

Christhard Schmid

Angewandte Stochastik und Versuchsplanung in den Natur- und Ingenieur- wissenschaften

mit Beispielen in R und SAS



Angewandte Stochastik und Versuchsplanung in den Natur- und Ingenieurwissenschaften

mit Beispielen in R und SAS

von Christhard Schmid

Gewidmet den Töchtern Friederike Schmid und Rotraut Schoop

WILEY

WILEY-VCH GmbH

Autor

Prof. Dr. Christhard Schmid

Feuerreiterweg 12
70597 Stuttgart
Germany

© Erhan Ergin / Fotolia.com für die in
der Randspalte verwendeten Symbole

■ Alle Bücher von Wiley-VCH werden sorgfältig erarbeitet. Dennoch übernehmen Autoren, Herausgeber und Verlag in keinem Fall, einschließlich des vorliegenden Werkes, für die Richtigkeit von Angaben, Hinweisen und Ratschlägen sowie für eventuelle Druckfehler irgendeine Haftung

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.d-nb.de>> abrufbar

© 2023 Wiley-VCH GmbH, Boschstraße 12, 69469 Weinheim, Germany

Alle Rechte, insbesondere die der Übersetzung in andere Sprachen, vorbehalten. Kein Teil dieses Buches darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Photokopie, Mikroverfilmung oder irgendein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsmaschinen, verwendbare Sprache übertragen oder übersetzt werden. Die Wiedergabe von Warenbezeichnungen, Handelsnamen oder sonstigen Kennzeichen in diesem Buch berechtigt nicht zu der Annahme, dass diese von jedermann frei benutzt werden dürfen. Vielmehr kann es sich auch dann um eingetragene Warenzeichen oder sonstige gesetzlich geschützte Kennzeichen handeln, wenn sie nicht eigens als solche markiert sind.

Print ISBN 978-3-527-34629-5

ePDF ISBN 978-3-527-82235-5

ePub ISBN 978-3-527-82237-9

Umschlaggestaltung Grafik-Design Schulz

Satz Straive, Chennai, India

Druck und Bindung

Inhaltsverzeichnis

1	Einleitung	1
1.1	Hauptbegriffe	1
1.2	Zielgruppen	2
1.2.1	Einsteiger/innen	2
1.2.2	Physiker/innen	3
1.2.3	SAS-Anwender/innen	4
1.3	Anmerkungen zur Notation	5
1.4	Symbole und Formelzeichen	6
1.5	Gliederung des Buches	11
1.6	Datenmaterial zu diesem Buch	12
2	Wahrscheinlichkeitstheorie	13
2.1	Was ist Wahrscheinlichkeit?	13
2.1.1	Ausgangssituation	13
2.1.2	Ermittlung von klassischen Wahrscheinlichkeiten	14
	<i>Beispiel „Münzwurf“</i>	14
	<i>Beispiel „Zwei Kugeln“</i>	14
	<i>Beispiel „N Münzwürfe“</i>	15
	<i>Beispiel „Verbogene Münze (Teil 1)“</i>	15
2.1.3	Formale Definition des Begriffs Wahrscheinlichkeit	16
2.2	Wahrscheinlichkeit abstrakt mathematisch	17
2.2.1	Der Wahrscheinlichkeitsraum	17
	<i>Beispiel „Verbogene Münze (Teil 2)“</i>	18
	<i>Beispiel „Maxwell-Boltzmann-Verteilung“</i>	18
2.2.2	Bedingte Wahrscheinlichkeit und statistische Unabhängigkeit	18
	<i>Beispiel „Bedingte Wahrscheinlichkeit“</i>	19
2.2.3	Zufallsvariablen	20
2.2.4	Die kumulative Verteilungsfunktion	20
2.2.5	Beispiele für kumulative Verteilungen	21
	<i>Beispiel „10-Euro Münze (Teil 1)“</i>	21
	<i>Beispiel „Kumulative Maxwell-Boltzmann-Verteilung“</i>	21
	<i>Beispiel „Kumulative Binomialverteilung“</i>	22
	<i>Beispiel „Kumulative Poisson-Verteilung“</i>	23
2.2.6	Die Dichtefunktion	23
	<i>Beispiel „10-Euro Münze (Teil 2)“</i>	24

	<i>Beispiel „Energiedichte der Maxwell-Boltzmann-Verteilung“</i>	24
2.2.7	Multidimensionale Verteilungen	25
2.3	Erwartungswert, Varianz, Korrelation	25
2.3.1	Varianz von Funktionen von Zufallsvariablen	26
2.3.2	Das Gaußsche Fehlerfortpflanzungsgesetz	28
2.4	Charakteristische Funktion und Momente	29
2.4.1	Die charakteristische Funktion einer Verteilung	29
2.4.2	Momente und Kumulanten	29
2.4.3	Kumulanten niedrigster Ordnung	30
2.4.4	Zentrale Momente niedrigster Ordnung	31
2.5	Die Berechnung von Verteilungen mit Hilfe der Deltafunktion	31
2.5.1	Verteilung von Funktionen einer Zufallsvariablen	31
2.5.2	Die Erwartungswerte von Delta- und Thetafunktion	32
2.5.3	Verallgemeinerungen auf mehrere Zufallsvariable	32
2.6	Grenzverteilungen und zentraler Grenzwertsatz	33
2.6.1	Historischer Exkurs zur Gauß-Verteilung	33
2.6.2	Zentraler Grenzwertsatz nach Lindenberg-Lévy	34
2.6.3	Lévy-Verteilungen	35
2.7	Quantile	37
2.8	Ergänzungen	38
2.8.1	Symmetrielerationen	38
2.8.2	Erwartungswerte von quadratischen Formen	38
2.8.3	Kombinatorik	38
3	Verteilungen	39
3.1	Eindimensionale diskrete Verteilungen	39
3.1.1	Hypergeometrische Verteilung	39
	<i>Beispiel „T-Shirts“</i>	40
	<i>Beispiel „6 aus 49“</i>	41
	<i>Beispiel „Komitee“</i>	41
	<i>Beispiel „Gewinner aus der falschen Stadt“</i>	42
3.1.2	Negative hypergeometrische Verteilung	42
3.1.3	Binomialverteilung	43
	<i>Beispiel „Zwei Grundaufgaben der Binomialverteilung“</i>	46
	<i>Beispiel „Glücksrad“</i>	47
3.1.4	Poisson-Verteilung	48
	<i>Beispiel „Ausfälle einer technischen Anlage“</i>	51
	<i>Beispiel „Giftpilze“</i>	51
3.1.5	Geometrische Verteilung	53
3.1.6	Negative Binomialverteilung	55
3.1.7	Tabellen der eindimensionalen diskreten Verteilungen	55
3.2	Eindimensionale kontinuierliche Verteilungen	58
3.2.1	Gauß-Verteilung	58
3.2.2	Cauchy-Verteilung	63
3.2.3	Chi-Quadrat-Verteilung	65
3.2.4	Nichtzentrale Chi-Quadrat-Verteilung	69
3.2.5	Student-Verteilung	70

3.2.6	Nichtzentrale Student-Verteilung	74
3.2.7	F-Verteilung	75
3.2.8	Nichtzentrale F-Verteilung	77
3.2.9	Exponentialverteilung	77
3.2.10	Weibull-Verteilung	79
3.2.11	Tabellen der eindimensionalen kontinuierlichen Verteilungen	82
3.3	Mehrdimensionale Verteilungen	85
3.3.1	n -dimensionale Gauß-Verteilung	85
3.3.2	Zweidimensionale Gauß-Verteilung	86
4	Mathematische Stichproben, Messreihen	87
4.1	Definition, Mittelwert und Stichprobenvarianz	87
4.1.1	Definition einer eindimensionalen Stichprobe und verteilungsunabhängige Eigenschaften	87
4.1.2	Einschub: Fehlerrechnung im physikalischen Praktikum	90
4.1.3	Verteilung der Stichprobenmomente bei Gauß-Verteilung	92
4.2	Konfidenzintervalle	93
4.2.1	Begriffe	93
4.2.2	Eine Tautologie als Basis	94
4.2.3	Die Tautologie für $\hat{\mu}$ und $\hat{\sigma}^2$	96
4.2.4	μ -Konfidenzintervalle bei bekanntem σ	96
4.2.5	Beispiele für μ -Konfidenzintervalle bei bekanntem σ	97
	<i>Beispiel „Benzinverbrauch“</i>	97
4.2.6	μ -Konfidenzintervall bei <i>unbekanntem</i> σ	97
4.2.7	Beispiele für μ -Konfidenzintervalle bei unbekanntem σ	99
	<i>Beispiel „Spülmittel (Teil 1)“</i>	101
	<i>Beispiel „Spülmittel (Teil 2)“</i>	101
	<i>Beispiel „Kontoführung (Teil 1)“</i>	102
	<i>Beispiel „Kontoführung (Teil 2)“</i>	102
	<i>Beispiel „Astronomische Distanz“</i>	103
	<i>Beispiel „Körpergröße“</i>	104
4.2.8	Konfidenzintervall für die Varianz σ^2	105
4.2.9	Beispiele für Konfidenzintervalle für die Varianz σ^2	107
	<i>Beispiel „Betonstahlmatten“</i>	108
4.2.10	Prognoseintervalle	109
4.2.11	Beispiele für Prognoseintervalle	109
	<i>Beispiel „$\hat{\mu}$-Prognoseintervall“</i>	109
	<i>Beispiel „$\hat{\sigma}^2$-Prognoseintervall“</i>	110
4.3	Parametertests	110
4.3.1	Begriffe: Nullhypothese, Prüfgröße, p-Wert	110
4.3.2	Der „N-Test“ (Test von μ bei bekanntem σ)	112
4.3.3	Beispiele für den „N-Test“	115
	<i>Beispiel „Zuckerpakete“</i>	115
	<i>Beispiel „Massenprodukte 1“</i>	115
	<i>Beispiel „Massenprodukte 2“</i>	117
4.3.4	Der t-Test (Test von μ bei unbekanntem σ)	119
4.3.5	Beispiele für den t-Test	120

	<i>Beispiel „Akkulaufzeit“</i>	120
	<i>Beispiel „Cholesterin 1“</i>	120
	<i>Beispiel „Wasserverbrauch“</i>	121
	<i>Beispiel „Autoreifen“</i>	122
	<i>Beispiel „Tankstelle“</i>	122
	<i>Beispiel „Cholesterin 2“</i>	123
4.3.6	Der t-Verschiebungstest	124
4.3.7	Der doppelte „N-Test“	125
4.3.8	Beispiele für den doppelten „N-Test“	126
	<i>Beispiel „Zweierlei Autoreifen (Teil 1)“</i>	126
4.3.9	Der doppelte t-Test	127
4.3.10	Beispiele für den doppelten t-Test	128
	<i>Beispiel „Zweierlei Autoreifen (Teil 2)“</i>	128
	<i>Beispiel „Vitamin C“</i>	128
4.3.11	Der asymptotische Test bei ungleichen Varianzen	129
4.3.12	Beispiele für den asymptotischen Test	130
	<i>Beispiel „Ziegel“</i>	130
4.3.13	Test der Varianz σ^2	131
4.3.14	Beispiele für Test der Varianz	132
	<i>Beispiel „Blumenzwiebeln“</i>	132
	<i>Beispiel „Abfüllanlage“</i>	132
4.4	Power-Analyse	133
4.4.1	Begriffe: β -Fehler, Power, Güte, OC	133
4.4.2	Das Forellenbeispiel	134
4.4.3	Illustrationen des „Forellenbeispiels“ mit R und SAS	136
	<i>Beispiel „Forellen: β-Fehler“</i>	136
	<i>Beispiel „Forellen: OC-Kurven“</i>	137
4.4.4	Die analytische linksseitige „ μ -power“-Theorie	138
4.4.5	R-Werkstatt für die allgemeine „ μ -power“-Theorie	140
4.4.6	Anwendungsbeispiele für die „ μ -power“-Theorie mit R	140
	<i>Beispiel „Forellen: μ-power-Analyse“</i>	140
	<i>Beispiel „Simple Power-Analyse elementar“</i>	142
	<i>Beispiel „Simple Power-Analyse mit pwr“</i>	143
4.4.7	Die analytische „ μ -sample-size“-Theorie	143
4.4.8	R-Werkstatt für die „ μ -sample-size“-Theorie	144
4.4.9	Beispiele für die „ μ -sample-size“-Theorie mit R	145
	<i>Beispiel „Stahlprüfung“</i>	145
	<i>Beispiel „Milchfettprüfung“</i>	147
5	Regression	149
5.1	Einführung	149
5.1.1	Was ist Regressionsanalyse?	149
5.1.2	Der lineare Regressionsansatz	150
5.1.3	Die Modellgleichung	150
5.1.4	(Ko-)Varianzanalyse und kategoriale Faktoren	150
5.2	Annahmen	151
5.2.1	Die acht Annahmen des klassischen linearen Modells	151

5.3	Modellparameter, Schätzung und Residuen	153
5.3.1	Deterministische Größen im klassischen linearen Modell . . .	153
5.3.2	Zufällige Größen im klassischen linearen Modell	154
5.3.3	Erwartungswert, Kovarianz und geschätzte Kovarianz der zufälligen Größen	155
5.3.4	Ein alternativer Ausdruck für die Varianzen von $\hat{\beta}$	156
5.3.5	Geometrische Eigenschaften aufgrund des KQ-Prinzips	157
5.3.6	Prognose für eine zukünftige Versuchseinstellung	158
5.4	Quadratsummen und Varianzanalyse	159
5.4.1	Quadratsummen und die „ANOVA-Tabelle“	159
5.4.2	Additive Zerlegung der korrigierten Gesamtstreuung	160
5.4.3	Berücksichtigung von Mehrfachversuchen	160
5.4.4	Zerlegung der Reststreuung bei Mehrfachversuchen	162
5.4.5	Das Bestimmtheitsmaß	162
5.4.6	Zerlegung der unkorrigierten Gesamtstreuung	163
5.5	Verteilungseigenschaften	163
5.5.1	Verteilung von Linearkombinationen der Stör-Variablen . . .	164
5.5.2	Drei grundlegende Verteilungseigenschaften	164
5.5.3	Verteilung der standardisierten Regressionsparameter	165
5.5.4	Verteilungseigenschaften der LOF-Quadratsummen	165
5.6	Hypothesentests	166
5.6.1	Der Overall-F-Test (Goodness-of-Fit-Test)	166
5.6.2	Der partielle F-Test	166
5.6.3	Der partielle t-Test	167
5.6.4	Der Lack-of-Fit-Test	167
5.6.5	Resumé: Die vier Tests der linearen Regression	168
5.6.6	Output von R und SAS	169
5.6.7	Beispiele für Regression und Hypothesentests	171
	<i>Beispiel „Blutdruck (Teil 1)“</i>	171
	<i>Beispiel „Blutdruck (Teil 2)“</i>	172
	<i>Beispiel „Blutdruck (Teil 3)“</i>	174
	<i>Beispiel „Stahlausbeute“</i>	175
	<i>Beispiel „Eukalyptusbäume“</i>	177
5.7	Konfidenzintervalle und Prognoseintervalle	179
5.7.1	Zusammenfassung: Konfidenz- und Prognoseintervalle	179
5.7.2	Beispiele	180
	<i>Beispiel „Stahlhärte“</i>	180
	<i>Beispiel „Sojabohnen (Teil 1)“</i>	184
	<i>Beispiel „Fallbeschleunigung“</i>	186
	<i>Beispiel „Federkonstante“</i>	189
5.8	Spezialfall: Einfache lineare Regression	191
5.8.1	Schwerpunktkoordinaten	191
5.8.2	Deterministische Größen	191
5.8.3	Zufällige Größen	192
5.8.4	Versuchsplanung bei der einfachen linearen Regression	192
5.8.5	Quadratsummen und Bestimmtheitsmaß	193
5.8.6	Konfidenz- und Prognoseintervall	193

5.8.7	Beispiele	193
	<i>Beispiel „Sojabohnen (Teil 2)“</i>	193
5.9	Modelldiagnose	195
5.9.1	Analyse der erklärenden Variablen	195
	<i>Beispiel „VIF“</i>	197
5.9.2	Analyse der Stör-Variablen	200
5.9.3	Einflussanalyse	203
5.9.4	Das Diagnose-Quartett <code>plot(lm(...))</code>	205
5.9.5	Beispiele zur Modelldiagnose	208
	<i>Beispiel „Fitness“</i>	208
	<i>Beispiel „Forbes“</i>	213
5.9.6	Alternative Modellierungsansätze nach Modelldiagnose	215
	<i>Beispiel „Reinigungsprozess“</i>	219
5.10	Nichtlineare Regression	220
5.10.1	Schätzverfahren der nichtlinearen Regression	220
5.10.2	R-Beispiele für Nichtlineare Regression	221
	<i>Beispiel „Pflanzenwachstum“</i>	221
	<i>Beispiel „Windgeschwindigkeit“</i>	223
6	Varianzanalyse	225
6.1	Einführung	225
6.1.1	Was ist Varianzanalyse?	225
6.1.2	Was ist Kovarianzanalyse?	226
6.1.3	Die Modelle der Varianzanalyse	226
6.1.4	Die hierarchische Fragestellung in der ANOVA und das Problem der mehrfachen Tests	227
6.1.5	Varianzanalyse mit R	228
6.2	Varianzanalyse mit festen Effekten	229
6.2.1	One-way ANOVA (einfaktorielle Varianzanalyse)	229
	<i>Beispiel „Stellen-Zugfestigkeit“</i>	232
	<i>Beispiel „Lehrmethoden-Zeugnis“</i>	235
6.2.2	Two-way ANOVA mit einfacher Besetzung	238
	<i>Beispiel „Glasart-Phosphorart-Helligkeit“</i>	239
6.2.3	Two-way ANOVA mit mehrfacher Besetzung	242
	<i>Beispiel „Materialien-Temperaturen-Lebensdauer“</i>	242
	<i>Beispiel „Preise-Werbungen-Absatz“</i>	244
	<i>Beispiel „Stoffe-Tage-Festigkeit-Block“</i>	248
	<i>Beispiel „Drucke-Chargen-Ausschuss-Block“</i>	250
6.2.4	Three-way und Four-way ANOVA (einfache Besetzung): Lateinische und Griechisch-Lateinische Quadrate	253
	<i>Beispiel „Treibstoff-Herstellung (Teil 1)“</i>	255
	<i>Beispiel „Treibstoff-Herstellung (Teil 2)“</i>	256
6.2.5	Kovarianzanalyse mit festen Effekten (ANCOVA)	257
	<i>Beispiel „Kilometerleistung“</i>	257
	<i>Beispiel „Truthahngewicht“</i>	258

7	Versuchsplanung	263
7.1	Einführung	263
7.1.1	Einsatzgebiet, Zielsetzung	263
7.1.2	Zwei Grundprinzipien der Versuchsplanung	264
7.1.3	Arten und Zielsetzungen von Versuchsplänen	265
7.1.4	Momentenmatrix und Design-Momente	267
7.2	Vollfaktorielle Pläne	269
7.2.1	Vollfaktorielle Versuchspläne (Definition)	269
7.2.2	Konstruktion faktorieller Pläne mit R	272
7.2.3	Auswertung vollfaktorieller Versuche mit FrF2	274
	<i>Beispiel „Lignit-Konversion“</i>	274
	<i>Beispiel „Prozessentwicklung“</i>	280
	<i>Beispiel „Wirbelschichtreaktor“</i>	285
7.3	Teilfaktorielle Pläne	288
7.3.1	Der Begriff Vermengung (Confounding)	288
7.3.2	Generatoren und Alias-Strukturen	288
7.3.3	Resolution, Aberration und andere Begriffe	290
7.3.4	Konstruktion teilfaktorieller Pläne mit R	293
	<i>Beispiel „rsm versus FrF2“</i>	295
	<i>Beispiel „Vier Aliasketten“</i>	297
7.3.5	Anzahl der Versuche versus möglichem Nutzen	300
7.3.6	Screening Fallstudien mit 2^{k-s} -Plänen	301
	<i>Beispiel „Schlagbiegefestigkeit“</i>	301
	<i>Beispiel „Radfahren“</i>	305
	<i>Beispiel „Spritzgießen“</i>	307
7.3.7	Aufspaltung vollfaktorieller 2^k -Pläne in Blöcke	311
	<i>Beispiel „Aufspaltung von 2^3-Plänen in Blöcke“</i>	313
	<i>Beispiel „Aufspaltung von 2^4-Plänen in Blöcke“</i>	313
	<i>Beispiel „Aufspaltung von 2^5-Plänen in Blöcke“</i>	315
7.4	Qualitätskriterien für Versuchspläne	316
7.4.1	Die Vorhersagevarianz	316
	<i>Beispiel „varfcn für (teil)faktorielle Pläne“</i>	319
7.4.2	Drehbarkeit, Orthogonalität und Uniformität	323
7.5	Response Surface Methodology (RSM)	325
7.5.1	Einführung	325
7.5.2	Bausteine der RSM, FO und SO Designs	328
7.5.3	Central Composite Designs (CCD)	330
7.5.4	RSM Werkzeuge	332
	<i>Beispiel „Vorhersagevarianz eines CCD-Plans“</i>	335
	<i>Beispiel „Schlechte Uniformität“</i>	336
	<i>Beispiel „Studie zur Uniformität“</i>	337
7.6	Beispiele für RSM-Anwendungen	339
7.6.1	Quadratische Regression mit rsm	339
	<i>Beispiel „2d-Optimum, orthogonal geblocktes CCD“</i>	339
	<i>Beispiel „3d-Optimum, ungeblocktes CCD“</i>	344
7.6.2	Umgang mit flachen Extrema und Sattelpunkten	347
	<i>Beispiel „2d-Sattel, Ridge-Design“</i>	347

	<i>Beispiel „Heli (4d-Sattel)“</i>	350
7.6.3	Eine Beispiel-Serie für eine sequentielle Optimierung bei drei Einflussgrößen	355
	<i>Beispiel „Backrezept (Teil 1)“</i>	355
	<i>Beispiel „Backrezept (Teil 2)“</i>	359
	<i>Beispiel „Backrezept (Teil 3)“</i>	366
7.7	Optimale Pläne	370
7.7.1	Einleitung	370
7.7.2	Optimalitätskriterien	371
7.7.3	Die Konstruktion optimaler Versuchspläne	373
	<i>Beispiel „D-optimale 2^{k-s}-Pläne“</i>	374
	<i>Beispiel „Optimale 3^3-Pläne und CCD“</i>	376
7.7.4	Verallgemeinerte (approximative) optimale Pläne	379
	<i>Beispiel „Approximativ optimaler Versuchsplan“</i>	379
7.8	Mixturpläne	381
7.8.1	Einleitung	381
	<i>Beispiel „Festigkeit von Polyblends“</i>	382
A	Mathematische Hilfsmittel	385
A.1	Substitutionsregel für Mehrfachintegrale	385
A.2	Integrale	385
A.3	Die Deltafunktion	386
A.4	Matrizen	387
A.4.1	Definitionen und Schreibweisen	387
A.4.2	Orthogonale und orthonormale Matrizen	388
A.4.3	Idempotente Matrizen	388
A.4.4	Mittelwertzentrierung mit idempotenter Matrix C	389
A.4.5	Der Rang einer Matrix	389
A.4.6	Der Nullraum einer $n \times p$ Matrix A	390
A.4.7	Die Inverse einer Matrix	390
A.4.8	Die Determinante einer quadratischen Matrix	391
A.4.9	Die Spur einer Matrix	391
A.4.10	Hadamard-Matrizen	392
R	Index der 95 R-Beispiele	393
S	Index der 61 SAS-Programme	397
	Literaturverzeichnis	399
	Sachregister	411

Tabellenverzeichnis

1.1	Symbole für diskrete Wahrscheinlichkeitsverteilungen	7
1.2	Symbole für kontinuierliche Wahrscheinlichkeitsverteilungen	8
1.3	Formelzeichen für Grundbegriffe wie Wahrscheinlichkeit, Zufallsvariable; Erwartungswert, Mittelwert; Varianz, Kovarianz, Streuung. . .	9
1.4	Formelzeichen für spezielle Begriffe der Stichproben- und Regressionsanalyse	10
2.1	Kumulanten von Binomial-, Poisson- und Gauß-Verteilung	31
3.1	Mittelwert, Streuung und charakteristische Funktion eindimensionaler diskreter Verteilungen	56
3.2	Dichtefunktionen eindimensionaler diskreter Verteilungen	57
3.3	Tabelle der kumulativen Normalverteilung	60
3.4	Mini-Tabelle der Quantilfunktion der Normalverteilung	61
3.5	Chi-Quadrat-Quantilfunktion	67
3.6	Mini-Tabelle der t-Verteilungsfunktion	71
3.7	t-Quantilfunktion	72
3.8	F-Quantilfunktion zu $p = 0.95$ und $p = 0.99$	76
3.9	Mittelwert, Streuung und charakteristische Funktion eindimensionaler kontinuierlicher Verteilungen	83
3.10	Dichtefunktionen eindimensionaler kontinuierlicher Verteilungen . . .	84
4.1	Bezeichnungen zu Mittelwert und Streuung von Stichproben	89
4.2	Verteilungsunabhängige Eigenschaften einer Stichprobe	90
4.3	Die linksseitige „ μ -power“-Theorie in zwölf Punkten	139
4.4	R-Werkstatt für die „ μ -power“-Analyse	140
4.5	R-Werkstatt für die „ μ -sample-size“-Analyse	144
5.1	Deterministische Größen im klassischen linearen Modell	153
5.2	Zufallsgrößen im klassischen linearen Modell	154
5.3	Eigenschaften der zufälligen Größen der Regressionsanalyse	155
5.4	Statistische Eigenschaften von prognostizierten Größen	158
5.5	Die ANOVA-Tabelle in Standardform	160
5.6	Die erweiterte ANOVA-Tabelle bei Mehrfachversuchen	161
5.7	Beispiel „Stahlhärte“ Teil 1: Programm	182

5.8	Beispiel „Stahlhärte“ Teil 2: Daten	182
5.9	Beispiel „Stahlhärte“ Teil 3: Output	183
5.10	Leave-one-out-Schätzer	202
6.1	Die ANOVA-Tabelle in der einfaktoriellen Varianzanalyse	231
7.1	Momentenmatrix und die Designmomente in einfachsten Fällen . . .	269
7.2	Ansatz für eine vollfaktorielle Versuchsanlage mit $k \leq 4$ Hauptfaktoren	270
7.3	Rekursive Entwicklung des vollfaktoriellen 2^4 -Versuchsplans	271
7.4	Zwei Beispiele von Generatoren mit Resolution 3 und 4	289
7.5	Ein 2^{4-1} -Versuchsplan mit Resolution 3	290
7.6	Ein 2^{4-1} -Versuchsplan mit Resolution 4	291
7.7	Resolution und Vermengung	292
7.8	Beispiel „Vier Aliasketten“	299
7.9	Aufspaltung von 2^k -Plänen in teilfaktorielle Blöcke	312
7.10	Verschiedene Modellgleichungen im Vergleich	318
7.11	Phasen der sequentiellen Optimierung eines Backrezepts	355
7.12	Optimalitätskriterien in der multiplen linearen Regression	372
7.13	D-optimale 2^{k-s} -Pläne, erzeugt mit <code>optFederov</code>	375
7.14	Optimale 3^3 -Pläne, erzeugt mit <code>optFederov</code> , und CCD-Pläne	378

Abbildungsverzeichnis

1.1	Screenshot nach Durchführung eines Programms in R-Studio	3
2.1	Elementarereignisse	15
2.2	Illustration zur bedingten Wahrscheinlichkeit	20
2.3	Kumulative Zweipunktverteilung „10-Euro-Münze“	21
2.4	Kumulative Maxwell-Boltzmann-Verteilung	22
2.5	Kumulative Binomialverteilung	22
2.6	Kumulative Poisson-Verteilung	23
2.7	Dichte der Zweipunktverteilung „10-Euro-Münze“	24
2.8	(Energie-)dichte der Maxwell-Boltzmann-Verteilung	24
3.1	Hypergeometrische Wahrscheinlichkeit	40
3.2	Binomialverteilung für $N = 8$, $p = 0.4$	43
3.3	Histogramm einer Binomialverteilung	44
3.4	Histogramm einer Poisson-Verteilung	45
3.5	Poissonverteilungen für verschiedene Parameter λ	48
3.6	Geometrische Verteilungen für $q = 0.4$	53
3.7	Negative Binomialverteilung	55
3.8	Gaußsche Glockenkurve	58
3.9	Dichte und kumulative Verteilung von drei Gauß-Verteilungen	59
3.10	Kumulative Normalverteilung und Umkehrfunktion	59
3.11	Quantilfunktion der Normalverteilung	61
3.12	Lorentzkurve	64
3.13	Dichten der Chi-Quadrat-Verteilung für verschiedene n	65
3.14	Kumulative Chi-Quadrat-Verteilung und Umkehrfunktion	66
3.15	Quantilfunktion der Chi-Quadrat-Verteilung für kleine p	66
3.16	Quantilfunktion der Chi-Quadrat-Verteilung für große p	67
3.17	Dichtefunktion der Student-Verteilung für verschiedene n	70
3.18	t-Quantilfunktion der Student-Verteilung	71
3.19	Weibull-Dichte und -Verteilungsfunktion	81
3.20	Zweidimensionale Normalverteilung	86
4.1	Illustration zur Tautologie	95
4.2	Stichprobenumfang versus relativer Messfehler	100
4.3	Stichprobenumfang versus relative Obergrenze der Varianz	107

4.4	Beispiel einer Qualitätsregelkarte	118
4.5	„Forellenbeispiel“: α - und β -Fehler	136
4.6	„Forellenbeispiel“: OC-Kurven	138
4.7	OC Kurven $\beta(\delta)$ für bekannte und unbekannte Varianz	142
4.8	Stichprobenumfang versus relative Verschiebung	146
5.1	Beispiel „Blutdruck“: Regressionsgerade	171
5.2	Beispiel „Sojabohnen“: Konfidenz- und Prognosetrichter	186
5.3	Beispiel „Fallbeschleunigung“: Regressionsgerade	188
5.4	Beispiel „Federkonstante“: Regressionsgerade	190
5.5	Beispiel „Fitness“ <i>mit</i> Ausreißer: Diagnoseplots	210
5.6	Beispiel „Fitness“ <i>ohne</i> Ausreißer: Diagnoseplots	211
5.7	Beispiel „Fitness“: Cook-Distanzen	212
5.8	Beispiel „Forbes“: Regressionsgeraden	214
5.9	Beispiel „Forbes“ ohne Logarithmierung: Diagnoseplots	216
5.10	Beispiel „Forbes“ mit Logarithmierung: Diagnoseplots	217
5.11	Beispiel „Pflanzenwachstum“: Nichtlineare Regression	222
5.12	Beispiel „Windgeschwindigkeit“: Nichtlineare Regression	224
6.1	Beispiel „ANOVA Stellen-Zugfestigkeit“: Boxplot	232
6.2	Beispiel „ANOVA Stellen-Zugfestigkeit“: Diagnoseplots	233
6.3	Beispiel „ANOVA Lehrmethoden-Zeugnis“: Boxplot	236
6.4	Beispiel „ANOVA Lehrmethoden-Zeugnis“: Diagnoseplots	237
6.5	Beispiel „ANOVA Glasart-Phosphorart-Helligkeit“: Boxplot	239
6.6	Beispiel „ANOVA Glasart-Phosphorart-Helligkeit“: Diagnoseplots	241
6.7	Beispiel „ANOVA Materialien-Temperaturen-Lebensdauer“: Boxplot	243
6.8	Beispiel „ANOVA Preise-Werbungen-Absatz“: Boxplots	245
6.9	Beispiel „ANOVA Preise-Werbungen-Absatz“: Wechselwirkungsplots	246
6.10	Beispiel „ANOVA Preise-Werbungen-Absatz“: Diagnoseplots	247
6.11	Beispiel „ANOVA Stoffe-Tage-Festigkeit-Block“: Boxplots	249
6.12	Beispiel „ANOVA-Druck-Chargen-Ausschuss-Block“: Boxplots	251
6.13	Beispiel „ANOVA Druck-Chargen-Ausschuss-Block“: Diagnoseplots	252
6.14	Beispiel „ANCOVA Truthahngewicht“: Scatterplot und Boxplot	259
6.15	Beispiel „ANCOVA Truthahngewicht“: Ausgleichsgerade	261
7.1	Beispiel „Lignit-Konversion“: Haupteffekte	277
7.2	Beispiel „Lignit-Konversion“: Wechselwirkungen	278
7.3	Beispiel „Lignit-Konversion“: (Halb)Normalplots	279
7.4	Beispiel „Prozessentwicklung“: Haupteffekte & Wechselwirkungen	281
7.5	Beispiel „Prozessentwicklung“: Halbnormalplots	282
7.6	Beispiel „Prozessentwicklung“: Würfeldiagramme	284
7.7	Beispiel „Schlagbiegefestigkeit“: Diverse Grafiken	303
7.8	Beispiel „Radfahren“: Diverse Grafiken	307
7.9	Beispiel „Spritzgießen“: Diverse Grafiken	310
7.10	Beispiel „ varfcn für vollfaktorielle Pläne“: Vorhersagevarianz	321
7.11	Beispiel „ varfcn für teilfaktorielle Pläne“: Vorhersagevarianz	322
7.12	Dreistufige Vorgehensweise in Response Surface Methoden	327

7.13	CCD-Designs mit zwei oder drei Hauptfaktoren	330
7.14	Beispiel „Vorhersagevarianz eines CCD-Plans“	336
7.15	Beispiel „Schlechte Uniformität“: Diverse Grafiken	337
7.16	Beispiel „Studie zu Uniformität“	338
7.17	Beispiel „3d-Optimum“: Contourplots	346
7.18	Beispiel „2d-Sattel 3 ² “ Ridge-Design: Contourplot	349
7.19	Beispiel „Heli (4d-Sattel)“: Contour- und Perspektivplots	353
7.20	Beispiel „Backrezept“: Hill-Climbing erster Ordnung	358
7.21	Beispiel „Backrezept“: Hill-Climbing zweiter Ordnung	366
7.22	Beispiel „Backrezept“: Finales Experiment	369
7.23	Beispiel „Festigkeit von Polyblends“: Mixtur-Plot	384

Kapitel 1

Einleitung

1.1 Hauptbegriffe

Statistik begegnet uns überall im täglichen Leben, wo Massendaten komprimiert werden.

Mathematische Statistik (auch: schließende Statistik, induktive Statistik, Inferenzstatistik oder inferentielle Statistik) ist das Teilgebiet der Statistik, das die Methoden und Verfahren der Statistik mit mathematischen Mitteln analysiert bzw. mit ihrer Hilfe erst begründet. Die mathematische Grundlage der Mathematischen Statistik ist die *Wahrscheinlichkeitstheorie*: Typischerweise werden die vorhandenen Daten einer Stichprobe (d.h. Beobachtungen, Messungen) als Realisierungen von stochastisch unabhängigen Zufallsvariablen interpretiert, so dass wahrscheinlichkeitstheoretische Methoden zur Untersuchung des stochastischen Verhaltens der Beobachtungen anwendbar sind. Die *Stochastik* fasst als Oberbegriff die Gebiete Wahrscheinlichkeitstheorie und Mathematische Statistik zusammen¹.

Versuchsplanung (Design of Experiments) ist eine Anwendung der Mathematischen Statistik (man sagt daher auch *statistische Versuchsplanung*), die – beispielsweise bei der industriellen Entwicklung von neuen Produkten und Fertigungsprozessen – helfen kann, mit möglichst geringem Aufwand reproduzierbare Ergebnisse zu erhalten.

Die Themen dieses Buchs sind Wahrscheinlichkeitstheorie, Mathematische Statistik und Versuchsplanung, aber nicht Statistische Physik. Letztere ist eine fundamentale physikalische Theorie, mit deren Hilfe u.a. Gesetze der Thermodynamik abgeleitet und begründet werden.

Die *mathematische Statistik* verwendet für manche grundlegenden Begriffe der Naturwissenschaft ihre *eigenen Namen*. Zum Beispiel benutzt der Statistiker für das naturwissenschaftliche „Messen (Beobachten)“ den Terminus „Ziehen einer Stichprobe“, oder er verwendet das Wort „Schätzen“ für das naturwissenschaftliche „Be-

¹Das Wort *Stochastik* stammt aus der Frühzeit der Wahrscheinlichkeitsrechnung (einer der Bernoullis hat es geprägt [Swoboda 1971, Seite 174]). Es wurde nach dem Zweiten Weltkrieg zum wissenschaftlichen Modewort. Günter Menges definiert in seinem „Grundriß der Statistik, 1“ lapidar: „Stochastik ist alles, was auf der Wahrscheinlichkeitsrechnung aufgebaut ist oder sonstwie mit ihr zu tun hat“ [Menges 1968]

stimmen (determine)“. Statistisches „Schätzen“ hat also nichts zu tun mit dessen umgangssprachlichen Gebrauch im Sinn von „Vermuten“. Was schließlich der Naturwissenschaftler mit „(statistischem) Messfehler“ oder „Rauschen“ bezeichnet, definiert die Statistik quantitativ mit Begriffen wie Konfidenzbereich, Konfidenzintervall, Intervallschätzwert, Reststreuung, innere Varianz.

1.2 Zielgruppen

Das Buch wendet sich an Naturwissenschaftler und Naturwissenschaftlerinnen und an Ingenieure und Ingenieurinnen in der Industrie.

1.2.1 Einsteiger/innen

Für **R-Einsteiger/innen** mit Windows-PC stehen 95 R-Stand-Alone-Beispiele bereit, die unabhängig und größtenteils isoliert voneinander lauffähig sind. (Lediglich das Programm `R.42.R` wird von einigen anderen Programmen eingebunden und verwendet). Steigen Sie direkt ein mit „learning-by-doing“:

- Installieren Sie RStudio (einmalig) – eine gute, gängige Oberfläche für interaktives R. Installationsanleitungen und Anleitungen zur Bedienung von RStudio finden Sie im Internet, z.B. https://www.youtube.com/watch?v=X_Mxya2Fis0 und <https://www.youtube.com/watch?v=tyvEHQszZJs>.
- Kopieren Sie das R-Skript-Verzeichnis gemäß der Anleitung von Abschnitt 1.6, Seite 12.
- Entscheiden Sie sich für ein attraktives Beispiel aus dem Lehrbuch, etwa in Abschnitt 7.2.3.1, auf S. 274.
- Öffnen Sie das entsprechende Programm `R.69.R` mit R-Studio.
- Starten Sie das Skript zeilenweise mit `Ctrl+Enter` (der Cursor wandert nach unten).
- Stoppen Sie bei `MEPlot(Fit)`
- Das Ergebnis sollte ungefähr so aussehen wie der Screenshot 1.1 auf Seite 3.

Für **Fortgeschrittene** mit Vorwissen in Höherer Mathematik (d.h. Vektoren und Matrizen, Differential- und Integralrechnung) bietet sich die Chance zu einem vertieften Verständnis der Statistik in den Kapiteln 4 bis 6.

Das Kapitel 4 (Stichprobenanalyse) erläutert Grundbegriffe der Mathematischen Statistik wie Statistische Messfehler, Konfidenzbereich in Abhängigkeit vom Stichprobenumfang, und liefert eine Einführung in Hypothesentests wie t-Test und andere, inklusive der Berechnung der Teststärke (Power-Analyse).

Kapitel 5 behandelt die Grundlagen der Regressionsanalyse: Annahmen, Schätzung, Residuen, Quadratsummen, Verteilungseigenschaften, Hypothesentests, Konfidenzintervalle, Prognoseintervalle, Modelldiagnose, Nichtlineare Regression.

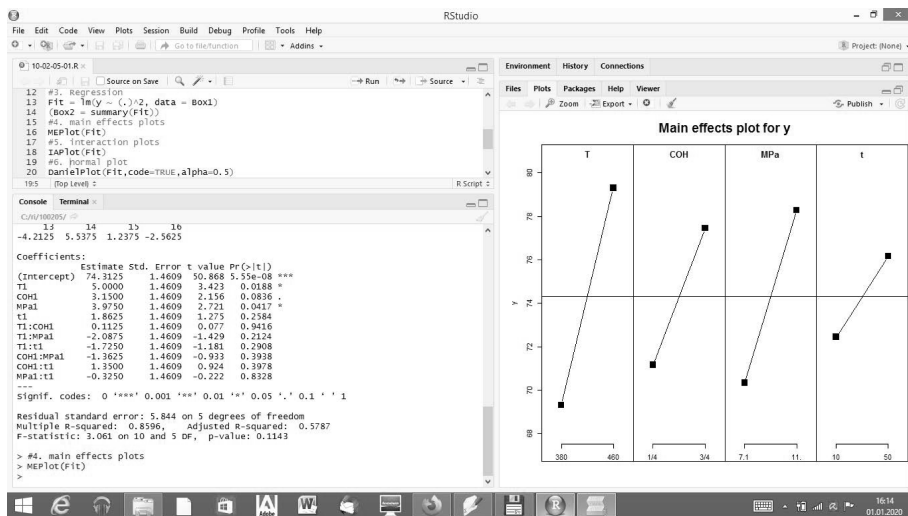


Abb. 1.1: Screenshot nach Durchführung eines Programms in R-Studio

Kapitel 6 ist der Varianzanalyse mit festen Effekten gewidmet. Die Themen sind einfache Varianzanalyse, zweifache Varianzanalyse mit einfacher und mehrfacher Besetzung, Lateinische und Griechisch-Lateinische Quadrate, Kovarianzanalyse.

1.2.2 Physiker/innen

Fortgeschrittene schließlich, die über Höhere Mathematik hinaus mit der **Dirac'schen Deltafunktion** in Berührung gekommen sind, finden im Lehrbuch die Grundlagen der Wahrscheinlichkeitstheorie in knapper Form dargestellt, in den Kapiteln 2 und 3:

Das Kapitel 2 beginnt mit einer kurzen Einführung in die Wahrscheinlichkeitstheorie (Auszug aus dem Vorlesungsmanuskript Theoretische Physik IV, „Thermodynamik und Statistische Physik“ von Prof. Dr. Friederike Schmid, Universität Mainz). Es wird ergänzt durch anwendungsbezogene Themen wie Kombinatorik, Quantile, Symmetrierelationen, quadratische Formen.

Das Kapitel 3 präsentiert acht eindimensionale diskrete Verteilungen (hypergeometrische Verteilung, Binomialverteilung, Poisson-Verteilung, geometrische Verteilung, negative Binomialverteilung u.a.), zehn eindimensionale kontinuierliche Verteilungen (Gauß-Verteilung, Cauchy-Verteilung, Chi-Quadrat-Verteilung, nichtzentrale Chi-Quadrat-Verteilung, Student-Verteilung, nichtzentrale Student-Verteilung, F-Verteilung, nichtzentrale F-Verteilung, Exponentialverteilung und Weibull-Verteilung) und zwei mehrdimensionale kontinuierliche Verteilungen (n -dimensionale und zwei-dimensionale Gauß-Verteilung).

Die Deltafunktion, ein von dem britischen Quantenphysiker Paul Dirac (1902-1984) erfundenes Werkzeug, dient hier dem Zweck, Statistiker von hochabstrakter

Mathematik (Maßtheorie) zu verschonen. Sie ist außergewöhnlich effektiv und elegant:

Diskrete und kontinuierliche Verteilungsdichten verschmelzen durch die δ -Funktion zu einer Einheit. Man könnte z.B. die Tabellen 3.2 und 3.10 auf den Seiten 57 und 84 zu einer einzigen Tabelle zusammenfassen (untereinander schreiben).

Chiquadrat-, Student-, u.a. Verteilungsdichten werden mit der δ -Funktion elegant hergeleitet. Siehe (3.53), (3.73) u.a.

Die Verteilung einer Funktion von Zufallsvariablen ergibt sich mit Hilfe der δ -Funktion wie von selbst (Abschnitt 2.5.1).

Für die Verallgemeinerung auf mehrere Zufallsvariable gilt Ähnliches (Abschnitt 2.5.3).

Die Wahrscheinlichkeitsdichte ist gleich dem Erwartungswert der δ -Funktion, und die kumulative Wahrscheinlichkeitsverteilung ist gleich dem Erwartungswert der Θ -Funktion (Abschnitt 2.5.2).

Eine halbe Seite (Anhang A.3) genügt, um die Definition und die Eigenschaften der Deltafunktion zu formulieren.

Es ist bemerkenswert, dass so gut wie alle Lehrbücher der Wahrscheinlichkeitstheorie auf den Gebrauch dieses effizienten Werkzeugs verzichten.

Schliesslich sei noch auf Beiträge hingewiesen, die einen Bezug zum physikalischen Praktikum haben: Man findet sie auf den Seiten 90 (Fehlerrechnung), 98 (Fehlerbalken), 184, 186 und 189 (ausführliche Beispiele mit einfacher linearer Regression), 193 (Regression von Hand lediglich mit Taschenrechner).

1.2.3 SAS-Anwender/innen

Zum vorliegenden Buch werden 61 SAS-Statistik-Programme zum Download bereitgestellt. Man kann sie nachschlagen in einem Index auf Seite 397. Dieser ist in zwei Gruppen 1 und 2 unterteilt:

Gruppe 1 enthält 37 SAS-Programme zur Erstellung von Grafiken und statistischen Tabellen in Kapitel 2 bis 4.

- SAS-Grafiken zur Wahrscheinlichkeitstheorie finden Sie zwischen Seite 21 und Seite 95.
- In statistischen Tabellen wird auch heute noch aus pädagogischen Gründen nachgeschlagen, um Standard-Aufgaben der schließenden Statistik manuell zu lösen. Solche Tabellen finden Sie zwischen Seite 60 und Seite 76.

Gruppe 2 enthält 24 alternative SAS-Lösungen zu ausgewählten R-Beispielen, (S.38.sas) bis (S.61.sas). Für Anwender/innen mit geringer R-, aber viel SAS-Erfahrung kann die Gegenüberstellung von R-Lösung und SAS-Lösung hilfreich sein beim Einstieg in die R-Programmierung.

1.3 Anmerkungen zur Notation

R-Code und R-Output hinterlegen wir grau.

Nummern von Gleichungen haben die Form (C.Lfdnr). Dabei ist C die Nummer eines Kapitels (Chapter) oder der Buchstabe eines Appendix, und Lfdnr eine laufende Nummer pro Kapitel. Die Klammern sind obligatorisch. Beispiele: (7.12) oder (A.5).

Nummern von Tabellen haben die Form „Tabelle C.Lfdnr“, also ohne Klammern, aber mit dem Wort Tabelle.

Nummern von Abbildungen haben die Form „Abbildung C.Lfdnr“

Nummern von Abschnitten haben die Form C.S, C.S.sS oder C.S.sS.ssS, ohne Klammern!. Dabei sind S, sS und ssS die Nummern von Section, Subsection und Subsubsection.

Matrizen schreiben wir fett und groß. Beispiele: \mathbf{X} , $\mathbf{\Sigma}$.

Zufallsvariablen unterstreichen wir. Beispiele: \underline{x} , \underline{y} , \underline{z} , \underline{t} . Ausnahmen: siehe unten.

Arithmetische Stichproben-Mittelwerte überstreichen wir. Obwohl sie selbst Zufallsvariablen sind, verzichten wir auf eine extra Unterstreichung. Beispiele: \bar{x} , \bar{y} , \bar{z} .

Zufallsvektoren sind klein, fett und unterstrichen. Beispiel: $\underline{\mathbf{y}} = (\underline{y}_1 \cdots \underline{y}_n)'$.

Operatoren der Wahrscheinlichkeitsrechnung kennzeichnen wir mit kalligraphischen Lettern \mathcal{P} , \mathcal{E} , \mathcal{D} .

$\mathcal{P}(A)$: Wahrscheinlichkeit für Ereignis A

$\mathcal{P}(A|B)$: Bedingte Wahrscheinlichkeit

$\mathcal{E}(\underline{z}_i)$: Erwartungswert von \underline{z}_i , $\mathcal{E}(\underline{\mathbf{z}})$: Vektor der Erwartungswerte von $\underline{\mathbf{z}}$

$\mathcal{D}^2(\underline{z}) = \mathcal{D}(\underline{z}, \underline{z})$: Varianz von \underline{z} (2.31),

$\mathcal{D}(\underline{z}_1, \underline{z}_2)$: Kovarianz von \underline{z}_1 und \underline{z}_2 (2.32);

$\mathcal{D}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')$: Kovarianzmatrix der Vektoren $\underline{\mathbf{x}}$ und $\underline{\mathbf{z}}$ (2.34)

Griechische Buchstaben verwenden wir (in der Wahrscheinlichkeitstheorie) für unbekannte – d.h. zu schätzende – Parameter. Beispiele: μ , σ^2 , β .

Schätzwerte aus einer Stichprobe kennzeichnen wir durch Überdachung. Beispiele: $\hat{\mu}$, $\hat{\sigma}$, $\hat{\mathcal{D}}^2(\underline{z})$, $\hat{\mathcal{D}}(\underline{x}, \underline{z})$, $\hat{\mathcal{D}}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')$. Auf eine extra Unterstreichung verzichten wir, obwohl Schätzwerte immer zufällig sind.

Bemerkung: In diesem Buch werden *nur erwartungstreue Schätzer* verwendet, d.h. wir fordern $\mathcal{E}(\hat{\beta}) = \beta$ für alle Schätzer $\hat{\beta}$ eines Parameters β .

(z.B. $\mathcal{E}(\hat{\mu}) = \mu$, $\mathcal{E}(\hat{\sigma}^2) = \sigma^2$, $\mathcal{E}(\hat{\mathcal{D}}^2(\underline{z})) = \mathcal{D}^2(\underline{z})$,

$\mathcal{E}(\hat{\mathcal{D}}(\underline{x}, \underline{z})) = \mathcal{D}(\underline{x}, \underline{z})$, $\mathcal{E}(\hat{\mathcal{D}}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')) = \mathcal{D}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')$ usw.)

Beweise: Folgt $A=B$ aus Gleichung x oder aus Abschnitt x , so schreiben wir, um Platz zu sparen, das x direkt über das Gleichheitszeichen. Im einzelnen:

$A \stackrel{(x)}{=} B$ heißt „Beweis siehe Gleichung oder Abschnitt (x) “.

$A \stackrel{(x)}{:=} B$ heißt „Definition siehe Gleichung oder Abschnitt (x) “ und

$A \stackrel{(x)}{\leadsto} B$ heißt „Aus A folgt B mittels Gleichung oder Abschnitt (x) “.

$A \stackrel{5.3.3\textcircled{8}}{=} B$ heißt „Beweis siehe Abschnitt 5.3.3 Zeile $\textcircled{8}$ “.

$A \stackrel{\text{Tab.5.3}\textcircled{c}}{=} B$ heißt „Beweis siehe Tabelle 5.3 Zeile $\textcircled{7}$ Spalte e“.

$A \stackrel{\checkmark}{=} B$ heißt „Beweis offensichtlich“.

$A \stackrel{zz}{=} B$ heißt „ist zu beweisen („zu zeigen“)“.

Wortwörtlich übernommene Textpassagen: (z.B. Aufgabentexte) werden in Anführungszeichen gesetzt, und der Kurztitel wird mit dem Hinweis „Ww.“ versehen. Beispiele

- Seite 101 „Spülmittel“, Aufgabe [Wiki 01, Ww.]
- Seite 108 „Betonstahlmatten“, Aufgabe [Dürr/Walter 1987, S. 136 Ww.]
- Seite 134 „Forellenbeispiel“, Aufgabe [Wiki 02, Ww.]

1.4 Symbole und Formelzeichen

Die Formelzeichen dieses Buches sind *eindeutig*, *systematisch* und können – dadurch bedingt – manchmal *recht kompliziert* erscheinen. Die Formelzeichen selbst werden in vier Tabellen vorgestellt:

Die erste Tabelle Tabelle 1.1 S. 7 enthält Symbole und Formelzeichen zu *sieben diskreten eindimensionalen Wahrscheinlichkeitsverteilungen*

Die zweite Tabelle Tabelle 1.2 S. 8 enthält Symbole und Formelzeichen mit Bezug zu *zwölf kontinuierlichen eindimensionalen Wahrscheinlichkeitsverteilungen*. Man beachte das *einheitliche Schema* für Verteilungsdichten, kumulative Verteilungsfunktionen und Quantilfunktionen. Ihre Bezeichnungen sind außerordentlich *effizient*. Mit „Effizienz“ ist gemeint, dass die Notation der Quantilfunktionen als eindeutige Umkehrfunktion vermöge der Beziehungen (2.91) und (2.92) quasi „von selbst rechnet“, d.h. dem Statistiker streckenweise das Denken abnimmt – ähnlich wie die Diracsche Bra- und Ket-Notation dem Physiker das Denken.

Die dritte Tabelle Tabelle 1.3 S. 9 enthält Formelzeichen zu *Grundbegriffen* wie Wahrscheinlichkeit, Zufallsvariable, Erwartungswert, Mittelwert, Streuung, Varianz und Kovarianz. Der Vollständigkeit halber werden darin auch die etwas weniger systematischen, aber auch gebräuchlichen Bezeichnungen $K_{\underline{z}_1 \underline{z}_2}$ und $\sigma_{\underline{z}}^2$ erwähnt. Bemerkung: Diese Schreibweisen verwenden wir aber nicht.

Die vierte Tabelle Tabelle 1.3 S. 9 enthält Formelzeichen zu *speziellen Begriffen* der Stichproben- und Regressionsanalyse.

Tabelle 1.1: Symbole zur Bezeichnung eindimensionaler diskreter Wahrscheinlichkeitsverteilungen.

Symbol	Name	Def.	Dichte
$\text{Hyp}_{n,N,M}$	Hypergeometrische Verteilung	3.1.1	$f_{\text{Hyp}_{n,N,M}}(z) = \sum_{x=\max(0,k-n)}^{\min(m,k)} \delta(z-x) P_{\text{Hyp}_{n,N,M}}(x)$
$\text{Nhyp}_{n,N,M}$	Negative Hypergeom. Verteilung	3.1.2	$f_{\text{Nhyp}_{n,N,M}}(z) = \sum_{x=m}^{m+N-m} \delta(z-x) P_{\text{Nhyp}_{n,N,M}}(x)$
$\text{Bin}_{N,p}$	Binomialverteilung	3.1.3	$f_{\text{Bin}_{N,p}}(z) = \sum_{x=0}^N \delta(z-x) P_{\text{Bin}_{N,p}}(x)$
Poi_{λ}	Poisson-Verteilung	3.1.4	$f_{\text{Poi}_{\lambda}}(z) = \sum_{x=0}^{\infty} \delta(z-x) P_{\text{Poi}_{\lambda}}(x)$
GeoA_q	geometr. Verteilung Variante 'A'	3.1.5	$f_{\text{GeoA}_q}(z) = \sum_{x=0}^{\infty} \delta(z-x) P_{\text{GeoA}_q}(x)$
GeoB_q	geometr. Verteilung Variante 'B'	3.1.5	$f_{\text{GeoB}_q}(z) = \sum_{x=0}^{\infty} \delta(z-x) P_{\text{GeoB}_q}(x)$
$\text{Nbin}_{q,v}$	negative Binomialv.	3.1.6	$f_{\text{Nbin}_{q,v}}(z) = \sum_{x=0}^{\infty} \delta(z-x) P_{\text{Nbin}_{q,v}}(x)$

Tabelle 1.2: Symbole zur Bezeichnung eindimensionaler kontinuierlicher Wahrscheinlichkeitsverteilungen

Sym- bol	Name der Verteilung	Defi- nition	Dichte (2.23)	kumulative Verteilung (2.12)	Quantil- funktion 2.7
*	allgemein		$f_*(z)$	$F_*(z)$	$F_*^{-1}(p)$
$N_{0,1}$	Normal- Verteilung	(2.87)	$f_{N_{0,1}}(z)$	$F_{N_{0,1}}(z)$	$F_{N_{0,1}}^{-1}(p)$
$N_{\mu,\sigma}$	Gauß- Verteilung	3.2.1	$f_{N_{\mu,\sigma}}(z)$	$F_{N_{\mu,\sigma}}(z)$	$F_{N_{\mu,\sigma}}^{-1}(p)$
$C_{\mu,\lambda}$	Cauchy- Verteilung	3.2.2	$f_{C_{\mu,\lambda}}(z)$	$F_{C_{\mu,\lambda}}(z)$	$F_{C_{\mu,\lambda}}^{-1}(p)$
χ_n^2	χ^2 -Verteilung	3.2.3	$f_{\chi_n^2}(z)$	$F_{\chi_n^2}(z)$	$F_{\chi_n^2}^{-1}(p)$
$\chi_{n,\lambda}^2$	nichtzentrale χ^2 -Verteilung	3.2.4	$f_{\chi_{n,\lambda}^2}(z)$	$F_{\chi_{n,\lambda}^2}(z)$	$F_{\chi_{n,\lambda}^2}^{-1}(p)$
t_n	t-Verteilung	3.2.5	$f_{t_n}(z)$	$F_{t_n}(z)$	$F_{t_n}^{-1}(p)$
$t_{n,\mu}$	nichtzentrale t-Verteilung	3.2.6	$f_{t_{n,\mu}}(z)$	$F_{t_{n,\mu}}(z)$	$F_{t_{n,\mu}}^{-1}(p)$
$G_{a,b}$	Gleich- verteilung	Tab. 3.10①	$f_{G_{a,b}}(z)$	$F_{G_{a,b}}(z)$	$F_{G_{a,b}}^{-1}(p)$
$F_{m,n}$	F-Verteilung	3.2.7	$f_{F_{m,n}}(z)$	$F_{F_{m,n}}(z)$	$F_{F_{m,n}}^{-1}(p)$
$F_{m,n,\lambda}$	nichtzentrale F-Verteilung	3.2.8	$f_{F_{m,n,\lambda}}(z)$	$F_{F_{m,n,\lambda}}(z)$	$F_{F_{m,n,\lambda}}^{-1}(p)$
E_β	Exponential- Verteilung	3.2.9	$f_{E_\beta}(z)$	$F_{E_\beta}(z)$	$F_{E_\beta}^{-1}(p)$
$W_{T,k}$	Weibull- Verteilung	3.2.10	$f_{W_{T,k}}(z)$	$F_{W_{T,k}}(z)$	$F_{W_{T,k}}^{-1}(p)$

Tabelle 1.3: Formelzeichen für Grundbegriffe wie Wahrscheinlichkeit, Zufallsvariable; Erwartungswert, Mittelwert; Varianz, Kovarianz, Streuung.

Formelzeichen	kurz	Bedeutung	Def.
$\mathcal{P}(A)$		Wahrscheinlichkeit für Ereignis A	(2.8)
$\mathcal{P}(A B)$		Bedingte Wahrscheinlichkeit	(2.9)
$\tilde{\alpha}$		Irrtumswahrscheinlichkeit, Signifikanzniveau	4.2.1
$\underline{z}, \underline{x}, \underline{y}$, etc.		Zufalls-Variable	2.2.3
$\underline{\mathbf{z}} = (z_1 \ z_2 \ \dots)'$		Zufalls-Vektor	
$\mathcal{E}(\underline{z})$	$= \mu$	Erwartungswert von \underline{z}	(2.29)
$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$	$= \hat{\mu}$	Stichproben-Mittelwert der z_i	(4.1)
$\mathcal{E}(\underline{\mathbf{z}})$	$= \underline{\boldsymbol{\mu}}$	Vektor der Erwartungswerte von $\underline{\mathbf{z}}$	
$\mathcal{D}(z_1, z_2)$	$\stackrel{n.v.}{=} K_{z_1 z_2}$	Kovarianz von z_1 und z_2	(2.32)
$\mathcal{D}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')$	$\stackrel{n.v.}{=} K_{\underline{\mathbf{z}}_1 \underline{\mathbf{z}}_2}$	Kovarianzmatrix von $\underline{\mathbf{x}}$ und $\underline{\mathbf{z}}$, Streuungsmatrix	(2.34)
$\hat{\mathcal{D}}(\underline{x}, \underline{z}), \hat{\mathcal{D}}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')$		Schätzer von $\mathcal{D}(\underline{x}, \underline{z})$ bzw. $\mathcal{D}(\underline{\mathbf{x}}, \underline{\mathbf{z}}')$	
$\mathcal{D}^2(\underline{z}) = \mathcal{D}(\underline{z}, \underline{z})$	$= \sigma^2$	Varianz von \underline{z}	(2.31)
$\hat{\mathcal{D}}^2(\underline{z}) = \hat{\mathcal{D}}(\underline{z}, \underline{z})$	$\stackrel{n.v.}{=} \sigma_{\underline{z}}^2$ $= \hat{\sigma}^2$	Stichprobenvarianz, Schätzer von $\mathcal{D}^2(\underline{z})$	(4.2)
$\mathcal{D}(\underline{z}) = \sqrt{\mathcal{D}(\underline{z}, \underline{z})}$	$= \sigma$	Standardabweichung von \underline{z}	2.31
$\hat{\mathcal{D}}(\underline{z}) = \sqrt{\hat{\mathcal{D}}(\underline{z}, \underline{z})}$	$= \hat{\sigma}$	Stichprobenstandardabweichung, standard error	Tab.4.1④

Tabelle 1.4: Formelzeichen für spezielle Begriffe der Stichproben- und Regressionsanalyse

Formelzeichen	Bedeutung	Def.
$\hat{\mu} = (\underline{z}_1 + \dots + \underline{z}_n)/n$	Stichproben-Mittelwert	(4.1)
$\hat{\sigma}$	Stichproben-Standardabweichung, standard error	(4.10)
$\hat{\mathcal{D}}(\hat{\mu})$	standard error of the mean	(4.11)
$\tilde{\alpha}$	Irrtumswahrscheinlichkeit, Signifikanzniveau	4.2.1
X	Design-Matrix bei der linearen Regression	(5.5),(5.16)
$\underline{\mathbf{y}} = (\underline{y}_1 \ \underline{y}_2 \ \dots \ \underline{y}_n)'$	Zufalls-Vektor der n Meßwerte	(5.28)
$\hat{\mathbf{y}} = (\hat{y}_1 \ \hat{y}_2 \ \dots \ \hat{y}_n)'$	Schätzer (Fit) für $\underline{\mathbf{y}}$	(5.33)
$\underline{\boldsymbol{\varepsilon}} = (\underline{\varepsilon}_1 \ \dots \ \underline{\varepsilon}_n)'$	Statistischer Fehler im Regressions- ansatz	(5.1),(5.27)
σ^2	unbekannte Varianz von $\underline{\varepsilon}_i$	(5.8),(5.27)
$\hat{\sigma}^2$	Schätzer von σ^2 , mean square error	(5.37)
$\boldsymbol{\beta} = (\beta_0 \ \beta_1 \ \dots \ \beta_p)'$	unbekannte Regressionsparameter	(5.1)
$\hat{\boldsymbol{\beta}} = (\hat{\beta}_0 \ \hat{\beta}_1 \ \dots \ \hat{\beta}_p)'$	Schätzer von $\boldsymbol{\beta}$	(5.30)
$\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1 \ \dots \ \hat{\varepsilon}_n)'$	Residuen	(5.34)
$\hat{\varepsilon}$	Prognosefehler	Tabelle 5.4(10)p
\underline{R}^2	Bestimmtheitsmaß	(5.48)
$\underline{R}_{\text{adjusted}}^2$	korrigiertes Bestimmtheitsmaß	(5.49)
$\xi, \underline{\eta}$	Schwerpunktskoordinaten bei der einfachen linearen Regression	5.8.1
$\ \xi\ ^2, \ \underline{\eta}\ ^2$	Norm von ξ , Norm von $\underline{\eta}$	5.8.1

1.5 Gliederung des Buches

Dieses Lehrbuch spannt einen anwendungsorientierten Bogen von den Grundlagen der Wahrscheinlichkeitstheorie (Kapitel 2 und 3) über zentrale Teile der Mathematischen Statistik (Kapitel 4 bis 6) bis zu praktischen Anwendungen der Statistischen Versuchsplanung (Kapitel 7). Im Inhaltsverzeichnis werden auch die Beispiele aufgelistet, eingebettet in ihren jeweiligen Kontext.

Das Kapitel 2 bildet den ersten Schwerpunkt. Es gibt eine kurze Einführung in die *Wahrscheinlichkeitstheorie*. Darin werden die Begriffe Zufallsversuch, Wahrscheinlichkeit eines Ereignisses, Zufallsvariable, Erwartungswert, u.a. erklärt. Die Wahrscheinlichkeitsverteilung diskreter oder kontinuierlicher Zufallsvariabler sowie von Funktionen derselben wird in eleganter Weise mit Hilfe der Dirac'schen Deltafunktion berechnet. Am Ende von Kapitel 2 findet man weitere für die Mathematische Statistik wichtige Teile der Wahrscheinlichkeitstheorie: Rechenregeln, Symmetrierelationen, Kombinatorik, Quantile. Letztere werden formal als Umkehrfunktionen notiert; das Unbehagen, das einen Statistik-Anfänger befällt, wenn vom Quantil die Rede ist, wird so an einen Formalismus delegiert.

Die Kapitel 3 behandelt die wichtigsten *statistischen Verteilungen*, die zur Beschreibung der (hypothetischen) Verteilung realer Daten verwendet werden. Eindimensionale diskrete Verteilungen, eindimensionale kontinuierliche Verteilungen und mehrdimensionale Gauß-Verteilungen.

Einen zweiten Schwerpunkt dieses Buches bilden die Kapitel 4 und 5 mit den Themen Stichproben und lineare Regression (Ausgleichsrechnung, Fitting).

Ein zentraler Begriff der mathematischen Statistik ist die *Schätzmethode*, nach der Schätzfunktionen für unbekannte Parameter einer statistischen Grundgesamtheit konstruiert werden. Ein klassisches Kriterium für die Beurteilung einer Schätzmethode ist die *Erwartungstreue* (vgl. Gleichungen (4.3), (4.4); ein klassisches Beispiel einer Schätzmethode ist die *Methode der kleinsten Quadrate* (vgl. Gl. (5.11), (5.30), (5.37)). Mit der Hilfe von Schätzmethoden werden Bereichsschätzungen von Parametern der Grundgesamtheit durch *Konfidenzintervalle* (im Abschnitt 4.2) ermittelt. In diesen Kontext gehört auch ein zentrales Thema der Versuchsplanung, die Ermittlung des notwendigen Stichprobenumfangs (sample size): Die Beispiele 4.2.7.1 und 4.2.9.1 demonstrieren den Zusammenhang zwischen sample size und Messfehler (Fehler erster Art, α -Fehler). Der zusätzliche Einfluss von Fehlern *zweiter* Art (β -Fehler) auf die sample size wird im Abschnitt 4.4 (Power-Analyse) behandelt, sowohl analytisch (Abschnitt 4.4.7) als auch mit R-Programmen (Abschnitt 4.4.8).

In den Abschnitten 4.3 und 5.6 wird gezeigt und durch zahlreiche Beispiele (manuell, R, SAS) demonstriert, wie konkrete Vermutungen über die Grundgesamtheit durch geeignete *statistische Tests* bestätigt oder abgelehnt werden.

Im Kapitel 5 wird die Methode der multiplen linearen Regression ausführlich hergeleitet und an Hand von Beispielen mit kommentiertem Computeroutput veranschaulicht.

Kapitel 6 behandelt die Themen Varianzanalyse mit festen Effekten (einfach, zweifach mit einfacher und mehrfacher Besetzung), lateinische bzw. griechisch-lateinische Quadrate und Kovarianzanalyse.

Den dritten, überwiegend praxisbezogenen Schwerpunkt dieses Buches bildet das Kapitel 7 Versuchsplanung mit den Haupt-Themen vollfaktorielle Pläne, teilfaktorielle Pläne und Response Surface Methoden. Weitere Themen sind optimale Pläne und Mixturpläne.

Anhang A enthält mathematische Sätze über Integrale, Matrizen etc., die bei den Beweisen im Hauptteil verwendet werden. Die Anhänge R und S sind Verzeichnisse der 95 R-Programme `R_1.R` bis `R_95.R` und der 61 statistischen SAS-Programme `S_1.sas` bis `S_61.sas`.

1.6 Datenmaterial zu diesem Buch im Internet

Für Leser/innen dieses Buches gibt es eine Möglichkeit, die R-Programme und SAS-Programme zu dem Buch (siehe Auflistung in den Anhängen R und S) aus dem Internet herunterzuladen. Man findet sie auf der Verlagsseite des Buches unter www.wiley-vch.de/ISBN9783527346295 zum Download.

Kapitel 2

Wahrscheinlichkeitstheorie

Teile des folgenden Kapitels sind angelehnt an das Vorlesungsmanuskript „Theoretische Physik IV“ von Prof. Friederike Schmid, Universität Mainz, mit freundlicher Genehmigung der Autorin. (Quelle: [Uni-Mainz 2018]).

2.1 Was ist Wahrscheinlichkeit?

2.1.1 Ausgangssituation

In Experimenten lässt sich das Ergebnis eines Versuchs in der Regel nicht genau vorhersagen. Das hat mehrere Gründe. Falls ein Kontinuum an Messwerten möglich ist, gilt prinzipiell: Punkte im Kontinuum sind physikalisch nicht messbar. Aber selbst wenn nur diskrete Messwerte möglich sind, lässt sich meistens nicht hundertprozentig vorhersagen, wie ein Versuch ausgehen wird. Besonders deutlich wird das zum Beispiel beim Wurf eines Würfels oder einer Münze. Üblicherweise sagt man, man erhält ein bestimmtes Ergebnis mit einer bestimmten „Wahrscheinlichkeit“. Doch was ist damit gemeint?

Ein möglicher Präzisierungsversuch wäre: Wir „definieren“ die „Wahrscheinlichkeit“ $p(E_i)$ eines Ereignisses E_i als relative Häufigkeit dieses Ergebnisses nach „unendlich“ vielen Versuchen, also in mathematischer Notation $p(E_i) = \lim_{N \rightarrow \infty} \frac{n(E_i)}{N}$. Doch diese „Definition“ ist in mehrfacher Hinsicht problematisch. Zum einen setzt sie voraus, dass im Prinzip unendlich viele Versuche durchgeführt werden könnten. Das ist de facto nicht möglich. Aber selbst wenn man davon absieht, ist die Definition nicht kompatibel mit der mathematischen Definition eines Grenzwertes: Das *Cauchy-Kriterium*

$$\lim_{n \rightarrow \infty} x_n = x \stackrel{\text{Def.}}{\hat{=}} \forall \varepsilon > 0 \exists N_0 > 0 : |x_n - x| < \varepsilon \forall n > N_0$$

kann bei einer Zufallsfolge nicht erfüllt sein! Wir müssen uns also zwei Fragen stellen:

Frage I: Was verstehen wir unter „Wahrscheinlichkeit“?

Frage II: Wie ermittelt man Wahrscheinlichkeiten?

Im Folgenden diskutieren wir zunächst die Frage II, wobei wir uns auf sogenannte klassische Wahrscheinlichkeiten konzentrieren. Im späteren Abschnitt 2.2 werden wir uns mit der Frage I beschäftigen.

2.1.2 Ermittlung von klassischen Wahrscheinlichkeiten

Es gibt keine rigorose Methode, ohne a-priori Wissen Wahrscheinlichkeiten zu bestimmen. Man muss stets Annahmen machen. Am einfachsten ist die sogenannte „klassische Annahme“: Die Wahrscheinlichkeit ist *gleichverteilt* auf sogenannte, vorher festzulegende „Elementarereignisse“.

Zur Erklärung des Begriffs Elementarereignis betrachten wir ein Experiment, dessen Ausgang nicht mit Sicherheit vorhergesagt werden kann. Obwohl der Ausgang des Experiments damit im Voraus nicht bekannt ist, wollen wir annehmen, dass wir alle *möglichen* Ausgänge des Experiments kennen. Die Menge aller möglichen Ausgänge eines Experiments wird als **Stichprobenraum** oder auch **Ereignisraum** des Experiments bezeichnet, und die einzelnen Elemente des Stichprobenraums werden **Elementarereignisse** genannt. Die Wahrscheinlichkeit des Eintretens eines beliebigen, nicht notwendigen elementaren Ereignisses E ergibt sich dann als

$$\mathcal{P}(E) = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} \quad (2.1)$$

Dazu betrachten wir drei Beispiele: 2.1.2.1 „Münzwurf“, 2.1.2.2 „Zwei Kugeln“ und 2.1.2.3 „ N Münzwürfe“.

Die klassische Annahme ist die einfachstmögliche Vermutung. Sie muss eventuell auf der Basis bisheriger Erfahrungen oder physikalischer Kenntnisse *modifiziert* werden. Ein Beispiel dafür ist 2.1.2.4 „Verbogene Münze“.

2.1.2.1 Beispiel „Münzwurf“

Beim Werfen einer Münze gibt es offensichtlich zwei Elementarereignisse: Kopf (K) oder Zahl (Z). Nach der klassischen Annahme sind dann die Wahrscheinlichkeiten für jedes dieser Ereignisse gegeben durch $\mathcal{P}(K) = \mathcal{P}(Z) = \frac{1}{2}$.

2.1.2.2 Beispiel „Zwei Kugeln“

Wenn wir zwei Kugeln aus einer Schüssel mit zwei schwarzen und zwei weißen Kugeln entnehmen, haben wir die Situation von Abbildung 2.1.

In diesem Fall gibt es 12 Elementarereignisse: Die Kugeln $i = 1, \dots, 4$ werden in der Reihenfolge (i, j) entnommen mit $(i, j) = (1, 2), (1, 3), \dots$. Die Wahrscheinlichkeit für jedes dieser Ereignisse ist demnach $\mathcal{P}((i, j)) = \frac{1}{12}$. Wir stellen nun die Frage: Mit welcher Wahrscheinlichkeit sind beide gezogenen Kugeln schwarz? Zwei Elementarereignisse führen hier zum Ziel: $(i, j) = (1, 2)$ und $(i, j) = (2, 1)$. Damit folgt mit Gleichung (2.1) die Wahrscheinlichkeit $\mathcal{P}((\bullet, \bullet)) = 2/12 = 1/6$. Fragen wir umgekehrt nach der Wahrscheinlichkeit, eine weiße und eine schwarze Kugel zu ziehen (in beliebiger Reihenfolge), dann gibt es 8 „günstige“ Fälle: $(i, j) = (1, 3), (3, 1), (1, 4), (4, 1), (2, 3), (3, 2), (2, 4), (4, 2)$. Dieser Fall tritt also mit der Wahrscheinlichkeit $\mathcal{P}((\bullet, \circ)) + \mathcal{P}((\circ, \bullet)) = 8/12 = 2/3$ ein.

2.1.2.3 Beispiel „ N Münzwürfe“

Wir fragen: Mit welcher Wahrscheinlichkeit $\mathcal{P}(n)$ erhält man n -mal Kopf?

- Anzahl der möglichen Fälle: Hier ist jedes Elementarereignis eine Folge von N Würfeln K oder Z, also z.B. (K,Z,K,K,...,Z,K). Da es für jeden einzelnen Wurf $M = 2$ verschiedene Ergebnisse gibt und der Wurf $m = N$ mal ausgeführt wird, gibt es $M^m = 2^N$ verschiedene solcher Folgen (sogenannte *Variationen mit Wiederholung*, vergl. Abschnitt 2.8.3, Gl. (2.104)).
- Anzahl der günstigen Fälle: Aus N verschiedenen Würfeln lassen sich ohne Rücksicht auf Anordnung n verschiedene Würfe des Wertes K (Kopf) herausgreifen auf $\binom{N}{k}$ verschiedene Weisen (sogenannte *Kombinationen*, vergl. Abschnitt 2.8.3, Gl. (2.101) mit $n = k$).

Beweis: Es gibt $N(N-1) \cdots (N-k+1)$ Möglichkeiten, k mal Kopf auf N Würfe zu verteilen. Da die Reihenfolge der Kopf-Würfe egal ist, sind je $k!$ Möglichkeiten äquivalent.

Damit folgt mit Gl. (2.1)

$$\mathcal{P}(k) = \left(\frac{1}{2}\right)^N \binom{N}{k} \quad (2.2)$$

wobei der Ausdruck

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \quad (2.3)$$

Binomialkoeffizient genannt wird.

2.1.2.4 Beispiel „Verbogene Münze“

Wir betrachten N Würfe einer Münze, die aber *gezinkt* ist. Nehmen wir an, wir vermuten aus vergangenen Beobachtungen, dass die Münze in einzelnen Würfeln mit einer Wahrscheinlichkeit p auf Kopf fallen wird und mit $(1-p)$ auf Zahl. Die Wahrscheinlichkeit, n mal Kopf zu werfen, ist dann die **Binomialverteilung**

$$\mathcal{P}(n) = \underbrace{p^n}_{\textcircled{1}} (1-p)^{N-n} \underbrace{\binom{N}{n}}_{\textcircled{2}} \quad (2.4)$$

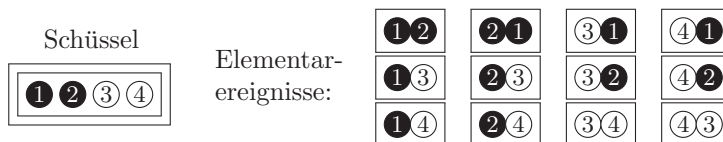


Abb. 2.1: Elementarereignisse bei der Entnahme zweier Kugeln aus einer Schüssel, die je zwei schwarze und weiße, unterscheidbare Kugeln enthält

Hier ist der erste Term ① die Wahrscheinlichkeit, überhaupt n -mal Kopf und $(N - n)$ -mal Zahl zu werfen, und der Binomialkoeffizient ② entspricht der Zahl der Möglichkeiten, in N Würfeln n -mal Kopf zu erhalten, die in Beispiel 2.1.2.3 (N Münzwürfe) oben berechnet wurde (der sog. *kombinatorische Faktor*).

Die Binomialverteilung ist eine der wichtigsten Verteilungen in der Statistik. Ebenso wichtig sind approximative Ausdrücke für die Binomialverteilung im Grenzfall sehr großer N :

- Für $N \rightarrow \infty$ und $p \rightarrow 0$ bei festgehaltenem Produkt $Np = \lambda$ geht die Binomialverteilung über in die **Poisson-Verteilung**

$$\mathcal{P}(n) \rightarrow e^{-\lambda} \frac{\lambda^n}{n!} \quad (2.5)$$

$$\begin{aligned} \text{Beweis: } \mathcal{P}(n) &= \left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n} \binom{N}{n} \\ &= \underbrace{\lambda^n \left(1 - \frac{\lambda}{N}\right)^N}_{\rightarrow e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{N}\right)^{-n}}_{\rightarrow 1} \underbrace{N^{-n} N(N-1) \cdots (N-n+1)}_{\rightarrow 1} \frac{1}{n!} \end{aligned}$$

Anwendung: Die Poisson-Verteilung $e^{-\lambda} \frac{\lambda^n}{n!}$ modelliert zum Beispiel die Anzahl n von Ereignissen, die bei konstanter mittlerer Rate λ unabhängig voneinander in einem festen Zeitintervall oder räumlichen Gebiet eintreten (Wahrscheinlichkeit für das Eintreten seltener Ereignisse).

- Falls p weder sehr nahe an $p = 0$ noch an $p = 1$ liegt, geht die Binomialverteilung für große N über in die **Gauß-Verteilung**

$$\mathcal{P}(n) \approx \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(n-\mu)^2}{2\sigma^2}} \quad (2.6)$$

mit $\mu = Np$, $\sigma^2 = Np(1-p)$. Der Beweis dieses Satzes und Anwendungen finden sich in Kapitel 2.7 (zentraler Grenzwertsatz).

2.1.3 Formale Definition des Begriffs Wahrscheinlichkeit

Wir wenden uns nun der Frage zu: Was bedeutet der Begriff „Wahrscheinlichkeit“ bzw. wie kann man Wahrscheinlichkeit definieren?

Eine pragmatische Antwort auf diese Frage wäre: Wahrscheinlichkeit ist ein *Schätzwert* für relative Häufigkeiten, also gewissermaßen eine *Prognose* für Messergebnisse. Dies liefert eine plausible Interpretation des Begriffes der Wahrscheinlichkeit, aber immer noch keine befriedigende Definition.

Alternativ kann man sich der Diskussion entziehen, indem man einfach eine Größe mit Namen „Wahrscheinlichkeit“ und bestimmten Eigenschaften *postuliert*. Dieser axiomatische Zugang soll im Folgenden dargestellt werden. Das Ziel ist die Definition einer mathematischen Struktur, die unsere intuitiven Begriffe von „Messergebnis“ und „Wahrscheinlichkeit“ präzisiert. Dabei müssen wir auch berücksichtigen, dass die physikalische Welt kontinuierlich ist, Punkte in einem Kontinuum in Messungen jedoch nicht beliebig genau eingegrenzt werden können. Wahrscheinlichkeiten können daher nur für *Intervalle* definiert sein und eine passende mathematische Struktur muss auf *Teilmengen* aufbauen (Kolmogoroff 1903 – 1987).

2.2 Wahrscheinlichkeit abstrakt mathematisch

2.2.1 Der Wahrscheinlichkeitsraum

Der Begriff des Wahrscheinlichkeitsraums wurde in den 1930er Jahren durch den russischen Mathematiker Andrei Nikolajewitsch Kolmogoroff (1903 – 1987) eingeführt, der damit einen axiomatisch basierte Wahrscheinlichkeitsbegriff begründete. Eine Grundannahme ist dabei, dass verschiedene Messergebnisse als isolierte „Punkte“ betrachtet werden können, die sich gegenseitig ausschließen: Das Ergebnis eines Münzwurfs ist immer K oder Z, niemals beides. Das Ergebnis einer einzelnen Geschwindigkeitsmessung ist eine einzelne Zahl. Mit dieser und weiteren Annahmen lässt sich eine Wahrscheinlichkeitsrechnung als mathematisches „Maß“ auf den Teilmengen der Menge aller möglichen Messergebnisse aufbauen. Die Teilmenge enthalten als Elemente also mögliche Messergebnisse. Den Teilmengen (und nicht den Messergebnissen selbst) werden als „Wahrscheinlichkeiten“ Zahlen zwischen Null und Eins zugeordnet. Die Kolmogoroff-Axiome besagen, grob gesagt:

- (i) Wahrscheinlichkeiten dürfen nicht negativ sein.
- (ii) Der Menge aller Messergebnisse ist die Wahrscheinlichkeit Eins zugeordnet.
- (iii) Wahrscheinlichkeiten von disjunkten Teilmengen – d.h. Teilmengen, die keine Elemente gemeinsam haben – addieren sich auf.

Daraus folgt unter anderem, dass Wahrscheinlichkeiten Zahlen zwischen Null und Eins sein müssen. Wir geben nun eine etwas präzisere formale Definition des sich daraus ergebenden mathematischen Objektes.

Formale Definition: Ein **Wahrscheinlichkeitsraum** $(\Omega, \mathcal{F}, \mu)$ ist ein *Messraum* mit Ergebnismenge Ω und messbaren Mengen (Ereignisalgebra) \mathcal{F} , auf dem zusätzlich ein Wahrscheinlichkeitsmaß μ gegeben ist. Im Einzelnen:

- Ω ist eine beliebige nichtleere Menge, genannt die **Ergebnismenge**. Ihre Elemente ω heißen **Ergebnisse**.
- \mathcal{F} ist eine Menge bestehend aus Teilmengen von Ω , die Ω enthält und *abgeschlossen* gegenüber der Bildung von Komplementen und abzählbaren Vereinigungen ist. Die Elemente von \mathcal{F} heißen **Ereignisse**. Andere Namen für \mathcal{F} sind: **Ereignissystem**, **Ereignisalgebra**, σ -Algebra über der Grundmenge Ω , σ -Mengenalgebra, abgeschlossenes Mengensystem, Sigmakörper, Borelscher Mengenkörper.
- $\mu : \mathcal{F} \rightarrow [0, 1]$ ist ein **Wahrscheinlichkeitsmaß**, das heißt eine Mengenfunktion, die den Ereignissen Zahlen zuordnet, derart dass $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$ (zweites Kolmogoroff-Axiom) und $\mu(A_1 \cup A_2 \cup \dots) = \mu(A_1) + \mu(A_2) + \dots$ gilt für paarweise disjunkte (d. h. sich gegenseitig ausschließende) Ereignisse A_1, A_2, \dots (drittes Kolmogoroff-Axiom).

Als *Beispiele* betrachten wir in den folgenden Unterkapiteln einen diskrete und einen kontinuierlichen Wahrscheinlichkeitsraum.

2.2.1.1 Beispiel eines diskreten Wahrscheinlichkeitsraums: „Verbogene Münze“

Die in Abschnitt 2.1.2.4 definierte Situation „Verbogene Münze“ ist charakterisiert durch den folgenden diskreten Wahrscheinlichkeitsraum:

- Die Ergebnismenge ist $\Omega = \{\text{Kopf}, \text{Zahl}\}$
- Die Ereignisalgebra \mathcal{F} besteht aus den Ereignissen „Kopf oder Zahl“, „weder Kopf noch Zahl“, „Zahl“ und „Kopf“, formal

$$\mathcal{F} = \{\{K, Z\}, \{\}, \{Z\}, \{K\}\}$$

- Damit lässt sich das diskrete Wahrscheinlichkeitsmaß erschöpfend angeben:
 $\mu(\{K, Z\}) = 1, \mu(\{\}) = 0, \mu(\{Z\}) = p, \mu(\{K\}) = 1 - p$

2.2.1.2 Beispiel eines kontinuierlichen Wahrscheinlichkeitsraums: Maxwell-Boltzmann-Verteilung

Die Maxwell-Boltzmann-Verteilung beschreibt die statistische Verteilung des Betrags der Teilchen-Impulse in einem idealen Gas. Sie spielt in der Thermodynamik, speziell der kinetischen Gastheorie, eine wichtige Rolle. In diesem Beispiel betrachten wir ein klassisches Teilchen in einem eindimensionalen Kasten der Länge L unter der Annahme, dass Ort x und Impuls p voneinander unabhängig sind.

Diese Situation ist charakterisiert durch den folgenden kontinuierlichen Wahrscheinlichkeitsraum:

- Die Ergebnismenge ist der klassische Phasenraum aus Orts- und Impulskoordinaten $\Omega = \{\Gamma\} = \{(x, p)\}$
- Die Ereignisalgebra \mathcal{F} ist die Menge aller (zweidimensionalen) Gebiete in Ω
- Ohne Beweis: Das kontinuierliche Wahrscheinlichkeitsmaß ergibt sich als Produkt von $dx dp$ und der Wahrscheinlichkeitsdichte, in diesem Fall der Maxwell-Boltzmann-Verteilung. Der Begriff der Wahrscheinlichkeitsdichte wird später noch genauer definiert werden.

$$d\mu(x, p) = dx dp e^{-p^2/2k_B T m} \underbrace{\frac{1}{L} \sqrt{\frac{1}{2\pi k_B T m}}}_{\text{Normierung}} \quad (2.7)$$

2.2.2 Bedingte Wahrscheinlichkeit und statistische Unabhängigkeit

Wir sprechen von der **unbedingten Wahrscheinlichkeit** $\mathcal{P}(A)$ eines Ereignisses A , die unabhängig vom Eintreten anderer Ereignisse allein durch das dem Ereignis A zugeordnete Wahrscheinlichkeitsmaß gegeben ist:

$$\mathcal{P}(A) = \mu(A) \quad (2.8)$$

Im Gegensatz dazu nennen wir die Wahrscheinlichkeit für das Eintreten von Ereignis A unter der Bedingung, dass Ereignis B bereits eingetreten ist, die **bedingte Wahrscheinlichkeit**, Schreibweise: $\mathcal{P}(A|B)$, Sprechweise „Wahrscheinlichkeit für A unter B “. Wir definieren

$$\mathcal{P}(A|B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)} \quad (2.9)$$

Begründung: Damit das Ereignis A eintritt, wenn das Ereignis B bereits eingetreten ist, muss das eingetretene Elementarereignis sowohl in A als auch in B , also in $A \cap B$ liegen. Da wir aber wissen, dass das eingetretene Elementarereignis in B liegen muss, brauchen wir nur den (reduzierten) Stichprobenraum B zu betrachten. Daher ist die Wahrscheinlichkeit des Ereignisses A , wenn B bereits eingetreten ist, gleich der Wahrscheinlichkeit von $A \cap B$ relativ zur Wahrscheinlichkeit von B .

Die Definition der bedingten Wahrscheinlichkeit in Gleichung (2.9) verträgt sich mit der Interpretation der Wahrscheinlichkeit als einer relativen Häufigkeit, wenn das Experiment nur genügend oft wiederholt wird. Um das zu zeigen, stellen wir uns eine große Zahl n von Wiederholungen eines Experiments vor. Da $\mathcal{P}(B)$ den Anteil der Experimente angibt, in denen B eintritt, wird B näherungsweise $n\mathcal{P}(B)$ -mal eintreten. Entsprechend werden in näherungsweise $n\mathcal{P}(A \cap B)$ dieser Experimente sowohl A als auch B eintreten. Betrachten wir die näherungsweise $n\mathcal{P}(B)$ Experimente, bei denen B eintritt, so tritt bei ihnen in näherungsweise $n\mathcal{P}(A \cap B)$ Fällen auch A ein. Für die Experimente, bei denen das Ereignis B eintritt, ergibt sich also ein Anteil, bei denen gleichzeitig das Ereignis A eintritt, nämlich näherungsweise

$$\frac{n\mathcal{P}(A \cap B)}{n\mathcal{P}(B)} = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}$$

Diese Näherung wird immer besser, je größer n wird. Daher gibt Gleichung (2.9) die passende Definition für die bedingte Wahrscheinlichkeit von A für den Fall, dass B bereits eingetreten ist.

Wir nennen zwei Ereignisse A und B **statistisch unabhängig** voneinander, falls gilt

$$\mathcal{P}(A|B) = \mathcal{P}(A) \quad (2.10)$$

oder äquivalent

$$\mathcal{P}(A \cap B) = \mathcal{P}(A) \cdot \mathcal{P}(B) \quad (2.11)$$

2.2.2.1 Beispiel für eine bedingte Wahrscheinlichkeit

Fragestellung: Gegeben ist eine Schüssel mit zwei weißen und zwei schwarzen Kugeln (Abb. 2.2 S. 20). Gesucht ist die Wahrscheinlichkeit, als zweites eine weiße Kugel zu ziehen, wenn die erste schon weiß war. Der Wahrscheinlichkeitsraum dieser Situation besteht aus den drei Elementen (a) der Ergebnismenge $\Omega = \{(\circ\circ), (\circ\bullet), (\bullet\circ), (\bullet\bullet)\}$, (b) den Wahrscheinlichkeiten $\mu(\{\circ\circ\}) = \mu(\{\bullet\bullet\}) = \frac{1}{6}$ und $\mu(\{\circ\bullet\}) = \mu(\{\bullet\circ\}) = \frac{1}{3}$, und (c) den beiden Ereignissen „1. Kugel weiß“ $E_1 = \{(\circ\circ), (\circ\bullet)\}$ und „2. Kugel weiß“ $E_2 = \{(\circ\circ), (\bullet\circ)\}$.

Schüssel

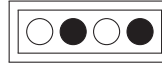


Abb. 2.2: Illustration zur bedingten Wahrscheinlichkeit: Angenommen, dieser Schüssel wird eine Kugel entnommen, die zufällig weiß ist. Mit welcher Wahrscheinlichkeit ist eine zweite zufällig entnommene Kugel dann auch weiß?

Mit Hilfe der Definition der bedingten Wahrscheinlichkeit erhalten wir

$$\mathcal{P}(E_2|E_1) = \frac{\mu(E_1 \cap E_2)}{\mu(E_1)} = \frac{\mu(\{(\circ \circ)\})}{\mu(\{(\circ \circ)\}, \{(\circ \bullet)\})} = \frac{1/6}{1/6 + 1/3} = \frac{1}{3}$$

2.2.3 Zufallsvariablen

Eine **Zufallsvariable** oder **Zufallsgröße** (englisch „*random variable*“) ist eine Größe, deren Wert vom Zufall abhängt.

Formal ist sie eine messbare Funktion von einem Wahrscheinlichkeitsraum in einen Messraum und hat Funktionswerte $\underline{z}(\omega)$, die von einer den Zufall repräsentierenden Größe ω abhängen. In diesem Buch wird sie im Folgenden durch Unterstreichung (z.B. $\underline{z}_1, \underline{z}_2, \dots$) gekennzeichnet. Ohne Unterstreichung (z_1, z_2, \dots) handelt es sich um eine *Realisierung*, d.h. einen einzelnen Wert, den die Zufallsvariable bei der Durchführung eines Zufallsexperiments annimmt. Eine Zufallsvariable ist also (formal) eine Zuordnungsvorschrift, die jedem möglichen Ergebnis eines Zufallsexperiments eine Größe zuordnet.

Ist diese Größe eine Zahl, so spricht man von einer **Zufallszahl**. Eine Zufallszahl ist (im reellen Fall, auf den wir uns hier beschränken) eine Funktion, die erstens jedem Ergebnis ω aus einer Ergebnismenge Ω eine reelle Zahl $\underline{z}(\omega)$ zuordnet (formal: eine Funktion $\underline{z} : \Omega \rightarrow \mathbb{R}$) und zweitens die folgende „Messbarkeitsbedingung“ erfüllt: Für jede beliebige Zahl z muss die Menge aller Ergebnisse ω , deren zugehörige Zufallszahl-Realisierung unterhalb von z liegt, einem Ereignis entsprechen (formal: $\forall z \in \mathbb{R} : \{\omega | \underline{z}(\omega) \leq z\} \in \mathcal{F}$).

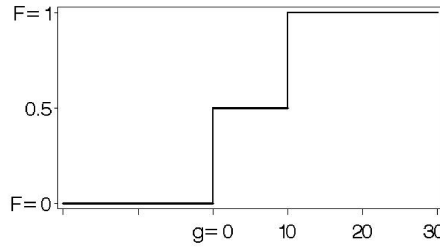
2.2.4 Die kumulative Verteilungsfunktion

Die kumulierte oder **kumulative Verteilungsfunktion** (*cumulative distribution function*) ist ein zentrales Konzept bei der Untersuchung von Wahrscheinlichkeitsverteilungen auf den reellen Zahlen. Jeder Wahrscheinlichkeitsverteilung und jeder reellwertigen Zufallsvariable kann eine Verteilungsfunktion zugeordnet werden. Anschaulich entspricht dabei der Wert der Verteilungsfunktion an der Stelle z der Wahrscheinlichkeit, dass die zugehörige Zufallsvariable \underline{z} einen Wert kleiner oder gleich z annimmt.

Ist beispielsweise die Verteilung der Schuhgrößen in Europa gegeben, so entspricht der Wert der entsprechenden Verteilungsfunktion bei 45 der Wahrscheinlichkeit, dass ein beliebiger Europäer die Schuhgröße 45 oder kleiner besitzt.

Wir definieren formal die **kumulative Verteilungsfunktion** als

$$F_{\underline{z}}(z) := \mathcal{P}(\underline{z}(\omega) \leq z) = \mu(\{\omega \in \Omega : \underline{z}(\omega) \leq z\}) \quad (2.12)$$

Abb. 2.3: Kumulative Zweipunktverteilung „10-Euro-Münze“ (SAS-Programm `S.1.sas`)

Aus der Definition lesen wir leicht die folgenden Eigenschaften der kumulativen Verteilungsfunktion ab:

$$F_{\underline{z}}(-\infty) = 0 \quad (2.13)$$

$$F_{\underline{z}}(\infty) = 1 \quad (2.14)$$

$$F_{\underline{z}}(z) = \text{nicht abnehmend} \quad (2.15)$$

$$1 - F_{\underline{z}}(z) = \mathcal{P}(\underline{z}(\omega) > z) \quad (2.16)$$

$$\mathcal{P}(a < \underline{z}(\omega) \leq b) = F_{\underline{z}}(b) - F_{\underline{z}}(a) \quad (2.17)$$

Beachte: Bei kontinuierlichen Verteilungen wie der Normalverteilung ist es egal, ob man in Gleichung (2.17) $<$ oder \leq schreibt. (2.18)

2.2.5 Beispiele für kumulative Verteilungen

2.2.5.1 Kumulative Zweipunktverteilung „10-Euro-Münze“

Wir erfinden eine 10-Euro-Münze. Der „Gewinn“ ist eine diskrete zufällige Variable \underline{G} mit Realisierungen $\underline{G}(\text{Zahl}) = 10$ Euro und $\underline{G}(\text{Kopf}) = 0$. Die kumulative Verteilungsfunktion kann ausgedrückt werden als

$$F_{\underline{G}}(g) = \mathcal{P}(\underline{G} \leq g) = \sum_{a=0,10} \frac{1}{2} \Theta(g - a) \quad (2.19)$$

Die Sprungfunktion $\Theta(x)$ ist im Anhang A definiert, Gleichung (A.15): $\Theta(x)$ springt an der Stelle $x = 0$ von Null auf Eins. An den Stellen $(g - a) = 0$, also bei $g = 0$ (Kopf) und bei $g = 10$ (Zahl der 10-Euro-Münze), bewirkt der Term $\frac{1}{2} \Theta(g - a)$ einen Sprung von jeweils 0.5 Einheiten. Diesen Sprung sieht man in Abb. 2.3 S. 21.

2.2.5.2 Kumulative Maxwell-Boltzmann-Verteilung

Im Beispiel des kontinuierlichen Wahrscheinlichkeitsraums 2.2.1.2 betrachten wir als Zufallsvariable die kinetische Energie $\underline{z} = \underline{E} = p^2/(2m)$. Die dazugehörige kumu-

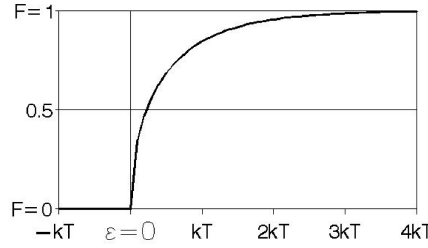


Abb. 2.4: Kumulative Maxwell-Boltzmann-Verteilung (SAS-Programm S.2.sas)

lative Verteilungsfunktion ergibt sich aus der Maxwell-Boltzmann-Verteilung (2.7)

$$\begin{aligned}
 F_{\underline{E}}(\varepsilon) &= \int_{\varepsilon < p^2/2m} d\mu(x, p) = \int_0^L dx \int_{-\sqrt{2m\varepsilon}}^{\sqrt{2m\varepsilon}} dp \frac{1}{L} e^{-p^2/2k_B T m} \frac{1}{\sqrt{2\pi k_B T m}} \\
 &= \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{\varepsilon/k_B T}} d\hat{p} e^{-\hat{p}^2} = \frac{2}{\sqrt{\pi}} \text{Erf}\left(\sqrt{\frac{\varepsilon}{k_B T}}\right) \quad (2.20)
 \end{aligned}$$

Sie ist in Abb. 2.4 S. 22 dargestellt.

2.2.5.3 Kumulative Binomialverteilung

Die Binomialverteilung $\mathcal{P}(n) = p^n(1-p)^{N-n} \binom{N}{n}$ haben wir bereits in Gleichung (2.4) kennengelernt. In Analogie zu Gleichung (2.19) schreiben wir die kumulative Binomialverteilung als:

$$F_{\underline{z}}(z) = \mathcal{P}(\underline{z} \leq z) = \sum_n \mathcal{P}(n) \Theta(z - n) = \sum_n p^n (1-p)^{N-n} \binom{N}{n} \Theta(z - n) \quad (2.21)$$

Abb. 2.5 S. 22 zeigt die kumulative Binomialverteilung für $N = 25$ und $p = 0.9$, und die zugehörige Gaußsche Approximation mit $\mu = Np$ und $\sigma^2 = Np(1-p)$, also konkret $\mu = 22.5$ und $\sigma^2 = 2.25$.

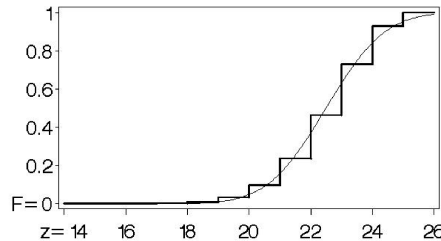


Abb. 2.5: Kumulative Binomialverteilung (SAS-Programm S.3.sas)

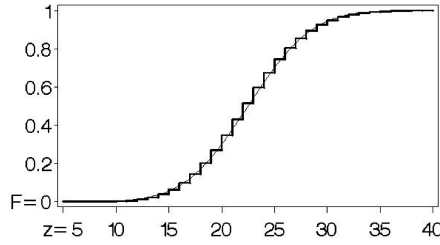


Abb. 2.6: Kumulative Poisson-Verteilung (SAS-Programm S.4.sas)

2.2.5.4 Kumulative Poisson-Verteilung

Auch die Poisson-Verteilung $\mathcal{P}(n) \rightarrow e^{-\lambda} \frac{\lambda^n}{n!}$ haben wir bereits in Gleichung (2.5) kennengelernt. In Analogie zu den Gleichungen (2.19) und (2.21) schreiben wir die kumulative Poisson-Verteilung als:

$$F_{\underline{z}}(z) = \mathcal{P}(\underline{z} \leq z) = \sum_n \mathcal{P}(n) \Theta(z - n) = \sum_n e^{-\lambda} \frac{\lambda^n}{n!} \Theta(z - n) \quad (2.22)$$

Abb. 2.6 S. 23 zeigt ein Beispiel für die kumulative Poisson-Verteilung, hier mit $\lambda = Np = 22.5$, und approximiert durch eine Gauß-Verteilung mit $\mu = \sigma^2 = Np$, konkret $\mu = \sigma^2 = 22.5$.

2.2.6 Die Dichtefunktion

Die **(Wahrscheinlichkeits-)Dichtefunktion** (*density function*) $f_{\underline{z}}(z)$ einer Zufallsvariablen ist die erste Ableitung der kumulativen Verteilung $F_{\underline{z}}(z)$. Sie kann einen visuellen Eindruck der Verteilung vermitteln: Wie der Name \underline{z} bereits andeutet, zeigt diese Funktion, in welchen Teilen sich die Werte der Zufallsvariablen am dichtesten scharen. Wir definieren sie durch die Gleichung

$$f_{\underline{z}}(z) dz := \mathcal{P}(z \leq \underline{z}(\omega) \leq z + dz) = \mu(\{\omega \in \Omega : \underline{z}(\omega) \in [z, z + dz]\}) \quad (2.23)$$

Daraus folgt unmittelbar

$$f_{\underline{z}}(z) = \frac{d}{dz} F_{\underline{z}}(z), \quad \int_{-\infty}^z d\xi f_{\underline{z}}(\xi) = F_{\underline{z}}(z), \quad \int_{-\infty}^{\infty} dz f_{\underline{z}}(z) = 1 \quad (2.24)$$

Die Wahrscheinlichkeitsdichte kann nur positive Werte annehmen, die aber durchaus größer als 1 sein können. Aus Gl. (2.24) (Mitte) kann man ablesen, dass die Wahrscheinlichkeitsdichte als Integrand der kumulativen Verteilungsfunktion aufgefasst werden kann, also als die Funktion „im Integral“. Die Fläche unter der Dichte hat den Inhalt 1. Das Tripel $(\mathbb{R}, B, f_{\underline{z}}(z) dz)$ ist wieder ein Wahrscheinlichkeitsraum, wobei die Ereignisalgebra B aus den Intervallen auf \mathbb{R} besteht.

In Abschnitt 2.5.2 werden wir einen eleganten Weg kennenlernen, die Wahrscheinlichkeitsdichte als Erwartungswert der Deltafunktion darzustellen.

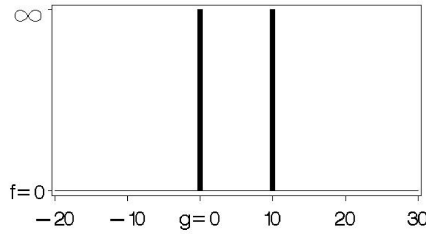


Abb. 2.7: Dichte der Zweipunktverteilung „10-Euro-Münze“ (SAS-Programm `S.5.sas`)
Die beiden „unendlich hohen“ Peaks stehen für Delta-Funktionen (siehe Anhang A.3).

2.2.6.1 Die Dichte der Zweipunktverteilung „10-Euro-Münze“

Wir zeichnen die erste Ableitung der kumulativen Zweipunktverteilung „10-Euro-Münze“, Gleichung (2.19):

$$f_{\underline{g}}(g) = \frac{d}{dg} F_{\underline{g}}(g) = \frac{d}{dg} \sum_{a=0,10} \frac{1}{2} \Theta(g-a) = \sum_{a=0,10} \frac{1}{2} \delta(g-a) \quad (2.25)$$

und erhalten die Abbildung 2.7.

2.2.6.2 Die (Energie-)dichte der Maxwell-Boltzmann-Verteilung

Wir zeichnen die erste Ableitung der kumulativen Maxwell-Boltzmann-Verteilung, Gleichung (2.20):

$$f_{\underline{\varepsilon}}(\varepsilon) = \frac{d}{d\varepsilon} F_{\underline{\varepsilon}}(\varepsilon) = \frac{2}{\sqrt{\pi}} \frac{d}{d\varepsilon} \int_0^{\sqrt{\varepsilon/k_B T}} d\hat{p} e^{-\hat{p}^2} = \frac{1}{\sqrt{\pi \varepsilon k_B T}} e^{-\varepsilon/k_B T} \quad (2.26)$$

und erhalten die Abbildung 2.8. Das Ergebnis für die Dichteverteilung entspricht der Gammaverteilung 3.10(8)b auf Seite 84 (mit $\alpha = 1/2, \beta = k_B T$).

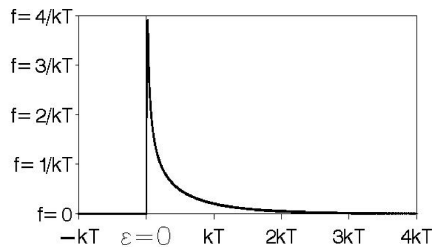


Abb. 2.8: (Energie-)dichte der Maxwell-Boltzmann-Verteilung (SAS-Programm `S.6.sas`)

2.2.7 Multidimensionale Verteilungen

Die Definition (2.23) kann auf mehrere Zufallsvariablen $\underline{z}_1 \cdots \underline{z}_n$ erweitert werden:

$$f_{\underline{z}_1 \cdots \underline{z}_n}(z_1 \cdots z_n) \, dz_1 \cdots dz_n = \mu(\{\omega \in \Omega : \underline{z}_i(\omega) \in [z_i, z_i + dz_i]\}) \quad (2.27)$$

Analog zu (2.11) heißen Zufallsvariablen **statistisch unabhängig**, wenn die Verteilung *faktoriisiert*:

$$f_{\underline{z}_1 \cdots \underline{z}_n}(z_1 \cdots z_n) = f_{\underline{z}_1}(z_1) \cdots f_{\underline{z}_n}(z_n) \quad (2.28)$$

2.3 Erwartungswert, Varianz, Korrelation

Wir führen nun einige in der Statistik wichtigen Begriffe ein.

Der **Erwartungswert** einer Zufallsvariablen \underline{z} ist definiert als

$$\mathcal{E}(\underline{z}) := \int_{\Omega} \underline{z}(\omega) \, d\mu(\omega) = \int_{-\infty}^{\infty} dz \, z \, f_{\underline{z}}(z) \quad (\text{falls Integral existiert}) \quad (2.29)$$

Der **Erwartungswert einer beliebigen Funktion** $g(\underline{z})$ ist

$$\mathcal{E}(g(\underline{z})) = \int_{-\infty}^{\infty} dz \, g(z) \, f_{\underline{z}}(z) \quad (\text{falls Integral existiert}) \quad (2.30)$$

Die **Varianz** (= **Streuung**) einer Zufallsvariablen \underline{z} ist definiert als

$$\mathcal{D}^2(\underline{z}) \equiv \mathcal{D}(\underline{z}, \underline{z}) := \mathcal{E}\left((\underline{z} - \mathcal{E}(\underline{z}))^2\right) \stackrel{\vee}{=} \mathcal{E}(\underline{z}^2) - \mathcal{E}(\underline{z})^2 \quad (2.31)$$

Die Wurzel davon, $\mathcal{D}(\underline{z}) := \sqrt{\mathcal{D}^2(\underline{z})}$, wird **Standardabweichung** genannt.

Die **Kovarianz** zweier Zufallsvariablen $\underline{z}_1, \underline{z}_2$ ist definiert als

$$\mathcal{D}(\underline{z}_1, \underline{z}_2) := \mathcal{E}\left((\underline{z}_1 - \mathcal{E}(\underline{z}_1))(\underline{z}_2 - \mathcal{E}(\underline{z}_2))\right) \stackrel{\vee}{=} \mathcal{E}(\underline{z}_1 \underline{z}_2) - \mathcal{E}(\underline{z}_1) \mathcal{E}(\underline{z}_2) \quad (2.32)$$

$$\text{mit } \mathcal{E}(\underline{z}_1 \underline{z}_2) = \int_{-\infty}^{\infty} dz_1 \, dz_2 \, z_1 \, z_2 \, f_{\underline{z}_1 \underline{z}_2}(z_1, z_2) \quad (2.33)$$

Die Kovarianz $\mathcal{D}(\underline{z}_1, \underline{z}_2)$ wird häufig als Indikator dafür benutzt, ob zwei Variablen statistisch abhängig sind. Für statistisch unabhängige Zufallsvariablen $\underline{z}_1, \underline{z}_2$ gilt $\mathcal{D}(\underline{z}_1, \underline{z}_2) = 0$. Die Umkehrung gilt jedoch nicht unbedingt, d.h. zwei unkorrelierte Zufallsvariablen können statistisch abhängig sein.

Anmerkung: Die Schreibweise $\mathcal{D}(\underline{x}, \underline{y})$ ist etwas ungewöhnlich, aber hat den Vorzug, einheitlich zu sein. In der Literatur werden oft zwei verschiedene Formelzeichen verwendet wie z.B. $\text{COVAR}(\underline{x}, \underline{y}) = \mathcal{D}(\underline{x}, \underline{y})$ und $\text{VAR}(\underline{z}) = \mathcal{D}(\underline{z}, \underline{z}) = \mathcal{D}^2(\underline{z})$.

Die **Varianz-Kovarianz-Matrix** zweier Zufallsvektoren ist

$$\mathcal{D}(\underline{x}, \underline{y}') := \mathcal{E}\left(\left(\underline{x} - \mathcal{E}(\underline{x})\right)\left(\underline{y} - \mathcal{E}(\underline{y})\right)'\right) = \mathcal{E}(\underline{x} \underline{y}') - \mathcal{E}(\underline{x})\mathcal{E}(\underline{y}') \quad (2.34)$$

mit den Vektoren $\underline{x} = (x_1 \ x_2 \ \dots)'$ und $\underline{y} = (y_1 \ y_2 \ \dots)'$

Der **Korrelationskoeffizient** zwischen z_1 und z_2 ist

$$\varrho_{z_1 z_2} := \mathcal{D}(z_1, z_2) / \sqrt{\mathcal{D}(z_1, z_1) \mathcal{D}(z_2, z_2)} \quad (2.35)$$

Aus der Schwarzschen Ungleichung kann man herleiten, dass

$$-1 \leq \varrho_{z_1 z_2} \leq 1 \quad (2.36)$$

Falls $\mathcal{E}(z_1) = \mathcal{E}(z_2) = 0$, ist der Korrelationskoeffizient gegeben durch

$$\varrho_{z_1 z_2} := \mathcal{E}(z_1 \cdot z_2) / \sqrt{\mathcal{E}(z_1^2) \cdot \mathcal{E}(z_2^2)} \quad (2.37)$$

2.3.1 Varianz von Funktionen von Zufallsvariablen

2.3.1.1 (Ko-)Varianz von Summen von Zufallsvariablen

Aus der Linearität der Definition (2.29) geht hervor, dass der Erwartungswert einer Summe von Zufallsvariablen gleich der Summe der Erwartungswerte dieser Zufallsvariablen ist. Das entsprechende Ergebnis gilt allerdings nicht allgemein auch für die Varianz, wie die folgenden Gleichungen demonstrieren.

$$\mathcal{D}(\underline{x} + \underline{z}, \underline{y}) = \mathcal{D}(\underline{x}, \underline{y}) + \mathcal{D}(\underline{z}, \underline{y}) \quad (2.38)$$

$$\mathcal{D}\left(\sum_{i=1}^n \underline{x}_i, \sum_{j=1}^m \underline{y}_j\right) = \sum_{i=1}^n \sum_{j=1}^m \mathcal{D}(\underline{x}_i, \underline{y}_j) \quad (2.39)$$

$$\mathcal{D}^2\left(\sum_{i=1}^n \underline{x}_i\right) = \sum_{i=1}^n \mathcal{D}^2(\underline{x}_i) + \sum_{i=1}^n \sum_{j \neq i}^n \mathcal{D}(\underline{x}_i, \underline{x}_j) \quad (2.40)$$

$$\mathcal{D}^2\left(\sum_{i=1}^n \underline{x}_i\right) = \sum_{i=1}^n \mathcal{D}^2(\underline{x}_i) \quad \text{falls } \underline{x}_i \text{ unabhängig} \quad (2.41)$$

Beweis: (2.31) bzw. (2.32) ausmultiplizieren oder [Ross 2006, Seite 111-114]

2.3.1.2 (Ko-)Varianz linearer Funktionen von Zufallsvariablen

Nützliche Ausdrücke für die Kovarianz ergeben sich, wenn man die rechte Seite der Definitionsgleichungen ausmultipliziert:

$$\mathcal{D}(a_1 z_1 + b_1, a_2 z_2 + b_2) = a_1 a_2 \mathcal{D}(z_1, z_2) \quad (2.42)$$

$$\begin{aligned} \text{Beweis: } \mathcal{D}(a_1 z_1 + b_1, a_2 z_2 + b_2) &= \mathcal{E}((a_1 z_1 + b_1)(a_2 z_2 + b_2)) - \mathcal{E}((a_1 z_1 + b_1))\mathcal{E}((a_2 z_2 + b_2)) \\ &= a_1 a_2 \mathcal{E}(z_1 z_2) + a_1 b_2 \mathcal{E}(z_1) + b_1 a_2 \mathcal{E}(z_2) + b_1 b_2 \\ &\quad - a_1 a_2 \mathcal{E}(z_1)\mathcal{E}(z_2) - a_1 b_2 \mathcal{E}(z_1) - b_1 a_2 \mathcal{E}(z_2) - b_1 b_2 \\ &= a_1 a_2 \mathcal{D}(z_1, z_2) \quad \checkmark \end{aligned}$$
