),
- (8) $A \subseteq B \Rightarrow P(A) \leq P(B)$ (Isotonie),
- (9) $A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A)$ (Subtraktivität),
- (10) A_\bullet p.d. $\Rightarrow P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i)$ (totale Additivität),
- (11) A_\bullet Quasipart. v. $\Omega \Rightarrow \sum_{i \in I} P(A_i) = 1$ (tot. Add. u. Normiertheit),
- (12) $P\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} P(A_i)$ (totale Subadditivität),
- (13) $A \subseteq \bigcup_{i \in I} A_i \Rightarrow P(A) \leq \sum_{i \in I} P(A_i)$,
- (14) $P(A_i) = 0$ für $i \in I \Rightarrow P\left(\bigcup_{i \in I} A_i\right) = 0$, $P(A_i) = 1$ für $i \in I \Rightarrow P\left(\bigcap_{i \in I} A_i\right) = 1$.

Insbesondere ist P ein Wahrscheinlichkeitsinhalt auf der Potenzmengenalgebra 2^Ω .

Die Wichtigkeit von (4) für konkrete Rechnungen kann nicht überbetont werden.

Beweis. $\Omega \neq \emptyset$: Im Fall $\Omega = \emptyset$ wäre \emptyset eine Quasipartition (tatsächlich die einzige Partition, neben der Quasipartition $\{\emptyset\}$) von Ω , und wir erhielten den Widerspruch $0 = \sum_{A \in \emptyset} P(A) = 1$, mit 3.2(2) im zweiten Schritt.

(2) ist 3.2(2) mit der Quasipartition $\mathcal{A} := \{\Omega\}$.

(3), (4): 3.2(2) mit der Quasipartition $\mathcal{A} := \{A, A^c\}$ liefert $P(A) + P(A^c) = 1$ (denn $A \neq A^c$ wegen $\Omega \neq \emptyset$), und mit 3.2(1) folgt $P(A) \leq 1$ und $P(A^c) \leq 1 < \infty$, also (3) und (4).

(1) folgt aus (4) mit $A := \emptyset$ unter Benutzung von (2).

(10): A_\bullet sei p.d. Wegen (1) kann ohne Einschränkung $A_i \neq \emptyset$ für $i \in I$, und damit die Injektivität von A_\bullet , angenommen werden. Mit $A := \bigcup_{i \in I} A_i$ liefert 3.2(2) angewandt auf die Quasipartition $\{A_i : i \in I\} \cup \{A^c\}$ folglich $1 = P(A^c) + \sum_{i \in I} P(A_i)$, und mit (4) folgt R.S.(10).

(11): Klar mit (10) und (2).

(5) ist ein Spezialfall von (10), mit $I := \{1, 2\}$, $A_1 := A$, $A_2 := B$.

P ist ein Wahrscheinlichkeitsinhalt wegen (1–3, 5).

(6): Die drei Mengen $A \setminus B$, $A \cap B$, $B \setminus A$ sind paarweise disjunkt und mit (5) folgt $P(A) = P(A \setminus B \cup A \cap B) = P(A \setminus B) + P(A \cap B)$, $P(B) = P(B \setminus A) + P(A \cap B)$, und $P(A \cup B) = P(A \setminus B \cup A \cap B \cup B \setminus A) = P(A \setminus B) + P(A \cap B) + P(B \setminus A)$. Einsetzen der drei erhaltenen Ausdrücke in (6) verifiziert die Behauptung. Dass die dabei auftretenden Subtraktionen erlaubt sind und überhaupt die rechte Seite von (6) definiert ist, ergibt sich aus der in (3) enthaltenen Endlichkeit von P .

(7), (8): Übung.

(9): Anwendung der Additivität (5) auf die disjunkten Mengen A und $B \setminus A$.

Beweis von (12) im Spezialfall I abzählbar³: Ohne Einschränkung kann angenommen werden, dass $I = \{1, \dots, n\}$ für ein $n \in \mathbb{N}_0$ oder $I = \mathbb{N}$ gilt. Mit

³Der allgemeine Fall kommt praktisch nie vor und ist anscheinend schwerer zu beweisen. Wir geben hier zwei Beweise. Der erste benutzt das folgende allgemeine

Disjunktifizierungslemma. Zu jeder Familie $(A_i : i \in I)$ von Mengen gibt es eine Familie $(B_i : i \in I)$ paarweise disjunkter Mengen mit $B_i \subseteq A_i$ für $i \in I$ und $\bigcup_{i \in I} B_i = \bigcup_{i \in I} A_i$.

Beweis: Mit der gegebenen Indexmenge I sei \mathcal{M} die Menge aller Familien $C_\bullet = (C_i : i \in I)$ paarweise disjunkter Mengen mit $C_i \subseteq A_i$ für $i \in I$. Für $C_\bullet, D_\bullet \in \mathcal{M}$ definieren wir $C_\bullet \leq D_\bullet$: $\Leftrightarrow C_i \subseteq D_i$ für $i \in I$. Damit ist (\mathcal{M}, \leq) eine partiell geordnete Menge. Nach dem Hausdorffschen Maximalkettensatz (siehe zum Beispiel Rudin 1987, p. 396, oder Halmos 1960, Section 16) enthält \mathcal{M} eine maximale Kette \mathcal{C} , d.h. es gilt $\mathcal{C} \subseteq \mathcal{M}$, \mathcal{C} ist bezüglich \leq total geordnet (\Leftrightarrow : \mathcal{C} ist **Kette**), und es gibt keine bezüglich \leq total geordnete Menge $\mathcal{D} \subseteq \mathcal{M}$ mit $\mathcal{D} \supsetneq \mathcal{C}$ (\Leftrightarrow : \mathcal{C} ist eine bezüglich der Mengeninklusion **maximale** Kette). Mit einem solchen \mathcal{C} setzen wir $B_i = \bigcup_{C_\bullet \in \mathcal{C}} C_i$ für $i \in I$. Wegen der Ketteneigenschaft von \mathcal{C} ist dann $(B_i : i \in I)$ eine Familie paarweise disjunkter Mengen. Offensichtlich gilt $\bigcup_{i \in I} B_i \subseteq \bigcup_{i \in I} A_i$, und wegen der Maximalität von \mathcal{C} kann hier keine echte Inklusion (d.h. “ \subsetneq ”) vorliegen. \square

Erster Beweis von (12) im allgemeinen Fall: Mit $(B_i : i \in I)$ wie im Disjunktifizierungslemma gilt (15). \square

Ein zweiter Beweis benutzt den abzählbaren Spezialfall und, ohne Zirkelschluss, den späteren Satz 3.8:

Zweiter Beweis von (12) im allgemeinen Fall: Nach Satz 3.8 gibt es eine abzählbare Menge

$B_i := A_i \setminus \bigcup_{j=1}^{i-1} A_j$ für $i \in I$, also $B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus A_1 \cup A_2$ usw., gilt $\bigcup_{i \in I} A_i = \bigcup_{i \in I} B_i$, denn “ \supseteq ” ist trivial und für $\omega \in \text{L.S.}$ und $i := \min\{i' \in I : \omega \in A_{i'}\}$ gilt $\omega \in B_i \subseteq \text{R.S.}$, und die B_i sind paarweise disjunkt, wegen $B_k \subseteq A_i^c \subseteq B_i^c$ für $k > i$. Es folgt, mit (10) im zweiten Schritt und (8) im dritten,

$$(15) \quad P\left(\bigcup_{i \in I} A_i\right) = P\left(\bigcup_{i \in I} B_i\right) = \sum_{i \in I} P(B_i) \leq \sum_{i \in I} P(A_i).$$

(Die endliche oder unendliche Folge B_\bullet heißt **Disjunktifizierung** der Folge A_\bullet .)

(13): Übung.

(14): Gilt $P(A_i) = 0$ für $i \in I$, so liefern (12) und (3)

$$0 \leq P\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} 0 = 0.$$

Damit und mit (4) ergibt sich im Fall von $P(A_i) = 1$ für $i \in I$

$$P\left(\bigcap_{i \in I} A_i\right) = 1 - P\left(\bigcup_{i \in I} A_i^c\right) = 1. \quad \square$$

3.4 Bemerkung. Die letzte Aussage in Satz 3.3 ist nicht umkehrbar:

Für den Wahrscheinlichkeitsinhalt P aus Aufgabe 2.17 ist $P(\{\omega\}) = P([\omega, \omega]) = \omega - \omega = 0$ für $\omega \in [0, 1] =: \Omega$, also $\sum_{\omega \in \Omega} P(\{\omega\}) = 0 \neq 1$ im Gegensatz zu 3.3(11). Nun ist P gar nicht auf ganz 2^Ω definiert, also auch schon deswegen gar kein elementares Wahrscheinlichkeitsmaß, lässt sich aber (zum Beispiel nach Elstrodt (2018, p. 4, Satz von Banach (1923)) zu einem Wahrscheinlichkeitsinhalt Q auf 2^Ω fortsetzen, der dann wegen $\sum_{\omega \in \Omega} Q(\{\omega\}) = 0 \neq 1$ kein elementares Wahrscheinlichkeitsmaß sein kann.

Übrigens ist jedes derartige Q auch kein (nichtelementares) Wahrscheinlichkeitsmaß im Sinne der MIT, siehe dazu etwa Elstrodt (2018, p. 5, Satz von Vitali (1905)) und Dudley (2002, p. 526, Satz von Banach und Kuratowski (1929)). \square

3.5 Aufgabe (Mögliche Durchschnittswahrscheinlichkeiten). Es seien A und B stets Ereignisse in einem Wahrscheinlichkeitsraum (Ω, P) mit $P(A) = 0.4$ und $P(B) = 0.75$.

(a) Zeigen Sie, dass $P(A \cap B) \geq 0.15$ gilt.

(b) Konstruieren Sie Ω, P, A und B so, dass $\#(\Omega) = 4$ und $P(A \cap B) = 0.2$ gilt.

(c) Geben Sie alle Werte für $P(A \cap B)$ an, die bei geeigneter Wahl von (Ω, P) auftreten können.

Hinweis: Satz 3.6(b) darf hier schon benutzt werden. \square

$A_0 \subseteq \bigcup_{i \in I} A_i =: A$ mit $P(A) = P(A_0)$. Wegen $A_0 \subseteq \bigcup_{i \in I} A_i$ können wir zu jedem $x \in A_0$ ein $i(x) \in I$ mit $x \in A_{i(x)}$ auswählen. Damit ist $I_0 := \{i(x) : x \in A_0\}$ abzählbar und $A_0 \subseteq \bigcup_{i \in I_0} A_i$. Unter Verwendung des abzählbaren Spezialfalls von (12) folgt $P(A) = P(A_0) \leq \sum_{i \in I_0} P(A_i) \leq \sum_{i \in I} P(A_i)$. \square

Wir zeigen nun, dass ein elementares Wahrscheinlichkeitsmaß schon durch seine Einzelwahrscheinlichkeiten festgelegt ist:

3.6 Satz. (a) *Es sei (Ω, P) ein Wahrscheinlichkeitsraum. Mit*

$$(1) \quad p(\omega) := P(\{\omega\}) \quad \text{für } \omega \in \Omega$$

gilt dann

$$(2) \quad p(\omega) \in [0, 1] \quad \text{für } \omega \in \Omega,$$

$$(3) \quad \sum_{\omega \in \Omega} p(\omega) = 1 \quad \text{(Normiertheit),}$$

$$(4) \quad P(A) = \sum_{\omega \in A} p(\omega) \quad \text{für } A \subseteq \Omega.$$

(b) *Es sei Ω eine Menge und es sei $p : \Omega \rightarrow [0, \infty]$ eine Funktion mit (3). Dann definiert (4), mit “=” ersetzt durch “:=”, ein Wahrscheinlichkeitsmaß P auf Ω , und es gilt (1) mit “:=” ersetzt durch “=”.*

Beweis. (a) (2) ist 3.3(3) mit $A := \{\omega\}$.

(4) ist 3.3(10) mit $A_\omega := \{\omega\}$ für $\omega \in A$.

(3) ist (4) mit $A := \Omega$, wegen 3.3(2).

(b) Es ist $P : 2^\Omega$ eine Funktion mit 3.2(1). Für jede Quasipartition \mathcal{A} von Ω ist $\Omega = \bigcup_{A \in \mathcal{A}} A$ mit id_A p.d., also mit dem Umordnungssatz 3.1(2) im zweiten Schritt

$$\sum_{A \in \mathcal{A}} P(A) = \sum_{A \in \mathcal{A}} \sum_{\omega \in A} p(\omega) = \sum_{\omega \in \Omega} p(\omega) = 1;$$

folglich gilt auch 3.2(2). Weiter gilt für $\omega \in \Omega$ auch $P(\{\omega\}) = \sum_{\omega' \in \{\omega\}} p(\omega') = p(\omega)$; also gilt (1) mit “:=” ersetzt durch “=”. \square

3.7 Definition. Jede Funktion $p : \Omega \rightarrow [0, \infty]$ mit 3.6(3) heißt **Dichte**, genauer **(elementare) (Wahrscheinlichkeits)zähldichte** oder auch **Wahrscheinlichkeitsfunktion**, auf Ω . Die Menge all dieser Dichten bezeichnen wir mit $\text{prob}(\Omega)$, die Menge aller Wahrscheinlichkeitsmaße auf Ω mit $\text{Prob}(\Omega)$. \square

Es ist also

$$\text{prob}(\Omega) = \left\{ p \in [0, 1]^\Omega : \sum_{\omega \in \Omega} p(\omega) = 1 \right\},$$

wobei zum Nachweis von $p \in \text{prob}(\Omega)$ statt der ersten Bedingung $p \in [0, 1]^\Omega$ nur $p \in [0, \infty]^\Omega$ nachgeprüft zu werden braucht, wegen Satz 3.6(b) und 3.6(2).

Nach Satz 3.6 sind bei gegebener Grundmenge Ω die dort betrachteten Abbildungen $\text{Prob}(\Omega) \ni P \mapsto p \in \text{prob}(\Omega)$ und $\text{prob}(\Omega) \ni p \mapsto P \in \text{Prob}(\Omega)$ zueinander inverse Bijektionen; daher rechnet man in konkreten Fällen je nach Bequemlichkeit mal direkt mit P , mal mit p .

Oft bezeichnet man die zu einem Wahrscheinlichkeitsmaß P gehörige Dichte auch mit einem anderen Buchstaben, z.B. gerne mit f oder, wenn noch weitere Wahrscheinlichkeitsmaße vorkommen, mit f_P .

Jede elementare Verteilung ist in abzählbar vielen Punkten konzentriert:

3.8 Satz. *Es sei (Ω, P) ein Wahrscheinlichkeitsraum mit zugehöriger Dichte p . Dann ist die Menge*

$$(1) \quad \Omega_0 := \{\omega \in \Omega : p(\omega) > 0\}$$

abzählbar, und es gilt

$$(2) \quad P(A) = \sum_{\omega \in A \cap \Omega_0} p(\omega) = P(A \cap \Omega_0) \quad \text{für } A \subseteq \Omega.$$

Beweis. Aus 3.6(3) und 3.1(3) ergibt sich die Abzählbarkeit von Ω_0 . Anwendung von 3.6(4) und 3.1(3) mit $I := A$ und $I_0 = A \cap \Omega_0$ liefert die erste Gleichung in (2). Nochmals 3.6(4), angewandt auf $A \cap \Omega_0$, liefert die zweite Gleichheit in (2). \square

Aus den Sätzen 3.6 und 3.8 ergibt sich leicht:

3.9 Bemerkung. Es sei Ω eine Menge und $P : 2^\Omega$ ein Funktion. Dann ist P genau dann ein elementares Wahrscheinlichkeitsmaß auf Ω , wenn es eine abzählbare Menge $\Omega_0 \subseteq \Omega$ und eine Funktion $p_0 : \Omega_0 \rightarrow [0, \infty]$ mit $\sum_{\omega \in \Omega_0} p_0(\omega) = 1$ und

$$(1) \quad P(A) = \sum_{\omega \in A \cap \Omega_0} p_0(\omega) \quad \text{für } A \subseteq \Omega$$

gibt. Dabei können die Summen rechts in (1) gemäß Satz B.15 auch als Grenzwerte unendlicher Reihen, jeweils bezüglich einer beliebigen Abzählung, gelesen werden, was manchem “elementarer” als die Verwendung des allgemeinen Summenbegriffs aus Bemerkung 3.1 erscheinen mag. \square

3.10 Definition. Es sei (Ω, P) ein Wahrscheinlichkeitsraum. Dann heißt jede Menge $A \subseteq \Omega$ mit $P(A) = 0$ eine **Nullmenge**, jede mit $P(A) = 1$ eine **Einsmenge**, oder ein **Träger (im weiteren Sinne)**, von P . Die bezüglich Mengeninklusion kleinste Einsmenge von P , also die Menge Ω_0 aus 3.8(1), heißt **der Träger (im engeren Sinn)**, oder **kleinster Träger**, von P , und wird mit $\text{Supp } P$ bezeichnet.⁴ \square

Rechenregel 3.3(14) sagt damit: Vereinigungen von Nullmengen sind Nullmengen, Schnitte von Einsmengen sind Einsmengen

In einem Wahrscheinlichkeitsraum (Ω, P) ist einerseits stets \emptyset eine Nullmenge, und andererseits $\Omega \setminus \text{Supp } P$ die bezüglich Mengeninklusion größte Nullmenge. Nichtleere Nullmengen werden oft als lästig empfunden, was eine Motivation für die folgende Vereinbarung ist.

⁴Von engl. *support* = dt. Träger. Oft “supp” geschrieben; hier groß da Menge, analog zu Argmax in Definition 3.21.

3.11 Vereinbarung. Sind (Ω_1, P_1) und (Ω_2, P_2) zwei Wahrscheinlichkeitsräume mit $P_1(\Omega_1 \cap \Omega_2) = 1 = P_2(\Omega_1 \cap \Omega_2)$ und

$$P_1(A) = P_2(A) \quad \text{für } A \subseteq \Omega_1 \cap \Omega_2,$$

so benennen wir oft P_1 und P_2 gleich, und ebenso die zugehörigen Dichten. Insbesondere bezeichnen wir bei gegebenem Wahrscheinlichkeitsraum (Ω_0, P_0) und gegebener Menge $\Omega \supseteq \Omega_0$ das durch

$$P(A) := P_0(A \cap \Omega_0) \quad \text{für } A \subseteq \Omega$$

definierte Wahrscheinlichkeitsmaß P auf Ω oft wieder mit P_0 . □

Zum Beispiel sehen wir in diesem Sinne die unten in 3.12 als ein Element von $\text{Prob}(\mathbb{N}_0)$ definierte Binomialverteilung $B_{n,p}$ gerne als ein Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ an, aber manchmal auch als eine Verteilung auf \mathbb{R} . In diesem Sinne schreiben wir für Grundmengen $\mathcal{X} \subseteq \mathcal{Y}$ gelegentlich auch $\text{Prob}(\mathcal{X}) \subseteq \text{Prob}(\mathcal{Y})$ und $\text{prob}(\mathcal{X}) \subseteq \text{prob}(\mathcal{Y})$, was wörtlich genommen im Fall $X \neq \mathcal{Y}$ Unsinn ist.

Ein wesentlicher Teil der Wahrscheinlichkeitsrechnung widmet sich dem Studium gewisser spezieller Verteilungen, welche sich für die verschiedensten inner- und außermathematischen Anwendungen als wichtig erwiesen haben. Nun können die meisten dieser Verteilungen auf natürliche Weise in einer einführenden Vorlesung wie dieser erst recht spät eingeführt werden, zum Beispiel die besonders wichtigen Binomialverteilungen als Verteilungen von Trefferzahlen in Bernoulli-Ketten erst nach der Einführung von Zufallsgrößen, Unabhängigkeit, bedingten Wahrscheinlichkeiten, und Kopplungsmodellen endlich in Kapitel 9.⁵ Andererseits sollten sie als Beispielmateriale möglichst früh zur Verfügung stehen. Wir lassen daher hier für den - im hiesigen elementaren Rahmen wichtigsten - Fall des Grundraumes \mathbb{N}_0 die allerwichtigsten speziellen Verteilungen per Definition durch ihre Dichten vom Himmel fallen:

3.12 Die wichtigsten Verteilungen auf \mathbb{N}_0 . Die Tabelle auf der folgenden Seite behauptet zunächst, dass die jeweilige Funktion f unter den angegebenen Restriktionen an die Parameter eine Wahrscheinlichkeitsdichte auf \mathbb{N}_0 ist, und insbesondere für $x \in -\mathbb{N}$ automatisch den Wert 0 liefert, mit den auf Seite xv eingeführten Indikatorschreibweisen, der allgemeinen Definition 4.2(1,3) der Binomialkoeffizienten $\binom{x}{k}$ mit $k \in \mathbb{Z}$, und der starken Nullkonvention B.7(2). Für Beweise siehe 3.13, 3.15, 4.8(b).

Die mit P bezeichnete Spalte gibt das von uns für die Verteilung verwendete Symbol an. Mit Ausnahme der Dirac-Verteilungen δ_a bezeichnen wir die zugehörigen Dichten mit den entsprechenden klein geschriebenen Buchstaben oder Buchstabenkombinationen, also $b_p, b_{n,p}, p_\lambda, h_{n,r,b}, g_p, nb_{r,p}$.

Gelegentlich nützlich sind noch streng genommen unlogische Schreibweisen wie etwa $B_{n=10,p=0.4}$ für $B_{10,0.4}$, oder $p_{\lambda=1}$ für p_1 , bei denen die Kenntnis von Parameterbenennungskonventionen einen Rückschluss auf das jeweils bezeichnete Objekt erleichtert.

⁵Das wird auch von anderen Autoren so gesehen, siehe etwa Henze (2021, Kapitel 18).

Name	P	Parameter	Dichte $\mathbb{Z} \ni x \mapsto f(x) :=$	Supp P	$\mu(P)$	$\sigma^2(P)$	$f(x)$ in R	Genese
Dirac-Verteilung	δ_a	$a \in \mathbb{N}_0$	$(x = a)$	$\{a\}$	a	0	<code>x==a</code>	5.8
Bernoulli-Vert.	B_p	$p \in [0, 1]$	$\mathbf{1}_{\{0,1\}}(x) \cdot p^x(1-p)^{1-x}$	$\{0, 1\}$ falls $p \in]0, 1[$	p	$p(1-p)$	<code>dbinom(x,1,p)</code>	5.9
Binomialvert.	$B_{n,p}$	$n \in \mathbb{N}_0, p \in [0, 1]$	$\binom{n}{x} p^x(1-p)^{n-x}$	$\{0, \dots, n\}$ falls $p \in]0, 1[$	np	$np(1-p)$	<code>dbinom(x,n,p)</code>	7.11 ($p \in \mathbb{Q}$), 9.5
Poisson-Vert.	P_λ	$\lambda \in [0, \infty[$	$\mathbf{1}_{\mathbb{N}_0}(x) \cdot e^{-\lambda} \frac{\lambda^x}{x!}$	\mathbb{N}_0 falls $\lambda > 0$	λ	λ	<code>dpois(x,lambda)</code>	4.11 ($\lambda=1$), 11.1
hypergeometr. V.	$H_{n,r,b}$	$n, r, b \in \mathbb{N}_0, n \leq r+b$	$\binom{r}{x} \binom{b}{n-x} / \binom{r+b}{n}$	$\{(n-b)_+, \dots, n \wedge r\}$	$\frac{nr}{r+b}$	$\frac{nr b(r+b-n)}{(r+b)^2(r+b-1)}$	<code>dhyper(x,r,b,n)</code>	5.15, 7.11
geometrische Vert.	G_p	$p \in]0, 1]$	$\mathbf{1}_{\mathbb{N}_0}(x) \cdot p(1-p)^x$	\mathbb{N}_0 falls $p < 1$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	<code>dgeom(x,p)</code>	9.20
Negativbinomialv.	$NB_{r,p}$	$r \in [0, \infty[, p \in]0, 1]$	$\binom{x+r-1}{x} p^r(1-p)^x$	\mathbb{N}_0 falls $r > 0, p < 1$	$r \frac{1-p}{p}$	$r \frac{1-p}{p^2}$	<code>dnbinom(x,r,p)</code>	9.20

Spezialfälle: $B_p = B_{1,p}$, $G_p = NB_{1,p}$.

Degenerierte Spezialfälle: δ_a , $B_0 = \delta_0$, $B_1 = \delta_1$, $B_{n,0} = B_{0,p} = \delta_0$, $B_{n,1} = \delta_n$, $P_0 = \delta_0$, $H_{0,r,b} = H_{n,0,b} = \delta_0$, $H_{n,r,0} = \delta_n$, $H_{r+b,r,b} = \delta_r$, $G_1 = \delta_0$, $NB_{r,1} = NB_{0,p} = \delta_0$; in allen anderen Fällen ist $\#\text{Supp } P \geq 2$, also P nichtdegeneriert.

Jede der hier betrachteten Verteilungen ist unimodal gemäß Definition 3.23, mit Modalstellen meist ungefähr gleich den in Definition 10.13 erklärten und in 10.18, 10.19, 10.48, 11.13(b) berechneten Erwartungswerten $\mu(P)$, siehe Beispiel 3.24 für die Fälle Dirac, Bernoulli, Binomial, Poisson, und Aufgabe 5.16 für den hypergeometrischen Fall. Natürliche Beispiele **bimodaler** Verteilungen sind die diskreten Arkussinusverteilungen aus den Aufgaben 3.30 und 4.8(e).

Die Varianzen $\sigma^2(P)$ werden erklärt in Definition 10.24 und berechnet in 10.26, 10.30, 10.48, 11.13(b).

In R kann man sich obige Dichten etwa in einem Binomialfall durch Eingabe von `n <- 10; p <- .3; x <- -n:(2*n); plot(x,dbinom(k,n,p))` veranschaulichen, oder auch als **Stabdiagramm** `plot(x,dbinom(x,n,p), type="h")`. Dabei steht das **d** für *density*; mit stattdessen **p** wie *probability* erhält man die Verteilungsfunktion, mit **q** die (untere) Quantilfunktion, gemäß 10.3 und 10.11. □

3.13 Beispiel. Es sei $n \in \mathbb{N}_0$ und $p \in [0, 1]$. Dann ist die Funktion

$$\Omega := \{0, \dots, n\} \ni k \mapsto \binom{n}{k} p^k (1-p)^{n-k} =: b_{n,p}(k)$$

nach der binomischen Formel eine Dichte auf Ω , und das zugehörige Wahrscheinlichkeitsmaß heißt **Binomialverteilung** mit den Parametern n und p , bezeichnet mit $B_{n,p}$. Dabei wird p auch **Trefferwahrscheinlichkeit** genannt, was durch Beispiel 7.11(b) und Satz 9.5 motiviert wird. Wegen $\binom{n}{k} = 0$ für $k \in \mathbb{N}_0$ mit $k > n$ nach 4.2(1) kann hier Ω durch \mathbb{N}_0 ersetzt werden, was wir ab (1) bequemerweise tun.

Zahlenbeispiel: $B_{4,1/2}(\{0, 2\}) = b_{4,1/2}(0) + b_{4,1/2}(2) = 1 \cdot 2^{-4} + 6 \cdot 2^{-4} = \frac{7}{16}$.

Interessanteres Beispiel: Mit dem Ereignis

(1) “Gerade” := $2\mathbb{N}_0$

erhalten wir mit 3.6(4) im ersten Schritt, 3.6(3) im dritten, und der binomischen Formel im vierten,

$$\begin{aligned} (2) \quad B_{n,p}(\text{“Gerade”}) &= \sum_{k \in 2\mathbb{N}_0} b_{n,p}(k) = \sum_{k \in \mathbb{N}_0} \frac{1+(-1)^k}{2} b_{n,p}(k) \\ &= \frac{1}{2} \left(1 + \sum_{k \in \mathbb{N}_0} \binom{n}{k} (-p)^k (1-p)^{n-k} \right) = \frac{1}{2} (1 + (1-2p)^n), \end{aligned}$$

und als Konsequenz also $\lim_{n \rightarrow \infty} B_{n,p}(\text{“Gerade”}) = \frac{1}{2}$ falls $p \in]0, 1[$. □

3.14 Aufgabe (Für Freunde der komplexen Einheit). Zeigen Sie

$$\lim_{n \rightarrow \infty} B_{n,p}(\text{“Durch 4 teilbar”}) = \frac{1}{4} \quad \text{für } p \in]0, 1[$$

mittels einer zum Ergebnis der Gleichungskette 3.13(2) analogen Darstellung. □

3.15 Aufgabe. Die in 3.12 definierten Funktionen $p_\lambda, g_p, nb_{r,p}$ sind tatsächlich Wahrscheinlichkeitsdichten, mit den angegebenen zugehörigen Trägern.

Hinweis: Binomialkoeffizientendefinition 4.2(1) und Binomialreihe 9.19(1). □

Welche Verteilungen sind nun für Modellierungen relevant? Rein mathematisch kommt *jede* (elementare) Verteilung als Verteilung eines Kollektivs gemäß den Definitionen 2.2 und 2.12 in Betracht:

3.16 Satz. *Es sei (Ω, P) ein Wahrscheinlichkeitsraum. Dann erfüllt P die Voraussetzungen des Widerspruchsfreiheitssatzes 2.16 von Feller.*

Beweis. Es sei $\Omega_0 := \text{Supp } P$, der Träger von P . Dann ist $\mathcal{A}_0 := \{B \subseteq \Omega_0 : B \text{ endlich}\}$ eine abzählbare Teilmenge der Potenzmengenalgebra $\mathcal{A} := 2^\Omega$, und für $A \in \mathcal{A}$ ist, mit 3.8(2) im ersten Schritt und der Definition der Summe aus 3.1(1) im zweiten,

$$\begin{aligned} P(A) &= \sum_{\omega \in A \cap \Omega_0} p(\omega) = \sup \left\{ \sum_{\omega \in B} p(\omega) : B \subseteq A \cap \Omega_0 \text{ endlich} \right\} \\ &= \sup \left\{ P(B) : B \in \mathcal{A}_0, B \subseteq A \right\}. \end{aligned} \quad \square$$

Direkt mit der Erfahrungswelt verknüpft sind hingegen zunächst nur empirisch erhaltene Verteilungen wie in den Beispielen 2.8, 2.9, 2.10, 2.14 (siehe dazu genauer unten 3.20(a)), wobei in 2.9, 2.10, 2.14 dann anzunehmen ist, dass die Längen n der betrachteten Folgen, also $n = 938\,223, 100, 10\,000$, die idealisiert geforderte Grenzwertbeziehung 2.2(1) in zufriedenstellender Genauigkeit realisieren. (Wir werden dies in späteren Beispielen gelegentlich mangels Daten für noch viel kleinere n annehmen.)

Andererseits legen oft Symmetriebetrachtungen die Annahme einer Gleichverteilung im Sinne der folgenden Definition nahe; ob sich so eine Annahme im Sinne der EWT dann bewährt ist natürlich in jedem Einzelfall eine andere Frage.

3.17 Definition. Es sei Ω eine endliche nichtleere Menge. Dann definiert

$$p(\omega) := \frac{1}{\#\Omega} \quad \text{für } \omega \in \Omega$$

eine Dichte p auf Ω . Das zugehörige Wahrscheinlichkeitsmaß heißt **Gleichverteilung** oder **Laplace-Verteilung** oder **uniforme Verteilung**⁶ auf Ω und wird mit U_Ω bezeichnet:

$$(1) \quad U_\Omega(A) := \frac{\#A}{\#\Omega} \quad \text{für } A \subseteq \Omega.$$

Es sei allgemeiner $\emptyset \neq \Omega_0 \subseteq \Omega$ mit Ω_0 endlich. Dann definiert

$$(2) \quad p(\omega) := \begin{cases} \frac{1}{\#\Omega_0} & \text{für } \omega \in \Omega_0, \\ 0 & \text{für } \omega \in \Omega \setminus \Omega_0 \end{cases}$$

eine Dichte p auf Ω . Das zugehörige Wahrscheinlichkeitsmaß heißt **Gleichverteilung** oder **Laplace-Verteilung** oder **uniforme Verteilung** in Ω_0 und wird mit U_{Ω_0} bezeichnet:

$$(3) \quad U_{\Omega_0}(A) = \frac{\#(A \cap \Omega_0)}{\#\Omega_0} \quad \text{für } A \subseteq \Omega.$$

Schließlich setzt man im Spezialfall einer einelementigen Menge $\Omega_0 = \{\omega_0\}$ noch $\delta_{\omega_0} := U_{\{\omega_0\}}$ und nennt dieses Wahrscheinlichkeitsmaß das **Dirac-Maß** oder die **Dirac-Verteilung** oder **Einpunktverteilung** in ω_0 . Es gilt also

$$(4) \quad \delta_{\omega_0}(A) = \begin{cases} 1 & \text{falls } \omega_0 \in A, \\ 0 & \text{falls } \omega_0 \notin A \end{cases} = \mathbf{1}_A(\omega_0) = (\omega_0 \in A) \quad \text{für } A \subseteq \Omega.$$

□

⁶Die Bezeichnung ‘uniforme Verteilung’ ist weniger verwechslungsanfällig als ‘Gleichverteilung’, siehe dazu die Fußnote auf p. 57 zu ‘identisch verteilt’, desweiteren Beispiel 5.6.

Die auf den ersten Blick uninteressant erscheinenden Dirac-Maße tauchen einerseits ständig als Randfälle interessanterer Verteilungen auf, zum Beispiel als degenerierte Binomialverteilungen $B_{n,0} = \delta_0$ und $B_{n,1} = \delta_n$, und sind andererseits notationell nützlich:

3.18 Bemerkung. Analog zu 2.5(c) ist jede Konvexkombination $\sum_{j \in J} \lambda_j P_j$ von elementaren Wahrscheinlichkeitsmaßen P_j auf einer Menge \mathcal{X} wieder ein solches. Ist nun $P \in \text{Prob}(\mathcal{X})$ mit Dichte f , so ist

$$(1) \quad P = \sum_{x \in \mathcal{X}} f(x) \delta_x,$$

denn $Q := \text{R.S}(1)$ erfüllt

$$Q(A) = \sum_{x \in \mathcal{X}} f(x) \delta_x(A) = \sum_{x \in \mathcal{X}} f(x) \cdot (x \in A) = \sum_{x \in A} f(x) = P(A) \quad \text{für } A \subseteq \mathcal{X}$$

nach Definition von Q gemäß 2.5(c) im ersten Schritt, und 3.6(4) im letzten.

Oft ist es bequem, ein Wahrscheinlichkeitsmaß P in der Form (1) anzugeben; zum Beispiel ist “Es sei $P := \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} \delta_x$ ” kürzer als “Es sei P die Verteilung zur Dichte $\{0, \dots, n\} \ni x \mapsto \binom{n}{x} p^x (1-p)^{n-x}$ ”. \square

Möchte man für eine “reale” Situation ein Modell angeben, so muss man natürlich neben der expliziten Angabe von (Ω, P) stets auch sagen, welche realen Ergebnisse welchen Ergebnissen ω im Modell entsprechen sollen, so wie unten in 3.19(1) beziehungsweise 3.19(2), oder auch wie in Ihrer Lösung der Aufgabe 3.28. Beispiel 3.19 zeigt insbesondere, dass ein bloßes “Alle Ergebnisse seien gleichwahrscheinlich” ohne weitere (oft jedoch als selbstverständlich erscheinende) Annahmen noch kein Modell spezifiziert.

3.19 Beispiel (Zweifacher Würfelwurf). Zwei ununterscheidbare Würfel werden einmal gleichzeitig geworfen.

(a) Als Grundraum könnte man

$$\begin{aligned} \Omega_a &:= \{ \omega \subseteq \{1, \dots, 6\} : \#\omega \in \{1, 2\} \} \\ &= \{ \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \dots, \{5, 6\} \} \end{aligned}$$

wählen, mit der Interpretation

$$(1) \quad \omega \cong \text{Menge der gewürfelten Augenzahlen.}$$

Damit wäre zum Beispiel

$$A := \text{“Augensumme} \geq 10\text{”} = \{ \{5\}, \{6\}, \{4, 6\}, \{5, 6\} \},$$

und unter der Gleichverteilungsannahme $P_a := U_{\Omega_a}$ erhielten wir

$$P_a(A) = \frac{\#A}{\#\Omega_a} = \frac{4}{6 + \binom{6}{2}} = \frac{4}{21}.$$

(b) **Standardmodell.** Alternativ könnte man

$$\Omega_b := \{1, \dots, 6\}^2 = \{(1, 1), (1, 2), \dots, (6, 6)\}$$

setzen, sich die beiden Würfel doch irgendwie unterscheidbar denkend, zum Beispiel den einen rot und den anderen grün, und dann mit der Interpretation

$$(2) \quad \omega = (\omega_1, \omega_2) \hat{=} \text{roter Würfel zeigt } \omega_1, \text{ grüner Würfel zeigt } \omega_2.$$

Damit wäre zum Beispiel

$$B := \text{“Augensumme } \geq 10\text{”} = \{(4, 6), (6, 4), (5, 5), (5, 6), (6, 5), (6, 6)\},$$

und unter der Gleichverteilungsannahme $P_b := U_{\Omega_b}$ erhielten wir

$$P_b(B) = \frac{\#B}{\#\Omega_b} = \frac{6}{36} = \frac{1}{6}.$$

□

Es ist natürlich prinzipiell unmöglich, innermathematisch zu entscheiden, welches der beiden obigen Modelle (Ω_a, P_a) und (Ω_b, P_b) besser die relativen Häufigkeiten bei realen Doppelwürfelwürfen beschreibt.⁷ In der Praxis, hier im Spezialfall von Gleichwahrscheinlichkeitsannahmen, folgt man meist dem heuristischen Prinzip:

Im Zweifelsfall wähle ein möglichst feines Modell.

Das soll heißen: Wähle Ω groß, also lieber wie in 3.19(b) als wie in 3.19(a), dabei im Modell möglichst viele Ergebnisse unterscheidend, auch wenn das in der Realität (etwa für einen Rot-Grün-Verwechsler) gar nicht möglich ist.

Es sei nicht verschwiegen, dass es “Gegenbeispiele” zur obigen Heuristik gibt: Bosonen, also gewisse Elementarteilchen wie zum Beispiel Photonen, verhalten sich angeblich analog zu dem Doppelwürfel nach 3.19(a), und nicht nach 3.19(b); siehe dazu Feller (1970, pp. 40–42).

Im Laufe der WR I und II lernen wir die folgenden, hier zur Übersicht angegebenen 8 Modellierungsmethoden kennen, wobei “Methode” durchweg als Euphemismus für “Annahme” gelesen werden kann.

⁷In der Zusammenfassung des hier einschlägigen §10, Abschnitt 1 bei von Mises (1990, p. 193): “Sätze der Logik oder Logistik, der reinen Mathematik und jeder axiomatisch aufgebauten Theorie sind nach der von L. Wittgenstein begründeten Auffassung »tautologisch«, d.h. sie besagen nichts über die Wirklichkeit, und sind lediglich Umformungen willkürlich festgesetzter Sprachregeln.”

3.20 Modellierungsmethoden.

- (a) Empirische Methode: In Beispiel 2.8 wähle $\Omega := \{1, \dots, 7\}$ und $P := \frac{685}{4800} \delta_1 + \frac{685}{4800} \delta_2 + \frac{687}{4800} \delta_3 + \frac{684}{4800} \delta_4 + \frac{688}{4800} \delta_5 + \frac{684}{4800} \delta_6 + \frac{687}{4800} \delta_7$.
 In Beispiel 2.9 wähle $\Omega := \{0, 1\}$ und $P := 0.484\delta_0 + 0.516\delta_1 = B_{0.516}$.
 In Beispiel 2.10 wähle $\Omega := \{0, 1\}$ und $P := B_{0.41}$ (Ergebnis Ihres Dozenten).
 In Beispiel 2.14 wähle $\Omega := \{0, 1\}$ und vielleicht $P := B_{\frac{1}{2}} = U_{\{0,1\}}$.
- (b) Gleichverteilungsannahme: Siehe Definition 3.17 und Beispiel 3.19. In den Beispielen 2.8, 2.9, 2.10, 2.14 wähle jeweils “aus Symmetriegründen” (?) $P := U_{\Omega}$.
- (c) Transformation mittels Zufallsgrößen: Siehe Kapitel 5.
- (d) Kopplung von Modellen: Siehe Kapitel 7.
- (e) Unabhängigkeitsannahme: Zum Beispiel Definition 9.1 und Satz 9.26.
- (f) Idealisierung durch Grenzwertbildung: Ein erstes Beispiel ist die Herleitung der speziellen Poisson-Verteilung P_1 in Beispiel 4.11. Siehe weiter 9.20(b) und 11.1.
- (g) Bedingen: Siehe Kapitel 6.
- (h) Charakterisierung durch plausible Eigenschaften: Herleitung der geometrischen Verteilungen in Aufgabe 10.45. \square

Eine Wahrscheinlichkeitsverteilung, selbst eine elementare auf \mathbb{N}_0 , ist ein kompliziertes mathematische Objekt, dessen möglichst einfache, durchaus vergrößernde Beschreibung eine spätere Hauptaufgabe der WR ist. Wir betrachten dazu hier im Vorgriff, neben dem bereits eingeführten Träger, nur die Maximalstellenmenge der zugehörigen Dichte, zumal diese in Beispiel 2.8 schon vorkam.

3.21 Definition (Argmax, argmax). Für Funktionen $f : \mathcal{X} \rightarrow \overline{\mathbb{R}}$ schreiben wir

$$\text{Argmax } f := \text{Menge aller Maximalstellen von } f.$$

Ist $\text{Argmax } f = \{x_0\}$ einelementig, so schreiben wir

$$\text{argmax } f := x_0.$$

Analog für Argmin und argmin. \square

Für die Dichte aus Beispiel 2.8 ist $\text{Argmax } f = \{5\}$ einelementig, also $\text{argmax } f = 5$. Dagegen ist in einem Binomialverteilungsbeispiel $\text{Argmax } b_{3, \frac{1}{2}} = \{1, 2\}$ zweielementig, also $\text{argmax } b_{3, \frac{1}{2}}$ nicht definiert; siehe hierzu systematischer Beispiel 3.24(c).

3.22 Bemerkung. Es sei \mathcal{X} eine Menge und $f \in \text{prob}(\mathcal{X})$. Dann ist $\text{Argmax } f$ eine endliche nichtleere Menge:

Es gilt $s := \sup_{x \in \mathcal{X}} f(x) > 0$ wegen $\sum_{x \in \mathcal{X}} f(x) = 1 > 0$. Im Fall von $\text{Argmax } f = \emptyset$, also $s > f(x)$ für $x \in \mathcal{X}$, wäre $\mathcal{X}_0 := \{x \in \mathcal{X} : f(x) > \frac{s}{2}\}$ eine unendliche Menge, und wir erhielten $1 = \sum_{x \in \mathcal{X}} f(x) \geq \sum_{x \in \mathcal{X}_0} \frac{s}{2} = \infty$. Wäre $\text{Argmax } f$ eine unendliche Menge, so erhielten wir $1 \geq \sum_{x \in \text{Argmax } f} s = \infty$. \square

3.23 Definition. Es sei $P \in \text{Prob}(\mathbb{Z})$ mit der Dichte f . Gibt es ein $x_0 \in \mathbb{Z}$ mit f wachsend auf $\{x \in \mathbb{Z} : x \leq x_0\}$ und fallend auf $\{x \in \mathbb{Z} : x \geq x_0\}$, so heißt P , oder auch f , **unimodal (auf \mathbb{Z})** mit der **Modalstellenmenge** $\text{Argmax } f$. \square

Eine Modalstellenmenge wie oben ist stets ein endliches nichtleeres Intervall in \mathbb{Z} , eventuell einelementig. Für Verteilungen auf Teilintervallen I von \mathbb{Z} , also beispielsweise $I = \mathbb{N}_0$ oder $I = \{0, \dots, n\}$, auf die Definition 3.23 nach Vereinbarung 3.11 ja anwendbar ist, braucht man die zugehörigen Dichten natürlich nur auf I zu betrachten.

3.24 Beispiele. Unimodal mit den angegebenen Maximalstellenmengen der Dichten sind:

(a) Dirac-Maße δ_a mit $a \in \mathbb{Z}$, einzige Maximalstelle a ,

(b) Poisson-Verteilungen P_λ für $\lambda \in [0, \infty[$,

$$\text{Argmax } p_\lambda = \{ \lceil \lambda - 1 \rceil, \lfloor \lambda \rfloor \} \setminus \{-1\} = \begin{cases} \{ \lfloor \lambda \rfloor \} & \text{falls } \lambda \notin \mathbb{N}, \\ \{ \lambda - 1, \lambda \} & \text{falls } \lambda \in \mathbb{N}, \end{cases}$$

(c) Binomialverteilungen für $n \in \mathbb{N}_0$ und $p \in [0, 1]$,

$$\text{Argmax } b_{n,p} = \{ \lceil (n+1)p - 1 \rceil, \lfloor (n+1)p \rfloor \} \setminus \{-1, n+1\}.$$

(Zahlenbeispiel, im Vorgriff auf die Herleitung in Satz 9.5 der Binomialverteilung als Verteilung der Trefferzahl in einer Bernoulli-Kette: Beim 64-fachen Würfelwurf ist 10 die wahrscheinlichste Anzahl der Sechsen.)

Beweis. (a) Trivial.

(b) Im degenerierten Fall $\lambda = 0$ ist $P_\lambda = \delta_0$ und die Behauptung folgt aus (a). Es sei nun $\lambda \in]0, \infty[$. Für $x \in \mathbb{N}_0$ gilt dann für die multiplikativen Zuwächse von p_λ die Äquivalenz

$$\frac{p_\lambda(x+1)}{p_\lambda(x)} = \frac{\lambda}{x+1} \left\{ \begin{array}{l} > \\ = \\ < \end{array} \right\} 1 \Leftrightarrow x \left\{ \begin{array}{l} < \\ = \\ > \end{array} \right\} \lambda - 1.$$

Also ist p_λ streng wachsend auf $\{0, \dots, \lceil \lambda - 1 \rceil\}$, denn liegen x und $x+1$ in dieser Menge, so folgt $x < \lambda - 1$ und damit $p_\lambda(x+1) > p_\lambda(x)$. Entsprechend ist p_λ konstant auf $\{\lceil \lambda - 1 \rceil, \lfloor \lambda \rfloor\}$ und streng fallend auf $\{\lfloor \lambda \rfloor, \dots\}$, und die Behauptung folgt.

(c) Übung. \square

Wir beenden dieses Kapitel mit einer wichtigen Ergänzung zu den Rechenregeln aus Satz 3.3: Wahrscheinlichkeiten $P(\bigcup_{i=1}^n A_i)$ nichtdisjunkter A_i sind oft schwer direkt zu handhaben, können aber nach Satz 3.25 mittels oft leichter zu bestimmenden Schnittwahrscheinlichkeiten berechnet oder abgeschätzt werden. Dabei sind von P nur die Inhaltseigenschaften und Endlichkeitsannahmen relevant, also weder die Normiertheit noch die die Totalität der Additivität, und tatsächlich wird Satz 3.25 im folgenden Kombinatorikkapitel 4 im Beweis von Satz 4.10 auf gewisse nichtnormierte P angewandt. Daher die allgemeine Formulierung:

3.25 Satz (Siebformel von Poincaré-Sylvester, Einschluss-Ausschluss-Formel). *Es sei P ein Inhalt auf einer Boole-Algebra \mathcal{A} . Für $n \in \mathbb{N}$ und $A_i \in \mathcal{A}$ mit $P(A_i) < \infty$ für $i \in \{1, \dots, n\}$ gilt dann*

$$(1) \quad P\left(\bigvee_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \dots + (-)^{n-1} P\left(\bigwedge_{i=1}^n A_i\right) \\ = \sum_{\ell=1}^n (-)^{\ell-1} \sum_{I \subseteq \{1, \dots, n\}, \#I=\ell} P\left(\bigwedge_{i \in I} A_i\right)$$

mit auf der rechten Seite abwechselnd zu großen und zu kleinen Partialsummen:

$$(2.1) \quad P\left(\bigvee_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i),$$

$$(2.2) \quad P\left(\bigvee_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j),$$

und so weiter⁸ (**Bonferroni-Ungleichungen**), siehe präziser unten (4)⁹.

Beweis. Für $n \in \mathbb{N}$ und $A_i \in \mathcal{A}$ mit $P(A_i) < \infty$ für $i \in \{1, \dots, n\}$ ist mit

$$(3) \quad S_\ell := S_\ell(A_1, \dots, A_n) := \sum_{I \subseteq \{1, \dots, n\}, \#I=\ell} P\left(\bigwedge_{i \in I} A_i\right) \quad \text{für } \ell \in \mathbb{N}$$

mur

$$(4) \quad P\left(\bigvee_{i=1}^n A_i\right) \begin{cases} \leq \\ \geq \end{cases} \sum_{\ell=1}^r (-)^{\ell-1} S_\ell \quad \text{für } r \in \mathbb{N}_0 \begin{cases} \text{ungerade} \\ \text{gerade} \end{cases}$$

zu zeigen, denn daraus ergibt sich auch die Identität (1) durch Betrachtung der $r \in \{n, n+1\}$ unter Verwendung von

$$(5) \quad S_\ell = 0 \quad \text{für } \ell > n.$$

Im Fall $r = 0$ folgt (4) trivial aus der Positivität von P .

In jedem Fall sind wegen $P(A_i) < \infty$ und der Isotonie 2.4(1) von P auch alle rechts in (1–4) auftretenden Inhalte endlich; die Endlichkeit von $P\left(\bigvee_{i=1}^n A_i\right)$ ergibt sich damit unten induktiv aus (2.1).

Für $n = 1$ gilt (4) wegen (5) und $S_1(A_1) = P(A_1)$.

Zur Vorbereitung des Induktionsschritts bemerken wir, dass 3.3(6) auch hier sinngemäß gilt, soll heißen

$$(6) \quad P(A \vee B) = P(A) + P(B) - P(AB) \quad \text{für } A, B \in \mathcal{A} \text{ mit } P(A), P(B) < \infty,$$

⁸Eine Floskel die, wie Heuser (2009, p. 48) bemerkt, oft anzeigt, dass man gerade nicht weiß wie es weiter geht. Zumindest in einem wissenschaftlichen Text sollte wie hier sogleich auf eine Präzisierung verwiesen werden.

⁹In einer Terminologie nach Pólya/Szegő (1970, p. 26, dort jedoch mit strikten Ungleichungen): Die Partialsummenfolge $(\sum_{\ell=1}^r (-)^{\ell-1} S_\ell)_{r \in \mathbb{N}_0}$ **umhüllt** $P\left(\bigvee_{i=1}^n A_i\right)$ **im engeren Sinne**.

also der Spezialfall $n = 2$ von (1), mit praktisch demselben Beweis wie für 3.3(6).¹⁰

Für den Induktionsschritt sei $n \in \mathbb{N}$, gelte (4) für beliebige $A_1, \dots, A_n \in \mathcal{A}$ endlichen Inhalts, und seien nun $A_1, \dots, A_{n+1} \in \mathcal{A}$ endlichen Inhalts. Dann gilt mit

$$(7) \quad B := \bigvee_{i=1}^n A_i, \text{ also auch } BA_{n+1} = \bigvee_{i=1}^n A_i A_{n+1},$$

für $r \in \mathbb{N}$, mit \geq falls r ungerade und \leq falls r gerade,

$$\begin{aligned} & \sum_{\ell=1}^r (-)^{\ell-1} S_\ell(A_1, \dots, A_{n+1}) \\ = & \sum_{\ell=1}^r (-)^{\ell-1} S_\ell(A_1, \dots, A_n) + P(A_{n+1}) - \sum_{\ell=1}^{r-1} (-)^{\ell-1} S_\ell(A_1 A_{n+1}, \dots, A_n A_{n+1}) \\ \left\{ \begin{array}{l} \geq \\ \leq \end{array} \right\} & P(B) + P(A_{n+1}) - P(BA_{n+1}) = P(B \vee A_{n+1}) = P\left(\bigvee_{i=1}^{n+1} A_i\right) \end{aligned}$$

wegen

$$\begin{aligned} S_1(A_1, \dots, A_{n+1}) &= S_1(A_1, \dots, A_n) + P(A_{n+1}), \\ S_\ell(A_1, \dots, A_{n+1}) &= S_\ell(A_1, \dots, A_n) + S_{\ell-1}(A_1 A_{n+1}, \dots, A_n A_{n+1}) \quad \text{für } \ell \geq 2 \end{aligned}$$

und einer Indexverschiebung im ersten Schritt, der Induktionsvoraussetzung und (7) im zweiten, und (6) im dritten. \square

Im Spezialfall dass die Schnittinhalte in (3) nur von $\#I$ abhängen, dass also $P(\bigwedge_{i \in I} A_i) = P(\bigwedge_{i \in J} A_i)$ für $I, J \subseteq \{1, \dots, n\}$ mit $\#I = \#J$ gilt, ergibt sich mit 4.4(3) die einfachere Darstellung

$$(8) \quad S_\ell = \binom{n}{\ell} P\left(\bigwedge_{i=1}^{\ell} A_i\right) \quad \text{für } \ell \in \{0, \dots, n\}.$$

3.26 Aufgabe. Es sei P ein Inhalt auf der Boole-Algebra \mathcal{A} . Für $n \in \mathbb{N}$ und $A_i \in \mathcal{A}$ mit $P(A_i) < \infty$ für $i \in \{1, \dots, n\}$ gilt dann¹¹

$$(1) \quad P\left(\bigvee_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} P(A_i A_{i+1}). \quad \square$$

¹⁰Hier merkt man also, dass man in einer systematischeren, aber eben auch zeitaufwändigeren Darstellung den Stoff etwas anders arrangieren sollte.

¹¹Das sich nach 3.5(c), 3.25(4), 3.26(1), 8.19(a) natürlich stellende Problem der Bestimmung optimaler unterer und oberer Schranken für die Wahrscheinlichkeit eines “Boole-Polynoms in Ereignissen A_1, \dots, A_n ” (Beispiele: $\bigvee_{i=1}^n A_i$, $\bigwedge_{i=1}^n A_i$, $A_1 A_2' \vee A_4$), gegebenen Schranken an die Wahrscheinlichkeiten von endlich vielen anderen, ist auf ein stets in endlich vielen Schritten lösbares lineares Optimierungsproblem zurückführbar, siehe Hailperin (1965).

Weitere Aufgaben

3.27 Aufgabe. Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. Which of these two alternatives is more probable?

- (1) Linda is a bank teller.
- (2) Linda is a bank teller and active in the feminist movement.

(a) Wie antworteten in einer Umfrage mehrheitlich Studenten (“undergraduates” der Stanford University und der University of British Columbia ohne Stochastik-Kenntnisse, aber auch Stanford-Studenten nach mehreren Stochastik-Vorlesungen, sowie auch Doktoranden)?

(b) Was sagen Sie nach Satz 3.3?

(c) Leider verpassen Sie ein nach langer Pause stattfindendes Abijahrgangstreffen, werden sich aber danach von einem Teilnehmer T so viel über die derzeitigen Berufe und Hobbies Ihrer Mitabiturienten $\nu \in \{1, \dots, n\}$ berichten lassen wie T dort erfahren hat. Welche der beiden Zahlen

$$p := \frac{1}{n} \sum_{\nu=1}^n x_{\nu} \quad \text{mit } x_{\nu} := (T \text{ sagt: "Beruf passt nicht zu } \nu\text{."}),$$

$$q := \frac{1}{n} \sum_{\nu=1}^n y_{\nu} \quad \text{mit } y_{\nu} := (T \text{ sagt: "Beruf passt nicht zu } \nu, \text{ aber Hobby."})$$

wird dann wohl die größere sein? □

3.28 Aufgabe (Will you play?). How about a small, friendly bet? Four playing cards, two red and two black, lie facedown on the table. Without looking, you choose any two cards. I'll bet you \$1 that you pick one red card and one black card. Shall we play?

Since you hesitate, I'll explain the natural fairness of this game. Only two outcomes are possible: the colors match or they do not. Your probability of winning is therefore one-half. The game is fair since neither of us has an advantage. Ready to put your money down?

You won't be fooled that easily, right? Your sense of caution is admirable. I'll explain further. Three, not two, outcomes are possible. Both cards you select are red, both are black, or one is black and the other red. The colors match in two of the three cases. Thus, your probability of winning is two-thirds. The game is in your favor. Now will you play?

Examine your chances more closely. When you choose your first card, it must be either red or black; the choice has no effect on your chances of winning. Now, of the remaining three cards, how many match the color of the card in your hand? Only one. Therefore, your probability of winning is only one-third.

Your suspicious nature is beginning to annoy me. Very well, here is my final explanation. Four (count 'em, four) possibilities exist. Consider the cards chosen one at a time. Both cards could be red; both could be black; or the first red and the second black; or the first black and the second red. The colors match in two of these four possible drawing sequences, giving you (again) 50% chance of winning the bet.

What is the correct answer?

(Aufgabe von R.T. Kurosaka, hier zitiert aus Bandelow (1989).) □

3.29 Aufgabe (Die Extrempunkte von $\text{Prob}(\Omega)$). Ist V ein \mathbb{R} -Vektorraum und $K \subseteq V$, so heißt K **konvex** falls die Implikation

$$x, y \in K, \lambda \in [0, 1] \Rightarrow \lambda x + (1 - \lambda)y \in K$$

gilt, und ein $e \in K$ heißt ein **Extrempunkt** von K falls die Implikation

$$x, y \in K, \lambda \in]0, 1[, \lambda x + (1 - \lambda)y = e \Rightarrow x = y = e$$

gilt. Beispiele in $V := \mathbb{R}^2$ von (konvexen) Mengen K mit zugehörigen Mengen E aller Extrempunkte: 1. $K := [0, 1]^2, E = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$; 2. $K :=]0, 1[^2, E = \emptyset$; 3. $K := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}, E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

Ist nun Ω eine nichtleere Menge, so ist $\text{Prob}(\Omega)$ eine gewisse Menge reellwertiger Funktionen auf 2^Ω , also eine Teilmenge des \mathbb{R} -Vektorraumes aller \mathbb{R} -wertigen Funktionen auf 2^Ω , und $\text{Prob}(\Omega)$ ist konvex, wie in Bemerkung 3.18 in allgemeinerer Form festgestellt. (Bis hier braucht noch nichts gezeigt zu werden.)

Was sind die Extrempunkte von $\text{Prob}(\Omega)$? □

3.30 Aufgabe (Diskrete Arcussinusverteilungen). Es sei $n \in \mathbb{N}$,

(1) $w_k := b_{2k, \frac{1}{2}}(k) \quad \text{für } k \in \mathbb{N}_0,$

(2) $p_k := w_k w_{n-k} \quad \text{für } k \in \{0, \dots, n\}.$

(a) Bestimmen Sie das Symmetrie- und Monotonieverhalten der Folge p_\bullet .

(b) Im Gleichverteilungsmodell für eine zufällige faire Münzwurffolge der Länge $2n$,

$$P := U_\Omega \text{ auf } \Omega := \{0, 1\}^{2n},$$

und mit

$$A_k := \left\{ \omega \in \Omega : \frac{1}{2k} \sum_{i=1}^{2k} \omega_i = \frac{1}{2}, \frac{1}{2\ell} \sum_{i=1}^{2\ell} \omega_i \neq \frac{1}{2} \text{ für } \ell \in \{k+1, \dots, n\} \right\} \text{ für } k \in \{0, \dots, n\}$$

gilt nach Feller (1970, p. 79, Theorem 1) oder Henze (2013 b, p. 16, Satz 1.4)

$$P(A_k) = p_k \quad \text{für } k \in \{0, \dots, n\},$$

d.h. p_k ist die Wahrscheinlichkeit dafür, dass nach genau $2k$ Münzwürfen das letzte Mal ein ‘‘Ausgleich’’ der Anzahlen der Nullen und Einsen beobachtet wird. Dass jedenfalls p_\bullet eine Dichte auf $\{0, \dots, n\}$ ist, wird in 4.8(e) gezeigt. Aufgrund des in den genannten Quellen diskutierten Konvergenzverhaltens für $n \rightarrow \infty$ wird $\sum_{k=0}^n p_k \delta_k$ als eine **diskrete Arkussinusverteilung** bezeichnet. Dies alles als gegeben annehmend:

Sind kleinere, mittlere, oder größere k am wahrscheinlichsten? Und um wieviel höher ist (ungefähr, bei großem n) der Maximalwert von p_\bullet verglichen mit dem zweitgrößten Wert? □

4 Elemente der Kombinatorik

Kombinatorik ist die Kunst des Zählens. Wir behandeln hier nur ihre elementarsten Aspekte als technisches Rüstzeug zur Berechnung von Wahrscheinlichkeiten in Gleichverteilungsmodellen.

4.1 Satz (Multiplikationsregel des Zählens.). *Es sei $k \in \mathbb{N}$ und es seien $j_1, \dots, j_k \in \mathbb{N}_0$. Eine Menge A von k -Tupeln $a = (a_1, \dots, a_k)$ sei wie folgt konstruierbar:*

1. *Es gebe j_1 Möglichkeiten, die erste Koordinate a_1 zu wählen.*
 2. *Ist a_1 gewählt, so gebe es j_2 Möglichkeiten, a_2 zu wählen. (Die Möglichkeiten dürfen dabei von a_1 abhängen, deren Anzahl j_2 jedoch nicht!)*
 3. *Sind a_1, a_2 gewählt, so gebe es j_3 Möglichkeiten, a_3 zu wählen.*
 - ⋮
 - k. *Sind a_1, \dots, a_{k-1} gewählt, so gebe es j_k Möglichkeiten, a_k zu wählen.*
- Dann gilt

$$(1) \quad \#(A) = j_1 \cdot \dots \cdot j_k = \prod_{\ell=1}^k j_\ell$$

Beweis. Induktion über k . □

4.2 Definition und Bemerkung (Faktorielle, Fakultäten, Binomialkoeffizienten).

(a) Für $x \in \mathbb{C}$ und $k \in \mathbb{N}_0$ setzt man neben $x^k := \prod_{i=1}^k x = \underbrace{x \cdot \dots \cdot x}_k$ noch

$$x^{\underline{k}} := \prod_{i=1}^k (x - i + 1) = x \cdot (x - 1) \cdot \dots \cdot (x - k + 1) \quad (\text{untere Faktorielle von } x),$$

$$x^{\overline{k}} := \prod_{i=1}^k (x + i - 1) = x \cdot (x + 1) \cdot \dots \cdot (x + k - 1) \quad (\text{obere Faktorielle von } x);$$

speziell ist also $x^0 = x^{\underline{0}} = x^{\overline{0}} = 1$. Für $n \in \mathbb{N}_0$ ist

$$n! := \prod_{i=1}^n i = n^{\underline{n}} = 1^{\overline{n}} = n \cdot (n - 1) \cdot \dots \cdot 2 \cdot 1$$

und speziell $0! = 1$.

(b) Für $x \in \mathbb{C}$ und $k \in \mathbb{N}_0$ definiert man

$$(1) \quad \binom{x}{k} := \frac{x^{\underline{k}}}{k!} = \frac{x \cdot (x - 1) \cdot \dots \cdot (x - k + 1)}{k \cdot (k - 1) \cdot \dots \cdot 1} = (-)^k \binom{k - x - 1}{k}.$$

Speziell für $n \in \mathbb{N}_0$ und $k \in \{0, \dots, n\}$ ist damit

$$(2) \quad \binom{n}{k} = \frac{n^{\underline{k}}}{k!} \cdot \frac{(n - k)!}{(n - k)!} = \frac{n!}{k!(n - k)!} = \frac{n^{\overline{n-k}}}{(n - k)!} = \binom{n}{n - k}.$$

Weiter setzt man noch

$$(3) \quad \binom{x}{k} := 0 \quad \text{für } x \in \mathbb{C} \text{ und } k \in -\mathbb{N} = \{-1, -2, \dots\}.$$

□

4.3 Definition (Tupel mit und ohne Wiederholungen, wachsende und streng wachsende Tupel, Teilmengen). Es sei M eine Menge und es sei $k \in \mathbb{N}_0$. Wir setzen

$$\begin{aligned} M^k &:= \{(x_1, \dots, x_k) : x_i \in M \text{ für } i \in \{1, \dots, k\}\} \\ &=: \text{Menge der } k\text{-Tupel aus } M \text{ (mit Wiederholung),} \end{aligned}$$

$$\begin{aligned} M_{\neq}^k &:= \{x \in M^k : x_i \neq x_j \text{ für } i, j \in \{1, \dots, k\} \text{ mit } i \neq j\} \\ &=: \text{Menge der } k\text{-Tupel aus } M \text{ ohne Wiederholung,} \end{aligned}$$

$$\binom{M}{k} := \{A \subseteq M : \#A = k\} \quad =: \text{Menge der } k\text{-Teilmengen aus } M$$

sowie, falls M durch eine Relation \leq total geordnet ist,

$$\begin{aligned} M_{\leq}^k &:= \{x \in M^k : x_1 \leq x_2 \leq \dots \leq x_k\} \\ &=: \text{Menge der } \mathbf{wachsenden } k\text{-Tupel aus } M, \end{aligned}$$

$$\begin{aligned} M_{<}^k &:= \{x \in M^k : x_1 < x_2 < \dots < x_k\} \\ &=: \text{Menge der } \mathbf{streng wachsenden } k\text{-Tupel aus } M. \end{aligned}$$

Im degenerierten Fall $k = 0$ erhalten wir

$$M^0 = M_{\neq}^0 = M_{\leq}^0 = M_{<}^0 = \{()\}$$

mit $() := \emptyset =$ leeres Tupel = einziges 0-Tupel, sowie $\binom{M}{0} = \{\emptyset\}$. □

Die obigen vier Arten von Tupeln heißen in der Literatur oft, zum Beispiel bei Henze (2021, pp. 52–53), **Permutationen mit Wiederholung**, **Permutationen ohne Wiederholung**, **Kombinationen mit Wiederholung**, **Kombinationen ohne Wiederholung**, jeweils **aus** M . Dabei müsste es statt “mit Wiederholung” eigentlich “ohne Wiederholungsverbot” heißen.

4.4 Satz (Grundformeln der Kombinatorik). *Es seien $n, k \in \mathbb{N}_0$ und es sei M eine Menge mit $\#M = n$. Dann gilt*

$$\begin{aligned} (1) \quad & \#(M^k) = n^k, \\ (2) \quad & \#(M_{\neq}^k) = n^{\underline{k}}, \\ (3) \quad & \#\binom{M}{k} = \binom{n}{k} \end{aligned}$$

und, falls M durch \leq total geordnet ist,

$$\begin{aligned} (4) \quad & \#(M_{\leq}^k) = \binom{n+k-1}{k}, \\ (5) \quad & \#(M_{<}^k) = \binom{n}{k}. \end{aligned}$$

Beweis. Im degenerierten Fall $k = 0$ steht jeweils “1=1”. Es gelte also ab jetzt $k \geq 1$.

(1): Anwendung der Multiplikationsregel 4.1 mit $j_1 = \dots = j_k = n$.

(2): Multiplikationsregel 4.1 mit $j_1 = n, j_2 = (n - 1)_+, \dots, j_k = (n - k + 1)_+$.

(4):¹ Zurückführung auf (5) mit der **Bijektionsmethode**: Ohne Einschränkung sei $M = \{1, \dots, n\}$ mit der üblichen Ordnung. Dann ist die Abbildung

$$\begin{aligned} M_{\leq}^k &\ni (x_1, x_2, \dots, x_k) \\ &\mapsto (x_1, x_2 + 1, \dots, x_k + k - 1) \\ &\in \{1, \dots, n + k - 1\}_{<}^k \end{aligned}$$

eine Bijektion; folglich ist $\#(M_{\leq}^k) = \#(\{1, \dots, n + k - 1\}_{<}^k)$, und die Anwendung von (5) mit $n + k - 1$ statt n liefert die Behauptung.

(5): Wieder 4.1 kombiniert mit dem “**Schäferprinzip**”:

$$\begin{aligned} n^k &= \#(M_{\neq}^k) && \text{[wegen (2)]} \\ &= \#(M_{<}^k) \cdot k! && \text{[4.1 mit } k = 2, \text{ und (2) mit } n = k] \end{aligned}$$

(3): Ohne Einschränkung sei $M = \{1, \dots, n\}$. Dann ist die Abbildung

$$M_{<}^k \ni (x_1, \dots, x_k) \mapsto \{x_1, \dots, x_k\} \in \binom{M}{k}$$

eine Bijektion, und mit (5) folgt die Behauptung. □

4.5 Beispiele. (a) 13er Wette. Bei der 13er Wette des Fußballtotos besteht ein Tipp aus der jeweiligen Vorhersage des Ausgangs $1 \hat{=}$ Heimsieg oder $0 \hat{=}$ Unentschieden oder $2 \hat{=}$ Gastsieg von 13 vorher festgelegten Fußballspielen. Wieviele verschiedene Tipps gibt es? (Lösung: 1 594 323.)

(b) Lotto “6 aus 49 mit Zusatzzahl”. Ein Tipp beim genannten Lotto² entsteht durch das Ankreuzen von genau sechs aus den Zahlen von 1 bis 49. Ein Ziehungsergebnis besteht aus genau sechs der Zahlen von 1 bis 49, sowie aus einer von diesen eigentlichen Lottozahlen verschiedenen Zusatzzahl aus den verbleibenden 43 Zahlen.

Wieviele verschiedene Tipps gibt es? (Lösung: 13 983 816.)

Wieviele verschiedene Ziehungsergebnisse gibt es? (Lösung: 601 304 088.)

Wie groß ist die Wahrscheinlichkeit für “5 richtige mit Zusatzzahl” bei einem Tipp und gleichverteiltem Ziehungsergebnis? (Lösung: $\frac{1}{2330636}$.)

(c) Mehrere gleiche Würfel. n gleiche Würfel werden gleichzeitig geworfen. Wieviele unterscheidbare Ergebnisse gibt es? (Lösung: $\binom{5+n}{n}$.)

Vorsicht, Lösungswegskizzen stehen oben auf der nächsten Seite!

¹Hier vor (5), da hier die Bijektionsmethode explizit ausgeführt wird, während sie im Beweis von (5) (der (4) nicht verwendet!) als selbstverständlich angenommen wird.

²In dieser Form in der BRD betrieben von 1956 bis 1991. Später Varianten mit “Superzahl”.

$1\,594\,323 = 3^{13}$, 4.4(1) mit $M = \{0, 1, 2\}$ und $k = 13$.

$13\,983\,816 = \binom{49}{6}$, 4.4(3) mit $M = \{1, \dots, 49\}$, $k = 6$.

$601\,304\,088 = \binom{49}{6} \cdot 43$, 4.1 mit $a_1 \in \binom{\{1, \dots, 49\}}{6}$ und $a_2 \in \{1, \dots, 49\} \setminus a_1$, 4.4(3).

$\frac{1}{2\,330\,636} = \binom{6}{5} \cdot 43 \cdot 1 / (\binom{49}{6} \cdot 43)$, Gleichverteilungsannahme, im Zähler 4.1 mit $a_1 \in \binom{\text{Tipp}}{5}$, $a_2 \in \{1, \dots, 49\} \setminus \text{Tipp}$, eigentliche Lottozahlen dann $= a_1 \cup \{a_2\}$, Zusatzzahl $a_3 \in \text{Tipp} \setminus a_1$.

$\binom{5+n}{n} = \binom{6+n-1}{n} = \#(\{1, \dots, 6\}_{\leq}^n)$, Würfel nach wachsenden Augenzahlen anordnen. □

4.6 Aufgabe (Wieviele fünfstellige Zahlen?). Wieviele fünfstellige Zahlen (positiv, mit eventuellen führenden Nullen hingeschrieben, also von 00000 bis 99999) gibt es, bei denen

- (a) alle Ziffern gleich sind,
- (b) vier Ziffern einander gleich sind und die verbleibende davon verschieden ist,
- (c) drei Ziffern einander gleich sind, während die beiden anderen davon und untereinander verschieden sind,
- (d) drei Ziffern gleich sind und die beiden anderen untereinander gleich, aber von den ersteren verschieden sind,
- (e) zwei Ziffern gleich sind und die drei anderen untereinander und von den ersteren verschieden sind,
- (f) zwei Paare gleicher Ziffern auftreten, die untereinander und von der fünften Ziffer verschieden sind,
- (g) fünf paarweise verschiedene Ziffern auftreten?

Empfohlen wird eine Probe durch Summation! □

4.7 Aufgabe. (a) Gitterpunktwege. Wieviele Wege gibt es vom Punkt $(0, 0)$ zu einem Punkt $(x, y) \in \mathbb{N}_0^2$, wenn nur Schritte der Länge 1 nach rechts oder nach oben erlaubt sind?

(b) **Kranke Bäume.** Entlang einer Straße stehen 30 Bäume. Genau 4 davon haben eine bestimmte Krankheit. Die 4 kranken Bäume sind benachbart. Kann das Benachbartsein Zufall sein, oder ist hier ein Indiz dafür zu sehen, dass die Krankheit sich vielleicht durch Ansteckung ausgebreitet hat? □

4.8 Aufgabe. (a) Geben Sie einen kombinatorischen Beweis für die **Vandermonde'sche Faltungsformel** für ganzzahlige positive Argumente

$$(1) \quad \sum_{k=0}^n \binom{m}{k} \binom{w}{n-k} = \binom{m+w}{n} \quad \text{für } n, m, w \in \mathbb{N}_0.$$

Hinweis: Auf wieviele Weisen läßt sich aus m männlichen und w weiblichen Studenten ein n -köpfiger Ausschuss bilden?

(b) Die in 3.12 eingeführten Funktionen $h_{n,r,b}$ sind tatsächlich Wahrscheinlichkeitszähldichten, mit den dort angegebenen Trägern.

(c) Aus (1) folgt die allgemeine Vandermonde'sche Faltungsformel

$$(2) \quad \sum_{k=0}^n \binom{x}{k} \binom{y}{n-k} = \binom{x+y}{n} \quad \text{für } n \in \mathbb{N}_0 \text{ und } x, y \in \mathbb{C}$$

mit Hilfe des Eindeutigkeitsatzes für Polynomfunktionen.

(d) Wird also wegen (c) für jede Wahl von $n \in \mathbb{N}_0$ und $r, b \in \mathbb{R}$ durch $f(x) := \binom{r}{x} \binom{b}{n-x} / \binom{r+b}{n}$ für $x \in \{0, \dots, n\}$ eine Wahrscheinlichkeitsdichte definiert?

(e) Die Identität (2) mit $x := y := -\frac{1}{2}$ liefert einen Beweis dafür, dass die Funktion p_\bullet aus 3.30(2) eine Wahrscheinlichkeitsdichte ist. \square

Das Ergebnis der folgenden Aufgabe wird in Aufgabe 5.17 zur Berechnung der Verteilungen von Würfelaugensummen benutzt.

4.9 Aufgabe (Mächtigkeiten diskreter Simplexe). Für $n, k \in \mathbb{N}_0$ gilt

$$(1) \quad \#\{x \in \mathbb{N}_0^n : \sum_{i=1}^n x_i = k\} = \binom{n+k-1}{k},$$

$$(2) \quad \#\{x \in \mathbb{N}_0^n : \sum_{i=1}^n x_i \leq k\} = \binom{n+k}{k}.$$

Hinweis: Es bezeichne $S_{n,k}$ die Menge links in (1), $T_{n,k}$ die links in (2). Geeignete Bijektionen dieser Mengen führen (1) auf Satz 4.4 zurück, und (2) auf (1). \square

Einen ganz anderen Beweis von 4.9(2) bringt Aufgabe 9.33(b).

Eine weitere elementare Technik der Kombinatorik, neben Multiplikationsregel, Bijektionsmethode, Schäferprinzip, und den vier Grundformeln, liefert die Siebformel:

4.10 Satz (Siebformel für Anzahlen). Für endliche Mengen A_1, \dots, A_n gilt

$$\begin{aligned} \#\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \#(A_i) - \sum_{(i,j) \in \{1, \dots, n\}^2_{<}} \#(A_i \cap A_j) + \sum_{(i,j,k) \in \{1, \dots, n\}^3_{<}} \#(A_i \cap A_j \cap A_k) \\ &\quad - + \dots + (-)^{n-1} \#\left(\bigcap_{i=1}^n A_i\right) \\ &= \sum_{\ell=1}^n (-)^{\ell-1} \sum_{I \subseteq \{1, \dots, n\}, \#I=\ell} \#\left(\bigcap_{i \in I} A_i\right) \end{aligned}$$

mit auf der rechten Seite abwechselnd zu großen und zu kleinen Partialsummen.

Beweis. Allgemeine Siebformel 3.25 spezialisiert auf den Inhalt $P := \#$ auf der Potenzmengenalgebra $2^{\mathcal{X}}$ mit $\mathcal{X} := \bigcup_{i=1}^n A_i$. \square

4.11 Beispiel (Rencontre-Problem). An einer Theatergarderobe werden nach Ende der Vorstellung die abgegebenen Mäntel rein zufällig zurückgegeben. Wieviele Besucher erhalten ihren eigenen Mantel?

Mathematische Berichtigung und Präzisierung der Fragestellung: Es sei $n \in \mathbb{N}_0$ und $k \in \{0, \dots, n\}$. Mit welcher Wahrscheinlichkeit hat eine zufällige Permutation

$$\omega \in \{1, \dots, n\}_{\neq}^n =: \Omega$$

der n -elementigen Menge $\{1, \dots, n\}$ genau k Fixpunkte? Zu berechnen ist also im Modell (Ω, U_Ω) die Wahrscheinlichkeit $p(k) := U_\Omega(A_k)$ von

$$A_k := \left\{ \omega \in \Omega : \sum_{i=1}^n (\omega_i = i) = k \right\}.$$

(a) Wir behandeln zunächst den Fall $k = 0$ durch Übergang zum Komplement und Anwendung der Siebformel für Anzahlen: Es ist

$$A_0^c = \bigcup_{i=1}^n B_i \quad \text{mit } B_i := \{ \omega \in \Omega : \omega_i = i \}.$$

Für $I \subseteq \{1, \dots, n\}$ mit $\#I = \ell$ ist nun

$$\# \left(\bigcap_{i \in I} B_i \right) = \# \left(\left(\{1, \dots, n\} \setminus I \right)_{\neq}^{n-\ell} \right) = (n - \ell)!$$

wegen 4.4(2) im zweiten Schritt, und damit ist nach 4.10 und 3.25(8) folglich

$$\begin{aligned} (1) \quad \#(A_0^c) &= \sum_{\ell=1}^n (-1)^{\ell-1} \sum_{I \subseteq \{1, \dots, n\}, \#I=\ell} \# \left(\bigcap_{i \in I} B_i \right) \\ &= \sum_{\ell=1}^n (-1)^{\ell-1} \binom{n}{\ell} (n - \ell)! = \sum_{\ell=1}^n (-1)^{\ell-1} \frac{n!}{\ell!} \end{aligned}$$

die Anzahl der Permutationen mit mindestens einem Fixpunkt. Also ist

$$(2) \quad \#(A_0) = \#\Omega - \#(A_0^c) = n! \sum_{\ell=0}^n \frac{(-1)^\ell}{\ell!}$$

die Anzahl der fixpunktfreien Permutationen, und damit ist

$$p(0) = U_\Omega(A_0) = \frac{\#A_0}{\#\Omega} = \sum_{\ell=0}^n \frac{(-1)^\ell}{\ell!},$$

also praktisch gleich e^{-1} ; zum Beispiel ergibt sich für $n \geq 5$ mit den Eigenschaften von **Leibnizreihen**³ oder den Bonferroni-Ungleichungen $\sum_{\ell=0}^5 \frac{(-1)^\ell}{\ell!} \leq p(0) <$

³Gemeint sind Reihen $(\sum_{k=k_0}^n a_k)_{n \geq k_0}$ mit reellen a_k mit $a_k a_{k+1} \leq 0$ und $|a_k| \downarrow 0$.

$\sum_{\ell=0}^4 \frac{(-)^\ell}{\ell!} = 0.375$, wegen $\frac{1}{5!} = \frac{1}{120} < 0.01$ also auf zwei Nachkommastellen korrekt gerundet $p(0) = 0.37$ unabhängig von n , und damit eine überraschend große Wahrscheinlichkeit von $1 - p(0) = 0.63$ dafür, dass mindestens ein Theaterbesucher seinen eigenen Mantel erhält.

(b) Wir schreiben ab jetzt genauer $\Omega_n, A_{n,k}, p_n(k)$ statt $\Omega, A_k, p(k)$. Für $k \in \{0, \dots, n\}$ ist damit

$$(3) \quad A_{n,k} = \bigcup_{I \subseteq \{1, \dots, n\}, \#I=k} \left\{ \omega \in \Omega_n : \omega_i = i \text{ für } i \in I, \omega_i \neq i \text{ für } i \in \{1, \dots, n\} \setminus I \right\}$$

eine Vereinigung von $\binom{n}{k}$ paarweise disjunkten Mengen der Mächtigkeit $\#A_{n-k,0}$, und mit (2) angewandt auf $n - k$ statt n ergibt sich

$$\#A_{n,k} = \binom{n}{k} (n - k)! \sum_{\ell=0}^{n-k} \frac{(-)^\ell}{\ell!} = \frac{n!}{k!} \sum_{\ell=0}^{n-k} \frac{(-)^\ell}{\ell!}$$

als Anzahl der Permutationen von $\{1, \dots, n\}$ mit genau k Fixpunkten, und damit

$$(4) \quad p_n(k) = \mathbb{U}_{\Omega_n}(A_{n,k}) = \frac{\#A_{n,k}}{\#\Omega_n} = \frac{1}{k!} \sum_{\ell=0}^{n-k} \frac{(-)^\ell}{\ell!}$$

als Wahrscheinlichkeit dafür, dass genau k Besucher ihren Mantel erhalten.

Offenbar ist $A_{n,\bullet}$ eine indizierte Quasipartition⁴ von Ω_n , also $\sum_{k=0}^n p_n(k) = \sum_{k=0}^n \mathbb{U}_{\Omega_n}(A_{n,k}) = 1$. Folglich ist p_n eine Wahrscheinlichkeitszähldichte auf $\{0, \dots, n\}$ oder, mit $p_n(k) := 0$ für $k \geq n + 1$, auch auf \mathbb{N}_0 ; das zugehörige Wahrscheinlichkeitsmaß nennen wir **Rencontre-Verteilung** mit dem Parameter n . Offenbar gilt

$$(5) \quad \lim_{n \rightarrow \infty} p_n(k) = \frac{e^{-1}}{k!} = p_1(k) \quad \text{für } k \in \mathbb{N}_0,$$

wobei $p_1 = p_{\lambda=1}$ die in 3.12 und Aufgabe 3.15 eingeführte Zähldichte der Poisson-Verteilung $P_1 = P_{\lambda=1}$ bezeichnet. Diese spezielle Poisson-Verteilung kann also wegen (5) interpretiert werden als approximative Verteilung der Anzahl der Fixpunkte einer Zufallspermutation einer großen endlichen Menge, ein Beispiel für die Modellierungsmethode 3.20(f). \square

Später lernen wir genauere Approximationskonzepte als die Konvergenz von Dichten wie in (5) kennen.

⁴Keine Partition, denn zwar gilt $A_{n,k} \neq \emptyset$ für jedes $k \neq n - 1$, was man etwa mit (3) und Betrachtung zyklischer Permutationen auf I^c sieht, aber $A_{n,n-1} = \emptyset$.

Weitere Aufgaben

4.12 Aufgabe (Lottozwillinge). Auf den ersten Blick überraschend häufig enthalten die real gezogenen Lottozahlen “6 aus 49” einen Zwilling $(j, j + 1)$. Wie wahrscheinlich ist dies?

Hinweis: Betrachten Sie die Menge aller Tupel aus $\{1, 2, \dots, 49\}_<^6$ welche keine Zwillinge enthalten, und wenden Sie darauf die Bijektionsmethode ähnlich wie im Beweis von 4.4(4) an. \square

4.13 Aufgabe. Wie viele Zahlen zwischen 1 und 3 000 002 sind durch 4 oder 6 oder 15 teilbar? \square

4.14 Aufgabe. Für endlichdimensionale Untervektorräume eines Vektorraumes gilt bekanntlich

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V).$$

Also auch, wie bei der Siebformel für drei Mengen,

$$\begin{aligned} \dim(U + V + W) = & \dim(U) + \dim(V) + \dim(W) \\ & - \dim(U \cap V) - \dim(U \cap W) - \dim(V \cap W) \\ & + \dim(U \cap V \cap W), \end{aligned}$$

oder? \square

4.15 Aufgabe (Mäntel und Hüte). In Abwandlung von Beispiel 4.11 werden nun unabhängig von den Mänteln auch die Hüte zurückgeben, d.h. alle Paare von Permutationen seien gleich wahrscheinlich. Mit welcher Wahrscheinlichkeit erhält dann mindestens einer der n Besucher seinen Mantel und seinen Hut zurück? Limes für $n \rightarrow \infty$? \square

4.16 Aufgabe (Bitte platznehmen.). (a) n unterscheidbare Stühle stehen um einen runden Tisch. Auf wie viele Weisen können k unterscheidbare Personen platznehmen, wenn keine zwei benachbarten Stühle besetzt sein dürfen?

(b) Auf wieviele Weisen können dagegen k gleichartige Gedecke nachbarfrei vor die n Stühle plaziert werden?

(c) Und auf wie viele Weisen können k unterscheidbare Personen nachbarfrei in einem Kettenkarussell mit n abgesehen von ihrer Lage zueinander ununterscheidbaren Plätzen fahren?

(d) Funktioniert Ihre Reduktion von (b) bzw. (c) auf (a) auch bei ununterscheidbaren Plätzen und zugleich ununterscheidbaren Personen? \square

5 Zufallsgrößen und ihre Verteilungen

Zufallsgrößen sind mathematische Hilfsmittel zur bequemen Beschreibung von Ereignissen und zur Definition neuer Wahrscheinlichkeitsmaße.

5.1 Definition. Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum, \mathcal{X} eine Menge, und $X : \Omega \rightarrow \mathcal{X}$ eine Funktion. Dann heißt X \mathcal{X} -wertige **Zufallsgröße**¹ auf (Ω, \mathbb{P}) , und die durch

$$(1) \quad P(B) := \mathbb{P}(X^{-1}[B]) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) \quad \text{für } B \subseteq \mathcal{X}$$

auf $2^{\mathcal{X}}$ definierte Funktion P heißt **Verteilung** von X (bezüglich \mathbb{P}), oder **Bildmaß** von \mathbb{P} unter X . □

5.2 Satz. In der Situation von Definition 5.1 ist P ein Wahrscheinlichkeitsmaß auf \mathcal{X} , dessen Dichte p gegeben ist durch

$$(1) \quad p(x) = \sum_{\omega \in X^{-1}[\{x\}]} \mathbb{P}(\{\omega\}) \quad \text{für } x \in \mathcal{X}.$$

Beweis. Die Positivität von P gemäß Definition 3.2 ergibt sich direkt mit 5.1(1) aus der von \mathbb{P} . Ist \mathcal{B} eine Quasipartition von \mathcal{X} , so ist $(X^{-1}[B] : B \in \mathcal{B})$ eine indizierte Quasipartition von Ω , und wir erhalten damit, unter Verwendung der totalen Additivität und Normiertheit von \mathbb{P} gemäß 3.3(11) im zweiten Schritt,

$$\sum_{B \in \mathcal{B}} P(B) = \sum_{B \in \mathcal{B}} \mathbb{P}(X^{-1}[B]) = 1;$$

also ist P ein Wahrscheinlichkeitsmaß auf \mathcal{X} . Für $x \in \mathcal{X}$ gilt

$$p(x) = P(\{x\}) = \mathbb{P}(X^{-1}[\{x\}]) = \sum_{\omega \in X^{-1}[\{x\}]} \mathbb{P}(\{\omega\})$$

wegen 3.6(1) im ersten Schritt, und 3.6(4) im letzten; also gilt (1). □

Der obige Wahrscheinlichkeitsraum (\mathcal{X}, P) heißt **Bildraum** von (Ω, \mathbb{P}) unter X . Die Verteilung P bezeichnen wir mit $X \square \mathbb{P}$, also, in Wiederholung der Definition 5.1(1),

$$(2) \quad (X \square \mathbb{P})(B) = \mathbb{P}(X^{-1}[B]) \quad \text{für } B \subseteq \mathcal{X}.$$

In der Literatur findet man statt $X \square \mathbb{P}$ unter anderem die Schreibweisen $\mathcal{L}(X, \mathbb{P})$, $\mathbb{P}X^{-1}$ und $X(\mathbb{P})$.

¹Leider meist **Zufallsvariable** genannt, eine unnötige Übersetzung aus dem Englischen (*random variable*) oder Französischen (*variable aléatoire*). Wir folgen mit *Zufallsgröße* lieber zum Beispiel Kolmogoroff (1933) als der Mehrheit.

Der spezielle Buchstabe “ \mathbb{P} ” wird üblicherweise nur für ein im jeweiligen Kontext “grundlegendes” Wahrscheinlichkeitsmaß verwendet; unüblich wäre zum Beispiel: “Es sei (Ω, P) ein Wahrscheinlichkeitsraum, $X : \Omega \rightarrow \mathcal{X}$, und $\mathbb{P} := X \square P$ ”.

Weiter bezeichnen man oft die Verteilung der Zufallsgröße X mit P und die einer weiteren Zufallsgröße Y dann mit Q , oder die erste mit P_X und die zweite mit P_Y .

Elemente $x \in \mathcal{X}$ heißen auch **Realisierungen** von X . Übliche Schreibweisen für $B \subseteq \mathcal{X}$ und $x \in \mathcal{X}$ sind

$$\begin{aligned} \{X \in B\} &:= X^{-1}[B], \\ \mathbb{P}(X \in B) &:= \mathbb{P}(\{X \in B\}) = P(B), \\ \{X = x\} &:= X^{-1}[\{x\}] = \{\omega \in \Omega : X(\omega) = x\}, \\ \mathbb{P}(X = x) &:= \mathbb{P}(X^{-1}[\{x\}]) = P(\{x\}) \end{aligned}$$

und analoge mehr. Ist zum Beispiel speziell $\mathcal{X} \subseteq \mathbb{R}$, so setzen wir

$$\begin{aligned} \{X \leq x\} &:= \{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}[] - \infty, x], \\ \mathbb{P}(X \leq x) &:= \mathbb{P}(\{X \leq x\}) = P(] - \infty, x]. \end{aligned}$$

Statt “ X hat die Verteilung P ”, also $X \square \mathbb{P} = P$, schreiben wir oft $X \sim P$ und sagen auch “ X ist nach P verteilt”. Zwei Zufallsgrößen X_1 und X_2 heißen **identisch verteilt**² wenn ihre Verteilungen P_1 und P_2 gleich sind. Man schreibt dafür auch $X_1 \sim X_2$, gelesen als “ X_1 ist verteilt wie X_2 ”. Dabei ist der Fall zugelassen, dass die X_i auf verschiedenen Wahrscheinlichkeitsräumen (Ω_i, \mathbb{P}_i) definiert sind, also $X_1 \sim X_2 :\Leftrightarrow X_1 \square \mathbb{P}_1 = X_2 \square \mathbb{P}_2$.

Ist $T : \mathcal{X} \rightarrow \mathcal{Y}$ eine Funktion, so wird die \mathcal{Y} -wertige Zufallsgröße $T \circ X$ oft mit $T(X)$ bezeichnet, wie zum Beispiel $\sin(X) := \sin \circ X$ und $|x| := |\cdot| \circ X$ in:

5.3 Aufgabe (Relationen zwischen Ereignissen und Wahrscheinlichkeiten). Es seien X und Y \mathbb{R} -wertige Zufallsgrößen auf (Ω, \mathbb{P}) . Welche Relation besteht jeweils zwischen den zwei Wahrscheinlichkeiten

- (a) $\mathbb{P}(\sin(X) = 0)$ und $\mathbb{P}(X \in \pi\mathbb{Z})$,
- (b) $\mathbb{P}(X \geq 2)$ und $\mathbb{P}(X^2 \geq 4)$,
- (c) $\mathbb{P}(|X| \leq 1, |Y| \leq 2)$ und $\mathbb{P}(|X + Y| \leq 3)$? □

5.4 Bemerkung. Es seien X und Y \mathcal{X} -wertige Zufallsgrößen auf demselben Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Trivialerweise gilt dann die Implikation

$$(1) \quad X = Y \quad \Rightarrow \quad X \sim Y$$

(aber im allgemeinen nicht die Umkehrung, siehe etwa X_1, X_2 in Beispiel 5.6).

²Die rein sprachlich äquivalente Bezeichnung “gleich verteilt” ist zu vermeiden, um Verwechslungen mit den Gleichverteilungen aus Definition 3.17 vorzubeugen.

Manchmal, etwa in 7.11(9), gilt statt L.S.(1) jedoch nur

$$(2) \quad \mathbb{P}(X \neq Y) = 0;$$

man nennt in diesem Fall X und Y **fast sicher** gleich, und man schreibt dafür “ $X = Y$ f.s.” oder “ $X =_{\mathbb{P}} Y$ ” oder ähnliches. Damit verallgemeinert sich (1) zu

$$(3) \quad X =_{\mathbb{P}} Y \Rightarrow X \sim Y,$$

denn mit L.S.(3) gilt für $B \subseteq \mathcal{X}$ wegen

$$\{X \in B\} \subseteq \{Y \in B\} \cup \{X \neq Y\}$$

dann, mit 3.3(13) im ersten Schritt und mit (2) im zweiten,

$$\mathbb{P}(X \in B) \leq \mathbb{P}(Y \in B) + \mathbb{P}(X \neq Y) = \mathbb{P}(Y \in B),$$

und mit Rollentausch auch $\mathbb{P}(Y \in B) \leq \mathbb{P}(X \in B)$; also gilt dann R.S.(3). \square

Später betrachten wir oft Verteilungen P von Zufallsgrößen X ohne einen Grundraum (Ω, \mathbb{P}) oder X explizit anzugeben, siehe etwa Aufgabe 5.17; dies ist dann gerechtfertigt durch:

5.5 Bemerkung. Es sei (\mathcal{X}, P) ein Wahrscheinlichkeitsraum. Dann gibt es eine Zufallsgröße mit der Verteilung P , nämlich zum Beispiel $X := \text{id}_{\Omega}$ auf $(\Omega, \mathbb{P}) := (\mathcal{X}, P)$, genannt die **kanonische Zufallsgröße** zu P . \square

5.6 Beispiel (Zweifacher Würfelwurf). In dessen Standardmodell (Ω, \mathbb{P}) gemäß Beispiel 3.19(b), also $\Omega = \{1, \dots, 6\}^2$ und $\mathbb{P} = U_{\Omega}$, sind die durch

$$X_1(\omega) := \omega_1, \quad X_2(\omega) := \omega_2 \quad \text{für } \omega \in \Omega$$

definierten Koordinatenprojektionen zwei \mathbb{R} -wertige Zufallsgrößen, genannt “Augenzahlen der einzelnen Würfel”. Damit ist mit

$$S := X_1 + X_2$$

(punktweise zu lesen, wie auch sonst bei Funktionen üblich) eine die Augensumme modellierende Zufallsgröße bequem hinschreibbar, ebenso wie das Ereignis

$$\text{“Pasch”} := \{X_1 = X_2\} := \{\omega \in \Omega : X_1(\omega) = X_2(\omega)\}.$$

Für $x \in \{1, \dots, 6\}$ ist

$$\mathbb{P}(X_1 = x) = \mathbb{P}(\{\omega \in \Omega : \omega_1 = x\}) = \frac{6}{36} = \frac{1}{6}.$$

Folglich ist $X_1 \square \mathbb{P} = U_{\{1, \dots, 6\}}$ und analog $X_2 \square \mathbb{P} = U_{\{1, \dots, 6\}}$. Die Zufallsgrößen X_1 und X_2 sind also sowohl gleich verteilt (identisch verteilt, $X_1 \square \mathbb{P} = X_2 \square \mathbb{P}$) als auch

gleichverteilt (uniform verteilt, $X_i \square \mathbb{P} = U_{\{1,\dots,6\}}$). Zur Bestimmung der Verteilung der Augensumme S berechnen wir für $s \in \{2, \dots, 12\}$

$$\mathbb{P}(S = s) = \frac{\#(S^{-1}[\{s\}])}{36} = \left\{ \begin{array}{ll} 1/36 & \text{falls } s = 2, \\ 2/36 & \text{falls } s = 3, \\ 3/36 & \text{falls } s = 4, \\ 4/36 & \text{falls } s = 5, \\ 5/36 & \text{falls } s = 6, \\ 6/36 & \text{falls } s = 7, \\ 5/36 & \text{falls } s = 8, \\ 4/36 & \text{falls } s = 9, \\ 3/36 & \text{falls } s = 10, \\ 2/36 & \text{falls } s = 11, \\ 1/36 & \text{falls } s = 12 \end{array} \right\} = \frac{6 - |7 - s|}{36}$$

wegen $S^{-1}[\{2\}] = \{(1, 1)\}$, $S^{-1}[\{3\}] = \{(1, 2), (2, 1)\}$, \dots , $S^{-1}[\{12\}] = \{(6, 6)\}$. Die Verteilung von S wird als eine **diskrete Dreiecksverteilung** bezeichnet.

Für $n \in \mathbb{N}$ statt 2 Würfel (oder Tetraeder, Oktaeder, Dodekaeder, Ikosaeder) siehe Aufgabe 5.17. □

5.7 Aufgabe (Verteilungen beim zweifachen Würfelwurf.). Wie in Beispiel 5.6 sei (Ω, \mathbb{P}) das Standardmodell für den zweifachen Würfelwurf, und X_1, X_2 seien die Zufallsgrößen, die das Ergebnis des ersten bzw. des zweiten Würfelwurfes angeben. Wir betrachten neben $X_1 + X_2$ die analog definierten Zufallsgrößen $X_1 - X_2$, $\min(X_1, X_2)$ und $\max(X_1, X_2)$.

(a) Geben Sie die drei Ereignisse $\{\min(X_1, X_2) = 4\}$, $\{\min(X_1, X_2) = \max(X_1, X_2)\}$, $\{\max(X_1, X_2) \leq 3\}$ in aufsteigender Form an, analog zu

$$\{X_1 + X_2 = 4\} = \{(1, 3), (2, 2), (3, 1)\}.$$

(b) In Beispiel 5.6 wurde die Verteilung von $X_1 + X_2$ durch Angabe der zugehörigen Dichte bestimmt. Bestimmen Sie entsprechend die Verteilungen der fünf Zufallsgrößen $X_1 - X_2$, $\min(X_1, X_2)$, $\max(X_1, X_2)$, (X_1, X_2) , $(\min(X_1, X_2), \max(X_1, X_2))$ ³. □

Besonders einfache, aber ständig vorkommende Beispiele von Zufallsgrößen sind Konstanten und Indikatoren:

5.8 Bemerkung (Konstante Zufallsgrößen und Dirac-Maße). Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum, \mathcal{X} eine Menge, und $a \in \mathcal{X}$. Für $X : \Omega \rightarrow \mathcal{X}$ definiert

durch $X(\omega) := a$ für $\omega \in \Omega$ gilt dann $X \sim \delta_a$, denn für $B \subseteq \mathcal{X}$ gilt $\{X \in B\} = \begin{cases} \Omega \\ \emptyset \end{cases}$

falls $a \begin{cases} \in \\ \notin \end{cases} B$, also $\mathbb{P}(X \in B) = \begin{cases} \mathbb{P}(\Omega) = 1 & \text{falls } a \in B \\ \mathbb{P}(\emptyset) = 0 & \text{falls } a \notin B \end{cases} = \delta_a(B)$.

Allgemeiner überlegt man sich ähnlich für eine beliebige Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ die Äquivalenz

$$X \sim \delta_a \Leftrightarrow X = a \text{ f.s.},$$

wobei eine in Bemerkung 5.4 eingeführte Schreibweise verwendet wurde. □

³Siehe p. xv für die Doppelklammernotation (...).

5.9 Definition und Bemerkung (Indikatoren und Bernoulli-Verteilungen).

Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum und $A \subseteq \Omega$. Dann heißt der Indikator $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$ auch die **Indikatorgröße** von A , und es gilt

$$\mathbb{P}(\mathbf{1}_A = 0) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A) \quad \text{und} \quad \mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A),$$

und folglich $\mathbf{1}_A \square \mathbb{P} =$

$$(1) \quad B_p := (1-p)\delta_0 + p\delta_1 \in \text{Prob}(\{0, 1\})$$

mit $p := \mathbb{P}(A) \in [0, 1]$. Das Wahrscheinlichkeitsmaß B_p heißt **Bernoulli-Verteilung** mit dem Parameter p . Wir bezeichnen ihre Zähl-dichte mit

$$(2) \quad b_p(k) := \begin{cases} 1-p & \text{falls } k=0 \\ p & \text{falls } k=1 \end{cases} = p^k(1-p)^{1-k} \quad \text{für } k \in \{0, 1\}.$$

Wie bei den allgemeineren Binomialverteilungen heißt p auch hier **Trefferwahrscheinlichkeit**. □

5.10 Satz (Rechenregeln für Indikatoren). *Es sei Ω eine Menge und es seien $A, B \subseteq \Omega$. Dann gilt*

$$\begin{aligned} (1) \quad & \mathbf{1}_\emptyset = 0, \quad \mathbf{1}_\Omega = 1, \\ (2) \quad & \mathbf{1}_{A^c} = 1 - \mathbf{1}_A, \\ (3) \quad & \mathbf{1}_{A \cap B} = \mathbf{1}_A \cdot \mathbf{1}_B, \\ (4) \quad & \mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B}, \\ (5) \quad & A, B \text{ disjunkt} \Leftrightarrow \mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B. \end{aligned}$$

Diese Beziehungen zwischen Funktionen sind natürlich wieder punktweise zu lesen, also zum Beispiel $\mathbf{1}_\emptyset(\omega) = 0$ und $\mathbf{1}_{A \cap B}(\omega) = \mathbf{1}_A(\omega) \cdot \mathbf{1}_B(\omega)$ für $\omega \in \Omega$.

Beweis. (3) ergibt sich leicht durch Fallunterscheidung:

Fall 1: Es sei $\omega \in A \cap B$. Dann ist nach der Definition der Schnittbildung von Mengen $\omega \in A$ und $\omega \in B$. Mit der Definition des Indikators ergibt sich $\mathbf{1}_{A \cap B}(\omega) = 1$, $\mathbf{1}_A(\omega) = 1$ und $\mathbf{1}_B(\omega) = 1$. Die zu beweisende Gleichung $\mathbf{1}_{A \cap B}(\omega) = \mathbf{1}_A(\omega) \cdot \mathbf{1}_B(\omega)$ stimmt also in diesem Fall.

Fall 2: Es sei $\omega \notin A \cap B$. Also gelten nicht beide der Beziehungen $\omega \in A$ und $\omega \in B$. Sei etwa $\omega \notin A$. Dann ist $\mathbf{1}_{A \cap B}(\omega) = 0$ und $\mathbf{1}_A(\omega) = 0$. Somit stimmt $\mathbf{1}_{A \cap B}(\omega) = \mathbf{1}_A(\omega) \cdot \mathbf{1}_B(\omega)$ auch hier.

Die übrigen Aussagen können ähnlich verifiziert werden. Eleganter ist jedoch folgendes Argument: Für $A \subseteq \Omega$ und $\omega \in \Omega$ ist nach Definition 3.17(4) offenbar

$$\mathbf{1}_A(\omega) = \delta_\omega(A).$$

Damit sind die noch zu zeigenden Rechenregeln (1,2,4,5) für Indikatoren mit Ausnahme der Rückrichtung in (5) Spezialfälle der Rechenregeln für Wahrscheinlichkeitsmaße 3.3(1,2,4,6,5), und besagte Rückrichtung folgt etwa durch Betrachtung von $\omega \in A \cap B$. □

Indikatoren dienen oft der einfacheren Darstellung komplizierterer Zufallsgrößen:

5.11 Bemerkung. Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum und es seien $A_1, \dots, A_n \subseteq \Omega$. Dann hat die zugehörige **Zählgröße**

$$(1) \quad X(\omega) := \#\{i \in \{1, \dots, n\} : \omega \in A_i\} \quad \text{für } \omega \in \Omega$$

die für viele Zwecke günstigere Darstellung

$$(2) \quad X = \sum_{i=1}^n \mathbf{1}_{A_i},$$

die sich zum Beispiel später bei der Berechnung von Erwartungswert und Varianz von X als nützlich erweist, siehe Korollar 10.29. \square

Der folgende Satz wird bei Verteilungsrechnungen oft stillschweigend benutzt:

5.12 Satz (Transitivität der Verteilungsbildung). *Neben dem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) und der Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ sei weiter eine Funktion $T : \mathcal{X} \rightarrow \mathcal{Y}$ gegeben. Damit ist $T \circ X$ eine \mathcal{Y} -wertige Zufallsgröße auf (Ω, \mathbb{P}) mit*

$$(1) \quad (T \circ X) \square \mathbb{P} = T \square (X \square \mathbb{P}).$$

Beweis. Für $C \subseteq \mathcal{Y}$ gilt

$$\begin{aligned} ((T \circ X) \square \mathbb{P})(C) &= \mathbb{P}((T \circ X)^{-1}[C]) = \mathbb{P}(X^{-1}[T^{-1}[C]]) \\ &= (X \square \mathbb{P})(T^{-1}[C]) = (T \square (X \square \mathbb{P}))(C). \end{aligned} \quad \square$$

5.13 Beispiel. Die beiden Zufallsgrößen⁴ $X = (X_i : i \in I)$ und $Y = (Y_i : i \in I)$ mit Werten in $\mathcal{X} = \prod_{i \in I} \mathcal{X}_i$ seien identisch verteilt. Dann gilt auch $X_i \sim Y_i$ für jedes $i \in I$, denn ist (Ω_1, \mathbb{P}_1) bzw. (Ω_2, \mathbb{P}_2) der Wahrscheinlichkeitsraum auf dem X bzw. Y definiert ist, und bezeichnet $\pi_i : \mathcal{X} \rightarrow \mathcal{X}_i$ die i -te Koordinatenprojektion, so folgt

$$X_i \square \mathbb{P}_1 = (\pi_i \circ X) \square \mathbb{P}_1 = \pi_i \square (X \square \mathbb{P}_1) = \pi_i \square (Y \square \mathbb{P}_2) = Y_i \square \mathbb{P}_2$$

unter Verwendung von 5.12 im zweiten und im vierten Schritt. In knappen Worten: Gleiche **Verbundverteilungen**⁵ impliziert gleiche **Randverteilungen**. Zur Nichtumkehrbarkeit dieser Aussage siehe Aufgabe 5.14. \square

5.14 Aufgabe (Gleiche Randverteilungen, verschiedene Verbundverteilungen).

(a) Geben Sie einen Wahrscheinlichkeitsraum (Ω, \mathbb{P}) und darauf zwei \mathbb{R}^2 -wertige Zufallsgrößen (X_1, X_2) und (Y_1, Y_2) an, für die $X_1 \sim Y_1$ und $X_2 \sim Y_2$ gilt, aber nicht $(X_1, X_2) \sim (Y_1, Y_2)$.

⁴Siehe wieder p. xv für die Doppelklammernotation (\dots) .

⁵Meist weniger treffend **gemeinsame Verteilungen** genannt.

(b) Wie viele Elemente muss Ω mindestens haben, damit Zufallsgrößen wie in (a) existieren? \square

Zufallsgrößen brauchen keineswegs immer \mathbb{R} -wertig oder \mathbb{R}^k -wertig zu sein:

5.15 Beispiel (Einfache Zufallsstichproben, natürliche Definition der hypergeometrischen Verteilungen). Es sei M eine endliche Menge, $N := \#M$, $n \in \{0, \dots, N\}$, und S eine auf

$$\mathcal{S} := \{s \subseteq M : \#s = n\}$$

uniform verteilte Zufallsgröße, d.h.

$$\mathbb{P}(S = s) = \frac{1}{\binom{N}{n}} \quad \text{für } s \in \mathcal{S}$$

wegen 4.4(3). Dann heißt S **einfache Zufallsstichprobe**⁶ aus M vom Umfang n . Es sei weiter $R \subseteq M$, $B := M \setminus R$, und

$$X := \#(S \cap R),$$

soll heißen $X(\omega) := \#(S(\omega) \cap R)$ für $\omega \in \Omega$. Dann ist X eine $\{0, \dots, n\}$ -wertige Zufallsgröße, und mit $r := \#R$ und $b := \#B$ gilt $N = r + b$ und

$$(1) \quad \mathbb{P}(X = k) = \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{r+b}{n}} =: h_{n,r,b}(k) \quad \text{für } k \in \{0, \dots, n\}$$

wegen $\#\{s \in \mathcal{S} : \#(s \cap R) = k\} = \binom{r}{k} \binom{b}{n-k}$. Die Verteilung von X , also

$$H_{n,r,b} := \sum_{k=0}^n h_{n,r,b}(k) \delta_k,$$

heißt **hypergeometrische Verteilung** mit den Parametern n, r, b .

Interpretation: $H_{n,r,b}$ ist die Verteilung der Anzahl der gezogenen roten Kugeln bei einer einfachen Zufallsstichprobe von Umfang n aus einer Urne mit r roten und b blauen Kugeln. \square

Weitere wichtige Beispiele mengenwertiger Zufallsgrößen sind Konfidenzbereiche, siehe Definition 13.4 und, ganz konkret, Satz 13.8.

5.16 Aufgabe. Es seien $n, r, b \in \mathbb{N}_0$ mit $n \leq r + b$.

(a) Für $k \in \mathbb{N}_0$ gilt $h_{n,r,b}(k) = h_{n,b,r}(n-k) = h_{r,n,r+b-n}(k)$. Können Sie diese Gleichungen auch ohne Benutzung der expliziten Formel 5.15(1) begründen?

⁶Engl. *simple random sample*; daher hier S wie *sample*.

(b) Aus Aufgabe 4.8(b) ist uns schon $\text{Supp } H_{n,r,b} = \{(n-b)_+, \dots, n \wedge r\}$ bekannt. Für die Menge M der Maximalstellen von $h_{n,r,b}$ gelten nun die Implikationen

$$\begin{aligned} \lceil m \rceil < (n-b)_+ &\Rightarrow M = \{(n-b)_+\}, \\ (n-b)_+ \leq \lceil m \rceil, \lceil m+1 \rceil \leq n \wedge r &\Rightarrow M = \{\lceil m \rceil, \lceil m+1 \rceil\}, \\ \lceil m+1 \rceil > n \wedge r &\Rightarrow M = \{n \wedge r\} \end{aligned}$$

mit $m := \frac{(n+1)(r+1)}{r+b+2} - 1$.

(c) Es sei X die Anzahl der Kreuz-Karten (ohne Kreuz-Bube) des Gebers beim Skat. Welches ist der wahrscheinlichste Wert von X , und wie groß ist seine Wahrscheinlichkeit? (Lösung mit (b), numerische Verifikation mit R.) \square

Es folgt eine weniger triviale Aufgabe zur Bestimmung von Verteilungen.

5.17 Aufgabe (Verteilungen allgemeinerer Würfelaugensummen). Es sei $(X_1, \dots, X_n) \sim U_{\{0, \dots, \ell-1\}^n}$ mit $\ell, n \in \mathbb{N}$. Dann gilt für $k \in \mathbb{Z}$

$$(1) \quad \mathbb{P}\left(\sum_{i=1}^n X_i = k\right) = \frac{1}{\ell^n} \sum_{j=0}^{\lfloor \frac{k}{\ell} \rfloor} (-1)^j \binom{n}{j} \binom{n+k-\ell j-1}{n-1},$$

$$(2) \quad \mathbb{P}\left(\sum_{i=1}^n X_i \leq k\right) = \frac{1}{\ell^n} \sum_{j=0}^{\lfloor \frac{k}{\ell} \rfloor} (-1)^j \binom{n}{j} \binom{n+k-\ell j}{n},$$

mit R.S.(1) $> 0 \Leftrightarrow k \in \{0, \dots, n(\ell-1)\}$, also R.S.(2) $\begin{cases} = 0 \\ \in]0, 1[\\ = 1 \end{cases} \Leftrightarrow \begin{cases} k < 0 \\ 0 \leq k < n(\ell-1) \\ k \geq n(\ell-1) \end{cases}$.

Für “normal” markierte “Würfel”, also mit $Y_i := X_i + 1$ statt X_i links in (1) und (2), hat man rechts jeweils k durch $k - n$ zu ersetzen.

Hinweis (Montmort 1708⁷): Zum Nachweis von (1) wende man die Siebformel für Anzahlen 4.10 an auf die Komplemente in $S_{n,k} := \{x \in \mathbb{N}_0^n : \sum_{i=1}^n x_i = k\}$ der $A_i := \{x \in S_{n,k} : x_i \leq \ell - 1\}$, und verwende 4.9(1). \square

⁷Laut Hald (1998, p. 34).

6 Bedingte Wahrscheinlichkeiten

6.1 Beispiel (Hochzeit beim Doppelkopf). Wir betrachten Kartenverteilungen beim Doppelkopf in der Version “mit Neunen”: 48 Karten, je 2 mal Karo 9 bis Kreuz As, werden an Spieler 1 bis Spieler 4 verteilt; dabei sieht jeder Spieler nur seine 12 Karten. Uns interessiert das Ereignis

$$H := \text{“Hochzeit”} := \text{“Beide Kreuz-Damen in einer Hand”}.$$

Formal sei $H \subseteq \Omega$ mit einem geeigneten Wahrscheinlichkeitsraum (Ω, P) . Wie groß ist $P(H)$?

Wir nehmen an¹, dass in (Ω, P) auch die Ereignisse

$$\begin{aligned} H_j &:= \text{“Spieler } j \text{ hat beide Kreuz-Damen”} && \text{für } j \in \{1, 2, 3, 4\} \\ B_k &:= \text{“Spieler 1 hat } k \text{ Kreuz-Damen”} && \text{für } k \in \{0, 1, 2\} \end{aligned}$$

so modelliert sind, dass H die Vereinigung der paarweise disjunkten und gleichwahrscheinlichen H_j ist, dass natürlich $H_1 = B_2$ und $H_2 \cup H_3 \cup H_4 \subseteq B_0$ gilt, und dass Ω die Vereinigung der paarweise disjunkten B_k ist, deren Wahrscheinlichkeiten sich nach der hypergeometrischen Verteilung $H_{n,r,b}$ mit $n := 12$, $r := 2$, $b := 46$ berechnen, mit 5.16(a) im lediglich zwecks anschließender Rechenvereinfachung vorgenommenen zweiten Schritt also

$$P(B_k) = h_{n,r,b}(k) = h_{r,n,r+b-n}(k) = \frac{\binom{12}{k} \binom{36}{2-k}}{\binom{48}{2}} = \begin{cases} \frac{36 \cdot 35}{48 \cdot 47} = 56\% & (k = 0) \\ \frac{2 \cdot 12 \cdot 36}{48 \cdot 47} = 38\% & (k = 1) \\ \frac{12 \cdot 11}{48 \cdot 47} = 6\% & (k = 2) \end{cases}$$

wobei hier und im folgenden Prozentzahlen individuell gerundet sind. Damit folgt

$$P(H) = \sum_{j=1}^4 P(H_j) = 4 P(H_1) = 4 P(B_2) = \frac{11}{47} = 23\%.$$

Dieses Ergebnis berücksichtigt jedoch nicht das Wissen von Spieler 1 nach Betrachtung seiner Karten, der im Fall von B_1 oder B_2 ja entscheiden kann, ob H eingetreten ist oder nicht. Wir fragen daher für ihn nach einer noch zu definierenden “Wahrscheinlichkeit von H gegeben B_0 ”. → 6.3

Die gesuchte Definition ergibt sich zwanglos als Verallgemeinerung der Definition 2.2 der Verteilung P eines Quasikollektivs φ_\bullet , und steht schon in Satz 2.18. In Spezialisierung auf den Fall dass P ein elementares Wahrscheinlichkeitsmaß ist, und vom Bezug zu φ_\bullet abstrahierend, erhalten wir nach 2.18(1):

6.2 Definition. Es sei (Ω, P) ein Wahrscheinlichkeitsraum, und es sei $B \subseteq \Omega$ mit $P(B) > 0$. Für $A \subseteq \Omega$ heißt dann

$$(1) \quad P(A|B) := \frac{P(AB)}{P(B)}$$

¹Siehe Aufgabe 6.9 für eine explizite Modellierung.

die (elementare)² bedingte Wahrscheinlichkeit von A gegeben das bedingende Ereignis B . Die damit auf 2^Ω definierte Funktion $P(\cdot | B)$ heißt P bedingt auf B . \square

6.3 Beispiel (6.1 ← Hochzeit beim Doppelkopf). In 6.1 ist, wegen $H_2 \cup H_3 \cup H_4 \subseteq B_0$ und $H_1 = B_2 \subseteq B_0^c$ im zweiten Schritt,

$$\begin{aligned} P(H | B_0) &= \frac{P((H_1 \cup H_2 \cup H_3 \cup H_4) \cap B_0)}{P(B_0)} \\ &= \frac{P(H_2 \cup H_3 \cup H_4)}{P(B_0)} = \frac{3P(B_2)}{P(B_0)} = \frac{11}{35} = 31\%. \end{aligned}$$

Wer häufig Doppelkopf spielt, muss also bei ungefähr 11 von 35 Blättern ohne eigene Kreuz-Dame mit einer Hochzeit rechnen.

Man erhält hier übrigens den gleichen Zahlenwert, wenn man sich erst eine Zuteilung von 12 Nichtkreuzdamen an Spieler 1 vorstellt und dann die Wahrscheinlichkeit für das Ereignis “Hochzeit” bei der Aufteilung der restlichen 36 Karten an Spieler 2 bis Spieler 4 analog zu Beispiel 6.1 berechnet, nämlich $3h_{n=12,r=2,b=34}(2) = \frac{11}{35}$. \square

Man kann versuchen, Definition 6.2 etwa wie folgt zu motivieren: Man erhält $P(\cdot | B)$ aus P , indem man erst die Wahrscheinlichkeit auf “dem durch das vorausgesetzte Eintreten von B ja unmöglichen” Ereignis B^c “entfernt”, und die dadurch erhaltene Funktion $2^\Omega \ni A \mapsto P(AB)$ so “hochskaliert”, dass sie an der Stelle Ω den Wert 1 erreicht. Obige Motivation über Satz 2.18 wirkt zwingender.

Aus Definition 6.2 ergeben sich leicht einige nützliche Rechenregeln:

6.4 Satz (Rechenregeln für elementare bedingte Wahrscheinlichkeiten). *Es sei (Ω, P) ein Wahrscheinlichkeitsraum.*

(a) *Es sei $B \subseteq \Omega$ mit $P(B) > 0$. Dann ist $P(\cdot | B)$ ein Wahrscheinlichkeitsmaß auf Ω , d.h. es gilt*

- (1) $P(A | B) \geq 0$ für $A \subseteq \Omega$,
- (2) $\sum_{A \in \mathcal{A}} P(A | B) = 1$ für jede Quasipartition \mathcal{A} von Ω .

Weiter gelten für $A \subseteq \Omega$ die Implikationen beziehungsweise die Gleichung

- (3) $A \supseteq B \Rightarrow P(A | B) = 1$,
- (4) $A \subseteq B^c \Rightarrow P(A | B) = 0$,
- (5) $P(A | \Omega) = P(A)$,

sowie die **Formel von der totalen Wahrscheinlichkeit**

$$(6) \quad P(A) = P(A | B)P(B) + P(A | B^c)P(B^c) \quad \text{falls} \quad P(B)P(B^c) > 0$$

²Das Adjektiv “elementar” dient hier zur Abgrenzung gegenüber den “allgemeinen bedingten Wahrscheinlichkeiten” der maßtheoretischen WT, welche Zufallsgrößen statt Zahlen sind, siehe etwa Billingsley (1995, Abschnitt 33).

und allgemeiner für Familien $B_\bullet = (B_j : j \in J)$ von Ereignissen $B_j \subseteq \Omega$

$$(7) \quad P(A) = \sum_{j \in J: P(B_j) > 0} P(A|B_j)P(B_j) \quad \text{für } B_\bullet \text{ p.d. mit } A \subseteq \bigcup_{j \in J} B_j,$$

und schließlich die **Bayes-Formel**

$$(8) \quad P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad \text{falls } P(A)P(B) > 0.$$

(b) Es sei $A \subseteq \Omega$, und $B_\bullet = (B_j : j \in J)$ eine paarweise disjunkte Familie in 2^Ω mit $P(B_j) > 0$ für $j \in J$, und $J \ni j \mapsto P(A|B_j)$ konstant. Dann gilt $P(A|\bigcup_{j \in J} B_j) = P(A|B_{j_0})$ für $j_0 \in J$.

Beweis. (a) Übungsaufgabe.

(b) Mit $B := \bigcup_{j \in J} B_j$ liefert (7) angewandt auf AB statt A im ersten Schritt

$$P(AB) = \sum_{j \in J} P(AB|B_j)P(B_j) = \sum_{j \in J} P(A|B_j)P(B_j) = P(A|B_{j_0})P(B).$$

□

Verwendete man Definition 6.2(1) auch im Fall von $P(B) = 0$, mit $P(A|B) := 0/0 := 0$, so blieben manche der obigen Rechenregeln erhalten, aber nicht alle. Daher lassen wir $P(A|B)$ im Fall von $P(B) = 0$ lieber undefiniert.

Statt (8) werden oft recht unübersichtliche Kombinationen dieser Formel mit (6) oder (7), wie etwa die erste Gleichung unten in 6.8(2), als Bayes-Formeln bezeichnet; man merkt sie sich jedoch besser getrennt.

Rechenregel 6.4(b) wird etwa in Beispiel 7.4 verwendet; über diese hinaus scheinen bedingte Wahrscheinlichkeiten als Funktion des zweiten Arguments kaum besondere Eigenschaften zu haben:

6.5 Aufgabe. (a) $P(A|\cdot)$ ist in aller Regel nicht monoton.

(b) **Simpson-Paradoxon**³. Es gibt Wahrscheinlichkeitsmaße P, Q auf einem selben Grundraum Ω und zwei Ereignisse $A, B \subseteq \Omega$ derart, dass einerseits

$$(1) \quad P(A|B) < Q(A|B) \quad \text{und} \quad P(A|B^c) < Q(A|B^c),$$

und andererseits

$$(2) \quad P(A) > Q(A)$$

gilt. (*Hinweis:* $\#\Omega = 4$ reicht.)

□

Die Wichtigkeit der Kenntnis von 6.5(b) wird illustriert durch Beispiel 6.7, vorbereitet durch:

³Nicht benannt nach dem Simpson der Ihnen vielleicht zuerst einfällt, und auch nicht nach dem der Ihnen nach Absolvierung der Numerik zuerst einfallen sollte (Thomas Simpson, 1710–1761), sondern nach Eduard H. Simpson (1922–2019).

6.6 Satz. Wenn das Simpson-Paradoxon 6.5(1,2) vorliegt, dann ist die für A günstigere der beiden Bedingungen B und B^c für P und Q dieselbe, und diese ist unter P echt wahrscheinlicher als unter Q , d.h. es gilt genau eine der beiden Bedingungen

- (1) $P(A|B) > P(A|B^c)$ und $Q(A|B) > Q(A|B^c)$ und $P(B) > Q(B)$,
- (2) $P(A|B^c) > P(A|B)$ und $Q(A|B^c) > Q(A|B)$ und $P(B^c) > Q(B^c)$.

Beweis. Sind $x_1, x_2, y_1, y_2 \in \mathbb{R}$ und $\lambda, \mu \in [0, 1]$ mit

- (3) $x_1 < y_1, \quad x_2 < y_2, \quad x := \lambda x_1 + (1 - \lambda)x_2 > \mu y_1 + (1 - \mu)y_2 =: y$,

so gilt genau eine der zwei Aussagen

- (4) $x_1 < y_1 < x_2 < y_2$ und $\lambda < \mu$,
- (5) $x_2 < y_2 < x_1 < y_1$ und $\lambda > \mu$,

denn im ohne Einschränkung (eventuell Umparametrisierung $1 \leftrightarrow 2, \lambda \leftrightarrow 1 - \lambda, \mu \leftrightarrow 1 - \mu$) vorliegenden Fall von $x_1 \leq x_2$ gilt zunächst $y_1 < x_2$ und damit die erste Hälfte von (4), da wir mit (3) sonst $x \leq x_1 \vee x_2 = x_2 \leq y_1 \wedge y_2 \leq y < x$ erhielten, und weiter gilt dann $\lambda < \mu$ und damit (4), da wir wieder mit (3) und der ersten Hälfte von (4) sonst $x = x_2 + \lambda(x_1 - x_2) \leq x_2 + \mu(x_1 - x_2) = \mu x_1 + (1 - \mu)x_2 < y < x$ erhielten.

Unter 6.5(1,2) gilt nun (3) mit $x_1 := P(A|B), y_1 := Q(A|B), x_2 := P(A|B^c), y_2 := Q(A|B^c), \lambda := P(B), \mu := Q(B)$, denn es ist dort $x = P(A)$ und $y = Q(A)$ nach der Formel von der totalen Wahrscheinlichkeit 6.4(6) angewandt auf P und auf Q , und die Alternative (5) oder (4) liefert die Alternative (1) oder, mit $\lambda < \mu \Leftrightarrow 1 - \lambda > 1 - \mu$, (2). □

6.7 Beispiel (Sex bias in graduate admissions). Beim Zulassungsverfahren zum Doktorandenstudium an der Universität Berkeley wurden im Herbst 1973 von den insgesamt 8442 männlichen Bewerbern 3738 zugelassen, also 44%, von den 4321 weiblichen dagegen 1494, also nur 35%. Es gab aber andererseits keinen Grund zu der Annahme, dass die weiblichen Bewerber insgesamt weniger qualifiziert gewesen sein sollten als die männlichen.

Diese scheinbare Diskriminierung aufgrund des Geschlechts verschwand jedoch bei separater Betrachtung der einzelnen Fächer für die sich die Kandidaten beworben hatten: Zwar gab es Fächer mit einer höheren Zulassungsrate für Männer, aber der umgekehrte Fall trat sogar häufiger oder ausgeprägter auf.⁴

Für den vereinfachten Fall von nur zwei Fächern werde in 6.5(b) das Ereignis B interpretiert als “Bewerber entscheidet sich für das erste Fach”, A als “Bewerber wird angenommen”, P als Verteilung im Kollektiv der männlichen Bewerber, entsprechend Q für die weiblichen. Dann ist, beide Geschlechter als gleich qualifiziert

⁴Siehe Bickel/Hammel/O’Connell (1975, erste Spalte auf p. 399) und die Darstellung bei Freedman/Pisani/Purves (2007, dort insbesondere die bei Bickel et al. nicht enthaltene Tabelle auf p. 18). Leider scheinen die Daten nicht vollständiger verfügbar zu sein.

angenommen, im Fall von 6.5(1,2) nach allem was wir wissen keine negative Diskriminierung der Frauen bei der Zulassung erkennbar, sondern höchstens bei den die Fachwahl beeinflussenden Umständen welche dazu geführt haben mögen, dass Frauen sich für Fächer mit schlechteren Zulassungschancen beworben haben:

Die scheinbare negative Diskriminierung der Frauen 6.5(1) kann bei gleichzeitigem Vorliegen von 6.5(2) nämlich dadurch erklärt werden, dass nach Satz 6.6 jedenfals 6.6(1) oder 6.6(2) gelten muss, und das bedeutet hier: Für jedes der beiden Geschlechter ist das Fach mit den schlechteren Zulassungschancen dasselbe, und Frauen wählen dieses mit höherer Wahrscheinlichkeit als Männer.⁵ \square

Ein weiteres reales Datenbeispiel für das Simpson-Paradoxon bringt Henze (2021, p. 109, 15.13): Im Februar/März 2020 war die Sterbewahrscheinlichkeit eines Covid-19-Infizierten in Italien zwar in jeder Altersgruppe (0-9 Jahre, . . . , 70-79 Jahre, ≥ 80 Jahre) geringer als in China, ohne Berücksichtigung des Alters jedoch fast doppelt so hoch. In Italien waren die Infizierten älter als die in China.

Oft wird das sorgfältige Hinschreiben eines Modells durch die bloße Angabe gewisser unbedingter und bedingter Wahrscheinlichkeiten ersetzt, wie etwa in:

6.8 Beispiel. Eine von einer Million Münzen habe “Zahl” auf beiden Seiten, die übrigen 999 999 seien normal. Eine zufällig ausgewählte Münze werde 20 mal geworfen und zeige jedesmal “Zahl”. Mit welcher Wahrscheinlichkeit ist sie dennoch normal?

Zur Beantwortung seien in einem geeigneten Wahrscheinlichkeitsraum (Ω, P) zwei Ereignisse $A =$ “Gewählte Münze ist normal” und $B =$ “Gewählte Münze zeigt 20 mal Zahl” gegeben, mit

$$(1) \quad P(A) = 1 - 10^{-6}, \quad P(B|A) = 2^{-20}, \quad P(B|A^c) = 1,$$

wobei sich die mittlere Wahrscheinlichkeit aus der naheliegenden Gleichwahrscheinlichkeitssannahme für alle 2^{20} Münzwurfresultate ergibt. Anwendung der Bayes-Formel und der Formel von der totalen Wahrscheinlichkeit liefert, am Ende gerundet,

$$(2) \quad \begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\ &= \frac{2^{-20}(1 - 10^{-6})}{2^{-20}(1 - 10^{-6}) + 1 \cdot 10^{-6}} = 0.49. \end{aligned} \quad \boxed{\rightarrow 7.4}$$

Überraschende Zahlenwerte bedingter Wahrscheinlichkeiten wie in Beispiel 6.8 treten analog in vielen praktisch wichtigen Fällen auf, im medizinischen Kontext

⁵Es wäre jedoch denkbar und jedenfalls rein mathematisch möglich, dass die Bewerbungen tatsächlich auf jeweils eines von zwei Unterfächern erfolgte, also $B = B_1 \cup B_2$ und $B^c = B_3 \cup B_4$ mit paarweise disjunkten B_i , bei denen dann wieder die Frauen benachteiligt erscheinen, d.h. $P(A|B_i) > P(A|B_j)$ für $i \in \{1, 2, 3, 4\}$. Daher oben die Einschränkung “nach allem was wir wissen”.

etwa bei der Fruchtwasseruntersuchung zur Diagnose des Down-Syndroms bei Ungeborenen, siehe Hesse (2009, p. 53), oder beim ELISA-Test zur Erkennung von Antikörpern gegen HIV, siehe Henze (2017, pp. 107–109), in Henze (2021, pp. 105–107) ohne Bezug auf einen konkreten Test.

Aber wie schreibt man denn nun etwa in Beispiel 6.8 ein Modell (Ω, P) sauber hin? Dieser Frage widmet sich Kapitel 7.

Weitere Aufgaben

6.9 Aufgabe (Ein explizites Modell für Beispiel 6.1). Definieren Sie formal Ereignisse $H, H_1, H_2, H_3, H_4, B_0, B_1, B_2$ mit den in Beispiel 6.1 geforderten Eigenschaften im Gleichverteilungsmodell auf der Menge

$$\Omega := \left\{ \omega \in \binom{M}{12}^4 : \bigcup_{j=1}^4 \omega_j = M \right\}$$

aller geordneten Partitionen der 48-elementigen Menge $M := \{1, \dots, 48\}$ in vier gleich große Teilmengen.

(Ein Element von Ω ist also ein Quadrupel $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$ mit $\omega_j \subseteq M$ und $\#\omega_j = 12$ für $j \in \{1, 2, 3, 4\}$, sowie $\bigcup_{j=1}^4 \omega_j = M$ und damit dann automatisch auch ω paarweise disjunkt.) □

6.10 Aufgabe. Wie verschärft sich Satz 2.18 für \mathcal{S} -Kollektive? □

7 Kopplungsmodelle

Wir beantworten in diesem Kapitel die hinter Beispiel 6.8 gestellte Frage. Die dortige Situation suggeriert eine “stufenweise Modellierung”, die wir der Einfachheit halber in Satz 7.2 zunächst nur für den (im betrachteten Beispiel ausreichenden) einfachsten Fall von lediglich zwei “Stufen” betrachten. Zur Vorbereitung notieren wir:

7.1 Definition. Es seien \mathcal{X} und \mathcal{Y} nichtleere Mengen. Jede Funktion $K : \mathcal{X} \times 2^{\mathcal{Y}}$ mit $K(x, \cdot) \in \text{Prob}(\mathcal{Y})$ für $x \in \mathcal{X}$ heißt (**elementarer**) **Markov-Kern** oder **Übergangskern**, jede Funktion $k : \mathcal{X} \times \mathcal{Y}$ mit $k(x, \cdot) \in \text{prob}(\mathcal{Y})$ für $x \in \mathcal{X}$ heißt **Übergangs(zähl)dichte**, jeweils von \mathcal{X} nach \mathcal{Y} . Die Menge all dieser K beziehungsweise k wird mit $\text{Mark}(\mathcal{X}, \mathcal{Y})$ beziehungsweise $\text{mark}(\mathcal{X}, \mathcal{Y})$ bezeichnet.

Konvention: Statt des Trennsymbols “,” in K bzw. k wird alternativ auch das Symbol “|” bei Umkehrung der Reihenfolge der beiden Argumente verwendet, also $K(B|x)$ statt $K(x, B)$ und $k(y|x)$ statt $k(x, y)$ geschrieben; dabei heißt weiter x das erste Argument und B bzw. y das zweite. \square

Unter Verwendung von Satz 3.6 ergibt sich dessen naheliegendes Analogon für $\text{Mark}(\mathcal{X}, \mathcal{Y})$ und $\text{mark}(\mathcal{X}, \mathcal{Y})$ statt $\text{Prob}(\mathcal{X})$ und $\text{prob}(\mathcal{X})$, mit $k(x, y) := K(x, \{y\})$ und $K(x, B) := \sum_{y \in B} k(x, y)$.

Unten in 7.2(3) steht links eine bedingte Wahrscheinlichkeit gemäß Definition 6.2(1), rechts der Wert des Kerns $P_{2|1} \in \text{Mark}(\Omega_1, \Omega_2)$ an der Stelle (ω_1, A_2) .

7.2 Satz. Es sei $\Omega = \Omega_1 \times \Omega_2$ mit Mengen Ω_1 und Ω_2 .

(a) Es sei $P_1 \in \text{Prob}(\Omega_1)$ mit der Dichte p_1 , und $P_{2|1} \in \text{Mark}(\Omega_1, \Omega_2)$ mit der Übergangsdichte $p_{2|1}$. Mit

$$(1) \quad p(\omega) := p_1(\omega_1)p_{2|1}(\omega_2|\omega_1) \quad \text{für } \omega \in \Omega$$

ist dann p die Dichte des einzigen $P \in \text{Prob}(\Omega)$ mit

$$(2) \quad P(A_1 \times \Omega_2) = P_1(A_1) \quad \text{für } A_1 \subseteq \Omega_1,$$

$$(3) \quad P(\Omega_1 \times A_2 | \{\omega_1\} \times \Omega_2) = P_{2|1}(A_2|\omega_1) \quad \text{für } P_1(\{\omega_1\}) > 0, A_2 \subseteq \Omega_2.$$

(b) Es sei $P \in \text{Prob}(\Omega)$. Genau für P_1 definiert durch (2) und $P_{2|1} : \Omega_1 \times 2^{\Omega_2}$ gewählt mit (3) und

$$(4) \quad P_{2|1}(\cdot|\omega_1) \in \text{Prob}(\Omega_2) \quad \text{für } P_1(\{\omega_1\}) = 0$$

ist $P_1 \in \text{Prob}(\Omega_1)$ und $P_{2|1} \in \text{Mark}(\Omega_1, \Omega_2)$ derart, dass das hiesige P gleich dem aus (a) ist.

Beweis. (a) Es ist $p \geq 0$, und mit der Normiertheit von $p_{2|1}(\cdot|\omega_1)$ im zweiten Schritt und der von p_1 im dritten folgt

$$(5) \quad \sum_{\omega \in \Omega} p(\omega) = \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} p_{2|1}(\omega_2|\omega_1)p_1(\omega_1) = \sum_{\omega_1 \in \Omega_1} p_1(\omega_1) = 1;$$

also ist p eine Wahrscheinlichkeitsdichte auf Ω .

Ist P das zu p gehörige Wahrscheinlichkeitsmaß auf Ω , so folgt analog zu (5)

$$\text{L.S.}(2) = \sum_{\omega \in A_1 \times \Omega_2} p(\omega) = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in \Omega_2} p_{2|1}(\omega_2|\omega_1)p_1(\omega_1) = \text{R.S.}(2) \quad \text{für } A_1 \subseteq \Omega_1,$$

und, mit der Definition 6.2(1) und mit $P(\{\omega_1\} \times \Omega_2) = P_1(\{\omega_1\}) > 0$ wegen (2),

$$\text{L.S.}(3) = \frac{P(\{\omega_1\} \times A_2)}{P(\{\omega_1\} \times \Omega_2)} = \frac{\sum_{\omega_2 \in A_2} p_1(\omega_1)p_{2|1}(\omega_2|\omega_1)}{p_1(\omega_1)} = \text{R.S.}(3)$$

für $\omega_1 \in \Omega_1$ mit $P_1(\{\omega_1\}) > 0$ und für $A_2 \subseteq \Omega_2$.

Ist umgekehrt $P \in \text{Prob}(\Omega)$ mit (2) und (3), so gilt mit p aus (1) für $\omega \in \Omega$ stets $P(\{\omega\}) = p(\omega)$, denn $P(\{\omega\}) \leq P(\{\omega_1\} \times \Omega_2) = P_1(\{\omega_1\}) = p_1(\omega_1)$ liefert dies falls $P_1(\{\omega_1\}) = 0$, und andernfalls gilt

$$\begin{aligned} P(\{\omega\}) &= P(\Omega_1 \times \{\omega_2\} | \{\omega_1\} \times \Omega_2)P(\{\omega_1\} \times \Omega_2) \\ &= P_{2|1}(\{\omega_2\}|\omega_1)P_1(\{\omega_1\}) = \text{R.S.}(1); \end{aligned}$$

also ist P das zu p gehörige Wahrscheinlichkeitsmaß.

(b) Für P_1 definiert durch (2) und $P_{2|1}$ gewählt mit (3) und (4) gilt offenbar¹ $P_1 \in \text{Prob}(\Omega_1)$ und $P_{2|1} \in \text{Mark}(\Omega_1, \Omega_2)$. Da das hier gegebene P die Bedingungen (2) und (3) erfüllt, ist es damit das eindeutige P aus (a).

Sind umgekehrt $P_1 \in \text{Prob}(\Omega_1)$ und $P_{2|1} \in \text{Mark}(\Omega_1, \Omega_2)$ derart, dass das gemäß (a) zugehörige Wahrscheinlichkeitsmaß gleich dem hier gegebenen P ist, so gelten (2), (3), (4). \square

7.3 Definition. In der Situation aus Satz 7.2(a) heißt

$$P_1 \otimes P_{2|1} := P$$

die **Kopplung** der **Startverteilung** P_1 mit dem Übergangskern $P_{2|1}$. \square

Nach Satz 7.2(a) ist damit jede Kopplung ein Wahrscheinlichkeitsmaß. Nach Satz 7.2(b) ist jedes Wahrscheinlichkeitsmaß auf einem kartesischen Produkt zweier Mengen eine Kopplung (**elementarer Desintegrationsatz**).

7.4 Beispiel (6.8 \leftarrow). Für Beispiel 6.8 sei

$$\Omega_1 := \{1, \dots, 10^6\}, \quad P_1 := U_{\Omega_1}$$

¹Erster Beweis: Direktes Nachrechnen der Definition 3.2.

Zweiter Beweis: Mit den beiden Koordinatenprojektionen $\pi_i : \Omega_1 \times \Omega_2 \rightarrow \Omega_i$ für $i \in \{1, 2\}$ ist $P_1 = \pi_1 \square P$ nach (2), und $P_{2|1}(\cdot|\omega_1) = \pi_2 \square P(\cdot | \{\omega_1\} \times \Omega_2)$ für $P_1(\{\omega\}) > 0$ nach (3), und die Behauptung folgt mit Satz 5.2 und der Wahrscheinlichkeitsmaßeigenschaft von $P(\cdot | \{\omega_1\} \times \Omega_2)$ nach Satz 6.4(a).

mit der Interpretation $1 \hat{=}$ Doppelzahlmünze wird im ersten Schritt gewählt, und

$$\Omega_2 := \{\text{Kopf, Zahl}\}^{20}, \quad z := (\text{Zahl}, \dots, \text{Zahl}) \in \Omega_2,$$

$$P_{2|1}(\cdot|\omega_1) := \begin{cases} U_{\Omega_2} \\ \delta_z \end{cases} \text{ falls } \omega_1 \begin{cases} \in \Omega_1 \setminus \{1\} \\ = 1 \end{cases}.$$

Mit (Ω, P) wie in Satz 7.2(a) und

$$A := (\Omega_1 \setminus \{1\}) \times \Omega_2, \quad B := \Omega_1 \times \{z\}$$

gilt dann 6.8(1), denn:

$P(A) = 1 - 10^{-6}$ ist 7.2(2) mit $A_1 := \Omega_1 \setminus \{1\}$, für $\omega_1 \in \Omega_1 \setminus \{1\}$ gilt wegen 7.2(3) mit $A_2 := \{z\}$ im ersten Schritt

$$P(B | \{\omega_1\} \times \Omega_2) = P_{2|1}(\{z\}|\omega_1) = 2^{-20}$$

und mit Satz 6.4(b) folgt $P(B | A) = 2^{-20}$, und wieder wegen 7.2(3) mit $A_2 := \{z\}$ sowie mit $A^c = \{1\} \times \Omega_2$ folgt $P(B | A^c) = P_{2|1}(\{z\}|1) = 1$. \square

Das sorgfältige Hinschreiben von Kopplungsmodellen mag in Beispielen wie 7.4 oder unten in 7.11 als eine unnötige Formalisierung des Offensichtlichen empfunden werden, denn dort ergibt sich ja - vielleicht mit Ausnahme von 7.11(5,6) - nichts Überraschendes. Jedoch erzwingt eine sorgfältige Modellierung die Formulierung sonst vielleicht unbewusst bleibender Annahmen, und bewahrt damit vor Scheinlösungen, oder vor der fruchtlosen Beschäftigung mit Scheinproblemen, wie durch die Aufgaben 7.6 und 7.12 illustriert.

7.5 Aufgabe. Von drei Spielkarten sei eine beidseitig weiß, eine beidseitig rot, und die verbleibende auf einer Seite weiß und auf der anderen rot. Die Karten werden rein zufällig unter ein schwarzes Tuch gelegt und gemischt. Nach Hervorziehen einer Karte sehen wir eine weiße Oberseite. Mit welcher Wahrscheinlichkeit ist auch die Unterseite weiß? \square

7.6 Aufgabe (Zwei-Jungen-Problem). Nachmittags von einer mehrtägigen Reise zurückkommend erfährt Herr Stoch vom Einzug einer vierköpfigen Familie ins Nachbarhaus. Am Abend berichtet ihm sein gerade vom Fußballtraining zurückkehrender Sohn, dass dort ein Neuer war, der jetzt im Nachbarhaus wohnt. Mit welcher Wahrscheinlichkeit sind nun für Herrn Stoch beide neuen Nachbarskinder Jungen?

(a) Beantworten Sie obige Frage durch Berechnung einer bedingten Wahrscheinlichkeit im naheliegenden Gleichverteilungsmodell für die a priori möglichen Geschlechterkombinationen.

(b) Modellieren Sie nun sorgfältiger als in Teil (a), indem Sie annehmen, dass die Nachbarskinder unabhängig voneinander zum Fußballtraining gehen, und zwar Jungen je mit Wahrscheinlichkeit p und Mädchen mit Wahrscheinlichkeit q . Präzisieren Sie das bedingende Ereignis zu "Genau ein neues Nachbarskind war beim Fußball, und zwar ein Junge". Welche Antworten sind bei geeigneter Wahl von p und q möglich? Welche unter der Restriktion $q = 0$? Was ergibt sich im Fall $p = q$?

(c) Was hätte Herr Stoch gerechnet, wenn ihm sein Sohn nach dem Training nichts von neuen Mitfussballspielern erzählt hätte? (Modell wie in (b), bedingendes Ereignis = ?) \square

7.7 Aufgabe (Optimale Aufteilung). Sie dürfen von zwei äußerlich ununterscheidbaren Urnen eine auswählen und dann aus der gewählten Urne rein zufällig eine der darin enthaltenen Kugeln ziehen. Ist die Kugel rot, so gewinnen Sie einen Hauptpreis, andernfalls gehen Sie leer aus. Die Kugeln, nämlich 100 rote und 100 blaue, dürfen Sie vorher selbst nach Belieben auf die Urnen verteilen, wobei allerdings jede Urne mindestens eine Kugel enthalten soll und alle Kugeln aufgeteilt werden sollen. Zwischen Aufteilung und Ziehung werden jedoch die Urnen ohne Ihr Zusehen umgestellt. Wie würden Sie die Kugeln aufteilen und wie groß wäre dann Ihre Wahrscheinlichkeit für einen Hauptpreis?

Hinweis: Stellen Sie zuerst für jede erlaubte Kugelaufteilung ein geeignetes Kopplungsmodell auf und berechnen Sie die zugehörige Gewinnwahrscheinlichkeit. Bestimmen Sie dann die optimale(n) Kugelaufteilung(en). \square

Die naheliegende Verallgemeinerung von Definition 7.3 auf $n \geq 2$ Stufen erfolgt induktiv:

7.8 Definition. Für $n \geq 2$ Mengen $\Omega_1, \dots, \Omega_n$ sowie für $P_1 \in \text{Prob}(\Omega_1)$ und $P_k \in \text{Mark}(\times_{j=1}^{k-1} \Omega_j, \Omega_k)$ für $k \in \{2, \dots, n\}$ heißt das ausgehend von Definition 7.3 induktiv definierte²

$$(1) \quad \bigotimes_{j=1}^n P_j := \left\{ \begin{array}{ll} P_1 \otimes P_2 & \text{falls } n = 2, \\ \left(\bigotimes_{j=1}^{n-1} P_j \right) \otimes P_n & \text{falls } n > 2 \end{array} \right\} \in \text{Prob}\left(\times_{j=1}^n \Omega_j\right)$$

die **Kopplung** von (P_1, \dots, P_n) , und $(\times_{j=1}^n \Omega_j, \bigotimes_{j=1}^n P_j)$ heißt **(n-stufiges) Kopplungsmodell**. \square

7.9 Bemerkungen. (a) Statt P_1, P_2, P_3, \dots schreiben wir in obiger Situation analog zur Schreibweise in Definition 7.3 auch umständlicher $P_1, P_{2|1}, P_{3|1,2}, \dots$, und allgemein statt P_j auch $P_{j|1, \dots, j-1}$ für $j \in \{1, \dots, n\}$, um damit die Art dieser Objekte notationell anzudeuten; aus $\bigotimes_{j=1}^n P_j$ in 7.8(1) wird damit $\bigotimes_{j=1}^n P_{j|1, \dots, j-1}$. Analog schreiben wir für die zu den $P_{j|1, \dots, j-1}$ gehörigen (Übergangs-)dichten neben p_j auch $p_{j|1, \dots, j-1}$. Damit hat dann $\bigotimes_{j=1}^n P_{j|1, \dots, j-1}$ per Induktion offenbar die Dichte

$$(1) \quad \begin{aligned} \times_{j=1}^n \Omega_j \ni \omega &\mapsto \prod_{j=1}^n p_{j|1, \dots, j-1}(\omega_j | \omega_1, \dots, \omega_{j-1}) \\ &= p_1(\omega_1) p_{2|1}(\omega_2 | \omega_1) p_{3|1,2}(\omega_3 | \omega_1, \omega_2) \dots p_{n|1, \dots, n-1}(\omega_n | \omega_1, \dots, \omega_{n-1}), \end{aligned}$$

manchmal **“erste Pfadregel”** genannt.

²Wir verwenden die üblichen Identifizierungen iterierter kartesischer Produkte, wie zum Beispiel die von $(\Omega_1 \times \Omega_2) \times \Omega_3$ mit $\Omega_1 \times \Omega_2 \times \Omega_3$. Zum Beispiel ist ja streng genommen $(P_1 \otimes P_2) \otimes P_3 \in \text{Prob}((\Omega_1 \times \Omega_2) \times \Omega_3)$ und nicht $\in \text{Prob}(\Omega_1 \times \Omega_2 \times \Omega_3)$, und damit dann $((P_1 \otimes P_2) \otimes P_3) \otimes P_4$ so eigentlich gar nicht definiert.

(b) **Konsistenz.** Gleichung 7.2(2) kann als eine Konsistenz bei einem Modellwechsel gelesen werden: Die dortigen Ereignisse A_1 und $A_1 \times \Omega_2$ werden ja gleich interpretiert, etwa in Beispiel 7.4 als “Gewählte Münze ist normal”, sind aber formal Ereignisse in verschiedenen Wahrscheinlichkeitsräumen; konsistenterweise stimmen jedoch ihre jeweiligen Wahrscheinlichkeiten überein. In Verallgemeinerung von 7.2(2) erhält man in der Situation von Definition 7.8 per Induktion leicht

$$(2) \quad \left(\bigotimes_{j=1}^n P_j \right) \left(A \times \prod_{j=k+1}^n \Omega_j \right) = \left(\bigotimes_{j=1}^k P_j \right) (A) \quad \text{für } 1 \leq k < n \text{ und } A \subseteq \prod_{j=1}^k \Omega_j$$

und damit speziell

$$(3) \quad \left(\bigotimes_{j=1}^n P_j \right) \left(A \times \prod_{j=2}^n \Omega_j \right) = P_1(A) \quad \text{für } A \subseteq \Omega_1.$$

(c) **Spezialfälle.** Wir verwenden ab hier bis Teil (f) nur die umständlichere Notation aus (a) und setzen

$$(4) \quad \mathbb{P} := \bigotimes_{j=1}^n P_{j|1,\dots,j-1}.$$

Gilt nun speziell

$$P_{j|1,\dots,j-1}(\cdot | \omega_1, \dots, \omega_{j-1}) = P_{j|j-1}(\cdot | \omega_{j-1}) \quad \text{für } 3 \leq j \leq n \text{ und} \\ \omega_1 \in \Omega_1, \dots, \omega_{j-1} \in \Omega_{j-1}$$

mit Übergangskernen $P_{j|j-1} \in \text{Mark}(\Omega_{j-1}, \Omega_j)$, so heißt \mathbb{P} **Markovsche Kopplung** und wird auch mit $\bigotimes_{j=1}^n P_{j|j-1}$ bezeichnet, mit $P_{1|0} := P_1$.

Gilt noch spezieller

$$P_{j|1,\dots,j-1}(\cdot | \omega_1, \dots, \omega_{j-1}) = P_j \quad \text{für } 2 \leq j \leq n \text{ und} \\ \omega_1 \in \Omega_1, \dots, \omega_{j-1} \in \Omega_{j-1}$$

mit Wahrscheinlichkeitsmaßen $P_j \in \text{Prob}(\Omega_j)$, so heißt \mathbb{P} **unabhängige Kopplung** oder **Produkt(wahrscheinlichkeits)maß** oder **Produktverteilung** und wird auch mit $\bigotimes_{j=1}^n P_j$ bezeichnet. In diesem Fall reduziert sich mit den zu den P_j gehörigen Dichten p_j die Formel (1) für die Dichte von \mathbb{P} zu

$$(5) \quad \prod_{j=1}^n \Omega_j \ni \omega \mapsto \prod_{j=1}^n p_j(\omega_j),$$

und damit ergibt sich in Verallgemeinerung von (3) leicht

$$(6) \quad \left(\bigotimes_{j=1}^n P_j \right) \left(\prod_{j=1}^n A_j \right) = \prod_{j=1}^n P_j(A_j) \quad \text{für } A_1 \subseteq \Omega_1, \dots, A_n \subseteq \Omega_n.$$

(d) Allgemeinere unabhängige Kopplungen. In (5) und (6) wird anders als in (1–4) die Ordnung auf der Indexmenge $\{1, \dots, n\}$ nicht wirklich benötigt:

Es sei $((\Omega_i, P_i) : i \in I)$ eine Familie von Wahrscheinlichkeitsräumen mit $\#I \in \mathbb{N}$ und den zugehörigen Dichten p_i . Durch

$$(7) \quad p(\omega) := \prod_{i \in I} p_i(\omega_i) \quad \text{für } \omega \in \prod_{i \in I} \Omega_i$$

wird dann auf $\prod_{i \in I} \Omega_i$ eine Wahrscheinlichkeitsdichte definiert, das zugehörige Wahrscheinlichkeitsmaß wird mit $\otimes_{i \in I} P_i$ bezeichnet und heißt das **Produkt** der P_i , und damit gilt

$$(8) \quad \left(\otimes_{i \in I} P_i \right) \left(\prod_{i \in I} A_i \right) = \prod_{i \in I} P_i(A_i) \quad \text{falls } A_i \subseteq \Omega_i \text{ für } i \in I.$$

(e) Koordinatengrößen. Ist (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum mit $\Omega = \prod_{i \in I} \Omega_i$, hier mit $\#I \in \mathbb{N}$, so heißt die i -te Koordinatenprojektion

$$\Omega \ni \omega \mapsto \omega_i \in \Omega_i$$

auch i -te **Koordinatengröße**, und wird zum Beispiel mit X_i bezeichnet. Damit schreibt sich im allgemeinen Fall von (4) die Konsistenz (3) als

$$(9) \quad X_1 \sim P_1$$

und allgemeiner (2) als

$$(10) \quad (X_1, \dots, X_k) \sim \otimes_{j=1}^k P_{j|1, \dots, j-1} \quad \text{für } 1 \leq k \leq n,$$

denn für $A \subseteq \prod_{j=1}^k \Omega_j$ ist $\mathbb{P}((X_1, \dots, X_k) \in A) = \text{L.S.}(2) = \text{R.S.}(2)$.

Ist schließlich $\mathbb{P} = \otimes_{i \in I} P_i$ eine unabhängige Kopplung wie in (d), so gilt

$$(11) \quad X_i \sim P_i \quad \text{für } i \in I,$$

$$(12) \quad \mathbb{P}(X_i \in A_i \text{ für } i \in I) = \prod_{i \in I} \mathbb{P}(X_i \in A_i) \quad \text{falls } A_i \subseteq \Omega_i \text{ für } i \in I,$$

denn es gilt

$$(13) \quad \text{L.S.}(12) = \text{L.S.}(8) = \text{R.S.}(8),$$

was für jedes $i_0 \in I$ mit $A_i := \Omega_i$ für $i \in I \setminus \{i_0\}$ zunächst $\mathbb{P}(X_{i_0} \in A_{i_0}) = P_{i_0}(A_{i_0})$ für $A_{i_0} \subseteq \Omega_{i_0}$ liefert. Letzteres ist (11), und ergibt durch Einsetzen in (13) auch (12).

(f) Zur gelegentlichen Vermeidung von Fallunterscheidungen definieren wir noch “einstufige Kopplungen” $\otimes_{i \in \{i_0\}} P_i := P_{i_0}$ für Wahrscheinlichkeitsmaße P_{i_0} , und sogar “nullstufige Kopplungen” $\otimes_{i \in \emptyset} P_i := \delta_\emptyset \in \text{Prob}(\{\emptyset\})$. \square

Ein einfaches Beispiel zu 7.9(d) ist:

7.10 Beispiel (Produkte von Gleichverteilungen sind Gleichverteilungen). Es sei $((\Omega_i, U_{\Omega_i,0}) : i \in I)$ mit $\#I \in \mathbb{N}$ eine Familie von Wahrscheinlichkeitsräumen mit Gleichverteilungen (siehe Definition 3.17). Dann ist

$$\bigotimes_{i \in I} U_{\Omega_i,0} = U_{\prod_{i \in I} \Omega_i,0}$$

auf $\Omega := \prod_{i \in I} \Omega_i$, denn für $\omega \in \Omega$ ist mit $\Omega_0 := \prod_{i \in I} \Omega_i,0$

$$\left(\bigotimes_{i \in I} U_{\Omega_i,0} \right) (\{\omega\}) = \prod_{i \in I} U_{\Omega_i,0} (\{\omega_i\}) = \prod_{i \in I} \frac{(\omega_i \in \Omega_i,0)}{\#\Omega_i,0} = \frac{(\omega \in \Omega_0)}{\#\Omega_0} = U_{\Omega_0} (\{\omega\}).$$

wegen 7.9(7) im ersten Schritt, 3.17(2) im zweiten und vierten, und 4.1(1) im dritten. \square

7.11 Beispiel (Sukzessives Ziehen). Es sei M eine Menge mit $N := \#M \in \mathbb{N}_0$ und es sei $n \in \mathbb{N}_0$. Wir modellieren “ n sukzessive Einzelziehungen aus M ” durch ein n -stufiges Kopplungsmodell \mathbb{P} auf

$$(1) \quad \Omega := M^n;$$

in der Notation aus 7.8 und 7.9 ist hier also $\Omega_j = M$ für $j \in \{1, \dots, n\}$. Es sei weiter $X := (X_1, \dots, X_n) := \text{id}_\Omega$, also X_j die j -te Koordinatengröße wie in 7.9(e).

Wir unterscheiden zwei Ziehungsvarianten. Im Trivialfall $n = 0$ sei dabei jeweils $\mathbb{P} := \delta_\emptyset$, im Einklang mit 7.9(f).

(a) **Ohne Zurücklegen**. Es sei $n \in \{1, \dots, N\}$,

$$(2) \quad P_1 := U_M$$

und, für $j \in \{2, \dots, n\}$,

$$(3) \quad P_{j|1, \dots, j-1}(\cdot | \omega_1, \dots, \omega_{j-1}) := U_{M \setminus \{\omega_1, \dots, \omega_{j-1}\}} \quad \text{für } (\omega_1, \dots, \omega_{j-1}) \in M^{j-1}.$$

Dann ist

$$(4) \quad \mathbb{P} := \bigotimes_{j=1}^n P_{j|1, \dots, j-1} = U_{M_{\neq}^n},$$

denn mit den zu (2) und (3) gehörigen (Übergangs-)dichten $p_{j|1, \dots, j-1}$ ist für $\omega \in \Omega$

$$\begin{aligned} \mathbb{P}(\{\omega\}) &= \prod_{j=1}^n p_{j|1, \dots, j-1}(\omega_j | \omega_1, \dots, \omega_{j-1}) = \prod_{j=1}^n \frac{(\omega_j \in M \setminus \{\omega_1, \dots, \omega_{j-1}\})}{\#(M \setminus \{\omega_1, \dots, \omega_{j-1}\})} \\ &= \frac{(\omega \in M_{\neq}^n)}{N^n} = U_{M_{\neq}^n}(\{\omega\}) \end{aligned}$$

wegen 7.9(1) im ersten Schritt, (2,3) und 3.17(2) im zweiten, der Äquivalenz $\omega \notin M_{\neq}^n \Leftrightarrow \exists j \in \{1, \dots, n\}$ mit $\omega_j \in \{\omega_1, \dots, \omega_{j-1}\}$ im dritten, und 3.17(3) im vierten.

Weiter gilt

$$(5) \quad ((X_{\sigma(1)}, \dots, X_{\sigma(n)}) \sim X \quad \text{für jede Permutation } \sigma \text{ von } \{1, \dots, n\}$$

(genannt **Permutationsinvarianz** von $(X_1, \dots, X_n) \square \mathbb{P}$, oder **Vertauschbarkeit**³ von (X_1, \dots, X_n) unter \mathbb{P}), denn für $y \in \Omega$ gilt

$$\begin{aligned} \mathbb{P}(\text{L.S.}(5) = y) &= \mathbb{P}(X = y \circ \sigma^{-1}) \\ &= \frac{(y \circ \sigma^{-1} \in M_{\neq}^n)}{N^n} = \frac{(y \in M_{\neq}^n)}{N^n} = \mathbb{P}(X = y). \end{aligned}$$

Damit gilt insbesondere

$$(6) \quad X_j \sim U_M \quad \text{für } j \in \{1, \dots, n\}$$

wegen $X_1 \sim U_M$ nach 7.9(9) und wegen $X_j \sim X_1$ nach (5) mit $\sigma(1) = j$. Weiter folgt⁴ für $s \subseteq M$ mit $\#s = n$, also $s = \{x_1, \dots, x_n\}$ für ein $x \in M_{\neq}^n$,

$$(7) \quad \mathbb{P}(\{X_1, \dots, X_n\} = s) = \mathbb{P}\left(\bigcup_{\sigma} \{X = x \circ \sigma\}\right) = \frac{n!}{N^n} = \frac{1}{\binom{N}{n}},$$

wegen der Gleichheit der beiden Ereignisse im ersten Schritt, und wegen der Additivität von \mathbb{P} und $X \sim U_{M_{\neq}^n}$ im zweiten, und somit ist $\{X_1, \dots, X_n\}$ eine einfache Zufallsstichprobe aus M vom Umfang n .⁵

Insbesondere gilt für

$$(8) \quad R \subseteq M \text{ mit } r := \#R \text{ und } b := N - r$$

dann, mit der in Bemerkung 5.4 eingeführten Notation im ersten Schritt und wegen 5.15(1) im zweiten Schritt,

$$(9) \quad \sum_{j=1}^n (X_j \in R) \stackrel{=}{=}_{\mathbb{P}} \#(\{X_1, \dots, X_n\} \cap R) \sim H_{n,r,b};$$

dabei gilt im ersten Schritt tatsächlich Gleichheit in jedem Punkt $\omega \in \Omega$ für den $\{X_1(\omega), \dots, X_n(\omega)\}$ n -elementig ist, also mit Wahrscheinlichkeit 1. Interpretation analog zu der in 5.15: Auch beim sukzessiven Ziehen ohne Zurücklegen von n Kugeln aus einer Urne mit r roten und b blauen Kugeln ist $H_{n,r,b}$ die Verteilung der Anzahl der gezogenen roten Kugeln.

³Leider meist weniger treffend **Austauschbarkeit** genannt, als ob zum Beispiel im Fall $n = 2$ statt $(X_1, X_2) \sim (X_2, X_1)$ etwa $(X_1, X_2) \sim (X_1, Y_2)$ mit einer gewissen dritten Zufallsgröße Y_2 gelten sollte. Wir folgen hier Chow/Teicher (1997, Titel und p. xiii) und nicht der Mehrheit.

⁴Dabei ist $\{X_1, \dots, X_n\}$ hier zu lesen als $\Omega \ni \omega \mapsto \{X_1(\omega), \dots, X_n(\omega)\}$, und damit $\mathbb{P}(\{X_1, \dots, X_n\} = s)$ als $\mathbb{P}(\{\omega \in \Omega : \{X_1(\omega), \dots, X_n(\omega)\} = s\})$.

⁵Was anschaulich kaum überrascht. Aber: "Was beweisbar ist, soll in der Wissenschaft nicht ohne Beweis geglaubt werden." (Dedekind 1893, Vorwort.)

(b) **Mit Zurücklegen.** Es seien $n, N \in \mathbb{N}$; dabei ist jetzt also auch $n > N$ erlaubt. Mit

$$P_j := U_M \quad \text{für } j \in \{1, \dots, n\}$$

erhalten wir mit 7.10 die unabhängige Kopplung

$$(10) \quad \mathbb{P} := \bigotimes_{j=1}^n P_j = U_{M^n}.$$

Wegen 7.9(11) gilt auch hier

$$(11) \quad X_j \sim U_M \quad \text{für } j \in \{1, \dots, n\};$$

also hat jedes X_j hier dieselbe Verteilung wie in (a), aber die beiden Verteilungen von X sind im Fall von $n \geq 2$ verschieden.

Gilt schließlich wieder (8), so haben wir für $k \in \{0, \dots, n\}$

$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^n (X_j \in R) = k\right) &= \frac{1}{\#\Omega} \#\{\omega \in \Omega : \sum_{j=1}^n (\omega_j \in R) = k\} \\ &= \frac{1}{N^n} \binom{n}{k} \#(R^k) \#((M \setminus R)^{n-k}) \\ &= \binom{n}{k} \frac{r^k b^{n-k}}{(r+b)^n} \\ &= b_{n, \frac{r}{r+b}}(k) \end{aligned}$$

wobei im zweiten Schritt der Binomialkoeffizient die Auswahl der k Indizes j zählt, für die $\omega_j \in R$ gelten soll; also gilt hier

$$(12) \quad \sum_{j=1}^n (X_j \in R) \sim B_{n, \frac{r}{r+b}}$$

(stochastische Herleitung der Binomialverteilung $B_{n,p}$ mit rationaler Trefferwahrscheinlichkeit p). \square

7.12 Aufgabe (Das Ziegenproblem). Ein Kandidat hat am Ende einer Fernsehshow die Wahl zwischen drei verschlossenen Türen. Er weiß, dass sich hinter zweien als Trostpreis jeweils eine Ziege verbirgt, während hinter der verbleibenden Tür als Hauptpreis ein Auto bereitsteht. Der Kandidat wählt nun eine Tür, ohne diese zu öffnen. Daraufhin öffnet der Spielleiter eine der beiden anderen Türen und eine Ziege erscheint. Der Spielleiter gibt nun dem Kandidaten die Chance, seine bisherige Wahl zu überdenken und erneut eine der drei Türen auszuwählen, um dann den hinter dieser Tür befindlichen Preis zu erhalten. Welches ist die beste Strategie für den Kandidaten um den Hauptpreis zu gewinnen, und mit welcher Wahrscheinlichkeit bekommt er ihn? \square

7.13 Aufgabe (zu einer möglichen Fehlinterpretation der Markov-Eigenschaft). Es sei $(\Omega, P) = (\Omega_1 \times \Omega_2 \times \Omega_3, P_1 \otimes P_{2|1} \otimes P_{3|2})$ ein Markovsches Kopplungsmodell, und es seien $A = A_1 \times \Omega_2 \times \Omega_3, B = \Omega_1 \times B_2 \times \Omega_3, C = \Omega_1 \times \Omega_2 \times C_3$ Ereignisse in Ω mit $P(AB) > 0$. Dann gilt

$$(1) \quad P(C|AB) = P(C|B)$$

(a) stets wenn B_2 einelementig ist,

(b) nicht zum Beispiel für geeignete A, B, C im folgenden Modell für die Trefferzahlensummen bei 3 sukzessiven fairen Münzwürfen: $\Omega_j := \{0, \dots, j\}$ für $j \in \{1, 2, 3\}$, $P_1 := B_{\frac{1}{2}} = \frac{1}{2}(\delta_0 + \delta_1)$, $P_{j|j-1}(\cdot | x) := \frac{1}{2}(\delta_x + \delta_{x+1})$ für $j \in \{2, 3\}$ und $x \in \Omega_{j-1}$. \square

Die folgende Definition fasst noch gewisse sich aus Zufallsgrößen ergebende bedingte Wahrscheinlichkeiten in oft bequemer Weise zusammen, siehe etwa die damit mögliche konzise Definition 8.1 der stochastischen Unabhängigkeit.

7.14 Definition. Es seien $X : \Omega \rightarrow \mathcal{X}$ und $Y : \Omega \rightarrow \mathcal{Y}$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Ist $K \in \text{Mark}(\mathcal{Y}, \mathcal{X})$ mit

$$(1) \quad \mathbb{P}(X \in A | Y = y) = K(A | y) \quad \text{für } A \subseteq \mathcal{X} \text{ und } \mathbb{P}(Y = y) > 0,$$

so heißt K eine **faktorierte bedingte Verteilung** von X gegeben Y . Die Menge all dieser bezeichnen wir mit $\mathbb{P}(X \in \cdot | Y = \cdot)$. \square

7.15 Beispiele. (a) In Satz 7.2 ist $P_{2|1}$ eine faktorierte bedingte Verteilung der zweiten Koordinatengröße gegeben die erste, siehe 7.2(3).

(b) Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum, $X : \Omega \rightarrow \mathcal{X}$ eine Zufallsgröße, und $B \subseteq \Omega$ ein Ereignis mit $\mathbb{P}(B) > 0$. Dann heißt, in Ergänzung zu Definition 6.2, $\mathbb{P}(X \in \cdot | B) \in \text{Prob}(\mathcal{X})$ die **elementare bedingte Verteilung** von X gegeben B . Ist nun auch $\mathbb{P}(B^c) > 0$, so definiert mit $Y := \mathbf{1}_B : \Omega \rightarrow \{0, 1\}$ dann

$$K(A | y) := \left\{ \begin{array}{ll} \mathbb{P}(X \in A | B) & \text{falls } y = 1, \\ \mathbb{P}(X \in A | B^c) & \text{falls } y = 0 \end{array} \right\} \quad \text{für } A \subseteq \mathcal{X} \text{ und } y \in \{0, 1\}$$

eine faktorierte bedingte Verteilung von X gegeben Y . \square

Interessantere und nützliche Beispiele faktorisierter bedingter Verteilungen bringen die Aufgaben 8.3, 9.15, 9.32, 9.33.

7.16 Bemerkung. Gegeben sei die Situation aus Definition 7.14.

(a) In Verallgemeinerung von Beispiel 7.15(a) ist mit $P := (Y, X) \square \mathbb{P}$, $\Omega_1 := \mathcal{Y}$, und $\Omega_2 := \mathcal{X}$ jedes $P_{2|1}$ gemäß 7.2(b) eine faktorierte bedingte Verteilung von X gegeben Y , denn für $A \subseteq \mathcal{X}$ und $\mathbb{P}(Y = y) > 0$ gilt mit 7.2(3) im letzten Schritt

$$\begin{aligned} \text{L.S.7.14(1)} &= \frac{\mathbb{P}(Y = y, X \in A)}{\mathbb{P}(Y = y)} = \frac{P(\{y\} \times A)}{P(\{y\} \times \mathcal{X})} \\ &= P(\Omega_1 \times A | \{y\} \times \mathcal{X}) = P_{2|1}(A | y). \end{aligned}$$

Insbesondere ist $\mathbb{P}(X \in \cdot | Y = \cdot)$ nichtleer.

(b) Sind $K, L \in \mathbb{P}(X \in \cdot | Y = \cdot)$, so gilt nach 7.14(1) mit einer Schreibweise aus 5.4

$$(1) \quad K(A | \cdot) =_{Y \square \mathbb{P}} L(A | \cdot) \quad \text{für } A \subseteq \mathcal{X},$$

und damit sogar, wegen (1) und 3.3(14) im letzten Schritt,

$$\begin{aligned} & (Y \square \mathbb{P})(\{y \in \mathcal{Y} : K(\cdot | y) = L(\cdot | y)\}) \\ &= (Y \square \mathbb{P})\left(\bigcap_{A \subseteq \mathcal{X}} \{y \in \mathcal{Y} : K(A | y) = L(A | y)\}\right) = 1. \end{aligned}$$

Speziell im Fall von $\mathbb{P}(Y = y) > 0$ für jedes $y \in \mathcal{Y}$ ist also $\mathbb{P}(X \in \cdot | Y = \cdot)$ einelementig.

(c) **Vereinbarung.** Es sei $K \in \mathbb{P}(X \in \cdot | Y = \cdot)$. Man schreibt dann, etwas unlogisch aber suggestiv, auch

$$(2) \quad \begin{aligned} \mathbb{P}(X \in \cdot | Y = \cdot) &:= K, \\ \mathbb{P}(X \in A | Y = \cdot) &:= K(A | \cdot) \quad \text{für } A \subseteq \mathcal{X}, \\ \mathbb{P}(X \in \cdot | Y = y) &:= K(\cdot | y) \quad \text{für } y \in \mathcal{Y}, \end{aligned}$$

und man nennt dann auch noch, unter Verwendung der Schreibweise (2), die $\text{Prob}(\mathcal{X})$ -wertige Zufallsgröße

$$(3) \quad \mathbb{P}(X \in \cdot | Y) := (\mathcal{Y} \ni y \mapsto K(\cdot | y)) \circ Y$$

eine (**nichtfaktorisierte**) **bedingte Verteilung**⁶ von X gegeben Y . □

⁶Dieser Begriff wird in diesen Vorlesungen nicht weiter benötigt und erscheint hier eigentlich nur, um einen Anschluss an weiterführende Bücher oder Vorlesungen zur allgemeinen WT herzustellen. Dort ist jedoch die Terminologie und Notation nicht ganz einheitlich; wir folgen hier etwa Feller (1971, pp. 160–161).

8 Stochastische Unabhängigkeit

Das didaktische Problem einer natürlichen Unabhängigkeitsdefinition, siehe unten Bemerkung 8.21, wird mit 8.1 elegant gelöst, jedoch um den Preis einer Verwendung der etwas abstrakten Definition 7.14 (was nach deren Einsetzen zur Bedingung 8.2(1) führt).

8.1 Definition. Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum. Eine Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ heißt von einer Zufallsgröße $Y : \Omega \rightarrow \mathcal{Y}$ (**stochastisch**) **unabhängig**, in Zeichen $X \perp\!\!\!\perp Y$, falls $X \square \mathbb{P}$, aufgefasst als eine im ersten Argument konstante Funktion auf $\mathcal{Y} \times 2^{\mathcal{X}}$, eine faktorisierte bedingte Verteilung von X gegeben Y ist. Eine Familie $X_{\bullet} = (X_i : i \in I)$ von Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$ heißt (**stochastisch**) **unabhängig** falls $X_{i_0} \perp\!\!\!\perp (X_i : i \in I \setminus \{i_0\})$ für jedes $i_0 \in I$ gilt. Ereignisse heißen (**stochastisch**) **unabhängig** falls ihre Indikatoren es sind. \square

Man verwendet Sprechweisen wie in der Linearen Algebra für die lineare Unabhängigkeit von Vektoren, wie zum Beispiel: Für $X \perp\!\!\!\perp Y$ sagt man auch, in einer durch 8.2(i) \Leftrightarrow (ii) unten gerechtfertigten symmetrischen Sprechweise, “ X und Y sind voneinander unabhängig”. Für “ $(X_i : i \in I)$ ist unabhängig” sagt man auch, etwas ungenau, “die X_i mit $i \in I$ sind unabhängig”.

Damit kann man Definition 8.1 auch kurz und nur ein wenig unpräzise so formulieren: 1. Zwei Zufallsgrößen X und Y heißen unabhängig wenn die bedingte Verteilung von X gegeben Y gleich der unbedingten Verteilung von X ist. 2. Beliebige viele Zufallsgrößen heißen unabhängig wenn eine jede von ihnen unabhängig von allen¹ anderen ist.

Wir studieren zunächst den Fall zweier Zufallsgrößen:

8.2 Satz. *Es seien $X : \Omega \rightarrow \mathcal{X}$ und $Y : \Omega \rightarrow \mathcal{Y}$ Zufallsgrößen auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Dann sind die folgenden sechs Aussagen äquivalent:*

- (i) $X \perp\!\!\!\perp Y$, (ii) $Y \perp\!\!\!\perp X$, (iii) (X, Y) unabhängig,
- (iv) $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ für $A \subseteq \mathcal{X}$ und $B \subseteq \mathcal{Y}$,
- (v) $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ für $x \in \mathcal{X}$ und $y \in \mathcal{Y}$,
- (vi) $(X, Y) \square \mathbb{P} = (X \square \mathbb{P}) \otimes (Y \square \mathbb{P})$.

Beweis. Die Bedingung (i) bedeutet ausgeschrieben

$$(1) \quad \mathbb{P}(X \in A | Y = y) = \mathbb{P}(X \in A) \quad \text{für } A \subseteq \mathcal{X} \text{ und } \mathbb{P}(Y = y) > 0,$$

also, nach Einsetzen der Definition der bedingten Wahrscheinlichkeit,

$$\mathbb{P}(X \in A, Y = y) = \mathbb{P}(X \in A)\mathbb{P}(Y = y) \quad \text{für } A \subseteq \mathcal{X} \text{ und } \mathbb{P}(Y = y) > 0,$$

und letzteres ist äquivalent zu

$$(2) \quad \mathbb{P}(X \in A, Y = y) = \mathbb{P}(X \in A)\mathbb{P}(Y = y) \quad \text{für } A \subseteq \mathcal{X} \text{ und } y \in \mathcal{Y},$$

¹Nicht “jeder”, denn damit würden wir nur “ $X_{i_0} \perp\!\!\!\perp X_i$ für jedes $i_0 \in I$ und $i \in I \setminus \{i_0\}$ ” fordern, was für $\#I \geq 3$ eine echt schwächere Bedingung ist; siehe dazu Aufgabe 8.18(c).

da die Gleichung in (2) für die $y \in \mathcal{Y}$ mit $\mathbb{P}(Y = y) = 0$ sowieso gilt.

(i) \Rightarrow (iv): Summation $\sum_{y \in B} \dots$ in (2).

(iv) \Rightarrow (v): Spezialisierung $A := \{x\}$, $B := \{y\}$.

(v) \Rightarrow (i): Summation $\sum_{x \in A} \dots$ in (v) liefert (2).

(i) \Leftrightarrow (ii): Klar mit (i) \Leftrightarrow (iv), da (iv) invariant unter dem Rollentausch $X \leftrightarrow Y$ ist.

(v) \Leftrightarrow (vi): Für $z = (x, y) \in \mathcal{X} \times \mathcal{Y}$ gilt

$$\left((X, Y) \square \mathbb{P} \right) (\{z\}) = \text{L.S.}(v),$$

und, im ersten Schritt mit der Definition der unabhängigen Kopplung \otimes in 7.9(c), siehe 7.9(5) mit $n = 2$,

$$\left((X \square \mathbb{P}) \otimes (Y \square \mathbb{P}) \right) (\{z\}) = (X \square \mathbb{P}) (\{x\}) \cdot (Y \square \mathbb{P}) (\{y\}) = \text{R.S.}(v).$$

Also ergibt sich die behauptete Äquivalenz (v) \Leftrightarrow (vi) mit Satz 3.6.

(i) \wedge (ii) \Leftrightarrow (iii): Mit der Definition der Unabhängigkeit der Familie (X, Y) im ersten Schritt ergibt sich die Äquivalenzkette

$$(iii) \Leftrightarrow X \perp\!\!\!\perp (Y) \text{ und } Y \perp\!\!\!\perp (X) \Leftrightarrow X \perp\!\!\!\perp Y \text{ und } Y \perp\!\!\!\perp X,$$

wobei der letzte Klammerweglassschritt etwa mit (i) \Leftrightarrow (v) gerechtfertigt werden kann. □

8.3 Aufgabe. Es seien $X \sim B_{n,p}$ und $Y \sim B_{m,p}$ unabhängige \mathbb{N}_0 -wertige Zufallsgrößen, mit $m, n \in \mathbb{N}$ und $p \in [0, 1]$. Dann gilt, wie weiter unten in Satz 9.8 und nochmals in Aufgabe 9.14 gezeigt und hier als bekannt angenommen, $X + Y \sim B_{m+n,p}$. Was ist hier $\mathbb{P}(X \in \cdot \mid X + Y = \cdot)$? □

Das - auch für sich interessante - Ergebnis von Aufgabe 8.3 wird in den Beweisen von Lemma 12.16 und Satz 13.15 benutzt.

8.4 Aufgabe. Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum. Gibt es dann eine Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ welche

(a) von sich selbst unabhängig ist,

(b) sogar von jeder Zufallsgröße $Y : \Omega \rightarrow \mathcal{Y}$ unabhängig ist? □

Unter anderem zur Vorbereitung von Satz 8.6 zeigen wir:

8.5 Lemma (Vererbung auf Umindizierungen und Unterfamilien). *Es sei $(X_i : i \in I)$ eine Familie unabhängiger Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$ auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Dann ist für jede injektive Funktion $\varphi : J \rightarrow I$ auch $(X_{\varphi(j)} : j \in J)$ unabhängig. Speziell ist $(X_i : i \in J)$ für $J \subseteq I$ unabhängig.*

Beweis. Der Spezialfall ergibt sich durch Betrachtung von $\varphi := \text{id}_J$.

Für den allgemeinen Fall sei $\varphi : J \rightarrow I$ injektiv und es sei $j_0 \in J$. Mit $i_0 := \varphi(j_0)$ und $Z := (X_{\varphi(j)} : j \in J \setminus \{j_0\}) : \Omega \rightarrow \prod_{j \in J \setminus \{j_0\}} \mathcal{X}_{\varphi(j)} =: \mathcal{Z}$ ist dann

$$(1) \quad X_{i_0} \perp\!\!\!\perp Z$$

zu zeigen. Wir betrachten dazu $Y := ((X_i : i \in I \setminus \{i_0\}) : \Omega \rightarrow \times_{i \in I \setminus \{i_0\}} \mathcal{X}_i =: \mathcal{Y}$. Wegen der Injektivität von φ gilt $\varphi(j) \in I \setminus \{i_0\}$ für $j \in J \setminus \{j_0\}$, und daher ist $\pi : \mathcal{Y} \rightarrow \mathcal{Z}$ durch $\pi(y) := (y_{\varphi(j)} : j \in J \setminus \{j_0\})$ für $y \in \mathcal{Y}$ wohldefiniert, und es gilt $Z = \pi \circ Y$. Damit erhalten wir für $A \subseteq \mathcal{X}_{i_0}$ und $C \subseteq \mathcal{Z}$, unter Verwendung von $X_{i_0} \perp\!\!\!\perp Y$ und Kriterium 8.2(iv) im zweiten Schritt,

$$\begin{aligned} \mathbb{P}(X_{i_0} \in A, Z \in C) &= \mathbb{P}(X_{i_0} \in A, Y \in \pi^{-1}[C]) = \mathbb{P}(X_{i_0} \in A)\mathbb{P}(Y \in \pi^{-1}[C]) \\ &= \mathbb{P}(X_{i_0} \in A)\mathbb{P}(Z \in C); \end{aligned}$$

also gilt (1). □

In der nun folgenden Verallgemeinerung der Unabhängigkeitskriterien 8.2(iv,v,vi), und damit auch bei deren Anwendungen, beschränken wir uns der Einfachheit halber auf endliche Familien:

8.6 Satz. *Es sei $(X_i : i \in I)$ eine endliche Familie von Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$ auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Dann sind die folgenden fünf Aussagen äquivalent:*

- (i) $(X_i : i \in I)$ ist unabhängig,
- (ii) $\mathbb{P}(X_i \in A_i \text{ für } i \in I) = \prod_{i \in I} \mathbb{P}(X_i \in A_i)$ falls $A_i \subseteq \mathcal{X}_i$ für $i \in I$,
- (iii) $\mathbb{P}(X_i = x_i \text{ für } i \in I) = \prod_{i \in I} \mathbb{P}(X_i = x_i)$ für $x \in \times_{i \in I} \mathcal{X}_i$,
- (iv) $\mathbb{P}(X_i = x_i \text{ für } i \in I) = \prod_{i \in I} \mathbb{P}(X_i = x_i)$ falls die rechte Seite > 0 ist,
- (v) $(X_i : i \in I) \square \mathbb{P} = \bigotimes_{i \in I} (X_i \square \mathbb{P})$.

Beweis. (i) \Rightarrow (ii): Wegen Lemma 8.5 kann ohne Einschränkung $I = \{1, \dots, n\}$ für ein $n \in \mathbb{N}$ angenommen werden. Damit gelte (i). Ist $k \in \{1, \dots, n\}$, so ist (X_1, \dots, X_k) unabhängig wegen Lemma 8.5, und für $A_i \subseteq \mathcal{X}_i$ für $i \in \{1, \dots, k\}$ folgt mit Kriterium 8.2(iv) im zweiten Schritt

$$\begin{aligned} \mathbb{P}(X_i \in A_i \text{ für } i \in \{1, \dots, k\}) &= \mathbb{P}\left(\left((X_1, \dots, X_{k-1}) \in \times_{i=1}^{k-1} A_i, X_k \in A_k\right)\right) \\ &= \mathbb{P}\left(\left((X_1, \dots, X_{k-1}) \in \times_{i=1}^{k-1} A_i\right)\right) \mathbb{P}(X_k \in A_k) \\ &= \mathbb{P}(X_i \in A_i \text{ für } i \in \{1, \dots, k-1\}) \mathbb{P}(X_k \in A_k). \end{aligned}$$

Damit erhält man (ii) durch Induktion.

(ii) \Rightarrow (iii): Spezialisierung $A_i = \{x_i\}$.

(iii) \Rightarrow (i): (iii) gelte, und es sei $i_0 \in I$. Mit $Y := ((X_i : i \in I \setminus \{i_0\}) : \Omega \rightarrow \times_{i \in I \setminus \{i_0\}} \mathcal{X}_i =: \mathcal{Y}$ ist dann

$$(1) \quad X_{i_0} \perp\!\!\!\perp Y$$

zu zeigen. Summation $\sum_{x_{i_0} \in \mathcal{X}_{i_0}} \dots$ in (iii) liefert

$$(2) \quad \mathbb{P}(Y = y) = \prod_{i \in I \setminus \{i_0\}} \mathbb{P}(X_i = y_i) \quad \text{für } y \in \mathcal{Y}.$$

Ist nun $x \in \mathcal{X}_{i_0}$ und $y \in \mathcal{Y}$, so erhalten wir mit $z \in \times_{i \in I} \mathcal{X}_i$ definiert durch $z_{i_0} := x$ und $z|_{I \setminus \{i_0\}} := y$, unter Anwendung von (iii) im ersten Schritt und (2) im dritten,

$$\begin{aligned} \mathbb{P}(X_{i_0} = x, Y = y) &= \prod_{i \in I} \mathbb{P}(X_i = z_i) = \mathbb{P}(X_{i_0} = x) \cdot \text{R.S.}(2) \\ &= \mathbb{P}(X_{i_0} = x) \mathbb{P}(Y = y); \end{aligned}$$

also gilt (1) nach Kriterium 8.2(v).

(iii) \Leftrightarrow (iv): Klar, da im Fall von $\mathbb{P}(X_i = x_i) = 0$ für ein $i \in I$ die betrachtete Gleichung automatisch gilt.

(iii) \Leftrightarrow (v): Wie der Beweis von 8.2(v) \Leftrightarrow (vi). □

Der Genauigkeit halber halten wir noch fest:

8.7 Bemerkung. Die Unabhängigkeit einer Familie $(X_i : i \in I)$ von Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$ hängt nicht von den jeweils gewählten Zielmengen \mathcal{X}_i ab:

Ist $(X_i : i \in I)$ als Familie von Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$ unabhängig, und ist auch $\mathcal{Y}_i \supseteq \mathcal{X}_i[\Omega]$ für $i \in I$, so ist $(X_i : i \in I)$ auch als Familie von Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{Y}_i$ unabhängig.

Denn das Erfülltsein von Kriterium 8.6(iv) angewandt auf die \mathcal{X}_i beziehungsweise die \mathcal{Y}_i ändert sich nicht, wenn diese jeweils durch $X_i[\Omega]$ ersetzt werden. □

8.8 Beispiel. Koordinatengrößen in unabhängigen Kopplungsmodellen sind unabhängig, siehe 7.9(12).

Umgekehrt ist die Verbundverteilung einer endlichen unabhängigen Familie von Zufallsgrößen nach 8.6(i) \Rightarrow (v) das Produkt ihrer Randverteilungen, also insbesondere eine unabhängige Kopplung. □

Wegen der obigen Charakterisierung 8.6(v) legen, anders als im durch Aufgabe 5.14 oder 7.11(6,11,4,10) exemplifizierten allgemeinen Fall, unter der Unabhängigkeitsannahme die Randverteilungen schon die Verbundverteilung fest. Genauer:

8.9 Korollar. *Es seien $X = (X_i : i \in I)$ und $Y = (Y_i : i \in I)$ Zufallsgrößen mit Werten in $\times_{i \in I} \mathcal{X}_i$. Es sei $(X_i : i \in I)$ unabhängig und I endlich. Dann gilt die Äquivalenz*

$$X \sim Y \Leftrightarrow (Y_i : i \in I) \text{ unabhängig und } X_i \sim Y_i \text{ für } i \in I.$$

Beweis. Es sei (Ω_1, \mathbb{P}_1) bzw. (Ω_2, \mathbb{P}_2) der Wahrscheinlichkeitsraum auf dem X bzw. Y definiert ist.

“ \Leftarrow ”: Mit der jeweiligen Unabhängigkeitsannahme im ersten und im dritten Schritt, und mit $X_i \sim Y_i$ im zweiten, folgt

$$X \square \mathbb{P}_1 = \bigotimes_{i \in I} (X_i \square \mathbb{P}_1) = \bigotimes_{i \in I} (Y_i \square \mathbb{P}_2) = Y \square \mathbb{P}_2.$$

“ \Rightarrow ”: Aus $X \sim Y$ folgt nach Beispiel 5.13 zunächst $X_i \sim Y_i$ für $i \in I$, und damit im dritten Schritt auch

$$Y \square \mathbb{P}_2 = X \square \mathbb{P}_1 = \bigotimes_{i \in I} (X_i \square \mathbb{P}_1) = \bigotimes_{i \in I} (Y_i \square \mathbb{P}_2),$$

und damit die Unabhängigkeit von $(Y_i : i \in I)$ nach Satz 8.6(v) \Rightarrow (i). □

In konkreten Fällen ist der Nachweis einer Unabhängigkeit nicht immer leicht. Manchmal helfen dabei Lemma 8.5 und der folgende Satz.

8.10 Satz (Weitere Vererbungskriterien). *Es sei $(X_i : i \in I)$ eine endliche Familie unabhängiger Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$. Dann sind auch die folgenden Familien von Zufallsgrößen jeweils unabhängig.*

(a) **Blockungen:** $((X_i : i \in I_j) : j \in J)$ für jede paarweise disjunkte endliche Familie $(I_j : j \in J)$ von Teilmengen von I .

(b) **Funktionen:** $(T_i \circ X_i : i \in I)$ mit Funktionen $T_i : \mathcal{X}_i \rightarrow \mathcal{Y}_i$.

Beweis. (a) Mit $Y_j := ((X_i : i \in I_j) : \Omega \rightarrow \times_{i \in I_j} \mathcal{X}_i =: \mathcal{Y}_j$ für $j \in J$ gilt für $y \in \times_{j \in J} \mathcal{Y}_j$, also $y_j = (y_{ji} : i \in I_j) \in \mathcal{Y}_j$ für $j \in J$, dann

$$\begin{aligned} \mathbb{P}(Y_j = y_j \text{ für } j \in J) &= \mathbb{P}(X_i = y_{ji} \text{ für } j \in J \text{ und } i \in I_j) \\ &= \prod_{j \in J} \prod_{i \in I_j} \mathbb{P}(X_i = y_{ji}) = \prod_{j \in J} \mathbb{P}(Y_j = y_j) \end{aligned}$$

wegen 8.6(i) \Rightarrow (iii) im zweiten Schritt, und wegen der Unabhängigkeit jeder Familie $(X_{ji} : i \in I_j)$ nach Lemma 8.5 und wieder 8.6(i) \Rightarrow (iii) im dritten. Also ist $(Y_j : j \in J)$ unabhängig nach 8.6(iii) \Rightarrow (i).

(b) Ist $B_i \subseteq \mathcal{Y}_i$ für $i \in I$, so gilt

$$\begin{aligned} \mathbb{P}(T_i \circ X_i \in B_i \text{ für } i \in I) &= \mathbb{P}(X_i \in T_i^{-1}[B_i] \text{ für } i \in I) \\ &= \prod_{i \in I} \mathbb{P}(X_i \in T_i^{-1}[B_i]) = \prod_{i \in I} \mathbb{P}(T_i \circ X_i \in B_i) \end{aligned}$$

wegen 8.6(i) \Rightarrow (ii) im zweiten Schritt; also ist $(T_i \circ X_i : i \in I)$ unabhängig wegen 8.6(ii) \Rightarrow (i). □

8.11 Beispiel. Für unabhängige $X_1, \dots, X_8 : \Omega \rightarrow \mathbb{R}$ sind jeweils

$$\begin{aligned} X_8, X_4, X_2, X_1, X_3, X_5 &\text{ unabhängig,} \\ (X_8, X_2), (X_1, X_5), X_3 &\text{ unabhängig,} \\ X_8 + X_2, X_1 \cdot X_5, \sin(X_3) &\text{ unabhängig,} \end{aligned}$$

nach sukzessiver Anwendung von Lemma 8.5 und Satz 8.10(a,b). □

8.12 Aufgabe (Unabhängige Würfelsummen modulo 6). Auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) seien $X_1, X_2, X_3, X_4 : \Omega \rightarrow \mathcal{X} := \{0, \dots, 5\}$ unabhängig und jeweils $U_{\mathcal{X}}$ -verteilt. Es bezeichne \oplus die Addition modulo 6 auf \mathcal{X} .

- (a) Wie ist $X_1 \oplus X_3$ verteilt?
 (b) Sind $X_1 \oplus X_3$ und $X_2 \oplus X_4$ unabhängig?
 (c) Sind $X_1 \oplus X_3$ und $X_2 \oplus X_3$ unabhängig?
 (d) Sind $X_1 + X_3$ und $X_2 + X_3$ unabhängig?
 (e) Bleibt Ihre Antwort für (c) gültig, wenn in der Voraussetzung die Gleichverteilung $U_{\mathcal{X}}$ durch eine beliebige Verteilung $P \in \text{Prob}(\mathcal{X})$ ersetzt wird?

Hinweis: Für in (a) oder (c) zu betrachtende Ereignisse A könnte die Darstellung $A = \bigcup_{j=0}^5 A \cap \{X_3 = j\}$ nützlich sein. \square

Wir spezialisieren nun zur Betrachtung der Unabhängigkeit von endlich vielen Ereignissen. Nach Lemma 8.5 können wir uns dabei auf die Indexmengen $\{1, \dots, n\}$ mit $n \in \mathbb{N}$ beschränken.

8.13 Korollar. *Es seien A_1, \dots, A_n Ereignisse in einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Dann sind die folgenden Bedingungen äquivalent:*

- (i) A_1, \dots, A_n unabhängig,
 (ii) $\mathbb{P}(\mathbf{1}_{A_i} = \varepsilon_i \text{ für } i \in \{1, \dots, n\}) = \prod_{i=1}^n \mathbb{P}(\mathbf{1}_{A_i} = \varepsilon_i)$ für $\varepsilon \in \{0, 1\}^n$,
 (iii) Bedingung (i) mit anderer Reihenfolge der A_i , oder mit beliebigen der A_i durch A_i^c ersetzt.

Beweis. (i) \Leftrightarrow (ii): Definition 8.1, kombiniert mit Kriterium 8.6(iii) angewandt auf die Familie $(\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n})$ der $\{0, 1\}$ -wertigen Zufallsgrößen $\mathbf{1}_{A_i}$.

(ii) \Leftrightarrow (iii): Lemma 8.5, beziehungsweise Satz 8.10(b) kombiniert mit $\mathbf{1}_{A_i^c} = 1 - \mathbf{1}_{A_i}$ und $\mathbf{1}_{A_i} = 1 - \mathbf{1}_{A_i^c}$. \square

Die 2^n Gleichungen in 8.13(ii), von denen die *eine* mit $\varepsilon = (1, \dots, 1)$ sich nach der Rechenregel 5.10(3) für Indikatoren in der Form

$$(1) \quad \mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i)$$

schreiben lässt, und die übrigen $2^n - 1$ sich wegen $\mathbf{1}_{A_i} = 0 \Leftrightarrow \mathbf{1}_{A_i^c} = 1$ durch Ersetzung gewisser A_i durch A_i^c ergeben, können durch $2^n - n - 1$ Gleichungen ersetzt werden. Wir notieren zuerst ausführlich den wichtigen Spezialfall $n = 2$:

8.14 Satz. *Es seien A und B Ereignisse in einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Dann sind die folgenden 32 Bedingungen äquivalent:*

- (i) $A \perp B$, (ii) $\mathbf{1}_A \perp \mathbf{1}_B$,
 (iii) $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$,
 (iv) $\mathbb{P}(B) = 0$ oder $\mathbb{P}(A|B) = \mathbb{P}(A)$,

und die aus den obigen Bedingungen durch Vertauschung von A mit B , oder A mit A^c , oder B mit B^c hervorgehenden sieben Varianten.

Beweis. (i) \Leftrightarrow (ii) ist Definition 8.1.

(i) \Rightarrow (iii): 8.13(1) mit $n = 2$.

(iii) ist äquivalent zu jeder seiner sieben Varianten, trivialerweise bezüglich der Vertauschung von A mit B , und wegen $\mathbb{P}(A^c B) - \mathbb{P}(A^c)\mathbb{P}(B) = \mathbb{P}(B) - \mathbb{P}(AB) - (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)$ bezüglich der Vertauschung von A mit A^c .

(iii) \Rightarrow (i): (iii) und seine drei durch Komplementierungen entstehenden Varianten liefern die vier Gleichungen in 8.13(ii) mit $n = 2$.

(iii) \Leftrightarrow (iv) ist Definition 6.2. □

8.15 Beispiel (Rauchen und Lungenkrebs). Für ein Kollektiv von Menschen bezeichne A das Ereignis, im Laufe des Lebens an Lungenkrebs zu erkranken, und B das Ereignis, über einen längeren Zeitraum geraucht zu haben. Dann ist typischerweise A nicht unabhängig von B , sondern **positiv abhängig** im Sinne von $P(A|B) > P(A)$. Folgt *daraus* schon, dass Rauchen Lungenkrebs begünstigt?

Wenn ja, dann folgt wegen der Äquivalenz

$$(1) \quad P(A|B) > P(A) \Leftrightarrow P(B|A) > P(B)$$

auch, dass Lungenkrebs Rauchen begünstigt. □

8.16 Aufgabe. Gegen seiner Meinung nach voreilige Schlüsse auf eine Begünstigung von Lungenkrebs durch Rauchen wandte sich der berühmte Statistiker und Genetiker Sir Ronald A. Fisher (1890–1962) in Fisher (1958 a,b) mit Daten aus der Zwillingsforschung:

Es wird davon ausgegangen, dass die Entstehung von Lungenkrebs durch genetische Faktoren begünstigt wird. Wenn nun auch das Rauchverhalten genetisch beeinflusst würde, dann wäre es zumindest denkbar, dass gewisse genetische Veranlagungen eines Menschen als Ursache² sowohl Rauchen als auch Lungenkrebs begünstigen könnten, ohne dass Rauchen irgendwie den Krebs verursachte.

Zur Untersuchung der Abhängigkeit des Rauchverhaltens von der genetischen Veranlagung werden in Fisher (1958 a) Beobachtungen an 51 eineiigen und 31 zweieiigen Zwillingspaaren betrachtet. Mit Definition 2.2 naiv auf $n = 51$ bzw. $n = 31$ angewandt erhalten wir im Kollektiv der eineiigen Zwillingspaare bzw. dem der zweieiigen

$$\begin{aligned} P_1(\text{“Gleiche Rauchgewohnheiten”}) &= \frac{33}{51} = 0.65, \\ P_2(\text{“Gleiche Rauchgewohnheiten”}) &= \frac{11}{31} = 0.35, \end{aligned}$$

auf ganze Prozent gerundet. Wenn man nun annimmt, dass der Effekt ähnlicher Umwelteinflüsse unter eineiigen Zwillingspaaren ungefähr genauso ist wie unter zweieiigen, dann wäre die deutlich höhere Wahrscheinlichkeit der Gleichheit des Rauchverhaltens unter eineiigen nur mit ihrer genetischen Gleichheit erklärbar, und würde

²Wir gehen hier darüber hinweg, dass der Begriff “Ursache” nicht zu den wohldefiniertesten gehört.

also eine Beeinflussung der Rauchgewohnheiten durch die genetische Veranlagung belegen.

(a) Es wäre jedoch denkbar, dass sich eineiige Zwillinge untereinander stärker beeinflussen als zweieiige. In Fisher (1958 b) werden daher weitere Beobachtungen an teils gemeinsam, teils getrennt aufgewachsenen eineiigen Zwillingspaaren untersucht. Die dort betrachteten eineiigen Zwillingspaare als ein Kollektiv ansehend, ist in diesem das Ereignis “Gleiche Rauchgewohnheiten” vom Ereignis “Gemeinsam aufgewachsen” unabhängig?

(b) Der in Fisher (1958 b) betrachtete Datensatz enthält auch Beobachtungen an zweieiigen Zwillingspaaren. Damit ergibt sich hier, zur Unterscheidung einen Index “b” anbringend,

$$\begin{aligned} P_{b,1}(\text{“Gleiche Rauchgewohnheiten”}) &= \frac{44}{53} = 0.83, \\ P_{b,2}(\text{“Gleiche Rauchgewohnheiten”}) &= \frac{9}{18} = 0.50. \end{aligned}$$

Wie erklären Sie sich die Unterschiede zu den obigen Werten 0.65 und 0.35? □

Das Analogon von Kriterium 8.14(iii) für drei Ereignisse ist komplizierter als man vielleicht zunächst denken mag:

8.17 Aufgabe. Ereignisse A, B, C in einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) sind genau dann unabhängig, wenn neben ihrer **paarweisen Unabhängigkeit**

$$(1) \quad \mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(AC) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(BC) = \mathbb{P}(B)\mathbb{P}(C)$$

auch

$$(2) \quad \mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

gilt. □

8.18 Aufgabe (Einfache Beispiele zur Unabhängigkeit). (a) Geben Sie im Standardmodell des *einfachen* Würfelwurfes ein nichttriviales Beispiel zweier unabhängiger Ereignisse an. Geht so etwas auch beim einmaligen Roulette-Spiel (Ergebnismenge $\{0, \dots, 36\}$)?

(b) Geben Sie im Standardmodell des einfachen Würfelwurfes drei *abhängige* Ereignisse A, B, C mit $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ an.

(c) Ein regelmäßiges Tetraeder sei wie folgt koloriert: Eine Fläche nur rot, eine Fläche nur blau, eine Fläche nur gelb, und eine Fläche mit allen drei Farben. Er werde einmal “tetraedert”. Es sei R das Ereignis, dass dabei die untenliegende Fläche die Farbe rot enthält. Entsprechend seien die Ereignisse B und G definiert. Sind R, B unabhängig? Sind R, G unabhängig? Sind B, G unabhängig? Sind R, B, G unabhängig? □

8.19 Aufgabe (Unschädliche Maschine). Eine Maschine (etwa ein Flugzeug oder ein Kernkraftwerk) arbeite unschädlich wenn mindestens eines von drei wichtigen Teilen funktioniert. Deren Ausfallwahrscheinlichkeiten seien 0.01, 0.02 und 0.03. Was können Sie über die Wahrscheinlichkeit dafür, dass die Maschine unschädlich ist, sagen

- (a) falls keine weiteren Voraussetzungen getroffen werden,
- (b) falls angenommen werden kann, dass die drei Teile unabhängig voneinander ausfallen?

Bemerkung: Dieses einfache Beispiel illustriert den in der angewandten Wahrscheinlichkeitsrechnung vielleicht häufigsten und oft schwerwiegenden Fehler: Die ungerechtfertigte Annahme der Unabhängigkeit. \square

Die nach den Aufgaben 8.17 und 8.18(b,c) naheliegende Verallgemeinerung des Kriteriums 8.14(iii) auf beliebig endlich viele Ereignisse ist:

8.20 Satz. Ereignisse A_1, \dots, A_n in einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) sind genau dann unabhängig, wenn

$$(1) \quad \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i) \quad \text{für } I \subseteq \{1, \dots, n\} \text{ mit } \#I \geq 2$$

gilt.

Beweis. Sind A_1, \dots, A_n unabhängig, so ist für $I \subseteq \{1, \dots, n\}$ auch $(A_i : i \in I)$ unabhängig, und die Gleichung in (1) folgt mit 8.13(ii) angewandt auf I statt $\{1, \dots, n\}$.

Gilt umkehrt (1), so schließt man ähnlich wie in Aufgabe 8.17, mit Induktion über die Anzahl der Komplementzeichen, auf die Unabhängigkeit der A_i . \square

8.21 Bemerkung. Die Wahl einer möglichst direkten Definition der Unabhängigkeit von Ereignissen, also ohne Rückgriff auf eine vorherige Definition der Unabhängigkeit von Zufallsgrößen, ist anscheinend ein nichttriviales didaktisches Problem.

(a) Im einfachsten Fall zweier Ereignisse A, B ist vielleicht

$$(1) \quad \mathbb{P}(A|B) = \mathbb{P}(A) \quad \text{und} \quad \mathbb{P}(B|A) = \mathbb{P}(B)$$

die natürlichste Wahl, welche jedoch nur im Fall von $\mathbb{P}(A)\mathbb{P}(B) > 0$ sinnvoll ist. Die Hinzufügung von

$$(2) \quad \text{, oder } \mathbb{P}(A) = 0 \text{ oder } \mathbb{P}(B) = 0$$

müsste man dann durch mathematische Eleganz rechtfertigen, etwa wie im folgenden wohl überzeugend, aber mit nur einfachsten Vorkenntnissen nicht voll verständlich:

Mit nur (1) als Definition der Unabhängigkeit könnte man Folgen $(A_n), (B_n)$ von Ereignissen finden, die im Sinne der punktweisen Konvergenzen $\mathbf{1}_{A_n} \rightarrow \mathbf{1}_A$ und $\mathbf{1}_{B_n} \rightarrow \mathbf{1}_B$ gegen die Ereignisse A, B konvergieren, für die weiter $A_n \perp B_n$ für $n \in \mathbb{N}$

gilt, aber nicht $A \perp B$, nämlich wegen $\mathbb{P}(A)\mathbb{P}(B) = 0$. Mit (1) und (2) als Definition, äquivalent also 8.14(iii), kann das nicht passieren, denn aus den Voraussetzungen ergibt sich dann $\mathbb{P}(A_n) \rightarrow \mathbb{P}(A)$, $\mathbb{P}(B_n) \rightarrow \mathbb{P}(B)$, und $\mathbb{P}(A_n B_n) \rightarrow \mathbb{P}(AB)$, nach einer in Satz 3.3 hinzufügbaren Rechenregel, also $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$ und damit $A \perp B$.

Weiter würde man sich wohl wundern, dass weder $\mathbb{P}(A) = 1$ noch $\mathbb{P}(B) = 1$ ein Sonderfall wäre.

(b) Für mehrere Ereignisse A_1, \dots, A_n hätte man 8.13(1) und seine $2^n - 1$ Varianten oder 8.20(1) zu motivieren. Vielleicht, die Definition von $A \perp B$ aus (a) voraussetzend, mit einer Forderung wie (im Fall $n = 3$)

A, B, C unabhängig $\stackrel{!}{\Rightarrow}$ A, B, C paarweise unabhängig, und AB, C unabhängig,

denn die erzwingt auch $\mathbb{P}(ABC) = \mathbb{P}(AB)\mathbb{P}(C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. □

Eine weitere Aufgabe

8.22 Aufgabe (Nichttransitive Würfel). Ute und Volker haben drei normale Würfel mit den ungewöhnlichen Beschriftungen $(1, 1, 5, 5, 9, 9)$, $(2, 2, 6, 6, 7, 7)$, $(3, 3, 4, 4, 8, 8)$ versehen. Sie wollen nun folgendes Spiel spielen: Zuerst wählt einer der beiden einen der drei Würfel, dann wählt der andere einen der zwei verbleibenden Würfel. Danach würfeln sie beide je einmal, und die höhere Augenzahl gewinnt. Volker will nun Ute den Vortritt bei der Würfelwahl lassen. Ist Volker ein Kavalier?

Hinweis: Modellieren Sie die Gewinnwahrscheinlichkeit für Ute durch $\mathbb{P}(U > V)$, wobei die auf einem geeigneten Wahrscheinlichkeitsraum (Ω, \mathbb{P}) zu definierenden Zufallsgrößen U, V die von Ute beziehungsweise Volker erzielten Augenzahlen angeben. Dabei werden U und V , und darf (Ω, \mathbb{P}) , vom Ergebnis der Würfelwahl abhängen. Welchen Würfel würde Volker jeweils wählen? □

9 Bernoulli-Ketten, Binomial-, Negativbinomial- und Multinomialverteilungen, Faltungen

Wir erinnern uns der Definition und Bemerkung 5.9 zu Indikatoren und Bernoulli-Verteilungen.

9.1 Definition (Bernoulli-Ketten). Es sei $n \in \mathbb{N}_0, p \in [0, 1]$, und $X_\bullet = (X_1, \dots, X_n)$ eine Familie unabhängiger und identisch je B_p -verteilter Zufallsgrößen X_i . Dann heißt X_\bullet **Bernoulli-Kette** mit den Parametern n (**Länge**) und p (**Trefferwahrscheinlichkeit**). \square

Solche Bernoulli-Ketten existieren nach Beispiel 8.8 sowie 7.9(11) mit $P_i := B_p$ für $i \in \{1, \dots, n\}$. Wir erinnern uns der Schreibweise $=_{\mathbb{P}}$ aus Bemerkung 5.4.

9.2 Satz. Es sei $n \in \mathbb{N}_0, p \in [0, 1]$, $X_\bullet = (X_1, \dots, X_n)$ eine Familie von Zufallsgrößen X_i auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) , und $X := (X_1, \dots, X_n)$ der zugehörige Zufallsvektor. Dann sind die folgenden vier Aussagen äquivalent:

- (i) X_\bullet ist eine Bernoulli-Kette mit den Parametern n und p .
- (ii) $X =_{\mathbb{P}} (\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n})$ mit unabhängigen Ereignissen A_i , je mit $\mathbb{P}(A_i) = p$.
- (iii) $X \square \mathbb{P} = B_p^{\otimes n}$.
- (iv) $\mathbb{P}(X = x) = p^{\sum_{i=1}^n x_i} \cdot (1 - p)^{n - \sum_{i=1}^n x_i}$ für $x \in \{0, 1\}^n$.

Beweis. (i) \Rightarrow (iii) folgt aus 8.6(i) \Rightarrow (v).

(iii) \Leftrightarrow (iv): Für $x \in \{0, 1\}^n$ ist

$$\text{R.S. (iv)} = \prod_{i=1}^n p^{x_i} (1 - p)^{1 - x_i}$$

die Dichte von $B_p^{\otimes n}$ an der Stelle x , wegen 5.9(2) und 7.9(5); also gilt (iii) \Leftrightarrow (iv).

(iii) \Rightarrow (i): (iii) gelte. Dann ist X verteilt wie der Koordinatengrößenvektor im unabhängigen Kopplungsmodell $B_p^{\otimes n}$, also folgt (i) mit 7.9(e).

(i) \Rightarrow (ii): (i) gelte. Für $i \in \{1, \dots, n\}$ sei $A_i := X_i^{-1}[\{1\}]$. Dann sind die Zufallsgrößen $\mathbf{1}_{A_i} = (X_i = 1)$ als Funktionen der X_i unabhängig wegen 8.10(b). Weiter gilt für jedes i wegen $\{X_i \neq \mathbf{1}_{A_i}\} \subseteq \{X_i \notin \{0, 1\}\}$ dann

$$\mathbb{P}(X_i \neq \mathbf{1}_{A_i}) \leq 1 - \mathbb{P}(X_i = 0) - \mathbb{P}(X_i = 1) = 1 - (1 - p) - p = 0,$$

also $X_i =_{\mathbb{P}} \mathbf{1}_{A_i}$, und damit folgt $X =_{\mathbb{P}} (\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n})$ nach 3.3(14) angewandt auf die Ereignisse $\{X_i = \mathbf{1}_{A_i}\}$.

(ii) \Rightarrow (i): (ii) gelte. Dann sind die $\mathbf{1}_{A_i}$ unabhängig nach Definition 8.1, und wegen $X \sim (\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n})$ nach 5.4(3) folgt die Unabhängigkeit der X_i aus Korollar 8.9 \square

Wir nennen $B_p^{\otimes n}$ eine **Bernoulli-Kettenverteilung** obwohl dies, nach Definition 9.1 und Satz 9.2, nicht eine Verteilung der Bernoulli-Kette X_\bullet , sondern die eines

zugehörigen Zufallsvektors X ist, und wir nennen wohl gelegentlich auch nachlässigerweise so ein X eine Bernoulli-Kette.

Definition 9.1 und Satz 9.2 betreffen Bernoulli-Ketten endlicher Länge. Gibt es auch “unendliche Bernoulli-Ketten”?

9.3 Bemerkung. Es sei $p \in]0, 1[$. Dann gibt es keine Familie $X_\bullet = (X_i : i \in \mathbb{N})$ unabhängiger und je B_p -verteilter Zufallsgrößen auf einem elementaren Wahrscheinlichkeitsraum (Ω, \mathbb{P}) , denn für so ein X_\bullet wäre mit $X := (X_i : i \in \mathbb{N})$ dann für $x \in \{0, 1\}^{\mathbb{N}}$ und $n \in \mathbb{N}$, beziehungsweise $n \rightarrow \infty$, unter Verwendung der Unabhängigkeit von (X_1, \dots, X_n) nach 8.5 im zweiten Schritt,

$$\begin{aligned} \mathbb{P}(X = x) &\leq \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\ &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \leq (\max\{p, 1-p\})^n \rightarrow 0, \end{aligned}$$

also $\mathbb{P}(X = x) = 0$, und damit im vierten Schritt, und mit 3.3(14) im ersten,

$$1 = \mathbb{P}\left(\bigcap_{i \in \mathbb{N}} \{X_i \in \{0, 1\}\}\right) = \mathbb{P}(X \in \{0, 1\}^{\mathbb{N}}) = \sum_{x \in \{0, 1\}^{\mathbb{N}}} \mathbb{P}(X = x) = 0.$$

In der MIT und der WT betrachtet man “nichtelementare Wahrscheinlichkeitsräume”; in geeigneten solchen existieren dann auch unendliche Bernoulli-Ketten zu gegebener Trefferwahrscheinlichkeit p , siehe etwa Billingsley (1995, Theorem 5.3 oder 20.4 oder 36.2). □

Zurück zu endlichen Bernoulli-Ketten:

9.4 Beispiel. Im Beispiel 7.11(b) des Ziehens mit Zurücklegen ist die Folge der Indikatoren von “rot in der j -ten Ziehung”, also $((X_j \in R) : j \in \{1, \dots, n\})$, eine Bernoulli-Kette mit den Parametern n und $\mathbb{P}(X_j \in R) = \frac{r}{r+b}$, denn die wegen 7.11(11) identisch verteilten Indikatoren $(X_j \in R)$ sind als Funktionen der nach Beispiel 8.8 unabhängigen X_i unabhängig wegen 8.10(b).

Dagegen ist im Beispiel 7.11(a) des Ziehens ohne Zurücklegen im Fall von $n \geq 2$ und $1 \leq r \leq N - 1$ die entsprechende Indikatorenfolge $((X_j \in R) : j \in \{1, \dots, n\})$ keine Bernoulli-Kette: Zwar sind wegen 7.11(6) auch hier die X_j und damit auch die Indikatoren $(X_j \in R)$ identisch verteilt, aber letztere sind nicht unabhängig, denn für die ersten beiden zugehörigen Ereignisse erhalten wir

$$\begin{aligned} \mathbb{P}(X_1 \in R, X_2 \in R) &= U_{M_{\neq}^2}(R \times R) = U_{M_{\neq}^2}(R_{\neq}^2) = \frac{r(r-1)}{N(N-1)} \\ &\neq \frac{r^2}{N^2} = \mathbb{P}(X_1 \in R)\mathbb{P}(X_2 \in R), \end{aligned}$$

im ersten Schritt wegen 7.9(10) mit $k = 2$ und 7.11(4) mit $n = 2$. □

In Verallgemeinerung von 7.11(12) gilt:

9.5 Satz. Es sei $(X_i : i \in \{1, \dots, n\})$ eine Bernoulli-Kette mit der Trefferwahrscheinlichkeit p . Dann gilt $\sum_{i=1}^n X_i \sim B_{n,p}$.

Beweis. Für $k \in \{0, \dots, n\}$ gilt, mit $X := (X_1, \dots, X_n)$,

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n X_i = k\right) &= \sum_{x \in \{0,1\}^n : \sum x_i = k} \mathbb{P}(X = x) \\ &= p^k (1-p)^{n-k} \#\left\{x \in \{0,1\}^n : \sum_{i=1}^n x_i = k\right\} \\ &= \binom{n}{k} p^k (1-p)^{n-k} = b_{n,p}(k) \end{aligned}$$

wegen 9.2(iv) im zweiten Schritt, 4.4(3) im dritten, und der Definition der $b_{n,p}$ in Beispiel 3.13 im vierten. □

9.6 Aufgabe (Warten auf den Lottogewinn).

(a) Es sei $n \in \mathbb{N}$, $p \in]0, 1[$, $\alpha \in [0, 1[$ und es gelte $X \sim B_{n,p}$. Geben Sie eine möglichst einfache Formel für $\mathbb{P}(X \geq 1)$ an und beweisen Sie mit deren Hilfe die Äquivalenz

$$\mathbb{P}(X \geq 1) \geq \alpha \iff n \geq \frac{\log(1-\alpha)}{\log(1-p)}.$$

(b) Sie wollen versuchen, durch beharrliches Lottospielen ans große Geld zu kommen. Dazu nehmen Sie sich vor, jede Woche eine Tippreihe für “6 aus 49” abzugeben. Wie viele Wochen müssen Sie durchhalten, um mindestens mit der Wahrscheinlichkeit 0.001 mindestens einmal “6 richtige” zu haben? □

9.7 Aufgabe (Wie viele Düsen sind sicherer?). Ein Düsenflugzeug mit $n = 2k \in 2\mathbb{N}$ Düsen kann sich in der Luft halten, wenn mindestens die Hälfte der Düsen funktioniert. Jede Düse falle unabhängig von den anderen mit der Wahrscheinlichkeit $p \in]0, 1[$ aus. Würden Sie lieber in ein Flugzeug mit 2 Düsen oder in eines mit 4 Düsen steigen wollen? □

Direkt aus der obigen natürlichen Herleitung der Binomialverteilungen in Satz 9.5 ergibt sich, anschaulich ohne weiteres und auch streng formal ohne Rechnung:

9.8 Satz. Es seien $S \sim B_{m,p}$ und $T \sim B_{n,p}$ unabhängig mit $m, n \in \mathbb{N}_0$ und $p \in [0, 1]$. Dann gilt $S + T \sim B_{m+n,p}$.

Beweis. 1. Sind auch S' und T' unabhängige Zufallsgrößen mit den angegebenen Binomialverteilungen, so gilt $(S, T) \sim (S', T')$ nach 8.9“ \Leftarrow ”, und damit auch $S + T \sim S' + T'$; also sind S und T mit den angegebenen Eigenschaften frei wählbar.

2. Gerechtfertigt durch die Bemerkung im Anschluss an Definition 9.1 wählen wir $(X_1, \dots, X_m, Y_1, \dots, Y_n) \sim B_p^{\otimes(m+n)}$. Dann gilt auch $(X_1, \dots, X_m) \sim B_p^{\otimes m}$ und $(Y_1, \dots, Y_n) \sim B_p^{\otimes n}$, es sind $S := \sum_{i=1}^m X_i$ und $T := \sum_{j=1}^n Y_j$ unabhängig wegen Satz 8.10, und eine dreimalige Anwendung von Satz 9.5 liefert $S \sim B_{m,p}$, $T \sim B_{n,p}$, und $S + T \sim B_{m+n,p}$. □

Satz 9.8 bestimmt in einem Spezialfall die Verteilung der Summe zweier unabhängiger reellwertiger Zufallsgrößen bei gegebenen Verteilungen der Summanden. Allgemeiner, etwa im mehrdimensionalen Fall oder mit Multiplikation statt Addition, nennt man so ein Resultat eine Faltung:

9.9 Definition. Es sei (\mathcal{X}, \cdot) eine Halbgruppe und es seien $P, Q \in \text{Prob}(\mathcal{X})$. Die Verteilung von $X \cdot Y$ für unabhängige Zufallsgrößen $X \sim P$ und $Y \sim Q$ heißt dann die **Faltung** (bezüglich der Verknüpfung \cdot) von P mit Q , in Zeichen $P * Q$. \square

9.10 Beispiel. Es gilt

$$(1) \quad B_{m,p} * B_{n,p} = B_{m+n,p} \quad \text{für } m, n \in \mathbb{N}_0 \text{ und } p \in [0, 1]$$

bezüglich $+$ auf \mathbb{R} , nach Satz 9.8. \square

9.11 Bemerkungen. Es sei (\mathcal{X}, \cdot) eine Halbgruppe.

(a) In 9.9 ist $P * Q$ tatsächlich wohldefiniert, wegen der Existenz der dortigen X, Y etwa als Koordinatengrößen im unabhängigen Kopplungsmodell $P \otimes Q$ auf \mathcal{X}^2 , und wegen einer Argumentation völlig analog zu Beweisschritt 1 zu Satz 9.8. Insbesondere gilt

$$(1) \quad P * Q = \left(\mathcal{X}^2 \ni (x, y) \mapsto x \cdot y \right) \square (P \otimes Q) \text{ für } P, Q \in \text{Prob}(\mathcal{X}).$$

(b) $(\text{Prob}(\mathcal{X}), *)$ ist eine Halbgruppe, denn $*$ ist offenbar eine Verknüpfung auf $\text{Prob}(\mathcal{X})$, und für $P, Q, R \in \text{Prob}(\mathcal{X})$ folgt mit unabhängigen $X \sim P, Y \sim Q, Z \sim R$ aus $(X \cdot Y) \cdot Z = X \cdot (Y \cdot Z)$ trivialerweise $((X \cdot Y) \cdot Z) \square \mathbb{P} = (X \cdot (Y \cdot Z)) \square \mathbb{P}$, und wegen $X \cdot Y \perp Z$ ist $((X \cdot Y) \cdot Z) \square \mathbb{P} = ((X \cdot Y) \square \mathbb{P}) * (Z \square \mathbb{P}) = (P * Q) * R$, und analog ist $(X \cdot (Y \cdot Z)) \square \mathbb{P} = P * (Q * R)$; also ist $*$ auch assoziativ. Ist \cdot kommutativ, so folgt ähnlich die Kommutativität von $*$. Ist e neutral für \cdot , so ist wegen $e \cdot X = X \cdot e = X$ für $X \sim P \in \text{Prob}(\mathcal{X})$ dann auch $(\text{Prob}(\mathcal{X}), *)$ ein Monoid, mit dem neutralen Element δ_e .

Man verwendet übliche Produkt- und Potenznotationen, $*_{i=1}^n P_i$ und P^{*n} für $P_1, \dots, P_n, P \in \text{Prob}(\mathcal{X})$ und $n \in \mathbb{N}$ im allgemeinen Fall, beziehungsweise $n \in \mathbb{N}_0$ im Monoidfall, und zeigt per Induktion

$$(2) \quad *_{i=1}^n P_i = \text{Verteilung von } \prod_{i=1}^n X_i \text{ für unabhängige } X_i \sim P_i,$$

$$(3) \quad P^{*n} = \text{Verteilung von } \prod_{i=1}^n X_i \text{ für unabh. u. identisch } P\text{-verteilte } X_i.$$

(c) Die bei weitem wichtigsten Spezialfälle sind $(\mathcal{X}, \cdot) = (\mathbb{R}, +)$, Unterhalbgruppen davon, und die entsprechenden mehrdimensionalen Varianten. Dann ist natürlich das Produktzeichen \prod in (2) und (3) durch das Summenzeichen \sum zu ersetzen, und wir erhalten beispielsweise in Verallgemeinerung von 9.10(1)

$$(4) \quad *_{i=1}^k B_{n_i,p} = B_{\sum_{i=1}^k n_i,p} \quad \text{für } k \in \mathbb{N}_0, n \in \mathbb{N}_0^k, p \in [0, 1]$$

als Verteilung von $\sum_{i=1}^n X_i$ für unabhängige $X_i \sim B_{n_i,p}$.

(d) Aufgabe 8.12(a) fragt nach einer Faltung über der endlichen Gruppe $(\mathbb{Z}_6, +)$.

(e) Faltungen über nichtkommutativen endlichen Gruppen kommen beispielsweise bei der stochastischen Analyse des Kartenmischens vor, siehe dazu etwa Diaconis (1988). \square

Zur Vorbereitung des nächsten Satzes erinnern wir an einen Begriff der elementaren Algebra:

9.12 Definition. Es sei (\mathcal{X}, \cdot) eine Halbgruppe, und es seien $A, B \subseteq \mathcal{X}$. Dann heißt

$$A \cdot B := \{xy : x \in A, y \in B\}$$

das **Komplexprodukt** von A mit B , und bei additiver Notation (also nur im kommutativen Fall) auch **Minkowski-Summe**, $A + B := \{x + y : x \in A, y \in B\}$. \square

9.13 Satz (Faltungsformeln für Dichten).

(a) **Abstrakter Fall.** Es sei (\mathcal{X}, \cdot) eine Halbgruppe und es seien $P, Q \in \text{Prob}(\mathcal{X})$ mit Trägern (im weiteren Sinne) $\mathcal{X}_0, \mathcal{Y}_0 \subseteq \mathcal{X}$. Dann ist das das Komplexprodukt $\mathcal{X}_0 \cdot \mathcal{Y}_0$ ein Träger von $P * Q$, genauer gilt

$$(1) \quad \text{Supp } P * Q = (\text{Supp } P) \cdot (\text{Supp } Q),$$

und für die zugehörigen Dichten gilt

$$(2) \quad f_{P*Q}(z) = \sum_{(x,y) \in \mathcal{X}^2 : xy=z} f_P(x)f_Q(y) = \sum_{(x,y) \in \mathcal{X}_0 \times \mathcal{Y}_0 : xy=z} f_P(x)f_Q(y) \quad \text{für } z \in \mathcal{X}.$$

Ist (\mathcal{X}, \cdot) eine Gruppe, so gilt weiter

$$(3) \quad \begin{aligned} f_{P*Q}(z) &= \sum_{y \in \mathcal{X}} f_P(zy^{-1})f_Q(y) = \sum_{x \in \mathcal{X}} f_P(x)f_Q(x^{-1}z) \\ &= \sum_{y \in \mathcal{Y}_0} f_P(zy^{-1})f_Q(y) = \sum_{x \in \mathcal{X}_0} f_P(x)f_Q(x^{-1}z) \quad \text{für } z \in \mathcal{X}. \end{aligned}$$

(b) **Spezialfall** $(\mathbb{N}_0, +)$. Es seien $P, Q \in \text{Prob}(\mathbb{N}_0)$. Dann gilt

$$(4) \quad f_{P*Q}(k) = \sum_{\ell=0}^k f_P(k-\ell)f_Q(\ell) = \sum_{\ell=0}^k f_P(\ell)f_Q(k-\ell) \quad \text{für } k \in \mathbb{N}_0,$$

und $\text{Supp } P * Q = (\text{Supp } P) + (\text{Supp } Q)$.

Beweis. (a) Es seien $X \sim P$ und $Y \sim Q$ unabhängige Zufallsgrößen auf einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Wegen

$$\{X \cdot Y \in \mathcal{X}_0 \cdot \mathcal{Y}_0\} \supseteq \{X \in \mathcal{X}_0\} \cap \{Y \in \mathcal{Y}_0\}$$

und $\mathbb{P}(X \in \mathcal{X}_0) = P(\mathcal{X}_0) = 1$ sowie $\mathbb{P}(Y \in \mathcal{Y}_0) = 1$ folgt

$$(P * Q)(\mathcal{X}_0 \cdot \mathcal{Y}_0) = \mathbb{P}(X \cdot Y \in \mathcal{X}_0 \cdot \mathcal{Y}_0) \geq \mathbb{P}(\{X \in \mathcal{X}_0\} \cap \{Y \in \mathcal{Y}_0\}) = 1$$

nach 3.3(14), auch ohne $X \perp Y$ zu verwenden; also ist $\mathcal{X}_0 \cdot \mathcal{Y}_0$ ein Träger von $P * Q$, und speziell für $\mathcal{X}_0 = \text{Supp } P$ und $\mathcal{Y}_0 = \text{Supp } Q$ ergibt sich “ \subseteq ” in (1).

Für $z \in \mathcal{X}$ gilt, wegen der totalen Additivität 3.3(10) im zweiten Schritt,

$$f_{P*Q}(z) = \mathbb{P}(X \cdot Y = z) = \sum_{(x,y) \in \mathcal{X}^2 : xy=z} \mathbb{P}(X = x, Y = y),$$

und mit $X \perp Y$ folgt die erste Darstellung in (2), wobei nur Summanden mit $x \in \mathcal{X}_0$ und $y \in \mathcal{Y}_0$ nicht Null sind, woraus die zweite Darstellung folgt.

Ist $z \in \text{R.S.}(1)$, also $z = x_0 y_0$ für geeignete $x_0 \in \text{Supp } P$ und $y_0 \in \text{Supp } Q$, so liefert (2) nun $f_{P*Q}(z) \geq f_P(x_0) f_Q(y_0) > 0$, also $z \in \text{L.S.}(1)$; folglich gilt auch “ \supseteq ” in (1).

Im Gruppenfall liefert (2) wegen der Äquivalenzkette $xy = z \Leftrightarrow x = zy^{-1} \Leftrightarrow y = x^{-1}z$ auch (3).

(b) Teil (a) angewandt auf die Gruppe $(\mathbb{Z}, +)$, mit $\mathcal{X}_0 := \mathcal{Y}_0 := \mathbb{N}_0$, unter Beachtung von $f_P(k - \ell) = 0 = f_Q(k - \ell)$ für $\ell > k$. □

9.14 Aufgabe. (a) Poisson-Verteilungen. $P_\lambda * P_\mu = P_{\lambda+\mu}$ für $\lambda, \mu \in [0, \infty[$.

(b) Nochmal Binomialverteilungen. Zeigen Sie auch 9.10(1) mittels 9.13(4).

Hinweis: Die dabei benötigte Vandermonde’sche Faltungsformel ist aus Aufgabe 4.8(a) bekannt, oder äquivalent dazu aus der Normiertheit der hypergeometrischen Zähldichten. □

Unter Verwendung von Aufgabe 9.14 löst man:

9.15 Aufgabe (Binomialverteilungen als bedingte Verteilungen von Poisson-Summanden). Es seien X und Y unabhängige \mathbb{N}_0 -wertige Zufallsgrößen auf dem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Bestimmen Sie die¹ faktorisierte bedingte Verteilung $\mathbb{P}(X \in \cdot \mid X + Y = \cdot)$ im Fall von $X \sim P_\lambda, Y \sim P_\mu$ mit $\lambda, \mu \in]0, \infty[$. □

Die Ergebnisse der Aufgaben 9.15 und 8.3 werden in Lemma 12.16 benutzt.

Es gibt nur wenige wichtige Beispiele bei denen, wie in Aufgabe 9.14 oder unten im Fall von Negativbinomialverteilungen, eine Summe wie in 9.13(4) explizit ausgerechnet werden kann. Und die iterierte Anwendung von 9.13(4) auf ein Faltungsprodukt der Form $\ast_{i=1}^n P_i$ oder auch nur P^{*n} führt auf ein Ergebnis mit $n - 1$ Summenzeichen. Daher ist das folgende Ergebnis, mit nur einem Summenzeichen, bemerkenswert:

9.16 Beispiel (Faltungen identischer uniformer Verteilungen). Für $\ell, n \in \mathbb{N}$ ist die Dichte von $U_{\{0, \dots, \ell-1\}}^{*n}$ durch 5.17(1) gegeben, denn dort sind X_1, \dots, X_n unabhängig und je $U_{\{0, \dots, \ell-1\}}$ -verteilt. □

¹Salopper Sprachgebrauch wie in Bemerkung 7.16(c) erklärt: Eigentlich wäre ja eine Menge von Kernen zu bestimmen.

9.17 Aufgabe (Faltungen verschiedener uniformer Verteilungen). Bestimmen Sie für $m, n \in \mathbb{N}$ die Faltung $U_{\{0, \dots, m-1\}} * U_{\{0, \dots, n-1\}}$ und vergleichen Sie das Ergebnis mit der in Beispiel 5.6(a) berechneten Verteilung der Augensumme beim zweifachen Würfelwurf. \square

9.18 Aufgabe (Falten verringert die Konzentration). Es seien P, Q Wahrscheinlichkeitsmaße auf \mathbb{R} . Zeigen Sie, dass der nach Bemerkung 3.22 existierende Maximalwert der Dichte f_{P*Q} der Faltung $P * Q$ höchstens so groß wie die Maximalwerte der Dichten f_P, f_Q der Faktoren der Faltung ist:

$$\max_{x \in \mathbb{R}} f_{P*Q}(x) \leq \min \left\{ \max_{x \in \mathbb{R}} f_P(x), \max_{x \in \mathbb{R}} f_Q(x) \right\}$$

Gibt es nichttriviale Beispiele, in denen hier Gleichheit gilt? \square

Zur Vorbereitung von 9.20–9.22 notieren wir:

9.19 Erinnerung (Binomialreihe). Es gilt

$$(1) \quad (1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{für } \alpha, x \in \mathbb{R} \text{ mit } |x| < 1,$$

siehe etwa Heuser (2009, Nr. 62, dort in Nr. 65 auch Diskussion des Konvergenz- und Darstellungsverhaltens in den Randpunkten $x \in \{-1, 1\}$) oder Storch/Wiebe (2003, p. 346, dort allgemeiner mit \mathbb{C} statt \mathbb{R}), und für $\alpha, \beta \in \mathbb{R}$ und $|x| < 1$ folgt, unter Verwendung des Cauchy-Produktes von Potenzreihen im letzten Schritt,

$$\begin{aligned} \sum_{k=0}^{\infty} \binom{\alpha + \beta}{k} x^k &= (1+x)^{\alpha+\beta} = (1+x)^\alpha (1+x)^\beta \\ &= \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \sum_{k=0}^{\infty} \binom{\beta}{k} x^k = \sum_{k=0}^{\infty} \sum_{\ell=0}^k \binom{\alpha}{\ell} \binom{\beta}{k-\ell} x^k, \end{aligned}$$

und nach dem Eindeutigkeitsatz für Potenzreihen ergibt sich damit die schon in Aufgabe 4.8(c) auf anderem Wege hergeleitete allgemeine Vandermonde’sche Faltungsformel

$$(2) \quad \sum_{\ell=0}^k \binom{\alpha}{\ell} \binom{\beta}{k-\ell} = \binom{\alpha + \beta}{k} \quad \text{für } \alpha, \beta \in \mathbb{R} \text{ und } k \in \mathbb{N}_0$$

(und analog sogar für $\alpha, \beta \in \mathbb{C}$, was im folgenden aber nicht benötigt wird). \square

9.20 Beispiel (Zensierte Wartezeiten).

(a) Es seien $n, r \in \mathbb{N}$, $p \in [0, 1]$, und (X_1, \dots, X_{n+r}) eine Bernoulli-Kette mit der Trefferwahrscheinlichkeit p , mit der Einfachheit halber $\{0, 1\}$ -wertigen X_i . Es bezeichne S_k die Anzahl der Treffer bis zum Zeitpunkt k , und $W_r^{(n)}$ die durch n beschränkte Anzahl der Fehlversuche vor dem r -ten Treffer, also

$$\begin{aligned} S_k &:= \sum_{i=1}^k X_i \quad \text{für } k \in \{0, \dots, n+r\}, \\ W_r^{(n)} &:= \left(\inf \{k \in \{1, \dots, n+r\} : S_k \geq r\} - r \right) \wedge n \end{aligned}$$

mit, wie üblich, $\inf \emptyset := \infty$ und $a \wedge b := \min\{a, b\}$. Die Verteilung von $W_r^{(n)}$ nennen wir **in n rechtszensierte Negativbinomialverteilung** mit den Parametern r, p (und n), bezeichnen sie mit $\text{NB}_{r,p,n}$, und ihrer Dichte mit $\text{nb}_{r,p,n}$.

Offenbar ist $W_r^{(n)}$ $\{0, \dots, n\}$ -wertig, also gilt

$$(1) \quad \text{NB}_{r,p,n}(\{0, \dots, n\}) = 1.$$

Für $\ell \in \{0, \dots, n-1\}$ gilt

$$\{W_r^{(n)} = \ell\} = \{S_{\ell+r} \geq r, S_{\ell+r-1} < r\} = \{X_{\ell+r} = 1, S_{\ell+r-1} = r-1\}$$

und damit, wegen $X_{\ell+r} \perp S_{\ell+r-1}$ im zweiten Schritt sowie $X_{\ell+r} \sim B_p$ und $S_{\ell+r-1} \sim B_{\ell+r-1,p}$ im dritten,

$$(2) \quad \begin{aligned} \text{nb}_{r,p,n}(\ell) &= \mathbb{P}(W_r^{(n)} = \ell) = \mathbb{P}(X_{\ell+r} = 1)\mathbb{P}(S_{\ell+r-1} = r-1) \\ &= \binom{\ell+r-1}{r-1} p^r (1-p)^\ell; \end{aligned}$$

wegen (1) wird damit also die Angabe der Dichte von $\text{NB}_{r,p,n}$ vervollständigt durch

$$\text{nb}_{r,p,n}(n) = 1 - \sum_{\ell=0}^{n-1} \binom{\ell+r-1}{r-1} p^r (1-p)^\ell.$$

In den uninteressanten Randfällen $p \in \{0, 1\}$ ergibt sich erwartungsgemäß $\text{NB}_{r,0,n} = \delta_n$ beziehungsweise $\text{NB}_{r,1,n} = \delta_0$.

(b) Nach Teil (a) gilt trivialerweise

$$(3) \quad \lim_{n \rightarrow \infty} \text{nb}_{r,p,n}(\ell) = \binom{\ell+r-1}{r-1} p^r (1-p)^\ell \quad \text{für } r \in \mathbb{N}, p \in [0, 1], \ell \in \mathbb{N}_0,$$

und wir fragen uns, ob die rechte Seite von (3) als Funktion von ℓ eine Wahrscheinlichkeitsdichte auf \mathbb{N}_0 ist. Für $p = 0$ ist das wegen $p^r = 0$ offenbar nicht der Fall, für die restlichen p aber doch:

Es sei also $p \in]0, 1]$. Weiter sei jetzt allgemeiner $r \in [0, \infty[$. Dann definiert

$$(4) \quad \text{nb}_{r,p}(\ell) := \binom{\ell+r-1}{\ell} p^r (1-p)^\ell \quad \text{für } \ell \in \mathbb{N}_0$$

eine Verallgemeinerung der rechten Seite von (3), denn im Fall $r \in \mathbb{N}$ gilt $\binom{\ell+r-1}{\ell} = \binom{\ell+r-1}{r-1}$ für $\ell \in \mathbb{N}_0$. Nach Aufgabe 3.15 ist $\text{nb}_{r,p}$ eine Wahrscheinlichkeitsdichte auf \mathbb{N}_0 . □

Dies motiviert im Fall von $r \in \mathbb{N}$ die folgende Definition, für deren Nützlichkeit auch im allgemeinen Fall wir hier nur auf die Literaturangaben in Johnson et al. (2005, pp. 232–233) verweisen.

9.21 Definition. Sind $r \in [0, \infty[$ und $p \in]0, 1]$, so heißt die Verteilung mit der Zähldichte $\text{nb}_{r,p}$ aus 9.20(4) **Negativbinomialverteilung** mit den Parametern r und p , in Zeichen $\text{NB}_{r,p}$. Speziell $G_p := \text{NB}_{1,p}$ für $p \in]0, 1]$ heißt **geometrische Verteilung** zur Trefferwahrscheinlichkeit p . \square

Im Fall von $r \in \mathbb{N}$ mag man nun $\text{NB}_{r,p}$ als Verteilung der Anzahl der Fehlversuche vor dem r -ten Treffer in einer (nach Bemerkung 9.3 im Rahmen der elementaren Wahrscheinlichkeitsräume gar nicht existierenden) unendlichen Bernoulli-Kette interpretieren. Diese Interpretation suggeriert im Fall ganzer r, s die folgenden Faltingsformeln:

9.22 Aufgabe. Für $p \in]0, 1]$ gilt

$$\begin{aligned} (1) \quad & \text{NB}_{r,p} * \text{NB}_{s,p} = \text{NB}_{r+s,p} \quad \text{für } r, s \in [0, \infty[, \\ (2) \quad & \text{NB}_{r,p} = G_p^{*r} \quad \text{für } r \in \mathbb{N}_0. \end{aligned}$$

\square

Zur Vorbereitung der Einführung von Multinomialverteilungen dient

9.23 Definition. Es sei I eine endliche Menge und $\alpha \in \mathbb{N}_0^I$. Dann heißt α ein **Multiindex**, und man setzt

$$\begin{aligned} x^\alpha &:= \prod_{i \in I} x_i^{\alpha_i} \quad \text{für } x \in \mathbb{C}^I, & \alpha! &:= \prod_{i \in I} \alpha_i!, & |\alpha| &:= \sum_{i \in I} \alpha_i, \\ \binom{n}{\alpha} &:= (|\alpha| = n) \frac{n!}{\alpha!} \quad \text{Multinomialkoeffizient} & & \text{für } n \in \mathbb{N}_0. \end{aligned}$$

\square

Diese Notation wird oft mit $I = \{1, \dots, d\}$ für ein $d \in \mathbb{N}$ angewandt. Zum Beispiel ist für $n \in \mathbb{N}_0$ und $k \in \{0, \dots, n\}$ mit $I = \{1, 2\}$ dann

$$(1) \quad \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(k, n-k)!} = \binom{n}{(k, n-k)},$$

aber, mit $I = \{1\}$, dagegen $\binom{n}{(k)} = (k = n)$. Deswegen verwendet man nur im Fall von $d \geq 2$ auch die ein Klammernpaar weglassende Schreibweise

$$\binom{n}{\alpha_1, \dots, \alpha_d} := \binom{n}{\alpha} \quad \text{für } n \in \mathbb{N}_0, \alpha \in \mathbb{N}_0^d.$$

Mit Definition 9.23 schreibt sich die **multinomische Formel** konzise als

$$(2) \quad \left(\sum_{i \in I} x_i \right)^n = \sum_{k \in \mathbb{N}_0^I} \binom{n}{k} x^k \quad \text{für } I \text{ endliche Menge, } x \in \mathbb{C}^I, n \in \mathbb{N}_0.$$

Speziell gilt also, die Notation $\text{prob}(\dots)$ aus Definition 7.1(a) verwendend,

$$(3) \quad \sum_{k \in \mathbb{N}_0^I} \binom{n}{k} p^k = 1 \quad \text{für } I \text{ endliche Menge, } n \in \mathbb{N}_0, p \in \text{prob}(I).$$

9.24 Definition. Ist I eine endliche Menge, $p \in \text{prob}(I)$, und $n \in \mathbb{N}_0$, so heißt die Verteilung mit der Dichte

$$\mathbb{N}_0^I \ni k \mapsto \binom{n}{k} p^k =: m_{n,p}(k)$$

Multinomialverteilung mit den Parametern n und p , in Zeichen $M_{n,p}$. □

Da der obige Multinomialkoeffizient genau für $|k| = n$ nicht verschwindet, ist $\{k \in \mathbb{N}_0^I : |k| = n\}$ ein Träger von $M_{n,p}$, und zwar der kleinste falls $p_i > 0$ für jedes $i \in I$ gilt; insbesondere gilt also die Implikation

$$(1) \quad X \sim M_{n,p} \Rightarrow \sum_{i \in I} X_i =_{\mathbb{P}} n.$$

Im Fall von $I = \{1, \dots, d\}$ mit $d \in \mathbb{N}$ schreibt man meist M_{n,p_1, \dots, p_d} statt $M_{n,(p_1, \dots, p_d)}$ und dann auch m_{n,p_1, \dots, p_d} statt $m_{n,(p_1, \dots, p_d)}$.

9.25 Bemerkung. Es sei $n \in \mathbb{N}_0$ und $p \in [0, 1]$. Mit 9.23(1) ergibt sich

$$(1) \quad b_{n,p}(k) = m_{n,p,1-p}(k, n-k) \quad \text{für } k \in \{0, \dots, n\}.$$

Also gilt für jede \mathbb{N}_0 -wertige (oder auch \mathbb{R} -wertige) Zufallsgröße X die Äquivalenz

$$(2) \quad X \sim B_{n,p} \Leftrightarrow (X, n-X) \sim M_{n,p,1-p}$$

wegen $\{X = k\} = \{X = k, n-X = n-k\}$ und (1). □

Die natürliche Erzeugung der Binomialverteilungen aus Satz 9.5 überträgt sich auf Multinomialverteilungen. Wir begnügen uns hier mit einer knappen Formulierung, und wir fassen uns auch in den Beweisen der folgenden Sätze etwas kürzer als sonst.

9.26 Satz. *Es sei I eine endliche Menge, $p \in \text{prob}(I)$, $n \in \mathbb{N}_0$, und (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum. Für $j \in \{1, \dots, n\}$ sei $(A_{ij} : i \in I)$ eine Quasipartition von Ω mit*

$$\begin{aligned} \mathbb{P}(A_{ij}) &= p_i \quad \text{für } i \in I \text{ und } j \in \{1, \dots, n\}, \\ \left((\mathbf{1}_{A_{ij}} : i \in I) : j \in \{1, \dots, n\} \right) &\text{ unabhängig.} \end{aligned}$$

Dann gilt $(\sum_{j=1}^n \mathbf{1}_{A_{ij}} : i \in I) \sim M_{n,p}$.

Beweis. Analog zu Satz 9.5, siehe etwa Henze (2017, pp. 146–148). □

9.27 Beispiel (n -facher Würfelwurf). Es sei $I = \{1, \dots, 6\}$, $p = (\frac{1}{6}, \dots, \frac{1}{6})$, $\Omega := \{1, \dots, 6\}^n$ mit einem $n \in \mathbb{N}_0$, $\mathbb{P} = U_{\{1, \dots, 6\}}^{\otimes n} = U_{\{1, \dots, 6\}^n}$, sowie $X_j : \Omega \rightarrow \{1, \dots, 6\}$ die j -te Koordinatengröße und $A_{ij} := \{X_j = i\}$ für $i \in I$ und $j \in \{1, \dots, n\}$.

Dann sind die Voraussetzungen von Satz 9.26 erfüllt, und es ergibt sich zum Beispiel mit $n = 60$ speziell

$$\begin{aligned} \mathbb{P}(\text{“10 Einsen, } \dots, \text{ 10 Sechsen”}) &= \binom{60}{10, \dots, 10} \left(\left(\frac{1}{6} \right)^{10} \right)^6 = \frac{60!}{(10!)^6 6^{60}} \\ &= 0.000074\dots \end{aligned}$$

□

9.28 Satz. *Es sei $X = (X_i : i \in I) \sim M_{n,p}$ mit $n \in \mathbb{N}_0$, $p \in \text{prob}(I)$, I endlich.*

(a) **Randverteilungen.** $X_i \sim B_{n,p_i}$ für $i \in I$.

(b) **Klassenbildung.** *Es sei $(I_j : j \in J)$ eine endliche Quasipartition von I . Dann gilt*

$$\left(\sum_{i \in I_j} X_i : j \in J \right) \sim M_{n, \left(\sum_{i \in I_j} p_i : j \in J \right)}.$$

Beweis. (b) Ohne Einschränkung sei $X = (\sum_{k=1}^n \mathbf{1}_{A_{ik}} : i \in I)$ wie in Satz 9.26, mit $(A_{ik} : i \in I)$ Quasipartition für $k \in \{1, \dots, n\}$. Für $j \in J$ gilt dann, mit 5.10(5) im zweiten Schritt,

$$\sum_{i \in I_j} X_i = \sum_{k=1}^n \sum_{i \in I_j} \mathbf{1}_{A_{ik}} = \sum_{k=1}^n \mathbf{1}_{B_{jk}}$$

mit $B_{jk} := \cup_{i \in I_j} A_{ik}$ und $\mathbb{P}(B_{jk}) = \sum_{i \in I_j} p_i$; dabei ist $(B_{jk} : j \in J)$ eine Quasipartition von Ω für $k \in \{1, \dots, n\}$, und $((\mathbf{1}_{B_{jk}} : j \in J) : k \in \{1, \dots, n\})$ ist unabhängig wegen Satz 8.10(b). Also folgt die Behauptung mit Satz 9.26.

(a) Teil (b) angewandt auf $I_1 := \{i\}$, $I_2 := I \setminus \{i\}$, $J := \{1, 2\}$ liefert

$$\left(X_i, \sum_{i' \in I \setminus \{i\}} X_{i'} \right) \sim M_{n, p_i, 1-p_i}$$

und mit $\sum_{i' \in I \setminus \{i\}} X_{i'} =_{\mathbb{P}} n - X_i$ nach 9.24(1) folgt die Behauptung mit 9.25(2). □

9.29 Aufgabe (Mendelsche Vererbung). Aus zwei Elternpopulationen E_1 und E_2 einer diploiden Organismenart möge bei n rein zufälligen Paarungen zwischen E_1 und E_2 (der Einfachheit halber “mit Zurücklegen”; bei bisexueller Fortpflanzung soll etwa E_1 nur aus weiblichen und E_2 nur aus männlichen Individuen bestehen) jeweils ein Nachkomme entstehen. Für ein bestimmtes Gen mit den zwei Allelen A und a seien $p_{iAA}, p_{iAa}, p_{iaa}$ die Anteile der Individuen mit den Genotypen AA, Aa, aa in der Population E_i . Es bezeichne X_{AA}, X_{Aa}, X_{aa} die Anzahlen der Nachkommen mit den jeweils angezeigten Genotypen.

(a) Wie ist (bei geeigneter Präzisierung der Annahmen) (X_{AA}, X_{Aa}, X_{aa}) verteilt?

(b) Es sei A dominant über a . Wie ist dann die Anzahl der Nachkommen vom Phänotyp A verteilt? □

9.30 Aufgabe (Summen binomialverteilter Zufallsgrößen).

(a) Aus Satz 9.28 ergeben sich neben Satz 9.8 weitere Beispiele in denen eine Summe zweier binomialverteilter Zufallsgrößen wieder binomialverteilt ist. Haben diese neuen Beispiele etwas mit Faltungen zu tun?

(b) Finden Sie Beispiele von Zufallsgrößen $X \sim B_{n,p}$ und $Y \sim B_{n,p}$, auf einem selben Wahrscheinlichkeitsraum (Ω, \mathbb{P}) , mit $X + Y$ nicht binomialverteilt. \square

Multinomialverteilungen liefern offensichtliche Beispiele von Faltungsformeln über den Monoiden $(\mathbb{N}_0^I, +)$:

9.31 Satz. *Es seien I eine endliche Menge, $p \in \text{prob}(I)$, und $m, n \in \mathbb{N}_0$. Dann gilt $M_{m,p} * M_{n,p} = M_{m+n,p}$.*

Beweis. Analog zu Satz 9.8, unter Benutzung von Satz 9.26. \square

9.32 Aufgabe. Was ist die faktorisierte Verbundverteilung unabhängiger Poissonverteilter Zufallsgrößen gegeben ihre Summe? \square

9.33 Aufgabe (Bedingte Verteilungen geometrisch verteilter Summanden).

(a) Was ist $\mathbb{P}(X \in \cdot \mid X + Y = s)$ für unabhängige $X, Y \sim G_p$? Überrascht?

(b) Berechnen Sie allgemeiner $\mathbb{P}((X_1, \dots, X_n) \in \cdot \mid \sum_{j=1}^{n+1} X_j = s)$ für unabhängige $X_1, \dots, X_{n+1} \sim G_p$, und beweisen Sie dadurch erneut Formel 4.9(2). \square

10 Verteilungsfunktionen, Quantile, Erwartungswerte, Gesetz großer Zahlen

Wie können Verteilungen auf \mathbb{R} bequem beschrieben werden? Vorweg ein Nachtrag zu den Rechenregeln 3.3:

10.1 Satz. *Es sei (Ω, P) ein elementarer Wahrscheinlichkeitsraum, und es sei $\mathcal{A} \subseteq 2^\Omega$ total geordnet bezüglich Mengeninklusion, d.h. es gelte die Implikation*

$$(1) \quad A, B \in \mathcal{A} \Rightarrow A \subseteq B \text{ oder } B \subseteq A.$$

Dann gilt

$$(2) \quad P(\bigcup \mathcal{A}) = \sup_{A \in \mathcal{A}} P(A) \quad \text{und} \quad P(\bigcap \mathcal{A}) = \inf_{A \in \mathcal{A}} P(A).$$

Beweis. Für $A \in \mathcal{A}$ gilt $\bigcup \mathcal{A} \supseteq A$ und damit $P(\bigcup \mathcal{A}) \geq P(A)$; also gilt $P(\bigcup \mathcal{A}) \geq \sup_{A \in \mathcal{A}} P(A)$. Umgekehrt ist

$$\begin{aligned} P(\bigcup \mathcal{A}) &= \sum_{\omega \in \bigcup \mathcal{A}} P(\{\omega\}) = \sup \left\{ \sum_{\omega \in B} P(\{\omega\}) : B \subseteq \bigcup \mathcal{A}, B \text{ endlich} \right\} \\ &= \sup \left\{ P(B) : B \subseteq \bigcup \mathcal{A}, B \text{ endlich} \right\}, \end{aligned}$$

und zu jeder endlichen Menge $B \subseteq \bigcup \mathcal{A}$ existiert zu jedem $\omega \in B$ eine Menge $A_\omega \in \mathcal{A}$ mit $\omega \in A_\omega$, also $B \subseteq \bigcup_{\omega \in B} A_\omega \in \mathcal{A}$ wegen (1) im letzten Schritt; also $P(\bigcup \mathcal{A}) \leq \sup \{P(A) : A \in \mathcal{A}\}$.

Die zweite Behauptung in (2) folgt aus der ersten mittels Komplementierung und de Morgan. \square

10.2 Aufgabe. In Satz 10.1 kann die Voraussetzung 10.1(1)

- (a) nicht ersatzlos gestrichen werden,
- (b) abgeschwächt werden zu: \mathcal{A} sei durch \subseteq aufwärts **gerichtet**, d.h. es gelte

$$(1) \quad A, B \in \mathcal{A} \Rightarrow \exists C \in \mathcal{A} \text{ mit } A \subseteq C \text{ und } B \subseteq C$$

(für die erste Hälfte von 10.1(2)) und abwärts gerichtet (analog definiert, für die zweite Hälfte von 10.1(2)). \square

10.3 Definition. Für $P \in \text{Prob}(\mathbb{R})$ heißt die Funktion

$$\mathbb{R} \ni x \mapsto P([-\infty, x]) =: F_P(x)$$

die zugehörige **Verteilungsfunktion**. \square

Ist X eine reellwertige Zufallsgröße mit der Verteilung P , so schreiben wir statt F_P auch F_X .

10.4 Beispiel. Für $\ell \in \mathbb{N}$ ist

$$(1) \quad F_{U_{\{0, \dots, \ell-1\}}}(x) = \frac{\#\left(]-\infty, x] \cap \{0, \dots, \ell-1\}\right)}{\ell} = 0 \vee \frac{\lfloor x \rfloor + 1}{\ell} \wedge 1 \quad \text{für } x \in \mathbb{R}.$$

□

10.5 Aufgabe. Allgemeiner als in Beispiel 10.4 gilt nach Aufgabe 5.17 für $n \in \mathbb{N}$ und unabhängige, je $U_{\{0, \dots, \ell-1\}}$ -verteilte X_1, \dots, X_n

$$(1) \quad F_{\sum_{i=1}^n X_i}(x) = \text{R.S. 5.17(2) mit } k := \lfloor x \rfloor, \quad \text{für } x \in \mathbb{R}.$$

Passt (1) zu 10.4(1)?

□

10.6 Aufgabe. Geben Sie für die Verteilungen aus drei der sieben Zeilen der Tabelle in 3.12 die zugehörigen Verteilungsfunktionen in einfacher Form (ohne Summenzeichen, ohne Pünktchen, ohne Integralzeichen) an.

□

10.7 Satz. Es sei $P \in \text{Prob}(\mathbb{R})$ und $F := F_P$.

(a) Es ist $F : \mathbb{R} \rightarrow [0, 1]$ wachsend und rechtsstetig mit

$$(1) \quad \lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

(b) Mit $F(x-) := \lim_{y \uparrow x} F(y)$ für $x \in \mathbb{R}$ ist $F(\cdot-)$ wachsend und linksstetig, mit

$$(2) \quad P(]-\infty, x]) = F(x-),$$

$$(3) \quad P(]x, \infty]) = 1 - F(x),$$

$$(4) \quad P([x, \infty]) = 1 - F(x-),$$

$$(5) \quad P(\{x\}) = F(x) - F(x-)$$

für $x \in \mathbb{R}$ und, für $a, b \in \mathbb{R}$ mit $a \leq b$,

$$(6) \quad P(]a, b]) = F(b) - F(a),$$

$$(7) \quad P([a, b]) = F(b) - F(a-),$$

$$(8) \quad P(]a, b]) = F(b-) - F(a),$$

$$(9) \quad P([a, b]) = F(b-) - F(a-).$$

Die Funktion $\mathbb{R} \ni x \mapsto \text{L.S.}(4)$ heißt auch (**linksstetige**) **komplementäre Verteilungsfunktion** von P .

Beweis. (a) Die Isotonie von F ergibt sich direkt aus der Isotonie 3.3(8) von P . Damit ist die Rechtsstetigkeit von F äquivalent zu

$$(10) \quad F(x) = \inf_{y > x} F(y) \quad \text{für } x \in \mathbb{R},$$

und (10) gilt wegen 10.1(1) mit $\mathcal{A} := \{]-\infty, y] : y > x\}$. Analog folgt (1) wegen

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &= \inf_{y \in \mathbb{R}} F(y) = P\left(\bigcap_{y \in \mathbb{R}}]-\infty, y]\right) = P(\emptyset) = 0, \\ \lim_{x \rightarrow \infty} F(x) &= \sup_{y \in \mathbb{R}} F(y) = P\left(\bigcup_{y \in \mathbb{R}}]-\infty, y]\right) = P(\mathbb{R}) = 1. \end{aligned}$$

(b) (2) gilt wegen 10.1(1) mit $\mathcal{A} := \{]-\infty, y] : y < x\}$, und impliziert die Linksstetigkeit von $F(\cdot-)$. (3) folgt direkt aus der Definition von F mit der Komplementarität 3.3(4). (4) folgt analog aus (2). Die restlichen Aussagen (5–9) folgen mit der Subtraktivität 3.3(9). \square

10.8 Korollar. Die Funktion $\text{Prob}(\mathbb{R}) \ni P \mapsto F_P$ ist injektiv.

Beweis. Klar nach 10.7(5) und 3.6(4). \square

10.9 Aufgabe. Geben Sie ein elementares Wahrscheinlichkeitsmaß $P \in \text{Prob}(\mathbb{R})$ mit einer streng wachsenden Verteilungsfunktion an. \square

10.10 Satz. Es sei $P \in \text{Prob}(\mathbb{R})$, $F := F_P$, $u \in]0, 1[$,

- (1) $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$,
 (2) $F^{-1}(u) := \sup\{x \in \mathbb{R} : F(x-) \leq u\}$.

Dann gilt $-\infty < F^{-1}(u) \leq F^{-1}(u) < \infty$, und für $q \in \mathbb{R}$ die Äquivalenzkette

$$\begin{aligned} P(]-\infty, q]) &\leq u \quad \text{und} \quad P(]q, \infty[) \leq 1 - u \\ \Leftrightarrow P(]-\infty, q]) &\geq u \quad \text{und} \quad P([q, \infty[) \geq 1 - u \\ \Leftrightarrow F(q-) &\leq u \leq F(q) \\ \Leftrightarrow q &\in [F^{-1}(u), F^{-1}(u)]. \end{aligned}$$

Beweis. Aus 10.7(a) folgt $F^{-1}(u) > -\infty$ und

(3) $\{x \in \mathbb{R} : F(x) \geq u\} = [F^{-1}(u), \infty[$.

Mit der Linksstetigkeit von $F(\cdot-)$ und mit 10.7(1) für $F(\cdot-)$ statt F folgt analog

(4) $\{x \in \mathbb{R} : F(x-) \leq u\} =]-\infty, F^{-1}(u)]$.

Für $x < F^{-1}(u)$ gilt $F(x) < u$, nach Definition des Begriffs “untere Schranke”, also gilt $F(F^{-1}(u)-) \leq u$, und mit (4) folglich $F^{-1}(u) \leq F^{-1}(u)$.

Es sei nun $q \in \mathbb{R}$. Dann gilt die Äquivalenzkette

$$P(]-\infty, q]) \leq u \Leftrightarrow P([q, \infty[) \geq 1 - u \Leftrightarrow F(q-) \leq u \Leftrightarrow q \leq F^{-1}(u)$$

wegen der Komplementarität 3.3(4) im ersten Schritt, 10.7(2) im zweiten, und (4) im dritten, und analog

$$P(]q, \infty[) \leq 1 - u \Leftrightarrow P(]-\infty, q]) \geq u \Leftrightarrow F(q) \geq u \Leftrightarrow F^{-1}(u) \leq q. \quad \square$$

10.11 Definition. Es sei $P \in \text{Prob}(\mathbb{R})$ und $F := F_P$. Die durch 10.10(1) definierte Funktion¹ $F^{-1} :]0, 1[\rightarrow \mathbb{R}$ heißt **(untere) Quantilfunktion** von P , die durch 10.10(2) definierte Funktion $F^{-1} :]0, 1[\rightarrow \mathbb{R}$ heißt **obere Quantilfunktion** von P . Für $u \in]0, 1[$ heißt jedes $q \in [F^{-1}(u), F^{-1}(u)]$ ein **u -Quantil** von P , und im Spezialfall $u = \frac{1}{2}$ ein **Median**. \square

¹Wir vermeiden die leider allgemein übliche Bezeichnung F^{-1} , da F^{-1} hier nie, und daher in der allgemeineren WT auch nur manchmal, eine Umkehrfunktion von F ist.

Natürlich sind Quantilfunktionen, Quantile, und Mediane von reellwertigen Zufallsgrößen wieder über deren Verteilungen definiert.

10.12 Aufgabe (Symmetrien bei Binomialverteilungen).

(a) Für $X \sim B_{n,p}$ gilt $n - X \sim ?$

(b) Bestimmen Sie für $n \in \mathbb{N}$ die Menge aller Mediane von $B_{n, \frac{1}{2}}$. □

Verteilungs- und Quantilfunktionen sind die natürlichsten Abbilder einer Verteilung auf \mathbb{R} , aber oft nicht leicht zu berechnen. Dagegen charakterisieren die nun einzuführenden Erwartungswerte und Varianzen, sowie verwandte Funktionale, eine Verteilung bei weiten nicht vollständig, sind aber einerseits rechnerisch sehr gut handhabbar, und andererseits über Sätze wie die Chebyshev-Ungleichung 10.31 und den Zentralen Grenzwertsatz 12.18 in vielen Situationen erstaunlich informativ.

Wir setzen ab jetzt Anhang B als bekannt voraus.

10.13 Definition. Es sei $P \in \text{Prob}(\mathbb{R})$ mit der Dichte f . Ist $(xf(x) : x \in \mathbb{R})$ halbsummierbar, so heißt

$$(1) \quad \mu(P) := \sum_{x \in \mathbb{R}} xf(x) \in \overline{\mathbb{R}}$$

(möglicherweise unendlicher) **Erwartungswert** von P . □

Dabei gilt $\sum_{x \in \mathbb{R}} |x|f(x) < \infty$ genau dann, wenn $\mu(P)$ existiert und endlich ist, wegen Bemerkung B.12 angewandt auf die Familie $(xf(x) : x \in \mathbb{R})$.

Deutet man P mechanisch als eine Massenverteilung der Gesamtmasse 1 auf \mathbb{R} , so ist $\mu(P)$ ihr Schwerpunkt. Ein stochastische Deutung von $\mu(P)$ liefert jedoch zuerst Satz 10.31 weiter unten.

Die im folgenden aufzustellenden Rechenregeln für Erwartungswerte formulieren sich im Zweifel bequemer für Zufallsgrößen als für Verteilungen, daher:

10.14 Definition. Es sei X eine \mathbb{R} -wertige Zufallsgröße auf dem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) . Ist $(X(\omega)\mathbb{P}(\{\omega\}) : \omega \in \Omega)$ halbsummierbar, so heißt

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

(möglicherweise unendlicher) **Erwartungswert** von X . □

Damit gelten für jede \mathbb{R} -wertige Zufallsgröße X die Implikationen

$$(1) \quad X \geq_{\mathbb{P}} 0 \Rightarrow \mathbb{E}X \in [0, \infty] \text{ existiert,}$$

$$(2) \quad \mathbb{E}|X| < \infty \Leftrightarrow \mathbb{E}X \in \mathbb{R} \text{ existiert;}$$

dabei ergibt sich (1) direkt aus Definition B.11, und (2) aus Bemerkung B.12, jeweils angewandt auf die Familie $(X(\omega)\mathbb{P}(\{\omega\}) : \omega \in \Omega)$.

10.15 Satz (Verteilungsabhängigkeit). Für $X \sim P \in \text{Prob}(\mathbb{R})$ gelten die Implikationen

$$\begin{aligned} \mathbb{E}X \text{ existiert gemäß 10.14} &\Leftrightarrow \mu(P) \text{ existiert gemäß 10.13} \\ &\Rightarrow \mathbb{E}X = \sum_{x \in \mathbb{R}} xP(\{x\}) = \mu(P). \end{aligned}$$

Insbesondere gilt die Implikation

$$X \sim Y, \mathbb{E}X \text{ existiert} \Rightarrow \mathbb{E}X = \mathbb{E}Y.$$

Beweis. Stets gilt, wegen des Umordnungssatzes B.10(8) angewandt auf die paarweise disjunkte Familie $(X^{-1}[\{x\}] : x \in \mathcal{X})$ im ersten Schritt,

$$\begin{aligned} \sum_{\omega \in \Omega} (X(\omega)\mathbb{P}(\{\omega\}))_+ &= \sum_{x \in \mathbb{R}} \sum_{\omega \in X^{-1}[\{x\}]} (X(\omega)\mathbb{P}(\{\omega\}))_+ = \sum_{x \in \mathbb{R}} x_+ \sum_{\omega \in X^{-1}[\{x\}]} \mathbb{P}(\{\omega\}) \\ &= \sum_{x \in \mathbb{R}} x_+ \mathbb{P}(X^{-1}[\{x\}]) = \sum_{x \in \mathbb{R}} (xP(\{x\}))_+ \end{aligned}$$

und analog $\sum_{\omega \in \Omega} (X(\omega)\mathbb{P}(\{\omega\}))_- = \sum_{x \in \mathbb{R}} (xP(\{x\}))_-$, woraus alles folgt. \square

10.16 Aufgabe. (a) Es sei X eine \mathbb{Z} -wertige Zufallsgröße auf (Ω, \mathbb{P}) mit

$$\mathbb{P}(X = k) = \frac{1}{2|k|(|k| + 1)}$$

für $k \in \mathbb{Z} \setminus \{0\}$ und $\mathbb{P}(X = 0) = 0$. Zeigen Sie, dass $\mathbb{E}X$ nicht existiert.

(b) Geben Sie eine \mathbb{N} -wertige Zufallsgröße mit existierendem aber unendlichem Erwartungswert an. \square

10.17 Satz (Rechenregeln für Erwartungswerte). Es seien X, Y reellwertige Zufallsgrößen auf dem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) mit Erwartungswerten $\mathbb{E}X, \mathbb{E}Y \in \overline{\mathbb{R}}$, und es sei $a \in \mathbb{R}$ und $A \subseteq \Omega$. Dann gilt

- (1) $\mathbb{E}\mathbf{1}_A = \mathbb{P}(A)$,
- (2) $\mathbb{E}(aX) = a\mathbb{E}X$ (Homogenität),
- (3) $\mathbb{E}X, \mathbb{E}Y > -\infty \Rightarrow \mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$ (Additivität),
- (4) $X \leq Y \Rightarrow \mathbb{E}X \leq \mathbb{E}Y$ (Isotonie),
- (5) $|\mathbb{E}X| \leq \mathbb{E}|X|$ (Dreiecksungleichung),
- (6) $\mathbb{E}X \in \mathbb{R} \Leftrightarrow \mathbb{E}|X| < \infty$.

In (3) kann die Prämisse auch durch $\mathbb{E}X, \mathbb{E}Y < \infty$ ersetzt werden.

Häufig benutzte Spezialfälle von (1) und (2) sind $\mathbb{E}0 = 0$, $\mathbb{E}1 = 1$, $\mathbb{E}a = a$. Die Prämisse in (3) ist meistens wegen $\mathbb{E}X, \mathbb{E}Y \in \mathbb{R}$ oder $X, Y \geq 0$ erfüllt.

Beweis. (1): Mit Definition 10.14 im ersten Schritt, sowie mit B.10(3) im zweiten und 3.6(4) im dritten, ergibt sich $\mathbb{E}\mathbf{1}_A = \sum_{\omega \in \Omega} \mathbf{1}_A(\omega)\mathbb{P}(\{\omega\}) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \mathbb{P}(A)$. Alternativ folgt (1) aus $\mathbf{1}_A \sim B_{\mathbb{P}(A)}$ nach 5.9 mittels Satz 10.15 im ersten Schritt und B.10(3) im dritten:

$$\mathbb{E}\mathbf{1}_A = \mu(B_{\mathbb{P}(A)}) = \sum_{x \in \mathbb{R}} x \cdot b_{\mathbb{P}(A)}(x) = 1 \cdot \mathbb{P}(A) = \mathbb{P}(A)$$

(rückwärts zu lesen, da dies die Existenz von $\mu(B_{\mathbb{P}(A)})$ und dann von $\mathbb{E}\mathbf{1}_A$ liefert).

Die restlichen Aussagen ergeben sich direkt aus Definition 10.14 und den Rechenregeln für Summen, zum Beispiel im Fall von (3):

$$\begin{aligned} \mathbb{E}X + \mathbb{E}Y &= \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega)\mathbb{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega))\mathbb{P}(\{\omega\}) = \mathbb{E}(X + Y) \end{aligned}$$

wegen B.13(5) im zweiten Schritt. □

Übrigens kann man die beiden obigen Rechenregeln (2) und (3) spezialisiert auf den Fall endlicher Erwartungswerte in der Sprache der Linearen Algebra auch so ausdrücken: *Auf jedem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) bildet die Menge aller reellwertigen Zufallsgrößen mit endlichen Erwartungswerten zusammen mit der üblichen punktweisen Addition und der Multiplikation mit reellen Zahlen einen \mathbb{R} -Vektorraum, und die Abbildung $X \mapsto \mathbb{E}X$ ist eine Linearform auf diesem.*

10.18 Beispiele. Es gilt

- (1) $\mu(\delta_a) = a$ für $a \in \mathbb{R}$,
- (2) $\mu(B_{n,p}) = np$ für $n \in \mathbb{N}_0, p \in [0, 1]$,
- (3) $\mu(H_{n,r,b}) = n \frac{r}{r+b}$ für $n, r, b \in \mathbb{N}_0, n \leq r+b, \frac{0}{0} := 0$.

Beweis. (1): $\mu(\delta_a) = \mathbb{E}a = a$.

(2): Mit einer Bernoulli-Kette $(\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n})$ mit $\mathbb{P}(A_j) = p$ gilt $\sum_{j=1}^n \mathbf{1}_{A_j} \sim B_{n,p}$ und daher, mit Satz 10.15 im ersten Schritt, 10.17(3) im zweiten, und 10.17(1) im dritten,

$$\mu(B_{n,p}) = \mathbb{E} \sum_{j=1}^n \mathbf{1}_{A_j} = \sum_{j=1}^n \mathbb{E}\mathbf{1}_{A_j} = \sum_{j=1}^n \mathbb{P}(A_j) = np.$$

Nochmal (2):

$$\begin{aligned} (4) \quad \mu(B_{n,p}) &= \sum_{x \in \mathbb{R}} x b_{n,p}(x) = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np \sum_{k=0}^{n-1} b_{n-1,p}(k) = np; \end{aligned}$$

dies ist auch korrekt im Trivialfall $n = 0$, der aber auch aus (1) mit $a := 0$ folgt.

(3): In der Situation von Beispiel 7.11(a) gilt

$$(5) \quad \sum_{j=1}^n (X_j \in R) \sim H_{n,r,b}$$

nach 7.11(9); damit folgt, im ersten Schritt analog zum ersten Beweis von (2), und mit der Gleichverteilungsaussage 7.11(6) im zweiten,

$$\mu(H_{n,r,b}) = \sum_{j=1}^n \mathbb{P}(X_j \in R) = n \frac{r}{r+b}.$$

Dabei war in 7.11(a) eigentlich unnötigerweise $n \geq 1$ angenommen; im gegenteiligen Fall ist jedoch $H_{n,r,b} = H_{0,r,b} = \delta_0$ und damit (3) trivial nach (1). \square

10.19 Aufgabe (Erwartungswerte einiger Verteilungen). Bestimmen Sie die Erwartungswerte

- (a) der Gleichverteilungen $U_{\{0,\dots,n-1\}}$ für $n \in \mathbb{N}$,
- (b) der Faltungen von Gleichverteilungen $U_{\{0,\dots,m-1\}} * U_{\{0,\dots,n-1\}}$ für $m, n \in \mathbb{N}$,
- (c) der geometrischen Verteilungen G_p für $p \in]0, 1]$,
- (d) der Negativbinomialverteilungen $NB_{r,p}$ für $r \in \mathbb{N}$ und $p \in]0, 1]$.
(Hinweis: Nutzen Sie die Beziehung $NB_{r,p} = G_p^{*r}$.) \square

In Verallgemeinerung des Spezialfall $g = \text{id}_{\mathbb{R}}$ aus Satz 10.15 gilt:

10.20 Satz (Transformationsformel). *Es sei (Ω, \mathbb{P}) ein Wahrscheinlichkeitsraum, \mathcal{X} eine beliebige Menge, $X : \Omega \rightarrow \mathcal{X}$, und $g : \mathcal{X} \rightarrow \mathbb{R}$. Dann gelten die Implikationen*

$$\mathbb{E}g(X) \text{ existiert} \Leftrightarrow \sum_{x \in \mathcal{X}} g(x) \mathbb{P}(X = x) \text{ existiert} \Rightarrow \mathbb{E}g(X) = \sum_{x \in \mathcal{X}} g(x) \mathbb{P}(X = x).$$

Beweis. Ähnlich wie 10.15. \square

10.21 Aufgabe (Welche Funktionen kommutieren mit der Erwartungswertbildung?). Bestimmen Sie die Menge aller Funktionen $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mit der Eigenschaft, dass

$$\mathbb{E}\varphi(X) = \varphi(\mathbb{E}X)$$

immer dann gilt, wenn beide Seiten existieren.

Hinweis: Um zu zeigen, dass nur die Funktionen aus der von Ihnen vermuteten Menge die gewünschte Eigenschaft haben, reicht es, Zufallsgrößen X mit Zweipunktverteilungen zu betrachten. \square

10.22 Satz (Multiplikativität). *Es seien X, Y unabhängige reellwertige Zufallsgrößen mit $\mathbb{E}X, \mathbb{E}Y \in \mathbb{R}$ oder $X, Y \geq 0$. Dann gilt*

$$(1) \quad \mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y.$$

Beweis. Wir haben

$$\begin{aligned} \mathbb{E}X \cdot \mathbb{E}Y &= \left(\sum_{x \in \mathbb{R}} x \mathbb{P}(X = x) \right) \left(\sum_{y \in \mathbb{R}} y \mathbb{P}(Y = y) \right) \\ &= \sum_{(x,y) \in \mathbb{R}^2} xy \mathbb{P}(X = x) \mathbb{P}(Y = y) = \sum_{(x,y) \in \mathbb{R}^2} xy \mathbb{P}((X, Y) = (x, y)) \\ &= \mathbb{E}(XY) \end{aligned}$$

wegen Satz 10.15 im ersten Schritt, B.13(10) oder B.10(10) im zweiten, $X \perp Y$ im dritten, und Satz 10.20 mit (X, Y) statt X und $g(x, y) := g(x, y)$ im letzten. \square

Ab sofort schreiben wir auch kurz $\mathbb{E}XY := \mathbb{E}(XY)$ und insbesondere $\mathbb{E}X^2 := \mathbb{E}(X^2)$.

Erwartungswerte allein sagen noch nicht viel über eine Verteilung aus:

10.23 Bemerkung (Nichtinformativität des Erwartungswertes). Es seien $a, b, \mu \in \mathbb{R}$ mit $a \leq b$. Dann gilt

$$(1) \quad \{P([a, b]) : P \in \text{Prob}(\mathbb{R}) \text{ mit } \mu(P) = \mu\} \supseteq [0, 1[.$$

Beweis. Es sei $p \in [0, 1[$ gegeben. Für $c, d \in \mathbb{R}$ mit $c < a$ und $d > b$ sei P das Wahrscheinlichkeitsmaß auf \mathbb{R} mit der durch

$$f(x) := \begin{cases} p & \text{für } x = \frac{a+b}{2}, \\ \frac{1-p}{2} & \text{für } x \in \{c, d\} \end{cases}$$

gegebenen Dichte. Dann gilt $P([a, b]) = p$ und

$$\mu(P) = \frac{a+b}{2}p + (c+d)\frac{1-p}{2}.$$

Wegen $(1-p)/2 \neq 0$ und da die Summe $c+d \in \mathbb{R}$ beliebig wählbar ist, gilt $\mu(P) = \mu$ für geeignete c, d . \square

Jedoch erweisen sich Erwartungswerte im Zusammenspiel mit den jetzt zu betrachtenden Varianzen als sehr nützlich, insbesondere durch die Chebyshev-Ungleichung 10.31.

10.24 Definition. Für $X \sim P \in \text{Prob}(\mathbb{R})$ mit endlichem **zweiten Moment**² $\mathbb{E}X^2$ heißt

$$\sigma^2(P) := \text{Var}X := \mathbb{E}(X - \mathbb{E}X)^2$$

Varianz und $\sigma(P) := \sqrt{\text{Var}X}$ **Standardabweichung** von P oder von X . \square

²Allgemeiner heißt für $n \in \mathbb{N}_0$ im Falle der Existenz $\mu_n := \mathbb{E}X^n$ das n -te **Moment** der reellwertigen Zufallsgröße X ; also speziell $\mu_0 = 1, \mu_1 = \mathbb{E}X, \mu_2 = \mathbb{E}X^2$.

Wegen $|X| \leq (1 + X^2)/2$ folgt aus $\mathbb{E}X^2 < \infty$ zunächst $\mathbb{E}|X| < \infty$ nach 10.17(4), also $\mathbb{E}X \in \mathbb{R}$ nach 10.14(2), und damit oben die Wohldefiniertheit von $\text{Var}X$. Die in der Definition von $\sigma^2(P)$ vorausgesetzte ‘‘Verteilungsabhangigkeit’’ von $\text{Var}X$ ist klar, siehe unten 10.25(1). Im Fall von $\mathbb{E}X^2 = \infty$ setzt man ublicherweise noch Varianz und Standardabweichung gleich ∞ .

10.25 Satz. *Es seien X, Y reellwertige Zufallsgroen auf (Ω, \mathbb{P}) mit $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$, $a, b \in \mathbb{R}$, und $A \subseteq \Omega$. Dann gilt*

- (1) $\text{Var}X = \sum_{x \in \mathbb{R}} (x - \mathbb{E}X)^2 \mathbb{P}(X = x),$
- (2) $\text{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2,$
- (3) $\mathbb{E}(X - a)^2 = \text{Var}X + (a - \mathbb{E}X)^2$ (Verschiebungssatz),
- (4) X, Y unabhangig $\Rightarrow \text{Var}(X + Y) = \text{Var}X + \text{Var}Y,$
- (5) $\text{Var}(aX + b) = a^2 \text{Var}X,$
- (6) $\text{Var}X \geq 0$ mit ‘‘=’’ $\Leftrightarrow \exists c \in \mathbb{R}$ mit $\mathbb{P}(X = c) = 1,$
- (7) $\text{Var}\mathbf{1}_A = \mathbb{P}(A)(1 - \mathbb{P}(A)).$

Beweis. (4): Es gilt (wobei eigentlich alles ruckwarts zu lesen ist)

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}\left(X + Y - \mathbb{E}(X + Y)\right)^2 \\ &= \mathbb{E}\left((X - \mathbb{E}X)^2 + (Y - \mathbb{E}Y)^2 + 2(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right) \\ &= \text{Var}X + \text{Var}Y + 2\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y), \end{aligned}$$

und mit der Unabhangigkeit von X und Y im ersten Schritt ergibt sich

$$(8) \quad \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}(X - \mathbb{E}X)\mathbb{E}(Y - \mathbb{E}Y) = (\mathbb{E}X - \mathbb{E}X)(\mathbb{E}Y - \mathbb{E}Y) = 0.$$

Die restlichen Beweise sind einfacher. □

10.26 Beispiel. Fur $n \in \mathbb{N}_0$ und $p \in [0, 1]$ gilt $\sigma^2(B_{n,p}) = np(1 - p)$.

Beweis. Mit einer Bernoulli-Kette $(\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}) \sim B_p^{\otimes n}$ gilt wegen 10.25(4,7)

$$\sigma^2(B_{n,p}) = \text{Var}\left(\sum_{i=1}^n \mathbf{1}_{A_i}\right) = \sum_{i=1}^n \text{Var}\mathbf{1}_{A_i} = \sum_{i=1}^n \mathbb{P}(A_i)(1 - \mathbb{P}(A_i)) = np(1 - p).$$

□

Die zu obigem Beispiel analoge Aussage fur den Fall hypergeometrischer Verteilungen ist zunachst weniger offensichtlich, da sich fur die entsprechende Indikatortensordarstellung die Varianzen mangels Unabhangigkeit nicht aufzuaddieren brauchen. Dies motiviert die Betrachtung der Kovarianz: Das ist der Term, der oben in 10.25(8) im Unabhangigkeitsfall verschwand.

10.27 Definition. Für reellwertige Zufallsgrößen X, Y auf (Ω, \mathbb{P}) mit $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$ heißt

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

Kovarianz von X und Y . □

Unter den genannten Voraussetzungen an X und Y ist $\text{Cov}(X, Y)$ tatsächlich eine wohldefinierte reelle Zahl, was man mit $\mathbb{E}(X - \mathbb{E}X)^2 < \infty, \mathbb{E}(Y - \mathbb{E}Y)^2 < \infty$ und $|(X - \mathbb{E}X)(Y - \mathbb{E}Y)| \leq \frac{1}{2}((X - \mathbb{E}X)^2 + (Y - \mathbb{E}Y)^2)$ erkennen kann.

10.28 Satz (Rechenregeln für Kovarianzen). Für reellwertige Zufallsgrößen X, Y, X_i, Y_j auf (Ω, \mathbb{P}) mit endlichen Varianzen und für $a, b, c, d, a_i, b_j \in \mathbb{R}$ und $A, B \subseteq \Omega$ gilt

$$(1) \quad \text{Cov}(X, Y) = \sum_{(x,y) \in \mathbb{R}^2} (x - \mathbb{E}X)(y - \mathbb{E}Y)\mathbb{P}(X = x, Y = y),$$

$$(2) \quad \text{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y,$$

$$(3) \quad \text{Cov}(X, Y) = \text{Cov}(Y, X), \quad \text{Cov}(X, X) = \text{Var}X,$$

$$(4) \quad \text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y),$$

$$(5) \quad X, Y \text{ unabhängig} \Rightarrow \text{Cov}(X, Y) = 0,$$

$$(6) \quad \text{Cov}\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j) \quad (\text{Bilinearität}),$$

$$(7) \quad \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}X_i + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j),$$

$$(8) \quad \text{Cov}(\mathbf{1}_A, \mathbf{1}_B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B).$$

Beweis. Ähnlich wie in 10.25. Zum Beispiel folgt (8) mit (2) im ersten Schritt in

$$\text{Cov}(\mathbf{1}_A, \mathbf{1}_B) = \mathbb{E}\mathbf{1}_A\mathbf{1}_B - \mathbb{E}\mathbf{1}_A\mathbb{E}\mathbf{1}_B = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B).$$

□

10.29 Korollar (Erwartungswerte und Varianzen von Indikatorkorrekturen). Für $n \in \mathbb{N}$, Ereignisse A_1, \dots, A_n in (Ω, \mathbb{P}) , und $S := \sum_{i=1}^n \mathbf{1}_{A_i}$ gilt

$$(1) \quad \mathbb{E}S = \sum_{i=1}^n \mathbb{P}(A_i)$$

$$(2) \quad \text{Var}S = \sum_{i=1}^n \mathbb{P}(A_i)(1 - \mathbb{P}(A_i)) + 2 \sum_{1 \leq i < j \leq n} \left(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j) \right)$$

Beweis. (1): Klar mit 10.17(3,1), wie in Beispiel 10.18.

(2): Klar mit 10.28(7), 10.25(7), und 10.28(8). □

10.30 Beispiel. Es gilt

$$\sigma^2(H_{n,r,b}) = n \frac{r}{r+b} \frac{b}{r+b} \left(1 - \frac{n-1}{r+b-1} \right) \quad \text{für } n, r, b \in \mathbb{N}_0 \text{ mit } n \leq r+b, \frac{0}{0} := 0;$$

dabei heißt der Faktor in den Klammern **Endlichkeitskorrektur**.

Beweis. Mit der Darstellung 10.18(5), und darin $\mathbb{P}(X_i \in R) = \frac{r}{r+b}$ nach 7.11(9) sowie $\mathbb{P}(X_i \in R, X_j \in R) = \mathbb{P}(X_1 \in R, X_2 \in R) = \frac{r(r-1)}{(r+b)(r+b-1)}$ für $i \neq j$ nach der Vertauschbarkeit 7.11(5) und der Konsistenz 7.9(10) mit $k = 2$ sowie 7.11(4) mit 2 statt n , liefert 10.29(2)

$$\sigma^2(\mathbb{H}_{n,r,b}) = n \frac{r}{r+b} \frac{b}{r+b} + n(n-1) \left(\frac{r(r-1)}{(r+b)(r+b-1)} - \left(\frac{r}{r+b} \right)^2 \right),$$

und eine leichte Rechnung ergibt die Behauptung. □

10.31 Satz (Chebyshev-Ungleichung). *Es sei X eine reellwertige Zufallsgröße mit endlicher Varianz. Dann gilt*

$$(1) \quad \mathbb{P}(|X - \mathbb{E}X| \geq c) \leq \frac{\text{Var}X}{c^2} \quad \text{für } c \in]0, \infty[.$$

Beweis. Mit $\mu := \mathbb{E}X$ und $A := \{x \in \mathbb{R} : |x - \mu| \geq c\}$ gilt

$$(2) \quad \mathbf{1}_A(x) \leq \left(\frac{x - \mu}{c} \right)^2 \quad \text{für } x \in \mathbb{R},$$

was durch eine Skizze klar wird und mittels Unterscheidung der beiden Fälle $x \in A$ und $x \notin A$ bewiesen wird. Einsetzen von X in (2) und Erwartungswertbildung unter Berücksichtigung der Isotonie 10.17(4) liefert (1). □

Im Nichttrivialfall von $\text{Var}X > 0$ ist (1) mittels Ersetzen von c durch $c\sqrt{\text{Var}X}$ äquivalent zu

$$(3) \quad \mathbb{P}\left(\left| \frac{X - \mathbb{E}X}{\sqrt{\text{Var}X}} \right| \geq c \right) \leq \frac{1}{c^2} \quad \text{falls } \text{Var}X, c \in]0, \infty[.$$

10.32 Definition. Es sei X eine \mathbb{R} -wertige Zufallsgröße.

(a) Ist $\mathbb{E}X = 0$, so heißt X (**am Erwartungswert**)³ **zentriert**. Ist $\mathbb{E}|X| < \infty$, so heißt $X - \mathbb{E}X$ die **Zentrierte** von X .

(b) Ist $\mathbb{E}X = 0$ und $\text{Var}X = 1$, so heißt X **standardisiert**. Ist $\text{Var}X \in]0, \infty[$, so heißt

$$(1) \quad \frac{X - \mathbb{E}X}{\sqrt{\text{Var}X}}$$

die **Standardisierte** von X . □

Eine Zentrierte ist offenbar zentriert, eine Standardisierte standardisiert (und damit natürlich auch zentriert). Ist $P \in \text{Prob}(\mathbb{R})$, so heißt gegebenenfalls die Verteilung der Zentrierten von $X \sim P$ auch Zentrierte von P ; analog für Standardisierte.

³Manchmal, insbesondere falls $\mathbb{E}X$ gar nicht existiert, wird eine analoge Definition mit Medianen statt Erwartungswerten verwendet.

10.33 Zahlenbeispiel. Beim 6 000-maligen Würfeln liefert die Chebyshev-Ungleichung

$$\mathbb{P}(900 < \text{Anzahl der Sechsen} < 1100) \geq \frac{11}{12} = 0.916666\dots$$

Der exakte Wert der betrachteten Wahrscheinlichkeit ist laut R

```
> pbinom(1099,6000,1/6) - pbinom(900,6000,1/6)
[1] 0.9994305
```

□

10.34 Aufgabe (Grobe Abschätzungen mit Varianzen und Chebyshev).

(a) Beweisen Sie die folgenden Abschätzungen für die Varianzen der Binomial- und hypergeometrischen Verteilungen:

$$\begin{aligned} \sigma^2(\mathbf{B}_{n,p}) &\leq \frac{n}{4} && \text{für } n \in \mathbb{N}, p \in [0, 1], \\ \sigma^2(\mathbf{H}_{n,r,b}) &\leq \frac{n}{4} && \text{für } n, r, b \in \mathbb{N}_0, n \leq r + b. \end{aligned}$$

(b) Sie wollen von den N Wahlberechtigten Deutschlands n fragen, ob sie am nächsten Sonntag CDU/CSU wählen würden oder nicht. Es sei p der Anteil der CDU/CSU-Wähler unter den N Wahlberechtigten und es sei \hat{p} der zufällige Anteil in Ihrer Stichprobe vom Umfang n . Wie groß ist die Wahrscheinlichkeit für eine “kleine” Abweichung $\mathbb{P}(|\hat{p} - p| \leq 0.01)$ mindestens? Wie groß würden Sie n wählen, um sicherzustellen, dass letztere mindestens 0.9 ist? □

Die Chebyshev-Ungleichung betrifft **zweiseitige Abweichungswahrscheinlichkeiten**. Oft, etwa unten in Beispiel 10.36, sind dagegen **einseitige Abweichungswahrscheinlichkeiten** gefragt, und die können ein bisschen besser abgeschätzt werden:

10.35 Aufgabe (Cantelli-Ungleichung). Für jede reelle Zufallsgröße X mit endlicher Varianz gilt

$$(1) \quad \mathbb{P}(X \geq \mathbb{E}X + c) \leq \frac{\text{Var}X}{c^2 + \text{Var}X} \quad \text{für } c \in]0, \infty[.$$

Hinweis: Ohne Einschränkung kann $\mathbb{E}X = 0$ angenommen werden. □

10.36 Beispiel (Rencontre-Verteilung). Wie unwahrscheinlich ist es, dass im Rencontre-Problem 4.11 mehr als 3, beziehungsweise mehr als 10, Theaterbesucher jeweils ihren Sonnenhut zurückerhalten?

Antwort: Die Wahrscheinlichkeit ist höchstens $\frac{1}{10}$, beziehungsweise kleiner als $\frac{1}{100}$.

Beweis. Es sei $n \in \mathbb{N}$ und es bezeichne P_n die Rencontre-Verteilung aus 4.11(b). Dann gilt

$$(1) \quad \mu(P_n) = 1 \quad \text{und} \quad \sigma^2(P_n) = \begin{cases} 0 & \text{falls } n = 1, \\ 1 & \text{falls } n \geq 2, \end{cases}$$

denn mit der Notation aus 4.11(a) und den dort notierten Mächtigkeiten der Schnitte der $B_i = \{\omega \in \Omega : \omega_i = i\}$ gilt

$$\sum_{i=1}^n \mathbf{1}_{B_i} \sim P_n$$

mit $U_\Omega(B_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$ für $i \in \{1, \dots, n\}$ und $U_\Omega(B_i \cap B_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$ für $(i, j) \in \{1, \dots, n\}_{\neq}^2$, so dass Korollar 10.29 die Beziehung (1) liefert. Damit liefern die Cantelli-Ungleichung 10.35 und eine separate Betrachtung des trivialen Sonderfalls $P_1 = \delta_1$ die Behauptung:

$$P_n(\{4, \dots\}) = P_n([\mu(P_n)+3, \infty[) \leq \frac{1}{3^2 + 1} = \frac{1}{10}$$

und analog $P_n(\{11, \dots\}) \leq \frac{1}{101} < \frac{1}{100}$. □

In konkreten Spezialfällen sind Abschätzungen mittels Chebyshev oder Cantelli fast immer recht grob, aber schärfere oder gar bestmögliche Abschätzungen oft nicht leicht zu erhalten:

10.37 Problem. Für die Folge (P_n) der Rencontre-Verteilungen und für $2 \leq x \in \mathbb{N}$ soll $\sup_{n \in \mathbb{N}} P_n([x, \infty[)$ berechnet werden. □

10.38 Satz ((Schwaches) Gesetz großer Zahlen). *Es sei $P \in \text{Prob}(\mathbb{R})$ mit endlichem Erwartungswert $\mu := \mu(P)$. Dann gibt es zu beliebigen $\varepsilon, \delta > 0$ ein $n_0 = n_0(\varepsilon, \delta, P) \in \mathbb{N}$ mit der Eigenschaft: Ist $n \geq n_0$ und sind X_1, \dots, X_n unabhängige Zufallsgrößen mit $X_i \sim P$ für $i \in \{1, \dots, n\}$, so gilt*

$$(1) \quad \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) < \delta.$$

Beweis. Im Spezialfall endlicher Varianzen $\sigma^2 := \sigma^2(P)$: Unter den betrachteten Voraussetzungen an die X_i gilt

$$\mu = \mathbb{E}X_1 = \mathbb{E}\frac{1}{n} \sum_{i=1}^n X_i$$

und die Chebyshev-Ungleichung liefert

$$(2) \quad \text{L.S.}(1) \leq \frac{\text{Var}(\frac{1}{n} \sum_{i=1}^n X_i)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \cdot \frac{1}{n^2} \sum_{i=1}^n \text{Var}X_i = \frac{\sigma^2}{\varepsilon^2 n}.$$

Folglich kann man $n_0 := \lceil \frac{\sigma^2}{\varepsilon^2 \delta} \rceil + 1$ wählen.

Siehe zum Beispiel Feller (1970, p. 246) für den komplizierteren Beweis im allgemeinen Fall nicht notwendig endlicher Varianzen. □

Spezialisierung auf Bernoulli-Verteilungen liefert:

10.39 Satz. Zu beliebigen $\varepsilon, \delta > 0$ gibt es ein $n_0 = n_0(\varepsilon, \delta) \in \mathbb{N}$ mit der Eigenschaft: Ist $p \in [0, 1]$, $n \geq n_0$, und sind A_1, \dots, A_n unabhängige Ereignisse mit $\mathbb{P}(A_i) = p$ für $i \in \{1, \dots, n\}$, so gilt

$$(1) \quad \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{A_i} - p\right| \geq \varepsilon\right) < \delta.$$

Beweis. Wegen $\text{Var} \mathbf{1}_{A_i} = p(1-p) \leq \frac{1}{4}$ spezialisiert sich 10.38(2) hier zu L.S.(1) $\leq \frac{p(1-p)}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n}$, so dass in der Tat $n_0 := \lceil \frac{1}{4\varepsilon^2 \delta} \rceil + 1$ unabhängig von p gewählt werden kann. \square

10.40 Bemerkung. Satz 10.39 wird in vielen herkömmlichen, den Kollektiv-Begriff übergehenden Einführungen in die Wahrscheinlichkeitsrechnung als eine **Häufigkeitsinterpretation** der Wahrscheinlichkeit liefernd angesehen: Die Ungleichung 10.39(1) besage ja, dass das noch zu interpretierende Objekt p ungefähr gleich der doch offensichtlich interpretierbaren relativen Häufigkeit $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{A_i}$ sei.

Leider enthält jedoch 10.39(1) die so noch nicht interpretierte äußere Wahrscheinlichkeit $\mathbb{P}(\dots)$, und auf \mathbb{P} wird auch in der Unabhängigkeitsannahme Bezug genommen. Schließlich müsste, selbst nach erfolgter Interpretation von \mathbb{P} , noch der Unabhängigkeitsbegriff ohne Rückgriff auf Kollektive interpretiert werden.

Daher kann Ihr Dozent sich der herkömmlichen Ansicht selbstverständlich nicht anschließen. \square

Unter den vielen Verschärfungen des Schwachen Gesetzes großer Zahlen erwähnen wir:

10.41 Satz (Starkes Gesetz großer Zahlen). *Es sei $P \in \text{Prob}(\mathbb{R})$ mit endlichem Erwartungswert $\mu := \mu(P)$. Dann gibt es zu beliebigen $\varepsilon, \delta > 0$ ein $n_0 = n_0(\varepsilon, \delta, P) \in \mathbb{N}$ mit der Eigenschaft: Ist $N \geq n_0$ und sind X_1, \dots, X_N unabhängige Zufallsgrößen mit $X_i \sim P$ für $i \in \{1, \dots, N\}$, so gilt*

$$(1) \quad \mathbb{P}\left(\max_{n=n_0}^N \left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) < \delta.$$

Beweis. Siehe etwa Feller (1970, pp. 260, 234, 203). \square

10.42 Beispiel. Bei rund 10^6 Taufen (in London um das Jahr 1670, siehe Tabelle auf p. 20, dort $\sum_{i=1}^{82} n_i = 938223 \approx 10^6$, $82 \approx 100$, $n_i \approx 10\,000$) ist in den meisten der rund 100 Jahresblöcke von je rund 10 000 Taufen der

$$(1) \quad \text{Jungenanteil} \in]0.510, 0.522[$$

(Spalte x_i/n_i der Tabelle). Dies bestätigt empirisch 10.39(1) mit $n = n_i \approx 10\,000$, $p = 0.516$, $\varepsilon = 0.06$, $\frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{1}_{A_j}(\omega) = \frac{x_i}{n_i} = \text{Jungenanteil}$, $1 - \delta \hat{=} \text{“in den meisten$

Fällen”, und schließlich L.S.10.39(1) (häufigkeits-)interpretiert als L.S.2.2(1) mit dort $n = 100$ statt $n \rightarrow \infty$, $A := “(1)” := \{\frac{1}{m} \sum_{j=1}^m \mathbf{1}_{A_j} \in]0.510, 0.522[\}$ mit $m \approx 10\,000$, A_j unabhängig in einem Wahrscheinlichkeitsraum (\mathcal{X}, P) mit $P(A_j) = p$ für $j \in \{1, \dots, m\}$.

Auch ist wohl in den meisten der rund 10 Blöcke von je rund 100 000 Taufen der Jungenanteil von der zehntausendsten bis zur hunderttausendsten Taufe stets $\in]0.510, 0.522[$ (aus der Tabelle leider nur $\frac{x_1}{n_1}, \frac{x_1+x_2}{n_1+n_2}, \dots, \frac{x_1+\dots+x_{10}}{n_1+\dots+n_{10}}$ für den ersten Block, $\frac{x_{11}}{n_{11}}$ usw. für den zweiten, usw., berechenbar). Dies bestätigt empirisch 10.41(1), spezialisiert auf Bernoulli-Verteilungen analog zu 10.39(1), mit dann $n = 10$ statt $n \rightarrow \infty$ in der Häufigkeitsinterpretation 2.2(1). \square

Mangels Zeit nicht mehr behandelt, jedoch hier stehen gelassen um folgende Aufgabennummern nicht nochmals zu ändern, wurden:

10.43 Erinnerung (Konvexe Funktionen).

10.44 Satz (Jensen-Ungleichung).

Weitere Aufgaben

10.45 Aufgabe (Diskrete Gedächtnislosigkeit). Gelegentlich hat man beim Warten, zum Beispiel auf eine 6 beim “Mensch ärgere Dich nicht”, das Gefühl, dass eine bereits verbrachte Wartezeit ein baldiges Eintreten des ersehnten Ereignisses auch nicht wahrscheinlicher macht. Kann das wirklich so sein?

Bestimmen Sie die Menge aller $P \in \text{Prob}(\mathbb{N}_0)$ mit der Eigenschaft

$$(1) \quad x, t \in \mathbb{N}, P([x, \infty[) > 0 \Rightarrow P([x+t, \infty[| [x, \infty[) = P([t, \infty[).$$

Hinweis: (1) ist eine (sehr einfache) Funktionalgleichung für die komplementäre Verteilungsfunktion von P , und aus letzterer ergibt sich dann auch die gewöhnliche Verteilungsfunktion, oder einfacher auch direkt die Zähldichte. \square

10.46 Aufgabe. (a) Der Erwartungswert als beste konstante Vorhersage. Es sei X eine \mathbb{R} -wertige Zufallsgröße mit $\mathbb{E}X^2 < \infty$. Dann ist der Erwartungswert von X die Minimalstelle der Funktion $\mathbb{R} \ni a \mapsto \mathbb{E}(X - a)^2$, und der Minimalwert also die Varianz.

(b) Beste affine Vorhersagen. Es seien X und Y \mathbb{R} -wertige Zufallsgrößen mit endlichen Varianzen auf demselben Wahrscheinlichkeitsraum. Bestimmen Sie die Minimalstellen und den Minimalwert der Funktion $\mathbb{R}^2 \ni (a, b) \mapsto \mathbb{E}(Y - (aX + b))^2$.

(Schreibt man letzteren im Fall von $\text{Var}X \cdot \text{Var}Y > 0$ als Funktion von $\text{Var}Y$ und des **Korrelationskoeffizienten**

$$(1) \quad \varrho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X \cdot \text{Var}Y}} \in [-1, 1],$$

so erscheint der Betrag oder auch das Quadrat von $\varrho(X, Y)$ als skalenbereinigtes Maß für die affine Vorhersagbarkeit von Y durch X . \square

Zwei \mathbb{R} -wertige Zufallsgrößen X, Y mit $\text{Cov}(X, Y) = 0$ heißen **unkorreliert**; dies sind sie also insbesondere unter der Voraussetzung von 10.28(5).

10.47 Aufgabe. Beweisen Sie die Bilinearität 10.28(6) der Kovarianz. \square

10.48 Aufgabe. Berechnen Sie für $r \in [0, \infty[$ und $p \in]0, 1]$ den Erwartungswert und die Varianz der Negativbinomialverteilung $\text{NB}_{r,p}$.

(*Hinweis* zur Berechnung der Varianz: 10.25(2) und $k^2 = k(k-1) + k$.) \square

11 Poisson-Approximation

11.1 Bemerkung (Lokaler Poisson-Grenzwertsatz für Binomialverteilungen). Wie läßt sich eine Binomialverteilung $B_{n,p}$ für große n approximieren? Für $X \sim B_{n,p}$ und $k \in \mathbb{N}_0$ mit $k < np$ gilt

$$\begin{aligned} B_{n,p}(\{0, \dots, k\}) &= \mathbb{P}(X \leq k) = \mathbb{P}(X - np \leq k - np) \\ &\leq \mathbb{P}(|X - np| \geq np - k) \leq \frac{np(1-p)}{(np-k)^2} \end{aligned}$$

wegen $k - np \leq 0$ im dritten Schritt und der Chebyshev-Ungleichung¹ im vierten. Folglich gilt

$$(1) \quad \lim_{n \rightarrow \infty} B_{n,p}(\{0, \dots, k\}) = 0 \quad \text{für } p \in]0, 1] \text{ und } k \in \mathbb{N}_0.$$

Da somit bei festem $p > 0$ und $n \rightarrow \infty$ die gesamte Binomialwahrscheinlichkeit zusammen mit dem Erwartungswert np im Unendlichen verschwindet, betrachten wir nun $p = p_n$ variabel mit $\mu(B_{n,p}) = np_n$ fest, oder gleich allgemeiner konvergent in \mathbb{R} :

Ist

$$(2) \quad p_\bullet \in [0, 1]^{\mathbb{N}} \text{ mit } \lim_{n \rightarrow \infty} np_n = \lambda \in [0, \infty[$$

sowie abkürzend $p := p_n$ und $\lambda_n := np_n$, so ergibt sich für $k \in \mathbb{N}_0$ zunächst

$$\begin{aligned} b_{n,p_n}(k) &= \frac{n^k}{k!} p^k (1-p)^{n-k} = \frac{(np)^k}{k!} (1-p)^n \frac{n^k}{n^k} (1-p)^{-k} \\ &= \frac{(\lambda_n)^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(\prod_{j=1}^k \frac{n-j+1}{n}\right) \left(1 - \frac{\lambda_n}{n}\right)^{-k}, \end{aligned}$$

und daraus unter Verwendung von $\lim_{n \rightarrow \infty} \lambda_n = \lambda$ aus (2) nach den Rechenregeln für konvergente Folgen und der auf kompakten Intervallen gleichmäßigen Konvergenz von $(1 + \frac{x}{n})^n$ gegen e^x , welch letztere bekanntlich mittels Logarithmierung aus $\frac{y}{1+y} \leq \log(1+y) \leq y$ für $y > -1$ folgt, dann

$$(3) \quad \lim_{n \rightarrow \infty} b_{n,p_n}(k) = \frac{\lambda^k}{k!} e^{-\lambda} = p_\lambda(k) \quad \text{für } k \in \mathbb{N}_0$$

mit der in 3.12 eingeführten Zähldichte p_λ der Poisson-Verteilung P_λ . □

11.2 Beispiel. Für $n = 365$ Ihnen unbekannte Personen sei X die Anzahl derer, die heute Geburtstag haben. Würden Sie auf $\{X \geq 2\}$ wetten wollen? Oder auf $\{X \text{ gerade}\}$? Können Sie schnell eine Antwort nur mit Bleistift und Papier geben? (Nehmen Sie dabei an, dass Sie 3.13(2) zwischenzeitlich vergessen haben.)

¹Alternativ könnte man statt des dritten Schrittes etwas schärfer mit der Cantelli-Ungleichung 10.35 abschätzen, oder?

Unter Vernachlässigung von Schaltjahren, Zwillingen und anderen Geschwistern, sowie der ungleichmäßigen Verteilung der Geburtstage über das Jahr ist $X \sim B_{n,p}$ mit $p = 1/365$ die naheliegende Modellannahme. Naive “Anwendung” des Poisson-Grenzwertsatzes 11.1(2) \Rightarrow (3) mit $\lambda := np = 1$ liefert

$$\begin{aligned} \mathbb{P}(X \geq 2) &= 1 - b_{n,p}(0) - b_{n,p}(1) \approx 1 - p_\lambda(0) - p_\lambda(1) = 1 - \frac{2}{e} < \frac{1}{2} \\ \mathbb{P}(X \text{ gerade}) &= \sum_{k \in \mathbb{N}_0} b_{n,p}(2k) \approx \sum_{k \in \mathbb{N}_0} p_\lambda(2k) = \cosh(\lambda) e^{-\lambda} = \frac{1}{2} + \frac{1}{2e^2} > \frac{1}{2} \end{aligned}$$

Würden Sie also gerne einerseits gegen $\{X \geq 2\}$ und andererseits auf $\{X \text{ gerade}\}$ wetten wollen? □

Offenbar werden derartige “praktische” Fragen durch einen reinen Grenzwertsatz wie 11.1(2) \Rightarrow (3) noch nicht beantwortet. Daher interessieren Ergebnisse wie die folgenden beiden:

11.3 Satz (Poisson-Approximationssatz für Summen unabhängiger Indikatoren). *Es seien X_1, \dots, X_n unabhängig mit $X_i \sim B_{p_i}$ und $p_i \in [0, 1]$ für $i \in \{1, \dots, n\}$, und es sei $S := \sum_{i=1}^n X_i$ und $\lambda := \sum_{i=1}^n p_i$. Dann gilt*

$$(1) \quad |\mathbb{P}(S \in A) - P_\lambda(A)| \leq \min(1, \frac{1}{\lambda}) \sum_{i=1}^n p_i^2 \quad \text{für } A \subseteq \mathbb{N}_0$$

und

$$(2) \quad \sup_{A \subseteq \mathbb{N}_0} \text{L.S.}(1) \geq \frac{1}{14} \text{R.S.}(1).$$

11.4 Korollar (Poisson-Approximation von Binomialverteilungen). *Für $n \in \mathbb{N}$, $p \in [0, 1]$ und $\lambda = np$ gilt*

$$(1) \quad \sup_{A \subseteq \mathbb{N}_0} |B_{n,p}(A) - P_\lambda(A)| \leq p \min(1, np)$$

und L.S.(1) $\geq \frac{1}{14}$ R.S.(1).

Beweis von Satz 11.3 und Korollar 11.4. Wir zeigen unten hinter Aufgabe 11.10 nur die leichtere Hälfte von 11.3(1), nämlich

$$(2) \quad \sup_{A \subseteq \mathbb{N}_0} |\mathbb{P}(S \in A) - P_\lambda(A)| \leq \sum_{i=1}^n p_i^2$$

und verweisen für den Rest des Beweises von 11.3 auf Barbour/Holst/Janson (1992, pp. 8, 61). Speziell für $p_1 = \dots = p_n = p$ ergibt sich daraus Korollar 11.4. □

11.5 Beispiel. In Beispiel 11.2 ist $p \min(1, np) = p = \frac{1}{365}$ und damit

$$\begin{aligned} \mathbb{P}(X \geq 2) &\leq 1 - \frac{2}{e} + \frac{1}{365} < \frac{1}{2}, \\ \mathbb{P}(X \text{ gerade}) &\geq \frac{1}{2} + \frac{1}{2e^2} - \frac{1}{365} > \frac{1}{2}. \end{aligned}$$

Dabei reicht die unten wirklich bewiesene Hälfte 11.4(2) der Fehlerabschätzung. \square

11.6 Definition. Es sei \mathcal{X} eine Menge. Für $P, Q \in \text{Prob}(\mathcal{X})$ heißt

$$D(P, Q) := \sup_{A \subseteq \mathcal{X}} |P(A) - Q(A)|$$

der **Supremumsabstand**² von P zu Q . \square

Zum Beispiel ist dann mit $\mathcal{X} = \mathbb{N}_0$

$$(1) \quad \text{L.S. 11.4(2)} = D(S \square \mathbb{P}, P_\lambda) = D\left(\bigstar_{i=1}^n B_{p_i}, P_\lambda\right).$$

Offenbar ist D eine Metrik auf $\text{Prob}(\mathcal{X})$, d.h. für $P, Q, R \in \text{Prob}(\mathcal{X})$ gilt

$$\begin{aligned} D(P, Q) &\in [0, \infty[, \quad \text{und} \quad D(P, Q) = 0 \Leftrightarrow P = Q, \\ D(P, Q) &= D(Q, P), \\ (2) \quad D(P, R) &\leq D(P, Q) + D(Q, R). \end{aligned}$$

11.7 Lemma. Es seien $P, Q \in \text{Prob}(\mathcal{X})$ mit den Dichten f, g . Mit

$$B := \{x \in \mathcal{X} : f(x) > g(x)\}$$

gilt

$$(1) \quad D(P, Q) = P(B) - Q(B) = \frac{1}{2} \sum_{x \in \mathcal{X}} |f(x) - g(x)|.$$

Beweis. Für $A \subseteq \mathcal{X}$ gilt dann

$$\begin{aligned} |P(A) - Q(A)| &= \left| \sum_{x \in \mathcal{X}} (\mathbf{1}_A(x) - \frac{1}{2})(f(x) - g(x)) \right| \\ &\leq \sum_{x \in \mathcal{X}} |\mathbf{1}_A(x) - \frac{1}{2}| \cdot |f(x) - g(x)| = \frac{1}{2} \sum_{x \in \mathcal{X}} |f(x) - g(x)|, \end{aligned}$$

und speziell für $A = B$ gilt oben überall Gleichheit sowie $P(B) \geq Q(B)$. \square

²Daneben wird in der Literatur noch $2D(P, Q)$, aber leider von manchen Autoren auch $D(P, Q)$, als **Totalvariationsabstand** bezeichnet.

11.8 Beispiel. Wir betrachten den Spezialfall $n = 1$ in 11.6(1): Für $p \in [0, 1]$ und $x \in \mathcal{X} := \mathbb{N}_0$ ist

$$\begin{aligned} f(x) &:= b_p(x) = 1 - p, \quad p, 0, \dots \\ g(x) &:= p_p(x) = e^{-p}, pe^{-p}, \dots \end{aligned}$$

und unter Verwendung von $e^{-p} > 1 - p$ und $pe^{-p} < p$ für $p > 0$ folgt $B := \{f > g\} = \{1\}$ für $p \in]0, 1]$ und $B = \emptyset$ für $p = 0$, mit 11.7(1) im ersten Fall und trivialerweise für $p = 0$ also

$$(1) \quad D(B_p, P_p) = f(1) - g(1) = p - pe^{-p} \leq p^2$$

mit $e^{-p} \geq 1 - p$ im letzten Schritt. □

11.9 Lemma (Abstände von Faltungen). *Es sei $(\mathcal{X}, +)$ eine abelsche Gruppe, und es seien $n \in \mathbb{N}$ und $P_1, \dots, P_n, Q, Q_1, \dots, Q_n \in \text{Prob}(\mathbb{R})$. Dann gilt*

$$\begin{aligned} (1) \quad & D(P_1 * Q, P_2 * Q) \leq D(P_1, P_2) \\ (2) \quad & D\left(\bigstar_{i=1}^n P_i, \bigstar_{i=1}^n Q_i\right) \leq \sum_{i=1}^n D(P_i, Q_i) \end{aligned}$$

Beweis. Zum Nachweis von (1) seien f_1, f_2, g die Dichten von P_1, P_2, Q . Nach Satz 9.13 ist dann

$$\mathcal{X} \ni x \mapsto \sum_{y \in \mathcal{X}} f_i(x - y)g(y)$$

Dichte von $P_i * Q$ für $i \in \{1, 2\}$, und mit 11.7(1) folgt

$$\begin{aligned} \text{L.S.}(1) &= \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{X}} f_1(x - y)g(y) - \sum_{y \in \mathcal{X}} f_2(x - y)g(y) \right| \\ &\leq \sum_{y \in \mathcal{X}} \frac{1}{2} \sum_{x \in \mathcal{X}} |f_1(x - y) - f_2(x - y)| g(y) = \text{R.S.}(1). \end{aligned}$$

Die Ungleichung (2) ist für $n = 1$ trivial und folgt per Induktion aus dem Spezialfall $n = 2$, der sich aus der Dreiecksungleichung 11.6(2) und aus (1) ergibt:

$$\begin{aligned} D(P_1 * P_2, Q_1 * Q_2) &\leq D(P_1 * P_2, Q_1 * P_2) + D(Q_1 * P_2, Q_1 * Q_2) \\ &\leq D(P_1, Q_1) + D(P_2, Q_2). \end{aligned} \quad \square$$

11.10 Aufgabe. Führen Sie den im Beweis von 11.9 genannten Induktionsbeweis aus. □

Beweis von 11.4(2). Mit $S \sim \bigstar_{i=1}^n B_{p_i}$ im ersten Schritt, $\lambda = \sum_{i=1}^n p_i$ und Aufgabe 9.14(a) im zweiten, Lemma 11.9 mit $\mathcal{X} = \mathbb{Z}$ im dritten, und Beispiel 11.5 im vierten folgt

$$\begin{aligned} \text{L.S. 11.4(2)} &= D\left(\bigstar_{i=1}^n B_{p_i}, P_\lambda\right) = D\left(\bigstar_{i=1}^n B_{p_i}, \bigstar_{i=1}^n P_{p_i}\right) \leq \sum_{i=1}^n D(B_{p_i}, P_{p_i}) \\ &\leq \sum_{i=1}^n p_i^2. \end{aligned} \quad \square$$

Weitere Aufgaben

11.11 Aufgabe (Modellierung mit Poisson-Verteilungen). **(a)** Aus einer Urne mit einer roten und 999 999 blauen Kugeln wird 1 000 000 mal mit Zurücklegen rein zufällig eine Kugel gezogen. Wie groß ist die Wahrscheinlichkeit, dass insgesamt genau zwei mal eine rote Kugel gezogen wird?

(b) An einem Sommerabend werde durchschnittlich alle 10 Minuten eine Sternschnuppe beobachtet. Wie groß ist die Wahrscheinlichkeit dafür, dass während einer Viertelstunde zwei Sternschnuppen beobachtet werden?

Hinweis: Siehe 3.12 für $\mu(P_\lambda)$, bewiesen in Aufgabe 11.13. □

11.12 Aufgabe (Einfache Beispiele zum Supremumsabstand). Vervollständigen Sie die folgenden Gleichungen:

$$\begin{aligned} D(U_{\{0, \dots, m-1\}}, U_{\{0, \dots, n-1\}}) &= && \text{für } m, n \in \mathbb{N}, \\ D(B_p, B_q) &= && \text{für } p, q \in [0, 1]. \end{aligned}$$

□

11.13 Aufgabe. Schon der lokale Grenzwertsatz 11.1(2) \Rightarrow (3) und dann erst recht Korollar 11.4 lassen den ersten Schritt in der Gleichungskette

$$(1) \quad \sigma^2(P_\lambda) = \lim_{n \rightarrow \infty} \sigma^2(B_{n, \frac{\lambda}{n}}) = \lim_{n \rightarrow \infty} \lambda \left(1 - \frac{\lambda}{n}\right) = \lambda \quad \text{für } \lambda \in [0, \infty[$$

plausibel erscheinen, beweisen ihn aber nicht.

(a) Geben Sie Wahrscheinlichkeitsmaße $P, P_n \in \text{Prob}(\mathbb{N}_0)$ mit endlichen Varianzen, $\lim_{n \rightarrow \infty} D(P_n, P) = 0$, aber nicht $\lim_{n \rightarrow \infty} \sigma^2(P_n) = \sigma^2(P)$ an.

Hinweis: Es reichen jeweils sehr wenige Trägerpunkte.

(b) Beweisen Sie (1).

Hinweis: Berechnen Sie zunächst $\mathbb{E}X$ und $\mathbb{E}X(X-1)$ für $X \sim P_\lambda$. □

12 Ein Zentraler Grenzwertsatz

Viele praktische Fragen zur “Schätzung” einer Wahrscheinlichkeit oder eines Anteils – eines Grundproblems der in Kapitel 13 näher betrachteten Mathematischen Statistik – führen in der Planungsphase, also vor einer Datenerhebung, auf Probleme wie das folgende:

12.1 Beispiel. Es seien $c \in]0, \infty[$ und $\alpha \in]0, 1[$. Gesucht ist ein möglichst kleines $n \in \mathbb{N}$ derart, dass

$$(1) \quad \mathbb{P}\left(\left|\frac{1}{n}X - p\right| < c\right) \geq 1 - \alpha$$

immer dann gilt, wenn

(a) $p \in [0, 1]$ und $X \sim B_{n,p}$, beziehungsweise

(b) $r, b \in \mathbb{N}_0$ mit $r + b \geq n$, $p := \frac{r}{r+b}$, und $X \sim H_{n,r,b}$.

Wir betrachten zur Illustration insbesondere das Zahlenbeispiel

$$(2) \quad c = 0.02, \quad \alpha = 0.05.$$

Wie in Aufgabe 10.34 gesehen, gilt in (a) und in (b) nach Chebyshev stets

$$(3) \quad \text{L.S.}(1) \geq 1 - \frac{\text{Var}\left(\frac{1}{n}X\right)}{c^2} \geq 1 - \frac{p(1-p)}{nc^2} \geq 1 - \frac{1}{4nc^2},$$

so dass

$$(4) \quad n \geq \frac{1}{4c^2\alpha}$$

für (1) reicht; speziell für (2) ist also $n = 12\,500$ genug. Aber muss n wirklich so groß sein? \square

Für gegebenes c und α lässt sich für jedes $n \in \mathbb{N}$ die Frage der Zugehörigkeit zu

$$(5) \quad \mathcal{N}_{B,c,\alpha} := \{n \in \mathbb{N} : (1) \text{ gilt unter } 12.1(a)\},$$

$$(6) \quad \mathcal{N}_{H,c,\alpha} := \{n \in \mathbb{N} : (1) \text{ gilt unter } 12.1(b)\}$$

sowie, relevant für die Schätzung eines Anteils in einer Population mit bekannter Gesamtgröße, bei gegebenem $N \in \mathbb{N}$

$$(7) \quad \mathcal{N}_{H,N,c,\alpha} := \{n \in \mathbb{N} : (1) \text{ gilt unter } 12.1(b) \text{ mit } r + b = N\}$$

im Prinzip leicht numerisch untersuchen, etwa unter Verwendung von R. Dennoch hätte man gerne analytische Bedingungen an n , wie (4) aber weniger restriktiv, welche, notfalls bis auf kleine Approximationsfehler, die Zugehörigkeit zu den jeweiligen Mengen in (5–7) implizieren oder sogar dazu äquivalent sind. Mit derartigen analytischen Bedingungen ließe sich der Einfluss der Wahl von c und α zumindest im Groben besser überblicken als durch punktuelle numerische Rechnungen. Außerdem hätte man damit eine gewisse Kontrolle einer rundungs-, abbruch-, und vor allem programmier- oder auch nur tippfehleranfälligen Numerik. Analytische Bedingungen der gewünschten Art ergeben sich aus dem unten formulierten Satz 12.6(b).

12.2 Definition. Die durch

$$\begin{aligned}\varphi(x) &:= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} && \text{für } x \in \mathbb{R}, \\ \Phi(z) &:= \int_{-\infty}^z \varphi(x) \, dx && \text{für } z \in \mathbb{R}\end{aligned}$$

definierten Funktionen φ und Φ sowie der letzteren Umkehrfunktion Φ^{-1} heißen **Dichte** beziehungsweise **Verteilungsfunktion** beziehungsweise **Quantilfunktion** der **Standardnormalverteilung**. \square

12.3 Bemerkung. Die Funktion φ ist keine Dichte eines elementaren Wahrscheinlichkeitsmaßes auf \mathbb{R} , denn es gilt ja $\sum_{x \in \mathbb{R}} \varphi(x) = \infty \neq 1$, sondern sie ist im Sinne der Maß- und Integrationstheorie eine Wahrscheinlichkeitsdichte bezüglich des Lebesgue-Maßes auf \mathbb{R} , was gerade $\varphi \geq 0$ und 12.4(1) bedeutet, und sie definiert daher ein Wahrscheinlichkeitsmaß auf der Borel- σ -Algebra auf \mathbb{R} , welches dann Standardnormalverteilung genannt wird. Dieses alles nicht als aus einer MIT oder WT bekannt annehmend, fassen wir hier “Dichte der Standardnormalverteilung” usw. als bloße zusammengesetzte Namen der drei analytischen Objekte φ, Φ, Φ^{-1} auf. \square

12.4 Satz. *Es ist φ gerade, überall > 0 , auf $[0, \infty[$ streng fallend, mit*

$$(1) \quad \int_{-\infty}^{\infty} \varphi(x) \, dx = 1.$$

Φ ist streng wachsend mit

$$(2) \quad \lim_{z \rightarrow -\infty} \Phi(z) = 0 =: \Phi(-\infty), \quad \lim_{z \rightarrow \infty} \Phi(z) = 1 =: \Phi(\infty)$$

Es ist $\Phi(0) = \frac{1}{2}$, und $\Phi - \frac{1}{2}$ ist ungerade:

$$(3) \quad \Phi(z) = 1 - \Phi(-z) \quad \text{für } z \in \mathbb{R},$$

$$(4) \quad \Phi^{-1}(u) = -\Phi^{-1}(1 - u) \quad \text{für } u \in]0, 1[.$$

Weiter gilt die **Standardflankenungleichung**¹ für Φ ,

$$(5) \quad \Phi(-z) = 1 - \Phi(z) < \frac{\varphi(z)}{z} \quad \text{für } z \in]0, \infty[.$$

Beweis. Für (1) siehe Storch/Wiebe (2003) oder Elstrodt (2018), oder überhaupt jedes gute Lehrbuch der Analysis. Damit ist bis (4) alles klar, wobei (2) auch aus (5) folgt, und (5) zum Beispiel aus $1 - \Phi(z) = \int_z^{\infty} \varphi(x) \, dx = \int_z^{\infty} \frac{1}{x} x \varphi(x) \, dx < \frac{1}{z} \int_z^{\infty} (-\varphi'(x)) \, dx = \frac{\varphi(z)}{z}$. \square

¹Ist F Verteilungsfunktion eines (elementaren) Wahrscheinlichkeitsmaßes $P \in \text{Prob}(\mathbb{R})$, so bezeichnet man F eingeschränkt auf “kleine Argumente” als die **linke Flanke** von P oder von F , und $1 - F$, oder auch die komplementäre Verteilungsfunktion $1 - F(\cdot)$ aus 10.7(4), eingeschränkt auf “große Werte” als **rechte Flanke**. Gleiche Terminologie für allgemeine Wahrscheinlichkeitsmaße gemäß der Bemerkungen 12.3 und 12.21, und damit dann auch für die (nichtelementare) Verteilungsfunktion Φ .

Φ ist in vielen Stochastik-Lehrbüchern tabelliert², siehe etwa Henze (2021, p. 362) oder Krenzel, in \mathbb{R} ist

$$(6) \quad \Phi = \text{pnorm}, \quad \varphi = \text{dnorm}, \quad \Phi^{-1} = \text{qnorm},$$

und wir merken uns die beiden fettgedruckten der speziellen Quantile

$$(7) \quad \begin{array}{c|ccccc} u & .9 & \mathbf{.95} & \mathbf{.975} & .99 & .995 \\ \hline \Phi^{-1}(u) & 1.282 & \mathbf{1.645} & \mathbf{1.96} & 2.326 & 2.576 \end{array}$$

(jedes auf drei Nachkommastellen korrekt gerundet, dabei jedes außer $\Phi^{-1}(.99)$ aufgerundet).

12.5 Aufgabe. (a) Zeigen Sie mit der Regel von de L'Hospital, dass die Standardflankenungleichung 12.4(5) für Φ asymptotisch scharf ist:

$$(1) \quad \Phi(-z) \sim \frac{\varphi(z)}{z} \sim 1 - \Phi(z) \quad \text{für } z \rightarrow \infty.$$

(b) Zeigen Sie, offenbar in Präzisierung von (1),

$$(2) \quad \Phi(-z) = 1 - \Phi(z) > \varphi(z) \cdot \left(\frac{1}{z} - \frac{1}{z^3} \right) \quad \text{für } z \in]0, \infty[$$

(etwa mit partieller Integration im obigen Beweis von 12.4 (5) oder, leichter auch weitere Verschärfungen beider Ungleichungen liefernd, mittels Transformation von \int_z^∞ auf \int_0^∞ , Faktorisierung des Integranden, und der Kenntnis der Vorzeichen der Taylor-Restglieder um Null von e^{-y} für $y > 0$). \square

12.6 Satz (Zentraler Grenzwertsatz für Poisson-, Binomial- und hypergeometrische Verteilungen, mit Fehlerschranken vom Berry-Esseen-Typ). *Es sei*

$$\begin{aligned} \mathcal{P} := & \{ P_\lambda : \lambda \in]0, \infty[\} \cup \{ B_{n,p} : n \in \mathbb{N}, p \in]0, 1[\} \\ & \cup \{ H_{n,r,b} : n, r, b \in \mathbb{N}, n < r + b \}, \end{aligned}$$

also die Menge der Poisson-, Binomial- und hypergeometrischen Verteilungen mit strikt positiven Varianzen.

(a) **Lokaler ZGWS.** *Es gibt ein $A \in]0, \infty[$ mit der Eigenschaft: Für*

$$(1) \quad P \in \mathcal{P}, \quad \mu := \mu(P), \quad \sigma := \sigma(P), \quad k \in \mathbb{Z}$$

gilt

$$(2) \quad \left| P(\{k\}) - \frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right) \right| \leq \frac{A}{\sigma^2}.$$

²Dass bisher kein Integrations-Trick zur “expliziten, integralfreien” Darstellung von Φ gefunden wurde, liegt daran, dass Φ nach einem Satz von Liouville in einem präzisen Sinne keine **elementare Funktion** ist; siehe dazu Rosenlicht (1972) und, für eine einerseits behutsamere und ausführlichere aber andererseits nicht ganz exakte Darstellung, Behrends (2007, pp. 149–161).

(b) **Globaler ZGWS.** Es gibt ein $A \in]0, \infty[$ mit der Eigenschaft: Für

$$(3) \quad P \in \mathcal{P}, \quad \mu := \mu(P), \quad \sigma := \sigma(P), \\ I \subseteq \mathbb{R} \text{ Intervall mit den Randpunkten } a, b \in \overline{\mathbb{R}} \text{ mit } a \leq b$$

gilt

$$(4) \quad \left| P(I) - \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right) \right| \leq \frac{A}{\sigma}.$$

Beweis. Siehe unten hinter Lemma 12.15. □

12.7 Bemerkung. (a) Globaler Fall. Es bezeichne $A_{12.6(b)}$ die optimale Konstante in 12.6(b). Diese ist unbekannt. Es gilt jedenfalls

$$(1) \quad 0.4288\dots = \frac{1}{\sqrt{2e}} \leq A_{12.6(b)} \leq 1.1166.$$

Weiter gilt für jedes $P \in \mathcal{P}$ wie in Satz 12.6

$$(2) \quad \frac{1}{\sqrt{1+12\sigma^2}} \leq \sup_I \text{L.S. 12.6(4)} \leq \frac{A_{12.6(b)}}{\sigma};$$

die sich aus 12.6(4) gleichmäßig für alle Intervalle ergebende obere Abschätzung ist also für von Null wegbeschränkte σ von der richtigen Ordnung.

Die obere Abschätzung in (1) und die untere Abschätzung in (2) werden bei Mattner/Schulz (2014, Absatz um Formel (17)) erklärt.

Aus Schulz (2016) ergibt sich, dass bei Beschränkung auf Poisson- und Binomialverteilungen die Schranke 1.1166 in (1) durch 0.8195 ersetzt werden kann.

(b) **Lokaler Fall.** Wiederum bei Beschränkung auf Poisson- und Binomialverteilungen ergibt sich aus Zolotukhin/Nagaev/Chebotarev (2018, p. 393, Lemma 5), dass 12.6(2) mit $A = 0.516$ gilt.

Im symmetrischen Binomialfall von $p = \frac{1}{2}$ kann R.S.12.6(2) zu $\frac{A}{\sigma^3}$ verschärft werden, was zumindest im Prinzip spätestens seit Laplace (1812) bekannt ist, siehe dazu Hald (1990, p. 498, (1)), aber erst van Nerven (2023) zeigte, dass dabei dann $A = \frac{1}{16\sqrt{2\pi}} = 0.0249\dots$ optimal ist. □

12.8 Aufgabe. Zeigen Sie die untere Abschätzung in 12.7(1) durch Betrachtung von Einpunktintervallen und Poisson-Verteilungen. □

Satz 12.6(b) wird häufig in folgender Umformulierung für standardisierte Zufallsgrößen gemäß 10.32(1) verwendet:

12.9 Bemerkung. Es sei $X \sim P$ mit $P \in \mathcal{P}$ sowie $\mu := \mu(P)$ und $\sigma := \sigma(P)$ wie wie in Satz 12.6(b). Für $-\infty \leq a \leq b \leq \infty$, mit nicht $a = b \in \{-\infty, \infty\}$, gilt dann

$$(1) \quad \left| \mathbb{P}\left(a \underset{(-)}{\leq} \frac{X-\mu}{\sigma} \underset{(-)}{\leq} b\right) - (\Phi(b) - \Phi(a)) \right| \leq \frac{A}{\sigma}$$

wegen 12.6(4) mit $\mu + \sigma a, \mu + \sigma b$ statt der dortigen a, b , und der Spezialfall $a = -b$ von (1) liefert für $b \geq 0$

$$(2) \quad \left| \mathbb{P} \left(\left| \frac{X - \mu}{\sigma} \right| \underset{(-)}{\leq} b \right) - (2\Phi(b) - 1) \right| \leq \frac{A}{\sigma}$$

wegen 12.4(3). □

12.10 Beispiel. In Beispiel 12.1 ist

$$(1) \quad \mathbb{P} \left(\left| \frac{1}{n}X - p \right| < c \right) = \mathbb{P} \left(\left| \frac{X - np}{\sigma} \right| < \frac{nc}{\sigma} \right) \approx 2\Phi\left(\frac{nc}{\sigma}\right) - 1,$$

wobei das streng genommen bedeutungslose Zeichen “ \approx ” hier für eine naive “Anwendung” von 12.9(2) steht. Nun ist

$$(2) \quad 2\Phi\left(\frac{nc}{\sigma}\right) - 1 \geq 1 - \alpha$$

wegen der strengen Isotonie von Φ äquivalent zu $n \geq \frac{\sigma}{c}\Phi^{-1}(1 - \frac{\alpha}{2})$, und letztere Ungleichung gilt, wegen $\sigma^2 \leq \frac{n}{4}$ wie aus Aufgabe 10.34(a) bekannt, immer dann wenn $n \geq \frac{1}{2c}\sqrt{n}\Phi^{-1}(1 - \frac{\alpha}{2})$ gilt, was, unter Berücksichtigung von $\Phi^{-1}(1 - \frac{\alpha}{2}) > 0$ wegen $1 - \frac{\alpha}{2} > \frac{1}{2}$, zu

$$(3) \quad n \geq \left(\frac{\Phi^{-1}(1 - \frac{\alpha}{2})}{2c} \right)^2$$

äquivalent ist. Bei Vernachlässigung des Approximationsfehlers in (1) wäre also (3) eine hinreichende Bedingung für 12.1(1), der man insbesondere die Abhängigkeit des kleinstmöglichen n von c gut ansieht. Für $\alpha = 0.05$ und mit $\Phi^{-1}(.975) = 1.96$ sicherheitshalber zu 2 gerundet ergibt (3) die Bedingung

$$(4) \quad n \geq \frac{1}{c^2},$$

während hier 12.1(4) einen fünf mal so großen Mindeststichprobenumfang verlangt. Speziell für 12.1(2) müsste also $n \geq 2500$ hinreichend für 12.1(1) sein. □

12.11 Aufgabe. Überprüfen Sie numerisch in \mathbb{R} , ob 12.1(1) im Binomialfall und mit 12.1(2) für $n = 2500$ gilt. Benutzen Sie dabei etwa die Kommandos `dbinom`, `pbinom`, `plot`. Hat die linke Seite von 12.1 ein einfaches Monotonieverhalten? Wenn man am Ende von Beispiel 12.10 mit 1.96 statt 2 rechnet, gilt dann 12.1(1)? Und wenn man dabei, wie Henze (2017, p. 255), das strikte Kleinerzeichen links in 12.1(1) durch “ \leq ” ersetzt? □

12.12 Aufgabe (Naives Rechnen mit dem ZGWS). Es sei $X \sim B_{n,p}$ eine Binomialverteilte Zufallsgröße. Beantworten Sie die folgenden Fragen mittels naiver Approximationen nach Satz 12.6(b) oder Bemerkung 12.9.

- (a) Es gelte $p = 0.4$ und $n = 1500$. Wie groß ist $\mathbb{P}(0.40 \leq \frac{X}{n} \leq 0.44)$?
- (b) Es sei $p = 0.375$ fest. Wie groß muss n gewählt werden, damit die Wahrscheinlichkeit dafür, dass $|\frac{X}{n} - p| \leq 0.01$ ist, mindestens 0.995 beträgt?
- (c) Es sei $p = \frac{2}{3}$, $n = 1200$. Wie groß muss ε gewählt werden, damit die Wahrscheinlichkeit für $|\frac{X}{n} - p| < \varepsilon$ mindestens 0.985 beträgt?
- (d) Es sei $n = 14400$ fest. Für welche Werte von p wird die Wahrscheinlichkeit für $|\frac{X}{n} - p| < 0.01$ mindestens 0.99? □

Die klassische Beweismethode zu Satz 12.6 benutzt:

12.13 Stirling-Formel. Für jedes $n \in \mathbb{N}$ ist

$$(1) \quad n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} e^{\varrho_n} \quad \text{mit} \quad 0 < \varrho_n < \frac{1}{12n}.$$

Beweis. Siehe zum Beispiel Krengel (2005) oder Storch/Wiebe (2003). □

12.14 Aufgabe. Wie viele Stellen hat 1000! im Dezimalsystem? Und welches sind die drei führenden Ziffern?

(An elektronischen Hilfsmitteln dürfen hier nur die Möglichkeiten eines “einfachen Taschenrechners” (mit Logarithmus-Taste) verwendet werden; *nicht* jedoch zum Beispiel <https://www.wolframalpha.com/>.) □

12.15 Lemma. Es gibt eine Konstante $c \in]0, \infty[$ mit

- (1) $|e^x - 1| \leq |x|e^{|x|}$ für $x \in \mathbb{R}$,
- (2) $\left| \log(1+x) - \left(x - \frac{x^2}{2}\right) \right| \leq c|x|^3$ für $|x| \leq \frac{1}{2}$,
- (3) $\left| \frac{x_1 x_2}{x_3} - 1 \right| \leq 2 \cdot (1 \vee |x_1|) \cdot \sum_{i=1}^3 |x_i - 1|$ für $x \in \mathbb{R}^3$ mit $|x_3| \geq \frac{1}{2}$.

Beweis. (1) und (2) sind einfache Taylor-Restglied-Abschätzungen nullten beziehungsweise zweiten Grades.

$$(3): \text{L.S.} = \frac{1}{|x_3|} |x_1 x_2 - x_1 + x_1 - 1 - (x_3 - 1)| \leq 2 \cdot (|x_1| \cdot |x_2 - 1| + |x_1 - 1| + |x_3 - 1|) \leq \text{R.S.} \quad \square$$

Auch wenn es uns hier auf die bestmögliche Konstante c oben in (2), also auf den Wert von

$$(4) \quad c_0 := \sup_{0 < |x| \leq \frac{1}{2}} \left| \frac{R(x)}{x^3} \right| \quad \text{mit} \quad R(x) := \log(1+x) - \left(x - \frac{x^2}{2}\right),$$

eigentlich nicht ankommen soll, sei zwecks Explizitheit hier und im Beweis von Lemma 12.16 bemerkt, dass das Lagrange-Restglied $c_0 \leq \frac{1}{3!} \sup_{0 < |\xi| < \frac{1}{2}} |\log'''(1+\xi)| = \frac{1}{3!} \sup_{0 < |\xi| < \frac{1}{2}} \left| \frac{2}{(1+\xi)^3} \right| = \frac{8}{3}$ liefert, die geschicktere einfache Anwendung des verallgemeinerten Mittelwertsatzes der Differentialrechnung immerhin schon $c_0 \leq \sup_{0 < |\xi| < \frac{1}{2}} \left| \frac{R'(\xi)}{3\xi^2} \right| = \frac{1}{3} \sup_{0 < |\xi| < \frac{1}{2}} \left| \frac{1}{1+\xi} \right| = \frac{2}{3}$, man aber mit der

Integraldarstellung $R(x) = \int_0^x \log'''(1+y) \frac{1}{2!} (x-y)^2 dy = x^3 \int_0^1 \frac{1}{(1+xy)^3} (1-y)^2 dy$ sieht, dass c_0 an $x = -\frac{1}{2}$ angenommen wird, und folglich gilt:

$$(5) \quad c_0 = 8 |R(-\frac{1}{2})| = 8 \log(2) - 5 = 0.5411 \dots$$

Wir beweisen nun zunächst Satz 12.6(a) mit einem Zusatz, der weiter unten im Beweis von Satz 12.6(b) verwendet wird.

12.16 Lemma. *Es gibt Konstanten $A, \alpha, c \in]0, \infty[$ mit der Eigenschaft: Für 12.6(1), also*

$$(1) \quad P \in \mathcal{P}, \quad \mu := \mu(P), \quad \sigma := \sigma(P), \quad k \in \mathbb{Z},$$

sowie mit

$$(2) \quad z := \frac{k - \mu}{\sigma}, \quad \text{also} \quad k = \mu + \sigma z,$$

und

$$(3) \quad \Delta := \text{L.S. 12.6(2)} = \left| P(\{k\}) - \frac{1}{\sigma} \varphi(z) \right|$$

gilt 12.6(2), also

$$(4) \quad \Delta \leq \frac{A}{\sigma^2},$$

sowie die Implikation

$$(5) \quad |z| \leq c\sigma \Rightarrow \Delta \leq \frac{A}{\sigma^2} e^{-\alpha z^2}.$$

Beweis. Wir zeigen, dass die Behauptung in endlich vielen (nämlich fünf), insgesamt alle Möglichkeiten erfassenden Spezialfällen gilt. Dabei wird die gewünschte Konstante c sowie ein die Fallunterscheidung festlegender Parameter σ_0 am Ende der Behandlung des ersten Falls gewählt, während A und α sich von Fall zu Fall ändern können. Die volle Behauptung folgt dann mit dem Minimum der erhaltenen α und dem Maximum der erhaltenen A . Die ersten drei Fälle betreffen nacheinander die drei betrachteten Verteilungstypen unter den Zusatzannahmen (10) und (11), die letzten beiden Fälle sind dann die verbleibenden “trivialen Randfälle”.

Mit c_1, c_2, \dots werden im Verlauf des Beweises absolute Konstanten $\in]0, \infty[$ bezeichnet, d.h. Konstanten, die man im Prinzip schon an dieser Stelle ohne Bezug auf die unten als gegeben betrachteten P oder k wählen könnte, nämlich

$$c_1 \text{ die Konstante } c \text{ bzw. } c_0 \text{ aus 12.15(2,4,5), } c_2 := \left(2c_1 + \frac{1}{2}\right) \vee \frac{5}{8}, \quad c_3 := 2c_2, \\ c_4 := \sup_{t \geq 0} (t^3 + 1) \exp\left(-\frac{t^2}{32}\right), \quad c_5 := c_4 e^{c_3}, \quad \dots$$

und auf deren mögliche oder gar bestmögliche Werte es hier nicht ankommt.

Es werde nun (1-3) angenommen³, und es sei

$$(6) \quad q := \frac{P(\{k\})}{\frac{1}{\sigma}\varphi(z)}.$$

1. Poisson-Fall. Ist $P = P_\lambda$ mit $\lambda \in]0, \infty[$, also $\mu = \sigma^2 = \lambda$ nach Aufgabe 11.13(b), und gilt

$$(7) \quad \sigma \geq 1,$$

$$(8) \quad |z| \leq \frac{\sigma}{2},$$

so folgt mit (2) im ersten und (8) im dritten Schritt zunächst

$$(9) \quad k = \mu + \sigma z \geq \mu - \sigma|z| \geq \mu - \frac{\sigma^2}{2} = \frac{\sigma^2}{2},$$

insbesondere also $k > 0$ und damit $k \in \mathbb{N}$, und daher mit der Notation aus 12.13 im zweiten Schritt

$$P(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} = \frac{\lambda^k e^{-\lambda}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k} e^{\varrho_k}} = \frac{1}{\sqrt{2\pi\lambda}} e^{k-\lambda} \left(\frac{k}{\lambda}\right)^{-(k+\frac{1}{2})} e^{-\varrho_k},$$

und mit (6) im zweiten und $g(x) := \log(1+x) - \left(x - \frac{x^2}{2}\right)$ für $x > -1$ im fünften Schritt dann

$$\begin{aligned} r := \log(q) &= \log\left(\frac{e^{k-\lambda} \left(\frac{k}{\lambda}\right)^{-(k+\frac{1}{2})} e^{-\varrho_k}}{e^{-z^2/2}}\right) = \frac{z^2}{2} + k - \lambda - \left(k + \frac{1}{2}\right) \log \frac{k}{\lambda} - \varrho_k \\ &= \frac{z^2}{2} + z\sigma - \left(\sigma^2 + z\sigma + \frac{1}{2}\right) \log\left(1 + \frac{z}{\sigma}\right) - \varrho_k \\ &= \frac{z^2}{2} + z\sigma - \left(\sigma^2 + z\sigma + \frac{1}{2}\right) \left(\frac{z}{\sigma} - \frac{z^2}{2\sigma^2} + g\left(\frac{z}{\sigma}\right)\right) - \varrho_k \\ &= \frac{z^3}{2\sigma} - \frac{1}{2} \left(\frac{z}{\sigma} - \frac{z^2}{2\sigma^2}\right) - \left(\sigma^2 + z\sigma + \frac{1}{2}\right) g\left(\frac{z}{\sigma}\right) - \varrho_k; \end{aligned}$$

dabei gilt

$$|\varrho_k| \leq \frac{1}{12k} \leq \frac{1}{6\sigma^2} \leq \frac{1}{6\sigma}$$

³Damit sind also P und k im Prinzip für den Rest des Beweises fixiert; dennoch erlauben wir uns in den Beweisteilen 2. und 3. jeweils eine ‘Anwendung des bereits bewiesenen’ auf gewisse dort hilfweise betrachtete P_i und k_i .

wegen 12.13 im ersten Schritt, (9) im mittleren, und (7) im letzten, und es folgt, mit (8), (7), und 12.15(2) im ersten Schritt,

$$\begin{aligned} |r| &\leq \frac{|z|^3}{2\sigma} + \frac{1}{2} \left(\frac{|z|}{\sigma} + \frac{|z|}{4\sigma} \right) + \left(\sigma^2 + \frac{\sigma^2}{2} + \frac{\sigma^2}{2} \right) c_1 \left| \frac{z}{\sigma} \right|^3 + \frac{1}{6\sigma} \\ &\leq \frac{c_2}{\sigma} (|z|^3 + |z| + 1) \\ &\leq \frac{c_3}{\sigma} (|z|^3 + 1) \leq \frac{c_4}{\sigma} \exp\left(\frac{z^2}{32}\right), \end{aligned}$$

und weiter, mit den letzten beiden Schranken für $|r|$ im dritten Schritt, sowie mit 12.15(1) zweiten und (7) im vierten,

$$\begin{aligned} |q - 1| &= |e^r - 1| \leq |r|e^{|r|} \leq \frac{c_4}{\sigma} \exp\left(\frac{z^2}{32}\right) \exp\left(\frac{c_3}{\sigma} (|z|^3 + 1)\right) \\ &\leq \frac{c_5}{\sigma} \exp\left(\frac{z^2}{32} + \frac{c_3}{\sigma} |z|^3\right). \end{aligned}$$

Also gibt es ein $c \in]0, \frac{1}{2}]$ und ein $\sigma_0 \in [1, \infty[$, die für den Rest des Beweises fixiert seien, derart, dass im Fall $P = P_\lambda$ und unter den (7) und (8) verschärfenden Annahmen

$$(10) \quad \sigma \geq \sigma_0,$$

$$(11) \quad |z| \leq c\sigma$$

stets

$$(12) \quad |q - 1| \leq \frac{c_5}{\sigma} \exp\left(\frac{z^2}{16}\right),$$

$$(13) \quad \frac{c_5}{\sigma} \leq \frac{1}{24}$$

gilt, und als Konsequenz von (12) insbesondere

$$(14) \quad \Delta = |q - 1| \frac{1}{\sigma} \varphi(z) \leq \frac{c_6}{\sigma^2} \exp\left(-\frac{7}{16} z^2\right),$$

also die Behauptung im Poisson-Fall mit (10) und (11).

2. Binomialfall. Im Fall $P = B_{n,p}$ mit $n \in \mathbb{N}$, $p \in]0, 1[$, und mit (10) und (11) gilt $\mu = np$, $\sigma^2 = np(1-p)$, und mit $\lambda := n(1-p)$ folgt

$$(15) \quad \mu + \lambda = n, \quad \frac{\mu}{\mu + \lambda} = p, \quad \frac{\mu\lambda}{\mu + \lambda} = \sigma^2, \quad \frac{1}{\mu} + \frac{1}{\lambda} = \frac{1}{\sigma^2},$$

$$(16) \quad \min \left\{ \sqrt{\lambda}, \sqrt{\mu}, \sqrt{n} \right\} \geq \sigma,$$

und mit der aus Aufgabe 9.15 bekannten, aber auch so leicht durch Kürzen verifizierbaren Darstellung

$$P(\{k\}) = b_{n,p}(k) = \frac{p_\mu(k)p_\lambda(n-k)}{p_{\lambda+\mu}(n)}$$

erhalten wir mit (15)

$$(17) \quad q = \frac{P(\{k\})}{\frac{1}{\sigma}\varphi(z)} = \frac{q_1 q_2}{q_3}$$

wobei

$$z_1 := \frac{k - \mu}{\sqrt{\mu}}, \quad z_2 := \frac{n - k - \lambda}{\sqrt{\lambda}} = -\frac{k - \mu}{\sqrt{\lambda}}, \quad z_3 := \frac{n - n}{\sqrt{n}} = 0,$$

$$q_1 := \frac{p_\mu(k)}{\frac{1}{\sqrt{\mu}}\varphi(z_1)}, \quad q_2 := \frac{p_\lambda(n-k)}{\frac{1}{\sqrt{\lambda}}\varphi(z_2)}, \quad q_3 := \frac{p_n(n)}{\frac{1}{\sqrt{n}}\varphi(z_3)};$$

dabei gilt wegen (16) und (11)

$$\begin{aligned} |z_1| &\leq |z| \leq c\sigma \leq c\sqrt{\mu}, \\ |z_2| &\leq |z| \leq c\sigma \leq c\sqrt{\lambda}, \\ |z_3| &= 0 \leq c\sqrt{n}, \end{aligned}$$

und damit nach 1. angewandt auf $P_\mu, k, P_\lambda, n-k, P_n, n$

$$\begin{aligned} |q_1 - 1| &\leq \frac{c_5}{\sqrt{\mu}} \exp\left(\frac{z_1^2}{16}\right) \leq \frac{c_5}{\sigma} \exp\left(\frac{z^2}{16}\right), \quad |q_2 - 1| \leq \frac{c_5}{\sigma} \exp\left(\frac{z^2}{16}\right), \\ |q_3 - 1| &\leq \frac{c_5}{\sqrt{n}} \exp\left(\frac{0^2}{16}\right) \leq \frac{c_5}{\sigma} \leq \frac{1}{24} \wedge \left(\frac{c_5}{\sigma} \exp\left(\frac{z^2}{16}\right)\right) \end{aligned}$$

und damit sicher $|q_3| \geq \frac{1}{2}$ und folglich, mit 12.15(3) im zweiten Schritt,

$$\begin{aligned} |q - 1| &= \left| \frac{q_1 q_2}{q_3} - 1 \right| \leq 2 \cdot (1 \vee |q_1|) \cdot \sum_{i=1}^3 |q_i - 1| \\ &\leq 2 \cdot (1 + |q_1 - 1|) \cdot \sum_{i=1}^3 |q_i - 1| \\ &\leq 2 \cdot \left(1 + \frac{c_5}{\sigma} \exp\left(\frac{z^2}{16}\right)\right) \frac{3c_5}{\sigma} \exp\left(\frac{z^2}{16}\right) \\ &\leq \frac{12c_5}{\sigma} \exp\left(\frac{z^2}{8}\right). \end{aligned}$$

Also gilt mit $c_7 := 12c_5$ im Fall $P = B_{n,p}$ mit (10) und (11) stets

$$(18) \quad |q - 1| \leq \frac{c_7}{\sigma} \exp\left(\frac{z^2}{8}\right),$$

$$(19) \quad \frac{c_7}{\sigma} \leq \frac{1}{2}$$

und, als Konsequenz von (18) insbesondere

$$(20) \quad \Delta = |q - 1| \frac{1}{\sigma} \varphi(z) \leq \frac{c_8}{\sigma^2} \exp\left(-\frac{3}{8}z^2\right),$$

also die Behauptung im hier betrachteten Binomialfall mit (10) und (11).

3. Hypergeometrischer Fall. Im Fall $P = H_{n,r,b}$ mit $n, r, b \in \mathbb{N}$ mit $n < r + b$, und mit (10) und (11) gilt mit $N := r + b$ und $p := \frac{r}{N}$ nach der aus Aufgabe 8.3 bekannten und wieder durch Kürzen leicht verifizierbaren Darstellung

$$P(\{k\}) = h_{n,r,b}(k) = \frac{b_{r,p}(k)b_{b,p}(n-k)}{b_{N,p}(n)}$$

wieder (17), aber hier mit

$$\begin{aligned} z_1 &:= \frac{k - rp}{\sqrt{rp(1-p)}}, & z_2 &:= \frac{n - k - bp}{\sqrt{bp(1-p)}}, & z_3 &:= \frac{n - Np}{\sqrt{Np(1-p)}} = 0, \\ q_1 &:= \frac{b_{r,p}(k)}{\frac{1}{\sqrt{rp(1-p)}}\varphi(z_1)}, & q_2 &:= \frac{b_{b,p}(k)}{\frac{1}{\sqrt{bp(1-p)}}\varphi(z_2)}, & q_3 &:= \frac{b_{N,p}(n)}{\frac{1}{\sqrt{Np(1-p)}}\varphi(z_3)} \end{aligned}$$

...

4. Im Fall $\sigma \leq \sigma_0$ gilt, diese Voraussetzung im vorletzten Schritt ausnutzend,

$$\Delta \leq P(\{k\}) + \frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right) \leq 1 + \frac{\varphi(0)}{\sigma} \leq \frac{\sigma_0^2 + \frac{\sigma_0}{\sqrt{2\pi}}}{\sigma^2} = \frac{c_{12}}{\sigma^2}$$

und damit auch die Implikation

$$|z| \leq \sigma \Rightarrow \Delta \leq \frac{c_{12}}{\sigma^2} e^{z^2} e^{-z^2} \leq \frac{c_{12} e^{\sigma_0^2}}{\sigma^2} e^{-z^2};$$

also kann in diesem Fall $(A, \alpha) = (c_{12} e^{\sigma_0^2}, 1)$ gewählt werden.

5. Im Fall $|z| > c\sigma$ erhalten wir einerseits mit Chebyshev im zweiten Schritt

$$P(\{k\}) \leq P\left(\left\{k' : \left|\frac{k' - \mu}{\sigma}\right| \geq |z|\right\}\right) \leq \frac{1}{z^2} \leq \frac{c^{-2}}{\sigma^2}$$

und andererseits

$$\frac{1}{\sigma} \varphi\left(\frac{k - \mu}{\sigma}\right) = \frac{1}{\sigma|z|} |z| \varphi(z) \leq \frac{c_{13}}{\sigma|z|} \leq \frac{c^{-1} c_{13}}{\sigma^2},$$

zusammen also $\Delta \leq (c^{-2} + c^{-1} c_{13})/\sigma^2$; also kann in diesem Fall, (A, α) mit $A := c^{-2} + c^{-1} c_{13}$ und α beliebig, etwa $\alpha := 1$, gewählt werden. \square

Beweis zu Satz 12.6(b). Wir schreiben wieder $c_1, c_2 \dots$ für (neue) absolute Konstanten $\in]0, \infty[$.

1. Für $(k_1, k_2) \in \mathbb{Z}_{\leq}^2$ gilt für jede stetig differenzierbare Funktion $f : [k_1, k_2] \rightarrow \mathbb{R}$

$$\sum_{k=k_1}^{k_2} f(k) = \int_{k_1}^{k_2} f(x) dx + \frac{f(k_1) + f(k_2)}{2} + \int_{k_1}^{k_2} \left(x - [x] - \frac{1}{2}\right) f'(x) dx$$

(Euler-Maclaurin-Summenformel im einfachsten Fall, siehe zum Beispiel Heuser (2009, Nr. 95), Beweis durch partielle Integration, etwa $\int_k^{k+1} 1 \cdot f(x) dx = (x - k + \frac{1}{2})f(x) \Big|_k^{k+1} - \int_k^{k+1} (x - [x] + \frac{1}{2})f'(x) dx$, dann $\sum_{k=k_1}^{k_2-1} \dots$), und damit

$$(1) \quad \left| \sum_{k=k_1}^{k_2} f(k) - \int_{k_1}^{k_2} f(x) dx \right| \leq \frac{1}{2} \int_{k_1}^{k_2} |f'(x)| dx + \frac{|f(k_1)| + |f(k_2)|}{2},$$

speziell also für $\mu \in \mathbb{R}$ und $\sigma \in]0, \infty[$, und dann ab jetzt die Abkürzung

$$(2) \quad \tilde{x} := \frac{x - \mu}{\sigma} \quad \text{für } x \in \mathbb{R}$$

verwendend,

$$(3) \quad \left| \sum_{k=k_1}^{k_2} \frac{1}{\sigma} \varphi(\tilde{k}) - \int_{k_1}^{k_2} \frac{1}{\sigma} \varphi(\tilde{x}) dx \right| \leq \frac{1}{2} \int_{-\infty}^{\infty} \left| \frac{1}{\sigma^2} \tilde{x} \varphi(\tilde{x}) \right| dx + \frac{1}{\sigma \sqrt{2\pi}} = \frac{c_1}{\sigma}$$

mit $\varphi'(x) = -x\varphi(x)$ und $\tilde{x}' = \frac{1}{\sigma}$ im ersten Schritt, und mit der Substitution $x \mapsto \sigma x + \mu$ im zweiten, und mit derselben Substitution

$$(4) \quad \int_{k_1}^{k_2} \frac{1}{\sigma} \varphi(\tilde{x}) dx = \Phi(\tilde{k}_2) - \Phi(\tilde{k}_1).$$

2. Der Mittelwertsatz und $\|\varphi\|_{\infty} = \varphi(0) = \frac{1}{\sqrt{2\pi}}$ liefern, mit der Notation (2),

$$(5) \quad |\Phi(\tilde{x}) - \Phi(\tilde{y})| \leq \frac{|x - y|}{\sqrt{2\pi}\sigma} \quad \text{für } x, y, \mu \in \mathbb{R} \text{ und } \sigma \in]0, \infty[.$$

3. Es gelte nun 12.6(3), also

$$P \in \mathcal{P}, \quad \mu := \mu(P), \quad \sigma := \sigma(P), \\ I \subseteq \mathbb{R} \text{ Intervall mit den Randpunkten } a, b \in \overline{\mathbb{R}} \text{ mit } a \leq b.$$

Ist $\sigma \leq 1$, so folgt

$$(6) \quad \text{L.S.12.6(4)} \leq 1 \leq \frac{1}{\sigma} = \frac{c_2}{\sigma};$$

wir können daher ab jetzt bei Bedarf

$$(7) \quad \sigma \geq 1$$

annehmen. Ist nun weiter $I \cap \mathbb{Z} = \emptyset$, so gilt $P(I) = 0$ und $0 \leq b - a \leq 1$, mit (5) also

$$(8) \quad \text{L.S.12.6(4)} = |\Phi(\tilde{b}) - \Phi(\tilde{a})| \leq \frac{1}{\sqrt{2\pi}\sigma} = \frac{c_3}{\sigma}.$$

Ist dagegen $I \cap \mathbb{Z} \neq \emptyset$, so ist

$$k_1 := \lceil a \rceil \leq \lfloor b \rfloor =: k_2,$$

und wir erhalten mit A, α, c aus Lemma 12.16 unter der Zusatzannahme

$$(9) \quad -c\sigma \leq \tilde{a}, \quad \tilde{b} \leq c\sigma$$

auch $|\tilde{k}| \leq c\sigma$ für $k \in \{k_1, \dots, k_2\} = I \cap \mathbb{Z}$, und folglich

$$(10) \quad \left| P(I) - \sum_{k=k_1}^{k_2} \frac{1}{\sigma} \varphi(\tilde{k}) \right| \leq \sum_{k=k_1}^{k_2} \left| P(\{k\}) - \frac{1}{\sigma} \varphi(\tilde{k}) \right| \\ \leq \sum_{k=k_1}^{k_2} \frac{A}{\sigma^2} \exp(-\alpha \tilde{k}^2) \leq \frac{c_4}{\sigma},$$

im zweiten Schritt mit Lemma 12.16, und im letzten Schritt mit (1), ähnlich wie bei (3), aber auch (7) benutzend, und damit

$$(11) \quad \text{L.S.12.6(4)} = \left| P(I) - (\Phi(\tilde{b}) - \Phi(\tilde{a})) \right| \\ \leq \text{L.S.(10)} + \text{L.S.(3)} + \left| \Phi(\tilde{k}_2) - \Phi(\tilde{k}_1) - (\Phi(\tilde{b}) - \Phi(\tilde{a})) \right| \leq \frac{c_5}{\sigma}$$

wegen (4) im zweiten Schritt, und mit (10,3,5) und $|k_2 - b| \leq 1$, $|k_1 - a| \leq 1$ im letzten Schritt.

4. Ist $x \in \mathbb{R}$ mit $\tilde{x} \leq -c\sigma$, so folgt

$$(12) \quad \Phi(\tilde{x}) \leq \Phi(-c\sigma) = \frac{1}{c\sigma} c\sigma \Phi(-c\sigma) \leq \frac{1}{c\sigma} \sup_{z \geq c} z \Phi(-z) \leq \frac{\varphi(c)}{c\sigma} = \frac{c_6}{\sigma}$$

mit (7) im dritten Schritt und 12.4(5) im vierten. Ist $x \in \mathbb{R}$ mit $\tilde{x} \geq c\sigma$, so liefert (12)

$$(13) \quad 1 - \Phi(\tilde{x}) = \Phi(-\tilde{x}) \leq \frac{c_6}{\sigma}.$$

5. Gilt nun 12.6(3) ohne dass (9) angenommen wird, so betrachten wir mit $I_0 := [a_0, b_0] := [\mu - c\sigma^2, \mu + c\sigma^2]$ das (möglicherweise leere) Intervall $J := I \cap I_0$. In jedem Fall gilt

$$(14) \quad 0 \leq P(I) - P(J) \leq P(\mathbb{R} \setminus I_0) \leq \frac{\sigma^2}{(c\sigma^2)^2} \leq \frac{c_7}{\sigma}$$

wegen der Isotonie und der Subtraktivität 3.3(8,9) in den ersten beiden Schritten, der Chebyshev-Ungleichung 10.31(1) im dritten Schritt, und wegen (7) im letzten.

Im Fall $J = \emptyset$ folgt $P(I) \leq \frac{c_7}{\sigma}$ und entweder $\tilde{a} \leq \tilde{b} \leq -c\sigma^2$ oder $c\sigma^2 \leq \tilde{a} \leq \tilde{b}$, und im ersten dieser beiden Unterfälle erhalten wir

$$(15) \quad \text{L.S.12.6(4)} \leq P(I) + \Phi(\tilde{b}) + \Phi(\tilde{a}) \leq \frac{c_8}{\sigma}$$

wegen(12), im zweiten

$$(16) \quad \text{L.S.12.6(4)} = \left| P(I) - \left(1 - \Phi(\tilde{b})\right) + \left(1 - \Phi(\tilde{a})\right) \right| \leq \frac{c_8}{\sigma}$$

wegen(13).

Im Fall $J \neq \emptyset$ hat J die Randpunkte $a \vee a_0 =: c \leq d := b \wedge b_0$, und mit

$$0 \leq \Psi := \Phi(\tilde{c}) - \Phi(\tilde{a}) + \Phi(\tilde{b}) - \Phi(\tilde{d}) \leq \Phi(\tilde{a}_0) + 1 - \Phi(\tilde{b}_0) \leq \frac{c_9}{\sigma}$$

und mit (14) folgt

$$(17) \quad \text{L.S.12.6(4)} \leq \left| P(J) - \left(\Phi(\tilde{d}) - \Phi(\tilde{c})\right) \right| + |P(I) - P(J)| + |\Psi| \leq \frac{c_{10}}{\sigma}$$

unter Benutzung von (11) mit J, c, d statt I, a, b .

6. Damit ist 12.6(4) in jedem Fall mit $A := \max\{c_2, c_3, c_5, c_8, c_{10}\}$ bewiesen. \square

12.17 Aufgabe. Führen Sie die im obigen Beweis zu Satz 12.6(b) angedeutete Begründung des letzten Schrittes in der dortigen Ungleichungskette (10) genauer aus. \square

Satz 12.6(b) kann als Spezialfall des folgenden allgemeinen Satzes (hier im Rahmen elementarer Wahrscheinlichkeitsräume zu lesen, aber wörtlich so auch im allgemeinen maßtheoretischen Rahmen gültig) aufgefasst werden:

12.18 Satz (ZGWS, Berry-Esseen-Version mit den Konstanten von Shevtsova (2013)).
Es sei $n \in \mathbb{N}$, es seien X_1, \dots, X_n unabhängige reellwertige Zufallsgrößen mit

$$\mu_i := \mathbb{E}X_i \in \mathbb{R}, \quad \sigma_i^2 := \text{Var}X_i < \infty, \quad \beta_i := \mathbb{E}|X_i - \mu_i|^3 < \infty$$

für $i \in \{1, \dots, n\}$, und es sei

$$\mu := \sum_{i=1}^n \mu_i, \quad \sigma^2 := \sum_{i=1}^n \sigma_i^2 > 0.$$

Weiter sei $I \subseteq \mathbb{R}$ ein Intervall mit den Randpunkten $a, b \in \overline{\mathbb{R}}$ mit $a \leq b$. Dann gilt

$$(1) \quad \left| \mathbb{P}\left(\sum_{i=1}^n X_i \in I\right) - \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right) \right| \leq 1.116 \frac{\sum_{i=1}^n \beta_i}{\sigma^3}.$$

Im Spezialfall identisch verteilter X_i kann die Konstante 1.116 durch 0.938 ersetzt werden, d.h. es gilt dann

$$(2) \quad \text{L.S.}(1) \leq \frac{0.938}{\sqrt{n}} \cdot \frac{\beta_1}{\sigma_1^3}.$$

Im Fall $a = -\infty$ oder $b = \infty$ können die Konstanten in (1) und (2) jeweils halbiert werden, also 1.116 durch 0.5583 sowie 0.938 durch 0.469 ersetzt werden.

Beweis. Übersichtliche Beweise für einseitig unbeschränkte Intervalle, jedoch mit schlechteren Konstanten an Stelle von 0.5583, findet man zum Beispiel bei Feller (1971, Kapitel XVI) und Gänsler/Stute (1977, pp. 162–167). Der Fall beliebiger Intervalle mit verdoppelter Konstante folgt daraus leicht mittels Dreiecksungleichung. Der schwerer nachzuvollziehende Beweis mit den angegebenen Konstanten von Shevtsova (2013) kombiniert, wie schon einige Vorläufer, langwierige analytische Überlegungen mit computergestützten numerischen Rechnungen. \square

12.19 Aufgabe. Ungleichung 12.18(2) für $X_i \sim B_p$ liefert 12.6(4) mit $A = 0.938$ im Binomialfall. \square

12.20 Bemerkung. (a) Es gibt Indizien dafür, dass in Satz 12.18 die Konstante 0.469 (und vielleicht sogar 0.5583) durch $(3 + \sqrt{10})/(6\sqrt{2\pi}) = 0.4097\dots$ ersetzt werden kann, und 0.938 (und vielleicht sogar 1.1166) durch $\sqrt{2/\pi} = 0.7978\dots$

Jedenfalls weiß man, dass es nicht besser geht, und die naheliegende Technik zum Erhalt unterer Schranken für die besten Konstanten liefert keine weitere Information, siehe Dinev/Mattner (2012). Außerdem konnte Schulz (2016) zeigen, dass im als ungünstigst vermuteten Fall Bernoulli-verteilter X_i tatsächlich 0.469 durch $(3 + \sqrt{10})/(6\sqrt{2\pi}) = 0.4097\dots$ ersetzt werden kann, und damit immerhin auch 0.938 durch $2(3 + \sqrt{10})/(6\sqrt{2\pi}) = 0.8194\dots$

(b) Man kann zeigen⁴: Zu jeder hypergeometrischen Verteilung $H_{n,r,b}$ gibt es Zahlen $p_1, \dots, p_n \in [0, 1]$ mit $H_{n,r,b} = \ast_{i=1}^n B_{p_i}$. Damit liefert eine Betrachtung analog zu Aufgabe 12.19, mit 12.18(1) statt 12.18(2), einen Beweis von 12.6(4) mit $A = 1.1166$ auch im hypergeometrischen Fall. \square

12.21 Bemerkung. Satz 12.18, ob in der hier gewählten Formulierung oder einer anderen, ist der wichtigste unter vielen Sätzen welche Anlass zur Betrachtung allgemeiner, nicht notwendig elementarer Wahrscheinlichkeitsmaße geben: Das sind Funktionen P auf einer (Mengen-) σ -Algebra \mathcal{B} auf einer Grundmenge Ω , für welche die Forderungen 3.2(1,2) abgeschwächt werden zu

- (1) $P(A) \in [0, \infty]$ für $A \in \mathcal{B}$ (**Positivität**),
- (2) $\sum_{A \in \mathcal{A}} P(A) = 1$ für jede abzählbare Quasipartition $\mathcal{A} \subseteq \mathcal{B}$ von Ω
(**σ -Additivität und Normiertheit**).

Die bisher betrachteten elementaren Wahrscheinlichkeitsmaße bilden dann einen Sonderfall unter denjenigen Wahrscheinlichkeitsmaßen mit $\mathcal{B} = 2^\Omega$.

Speziell mit $\Omega = \mathbb{R}$ und \mathcal{B} gleich der Borel- σ -Algebra auf \mathbb{R} definiert dann, wie man in der MIT zeigt,

$$N_{\mu, \sigma^2}(A) := \int_A \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) dx \quad \text{für } A \in \mathcal{B}$$

ein Wahrscheinlichkeitsmaß N_{μ, σ^2} auf \mathcal{B} , genannt **Normalverteilung** mit den Parametern μ und σ^2 , und für jedes Intervall $I \subseteq \mathbb{R}$ mit den Randpunkten $a \leq b$ gilt

$$N_{\mu, \sigma^2}(I) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Satz 12.18 macht also eine Aussage zur Approximation der Verteilung einer Summe durch die nichtelementare Verteilung N_{μ, σ^2} , wobei dort nur Intervalle und damit nur spezielle Borel-Mengen betrachtet werden.

Eine systematische Betrachtung allgemeiner, nicht notwendig elementarer Verteilungen erfolgt in der MIT und der WT. \square

⁴Vatutin und Mikhailov (1983, Corollary 5 mit $n = 2$, also F_2 erzeugende Funktion von $H_{s_1, s_2, N-s_2} = H_{s_2, s_1, N-s_1}$), ausführlicher bei Roos (1996, pp. 8–10).

13 Statistische Modelle, Schätzprobleme, Tests

Statistische Modelle sind Wahrscheinlichkeitsmodelle, welche noch von einem als unbekannt angesehenen Parameter abhängen. Formale Definition in zwei Varianten:

13.1 Definition. Es seien \mathcal{X} und Θ nichtleere Mengen. In jedem der beiden folgenden Fälle heißt \mathcal{X} **Stichprobenraum**, und Θ **Parameterraum**.

(a) Für $\vartheta \in \Theta$ sei $P_\vartheta \in \text{Prob}(\mathcal{X})$. Dann heißt

$$(1) \quad P_\bullet := (P_\vartheta : \vartheta \in \Theta)$$

kanonisches statistisches Modell auf \mathcal{X} , mit der Beobachtungsgröße

$$(2) \quad X := \text{id}_{\mathcal{X}}.$$

(b) Für $\vartheta \in \Theta$ sei $(\Omega_\vartheta, \mathbb{P}_\vartheta)$ ein Wahrscheinlichkeitstraum und $X_\vartheta : \Omega_\vartheta \rightarrow \mathcal{X}$ eine Zufallsgröße. Dann heißt

$$(3) \quad (\mathcal{X}, ((\Omega_\vartheta, \mathbb{P}_\vartheta, X_\vartheta) : \vartheta \in \Theta))$$

unkanonisches statistisches Modell, mit den Beobachtungsgrößen X_ϑ für $\vartheta \in \Theta$, und dem zugehörigen kanonischen Modell

$$(4) \quad (X_\vartheta \square \mathbb{P}_\vartheta : \vartheta \in \Theta).$$

□

13.2 Beispiel. Es sei $n \in \mathbb{N}$. Für $N \in \mathbb{N}$ mit $N \geq n$ und für $r \in \{0, \dots, N\}$ sei, Beispiel 5.15 bezüglich der dort noch gegebenen Wahlmöglichkeiten konkretisierend,

$$\begin{aligned} M_N &:= \{1, \dots, N\}, & R_r &:= \{1, \dots, r\}, & \mathcal{S}_N &:= \{s \subseteq M_N : \#s = n\}, \\ \mathbb{P}_N &:= \text{U}_{\mathcal{S}_N}, & \mathcal{S}_N &:= \text{id}_{\mathcal{S}_N}, & X_{N,r} &:= \#(S_N \cap R_r), \\ \mathcal{X} &:= \{0, \dots, n\}, & \Theta &:= \{(r, N) \in \mathbb{N}_0^2 : r \wedge n \leq N\}. \end{aligned}$$

Dann ist

$$(1) \quad \left(\mathcal{X}, \left((\mathcal{S}_N, \mathbb{P}_N, X_{N,r}) : (r, N) \in \Theta \right) \right)$$

ein unkanonisches statistisches Modell, für die Anzahl der gezogenen roten Kugeln in einer einfachen Zufallsstichprobe vom Umfang n aus einer Urne unbekannter Größe N mit einer unbekanntem Anzahl r roter Kugeln. Das zugehörige kanonische Modell ist

$$(2) \quad \left(H_{n,r,N-r} : (r, N) \in \Theta \right),$$

wie in Beispiel 5.15 gezeigt beziehungsweise definiert. □

Kanonische Modelle reichen im Prinzip für die Statistik vollkommen aus, aber unkanonische sind zur Bildung kanonischer oft hilfreich. Ab jetzt betrachten wir hier nur noch kanonische statistische Modelle, kurz **Modelle** genannt. Statt “Modell” sagt man auch “**Experiment**” oder “**Verteilungsannahme**”.

Die Parametrisierung eines Modells ist meist nicht frei von Willkür, siehe etwa 13.2(2) versus 13.3(c).

13.3 Beispiele. Es sei $n \in \mathbb{N}$. Kanonische Modelle auf $\mathcal{X} := \{0, \dots, n\}$ sind:

(a) $B_{n,\bullet} = (B_{n,p} : p \in [0, 1])$, das **Binomialmodell** (zum bekannten Stichprobenumfangsparameter n).

(b) Für $N \in \mathbb{N}$ und $N \geq n$ das **hypergeometrische Modell** ($H_{n,r,N-r} : r \in \{0, \dots, N\}$) zur bekannten Populationsgröße N (und zum bekannten Stichprobenumfangsparameter n).

(c) Das **hypergeometrische Modell** ($H_{n,r,b} : r, b \in \mathbb{N}_0$ mit $r + b \geq n$), zu unbekannter Populationsgröße (und zum bekannten Stichprobenumfangsparameter n). \square

13.4 Definition. Es sei $P_\bullet = (P_\vartheta : \vartheta \in \Theta)$ ein Modell auf \mathcal{X} , und es sei $\kappa : \Theta \rightarrow \Gamma$ eine Funktion. Dann heißt (P_\bullet, κ) **Schätzproblem** mit dem **interessierenden Parameter** κ .

Ist $\hat{\kappa} : \mathcal{X} \rightarrow \Gamma$ eine Zufallsgröße mit $\Gamma \supseteq \kappa[\Theta]$, so heißt $\hat{\kappa}$ **Schätzer** für (P_\bullet, κ) , mit der **Verteilung** $(\hat{\kappa} \square P_\vartheta : \vartheta \in \Theta)$.

Ist $\beta \in [0, 1]$ und ist $K : \mathcal{X} \rightarrow 2^{\kappa[\Theta]}$ eine Zufallsgröße mit

$$P_\vartheta(K \ni \kappa(\vartheta)) \geq \beta \quad \text{für } \vartheta \in \Theta,$$

so heißt K **Konfidenzbereich** zum **(Konfidenz)niveau** β , kurz **β -Konfidenzbereich**, für (P_\bullet, κ) . \square

Üblich, aber nicht begründbar, ist die Wahl von $\beta := 0.95$ als Konfidenzniveau.

Interpretation: Von den β -Konfidenzbereichen die ein Statistiker in seinem Leben so ausrechnet enthalten mindestens $\beta \cdot 100\%$ den unbekanntem wahren Parameter. Jeder einzelnen der vom Statistiker getroffenen Aussage der Form “ $K(x) \ni \kappa(\vartheta)$ ” kommt dagegen keine Wahrscheinlichkeit zu, auch wenn dabei Formulierungen wie zum Beispiel “zu 95% sicher” gewählt werden mögen.

13.5 Beispiele. Interessierende Parameter und zugehörige plausible Schätzer in Beispiel 13.3 sind, mit $X := \text{id}_{\{0, \dots, n\}}$:

(a) $\text{id}_{[0,1]}$, $\hat{p} := \frac{1}{n}X$.

(b) $\text{id}_{\{0, \dots, N\}}$, $\hat{r} := \frac{N}{n}X$ (**Hochrechnung** von der Stichprobe auf die Population).

(c) $(r, b) \mapsto \frac{r}{r+b} =: p(r, b)$, $\hat{p} := \frac{1}{n}X$. \square

¹Oft ist dabei sogar $\Gamma = \kappa[\Theta]$ wählbar.

13.6 Beispiel (BSE² im Jahre 1998). Siehe Lorenzen (1998). Dort lesen wir:

“So trat ...kein einziger BSE-Fall bei den insgesamt 234 Nachkommen aller 126 britischen Fleischrind-Zuchtkühe auf, die stets im gleichen Betrieb lebten und schließlich an BSE erkrankten. ...Das Risiko, daß dieser Erreger von der Kuh auf ihr Kalb übertragen wird, ist somit gleich Null.”

Dies kann wie folgt interpretiert werden: In einer Situation wie in Beispiel 13.5(a) mit $n = 234$ ergibt die zu X gehörige reale Beobachtung den Wert 0, kurz “ $X = 0$ ”, und entsprechend “ $\hat{p} = 0$ ”, und daraus wird nun auf “ $p = 0$ ” geschlossen.

Letzterer Schluss, in dieser Schärfe von einem Schätzwert auf einen exakten Parameterwert schließen wollend und nicht etwa eine Formulierung wie “ist somit ungefähr Null” benutzend, ist offenkundig als absurd zu bewerten (ohne damit die Seriösität des Restes des Artikels in Zweifel ziehen zu wollen). \triangleright

Schätzer sind für Mononumeralisten³, der Gebildete betrachtet Konfidenzbereiche. Gewisse Konfidenzbereiche sind uns schon früher begegnet:

13.7 Beispiel. Es sei $n \in \mathbb{N}$, $\beta \in]0, 1[$, und $\alpha := 1 - \beta$. Im Schätzproblem $(B_{n,\bullet}, \text{id}_{[0,1]})$ auf $\mathcal{X} := \{0, \dots, n\}$ ist ein β -Konfidenzbereich

$$(1) \quad \mathcal{X} \ni x \mapsto \left\{ p \in [0, 1] : \left| \frac{x}{n} - p \right| < \frac{1}{\sqrt{4\alpha n}} \right\},$$

aufgrund der Chebyshev-Ungleichung, siehe Beispiel 12.1(a)(Auflösung von (4) mit “=” nach c), und speziell im Fall $n = 2500$ und $\beta = 0.95$ ist nach Aufgabe 12.11 auch

$$(2) \quad \mathcal{X} \ni x \mapsto \left\{ p \in [0, 1] : \left| \frac{x}{n} - p \right| < 0.02 \right\}$$

einer⁴, mit $n = 2401$ statt $n = 2500$ dagegen nicht ganz. \square

Ein Nachteil von (1) und den auf beliebige n verallgemeinerten Fassungen von (2) ist die Konstanz der Länge des jeweiligen Konfidenzintervalls, denn für $\frac{x}{n}$ nahe 0 oder 1 wünscht man sich doch ein kürzeres Intervall. Dieser Mangel kann bezüglich (2) behoben werden, indem man die bei der Herleitung verwendete Abschätzung $p(1-p) \leq \frac{1}{4}$ nicht vornimmt, sondern stattdessen die dann vorliegende quadratische Ungleichung in p auflöst, siehe etwa Henze (2017, Abschnitt 29.7). Da dabei der Zentrale Grenzwertsatzes für Binomialverteilungen naiv “angewendet” wird und man also nicht genau weiß ob das Ergebnis zumindest ungefähr korrekt ist, verwendet man in der Praxis vorzugsweise eines der folgenden Konfidenzintervalle 13.8(4,5,6).

²Bovine spongiforme Enzephalopathie, umgangssprachlich Rinderwahn genannt.

³Ein **Mononumeralist** ist jemand, der mit *einer* Zahl im Grunde schon überfordert ist, und der sich daher auf zwei oder gar drei Zahlen zugleich gar nicht einlassen mag.

⁴Dies wurde in der genannten Aufgabe nicht wirklich bewiesen, sondern nur numerisch.

13.8 Satz (Konfidenzintervalle in Binomialmodellen, Clopper/Pearson). *Es sei $n \in \mathbb{N}$ und $\mathcal{X} := \{0, \dots, n\}$. Für $\beta \in [0, 1]$ und $x \in \mathcal{X}$ existieren dann*

$$\underline{p}_\beta(x) := \begin{cases} 0 & \text{falls } x = 0, \\ \text{das } p \in [0, 1] \text{ mit } B_{n,p}(\{x, \dots, n\}) = 1 - \beta & \text{falls } x > 0 \end{cases}$$

und

$$\bar{p}_\beta(x) := \begin{cases} \text{das } p \in [0, 1] \text{ mit } B_{n,p}(\{0, \dots, x\}) = 1 - \beta & \text{falls } x < n, \\ 1 & \text{falls } x = n, \end{cases}$$

und damit gilt insbesondere

$$\begin{aligned} (1) \quad & \bar{p}_\beta(x) = 1 - \underline{p}_\beta(n - x), \\ (2) \quad & \underline{p}_\beta(1) = 1 - \beta^{1/n}, \quad \underline{p}_\beta(n) = (1 - \beta)^{1/n}, \\ (3) \quad & \bar{p}_\beta(0) = 1 - (1 - \beta)^{1/n}, \quad \bar{p}_\beta(n - 1) = \beta^{1/n}. \end{aligned}$$

Mit den obigen Definitionen und für⁵ $\beta \in]0, 1]$ sind drei β -Konfidenzintervalle für das Schätzproblem $((B_{n,p} : p \in [0, 1]), \text{id}_{[0,1]})$ gegeben durch

$$(4) \quad x \mapsto \begin{cases} [\underline{p}_\beta(0), 1] = [0, 1] & \text{falls } x = 0, \\]\underline{p}_\beta(x), 1] & \text{falls } x > 0, \end{cases}$$

$$(5) \quad x \mapsto \begin{cases} [0, \bar{p}_\beta(x)[& \text{falls } x < n, \\ [0, \bar{p}_\beta(n)] = [0, 1] & \text{falls } x = n, \end{cases}$$

$$(6) \quad x \mapsto \begin{cases} \left[\frac{\underline{p}_{1+\beta}}{2}(0), \bar{p}_{1+\beta}(0) \right] = [0, \bar{p}_{1+\beta}(0)[& \text{falls } x = 0, \\ \left] \frac{\underline{p}_{1+\beta}}{2}(x), \bar{p}_{1+\beta}(x) \right[& \text{falls } 0 < x < n, \\ \left[\frac{\underline{p}_{1+\beta}}{2}(n), \bar{p}_{1+\beta}(n) \right] = \left] \frac{\underline{p}_{1+\beta}}{2}(n), 1 \right[& \text{falls } x = n. \end{cases}$$

Für $\beta \in]0, 1[$ sind \underline{p}_β und \bar{p}_β streng wachsende Funktionen auf \mathcal{X} . Für $x \in \{1, \dots, n\}$ ist $\beta \mapsto \underline{p}_\beta(x)$ streng fallend. Für $x \in \{0, \dots, n - 1\}$ ist $\beta \mapsto \bar{p}_\beta(x)$ streng wachsend.

Beweis. 1. Für $x \in \mathcal{X}$ und $p \in [0, 1]$ zeigt eine einfache Rechnung

$$(7) \quad \partial_p b_{n,p}(x) = -n (b_{n-1,p}(x) - b_{n-1,p}(x - 1))$$

und damit per Teleskopeffekt

$$(8) \quad \partial_p B_{n,p}(\{x, \dots, n\}) = \sum_{k=x}^n \partial_p b_{n,p}(k) = n b_{n-1,p}(x - 1).$$

⁵Wir schließen hier genau den praktisch ohnehin uninteressanten Fall $\beta = 0$ aus, weil genau in den verbleibenden Fällen die hier unter (3) und (4) angegebenen Intervalle jeweils in einem bestimmten Sinne optimal sind.

Ist nun speziell $x \in \{1, \dots, n\}$, so folgt aus (8) mit $b_{n,p}(x-1) > 0$ für $p \in]0, 1[$, dass

$$[0, 1] \ni p \mapsto B_{n,p}(\{x, \dots, n\}) =: h_x(p) \in [0, 1]$$

eine stetige isotone Bijektion ist, denn es gilt $h_x(0) = 0$ and $h_x(1) = 1$. Für $\beta \in [0, 1]$ existiert also \underline{p}_β wie behauptet, es gilt

$$\underline{p}_\beta(x) = h_x^{-1}(1 - \beta) \quad \text{falls } x \in \{1, \dots, n\},$$

und es ergeben sich damit die behaupteten Monotonien von $\underline{p}_\beta(x)$, einerseits in β da h_x^{-1} als Umkehrfunktion einer isotonen Funktion isoton ist, und andererseits in x wegen der für $p \in [0, 1]$ gültigen Implikationskette

$$x \leq y \Rightarrow h_x(p) \geq h_y(p) \Rightarrow h_y^{-1}(h_x(p)) \geq p$$

angewandt auf $p := h_x^{-1}(1 - \beta)$.

Es bezeichne nun $K : \mathcal{X} \rightarrow 2^{[0,1]}$ die Funktion aus (4). Für $\beta \in]0, 1]$ und $p \in [0, 1]$ gilt, wegen $\beta > 0$ im dritten Schritt,

$$\begin{aligned} \{x \in \mathcal{X} : K(x) \ni p\} &= \{x \in \mathcal{X} : x = 0, \text{ oder } x \geq 1 \text{ und } h_x^{-1}(1 - \beta) < p\} \\ &= \{x \in \mathcal{X} : x = 0, \text{ oder } x \geq 1 \text{ und } 1 - \beta < h_x(p)\} \\ &= \{x \in \mathcal{X} : 1 - \beta < B_{n,p}(\{x, \dots, n\})\} \\ &= \{x \in \mathcal{X} : B_{n,p}(\{0, \dots, x-1\}) < \beta\} \\ &=: M, \end{aligned}$$

und wiederum wegen $\beta > 0$ und wegen $\{0, \dots, -1\} = \emptyset$ gilt $0 \in M$, und damit existiert $\max M$, und mit der Isotonie von $x \mapsto B_{n,p}(\{0, \dots, x-1\})$ folgt

$$M = \{0, \dots, \max M\}.$$

Ist nun $\max M = n$, so folgt sofort $B_{n,p}(K \ni p) = B_{n,p}(\mathcal{X}) = 1 \geq \beta$. Ist dagegen $\max M < n$, so ist $1 + \max M \in \mathcal{X} \setminus M$ und damit, nach der Definition von M im letzten Schritt,

$$B_{n,p}(K \ni p) = B_{n,p}(M) = B_{n,p}(\{0, \dots, (1 + \max M) - 1\}) \geq \beta$$

und damit die Behauptung bezüglich (4); die Formeln in (2) und (3) sind dabei offensichtlich.

2. Wegen $B_{n,p}(\{0, \dots, x\}) = B_{n,1-p}(\{n-x, \dots, n\})$ gilt (1); also läßt sich (5) auf (4) zurückführen, und ebenso die Monotonie von $\bar{p}_\beta(x)$ auf die von $\underline{p}_\beta(x)$.

3. Subadditivität angewandt auf (4) und (5) liefert (6): Bezeichnen K und Λ die Intervalle zum Konfidenzniveau $\frac{1+\beta}{2}$ gemäß (4) beziehungsweise (5), so folgt mit $P := B_{n,p}$

$$\begin{aligned} P(K \cap \Lambda \ni p) &= 1 - P(K \not\ni p \text{ oder } \Lambda \not\ni p) \\ &\geq 1 - P(K \not\ni p) - P(\Lambda \not\ni p) \geq 1 - 2 \left(1 - \frac{1+\beta}{2}\right) = \beta. \end{aligned}$$

□

13.9 Beispiel (BSE). Wird die in Beispiel 13.6 gegebene Situation durch das Schätzproblem 13.5(a) mit $n = 234$ modelliert, und möchte man “nachweisen”, dass die Übertragungswahrscheinlichkeit p klein ist, so sollte man von den drei Konfidenzintervallen aus 13.8 das Intervall 13.8(5) wählen, da aus der Monotonie der Funktionen h_x aus dem Beweis von Satz 13.8 die Ungleichungen $\bar{p}_\beta(x) < \bar{p}_{\frac{1+\beta}{2}}(x) < 1$ für $x \in \{0, \dots, n - 1\}$ und $\beta \in]0, 1[$ folgen, und daher mit 13.8(5) die schärfsten oberen Schranken für p erhalten werden.

Speziell für das übliche $\beta := 0.95$ und mit der Beobachtung $x = 0$ ergibt 13.8(5) das Intervall $[0, \bar{p}_\beta(0)[= [0, 0.0127\dots[$ unter Benutzung von 13.8(3), und damit die Aussage “ $p < 0.013$ ” statt “ $p = 0$ ”. \square

13.10 Bemerkung. Die Konfidenzintervalle aus 13.8 können bequem mit R berechnet werden: Mit konkreten Zahlen x und n , oder mit vorherigen Definitionen $x < \dots$ und $n < \dots$, ergeben die Eingaben

```
binom.test(x,n,alternative="greater"),
binom.test(x,n,alternative="less"),
binom.test(x,n,alternative="two.sided")
```

jeweils alles mögliche, und dabei insbesondere die Intervalle 13.8(4,5,6) in dieser Reihenfolge und mit $\beta = 0.95$; statt `binom.test(x,n,alternative="two.sided")` reicht auch `binom.test(x,n)`. Das Konfidenzniveau kann auch geändert werden, siehe `?binom.test`. Möchte man sich nur das Intervall und das Konfidenzniveau ausgeben lassen, so füge man `$conf.int` dem jeweiligen obigen Befehl hinten an. \square

13.11 Aufgabe. Es sei $P_\bullet = (P_\vartheta : \vartheta \in \Theta)$ ein Modell auf \mathcal{X} .

(a) Für $i \in \{1, 2\}$ sei $\kappa_i : \mathcal{X}$ ein interessierender Parameter, $\beta_i = 1 - \alpha_i \in [0, 1]$, und K_i ein β_i -Konfidenzbereich für (P_\bullet, κ_i) . Es sei $\kappa := (\kappa_1, \kappa_2)$. Dann ist $K := (K_1 \times K_2) \cap \kappa[\Theta]$, genauer definiert durch

$$K(x) := \left(K_1(x) \times K_2(x) \right) \cap \kappa[\Theta] \quad \text{für } x \in \mathcal{X},$$

ein Konfidenzbereich zum Niveau $\beta_1 + \beta_2 - 1 = 1 - \alpha_1 - \alpha_2$ für (P_\bullet, κ) .

(b) Kann man die Konklusion von Teil (a) verschärfen, falls zusätzlich angenommen wird, dass für jedes $\vartheta \in \Theta$ die Zufallsgrößen K_1, K_2 P_ϑ -unabhängig sind?

(c) Es sei $\kappa : \Theta$ ein interessierender Parameter, und K ein β -Konfidenzbereich für (P_\bullet, κ) . Es sei weiter $\lambda : \kappa[\Theta]$ eine Funktion. Was ergibt sich dann für das Schätzproblem $(P_\bullet, \lambda \circ \kappa)$? \square

13.12 Aufgabe. Wie groß ist die Wahrscheinlichkeit dafür, dass eineiige Zwillinge die gleichen Rauchgewohnheiten haben? Beantworten Sie diese Frage aufgrund der Daten aus Aufgabe 8.16 nun weniger naiv durch Angabe eines 95%-Konfidenzintervalls. Analog für zweieiige Zwillinge.

Können Sie sich damit “zu 95% sicher sein”, dass es hier zwischen den ein- und den zweieiigen Zwillingspaaren einen Unterschied gibt? (Zwei Begründungen!) \square

Anders als beim logischen Schließen kann es beim **statistischen Schließen** (auch **statistische Inferenz** genannt), womit meist die Anwendung eines Konfidenzbereichs gemäß Definition 13.4 oder eines Tests gemäß Definition 13.13 gemeint ist, durchaus schaden, zu präzise zu fragen: In Aufgabe 13.12 führte zum üblichen Konfidenzniveau 0.95 die Frage

$$(1) \quad \text{“Wo liegt } (p_1, p_2)\text{?”}$$

gemäß Aufgabe 13.11 (a oder b) auf ein Konfidenzrechteck, welches keine nichttriviale Antwort auf die Frage

$$(2) \quad \text{“} p_1 \neq p_2 \text{?”}$$

mehr gestattet. Verglichen mit (1) unpräzise Fragen wie (2), sowie zu ihrer Beantwortung passende statistische Verfahren, werden formalisiert durch:

13.13 Definition. Es sei $P_\bullet = (P_\vartheta : \vartheta \in \Theta)$ ein Modell auf \mathcal{X} , und es sei⁶ $\Theta_0 \subseteq \Theta$. Dann heißt (P_\bullet, Θ_0) **Testproblem** mit der **Hypothese** Θ_0 und der **Alternative** $\Theta \setminus \Theta_0$.

Ist weiter $\alpha \in [0, 1]$ und $\psi : \mathcal{X} \rightarrow \{0, 1\}$ eine Funktion mit

$$(1) \quad P_\vartheta(\psi = 1) \leq \alpha \quad \text{für } \vartheta \in \Theta_0,$$

so heißt ψ **Test** zum (**Fehler-**)**Niveau** α für (P_\bullet, Θ_0) , mit dem **kritischen Bereich** $\{x \in \mathcal{X} : \psi(x) = 1\}$. □

Interpretation: Nach Beobachtung von $x \in \mathcal{X}$ behaupte

$$\text{“} \vartheta \in \left\{ \begin{array}{c} \Theta \setminus \Theta_0 \\ \Theta \end{array} \right\} \text{” falls } \psi(x) = \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\};$$

d.h. im ersten Fall wird die Hypothese Θ_0 verworfen, während im zweiten Fall nichts behauptet wird was nicht schon in der Modellannahme steckt. Damit ist die linke Seite von (1) die Fehlerwahrscheinlichkeit der Behauptung im Fall von $\vartheta \in \Theta_0$, während für $\vartheta \in \Theta \setminus \Theta_0$ die Fehlerwahrscheinlichkeit Null ist. Tests können als spezielle grobe Konfidenzbereiche aufgefasst werden:

13.14 Bemerkung. Es sei (P_\bullet, Θ_0) ein Testproblem auf \mathcal{X} mit dem Parameterraum Θ , und es sei $\alpha = 1 - \beta \in [0, 1]$. Eine Funktion $\psi : \mathcal{X} \rightarrow \{0, 1\}$ ist genau dann ein Test zum Niveau α für (P_\bullet, Θ_0) , wenn

$$(1) \quad K(x) := \left\{ \begin{array}{c} \Theta \setminus \Theta_0 \\ \Theta \end{array} \right\} \quad \text{falls } \psi(x) \in \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\}$$

einen β -Konfidenzbereich für $(P_\bullet, \text{id}_\Theta)$ definiert, denn mit (1) gilt $\{K \not\neq \vartheta\} = \{\psi = 1\}$ falls $\vartheta \in \Theta_0$, und $\{K \not\neq \vartheta\} = \emptyset$ falls $\vartheta \in \Theta \setminus \Theta_0$, und damit $P_\vartheta(K \not\neq \vartheta) \leq \alpha$ für $\vartheta \in \Theta$ genau dann, wenn 13.13(1) gilt. □

⁶Man denke sich die Fälle $\Theta_0 = \emptyset$ und $\Theta_0 = \Theta$ ausgeschlossen; in einer ausführlicheren Darstellung als hier erweist sich ihre Zulassung jedoch als mathematisch bequem.

Wir begnügen uns zum Abschluss dieses Kapitels mit einem wichtigen Beispiel eines Tests, welcher auf die Frage 13.12(2) anwendbar ist:

13.15 Satz (Fishers exakter Test). *Es sei $(n_1, n_2) \in \mathbb{N}^2$,*

$$(1) \quad P_\bullet := \left(B_{n_1, p_1} \otimes B_{n_2, p_2} : p \in [0, 1]^2 \right)$$

das kanonische Modell für zwei unabhängige binomialverteilte Beobachtungen mit den bekannten Stichprobenumfangparametern n_1 und n_2 , auf dem Stichprobenraum $\mathcal{X} := \{0, \dots, n_1\} \times \{0, \dots, n_2\}$, und es sei

$$\Theta_0 := \{p \in [0, 1]^2 : p_1 = p_2\},$$

die Hypothese gleicher Trefferwahrscheinlichkeiten, sowie $\alpha \in [0, 1]$. Dann definiert

$$\psi(x) := \left(H_{x_1+x_2, n_1, n_2}(\{0, \dots, x_1\}) \leq \frac{\alpha}{2} \text{ oder} \right. \\ \left. H_{x_1+x_2, n_1, n_2}(\{x_1, \dots, x_1 + x_2\}) \leq \frac{\alpha}{2} \right) \quad \text{für } x \in \mathcal{X}$$

einen Test ψ zum Niveau α für (P_\bullet, Θ_0) .

Beweis. Es sei $p = (\pi, \pi) \in \Theta_0$ gegeben und $P := B_{n_1, \pi} \otimes B_{n_2, \pi}$. Zu zeigen ist

$$(2) \quad P(\psi = 1) \leq \alpha.$$

Wir setzen $(X_1, X_2) := \text{id}_{\mathcal{X}}$, $S := X_1 + X_2$, $N := n_1 + n_2$,

$$(3) \quad H_s := H_{s, n_1, n_2} \quad \text{für } s \in \{0, \dots, N\},$$

$$(4) \quad \psi_1(x) := \left(H_{x_1+x_2}(\{0, \dots, x_1\}) \leq \frac{\alpha}{2} \right) \quad \text{und}$$

$$(5) \quad \psi_2(x) := \left(H_{x_1+x_2}(\{x_1, \dots, x_1 + x_2\}) \leq \frac{\alpha}{2} \right) \quad \text{für } x \in \mathcal{X}.$$

Unter P sind nach 8.8 und 7.9(11) dann

$$(6) \quad X_i \sim B_{n_i, \pi} \quad \text{unabhängig für } i \in \{1, 2\}.$$

1. Es gelte zunächst $\pi \in]0, 1[$. Für $s \in \{0, \dots, N\}$ ist dann

$$P(X_1 = \xi | S = s) = h_{s, n_1, n_2}(s) \quad \text{für } \xi \in \mathbb{Z},$$

nach Aufgabe 8.3 oder hier direkt durch leichtes Nachrechnen, und mit den Eigenschaften bedingter Wahrscheinlichkeiten aus 6.4(a) und der Abkürzung (3) folgt

$$(7) \quad P(X_1 \leq c | S = s) = H_s(\{0, \dots, c\}) \quad \text{und}$$

$$(8) \quad P(X_1 \geq c | S = s) = H_s(\{c, \dots, N\}) \quad \text{für } c \in \mathbb{Z},$$

und damit

$$\begin{aligned}
 P(\psi_1 = 1, S = s) &= P(\mathbb{H}_s(\{0, \dots, X_1\}) \leq \frac{\alpha}{2}, S = s) \\
 &= P(X_1 \leq c, S = s) \text{ mit } c := \max\{\xi \in \mathbb{Z} : \mathbb{H}_s(\{0, \dots, \xi\}) \leq \frac{\alpha}{2}\} \\
 &= P(X_1 \leq c | S = s)P(S = s) \\
 &= \mathbb{H}_s(\{0, \dots, c\})P(S = s) \leq \frac{\alpha}{2}P(S = s)
 \end{aligned}$$

mit der Definition von ψ_1 und einem Einsetzen von s für S im ersten Schritt, der Isotonie von $F(\xi) := \mathbb{H}_s(\{0, \dots, \xi\})$ sowie $F(-1) = 0 \leq \frac{\alpha}{2}$ und $F(s) = 1 > \frac{\alpha}{2}$ im zweiten, (7) im vierten, und der Definition von c im fünften. Folglich ist

$$P(\psi_1 = 1) = \sum_{s=0}^N P(\psi_1 = 1, S = s) \leq \frac{\alpha}{2} \sum_{s=0}^N P(S = s) = \frac{\alpha}{2},$$

und analog, mit (8) und dann $c := \min\{\xi \in \mathbb{Z} : \mathbb{H}_s(\{\xi, \dots, N\}) \leq \frac{\alpha}{2}\}$, ergibt sich $P(\psi_2 = 1) \leq \frac{\alpha}{2}$, und mit

$$(9) \quad P(\psi = 1) \leq P(\psi_1 = 1) + P(\psi_2 = 1)$$

erhalten wir (2).

2. Erster Beweis für die Randfälle: Im Fall $\pi = 0$ ist $P = \delta_0 \otimes \delta_0$, d.h. für P -fast jedes $x = (x_1, x_2) \in \mathcal{X}$ ist $x = 0$, also $\mathbb{H}_{x_1+x_2} = \mathbb{H}_0 = \delta_0$, für $i \in \{1, 2\}$ also $\psi_i(x) = (\delta_0(\{0\}) \leq \frac{\alpha}{2}) = (1 \leq \frac{\alpha}{2}) = 0$; also gilt hier $P(\psi_i = 1) = 0$ für $i \in \{1, 2\}$, und mit (9) folgt wieder (2). Im Fall $\pi = 1$ ist für P -fast jedes x analog $\mathbb{H}_{x_1+x_2} = \mathbb{H}_N = \delta_N$ und damit wieder $\psi_i(x) = 0$ für $i \in \{1, 2\}$.

3. Alternativer Beweis für die Randfälle: Mit Satz 10.20 ist

$$\text{L.S.}(2) = \sum_{x \in \mathcal{X}} (\psi(x) = 1) b_{n_1, \pi}(x_1) b_{n_2, \pi}(x_2)$$

polynomiell in $\pi \in [0, 1]$, also stetig, und folglich liefert 1. schon die allgemeine Gültigkeit von (2). □

13.16 Beispiel. In Aufgabe 8.16 ist $n_1 = 51, x_1 = 33, n_2 = 31, x_2 = 11$, also

$$\begin{aligned}
 \mathbb{H}_{x_1+x_2, n_1, n_2}(\{x_1, \dots, x_1 + x_2\}) &= \mathbb{H}_{44, 51, 31}(\{33, \dots, 44\}) \\
 &= 1 - \mathbb{H}_{44, 51, 31}(\{0, \dots, 32\}) \\
 &= 1\text{-phyper}(32, 51, 31, 44) \\
 &= 0.009333369 \\
 &\leq \frac{0.05}{2},
 \end{aligned}$$

also $\psi(x) = 1$ bei Verwendung des üblichen Fehlerniveaus $\alpha = 0.05$, für welches somit die Hypothese “ $p_1 = p_2$ ” verworfen wird, also “unterschiedliche Rauchgewohnheiten” behauptet wird. □

Statt oben “1-phyper(32,51,31,44)” kann man, mit etwas Konzentration, auch geringfügig weniger in R eintippen:

13.17 Bemerkung. Für $(n, r, b) \in \mathbb{N}_0^3$ mit $n \leq r + b$ und für $x \in \mathbb{R}$ gilt, tatsächlich beziehungsweise numerisch in R siehe 5.16 und 3.12,

$$(1) \quad H_{n,r,b}([x, \infty]) = H_{n,b,r}([-\infty, n - x]),$$

$$(2) \quad H_{n,r,b}([-\infty, x]) = \text{phyper}(x, r, b, n).$$

Damit erhält man in der Notation der Sätze 13.15 und, unten, 13.18

$$(3) \quad H_{x_1+x_2}(\{0, \dots, x_1\}) = \text{phyper}(x_1, n_1, n_2, x_1+x_2),$$

$$(4) \quad H_{x_1+x_2}(\{x_1, \dots, x_1 + x_2\}) = \text{phyper}(x_2, n_2, n_1, x_1+x_2)$$

zur numerischen Berechnung von $\psi_1(x)$ und $\psi_2(x)$. □

Nun wird man die oben in Beispiel 13.16 gemäß der Interpretation bei Definition 13.13 erhaltene Aussage wohl etwas zu grob finden: Erstens müsste es wegen $\frac{33}{51} > \frac{11}{31}$ doch eigentlich möglich sein, immer noch im Rahmen einer Fehlerwahrscheinlichkeit $\leq 1 - \alpha$, sogar “ $p_1 > p_2$ ” zu behaupten. Und zweitens: Verschiedenes ist nun mal verschieden, und folglich sind unterschiedliche Rauchgewohnheiten bei unterschiedlichen Zwillingspaartypen doch sowieso zu erwarten, und die außermathematische Interpretation des mathematisch durchaus nichttrivialen Ergebnisses des Beispiels 13.16 erscheint trivial.

Tatsächlich gilt folgende Verfeinerung von Satz 13.15, welche insbesondere im obigen Beispiel “ $p_1 > p_2$ ” liefert:

13.18 Satz (Fishers exakter Test mit Richtungsentscheidung). *In der Situation von Satz 13.15, mit ψ_1, ψ_2 definiert in 13.15(4,5), sowie mit*

$$A(x) := \left\{ \begin{array}{l} \text{“}p_1 < p_2\text{”} := \{p \in [0, 1]^2 : p_1 < p_2\} \\ \text{“}p_2 < p_1\text{”} := \{p \in [0, 1]^2 : p_2 < p_1\} \\ \text{“Keine Aussage”} := [0, 1]^2 \end{array} \right\} \text{ falls } \left\{ \begin{array}{l} \psi_1(x) = 1 \\ \psi_2(x) = 1 \\ \psi_1(x) = \psi_2(x) = 0 \end{array} \right\},$$

ist $A : \mathcal{X} \rightarrow 2^{[0,1]^2}$ wohldefiniert mit

$$(1) \quad (B_{n_1, p_1} \otimes B_{n_2, p_2})(A \ni p) \geq 1 - \alpha \quad \text{für } p \in [0, 1]^2.$$

Beweis. Es handelt sich um einen Spezialfall allgemeinerer Aussagen (über bedingte zweiseitige Tests in exponentiellen Familien wie in Lehmann/Romano (2005, Abschnitte 4.4 und 4.5) und Witting (1985, Abschnitte 3.3.2 und 3.3.3)), welcher üblicherweise in Vorlesungen zur Mathematischen Statistik (in Trier voraussichtlich das nächste mal im Wintersemester 2022/23 oder im Sommersemester 2023) bewiesen wird. Wir bringen hier nur den ersten Beweisschritt der Wohldefiniertheit:

1. Mit der Abkürzung aus 13.15(3) gilt für $x \in \mathcal{X}$ nach der aus der Tabelle in 3.12 und aus Aufgabe 4.8(b) bekannten Formel für die Träger hypergeometrischer Verteilungen

$$\text{Supp } H_{x_1+x_2} = \{(x_1 + x_2 - n_2)_+, \dots, (x_1 + x_2) \wedge n_1\} \ni x_1,$$

also $H_{x_1+x_2}(\{0, \dots, x_1\}) + H_{x_1+x_2}(\{x_1, \dots, x_1+x_2\}) = 1 + H_{x_1+x_2}(\{x_1\}) > 1 \geq \frac{\alpha}{2} + \frac{\alpha}{2}$, also nicht $\psi_1(x) = \psi_2(x) = 1$. Folglich ist A wohldefiniert. \square

Eine typische Anwendung von Satz 13.18 wäre ein Vergleich zweier Behandlungsmethoden mit nur zwei möglichen Ergebnissen:

13.19 Beispiel (Wirkt Clofibrat nach einem Herzinfarkt lebensverlängernd?). Dieses Beispiel folgt Freedman/Pisani/Purves (2007) und der Originalquelle, Coronary Drug Project Research Group (1980); letztere enthält Verweise auf frühere Publikationen (1973, 1974, 1975).

Stichprobe: Alle 8341 geeigneten Patienten (Männer, 30-64 Jahre alt, mit Herzinfarkt in den letzten drei Monaten) in mehreren teilnehmenden Kliniken vom März 1966 bis Oktober 1969. Je ungefähr 1100 von diesen bekamen eines von fünf neuen Medikamenten zufällig und doppelblind zugewiesen, die restlichen ein Placebo. Gefragt ist nach der Auswirkung auf die 5-Jahres-Überlebensrate. Wir betrachten hier nur das eine Medikament Clofibrat und das Placebo, mit insgesamt 3892 Patienten.

Ergebnis:

	Clofibrat	Placebo
(1)	# Patienten	1103 2789
	Davon tot nach 5 Jahren	20.0% 20.9%

Also wohl fast keine Verbesserung, wenn überhaupt, durch Clofibrat?

Aber nicht alle Patienten nahmen die ihnen zugewiesenen Medikamente so, wie sie es sollten. Durch Beobachtungen und Befragungen wurde auch die Adhärenz⁷ von Patienten eingeschätzt und möglichst als “ $\geq 80\%$ ” oder “ $< 80\%$ ” klassifiziert, je nach dem ob im Beobachtungszeitraum 80% der vorgesehenen Pillen eingenommen wurden oder nicht. Für die Clofibrat-Gruppe ergab sich mit

	Clofibrat mit Adhärenz $\geq 80\%$	
(2)	# Patienten	708
	Davon tot nach 5 Jahren	15.0%

eine doch deutlich geringere Sterblichkeitsrate unter denen, die ihre Medikamente im wesentlichen wie vorgeschrieben einnahmen. Also hilft Clofibrat doch etwas? Aber wem eigentlich?

⁷Ausmaß der Befolgung einer medizinisch verordneten Behandlung.

Grundgesamtheit: Gedachte Menge aller mittelalten männlichen Amerikaner mit Herzinfarkt irgendwann zwischen März 1966 und zum Beispiel Dezember 1985, denn eigentlich interessieren, bei Publikation der Hauptergebnisse der Studie im Jahre 1975 (Zulassung des letzten Studienteilnehmers Oktober 1969, fünf Jahre Wartezeit um Erleben bzw. Nichterleben beobachten zu können, ein Jahr bis zur Publikation), “alle” künftigen solchen Patienten, aber einerseits beginnt nun mal die Stichprobenentnahme im März 1966, und andererseits wird man weder bis in alle Ewigkeit von einer homogenen Grundgesamtheit ausgehen, noch die Entwicklung besserer Medikamente etwa in den 10 Jahren nach 1975 vernachlässigen wollen. Denkt man sich die insgesamt N Patienten der gedachten Menge etwa nach den Zeitpunkten ihrer Herzinfarkte durchnummeriert, so stellt man sich den ν -ten Patienten für die gegenwärtigen Zwecke durch zwei Zahlen

$$\left\{ \begin{matrix} t_{\nu,1} \\ t_{\nu,2} \end{matrix} \right\} = \text{Überlebenszeit nach Herzinfarkt mit } \left\{ \begin{matrix} \text{Placebo} \\ \text{Clofibrat} \end{matrix} \right\}$$

charakterisiert vor, und setzt

$$\text{5-Jahres-Sterbewahrscheinlichkeit mit } \left\{ \begin{matrix} \text{Placebo} \\ \text{Clofibrat} \end{matrix} \right\} := \left\{ \begin{matrix} p_1 \\ p_2 \end{matrix} \right\}$$

mit

$$p_i := \frac{1}{N} \sum_{\nu=1}^N (t_{\nu,i} \leq 5) \quad \text{für } i \in \{1, 2\}.$$

Man stellt sich weiter vor, dass sich die p_i durch “nichthellseherische” Zuordnung jedes einzelnen Patienten zur Placebo- beziehungsweise Clofibratbehandlung praktisch nicht ändern.⁸ Dann bedeutet $p_2 < p_1$ einen Nutzen von Clofibrat, bezogen auf die 5-Jahres-Überlebenswahrscheinlichkeit in der betrachteten Grundgesamtheit. Damit

Verteilungsannahme an die Stichprobe: Daten $x := (x_1, x_2) \in \{0, 1\}^{n_1} \times \{0, 1\}^{n_2}$ mit

$$x_{i,j} := 1 \Leftrightarrow j\text{-ter Patient in Stichprobenbehandlungsgruppe } i \text{ stirbt,}$$

mit $n_1 := 2789$ (Placebo-Patienten) und $n_2 := 708$ (Clofibrat-Adhärenzte), wobei⁹

$$\begin{aligned} x_{1+} &:= \sum_{j=1}^{n_1} x_{1,j} = 2789 \cdot 20.9\% = 583, \\ x_{2+} &:= \sum_{j=1}^{n_2} x_{2,j} = 708 \cdot 15.0\% = 106. \end{aligned}$$

⁸Präzisierung möglich durch Idealisierung $N \rightarrow \infty$ und den Kollektivbegriff aus Kapitel 2.

⁹Genauer hat man zwar $105 < 105.8\dots = 708 \cdot 0.1495 \leq x_{2+} \leq 708 \cdot 0.1505 = 106.5\dots < 107$, also wirklich $x_{2+} = 106$, aber $581.5\dots = 2789 \cdot 0.2085 \leq x_{1+} \leq 2789 \cdot 0.2095 = 584.2\dots$ und damit nur $x_{1+} \in \{582, 583, 584\}$.

Modell für x_1 eigentlich über Ziehen ohne Zurücklegen aus der Population der Placebobehandelten, und entsprechend, stochastisch unabhängig von diesen, x_2 aus den Clofibratadhärenten. Vernachlässigung des Zurücklegens (N groß¹⁰) führt für (x_{1+}, x_{2+}) in der Rolle von $x = (x_1, x_2)$ auf das Modell in 13.15(1).

Modellbasierte Inferenz: Satz 13.18 mit $\alpha := 0.05$, angewandt auf $(n_1, n_2, x_{1+}, x_{2+}) = (2789, 708, 583, 106)$ aus Beispiel 13.19 statt (n_1, n_2, x_1, x_2) wie in 13.18, liefert mit Bemerkung 13.17

$$\begin{aligned} \psi_2(x_{1+}, x_{2+}) &= (\text{phyper}(106, 708, 2789, 583+106) \leq 0.025) \\ &= (0.000176 \dots \leq 0.025) \\ &= 1 \end{aligned}$$

und damit die Aussage “ $p_2 < p_1$ ”, also eine lebensverlängernde Wirkung von Clofibrat. → 13.21

Natürlich sind solche Erfolge statistischer Auswertungen nur mit einigermaßen großen Stichprobenumfängen möglich, wie illustriert durch:

13.20 Aufgabe. In Beispiel 13.19 war

$$\frac{x_{1+}}{n_1} = \frac{583}{2789} \approx \frac{1}{5}, \quad \frac{x_{2+}}{n_2} = \frac{106}{708} \approx \frac{1}{7}, \quad \frac{n_1}{n_2} = \frac{2789}{708} \approx 4.$$

Für $k \in 35\mathbb{N}$ betrachten wir die insofern ähnlichen fiktiven Daten

	Clofibrat-Adhärente	Placebo
# Patienten	k	4k
# tot nach 5 Jahren	$\frac{k}{5}$	$\frac{4k}{7}$

Welche Aussage liefert Satz 13.18 mit $\alpha = 0.05$ dann jeweils? (Numerisch zu lösen mit R.) □

13.21 Beispiel (13.19 ←). Die Daten 13.19(1,2) sind jedoch nur ein Teil der Wahrheit.

Vollständigeres Ergebnis:

	Clofibrat, Adhärenz			Placebo, Adhärenz		
	≥ 80%	< 80%	?	≥ 80%	< 80%	?
(1) # Patienten	708	357	38	1813	882	94
Davon tot nach 5 Jahren	15.0%	24.6%		15.1%	28.2%	
# Patienten	1103			2789		
Davon tot nach 5 Jahren	20.0%			20.9%		

¹⁰Wir verzichten hier auf eine genauere Rechtfertigung mit Sätzen von Freedman (1977), Ehm (1991), und anderen.

38 + 94 Patienten wurden nicht bezüglich Adhärenz klassifiziert.

Ist also auch das Placebo für die Adhärenten lebensverlängernd? Jedenfalls wirkt Clofibrat gegenüber dem Placebo nicht lebensverlängernd unter den Adhärenten, und kaum unter den Nichtadhärenten. Vielmehr leben die Adhärenten allgemein, unabhängig von der Behandlung, länger als die Nichtadhärenten. Also ist obiger Vergleich zwischen x_{1+} und x_{2+} unsinnig, da sich x_{2+} aus eine Stichprobe aus allen Patienten ergibt, x_{1+} dagegen aus einer aus der Schicht der Adhärenten. \square

Allgemein ist eine in sich noch so konsistente statistische Inferenz, notwendig ausgehend von einer Verteilungsannahme, oft schädlich, wenn besagte Annahme unangemessen ist. Speziell beim *Vergleich* medizinischer Behandlungen zählt nicht die Behandlung, sondern die Absicht dazu (auch auf deutsch oft **intention to treat** genannt).

A Boole-Algebren

A.1 Definition. Es sei \mathcal{A} eine Menge mit zwei assoziativen und kommutativen binären Verknüpfungen \wedge und \vee , zugehörigen neutralen Elementen I bzw. O mit $O \neq I$, und einer unären Verknüpfung $'$ mit

$$(1) \quad A \wedge A' = O \quad \text{und} \quad A \vee A' = I \quad \text{für } A \in \mathcal{A}.$$

Weiter sei \wedge distributiv über \vee und umgekehrt, d.h. für $A, B, C \in \mathcal{A}$ gelte

$$(2) \quad A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C) \quad \text{und} \quad A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C).$$

Dann heißt $(\mathcal{A}, \wedge, \vee, ', O, I)$ **Boole-Algebra**, mit der **Trägermenge**¹ \mathcal{A} , der² **Schnittbildung** \wedge , der **Vereinigungsbildung** \vee , der **Komplementierung** $'$, der **Null**³ O , und der **Eins** I . Sie heißt **nichtdegeneriert** falls $\#\mathcal{A} > 2$ gilt. \square

Wir lesen $A \wedge B$, $A \vee B$, A' kurz etwa als “ A geschnitten B , A vereinigt B , A Komplement”, wie in den Beispielen A.3(a) und A.4(a); siehe dazu auch die Sätze A.16 und A.17. Alternativ lesen auch “ A und B , A oder B , nicht A ”, motiviert durch Satz 1.11(a).

Mit etwas anderen Worten lautet die Definition A.1: Eine Boole-Algebra ist ein Sextupel $(\mathcal{A}, \wedge, \vee, ', O, I)$ mit $O \neq I$ und derart, dass (\mathcal{A}, \wedge, I) und (\mathcal{A}, \vee, O) kommutative Monoide sind, die Selbstabbildung $' : \mathcal{A} \rightarrow \mathcal{A}$ die Bedingung (1) erfüllt, und \wedge und \vee zueinander distributiv sind.

Wird eine “mathematische Struktur” zur Modellierung einer Anwendungssituation eingeführt, wie in unserem Fall die Struktur “Boole-Algebra” zur Modellierung der von uns im täglichen Leben praktizierten Aussagenlogik, wie in Kapitel 1 ausgeführt, so sollte die betreffende Struktur idealerweise durch möglichst wenige Eigenschaften charakterisiert werden, da dann nur diese für die Anwendungssituation zu rechtfertigen sind. In dieser Hinsicht ist Definition A.1 sicher nicht ideal, da zwei der dort gestellten Forderungen weggelassen werden können, was hier der Vollständigkeit halber bemerkt sei:

A.2 Satz. *In Definition A.1 können die beiden Assoziativitätsbedingungen ersatzlos gestrichen werden.*

Beweis. Siehe Huntington (1904, Behauptungen XIIIa und XIIIb auf p. 294, Beweise auf pp. 295–296). \square

¹Eine allgemein übliche Bezeichnung für die Menge, auf deren kartesischen Potenzen die Verknüpfungen einer “algebraischen Struktur” definiert sind. Die stattdessen auch oft benutzte Bezeichnung “Grundmenge” versuchen wir hier zu vermeiden, da sie in den häufig vorkommenden Beispielen A.3(b) und A.4(a) zu Verwirrungen führen könnte: Die Grundmenge ist dort Ω , die Trägermenge dagegen $\mathcal{A} = 2^\Omega$ beziehungsweise allgemeiner $\mathcal{A} \subseteq 2^\Omega$.

²Die folgenden fünf Namen können bei Bedarf durch den Vorsatz “**Boole-**” präzisiert werden.

³Man kann O auch als “römisch Null” und I als “römisch Eins” lesen.

Die oben gewählte Version der Definition A.1 ist daher als ein Kompromiss anzusehen, indem hier einerseits die weiteren Rechenregeln aus Satz A.8 bewiesen werden, und andererseits aus Zeitgründen auf einen Beweis von Satz A.2 verzichtet wird. Für unsere konkrete Anwendungssituation in Kapitel 1 kommen wir dem oben angedeuteten Ideal jedoch mit dem Beweis des Satzes A.11 wohl sehr nahe, ohne im folgenden den Satz A.2 jemals zu verwenden.

A.3 Beispiele. (a) Für $O \neq I$ und $O \wedge O := O \wedge I := I \wedge O := O$, $I \wedge I := I$, $O \vee O := O$, $O \vee I := I \vee O := I \vee I := I$, $O' := I$, $I' := O$ ist $(\{O, I\}, \wedge, \vee, ', O, I)$ eine degenerierte Boole-Algebra.

(b) Es sei Ω eine nichtleere Menge und $\mathcal{A} := 2^\Omega$. Mit den üblichen binären Mengenoperationen \cap und \cup sowie mit der Komplementbildung c ist dann $(\mathcal{A}, \cap, \cup, ^c, \emptyset, \Omega)$ eine Boole-Algebra, die **Potenzmengenalgebra** von Ω . \square

Potenzmengenalgebren sind Spezialfälle von Mengenalgebren:

A.4 Aufgabe. (a) Es sei Ω eine nichtleere Menge und es sei $\mathcal{A} \subseteq 2^\Omega$ nichtleer, \cup -stabil, und **komplementstabil**, d.h. für $A, B \in \mathcal{A}$ gelte $A \cup B \in \mathcal{A}$ und $A^c \in \mathcal{A}$. Dann ist $(\mathcal{A}, \cap, \cup, ^c, \emptyset, \Omega)$ eine Boole-Algebra, eine sogenannte **Mengen-Algebra** auf der **Grundmenge** Ω .

(b) Wie viele Mengenalgebren auf der Grundmenge $\{1, 2, 3\}$ gibt es? \square

Zum Beispiel ist die Menge aller endlichen Vereinigungen von Intervallen (die auch einelementig oder gar leer sein dürfen) in \mathbb{R} eine Mengenalgebra.

A.5 Bemerkung. Gilt in der Situation von Aufgabe A.4(a) auch $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$ für jede Folge (A_n) in \mathcal{A} , so heißt \mathcal{A} **Mengen- σ -Algebra**⁴. Zum Beispiel ist die Potenzmengenalgebra aus A.3(b) stets eine Mengen- σ -Algebra, die Algebra \mathcal{A} der endlichen Vereinigungen von Intervallen in \mathbb{R} aber nicht, da ja zum Beispiel $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}$ eine abzählbare Vereinigung von Einpunktintervallen ist, aber nicht als endliche Vereinigung von Intervallen geschrieben werden kann. Mit letzterem \mathcal{A} wird in der MIT insbesondere die kleinste σ -Algebra \mathcal{B} mit $\mathcal{A} \subseteq \mathcal{B}$ betrachtet, die sogenannte **Borel- σ -Algebra** von \mathbb{R} . \square

A.6 Bemerkung (Dualität). Mit $(\mathcal{A}, \wedge, \vee, ', O, I)$ ist offenbar auch $(\mathcal{A}, \vee, \wedge, ', I, O)$ ein Boole-Algebra. \square

A.7 Aufgabe. Für $n \in \mathbb{N}$ sei $T_n := \{x \in \mathbb{N} : \frac{n}{x} \in \mathbb{N}\}$ die Menge aller Teiler von n , und damit $x \wedge y := \text{ggT}(x, y) := \max T_x \cap T_y$, $x \vee y := \text{kgV}(x, y) := \min(x\mathbb{N}) \cap (y\mathbb{N})$, und $x' := \frac{n}{x}$ für $x, y \in T_n$. Für welche n ist $(T_n, \wedge, \vee, ', 1, n)$ eine Boole-Algebra? \square

⁴Grundbegriff der ursprünglich rein analytisch-geometrisch motivierten, um 1900 von Émile Borel (1871–1956) und Henri Léon Lebesgue (1875–1941) initiierten Maß- und Integrationstheorie (MIT), dort kürzer “ σ -Algebra” genannt. Siehe etwa Elstrodt (2018, Kapitel 1). Die MIT wurde dann recht schnell als eine geeignete Grundlage der WT erkannt, siehe Shafer/Vovk (2006).

A.8 Satz. *Es sei $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra. Dann sind \wedge und \vee **idempotent** mit den **absorbierenden** Elementen O bzw. I , d.h. es gilt*

$$(1) \quad A \wedge A = A, \quad A \wedge O = O, \quad A \vee A = A, \quad A \vee I = I \quad \text{für } A \in \mathcal{A}.$$

Weiter gilt

$$(2) \quad O' = I, \quad I' = O,$$

und, für $A, B \in \mathcal{A}$,

$$(3) \quad B = A' \Leftrightarrow A \wedge B = O \text{ und } A \vee B = I \quad (\text{Charakterisierung der Komplementarität}),$$

$$(4) \quad A'' = A \quad (\text{Involutivität der Komplementierung}),$$

$$(5) \quad (A \wedge B)' = A' \vee B' \text{ und } (A \vee B)' = A' \wedge B' \quad (\text{De Morgansche Regeln}),$$

sowie noch die Johnson-Identität 1.7(1), also

$$(6) \quad (A \wedge B)' \wedge (A \wedge B'')' = A'.$$

Weiter ist $'$ fixpunktfrei, d.h. es gilt

$$(7) \quad A' \neq A \quad \text{für } A \in \mathcal{A}.$$

Beweis. Für $A \in \mathcal{A}$ gilt

$$A = A \wedge I = A \wedge (A \vee A') = (A \wedge A) \vee (A \wedge A') = (A \wedge A) \vee O = A \wedge A$$

und damit auch $A \wedge O = A \wedge (A \wedge A') = (A \wedge A) \wedge A' = A \wedge A' = O$, zusammen also die erste Hälfte von (1). Daraus folgt die zweite Hälfte wegen A.6.

(2) ergibt sich aus A.1(1) mit $A := I$ in der ersten Identität und $A := O$ in der zweiten, unter Ausnutzung der jeweiligen Neutralität.

In (3) gilt “ \Rightarrow ” nach A.1(1). Für die Rückrichtung bemerken wir, dass für $A, C_1, C_2 \in \mathcal{A}$ mit $A \wedge C_j = O$ und $A \vee C_j = I$ für $j \in \{1, 2\}$ sich

$$\begin{aligned} C_1 &= C_1 \wedge I = C_1 \wedge (A \vee C_2) = (C_1 \wedge A) \vee (C_1 \wedge C_2) = O \vee (C_1 \wedge C_2) \\ &= C_1 \wedge C_2 \end{aligned}$$

und mit Rollentausch und Kommutativität auch $C_2 = C_1 \wedge C_2$ ergibt, also $C_1 = C_2$, und dies angewandt auf $C_1 := A'$, siehe A.1(1), und $C_2 := B$, siehe R.S.(3), liefert das gewünschte.

Mit (3) angewandt auf (A', A) statt (A, B) im ersten Schritt und mit der Kommutativität von \wedge und \vee im zweiten ergibt sich für $A \in \mathcal{A}$ die Implikationskette

$$\begin{aligned} A = (A')' &\Leftrightarrow A' \wedge A = O \text{ und } A' \vee A = I \\ &\Leftrightarrow A \wedge A' = O \text{ und } A \vee A' = I \Leftrightarrow \text{A.1(1)}, \end{aligned}$$

also (4). (Warum im letzten Schritt obiger Implikationskette nicht auch \Leftrightarrow statt \Leftarrow ? Weil links A fixiert, rechts dagegen jedquantisiert ist.)

Zum Nachweis von (5) für $A, B \in \mathcal{A}$ betrachten wir $C := A' \vee B'$. Mit jeweils der Distributivität im ersten Schritt und einer Absorbtionseigenschaft aus (1) im dritten erhalten wir

$$\begin{aligned}(A \wedge B) \wedge C &= (A \wedge B \wedge A') \vee (A \wedge B \wedge B') = (B \wedge O) \vee (A \wedge O) = O \vee O = O, \\ (A \wedge B) \vee C &= (A \vee A' \vee B') \wedge (B \vee A' \vee B') = (I \vee B') \wedge (I \vee A') = I \wedge I = I,\end{aligned}$$

und damit nach (3) angewandt auf $A \wedge B$ und C statt A und B also die erste Gleichung in (5). Daraus folgt die zweite wegen A.6.

Mit De Morgan im ersten Schritt erhalten wir schließlich für $A, B \in \mathcal{A}$

$$\begin{aligned}(A \wedge B)' \wedge (A \wedge B)' &= (A' \vee B') \wedge (A' \vee B') = (A' \vee B') \wedge (A' \vee B) \\ &= A' \vee (B' \wedge B) = A' \vee O = A';\end{aligned}$$

also gilt auch die Johnson-Identität (6).

Zu (7): Aus $A = A'$ würde sich mit (3) angewandt auf $B := A$ sowohl $A \wedge A = O$ als auch $A \vee A = I$ ergeben, mit der Idempotenz der beiden binären Verknüpfungen also $O = A = I$, im Widerspruch zu Definition A.1. \square

A.9 Notationskonventionen. (a) Ist von einer Boole-Algebra \mathcal{A} (oder $\mathcal{B}, \mathcal{C}, \dots$) ohne Benennung der Verknüpfungen die Rede, so sollen diese in der Regel wie oben mit $\wedge, \vee, ', O, I$ bezeichnet werden. (Ausnahmen: Mengenalgebren, dort $\cap, \cup, ^c, \emptyset$, Grundmenge.)

(b) Beim Rechnen in einer Boole-Algebra lassen wir das Schnittzeichen \wedge oft weg und verwenden neben der Klammervermeidungskonvention “unär geht vor binär” und damit “’ vor \wedge und vor \vee ” dann auch “ \wedge vor \vee ”, also zum Beispiel

$$AB := A \wedge B, \quad AB'C := A \wedge (B') \wedge C, \quad A \vee BC := A \vee (B \wedge C).$$

Analog unten in der (zunächst) allgemeineren Situation aus Satz A.11.

(c) Wie in jedem kommutativen Monoid ist für (\mathcal{A}, \vee, O) und jede endliche Familie⁵ $A_\bullet = (A_j)_{j \in J}$ in \mathcal{A} deren (Boole-)Vereinigung $\vee A_\bullet = \vee_{j \in J} A_j$ per Induktion über $\#J$ erklärt: $\vee_{j \in \emptyset} A_j := O$, $\vee_{j \in J} A_j := \left(\vee_{j \in J \setminus \{k\}} A_j \right) \vee A_k$ für $\#J \in \mathbb{N}$ und $k \in J$. Siehe B.1 oder gleich Lang (2002, Chapter I, §1) für eine präzisere Ausführung der allgemeinen Definition. Diese angewandt auf das Monoid (\mathcal{A}, \wedge, I) ergibt für den (Boole-)Schnitt $\wedge A_\bullet = \wedge_{j \in J} A_j$ analog $\wedge_{j \in \emptyset} A_j := I$ und $\wedge_{j \in J} A_j := \left(\wedge_{j \in J \setminus \{k\}} A_j \right) \wedge A_k$ für $\#J \in \mathbb{N}$ und $k \in J$. \square

⁵Eine Familie, d.h. eine Funktion, also eine Menge von geordneten Paaren, ist genau dann endlich, wenn ihre Indexmenge es ist.

A.10 Aufgabe. Es sei \mathcal{A} eine Boole-Algebra. Dann gilt für $A, B \in \mathcal{A}$

- (1) $A = B \Leftrightarrow A'B = O \text{ und } A' \vee B = I,$
 (2) $A(A \vee B) = A = A \vee AB$ (**Absorbtionsgesetz**).

(*Hinweis:* Zum Nachweis einer vielleicht weniger offensichtlichen Identität in (2) mag die Äquivalenz (1) nützlich sein.) \square

In der von uns in Kapitel 1 mit Definition 1.6 angestrebten booleschen Analyse des Aussagenkalküls wird der folgende Satz A.11 schließlich in Satz 1.11(a) auf $(\mathcal{A}, \wedge, ')$ aus Satz 1.7 angewandt.

A.11 Satz (Johnson 1892, Huntington 1933)^{6,7} *Es sei \mathcal{A} eine mindestens zweielementige Menge mit einer assoziativen und kommutativen binären Verknüpfung \wedge und einer involutiven unären Verknüpfung $'$. Es gelte die Johnson-Identität A.8(6). Dann gibt es genau ein Tripel (\vee, O, I) derart, dass $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra ist, nämlich (\vee, O, I) definiert durch*

- (1) $A \vee B := (A' \wedge B)'$ für $A, B \in \mathcal{A},$
 (2) $O := A \wedge A'$ und $I := O'$ mit einem beliebigen $A \in \mathcal{A}.$

Beweis. 1. Eindeutigkeit: Ist (\vee, O, I) derart, dass $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra ist, so gelten die Gleichungen in (1) und (2) ohne Definitionsdoppelpunkte, wegen A.1(1) und A.8(2,4,5).

Es bleibt also nur noch zu zeigen, dass O und I durch (2) wohldefiniert sind, und dass damit und mit (1) tatsächlich $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra ist.

2. Wohldefiniertheit von O und I : Mit der Konvention A.9(b) wird die Johnson-Identität A.8(6) zu

$$(3) \quad (AB)'(AB) = A' \quad \text{für } A, B \in \mathcal{A}.$$

Für $A, B \in \mathcal{A}$ gilt, wegen $A'' = A$ und der Kommutativität im ersten Schritt, der Johnson-Identität (3) angewandt auf (A, B) und auf (A', B) im zweiten,

⁶Bei Johnson (1892, p. 10) steht noch die Zusatzforderung der Idempotenz, $A \wedge A = A$ für $A \in \mathcal{A}$. Huntington (1933, pp. 557–558) beweist deren Überflüssigkeit sowie die Ersetzbarkeit der beiden Gleichungen $A'' = A$ und (3) durch die offenbar daraus folgende eine Gleichung $(A'B)'(A'B) = A$; diese ist jedoch wohl kaum leichter zu motivieren als die beiden einzelnen, daher die hiesige Formulierung von Satz A.11 statt Huntingtons formal schärferer. Meredith/Prior (1968, pp. 213, 226) erwähnen besagte Resultate, Huntington zitiert Johnson nicht.

⁷Anders als in Definition A.1 kann hier die Assoziativitätsbedingung an \wedge nicht wie in A.2 ersatzlos gestrichen werden, da nach Diamond und McKinsey (1947) jede Charakterisierung nur durch Gleichungen des Begriffs “wie Boole-Algebra, aber eventuell doch einelementig oder leer”, hier “ $A(BC) = (AB)C, AB = BA, A'' = A, A.8(6)$ ”, eine Gleichung mit mindestens drei Variablen enthalten muss; im vorliegenden Fall kann danach die Assoziativität nicht gestrichen werden.

der Kommutativität und der Assoziativität in den Formen $(ab)(cd) = (ac)(bd)$ und $\alpha\beta = \beta\alpha$ im dritten, sowie (3) angewandt auf (B, A) und auf (B', A) im vierten,

$$\begin{aligned} AA' &= A'A'' = ((AB)'(AB')')((A'B)'(A'B')') \\ &= ((BA)'(BA')')((B'A)'(B'A')') = B'B'' = BB'. \end{aligned}$$

Also ist O , und damit auch I , wohldefiniert, mit

$$(4) \quad AA' = O \quad \text{für jedes } A \in \mathcal{A},$$

$$(5) \quad OO' = O.$$

3. Assoziativität und Kommutativität von \vee : Für $A, B, C \in \mathcal{A}$ liefert die viermalige Anwendung von (1), zusammen mit der Involutivität von $'$ und der Assoziativität von \wedge ,

$$\begin{aligned} (A \vee B) \vee C &= (A' \wedge B')' \vee C = ((A' \wedge B') \wedge C')' \\ &= (A' \wedge (B' \wedge C'))' = A \vee (B' \wedge C')' = A \vee (B \vee C); \end{aligned}$$

also ist auch \vee assoziativ. Die Kommutativität von \vee ist offensichtlicher.

4. Neutralität von I und O , A.1(1): Mit (3), angewandt auf $A := B := O$ im ersten Schritt, und mit (5) im zweiten, erhalten wir

$$(6) \quad O' = (OO)'(OO')' = (OO)'O'.$$

Mit (5) im ersten und im vierten Schritt, sowie mit (6) im zweiten und Kommutativität und Assoziativität im dritten, folgt

$$(7) \quad O = OO' = O((OO)'O') = (OO')(OO)' = O(OO)'.$$

Mit (7) im ersten Schritt, Assoziativität im zweiten, und (4) im dritten folgt

$$(8) \quad OO = O(O(OO)') = (OO)(OO)' = O.$$

Einsetzen von (8) in (6) liefert

$$O' = O'O'$$

und mit (4) folglich

$$(9) \quad (A'A)' = (A'A)'(A'A)' \quad \text{für } A \in \mathcal{A}.$$

Also erhalten wir für $A \in \mathcal{A}$ mit (4) im zweiten Schritt, (3) angewandt auf (A', A) im dritten und sechsten, Assoziativität und Kommutativität im vierten, und (9) im fünften,

$$\begin{aligned} AI &= A''O' = A''(A'A)' = ((A'A)'(A'A')')(A'A)' \\ &= ((A'A)'(A'A')')(A'A)' = (A'A)'(A'A)' = A'' = A. \end{aligned}$$

Damit gilt für $A \in \mathcal{A}$ auch $A \vee O = (A'I)' = A'' = A$, und zusammen ist also I neutral für \wedge , und O neutral für \vee .

Weiter gilt für $A \in \mathcal{A}$ mit (4) im dritten Schritt

$$A \vee A' = (A'A'')' = (A'A)' = O' = I,$$

also gilt A.1(1).

5. $O \neq I$: Für $A \in \mathcal{A}$ gilt mit (3) angewandt auf $B := A$ im ersten Schritt

$$A' = (AA)'(AA')' = (AA)'O' = (AA)'I = (AA)',$$

und mit der Involutivität von $'$ folgt die Idempotenz

$$(10) \quad AA = A \quad \text{für } A \in \mathcal{A}.$$

Damit ergibt sich $O \neq I$, denn sonst erhielten wir mit (10) im vierten Schritt

$$A = AI = AO = AAA' = AA' = O \quad \text{für } A \in \mathcal{A},$$

im Widerspruch zur Voraussetzung $\#\mathcal{A} \geq 2$.

6. Distributivität: Mit (3) im ersten und im dritten Schritt und mit (10) im zweiten erhalten wir

$$(11) \quad A' = (AB)'(AB')' = (AB)'(AB)'(AB')' = (AB)'A' \quad \text{für } A, B \in \mathcal{A},$$

mit (B', A) statt (A, B) also

$$(12) \quad B = (B'A)'B \quad \text{für } A, B \in \mathcal{A}.$$

Mit (3) im ersten Schritt und mit (12) im zweiten erhalten wir

$$(13) \quad A'B = (AB)'(AB')'B = (AB)'B \quad \text{für } A, B \in \mathcal{A}.$$

Damit erhalten wir für $A, B, C \in \mathcal{A}$ mit (3) im ersten und im vierten Schritt, (13) und (11) im zweiten, und (11) im dritten

$$(14) \quad \begin{aligned} (A(BC)')' &= (A(BC)'C)'(A(BC)'C')' = (AB'C)'(AC')' \\ &= (AB'C)'(AC'B')'(AC')' = (AB')'(AC')'. \end{aligned}$$

Durch Anwendung der Involution $'$ und Ersetzung von B, C durch B', C' , sowie schließlich mit Definition (1), ergibt sich

$$A(B \vee C) = AB \vee AC \quad \text{für } A, B, C \in \mathcal{A},$$

also die erste Hälfte von A.1(2), und die zweite folgt daraus durch Ersetzung von A, B, C durch A', B', C' und Anwendung von $'$. \square

A.12 Aufgabe. Studieren Sie den in der Vorlesung nicht vorgetragenen Beweisschritt 6 zu Satz A.11, und führen Sie ihn beginnend mit A.11(14) detaillierter aus. \square

Im Satz von Johnson-Huntington ist eine Boole-Algebra, also eine Menge mit fünf Verknüpfungen wenn man die Konstanten O und I als “nulläre Verknüpfungen” auffasst, schon durch ein Tripel $(\mathcal{A}, \wedge, ')$ mit nur zwei Verknüpfungen, und auch mit weniger postulierten Identitäten, gegeben. Man fragt sich nun natürlich, ob man diese nicht auch durch andere möglichst kleine Mengen von Verknüpfungen bzw. Identitäten ersetzen kann. Die Aufgaben A.21 und A.22 betreffen zwei überraschende Antworten.

Ähnlich wie man nach Aufgabe A.21 eine Boole-Algebra als einen Ring mit gewissen Zusatzeigenschaften auffassen kann, so kann man eine Boole-Algebra auch als eine (partiell) geordnete Menge mit gewissen Zusatzeigenschaften auffassen (“**boolescher Verband**”). Hierzu betrachten wir in der folgenden Aufgabe jedoch nur die sich in jeder Boole-Algebra auf natürliche Weise ergebende Ordnung:

A.13 Aufgabe. Es sei $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra. Auf \mathcal{A} sei die (binäre) Relation \leq definiert durch

$$A \leq B \quad :\Leftrightarrow \quad A \wedge B = A.$$

(a) (\mathcal{A}, \leq) ist eine (partiell) geordnete Menge, mit kleinstem Element O und größtem Element I .

(b) Für $A, B \in \mathcal{A}$ gilt die Äquivalenzkette

$$A \leq B \quad \Leftrightarrow \quad A \vee B = B \quad \Leftrightarrow \quad A' \geq B'.$$

(Hinweis: Absorbtionsgesetz A.10(2).)

(c) Die binären Verknüpfungen \wedge und \vee sind isoton, d.h. für $A_1, A_2, B_1, B_2 \in \mathcal{A}$ mit $A_1 \leq A_2$ und $B_1 \leq B_2$ gilt $A_1 \wedge B_1 \leq A_2 \wedge B_2$ und $A_1 \vee B_1 \leq A_2 \vee B_2$.

(d) Zu $A, B \in \mathcal{A}$ existieren bezüglich \leq stets

$$\begin{aligned} \inf\{A, B\} &:= \text{größte untere Schranke von } \{A, B\}, \\ \sup\{A, B\} &:= \text{kleinste obere Schranke von } \{A, B\}, \end{aligned}$$

d.h. die geordnete Menge (\mathcal{A}, \leq) ist ein **Verband**. \square

Wird bei einer gegebenen Boole-Algebra kommentarlos auf eine Ordnung Bezug genommen, wie zum Beispiel in A.15(g), oder “ \leq ” geschrieben, so ist damit die Ordnung \leq aus obiger Aufgabe A.13 gemeint.

Wie etwa aus der Betrachtung von Gruppen oder Ringen bekannt, definiert man bei jeder “Spezies algebraischer Strukturen” Homomorphismen zwischen zwei “Individuen” als Funktionen φ von der ersten Trägermenge in die zweite derart, dass

sich alle in der Speziesdefinition vorkommenden Verknüpfungen und Konstanten “mittels φ übertragen”.⁸ Für Boole-Algebren ausformuliert, vorsichtshalber einmal pingelig ohne den in A.9(a) genannten Notationsmissbrauch, ergibt das:

A.14 Definition. Es seien $\boxed{\mathcal{A}} = (\mathcal{A}, \wedge, \vee, ', O, I)$ und⁹ $\boxed{\mathcal{B}} = (\mathcal{B}, \sqcap, \sqcup, *, Z, U)$ Boole-Algebren. Eine Funktion $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ mit

- (1) $\varphi(A \wedge B) = \varphi(A) \sqcap \varphi(B),$
- (2) $\varphi(A \vee B) = \varphi(A) \sqcup \varphi(B),$
- (3) $\varphi(A') = \varphi(A)^*,$
- (4) $\varphi(O) = Z,$
- (5) $\varphi(I) = U$

für $A, B \in \mathcal{A}$ heißt **(Boole-Algebra-Homo)-morphismus** von $\boxed{\mathcal{A}}$ nach $\boxed{\mathcal{B}}$; sie heißt **Isomorphismus** falls es einen Morphismus ψ von $\boxed{\mathcal{B}}$ nach $\boxed{\mathcal{A}}$ mit $\psi \circ \varphi = \text{id}_{\mathcal{A}}$ und $\varphi \circ \psi = \text{id}_{\mathcal{B}}$ gibt, und **Automorphismus** falls darüber hinaus $\boxed{\mathcal{A}} = \boxed{\mathcal{B}}$ ist. \square

Statt “ $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ Morphismus von $\boxed{\mathcal{A}}$ nach $\boxed{\mathcal{B}}$ ” schreibt man auch “ $\varphi : \boxed{\mathcal{A}} \rightarrow \boxed{\mathcal{B}}$ Morphismus”, und dies auch mit den Abkürzungen \mathcal{A}, \mathcal{B} statt $\boxed{\mathcal{A}}, \boxed{\mathcal{B}}$.

A.15 Aufgabe. (a) Definition A.14 bleibt äquivalent wenn von den dortigen Zeilen (1-5) eine der ersten beiden und die beiden letzten gestrichen werden.

(b) Ein Boole-Algebren-Homomorphismus ist genau dann ein Isomorphismus, wenn er eine Bijektion der betreffenden Trägermengen ist.

(c) Es seien \mathcal{X} und \mathcal{Y} zwei nichtleere Mengen und es sei $f : \mathcal{Y} \rightarrow \mathcal{X}$ eine Funktion. Dann ist die Urbildabbildung $\varphi := f^{-1}[\cdot]$ ein Morphismus von der Potenzmengenalgebra $2^{\mathcal{X}}$ in die Potenzmengenalgebra $2^{\mathcal{Y}}$. Für welche f ist φ ein Isomorphismus?

(d) Es seien \mathcal{A} und \mathcal{B} Mengenalgebren auf den nichtleeren Mengen \mathcal{X} bzw. \mathcal{Y} , und es sei $f : \mathcal{Y} \rightarrow \mathcal{X}$ eine Funktion mit $f^{-1}[A] \in \mathcal{B}$ für $A \in \mathcal{A}$. Dann ist offenbar in Verallgemeinerung der Aussage in (c) auch hier $\varphi := f^{-1}[\cdot]$ ein Morphismus von \mathcal{A} nach \mathcal{B} . Ist hier für bijektives f notwendig φ ein Isomorphismus?

(e) Es sei $(\mathcal{A}, \wedge, \vee, ', O, I)$ ein Boole-Algebra. Ist dann $'$ ein Isomorphismus? Also ein Automorphismus?

(f) Pingelig aufgeschrieben: Für jede Boole-Algebra $\boxed{\mathcal{A}} = (\mathcal{A}, \wedge, \vee, ', O, I)$ ist die Identität $\text{id}_{\mathcal{A}}$ ein Automorphismus von $\boxed{\mathcal{A}}$.

Mit der Schreibweise hinter Definition A.14: Sind $\boxed{\mathcal{A}}, \boxed{\mathcal{B}}, \boxed{\mathcal{C}}$ Boole-Algebren

⁸Eine präzisere Formulierung, die zum Beispiel Gruppen, Ringe, Boole-Algebren erfasst, liefert die Universelle Algebra, siehe dazu etwa Brandenburg (2017, Kapitel 4).

⁹ Z wie zero, U wie Universum oder engl. *unit*.

und sind $\varphi : \boxed{\mathcal{A}} \rightarrow \boxed{\mathcal{B}}$ und $\psi : \boxed{\mathcal{B}} \rightarrow \boxed{\mathcal{C}}$ Morphismen, so ist auch $\psi \circ \varphi : \boxed{\mathcal{A}} \rightarrow \boxed{\mathcal{C}}$ ein Morphismus.¹⁰

(g) Jeder Morphismus von Boole-Algebren ist isoton.

(h) Zwei Elemente A, B einer Boole-Algebra heißen zueinander

(1) **disjunkt**¹¹ $:\Leftrightarrow A \wedge B = O$,

(2) **komplementär** $:\Leftrightarrow A \wedge B = O$ und $A \vee B = I$.

Jeder Morphismus von Boole-Algebren bildet disjunkte auf disjunkte Elemente ab, und komplementäre auf komplementäre.

Gilt eine analoge Aussage für nicht disjunkte Elemente? □

Wir geben noch ohne Beweis zwei Darstellungssätze für Boole-Algebren an, welche in dieser Vorlesung zwar nirgends benutzt werden, aber doch zum Teil erklären, warum wir in konkreten Fällen meist Mengenalgebren betrachten.

A.16 Satz (Huntington 1904). *Jede endliche Boole-Algebra ist isomorph zu einer Potenzmengenalgebra.*

Beweis. Siehe etwa Rényi (1979, Kapitel 1, §4) oder, in allgemeinerem Kontext, Givant/Halmos (2009, Chapter 15). □

A.17 Satz (Stone 1936). *Jede Boole-Algebra ist isomorph zu einer Mengenalgebra.*

Beweis. Siehe etwa Rényi (1979, Kapitel 1, §5) oder Givant/Halmos (2009, p. 190). □

Eine Benutzung von Satz A.17 erlaubt es, sich beim Lösen von Aufgaben wie KV 3 auf p. 178 auf den Fall von Mengenalgebren zu beschränken.

A.18 Aufgabe. (a) Ist in der Situation aus Satz 1.11(a) die Boole-Algebra S^*/\equiv schon eine Mengenalgebra?

(b) Geben Sie Isomorphismen der Mengenalgebren aus Aufgabe A.4(b) auf Potenzmengenalgebren an.

(c) Geben Sie einen Isomorphismus von T_{30} aus Aufgabe A.7 auf eine Mengenalgebra an. □

¹⁰In der Terminologie der Kategorientheorie, siehe etwa Brandenburg (2017) oder Herrlich und Strecker (2007), ergeben beide Aussagen zusammen: Die Klasse der Boole-Algebren als Objekte zusammen mit den Boole-Algebra-Homomorphismen als Morphismen, und mit den gewöhnlichen Hintereinanderausführungen von Funktionen als Verknüpfungen, bilden eine Kategorie.

¹¹Genauer wäre **Boole-disjunkt**. Denn in der Situation von Satz 1.11(a) sind je zwei verschiedene $A, B \in S^*/\equiv$ als Äquivalenzklassen zwar disjunkte Mengen, aber zum Beispiel sind die Äquivalenzklassen zweier verschiedener Elementaraussagen stets verschieden nach Korollar 1.14, nach 1.17(2) aber nicht Boole-disjunkt.

A.19 Aufgabe. Es sei $\mathcal{A} := \{A \subseteq \mathbb{N} : A \text{ endlich oder } A^c \text{ endlich}\}$.

(a) \mathcal{A} ist eine Mengenalgebra.

(b) \mathcal{A} ist nicht isomorph zu einer Potenzmengenalgebra.

Hinweis: Wieviele Elemente hat \mathcal{A} ? □

A.20 Aufgabe. Es sei jeweils $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra.

(a) Eine Menge $\mathcal{B} \subseteq \mathcal{A}$ heißt **Unter-(Boole-)Algebra** von \mathcal{A} falls $(\mathcal{B}, \wedge|_{\mathcal{B} \times \mathcal{B}}, \vee|_{\mathcal{B} \times \mathcal{B}}, '|_{\mathcal{B}}, O, I)$ eine Boole-Algebra ist. Dies gilt offenbar¹² genau dann, wenn wir

$$(1) \quad A \wedge B, A \vee B, A', O, I \in \mathcal{B} \quad \text{für } A, B \in \mathcal{B}$$

haben, und nach de Morgan kann offenbar ähnlich zu Aufgabe A.15(a) in der Bedingung (1) einer der ersten beiden und einer der letzten beiden der fünf zu betrachtenden Ausdrücke gestrichen werden.

(b) Für jede Menge $\mathcal{E} \subseteq \mathcal{A}$ ist

$$\alpha(\mathcal{E}) := \bigcap \{\mathcal{B} : \mathcal{B} \supseteq \mathcal{E}, \mathcal{B} \text{ Unteralgebra von } \mathcal{A}\}$$

eine Unteralgebra von \mathcal{A} , die von \mathcal{E} **erzeugte Unteralgebra**.

Es sei nun $\mathcal{E} \subseteq \mathcal{A}$ zweielementig. Wie viele Elemente hat dann $\alpha(\mathcal{E})$ höchstens?

(c) Es sei \mathcal{B} eine weitere Boole-Algebra, und es sei $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ ein Morphismus. Dann ist dessen Bild $\varphi[\mathcal{A}] = \{\varphi(A) : A \in \mathcal{A}\}$ offenbar eine Unteralgebra von \mathcal{B} . □

¹²Das Adjektiv “**offenbar**” in einer Aufgabe signalisiert, dass die betreffende Teilaussage im Falle ihrer Richtigkeit in der Lösung nicht diskutiert zu werden braucht. Hier braucht also für Teil (a), falls der richtig ist, nichts getan zu werden.

Weitere Aufgaben¹³

A.21 Aufgabe. Bekanntlich ist ein **Ring** ein Sextupel¹⁴ $R = (R, +, \cdot, (-), 0, 1)$ mit $0 \neq 1$ derart^{15,16}, dass $(R, +, (-), 0)$ eine Gruppe ist, $(R, \cdot, 1)$ ein Monoid, und die **Multiplikation** \cdot distributiv über der **Addition** $+$ ist:

$$x \cdot (y + z) = xy + xz \quad \text{und} \quad (y + z) \cdot x = yx + zx \quad \text{für } x, y, z \in R.$$

Ein Ring heißt **kommutativ** bzw. **idempotent** falls seine Multiplikation dies ist. Idempotente Ringe heißen auch **boolesch**.

(a) Es sei $(R, +, \cdot, (-), 0, 1)$ ein idempotenter Ring. Dann ist R kommutativ, und die Negation $(-)$ ist die Identität id_R , d.h. es gilt

$$(1) \quad x + x = 0 \quad \text{für } x \in R.$$

(b) Es sei $(R, +, \cdot, \text{id}_R, 0, 1)$ ein boolescher Ring. Mit

$$\begin{aligned} x \vee y &:= x + y + xy \quad \text{für } x, y \in R, \\ x' &:= 1 + x \quad \text{für } x \in R \end{aligned}$$

ist dann $(R, \cdot, \vee, ', 0, 1)$ eine Boole-Algebra.

(c) Es sei $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra. Mit der (**booleschen**) **symmetrischen Differenz**

$$A \Delta B := AB' \vee BA' \quad \text{für } A, B \in \mathcal{A}$$

als Addition ist dann $(\mathcal{A}, \Delta, \wedge, \text{id}_R, O, I)$ ein boolescher Ring.

(d) Die mit (b) und (c) gegebenen Abbildungen von der Klasse **BooleRing** aller booleschen Ringe in die Klasse **BooleAlg** aller Boole-Algebren bzw. umgekehrt sind zueinander inverse Bijektionen. \square

¹³Deren Definitionen und Resultate im weiteren Vorlesungsverlauf wohl kaum vorkommen.

¹⁴Üblicher Notationsmissbrauch analog zu A.9(a): Offiziell gehören bei Ihrem Dozenten alle in den Axiomen auftauchenden Verknüpfungen oder Konstanten in das Tupel, also hier auch die **Negation** $(-)$, d.h. die Inversenbildung bezüglich $+$; abkürzend können sie weggelassen werden.

¹⁵Die oft noch zusätzlich geforderte Kommutativität der Addition ergibt sich aus den übrigen Annahmen, indem man auf $(1+1)(x+y)$ auf zwei Weisen erst die eine und dann die andere Hälfte der Distributivität anwendet, und bei der sich ergebenden Identität x links- und y rechtskürzt. Diese Tatsache ist mindestens seit Hilbert (1900, p. 183) bekannt, wie man aus Hilbert (1977, p. 243) lernt.

¹⁶Es gibt (mindestens) zwei dazu und untereinander inäquivalente Definitionsvarianten, nämlich erstens den Verzicht auf die Forderung $0 \neq 1$ und damit genau die “Nullringe” mit $R = \{0\}$ zusätzlich zulassend, und zweitens den Verzicht auf die 1, aber die Kommutativität von $+$ fordernd, also schon einen **Rng** $(R, +, \cdot, (-), 0)$ (kein Tippfehler!), mit (R, \cdot) Halbgruppe aber nicht notwendig Monoid, als Ring bezeichnend.

A.22 Aufgabe. Sheffer (1913) zeigte¹⁷, dass sich Boole-Algebren schon mit nur einer Verknüpfung beschreiben lassen:

(a) Es sei $(\mathcal{A}, \wedge, \vee, ', O, I)$ eine Boole-Algebra, und damit

$$A|B := A' \wedge B' \quad \text{für } A, B \in \mathcal{A}.$$

Dann lassen sich $\wedge, \vee, ', O, I$ aus der binären Verknüpfung $|$ rekonstruieren.

(b) In der Situation aus Teil (a) gelten für $A, B, C \in \mathcal{A}$ die Identitäten von Sheffer (1913)

- (1) $(A|A)|(A|A) = A,$
- (2) $A|(B|(B|B)) = (A|A),$
- (3) $(A|(B|C))|(A|(B|C)) = ((B|B)|A)|((C|C)|A),$

von Veroff (2003)

- (4) $(A|B)|(A|(B|C)) = A,$
- (5) $A|B = B|A,$

und von McCune et al. (2002)

- (6) $(A|((B|A)|A))|(B|(C|A)) = B.$

(c) Sheffer (1913), Veroff (2003), und McCune et al. (2002) zeigten umgekehrt jeweils: Ist \mathcal{A} eine mindestens zweielementige Menge mit einer binären Verknüpfung $|$ welche die entsprechenden obigen Identitäten erfüllt, so ist $(\mathcal{A}, \wedge, \vee, ', O, I)$ mit den von Ihnen in Teil (a) angegebenen $\wedge, \vee, ', O, I$ eine Boole-Algebra.

Welches Hilfsmittel benutzten Veroff (2003) und McCune et al. (2002) um ihre gegenüber Sheffer (1913) verbesserten Resultate zu finden und zu beweisen? Hätte Huntington das nicht in seinen Arbeiten zu Boole-Algebren miterledigen können?

(*Hinweis:* Dieser Teil ist eine Literaturrecherche- und Diagonalleseaufgabe. Die Arbeit von Sheffer (1913) steht vorne auf Seite [xiii](#); die absichtlich dort nicht angegebenen Arbeiten von Veroff (2003) und McCune et al. (2002) findet man damit leicht unter Verwendung von Google Scholar. Als Antwort reichen sehr wenige Zeilen.) \square

Wer nun auf den Geschmack für Sätze in der Art von [A.2](#), [A.8](#), [A.11](#), [A.22\(c\)](#) gekommen sein sollte, kann mit Padmanabhan/Rudeanu (2008) einige amüsante Stunden verbringen.

¹⁷Und vor ihm Peirce (1880), laut Hilbert/Bernays (1968, p. 48).

A.23 Aufgabe. (a) Es sei \mathcal{A} eine Boole-Algebra und es sei $\mathcal{F} \subseteq \mathcal{A}$. Dann sind die folgenden drei Bedingungen äquivalent:

(i) Für $A, B \in \mathcal{A}$ gilt die Äquivalenz

$$(1) \quad A, B \in \mathcal{F} \Leftrightarrow A \wedge B \in \mathcal{F}.$$

(ii) Für $A, B \in \mathcal{A}$ gelten die Implikationen

$$(2) \quad A, B \in \mathcal{F} \Rightarrow A \wedge B \in \mathcal{F},$$

$$(3) \quad A \in \mathcal{F}, A \leq B \Rightarrow B \in \mathcal{F}.$$

(iii) Für $A, B \in \mathcal{A}$ gelten die Implikationen (2) und

$$(4) \quad A \in \mathcal{F} \Rightarrow A \vee B \in \mathcal{F}.$$

(b) Es gelte eine, und damit jede, der Bedingungen aus (a). Dann gelten die Äquivalenzen

$$(5) \quad \mathcal{F} \neq \emptyset \Leftrightarrow I \in \mathcal{F},$$

$$(6) \quad \mathcal{F} \neq \mathcal{A} \Leftrightarrow O \notin \mathcal{F}.$$

Gilt nun $\mathcal{F} \neq \emptyset$, so heißt \mathcal{F} **Filter** in \mathcal{A} ; gilt darüber hinaus $O \notin \mathcal{F}$, so heißt \mathcal{F} **eigentlich**. \square

A.24 Aufgabe. Es sei \mathcal{A} eine Boole-Algebra. Ein bezüglich Mengeninklusion maximaler eigentlicher Filter in \mathcal{A} , d.h. ein eigentlicher Filter \mathcal{F} in \mathcal{A} mit der Eigenschaft

$$\mathcal{G} \text{ eigentlicher Filter in } \mathcal{A}, \mathcal{F} \subseteq \mathcal{G} \Rightarrow \mathcal{F} = \mathcal{G},$$

heißt **Ultrafilter** in \mathcal{A} .

Es sei nun \mathcal{F} ein Filter in \mathcal{A} . Dann sind die folgenden vier Aussagen äquivalent:

(i) \mathcal{F} ist ein Ultrafilter in \mathcal{A} .

(ii) Für jedes $A \in \mathcal{A}$ gilt entweder $A \in \mathcal{F}$ oder $A' \in \mathcal{F}$.

(iii) \mathcal{F} ist eigentlich, und für jedes $A \in \mathcal{A}$ gilt $A \in \mathcal{F}$ oder $A' \in \mathcal{F}$.

(iv) Für jede endliche und paarweise disjunkte Familie¹⁸ $(A_j)_{j \in J}$ in \mathcal{A} gilt mit der Notation aus A.9(c) die Äquivalenz

$$\bigvee_{j \in J} A_j \in \mathcal{F} \Leftrightarrow \text{für genau ein } j \in J \text{ gilt } A_j \in \mathcal{F}.$$

\square

Wegen A.24(i) \Leftrightarrow (ii) sind die Realitäten gemäß Definition 1.20 genau die Ultrafilter der Definition in A.24.

¹⁸Eine Familie $A_\bullet = (A_j)_{j \in J}$ in einer Boole-Algebra heißt nach Aufgabe A.15(h) **paarweise disjunkt**, abgekürzt **p.d.**, wenn $A_j \wedge A_k = O$ für $(j, k) \in J_{\neq}^2$ gilt.

B Summen

In der Stochastik und anderswo brauchen wir immer wieder Summen mit unendlich vielen Summanden ohne ausgezeichnete Reihenfolge. Deren Theorie ab Definition B.9 ist das Ziel dieses Kapitels. Da wir auch ∞ und $-\infty$ als Summanden zulassen wollen, beginnen wir mit Summen mit nur endlich vielen Summanden in abelschen Monoiden, gefolgt von einer kurzen Einführung in die erweiterten reellen Zahlen.

B.1 Definition. Es sei $(H, +, 0)$ ein abelsches Monoid. Für $n \in \mathbb{N}_0$ und Familien $(x_i : i \in \{1, \dots, n\})$ in H definiert man induktiv

$$\sum_{i=1}^n x_i := \begin{cases} 0 & (n = 0) \\ \sum_{i=1}^{n-1} x_i + x_n & (n \geq 1) \end{cases}$$

Für endliche Familien $(x_i : i \in I)$ in H sei

$$(1) \quad \sum_{i \in I} x_i := \sum_{i=1}^n x_{\varphi(i)}$$

mit $n := \#I$ und einer Bijektion $\varphi : \{1, \dots, n\} \rightarrow I$. □

Wird, etwa im Beispiel des Monoids $(\mathbb{C}, \cdot, 1)$, die Gruppenverknüpfung multiplikativ geschrieben, so ist oben das Summenzeichen \sum jeweils durch das Produktzeichen \prod zu ersetzen.

Definition (1) benötigt einen Wohldefiniertheitsnachweis: Zu zeigen ist, dass die rechte Seite in (1) nicht von der Wahl von φ abhängt. Dies ist wegen der vorausgesetzten Kommutativität von H anschaulich klar und ergibt sich formal relativ leicht mittels Induktion über die Mächtigkeit von I , siehe etwa Lang (2002, p. 5). Die folgenden Rechenregeln sind ebenso anschaulich einleuchtend:

B.2 Satz. *Es sei $(H, +, 0)$ ein abelsches Monoid. Es seien I, J, A, I_α ($\alpha \in A$) endliche Mengen und es seien $(x_i : i \in I), \dots$ Familien in H . Dann gelten die Gleichungen beziehungsweise Implikationen*

- (1) $\sum_{i \in \emptyset} x_i = 0,$ (2) $\sum_{i \in I} 0 = 0,$ (3) $\sum_{i \in I} x_i = \sum_{i \in I, x_i \neq 0} x_i,$
- (4) $\varphi : J \rightarrow I$ Bijektion $\Rightarrow \sum_{i \in I} x_i = \sum_{j \in J} x_{\varphi(j)}$ **(Substitutionsregel,**
im Fall $I = J$ **Permutationsinvarianz),**
- (5) $\sum_{i \in I} (x_i + y_i) = \sum_{i \in I} x_i + \sum_{i \in I} y_i$ **(Additivität),**
- (6) $\sum_{(i,j) \in I \times J} x_{ij} = \sum_{i \in I} \sum_{j \in J} x_{ij} = \sum_{j \in J} \sum_{i \in I} x_{ij}$ **(Doppelsummensatz),**
- (7) I, J disjunkt $\Rightarrow \sum_{k \in I \cup J} x_k = \sum_{i \in I} x_i + \sum_{j \in J} x_j$ **(Additivität),**
- (8) $I = \bigcup_{\alpha \in A} I_\alpha, I_\alpha$ p.d. $\Rightarrow \sum_{i \in I} x_i = \sum_{\alpha \in A} \sum_{i \in I_\alpha} x_i$ **(Umordnungssatz).**

Beweis-Skizze: (1): Direkt nach den Definitionen ist $\sum_{i \in \emptyset} x_i = \sum_{i=1}^0 x_{\varphi(i)} = 0$, wobei hier $\varphi := \emptyset$ die leere Funktion ist.

(2): Es gilt $\sum_{i \in I} 0 = \sum_{i=1}^{\#(I)} 0 = 0$, wegen Definition B.1(1) und Induktion über $\#(I)$.

(4) ergibt sich aus der Wohldefiniiertheit von B.1(1): Ist $n := \#(I)$ und $\psi : \{1, \dots, n\} \rightarrow I$ eine Bijektion, so ist auch $\varphi^{-1} \circ \psi : \{1, \dots, n\} \rightarrow J$ eine Bijektion, und zweimalige Anwendung von B.1(1) liefert $\sum_{i \in I} x_i = \sum_{i=1}^n x_{\psi(i)} = \sum_{i=1}^n x_{\varphi(\varphi^{-1}(\psi(i)))} = \sum_{j \in J} x_{\varphi(j)}$.

(7) zeigt man mit Induktion über $\#J$.

(3): (7) mit $\{i \in I : x_i \neq 0\}$ und $\{i \in I : x_i = 0\}$ statt I und J , dann (2).

(8) folgt aus (7) mit Induktion über $\#A$.

(6) ist ein Spezialfall von (8): Man schreibt $I \times J = \bigcup_{i \in I} \{i\} \times J = \bigcup_{j \in J} I \times \{j\}$ und wendet (8) einmal mit $A := I$ und einmal mit $A := J$ an.

(5) folgt aus der zweiten Gleichheit in (6), mit $J := \{1, 2\}$ und $x_{ij} := x_i$ für $j = 1$, $x_{ij} := y_i$ für $j = 2$. □

Als Übungsaufgabe zu B.1 und B.2 kann man etwas Mengenrechnung üben:

B.3 Aufgabe. Es sei \mathcal{X} eine Menge. Mit der üblichen Definition

$$A \cup B := \{x : x \in A \text{ oder } x \in B\} \quad \text{für } A, B \in 2^{\mathcal{X}}$$

ist $(2^{\mathcal{X}}, \cup, \emptyset)$ eine abelsches Monoid. Folglich sind Definition B.1 und Satz B.2 anwendbar. Statt \sum schreiben wir vorübergehend \bigsqcup .

(a) Zeigen Sie: Für jede endliche Familie $(A_i : i \in I)$ in $2^{\mathcal{X}}$ gilt $\bigsqcup_{i \in I} A_i = \bigcup_{i \in I} A_i$ mit der üblichen Definition $\bigcup_{i \in I} A_i := \{x : \exists i \in I \text{ mit } x \in A_i\}$. Wir schreiben daher ab sofort \cup statt \bigsqcup .

(b) Welche der Rechenregeln B.2(1–8) vereinfachen sich im vorliegenden Spezialfall?

(c) Was ist das Analogon zu obigem mit \cap statt \cup ? □

B.4 Definition (Die erweiterten reellen Zahlen).

(a) **Die Menge $\overline{\mathbb{R}}$.** Man wählt irgend zwei verschiedene und nicht zu \mathbb{R} gehörige Objekte, nennt sie $-\infty$ und ∞ , und setzt

$$(1) \quad \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}.$$

Die Elemente von $\overline{\mathbb{R}}$ heißen **erweiterte reelle Zahlen**.

(b) **Anordnung auf $\overline{\mathbb{R}}$.** Für $a, b \in \overline{\mathbb{R}}$ definiert man

$$\begin{aligned} a \leq b &: \Leftrightarrow a, b \in \mathbb{R} \text{ und } a \leq b, \text{ oder } a = -\infty, \text{ oder } b = \infty, \\ a < b &: \Leftrightarrow a \leq b \text{ und } a \neq b, \\ [a, b] &:= \{x \in \overline{\mathbb{R}} : a \leq x \leq b\}, \\ [a, b[&:= \{x \in \overline{\mathbb{R}} : a \leq x < b\}, \\]a, b] &:= \{x \in \overline{\mathbb{R}} : a < x \leq b\}, \\]a, b[&:= \{x \in \overline{\mathbb{R}} : a < x < b\}. \end{aligned}$$

□

B.5 Satz. Jede Menge $A \subseteq \overline{\mathbb{R}}$ hat ein Supremum ($:=$ kleinste obere Schranke) und ein Infimum ($:=$ größte untere Schranke) in $\overline{\mathbb{R}}$. Bezeichnung: $\sup A$ beziehungsweise $\inf A$.

Beweis. Ist $A = \emptyset$ oder $A = \{-\infty\}$, so ist $\sup A = -\infty$. Ist $\emptyset \neq A \setminus \{-\infty\} \subseteq \mathbb{R}$ und $A \setminus \{-\infty\}$ nach oben beschränkt in \mathbb{R} , so ist $\sup A = \sup_{\mathbb{R}} A \setminus \{-\infty\}$, wobei rechts mit $\sup_{\mathbb{R}} \dots$ das bekannte Supremum in \mathbb{R} gemeint ist. Ist schließlich $\infty \in A$ oder ist $A \setminus \{-\infty\} \subseteq \mathbb{R}$ eine in \mathbb{R} nach oben unbeschränkte Menge, so ist $\sup A = \infty$. Für Infima statt Suprema ist alles analog. □

B.6 Bemerkung. Es sei nun speziell $A \subseteq \mathbb{R}$. Dann sind Aussagen wie “ A ist nach oben beschränkt” oder “ A hat ein Supremum” mehrdeutig: Beziehen wir uns dabei auf A als Teilmenge von \mathbb{R} oder als Teilmenge von $\overline{\mathbb{R}}$? Das müsste eigentlich immer dazu geschrieben werden. Meist meint man mit “ A ist nach oben beschränkt”, dass A in \mathbb{R} nach oben beschränkt sein soll, während $\sup A$ und $\inf A$ auch gleich ∞ oder $-\infty$ sein dürfen; mit dieser Konvention gilt dann

$$A \text{ nach oben beschränkt} \Leftrightarrow \sup A < \infty,$$

$$A \text{ nach unten beschränkt} \Leftrightarrow \inf A > -\infty,$$

$$A \text{ beschränkt} \Leftrightarrow A \text{ nach oben und nach unten beschränkt}$$

$$\Leftrightarrow \sup A \text{ und } \inf A \text{ endlich, oder (Trivialfall) } A = \emptyset.$$

□

B.7 Definition (Rechnen in $\overline{\mathbb{R}}$). Wir verwenden praktisch immer die folgenden allgemein üblichen Definitionen:

$$\begin{aligned} a + \infty &:= \infty + a := \infty && (a \in]-\infty, \infty]) \\ a - \infty &:= a + (-\infty) := -\infty + a := -\infty && (a \in [-\infty, \infty[) \\ -\infty + \infty \text{ und } \infty - \infty &&& \text{ bleiben undefiniert} \end{aligned}$$

$$\begin{aligned} a \cdot \infty &:= \infty \cdot a := \begin{cases} \infty & (a \in]0, \infty]) \\ -\infty & (a \in [-\infty, 0[) \end{cases} \\ a \cdot (-\infty) &:= (-\infty) \cdot a := \begin{cases} -\infty & (a \in]0, \infty]) \\ \infty & (a \in [-\infty, 0[) \end{cases} \\ \frac{a}{\infty} &:= \frac{a}{-\infty} := 0 && (a \in \mathbb{R}) \\ &&& \frac{a}{0} \text{ bleibt für } a \neq 0 \text{ undefiniert} \end{aligned}$$

Weiter verwenden wir fast immer die **Nullkonvention**

$$(1) \quad 0 \cdot \infty := \infty \cdot 0 := (-\infty) \cdot 0 := 0 \cdot (-\infty) := 0$$

und manchmal auch die **starke Nullkonvention**

$$\begin{aligned} 0 \cdot (\text{irgendwas, ob definiert oder nicht}) &:= (\text{irgendwas, ob definiert oder nicht}) \cdot 0 \\ (2) \qquad \qquad \qquad &:= 0 \end{aligned}$$

und dann zum Beispiel $\frac{0}{0} := 0 \cdot \frac{1}{0} = 0$. □

Mit den obigen Definitionen ist $+$ keine Verknüpfung auf $\overline{\mathbb{R}}$, da zum Beispiel $-\infty + \infty$ gar nicht definiert ist. Insbesondere ist $(\overline{\mathbb{R}}, +)$ keine Halbgruppe, erst recht also kein Monoid. Dagegen ist, wie man leicht nachweist, jede der Mengen $[-\infty, \infty[$, $] - \infty, \infty]$, $[-\infty, 0]$ und $[0, \infty]$ ein abelsches Monoid bezüglich (der jeweiligen Restriktion von) $+$, mit dem neutralem Element 0.

B.8 Satz (Summen in $[0, \infty]$ mit endlich vielen Summanden). *Es seien I, J, A, I_α ($\alpha \in A$) endliche Mengen, es seien $(x_i : i \in I)$, ... Familien in $[0, \infty]$, und es sei $c \in [0, \infty]$. Dann gelten die Gleichungen beziehungsweise Implikationen*

(1–8) wie B.2(1–8),

$$(9) \qquad \sum_{i \in I} cx_i = c \sum_{i \in I} x_i \quad (\text{Homogenität oder Distributivität}),$$

$$(10) \quad \left(\sum_{i \in I} x_i \right) \cdot \left(\sum_{j \in J} y_j \right) = \sum_{(i,j) \in I \times J} x_i y_j \quad (\text{Distributivität}),$$

$$(11) \quad x_i \leq y_i \text{ für } i \in I \Rightarrow \sum_{i \in I} x_i \leq \sum_{i \in I} y_i \quad (\text{Isotonie}),$$

$$(12) \qquad I \subseteq J \Rightarrow \sum_{i \in I} x_i \leq \sum_{j \in J} x_j \quad (\text{Isotonie}).$$

Beweis-Skizze: Die Rechenregeln (1–8) gelten nach Satz B.2 angewandt auf das abelsche Monoid $([0, \infty], +, 0)$.

(9): Induktion über $\#I$.

(10): Zwei mal (9), mit $c = \sum x_i$ beziehungsweise $c = y_j$.

(11): Induktion über $\#I$.

(12): Induktion über $\#(J \setminus I)$. □

B.9 Definition (Summen in $[0, \infty]$ mit beliebig vielen Summanden). Ist $(x_i : i \in I)$ eine Familie in $[0, \infty]$, so heißt

$$\sum_{i \in I} x_i := \sup \left\{ \sum_{i \in I_0} x_i : I_0 \subseteq I, I_0 \text{ endlich} \right\}$$

ihre **Summe**. □

B.10 Satz. *Es seien I, J, A, I_α ($\alpha \in A$) Mengen, es seien $(x_i : i \in I)$, ... Familien in $[0, \infty]$, und es sei $c \in [0, \infty]$. Dann gelten die Gleichungen beziehungsweise Implikationen*

(1–12) wie B.8(1–12).

Exemplarischer Beweis der Additivität (5): Es seien $(x_i : i \in I)$ und $(y_i : i \in I)$ Familien in $[0, \infty]$.

Für jede endliche Menge $I_0 \subseteq I$ gilt dann, wegen der Additivität B.2(5) für Summen mit endlich vielen Summanden im ersten Schritt, und Definition B.9 im zweiten

$$\sum_{i \in I_0} (x_i + y_i) = \sum_{i \in I_0} x_i + \sum_{i \in I_0} y_i \leq \sum_{i \in I} x_i + \sum_{i \in I} y_i;$$

und nach Definition B.9 folgt $\sum_{i \in I} (x_i + y_i) \leq \sum_{i \in I} x_i + \sum_{i \in I} y_i$.

Sind zum Nachweis der umgekehrten Ungleichung $s, t \in [0, \infty]$ mit

$$(1) \quad s < \sum_{i \in I} x_i \quad \text{und} \quad t < \sum_{i \in I} y_i$$

vorgegeben, so gibt es nach Definition B.9 dann endliche Mengen $I_1, I_2 \subseteq I$ mit $\sum_{i \in I_1} x_i > s$ und $\sum_{i \in I_2} y_i > t$, es ist damit also $I_0 := I_1 \cup I_2 \subseteq I$ eine endliche Menge, und mit der Isotonie B.8(4) für Summen mit endlich vielen Summanden im zweiten Schritt folgt

$$\sum_{i \in I_0} (x_i + y_i) = \sum_{i \in I_0} x_i + \sum_{i \in I_0} y_i \geq \sum_{i \in I_1} x_i + \sum_{i \in I_2} y_i > s + t,$$

und nach Definition B.9 folgt daher $\sum_{i \in I} (x_i + y_i) > s + t$. Wegen der Beliebigkeit von s, t mit (1) folgt die Behauptung. \square

B.11 Definition. Ist $(x_i : i \in I)$ eine Familie in $\overline{\mathbb{R}}$ mit $\sum_{i \in I} (x_i)_+ < \infty$ oder $\sum_{i \in I} (x_i)_- < \infty$, so heißt sie **halbsummierbar**, und

$$\sum_{i \in I} x_i := \sum_{i \in I} (x_i)_+ - \sum_{i \in I} (x_i)_-$$

heißt ihre **Summe**. Gilt sogar $\sum_{i \in I} (x_i)_+ < \infty$ und $\sum_{i \in I} (x_i)_- < \infty$, so heißt die Familie $(x_i : i \in I)$ **summierbar**. \square

Dabei sei auch für erweiterte reelle Zahlen $x_+ := \max(x, 0)$ und $x_- := \max(-x, 0)$, also $x = x_+ - x_-$ und $|x| := \max(x, -x) = x_+ + x_-$.

B.12 Bemerkung. Für jede Familie $x_\bullet = (x_i : i \in I)$ gelten die Implikationen

$$x_\bullet \text{ summierbar} \Leftrightarrow \sum_{i \in I} |x_i| < \infty \Leftrightarrow \sum_{i \in I} x_i \in \mathbb{R} \text{ existiert} \Rightarrow x_i \in \mathbb{R} \text{ für } i \in I,$$

denn: Die erste Bedingung impliziert die zweite wegen $(x_i)_+ + (x_i)_- = |x_i|$ und B.10(5). Die zweite Bedingung impliziert die erste wegen $(x_i)_+ \leq |x_i|$, $(x_i)_- \leq |x_i|$, und B.10(11). Die Äquivalenz der dritten Bedingung mit der ersten ergibt sich direkt aus Definition B.9. Die vierte Bedingung folgt aus der zweiten wegen $|x_{i_0}| = \sum_{i \in \{i_0\}} |x_i| \leq \sum_{i \in I} |x_i|$ für $i_0 \in I$ nach B.10(12). \square

Beweis. Nach Voraussetzung ist $M := \sum_{i \in I} |x_i| < \infty$. Für $n \in \mathbb{N}$ und $I_n := \{i \in I : |x_i| \geq \frac{1}{n}\}$ erhalten wir

$$M \geq \sum_{i \in I_n} |x_i| \geq \sum_{i \in I_n} \frac{1}{n} = \frac{1}{n} \#I_n$$

und damit $\#I_n \leq nM < \infty$. Folglich ist $I_0 = \bigcup_{n \in \mathbb{N}} I_n$ abzählbar, und die behauptete Identität ist ein Spezialfall von B.13(3). \square

B.15 Satz (Spezialisierung auf unendliche Reihen). *Es sei $x_\bullet = (x_k)_{k \in \mathbb{N}}$ eine Folge in \mathbb{R} . Dann gelten die Implikationen*

- (1) $x_k \geq 0$ für $k \in \mathbb{N} \Rightarrow \sum_{k \in \mathbb{N}} x_k = \sum_{k=1}^{\infty} x_k \in [0, \infty]$,
- (2) x_\bullet summierbar $\Leftrightarrow \sum_{k=1}^{\infty} x_k$ absolut konvergent $\Leftrightarrow \sum_{k \in \mathbb{N}} x_k = \sum_{k=1}^{\infty} x_k \in \mathbb{R}$.

Beweis. (1): Mit der Prämisse von (1) im zweiten und im dritten Schritt ergibt sich

$$\begin{aligned} \sum_{k=1}^{\infty} x_k &= \lim_{n \rightarrow \infty} \sum_{k=1}^n x_k = \sup_{n \in \mathbb{N}} \sum_{k \in \{1, \dots, n\}} x_k = \sup \left\{ \sum_{k \in I_0} x_k : I_0 \subseteq I \text{ endlich} \right\} \\ &= \sum_{k \in \mathbb{N}} x_k; \end{aligned}$$

dabei gilt im dritten Schritt “ \leq ” da jede der Mengen $\{1, \dots, n\}$ links eine der rechts auftretenden Mengen I_0 ist, und “ \geq ” wegen $I_0 \subseteq \{1, \dots, n\}$ mit $n := \max I_0$.

(2): Es bezeichnen A, B, C die drei als äquivalent behaupteten Aussagen. Dann ergibt sich $A \Leftrightarrow B$ durch (1) angewandt auf die Folge $(|x_k|)_{k \in \mathbb{N}}$, $B \Rightarrow C$ durch (1) angewandt auf die Folgen $((x_k)_+)_{k \in \mathbb{N}}$ und $((x_k)_-)_{k \in \mathbb{N}}$, und $C \Rightarrow A$ unter alleiniger Verwendung der Definiertheit und Endlichkeit von $\sum_{k \in \mathbb{N}} x_k \in \mathbb{R}$ aus der Definition der Summierbarkeit. \square

B.16 Aufgabe (Iterierte Summen). Es sei für $(i, j) \in \mathbb{N}^2$

$$x_{ij} := \begin{cases} 1 & \text{falls } j - i = 0, \\ -1 & \text{falls } j - i = 1, \\ 0 & \text{sonst.} \end{cases}$$

Berechnen Sie die beiden iterierten Summen $\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} x_{ij}$ und $\sum_{j \in \mathbb{N}} \sum_{i \in \mathbb{N}} x_{ij}$. Wie steht es mit der Summe $\sum_{(i,j) \in \mathbb{N}^2} x_{ij}$? \square

B.17 Aufgabe (Eine Umordnung der alternierenden harmonischen Reihe). Bekanntlich ist

$$s := \sum_{i=1}^{\infty} \frac{(-)^{i-1}}{i} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \dots \in]0, \infty[.$$

Betrachten Sie nun den Wert

$$t := 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + \frac{1}{13} + \frac{1}{15} - \frac{1}{8} + \dots$$

Welches Ergebnis erhalten Sie für t ? Hinweis:

$$t = \sum_{k=0}^{\infty} \left(\left(\frac{1}{4k+1} - \frac{1}{4k+2} + \frac{1}{4k+3} - \frac{1}{4k+4} \right) + \frac{1}{2} \cdot \left(\frac{1}{2k+1} - \frac{1}{2k+2} \right) \right).$$

□

Literaturhinweise. Siehe Lang (2002) für die Wohldefiniertheit endlicher Summen über ungeordneten Indextmengen, sowie die drei angegebenen Lehrbücher zur Analysis für Darstellungen der von Moore (1915, 1922) stammenden Theorie unendlicher Summen. Deren Nichtbehandlung in den meisten anderen Analysis-Büchern scheint purer Traditionismus zu sein.

BARNER, M. & FLOHR, F. (2000). *Analysis I*. 5. Auflage, de Gruyter, Berlin.

KÖNIGSBERGER, K. *Analysis I*. Springer, Berlin.

LANG, S. (2002). *Algebra*. Rev. 3rd. ed., Springer.

MOORE, E.H. (1915). Definition of limit in general integral analysis. *Proc. Nat. Acad. Sci. USA* **1**, 628.

MOORE, E.H. (1922). On power series in general analysis. *Math. Annalen* **86**, 30-39.

STORCH, U. & WIEBE, H. (2003). *Lehrbuch der Mathematik. Band I: Analysis einer Veränderlichen*. 3. Auflage, Spektrum Akademischer Verlag, Heidelberg.

Klausurvorbereitungsaufgaben zur WR I

Die Klausur zur WR I besteht aus 16+6 Multiple-Choice-Fragen. Bei jeder Frage sind bis zu fünf Antwortmöglichkeiten gegeben, von denen keine, eine, mehrere, oder gar alle richtig sein können. Sie erhalten bei einer Frage genau dann einen Punkt, wenn alle richtigen Antwortmöglichkeiten angekreuzt sind, und alle falschen Antwortmöglichkeiten nicht angekreuzt sind. Andernfalls erhalten Sie auf die betreffende Frage keinen Punkt. Einziges erlaubtes Hilfsmittel in der Klausur ist ein handgeschriebener “Spickzettel”, nicht größer als DIN A4, beide Seiten dürfen beschrieben sein.

Die Fragen 1 bis 16 der Klausur sind vergleichbar zu den weiter unten angegebenen Klausurvorbereitungsaufgaben KV 1 – KV 17, mit eventuellen Änderungen im Vorlesungsverlauf. Die sechs Fragen 17 bis 22 der Klausur sind identisch zu den folgenden Fragen mit nur je einer Antwortmöglichkeit:

17

$0 + 1 = 1$

18

$1 + 1 = 2$

19

$2 + 1 = 3$

20

$3 + 1 = 4$

21

$4 + 1 = 5$

22

$5 + 1 = 6$

Es sind also insgesamt $16 + 6 = 22$ Punkte erreichbar. Es sei $x + 6$ die von Ihnen erreichte Punktzahl. Im Fall $x + 6 \geq 22/2 = 11$, d.h. $x \geq 5$, haben Sie bestanden; ab $x + 6 - 11 \geq 0.9 \cdot (22 - 11)$, d.h. $x \geq 14.9$, d.h. $x \geq 15$, erhalten Sie die Bestnote 1.0. Genauer wird im Einklang mit der Allgemeinen Bachelor-Prüfungsordnung (daher der Zirkus mit den Fragen 17 bis 22) folgendes Bewertungsschema verwendet:

Punktzahl:	≤ 10	11	12	13	14	15	16	17	18	19	20	21	22
Punktzahl – 6:	≤ 4	5	6	7	8	9	10	11	12	13	14	15	16
Note:	n.b.	4.0	4.0	3.7	3.3	3.0	2.7	2.3	2.0	1.7	1.3	1.0	1.0

Lösungen zu den folgenden Klausurvorbereitungsaufgaben stehen auf Seite 184 und werden bei Bedarf in den Übungen besprochen. Einige der tatsächlichen Klausuraufgaben werden Variationen von Beispielaufgaben sein. Die Reihenfolge der Aufgaben in der Klausur entspricht ungefähr der Reihenfolge des Stoffes in den Vorlesungen und Übungen.

KV 1 Im Aussagenkalkül (S^*, \equiv) zur Grundmenge S und für $s, t \in S$ mit $s \neq t$ sind sinnvolle und richtige Äquivalenzaussagen

- $N\&s \equiv \&Ns$
- $N\&st \equiv \&st$
- $N\&st \equiv N\&ts$
- $\&N\&stN\&sNt \equiv Ns$

KV 2 Es seien R und S zwei Relationen auf der Menge X , d.h. Teilmengen von X^2 , und es sei $T := R \cap S$ und $U := R \cup S$. Dann gelten die Implikationen

- R transitiv $\Rightarrow T$ transitiv
- R, S transitiv $\Rightarrow T$ transitiv
- R reflexiv $\Rightarrow T$ reflexiv
- R reflexiv $\Rightarrow U$ reflexiv

KV 3 In jeder Boole-Algebra $(\mathcal{A}, \wedge, \vee, ', O, I)$ gilt für $A, B \in \mathcal{A}$ stets

- $A \wedge A' = A \vee A'$
- $(A' \vee B) = A \wedge B'$
- $(A \wedge B) \wedge C = A \wedge (B \wedge C)$
- $(A \wedge B) \vee C = A \wedge (B \vee C)$

KV 4 Es sei Ω eine mindestens zweielementige Menge. Für die Mengenalgebren $\mathcal{A} := \{\emptyset, \Omega\}$ und $\mathcal{B} := 2^\Omega$ und die Urbildabbildung $\varphi := \text{id}_\Omega^{-1}[\cdot]$ gilt dann

- $\varphi|_{\mathcal{A}}$ ist ein Morphismus von \mathcal{A} nach \mathcal{B}
- $\varphi|_{\mathcal{A}}$ ist ein Isomorphismus von \mathcal{A} nach \mathcal{B}
- φ ist ein Morphismus von \mathcal{B} nach \mathcal{A}
- φ ist ein Isomorphismus von \mathcal{B} nach \mathcal{A}

KV 5 Es sei φ_\bullet ein Quasikollektiv in der Boole-Welt $(\mathcal{C}, \mathcal{R})$ mit der Verteilung P . Dann gilt stets:

- P ist ein Wahrscheinlichkeitsinhalt auf der Boole-Algebra \mathcal{R} .
- P ist ein Wahrscheinlichkeitsinhalt auf der Boole-Algebra \mathcal{C} .
- P ist ein Wahrscheinlichkeitsinhalt auf einer Boole-Algebra.
- P ist endlich additiv.
- Ist \mathcal{S} eine Menge von Auswahlregeln und ist φ_\bullet kein \mathcal{S} -Kollektiv, so ist P nicht endlich additiv.

KV 6 Es sei \mathbf{x} ein Zahlenvektor der Länge 100. In R gilt:

- Der Befehl `seq(1, 10)` liefert einen Vektor der Länge 10 mit jedem Eintrag 1.
- Der Befehl `x[(1:10)^2]` liefert einen Vektor mit genau den Einträgen von \mathbf{x} , deren Indizes Quadratzahlen sind.
- Der Befehl `sum(x==0)` liefert die Anzahl der Nullen in \mathbf{x} .
- Der Befehl `sapply(1:2, function(y) mean(x[y+2*(0:49)]))` liefert einen Vektor der Länge 2 bei dem der erste Eintrag der Mittelwert aller Einträge von \mathbf{x} mit ungeraden Indizes ist.

KV 7 Für beliebige Ereignisse A, B in einem (elementaren) Wahrscheinlichkeitsraum (Ω, P) gelten die (Un-)gleichungen beziehungsweise Implikationen

- $P(A \setminus B) = P(A) - P(B)$
- $P(A \cup B) = P(A) + P(B)$
- $P(A) \leq P(A^c)$
- $P(A) \geq 2/3, P(B) \geq 2/3 \Rightarrow P(A \cap B) \geq 4/9$
- $P(A) \geq 2/3, P(B) \geq 2/3 \Rightarrow P(A \cap B) \geq 1/3$

KV 8 Es sei $\Omega := \{1, \dots, n\}$ mit $n \geq 2$. In welchen der folgenden Fälle ist p eine Wahrscheinlichkeitszähldichte auf Ω ?

- $p(k) := \frac{k}{n}$ für $k \in \Omega$
- $p(k) := \frac{k}{\frac{n(n+1)}{2}}$ für $k \in \Omega$
- $p(k) := 1$ für $k \in \Omega$
- $p(k) := \frac{k - \frac{n(n+1)}{4}}{\frac{n(n+1)}{4}}$ für $k \in \Omega$

KV 9 Mit den Notationen der Vorlesung für die Gleich-, Binomial-, und Poissonverteilungen gilt

- $U_{\{1, \dots, 5\}}(2\mathbb{N}) = \frac{1}{2}$
- $b_{3, \frac{1}{3}}(0) = \frac{8}{27}$
- $B_{4, \frac{1}{2}}(\{1, \dots, 4\}) = 1$
- $p_1(1) = e^{-1}$

KV 10 Im Standardmodell des einmaligen gleichzeitigen Werfens eines roten und eines grünen Würfels ist die Wahrscheinlichkeit dafür, dass die Augenzahl des roten Würfels echt größer als die des grünen Würfels ist, gleich

- 5/6
- 1/3
- 5/12
- 1/4
- 1/2

KV 11 Es seien A, B, C, D Ereignisse in einem elementaren Wahrscheinlichkeitsraum (Ω, P) . Dann gilt stets

- $P(A \cup B \cup C \cup D) \leq P(A) + P(B) + P(C) + P(D) - P(AB) - P(BC) - P(CD)$.
- $P(A \cup B \cup C \cup D) \geq P(A) + P(B) + P(C) + P(D) - P(AB) - P(AC) - P(AD) - P(BC) - P(BD) - P(CD)$
- $P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D) - P(AB) - P(AC) - P(AD) - P(BC) - P(BD) - P(CD) + P(ABC) + P(ABD) + P(ACD) + P(BCD)$

KV 12 Eine Arbeitsgruppe aus 6 Personen soll in zwei gleichberechtigte Gruppen zu je 3 Personen geteilt werden. Wieviele Möglichkeiten gibt es dafür?

- 18
- 6
- 3
- 20
- 10

KV 13 Die Anzahl derjenigen “vierstelligen” Zahlen zwischen 0000 und 9999, die genau drei verschiedene Ziffern enthalten, ist

- 360
- 720
- 2880
- 4320
- 4960

KV 14 Aus einer Urne mit $w \geq 2$ weißen und $s \geq 1$ schwarzen Kugeln werden drei Kugeln zufällig entnommen. Dann ist unter der naheliegenden Gleichverteilungsannahme die Wahrscheinlichkeit dafür, dass genau zwei davon weiß sind, stets

- $\frac{\binom{w}{2}}{\binom{w+s}{3}}$
- $\frac{2}{3}$
- $\frac{s \binom{w}{2}}{\binom{w+s}{3}}$
- $\frac{2}{w+s}$
- $\frac{w \binom{s}{2}}{\binom{w+s}{3}}$

KV 15 Es sei $\mathbb{P} := U_{\{(0,0),(1,1)\}}$ auf $\Omega := \{0, 1\}^2$. Mit den durch

$$X(\omega) := \omega_1, \quad Y(\omega) := \omega_2 \quad \text{für } \omega \in \Omega$$

definierten Zufallsgrößen X und Y auf (Ω, \mathbb{P}) gilt dann

- $X \sim U_{\{0,1\}}$
- $(X, Y) \sim (X, X)$
- $\mathbb{P}(X - Y = 0) = 1$
- $(X, Y) \sim U_\Omega$

KV 16 Es sei (Ω, P) ein elementarer Wahrscheinlichkeitsraum und es seien $A, B, C \subseteq \Omega$ mit $P(B)P(B^c)P(C) > 0$. Dann gelten die Identitäten bzw. Implikationen

- $P(A|B) + P(A^c|B) = P(B)$
- $P(A|B) + P(A^c|B) = 1$
- $P(A|B) = P(A|B^c) \Rightarrow P(A|B) = P(A)$
- $B \subseteq C \Rightarrow P(A|B) \leq P(A|C)$
- $B \subseteq C \Rightarrow P(A|B) \geq P(A|C)$

KV 17 Es sei $\Omega = \Omega_1 \times \Omega_2$ mit Mengen Ω_1 und Ω_2 . Weiter seien $P_1, Q_1 \in \text{Prob}(\Omega_1)$ und $P_{2|1}, Q_{2|1} \in \text{Mark}(\Omega_1, \Omega_2)$ mit

$$P_1 \otimes P_{2|1} = Q_1 \otimes Q_{2|1}.$$

Dann gelten die Identitäten bzw. Implikationen

- $P_1 = Q_1$
- $P_{2|1} = Q_{2|1}$
- $\omega_1 \in \Omega_1$ mit $P_1(\{\omega_1\}) > 0 \Rightarrow P_{2|1}(\cdot|\omega_1) = Q_{2|1}(\cdot|\omega_1)$
- $\omega_1 \in \Omega_1$ mit $P_1(\{\omega_1\}) = 0 \Rightarrow P_{2|1}(\cdot|\omega_1) \neq Q_{2|1}(\cdot|\omega_1)$

(Wie gesagt gibt es in der Klausur nur 16 Aufgaben vom Typ KV 1 bis KV 17.)
Lösungen zu den Klausurvorbereitungsaufgaben auf Seite 184:

1cd, 2bd, 3bc, 4a, 5cd, 6bcd, 7e, 8b, 9bd, 10c, 11ab, 12e, 13d, 14c, 15abc, 16bc, 17ac

Dabei bedeutet zum Beispiel “1cd” dass bei KV1 genau die dritte und die vierte Antwort richtig sind.

Hinweise zu den mündlichen Prüfungen der WR

1. Dort sollte jede Frage inhaltlich und sprachlich korrekt beantwortet werden, idealerweise durch den Prüfling. Gelegentliches Abweichen vom Hochdeutschen, etwa “holen” statt “nehmen”, ist dabei kein Problem, eine Vermengung von Dativ und Akkusativ, wie zum Beispiel in “Der Schnitt ist distributiv über *die* Vereinigung” (leider schon oft gehört), oder auch die fälschliche Verwendung eines bestimmten Artikels¹, gerade bei Lehramtskandidaten schon. Das Korrigieren fehlerhafter oder unvollständiger Antworten verringert den in der Prüfung behandelbaren Stoff und kann dementsprechend zu einer weniger guten Note führen. Eigentlich sollten die Übungsgruppen sowie Sprechstunden, neben Gesprächen mit Kommilitonen, über die Jahre ausreichend Gelegenheiten bieten, ein präzises mathematisches Formulieren zu üben; wer diese Gelegenheiten nicht oder nur wenig nutzt, dem ist nicht zu helfen. Und wer es als Hörer der WR nicht schafft, zum Beispiel “Binomialverteilung” richtig zu sagen, der sollte vielleicht nicht gerade Lehrer werden wollen – egal in welchem Fach.

2. So gut wie keine Prüfung, auch wenn sie am Ende mit 1.0 benotet wird, verläuft fehlerfrei. Leicht verheddert man sich bei einer geforderten kleinen Rechnung oder steht vielleicht bei einer Frage zur Kombinatorik wie der Ochs vorm Berg, auch wenn man eigentlich außerhalb der Prüfungssituation den jeweiligen Stoff beherrscht. In so einem Fall ist es wichtig, dass Sie als Prüfling die Prüfung weiterhin sportlich sehen: Ja, es gibt Fehler, die so gravierend sind, dass die Prüfung eigentlich nach 5 oder 10 Minuten schon vorbei ist, aber Ihre Aufgabe ist es nicht, sich darüber Gedanken zu machen, sondern auf die weiteren Fragen möglichst gut zu antworten. So mancher Prüfling ist am Ende überrascht, trotz von ihm selbst bemerkter Mängel eine gute bis sehr gute Note zu erhalten.

3. Das Prüfungsanforderungsniveau ist viel niedriger, als man vielleicht aufgrund des Skriptes befürchten mag: Man braucht nicht lange, komplizierte Beweise wiederzugeben um eine Eins zu erhalten (kürzere aber schon). Und selbstverständlich wird kein Fußnotenwissen abgefragt.

4. In der Vorlesung nicht vorgetragene Teile des Skriptes sowie nicht oder nur mit * gestellte Aufgaben sind nicht unmittelbar prüfungsrelevant. Mittelbar dagegen natürlich schon, denn je mehr man lernt, desto leichter fällt einem alles andere.

5. Sie dürfen Ihr Skript (auch mit eigenen Anmerkungen versehen) in der Prüfung zum *gelegentlichen* Nachschlagen verwenden. Zum Beispiel führt es nicht zu einer schlechteren Benotung, wenn Sie die genaue Formalisierung des Begriffs “ \mathcal{A} -Auswahlregel” nochmal nachschlagen müssen. Andererseits fällt natürlich durch, wer schon bei Fragen wie “Was ist ein elementarer Wahrscheinlichkeitsraum?” alles nachschlagen muss.

¹Leider unterlaufen nicht nur Ihrem Dozenten, sondern auch wirklich gut schreibenden Autoren wie etwa Heuser (2009), manchmal Formulierungen wie “Es sei X eine Zufallsgröße auf dem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) ”, ohne (Ω, \mathbb{P}) vorher einzuführen. Derartiges sollte in diesem Skript bei künftigen Verbesserungen möglichst ganz eliminiert werden.

6. Die erste Frage der Prüfung wird “Wo sollen wir anfangen?” sein, die zweite dann zu Ihrer Antwort passend möglichst wörtlich eine der untenstehenden. Die weiteren Fragen ergeben sich - größtenteils in Anlehnung an untenstehende Beispielfragen - aus dem Prüfungsverlauf, mit den in Anbetracht der Kürze der zur Verfügung stehenden Zeit konkurrierenden Zielen der exemplarischen Klärung von Details, der eventuellen Einräumung der Gelegenheit zu Berichtigungen, und der Abdeckung von möglichst viel Stoff.

7. Modifikationen dieser Hinweise im Vorlesungsverlauf sind möglich.

Beispiele möglicher Prüfungsfragen:

- Was ist eine Boole-Algebra?
 - Beispiele?
 - Welche weiteren Rechenregeln gelten in einer Boole-Algebra?
 - * Könnten Sie bitte versuchen, diese ... zu beweisen?
 - Wieso haben wir Boole-Algebren betrachtet?
 - * Was ist der Aussagenkalkül zur Grundmenge S ?
 - * Und was hat der nun mit Boole-Algebren zu tun?
- Was ist ein Boole-Algebra-Morphismus?
 - Beispiele?
 - Was haben die für weitere Eigenschaften?
- Was ist eine Boole-Welt?
- Was ist ein Quasikollektiv?
 - Welche Eigenschaften hat dessen Verteilung? Siebformel?
 - Warum ist man mit dem Begriff des Quasikollektivs noch nicht ganz zufrieden?
 - * Daher welche Verfeinerung?
 - * Problem damit? Widerspruchsfrei?
- Was ist ein elementarer Wahrscheinlichkeitsraum?
 - Was für Rechenregeln gelten da?
 - * Könnten Sie bitte versuchen, diese ... zu beweisen?
 - Alternative Beschreibung durch einfacheres mathematisches Objekt?
 - Konkrete Beispiele?
- Welche Methoden der Kombinatorik haben wir kennengelernt?
- Was ist eine Zufallsgröße?
 - Was ihre Verteilung?
 - Beispiel?
 - Und wozu sind Zufallsgrößen gut?
- Was ist eine hypergeometrische Verteilung? Und wo kommen die vor?

- Was ist eine bedingte Wahrscheinlichkeit?
 - Beispiel?
 - Rechenregeln?
 - Simpson-Paradoxon?
- Von den Ereignissen A, B in einem Wahrscheinlichkeitsraum (Ω, \mathbb{P}) seien Ihnen nur die Wahrscheinlichkeiten $\mathbb{P}(A)$ und $\mathbb{P}(B)$ bekannt.
 - Was können Sie über $\mathbb{P}(A \cup B)$ sagen?
 - Was können Sie über $\mathbb{P}(A \cap B)$ sagen?
- Was ist ein Kopplungsmodell?
 - Beispiel?
 - Eigenschaften?
 - Universalität?
- Wie ist die stochastische Unabhängigkeit definiert?
 - Und was ist eine faktorisierte bedingte Verteilung?
 - Wie kann man Unabhängigkeit von Zufallsgrößen verifizieren?
 - Wie charakterisiert sich die Unabhängigkeit von zwei bzw. von drei Ereignissen?
 - Beispiele, Gegenbeispiele?
- Was ist eine Bernoulli-Kette?
 - Und was hat das mit Binomialverteilungen zu tun?
 - Und mit Negativbinomialverteilungen?
 - Und mit hypergeometrischen Verteilungen?
- Was ist eine Faltung?
 - Beispiele?
 - * Und wie beweist man Beispiel ... ?
- Was ist eine Multinomialverteilung?
 - Eigenschaften?
- Was ist eine Verteilungsfunktion?
 - Beispiel?
 - Wozu gut?
- Was ist ein Median?
 - Beispiel?
- Was ist ein Erwartungswert?
 - Beispiele?
 - Eigenschaften?
 - Und was sind Varianzen?

- Wozu sind Erwartungswerte und Varianzen gut?
- Wie sind allgemeine Summen definiert?
 - Wie steht es mit der Rechenregel $\sum_{i \in I} \sum_{j \in J} x_{ij} = \sum_{j \in J} \sum_{i \in I} x_{ij}$?
- Was sagt das schwache Gesetz großer Zahlen?
 - Beweis unter einer Zusatzvoraussetzung?
- Was wissen wir zur Poisson-Approximation?
- Was sind φ und Φ ?
 - $\Phi^{-1}(0.95) = ?$, $\Phi^{-1}(0.975) = ?$
 - Und wie haben wir das benutzt?
 - Wie lautet der ZGWS für gewisse, von uns betrachtete spezielle Verteilungen?
- Stirling-Formel?
 - Wozu gut?
- Was ist ein Konfidenzbereich?
 - Beispiel?
- Was ist ein Test?
 - Beispiel?
- Erklären Sie mir doch bitte das eine oder andere R-Kommando welches Sie in einer der Übungen verwendet haben.

Weiter kann es Fragen folgender Art zum Skript geben:

- Können Sie mir hier diesen Beweisschritt erklären?
- Wie löst man diese Aufgabe?