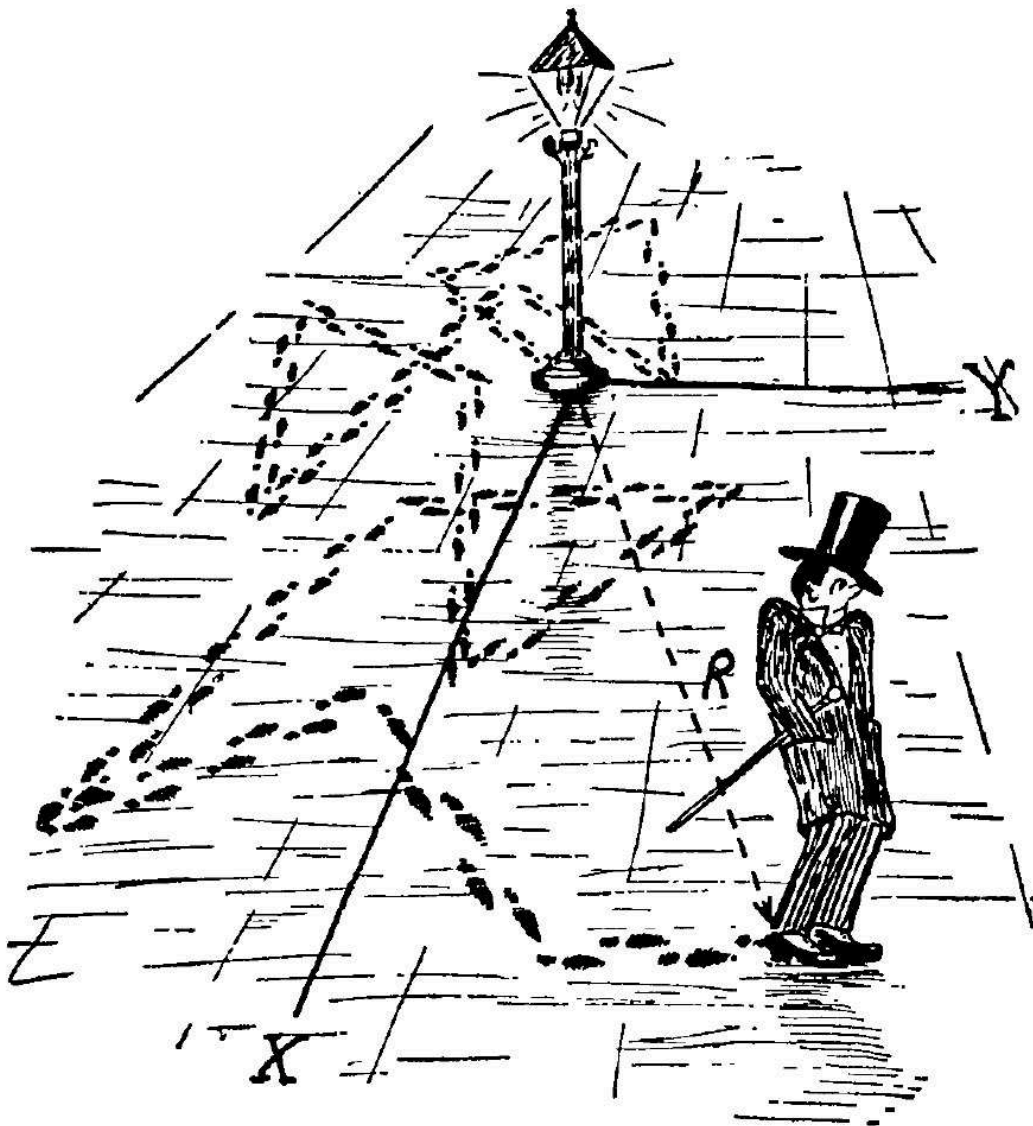


Mirko Birbaumer
Sandro Schmid
Peter Büchel

Stochastik

Vorlesungsskript HS 16



Hochschule Luzern Technik & Architektur

Inhaltsverzeichnis

| | |
|--|-----------|
| 1. Einführung | 1 |
| 1.1. Was ist Statistik? | 1 |
| 1.2. Kann ich Statistik überhaupt brauchen? | 4 |
| 1.3. Was ist der Inhalt dieses Kurses? | 5 |
| 1.4. Software | 6 |
| 1.5. Literaturverzeichnis | 7 |
| 2. Deskriptive Statistik | 8 |
| 2.1. Deskriptive Statistik eindimensionaler Daten | 8 |
| 2.1.1. Messungen der Schmelzwärme von Eis | 8 |
| 2.1.2. Darstellung von Messwerten | 9 |
| 2.1.3. Kennzahlen | 11 |
| 2.1.4. Weitere Kennzahlen | 14 |
| 2.1.5. Graphische Methoden | 20 |
| 2.2. Deskriptive Statistik zweidimensionaler Daten | 29 |
| 2.2.1. Graphische Darstellung: Streudiagramm | 30 |
| 2.2.2. Einfache lineare Regression | 32 |
| 2.2.3. Empirische Korrelation | 41 |
| 3. Modelle für Zähldaten | 45 |
| 3.1. Einführung | 45 |
| 3.2. Wahrscheinlichkeitsmodelle | 48 |
| 3.3. Der Begriff der Unabhängigkeit | 58 |
| 3.4. Wahrscheinlichkeitsbäume | 61 |
| 3.5. Bedingte Wahrscheinlichkeit | 65 |
| 3.6. Zufallsvariable | 74 |
| 3.7. Binomialverteilung | 78 |
| 3.8. Kennzahlen einer Verteilung | 88 |
| 3.9. Kumulative Verteilungsfunktion | 91 |
| 3.10. Poisson-Verteilung | 93 |
| 3.10.1. Poisson-Approximation der Binomialverteilung | 95 |
| 3.10.2. Summen von Poisson-verteilten Zufallsvariablen | 96 |
| 4. Statistik für Zähldaten | 98 |
| 4.1. Drei Grundfragestellungen der Statistik | 98 |

Inhaltsverzeichnis

| | | |
|-----------|--|------------|
| 4.2. | (Punkt-)Schätzung bei Binomialverteilung und Poisson-Verteilung . . . | 100 |
| 4.2.1. | Momentenmethode bei Binomialverteilung | 101 |
| 4.2.2. | Maximum-Likelihood bei Binomialverteilung | 103 |
| 4.2.3. | Momentenmethode bei Poisson-Verteilung | 108 |
| 4.2.4. | Maximum-Likelihood bei Poisson-Verteilung | 109 |
| 4.3. | Statistischer Test | 113 |
| 4.3.1. | Motivation | 113 |
| 4.3.2. | Formales Vorgehen | 116 |
| 4.3.3. | Zweiseitiger statistischer Test | 121 |
| 4.3.4. | Fehler 1. und 2. Art beim statistischen Test | 125 |
| 4.3.5. | Die Macht eines statistischen Tests | 129 |
| 4.3.6. | Der P -Wert | 131 |
| 4.4. | Vertrauensintervall | 135 |
| 5. | Modelle für Messdaten | 146 |
| 5.1. | Stetige Zufallsvariablen und Wahrscheinlichkeitsverteilungen | 146 |
| 5.1.1. | Wahrscheinlichkeitsdichte | 150 |
| 5.1.2. | Kennzahlen von stetigen Verteilungen | 152 |
| 5.2. | Wichtige stetige Verteilungen | 154 |
| 5.2.1. | Uniforme Verteilung | 154 |
| 5.2.2. | Exponentialverteilung | 156 |
| 5.2.3. | Normalverteilung (Gauss-Verteilung) | 160 |
| 5.3. | Funktionen einer Zufallsvariable | 164 |
| 5.3.1. | Lineare Transformationen von Zufallsvariablen | 165 |
| 5.3.2. | Nichtlineare Transformationen von Zufallsvariablen | 171 |
| 5.4. | Funktionen von mehreren Zufallsvariablen | 172 |
| 5.4.1. | Unabhängigkeit und i.i.d. Annahme | 174 |
| 5.4.2. | Kennzahlen von S_n und \bar{X}_n | 174 |
| 5.4.3. | Verteilungen von S_n und \bar{X}_n | 178 |
| 5.4.4. | Fehlerrechnung bei Messreihen | 181 |
| 6. | Statistik für Messdaten | 187 |
| 6.1. | Überprüfen der (Normal-) Verteilungsannahme | 187 |
| 6.1.1. | Q-Q-Plot | 188 |
| 6.1.2. | Normal-Plot | 191 |
| 6.2. | Parameterschätzung für stetige Wahrscheinlichkeitsverteilungen | 193 |
| 6.2.1. | Momentenmethode | 193 |
| 6.2.2. | Maximum-Likelihood-Methode | 196 |
| 6.3. | Stat. Tests und Vertrauensintervall bei normalv. Daten | 199 |
| 6.3.1. | Der z -Test (σ_X bekannt) | 200 |
| 6.3.2. | Der t -Test (σ_X unbekannt) | 205 |
| 6.3.3. | Vertrauensintervall für μ | 211 |
| 6.4. | Statistische Tests bei nicht-normalverteilten Daten | 212 |
| 6.4.1. | Der Vorzeichentest | 213 |

Inhaltsverzeichnis

| | |
|--|------------|
| 6.4.2. Der Wilcoxon-Test | 216 |
| 6.5. Statistische Tests bei zwei Stichproben | 218 |
| 6.5.1. Gepaarte Stichproben | 219 |
| 6.5.2. Ungepaarte Stichproben | 221 |
| 6.6. Statistische Signifikanz und fachliche Relevanz bei statistischen Tests | 225 |
| 7. Stochastische Prozesse | 227 |
| 7.1. Vom Random Walk zum Thermischen Rauschen | 227 |
| 7.1.1. Random Walk | 227 |
| 7.1.2. Brownsche Bewegung | 232 |
| 7.2. Charakteristische Grössen von stochastischen Signalen | 237 |
| 7.3. Wahrscheinlichkeitsverteilungsfunktion und -dichtefunktion | 239 |
| 7.4. Autokorrelationsfunktion und Kreuzkorrelationsfunktion | 249 |
| 7.4.1. Die Autokorrelationsfunktion | 249 |
| 7.4.2. Die Kreuzkorrelationsfunktion | 257 |
| 7.5. Detektion eines deterministischen Signals in einem verrauschten Signal | 258 |
| 7.6. Zeitreihen und Autokorrelationsfunktionen mit R | 262 |
| A. Ergänzungen zur Binomialverteilung und Poissonverteilung | 265 |
| A.1. Herleitung des Binomialkoeffizienten | 265 |
| A.2. Erwartungswert und Varianz der Binomialverteilung | 266 |
| A.3. Herleitung der Poisson-Verteilung | 268 |
| A.4. Erwartungswert und Varianz der Poissonverteilung | 269 |
| A.5. Poisson-Verteilung als Grenzfall der Binomialverteilung | 270 |
| A.6. Berechnung von Erwartungswert und Varianz | 271 |
| A.7. Erwartungstreue der Varianz | 272 |
| A.8. Beweis des Zentralen Grenzwertsatzes | 275 |
| B. Aus der Normalverteilung hergeleitete Verteilungen | 277 |
| B.1. Dichtefunktion der Chi-Quadrat-Verteilung | 277 |
| B.2. t -Verteilung | 281 |
| B.3. Chi-Quadrat Test | 284 |
| C. Signaltheorie und das Wiener-Khintchine Theorem | 289 |
| C.1. Herleitung der Lösung der Diffusionsgleichung aus Random Walk | 289 |
| C.2. Korrelation von Signalen und Wiener-Khintchine-Theorem | 291 |
| C.3. Energie, Leistung, Korrelation | 291 |
| C.3.1. Ähnlichkeitsmass für Signale | 292 |
| C.4. Korrelation und Faltung | 294 |
| C.4.1. Kreuzkorrelationstheorem | 294 |
| C.5. Wiener-Khintchine Theorem | 296 |
| Index | 304 |

Tabellenverzeichnis

| | |
|---|-----|
| 2.1. Messungen der latenten Schmelzwärme von Eis | 9 |
| 2.2. Weinkonsumation | 30 |
| 2.3. Buchpreis und Seitenzahl | 32 |
| 2.4. Grössenvergleich von Vätern und Söhnen | 38 |
| 2.5. Verkehrstoten in aufeinanderfolgenden Jahren | 39 |
| 3.1. Anzahl Zerfälle von Americium 241 | 46 |
| 3.2. Operationen der Mengenlehre | 51 |
| 3.3. Wahrscheinlichkeiten für einen nicht-fairen Würfel. | 55 |
| 3.4. Wahrscheinlichkeit für eine Krankheit | 70 |
| 3.5. Wahrscheinlichkeitsverteilung von gezogenen Jasskarten. | 77 |
| 3.6. Werfen einer fairen Münze | 79 |
| 4.1. Zahlen für $X \sim \text{Binomial}(100, 1/2)$ | 114 |
| 4.2. Verwerfungsbereich für $X \sim \text{Bin}(100, 1/2)$ | 116 |
| 4.3. Verwerfungsbereich für $X \sim \text{Binomial}(100, 0.1)$ | 118 |
| 4.4. Verwerfungsbereich für $X \sim \text{Bin}(100, 0.5)$ für den zweiseitigen Test . . | 123 |
| 6.1. 20 Messungen der Betondruckfestigkeit. | 189 |
| 6.2. Messung der Schmelzwärme mit der Methode A. | 213 |
| 6.3. Vorzeichen der Schmelzwärme Messung in Bezug zum Median | 214 |
| 6.4. Daten und Ränge der Schmelzwärme-Messung | 217 |
| A.1. Symbole für Erfolg und Misserfolg | 265 |
| B.1. Anzahl Zerfälle in 7.5 Sekunden | 287 |

Abbildungsverzeichnis

| | |
|--|-----|
| 1.1. Konzept der Wahrscheinlichkeitsrechnung | 3 |
| 2.1. Histogramm IQ-Test | 21 |
| 2.2. Boxplot | 24 |
| 2.3. Boxplot Schmelzwärme von Eis | 26 |
| 2.4. Emp. kumulative Verteilungsfunktion von Schmelzwärme von Eis . . | 27 |
| 2.5. Emp. kumulative Verteilungsfunktion von Schmelzwärme von Eis . . | 28 |
| 2.6. Streudiagramm Mortalität und Weinkonsum | 31 |
| 2.7. Streudiagramm Seitenzahl - Buchpreis | 33 |
| 2.8. Gerade durch Streudiagramm | 34 |
| 2.9. Residuen | 35 |
| 2.10. Streudiagramm Körpergrössen Väter-Söhne | 38 |
| 2.11. Verkehrstote | 40 |
| 2.12. Regressionsgerade Weinkonsum-Sterblichkeit | 41 |
| 2.13. 21 verschiedene Datensätze | 44 |
| 3.1. Häufigkeitsverteilung des Alphazerfalles | 47 |
| 3.2. Venn-Diagramm Mengenlehre | 52 |
| 3.3. Wahrscheinlichkeitsbaum | 62 |
| 3.4. Wahrscheinlichkeitsbaum zweimaliges Werfen einer Münze | 64 |
| 3.5. Wahrscheinlichkeitsbaum Würfel und Jasskarte | 65 |
| 3.6. Wahrscheinlichkeitsbaum Augensumme | 66 |
| 3.7. Wahrscheinlichkeitsbaum bedingte Wahrscheinlichkeit | 67 |
| 3.8. Hilfsillustration für bedingte Wahrscheinlichkeiten | 68 |
| 3.9. Wahrscheinlichkeitsbaum bedingte Wahrscheinlichkeit | 73 |
| 3.10. Wahrscheinlichkeiten für die Anzahl Gewinne in 100 Losen | 85 |
| 3.11. Binomialwahrscheinlichkeiten in Form eines Stabsdiagramms | 87 |
| 3.12. Kumulative Verteilungsfunktion für $X \sim \text{Binomial}(100,0.5)$ | 92 |
| 4.1. Likelihood-Funktion für $X_i \sim \text{Pois}(\lambda)$ | 110 |
| 4.2. Log-Likelihood-Funktion für $X_i \sim \text{Pois}(\lambda)$ | 112 |
| 4.3. Beispiel Fehler 1. Art | 126 |
| 4.4. Fehler 2. Art für Binomialverteilung | 128 |
| 4.5. Macht des Tests für Binomialverteilung. | 129 |
| 4.6. P -Wert für Binomialverteilung | 133 |
| 4.7. Untere Grenze des 95 %-Vertrauensintervalls | 136 |

Abbildungsverzeichnis

| | | |
|-------|--|-----|
| 4.8. | Obere Grenze des 95 %-Vertrauensintervalls | 137 |
| 4.9. | Vergleich von 100 95 %-Vertrauensintervallen | 140 |
| 5.1. | Beispiel für eine kumulative Verteilungsfunktion. | 149 |
| 5.2. | Dichte und Wahrscheinlichkeit einer Zufallsvariablen | 152 |
| 5.3. | Illustration des Quantils | 153 |
| 5.4. | Dichte und Verteilungsfunktion der uniformen Verteilung | 155 |
| 5.5. | Dichte und Verteilungsfunktion der Exponentialverteilung | 158 |
| 5.6. | Dichte und Verteilungsfunktion der Normalverteilung | 161 |
| 5.7. | Dichte der Normalverteilung | 163 |
| 5.8. | 100 Messungen der Fallzeit eines Körpers. | 184 |
| 6.1. | QQ-Plot vom Datensatz der Betondruckfestigkeit | 192 |
| 6.2. | Normal-Plot von $\mathcal{N}(0, 1)$ und Cauchy-Verteilung | 193 |
| 6.3. | Normal-Plots | 194 |
| 6.4. | Dichtefunktion des zweiseitigen Z-Tests zum Niveau α | 201 |
| 6.5. | Dichten der t -Verteilung | 206 |
| 6.6. | Verschiedene Fälle (1 bis 5) von statistischer Signifikanz und fachlicher Relevanz. Die Vert | |
| 7.1. | Aktienkurs : SMI-Entwicklung 1989-2014. | 227 |
| 7.2. | Weisses Rauschen. | 228 |
| 7.3. | Random Walks | 229 |
| 7.4. | Schritte auf einem eindimensionalen Gitter bei einem Random-Walk. | 229 |
| 7.5. | Brownsche Bewegung | 233 |
| 7.6. | Wahrscheinlichkeitsdichte und mittleres Verschiebungssquadrat | 234 |
| 7.7. | Thermische Bewegung von Elektronen, Rauschspannung | 236 |
| 7.8. | Zwei diskrete stochastische Signale $S_1(t)$ und $S_2(t)$ | 237 |
| 7.9. | Relative Häufigkeiten von Signalwerten | 238 |
| 7.10. | Prozentualer Anteil von Signalwerten | 239 |
| 7.11. | Kontinuierlicher Zufallsprozess | 240 |
| 7.12. | Interpretation der Wahrscheinlichkeitsdichtefunktion. | 242 |
| 7.13. | Scharmittelwert | 243 |
| 7.14. | Gleich grosse Flächen | 245 |
| 7.15. | Zeitmittelwert | 246 |
| 7.16. | Maxwell-Boltzmann für das zweiatomige Stickstoffmolekül | 249 |
| 7.17. | Zeitreihe und Autokorrelationsfunktion | 254 |
| 7.18. | Rechtecksignal und Autokorrelationsfunktion | 259 |
| 7.19. | Realisierung eines verrauschten Rechtecksignals. | 260 |
| 7.20. | Autokorrelationsfunktion falls $r(t)$ in $S(t)$ enthalten ist | 261 |
| 7.21. | Blockschaltbild des Detektors für $r(t)$ | 261 |
| 7.22. | Korrelogramm für weisses Rauschen | 264 |
| B.1. | Chi-Quadrat-Verteilung | 281 |
| C.1. | Leistungsdichtespektrum von weissem bandbegrenztem Rauschen. | 297 |

Abbildungsverzeichnis

| | |
|--|-----|
| C.2. Rechtecksignal | 298 |
| C.3. Rücktransformation in den Zeitraum | 301 |
| C.4. Das T_0 -periodische Rechtecksignal | 301 |
| C.5. T_0 -periodische Fortsetzung | 303 |

Kapitel 1.

Einführung

It is easy to lie with statistics. It is hard to tell the truth without statistics.

(Andrejs Dunkels)

1.1. Was ist Statistik?

Das Wort *Wahrscheinlichkeit* taucht in der Alltagssprache häufig auf. Hier einige Beispiele:

- Wir hören im Wetterbericht: „Die Wahrscheinlichkeit, dass es heute morgen regnet, liegt bei 60 Prozent“.
- Weiter hört man: „Die Wahrscheinlichkeit, dass ich hundert Jahre alt werde, ist klein“.
- In Basel möchte ein Seismologe bestimmen, wie gross die Wahrscheinlichkeit ist, dass Geothermie-Bohrungen ein Erdbeben von einer bestimmten Grössenordnung auslösen.
- Ein Atomphysiker stellt sich andererseits die Frage: „Wie gross ist die Wahrscheinlichkeit, dass ein Geiger-Zähler in den nächsten 10 Sekunden 20 Zerfälle registriert?“.
- Ein Schweizer Politiker oder Nationalbanker interessiert sich momentan wohl für die Frage: „Wie gross ist die Wahrscheinlichkeit, dass der Wert vom Euro in diesem Jahr wieder über 1.20 Franken steigt?“
- Oder: „Wie gross ist die Wahrscheinlichkeit, dass es einen Börsencrash gibt?“

Wahrscheinlichkeiten geben wir im Zusammenhang mit Vermutungen an. Warum stellen wir Vermutungen an? Wir stellen Vermutungen an, wenn wir eine Aussage oder Vorhersage machen möchten, aber dazu nur über unvollständige Informationen oder

unsichere Kenntnisse verfügen. Wir stellen auch Vermutungen an, weil wir eine Entscheidung fällen möchten: „Soll ich heute morgen einen Regenschirm mitnehmen?“ „Soll ich mich bei einer Bank bewerben, oder selbstversorgender Bio-Bauer werden?“

In den Naturwissenschaften möchten wir mit Hilfe unserer beschränkten oder unvollständigen Kenntnissen ein physikalisches System so allgemein wie möglich beschreiben. Die Beschreibung eines physikalischen Systems stellt aber letztlich nichts anderes als eine Vermutung (Modell) dar, denn wir können ein (realistisches) physikalisches System niemals bis ins letzte Detail beschreiben. Nun gibt es bessere Vermutungen und schlechtere Vermutungen, wie ein physikalisches System beschaffen ist. Die *Stochastik* hilft uns dabei, bessere Vermutungen anzustellen.

Betrachten wir das Beispiel einer Münze: Wir möchten vorhersagen, ob ein Münzwurf das Ergebnis „Kopf“ oder „Zahl“ ergibt. Wüssten wir die genaue Massenverteilung der Münze, die genaue Anfangsgeschwindigkeit und Anfangsposition der Münze und die Positionen und Geschwindigkeiten aller Luftmoleküle zu jedem Zeitpunkt während des Wurfs, könnten wir wohl mit Hilfe der Mechanik vorhersagen, ob der Münzwurf mit Kopf oder Zahl auf dem Boden landet. Nun verfügen wir in der Praxis nie über alle diese Informationen. Aufgrund unserer Unkenntnis stellen wir die Vermutung an, dass die Massenverteilung der Münze dergestalt ist, dass wir diese als fair bezeichnen, d.h., die Anzahl Würfe mit „Kopf“ ist in etwa gleich der Anzahl Würfe mit „Zahl“. Je nach dem, wie stark sich die Anzahl Würfe mit „Kopf“ von der Anzahl Würfe mit „Zahl“ unterscheidet, können wir mit Hilfe der Stochastik aussagen, wie gut unsere Beobachtung mit der Vermutung zusammenpasst, dass die Münze fair ist, und ob wir an unserer Vermutung (dass die Münze fair ist) festhalten sollten.

Auch in der kinetischen Gastheorie, wo wir es mit der Grössenordnung von 10^{22} Gasmolekülen zu tun haben, können wir im besten Fall aussagen, wie wahrscheinlich es ist, dass ein Gasmolekül bei einer bestimmten Temperatur eine Geschwindigkeit in einem bestimmten Intervall hat. Denn es ist nicht realisierbar, jedem Molekül eine genaue Position und Geschwindigkeit zuzuordnen. Dies hat nicht nur mit der Komplexität des Problems zu tun; wir wissen mittlerweile, dass die Quantenmechanik verbietet, die genaue Position und Geschwindigkeit eines Atoms gleichzeitig zu bestimmen (*Heisenbergsche Unschärferelation*). Das Konzept von Wahrscheinlichkeiten ist essentiell, um das atomare Geschehen zu beschreiben.

Stochastik ist ein Teilgebiet der Mathematik und fasst als Oberbegriff die Gebiete Wahrscheinlichkeitsrechnung und Statistik zusammen. In der *Wahrscheinlichkeitsrechnung* geht man von einem Modell aus (man beschreibt einen sogenannten datengenerierenden Prozess) und leitet daraus entsprechende Eigenschaften ab. Wie in Abbildung 1.1 dargestellt, kann man sich unter einem Modell symbolisch eine Urne vorstellen, aus der man Kugeln (Daten) zieht. Zum Beispiel können wir uns die Frage stellen: „Wie gross ist die Wahrscheinlichkeit, eine rote Kugel zu ziehen?“ Diese Frage können wir beantworten, wenn wir wissen, wie viele rote und blaue Kugeln in

der Urne sind. Hat es drei rote und fünf blaue Kugeln in der Urne, so beträgt die Wahrscheinlichkeit, zufällig eine rote Kugel zu ziehen, $\frac{3}{8}$.

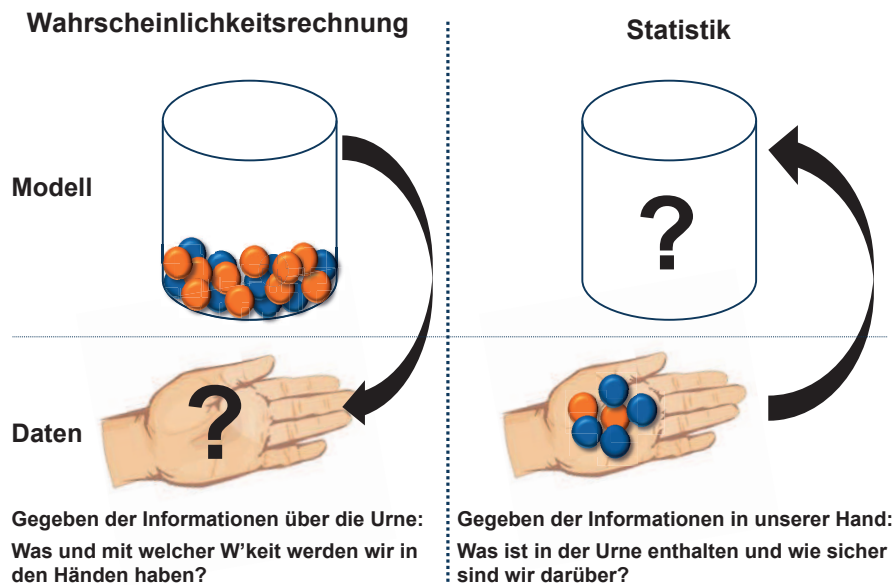


Abbildung 1.1.: Darstellung der Konzepte der Wahrscheinlichkeitsrechnung und der Statistik. Das Modell wird hier durch eine Urne symbolisiert

Betrachten wir ein anwendungsrelevantes Beispiel: Eine Gemeinde will für einen Bach einen Damm bauen, da dieser Bach oft über die Ufer tritt. Die Frage ist nun, wie hoch der Damm gebaut werden soll. Ein zu niedriger Damm ist zwar kostengünstig, dafür werden die Ausgaben für die Überschwemmungen sehr hoch. Auf der anderen Seite ist ein zu hoher Damm sehr teuer. Wie können wir nun versuchen, eine „gute“ Dammhöhe zu ermitteln? Eine „gute“ Dammhöhe zeichnet sich dadurch aus, dass der Damm genügend Sicherheit bietet, aber gleichzeitig auch noch finanzierbar ist. Hierzu müssen wir die Unsicherheit in Bezug auf den jährlichen maximalen Wasserstand quantifizieren können (z.B. in einer 30-Jahr Periode). Dazu stützen wir uns auf die Wahrscheinlichkeitsrechnung. Mit Hilfe der Wahrscheinlichkeitsrechnung können wir zum Beispiel die Wahrscheinlichkeit berechnen, dass in einer 30-Jahr Periode der maximale Wasserstand gewisse Höhen überschreitet. Oder wir können die zu erwartenden Kosten in einer 30-Jahr Periode aufgrund von Überschwemmungen bei einer gegebenen Dammhöhe berechnen. Solche Zahlen ermöglichen dann eine sinnvolle Kosten-Nutzen-Rechnung für den Bau eines Damms.

In der *Statistik* geht es darum, aus vorhandenen Daten auf das Modell zu schliessen, mit dem wir dann wieder Vorhersagen machen können. Wir denken also gerade „in die andere Richtung“. Ist in Abbildung 1.1 rechts die Anzahl der blauen und roten Kugeln in der Urne unbekannt, so können wir zum Beispiel 100-mal eine Kugel ziehen, die nach jeder Ziehung wieder zurückgelegt wird. Ziehen wir 40 rote Kugeln,

so können wir *vermuten*, dass in der Urne 40 % der Kugeln rot sind. Diese Vermutung können wir überprüfen, indem wir weitere 100 Kugeln ziehen. Stimmt das Resultat mehr oder weniger mit dem ersten Versuch überein, so halten wir an unserer Vermutung fest. Ziehen wir aber 60 rote Kugeln, so müssen wir unsere Vermutung überprüfen. Die Statistik hilft uns, quantitative Aussagen zu machen, wie gut eine Beobachtung mit einer solchen Vermutung (Modell) zusammenpasst.

Für unser Dammbispiel sehen die entsprechenden Überlegungen wie folgt aus: Wir analysieren Datenpunkte (z.B. jährliche Wasserstandsmessungen in den letzten 50 Jahren) und versuchen mit diesem beschränkten Wissen herauszufinden, was wohl ein gutes Modell für die „wahren“ jährlichen Höchststände des Wasserpegels ist. Das Modell sollte Vorhersagen über die höchstmöglichen Wasserstände machen können. Diese Vorhersagen werden, wie oben erwähnt, in der Regel in Wahrscheinlichkeiten angegeben. Ein solches Wahrscheinlichkeitsmodell ist dann die Grundlage, aufgrund derer wir die optimale Dammhöhe bestimmen werden.

In der Statistik können wir zusätzlich auch Angaben darüber machen, wie plausibel ein Modell aufgrund von Beobachtungen ist (was auf den ersten Blick erstaunlich erscheint). Werfen wir eine Münze 100-mal und erhalten 70-mal „Kopf“, können wir dann immer noch behaupten, dass die Münze fair ist? Theoretisch müssten wir 50-mal „Kopf“ erhalten, aber es wäre ja möglich, dass zufälligerweise 70-mal „Kopf“ geworfen wurde. Wann können wir den Zufall ausschliessen? Solche Fragen können wir mit der Statistik beantworten.

1.2. Kann ich Statistik überhaupt brauchen?

Die Statistik hat ihren Ursprung in der Mathematik, greift aber in viele Bereiche der modernen Wissenschaften über. Grosse Teile der Biologie, der Medizin, der Ingenieurwissenschaften und der Umweltforschung wären heute ohne Statistik undenkbar. Der Chefökonom von Google, Hal Varian, sagte vor einigen Jahren: Der sechste Beruf des kommenden Jahrzehnts ist der des Statistikers. In der Vergangenheit bis heute stammen viele Anregungen und Problemstellungen für die Statistik aus den Bereichen Biologie und Pharmazie.

Eine häufige Frage bei Tier und Mensch lautet: wie bestätigt man die Wirksamkeit eines neuen Medikamentes. Dabei erhält eine zufällig ausgewählte Gruppe (Medikamentengruppe) von Patienten das neue Medikament in Form einer Tablette. Eine andere zufällig ausgewählte Gruppe von Patienten (Kontrollgruppe) erhält ein Placebo¹. Die Medikamentengruppe hat gegenüber der Kontrollgruppe nach zwei Wochen eine deutliche Verbesserung der Symptome gezeigt. Das Medikament wirkt also. Oder doch nicht? Kann es sein, dass das Medikament gar nicht wirkt und alle

¹Tablette mit gleichem Aussehen und Geschmack wie das Medikament aber ohne Wirkstoff.

Personen der Medikamentengruppe unabhängig vom Medikament *zufällig* eine Verbesserung der Symptome hatte? Wie können wir „Zufall“ ausschliessen? Mit Statistik lässt sich diese Frage beantworten.

Oder in der Genetik stellt sich zum Beispiel folgendes Problem: Man beobachtet, dass in einer Gruppe von Krebspatienten gewisse Gene stärker aktiv sind als in einer Kontrollgruppe. Könnte es Zufall sein, dass alle Personen, bei denen diese Gene aktiver sind in derselben Gruppe gelandet sind? Hat also das aktive Gen gar nichts mit der Krebserkrankung zu tun, obwohl es bei allen Krebspatienten vorkommt? Wie können wir hier mathematisch Zufall ausschliessen?

Aber auch die Ingenieurwissenschaften liefern sehr interessante Aufgabenstellungen für die Statistik. Ein typisches Beispiel, das vielen Leuten gar nicht bekannt sein dürfte: Spracherkennungsprogramme oder Programme, mit denen Roboter visuell ihre Umwelt erkennen können – Stichwort „Machine Learning“ –, funktionieren nur mit Statistik. Auch da geht es um den Umgang mit Unsicherheiten und Variabilität. Eine Silbe oder ein Wort wird von jeder Person leicht anders ausgesprochen. Ein gutes System muss trotz diesen Variationen ein Wort oder einen Text erkennen können. Und ein Roboter wird kaum zweimal die genau gleiche Situation antreffen. Trotzdem muss er entscheiden können, ob eine angetroffene Situation einer gespeicherten Standardsituation ähnlich ist.

Diese Vorlesung soll Ihnen helfen, ein Fundament zu legen und die Grundbegriffe in der Statistik verstehen und anwenden zu können.

1.3. Was ist der Inhalt dieses Kurses?

In Kapitel 2 werden wir uns mit den sehr wahrscheinlich bereits bekannten Grössen der Statistik wie arithmetisches Mittel, Standardabweichung, Varianz, Quantil, Median und Korrelation auseinandersetzen. Dabei werden wir lernen, wie Daten mit der Statistiksoftware R graphisch dargestellt werden können.

In Kapitel 3 geht es zunächst darum, den Begriff „Zufall“ mathematisch genau zu definieren. Wenn das geschehen ist, kann man mit Wahrscheinlichkeiten einfach rechnen. Wir werden zunächst nur Zähldaten behandeln, also Situationen, in denen nur ganze Zahlen auftreten (z.B. die Anzahl Gewinne, wenn man 20 Lose kauft). Wir werden die Begriffe Unabhängigkeit, bedingte Wahrscheinlichkeit, Zufallsvariable, Erwartungswert und Standardabweichung kennenlernen. Ausserdem werden wir verschiedene Arten von Zufall gruppieren und dabei den Begriff der Verteilung kennenlernen.

In Kapitel 4 lernen wir die drei Grundfragen der Statistik - Punktschätzung, Hypothesentest, Vertrauensintervall - anhand von Zähldaten kennen. Damit können wir folgende generische Frage beantworten: Angenommen wir haben ein Medikament

an 100 Testpersonen ausprobiert. 67 Personen wurden gesund. Bei welchem Anteil der Gesamtbevölkerung wirkt das Medikament? In welchem Bereich liegt wohl der wahre Anteil mit grosser Wahrscheinlichkeit (z.B. mit 95%-Wahrscheinlichkeit zwischen 60% und 70%)? Der **Binomialtest** wird uns hier eine Antwort liefern.

In Kapitel 5 werden wir kontinuierliche Wahrscheinlichkeitsverteilungen kennenlernen, die wir zur Modellierung von kontinuierlichen, reellen Daten (z.B. Grösse, Gewicht, etc.) benötigen. Wir werden uns mit der berühmtesten Wahrscheinlichkeitsverteilung, der sogenannten **Normalverteilung** (auch bekannt als Gauss'sche Verteilung), eingehend beschäftigen. Dabei interessiert uns auch, wie der arithmetische Mittelwert einer Messreihe verteilt ist. Denn bei der Angabe des Mittelwertes geben wir in der Regel auch den Messfehler an (entweder den absoluten oder relativen Fehler).

In Kapitel 6 erweitern wir die statistischen Methoden, die wir in Kapitel 4 kennengelernt haben, auf kontinuierliche, reelle Daten. Angenommen, wir haben neue Augentropfen entwickelt, die den Augeninnendruck senken sollen. Wir wählen zufällig 20 Testpersonen und teilen sie zufällig in zwei Gruppen mit je 10 Personen auf. Gruppe *N* wird mit den neuen Augentropfen behandelt, die andere Gruppe *A* mit herkömmlichen Augentropfen. In Gruppe *N* scheint der Augeninnendruck stärker zu sinken als in der Gruppe *A*. Die zentrale Frage, die wir in diesem Kapitel beantworten, ist folgende: Könnte es sein, dass beide Medikamente gleich gut wirken, aber die Personen, die besonders gut auf Augentropfen ansprechen, zufällig in Gruppe *N* zusammengefasst wurden? Der **t-Test** und der **Wilcoxon-Test** werden uns hier eine Antwort liefern.

In Kapitel 7 beschäftigen uns **stochastische Prozesse** wie die Brownsche Bewegung oder weisses Rauschen. Diese finden häufige Anwendung in den Ingenieur- und Naturwissenschaften, z.B. für die Beschreibung von thermischem Rauschen an elektrischen Widerständen oder für die Erkennung eines deterministischen Signals in einem verrauschten Signal.

1.4. Software

Wir werden die Statistiksoftware **R** verwenden, insbesondere **R-Studio**. Die Beherrschung von **R** ist wesentlicher Bestandteil dieses Stochastik Moduls und auch prüfungsrelevant. **R-Studio** können Sie kostenlos über folgenden [Link](#) beziehen und auf Ihrem Laptop installieren. Unter [Docs](#) finden Sie auch eine sehr gut gestaltete Bedienungsanleitung zu **R-Studio**, die wir sehr empfehlen können. Andere weit verbreitete Statistikprogramme sind **SPSS** und **SAS**. Alle Methoden, die wir in diesem Kurs besprechen, sind in jeder Statistiksoftware implementiert. In der ersten Unterrichtseinheit wird es eine Einführung in die Benützung von **R** geben. [\[Dal08\]](#) ist ein

hervorragendes Buch, das auf Ilias abgelegt wird und das Sie sowohl als Nachschlagewerk zur Benützung von R als auch zu allen in der Vorlesung behandelten Themen verwenden können.

1.5. Literaturverzeichnis

- [Dal08] P. Dalgaard. *Introductory Statistics with R*. Statistics and Computing. Springer, 2008.
- [Kal] M. Kalisch. Statistik für Biologie und Pharmazeutische Wissenschaften. Vorlesungsskript.
- [MR14] D.C. Montgomery and G.C. Runger. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, 2014.
- [Pap06] Lothar Papula. *Mathematische Formelsammlung für Ingenieure und Naturwissenschaftler: mit zahlreichen Rechenbeispielen und einer ausführlichen Integraltafel*. Lothar Papula. Vieweg, 2006.
- [Pap08] L. Papula. *Mathematik für Ingenieure und Naturwissenschaftler 3: Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichsrechnung*. Viewegs Fachbücher der Technik. Vieweg + Teubner, 2008.
- [Ric06] J. Rice. *Mathematical Statistics and Data Analysis*. Number S. 3 in Advanced series. Cengage Learning, 2006.

Dieses Vorlesungsskript beruht auf dem von Markus Kalisch, Peter Bühlmann und Hansruedi Künsch verfassten und an der ETH Zürich verwendeten Skript Statistik für Biologie und Pharmazeutische Wissenschaften [[Kal](#)].

Wir empfehlen das bewährte Lehrbuch von Lothar Papula [[Pap08](#)].

Als Ergänzung oder als ausführlichere Alternative zum Skript empfehlen wir für besonders Interessierte die im angelsächsischen Raum verbreiteten und in englischer Sprache geschriebenen Lehrbücher [[MR14](#)] und [[Ric06](#)]. Zusätzlich empfehlen wir die Formelsammlung [[Pap06](#)], die auch an der Modulendprüfung benützt werden darf.

Kapitel 2.

Deskriptive Statistik

Definition of Statistics: The science of producing unreliable facts from reliable figures.

(Evan Esar)

2.1. Deskriptive Statistik eindimensionaler Daten

Die deskriptive Statistik befasst sich mit der Darstellung von Datensätzen (Zusammenstellung verschiedener Daten). Dabei werden diese Datensätze durch gewisse Zahlen charakterisiert (zum Beispiel den Mittelwert) und graphisch in einem Koordinatensystem dargestellt. Wir befassen uns zunächst mit *eindimensionalen* Daten, wo *eine* Messgrösse an einem Untersuchungsobjekt ermittelt wird. Anhand des folgenden Beispieles werden wir die wichtigen Begriffe und Vorgehensweisen genauer kennenlernen.

2.1.1. Messungen der Schmelzwärme von Eis

Als Einführungsbeispiel betrachten wir zwei *Datensätze*, bei welchen zwei Methoden zur Bestimmung der latenten Schmelzwärme von Eis verglichen werden. Wiederholte Messungen der freigesetzten Wärme beim Übergang von Eis bei -0.7°C zu Wasser bei 0°C ergaben die Werte (in cal/g), die in Tabelle 2.1 aufgeführt sind. Obwohl die Messungen mit der grösstmöglichen Sorgfalt durchgeführt und alle Störeinflüsse ausgeschaltet wurden, variieren die Messwerte innerhalb beider Methoden. Es stellen sich hier nun die folgenden Fragen:

- Gibt es einen Unterschied zwischen der Methode A und der Methode B?
- Falls ja, wie können wir diesen Unterschied ermitteln?

| | | | | | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Methode A | 79.98 | 80.04 | 80.02 | 80.04 | 80.03 | 80.03 | 80.04 | 79.97 | 80.05 |
| Methode A | 80.03 | 80.02 | 80.00 | 80.02 | | | | | |
| Methode B | 80.02 | 79.94 | 79.98 | 79.97 | 79.97 | 80.03 | 79.95 | 79.97 | |

Tabelle 2.1.: Messungen zur Bestimmung der latenten Schmelzwärme von Eis anhand von zwei Methoden.

Es fällt auf, dass bei beiden Methoden die Messwerte um 80 herum liegen. Bei Methode A liegen aber nur 2 Werte von 13 *unter* 80, während bei der Methode B nur 2 von 8 Werten *über* 80 liegen. Die Werte der Methode A sind also eher grösser als die der Methode B. Was heisst hier aber „eher“? Es ist also von Interesse, die Messreihen irgendwie so zusammenzufassen, dass wir die beiden Methoden miteinander vergleichen können.

Die *deskriptive Statistik* beschäftigt sich damit, auf welche Weisen (quantitative) Daten organisiert und zusammengefasst werden können. Dies hat zum Ziel, dass die Interpretation und darauffolgende statistische Analyse dieser Daten vereinfacht werden. Wir machen dies mit Hilfe von

- graphischen Darstellungen
- Zusammenfassungen von Daten, die die wichtigen Merkmale der Daten hervorheben sollen, wie eben zum Beispiel die mittlere Lage der Messwerte und die Streuung dieser Messwerte um die mittlere Lage.

Diese sogenannten *Kennzahlen* sollen die Daten numerisch zusammenfassen und grob charakterisieren.

Bei statistischen Analysen, wie wir sie im Laufe der Vorlesung kennenlernen werden, ist es ausserordentlich wichtig, nicht einfach blind ein Modell anzupassen oder ein statistisches Verfahren anzuwenden. Die Daten sollten immer mit Hilfe von geeigneten graphischen Mitteln *und* den Kennzahlen dargestellt werden, da man nur auf diese Weise (teils unerwartete) Strukturen und Besonderheiten entdecken kann.

Im Folgenden werden die Daten mit x_1, \dots, x_n bezeichnet, wobei n der *Umfang* der Messreihe genannt wird. Im Fall der Messreihe der Methode A ist dies für $n = 13$:

$$x_1 = 79.98, \quad x_2 = 80.04, \quad \dots, \quad x_{13} = 80.02$$

2.1.2. Darstellung von Messwerten

Bevor wir uns mit Kennzahlen und graphischen Darstellungen von Datensätzen auseinandersetzen, müssen wir Regeln für die Darstellung von Messwerten festlegen. Dazu benötigen wir die Begriffe *Nachkommastellen* und *signifikante Stellen*.

Als **Nachkommastellen** werden die in der dezimalen Darstellung einer Zahl verwendeten Ziffern rechts des Kommas bezeichnet. Im obigen Beispiel haben die Messpunkte

$$x_1 = 79.98, \quad x_2 = 80.04, \quad \dots, \quad x_{13} = 80.02$$

zwei Nachkommastellen.

Die **signifikanten Stellen** werden als die erste von Null verschiedene Stelle bis zur Rundungsstelle definiert. Die Rundungsstelle ist die letzte Stelle, die nach dem Runden noch angegeben werden kann. Im obigen Beispiel haben wir also **vier** signifikante Stellen.

Beispiel 2.1.1

| Zahl | Anzahl Signifikante Stellen | Anzahl Nachkommastellen |
|--------------------|-----------------------------|-------------------------|
| 98.76 | 4 | 2 |
| 0.009876 | 4 | 6 |
| $987.6 \cdot 10^4$ | 4 | 1 |
| $9.876 \cdot 10^6$ | 4 | 3 |

□

Bemerkungen:

- i. Ganze Zahlen haben keine Nachkommastellen.
- ii. In manchen Fällen ist die Bestimmung der signifikanten Stellen unklar: Besitzt 20 eine, zwei oder sogar mehr signifikante Stellen? Je nach Zusammenhang ist eine Zahl exakt zu werten, wenn sie z. B. als natürliche Zahl verwendet wird; oder sie ist als gerundete Zahl zu werten, wenn sie als Zahlenwert zu einer physikalischen Grösse verwendet wird. Zu einer exakten Zahl stellt sich die Frage nach der Signifikanz nicht, da sie mit beliebig vielen Nachkomma-Nullen verlängert werden kann.
- iii. Um zu einer mittels Messtechnik ermittelten Grösse beim Zahlenwert 20 eine Mehrdeutigkeit zu vermeiden, soll man die wissenschaftliche Schreibweise mit Zehnerpotenz-Faktor wählen. Im Fall von einer signifikanten Stelle also $2 \cdot 10^1$; im Fall von drei signifikanten Stellen $2.00 \cdot 10^1$.

Darstellung Rechenergebnis

Bei der Darstellung eines Rechenergebnis von Messwerten gelten folgende zwei Regeln:

1. Das Ergebnis einer **Addition/Subtraktion** bekommt genauso viele Nachkommastellen wie die Zahl mit den wenigsten Nachkommastellen.
2. Das Ergebnis einer **Multiplikation/Division** bekommt genauso viele signifikante Stellen wie die Zahl mit den wenigsten signifikanten Stellen.

Beispiel 2.1.2

| Zahlen | Kleinste Anzahl Signifikante Stellen | Kleinste Anzahl Nachkommastellen | Ergebnis |
|----------------------|--------------------------------------|----------------------------------|----------|
| $20.567 + 0.0007$ | | 3 | 20.568 |
| $12 + 1.234$ | | 0 | 13 |
| $12.00 + 1.234$ | | 2 | 13.23 |
| $12.000 + 1.234$ | | 3 | 13.234 |
| $1.234 \cdot 3.33$ | 3 | | 4.11 |
| $1.234 \cdot 0.0015$ | 2 | | 0.0019 |

□

Bemerkungen:

- i. Eine Rundung sollte erst möglichst spät innerhalb des Rechnungsgangs durchgeführt werden. Sonst können sich mehrere Rundungsabweichungen zu einer grösseren Gesamtabweichung zusammensetzen. Um diese Vergrößerung zu vermeiden, sollen in Zwischenrechnungen bekannte Grössen mit mindestens einer Stelle mehr eingesetzt werden als im Ergebnis angegeben werden kann.

2.1.3. Kennzahlen

Häufig ist es sinnvoll, Datensätze *numerisch* zusammenzufassen. Die Datensätze werden dabei auf eine oder mehrere Zahlen reduziert. Dazu verwenden wir meistens zwei *Kenngrössen*: Eine beschreibt die mittlere Lage der Messwerte und die andere die Variabilität oder Streuung dieser Messwerte. Mit Streuung meinen wir die „durchschnittliche“ Abweichung der Messwerte von der mittleren Lage.

Arithmetisches Mittel

Die bekannteste Grösse für die mittlere Lage ist der wohlbekannte Durchschnitt oder das

Arithmetische Mittel \bar{x}_n

$$\bar{x}_n = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Beispiel 2.1.3 Messung der Schmelzwärme von Eis mit Methode A

Das arithmetische Mittel der $n = 13$ Messungen ist

$$\bar{x}_{13} = \frac{79.98 + 80.04 + \cdots + 80.03 + 80.02 + 80.00 + 80.02}{13} = 80.02$$

Wir summieren also alle Werte auf und dividieren die Summe durch die Anzahl der Werte.

□

Mit R berechnen wir den Mittelwert wie folgt.

```
methodeA <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03,
             80.04, 79.97, 80.05, 80.03, 80.02, 80, 80.02)
mean(methodeA)

## [1] 80.02077
```

Empirische Varianz und Standardabweichung

Obwohl das arithmetische Mittel schon einiges über einen Datensatz aussagt, beschreibt er diesen aber nur unvollständig. Wir betrachten als Beispiel die folgenden beiden Datensätze von (fiktiven) Schulnoten:

2; 6; 3; 5 und 4; 4; 4; 4

Beide haben denselben Mittelwert 4, aber die Verteilung der Daten um den Mittelwert ist sehr unterschiedlich. Im ersten Fall gibt es zwei gute und zwei schlechte Schüler

und im zweiten Fall sind alle Schüler gleich gut. Wir sagen, die Datensätze haben eine unterschiedliche Streuung um die Mittelwerte.

Wir wollen diese Streuung numerisch erfassen. Ein erster Ansatz besteht darin, dass man den Durchschnitt der *Unterschiede zum Mittelwert* nimmt. Im ersten Fall wäre dies

$$\frac{(2-4) + (6-4) + (3-4) + (5-4)}{4} = \frac{-2 + 2 - 1 + 1}{4} = 0$$

Im zweiten Fall gibt dies auch 0. Diese Methode trägt also nicht viel zur Beschreibung der Streuung bei, da die Unterschiede zum Mittelwert *negativ* werden können und sich diese wie im obigen Fall aufheben können.

Der nächste Ansatz geht dahin, dass wir die Unterschiede zum Mittelwert durch die Absolutwerte der Unterschiede zum Mittelwert ersetzen. Im ersten Fall erhalten wir dann

$$\frac{|(2-4)| + |(6-4)| + |(3-4)| + |(5-4)|}{4} = \frac{2 + 2 + 1 + 1}{4} = 1.5$$

Die Noten weichen nun im Schnitt 1.5 Noten vom Mittelwert ab. Im zweiten Fall ist dieser Wert natürlich 0. Je grösser dieser Wert (der immer grösser gleich 0 ist), desto mehr unterscheiden sich die Daten bei gleichem Mittelwert voneinander. Dieser Wert für die Streuung heisst auch *mittlere absolute Abweichung*.

Da es sich mit Absolutwerten nicht einfach rechnen lässt (zum Beispiel Ableitungen), wird die (scheinbar kompliziertere) *empirische Varianz* und *empirische Standardabweichung* für das Mass der Variabilität oder Streuung der Messwerte verwendet. Diese sind definiert durch

Empirische Varianz $\text{Var}(x)$ und Standardabweichung s_x

$$\text{Var}(x) = \frac{(x_1 - \bar{x}_n)^2 + (x_2 - \bar{x}_n)^2 + \dots + (x_n - \bar{x}_n)^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

$$s_x = \sqrt{\text{Var}(x)} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}.$$

Bemerkungen:

- i. Bei der Varianz quadrieren wir die Abweichungen $x_i - \bar{x}_n$, damit sich die Abweichungen vom Mittelwert nicht gegenseitig aufheben können. Der Nenner $n-1$, anstelle von n , ist mathematisch begründet¹.

¹Eine genaue Begründung finden Sie im Anhang in Kapitel A.7

- ii. Die Standardabweichung ist die Wurzel der Varianz. Da wir für die Berechnung der Varianz die Quadrate der Abstände zum Mittelwert verwendet haben, bekommen wir durch das Wurzelziehen wieder dieselbe Einheit wie bei den Daten selbst. Der Wert der empirischen Varianz hat keine physikalische Bedeutung. Wir wissen nur, je grösser der Wert, desto grösser die Streuung.

Beispiel 2.1.4 Messung der Schmelzwärme von Eis mit Methode A

Das arithmetische Mittel der $n = 13$ Messungen ist $\bar{x}_{13} = 80.02$ (siehe oben) und die empirische Varianz ergibt

$$\begin{aligned}\text{Var}(x) &= \frac{(79.98 - 80.02)^2 + (80.04 - 80.02)^2 + \dots + (80.00 - 80.02)^2 + (80.02 - 80.02)^2}{13 - 1} \\ &= 0.0005744\end{aligned}$$

Die empirische Standardabweichung ist dann

$$s_x = \sqrt{0.000574} = 0.02397$$

Somit ist die „mittlere“ Abweichung vom Mittelwert 0.023 97 cal/g.

Für Methode B finden wir $\bar{x}_8 = 79.98$ und $s_x = 0.03137$ mit der analogen Interpretation.

□

Die empirische Varianz bzw. Standardabweichung (englisch: standard deviation) ist von Hand mühsam auszurechnen, deswegen benutzen wir R:

```
var(methodeA)

## [1] 0.000574359

sd(methodeA)

## [1] 0.02396579
```

2.1.4. Weitere Kennzahlen

Im Folgenden werden wir zwei alternative Kenngrössen studieren, und zwar den *Median* als Lagemass und die *Quartilsdifferenz* als Streuungsmass.

Median

Ein weiteres Lagemass für die mittlere Lage ist der *Median*. Es handelt sich dabei um den Wert, bei dem rund die Hälfte der Messwerte unterhalb von diesem Wert liegen. Ist beispielsweise bei einer Prüfung der Median 4.6, dann hat die Hälfte der Klasse eine Note unterhalb von 4.6. Umgekehrt liegen die Noten der anderen Hälfte *oberhalb* dieser Note.

Um den *Median* zu bestimmen, müssen wir die Daten zuerst der Grösse nach ordnen:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Für die Daten der Methode A ergibt dies

79.97; 79.98; 80.00; 80.02; 80.02; 80.02; 80.03; 80.03; 80.03; 80.04; 80.04; 80.04; 80.05

Der Median von diesen 13 Messungen ist dann der Wert der mittleren Beobachtung. Dies ist in diesem Fall der Wert der 7. Beobachtung:

79.97; 79.98; 80.00; 80.02; 80.02; 80.02; **80.03**; 80.03; 80.03; 80.04; 80.04; 80.04; 80.05

Der Median des Datensatzes der Methode A lautet somit 80.03. 6 Beobachtungen sind kleiner oder gleich 80.03 und 6 Messwerte sind grösser oder gleich 80.03. In diesem Beispiel ist die Anzahl der Daten ungerade, und somit gibt es eine mittlere Beobachtung. Ist die Anzahl der Daten gerade, so gibt es zwei gleichwertige mittlere Beobachtungen. Als Median benützen wir in diesem Fall den Mittelwert der beiden mittleren Beobachtungen. Der Datensatz der Methode B hat 8 Beobachtungen. Wir ordnen den Datensatz der Grösse nach und definieren als Median den Durchschnitt von der 4. und 5. Beobachtung:

79.94; 79.95; 79.97; **79.97; 79.97**; 79.94; 80.02; 80.03

$$\frac{79.97 + 79.97}{2} = 79.97$$

Der Median der Daten der Methode B ist 79.97. Das heisst, die Hälfte der Messwerte ist kleiner oder gleich diesem Wert und die andere Hälfte ist grösser oder gleich diesem Wert. Die Werte der beiden mittleren Beobachtungen sind hier zufällig gleich, dies ist aber im Allgemeinen nicht so.

Mit R bestimmen wir den Median wie folgt:

```
median(methodeA)
```

```
## [1] 80.03
```

```
methodeB <- c(80.02, 79.94, 79.98, 79.97, 79.97, 80.03,  
             79.95, 79.97)  
  
median(methodeB)  
  
## [1] 79.97
```

Wir haben nun zwei Lagemasse für die Mitte eines Datensatzes: das arithmetische Mittel und den Median. Welches sind die Vorzüge der jeweiligen Lagemasse? Eine Eigenschaft des Medians ist die *Robustheit*. Der Median wird weniger stark durch extreme Beobachtungen beeinflusst als das arithmetische Mittel.

Beispiel 2.1.5 Messung der Schmelzwärme von Eis mit Methode A

Bei der grössten Beobachtung ($x_9 = 80.05$) ist ein Tippfehler passiert und $x_9 = 800.5$ eingegeben worden. Das arithmetische Mittel ist dann

$$\bar{x}_{13} = 135.44$$

Der Median ist aber nach wie vor

$$x_{(7)} = 80.03$$

Das arithmetische Mittel wird also durch Veränderung einer Beobachtung sehr stark beeinflusst, während der Median hier gleich bleibt – er ist *robust*.

□

Quartile

Das **untere Quartil** ist derjenige Wert, bei welchem 25 % aller Beobachtungen kleiner oder gleich gross und 75 % grösser oder gleich gross wie dieser Wert sind. Dementsprechend ist das **obere Quartil** derjenige Wert, bei dem 75 % aller Beobachtungen kleiner oder gleich gross und 25 % grösser oder gleich gross sind wie dieser Wert.

Allerdings gibt es für die meisten Datensätze nicht *exakt* 25 % der Anzahl Beobachtungen, wie folgendes Beispiel zeigt: Die Methode A hat $n = 13$ Messpunkte und 25 % dieser Anzahl ist 3.25. Wir *wählen* in diesem Fall den nächstgrösseren Wert $x_{(4)}$ als unteres Quartil:

79.97; 79.98; 80.00; 80.02; 80.02; 80.02; 80.03; 80.03; 80.03; 80.04; 80.04; 80.04; 80.05

Das untere Quartil ist somit 80.02. Rund ein Viertel der Messwerte sind kleiner oder gleich gross wie dieser Wert. Für das obere Quartil wählen wir $x_{(10)}$, da für $0.75 \cdot 13 = 9.75$ die Zahl 10 der nächsthöhere Wert ist.

79.97; 79.98; 80.00; 80.02; 80.02; 80.02; 80.03; 80.03; 80.03; 80.04; 80.04; 80.05

Rund drei Viertel aller Messwerte sind also kleiner oder gleich 80.04. Bei der Methode *B* sind 25 % von 8 Werten zwei Werte. Dies ist eine ganze Zahl, und wir wählen den Durchschnitt von $x_{(2)}$ und $x_{(3)}$ als unteres Quartil. Dann sind 2 Beobachtungen kleiner und 6 Beobachtungen grösser als dieser Wert.

79.94; 79.95; 79.97; 79.97; 79.97; 79.94; 80.02; 80.03

$$\frac{79.95 + 79.97}{2} = 79.96$$

Das untere Quartil der Methode *B* ist also 79.96.

Bemerkungen:

- i. Wir haben hier für den Fall, dass die Anzahl Beobachtungen keine ganze Zahl ist, jeweils aufgerundet. Somit liegt das untere Quartil für Methode *A* bei der 4. Beobachtung, was etwa 31 % der Beobachtungen entspricht. Hätten wir abgerundet, so entspräche dies der 3. Beobachtung, also ungefähr 23 %, was eigentlich näher bei 25 % liegt als 31 %. Der Unterschied ist hier allerdings nur deswegen recht gross, da der Datensatz mit $n = 13$ ziemlich klein ist. Für grosse Datensätze spielt es praktisch keine Rolle, ob wir auf- oder abrunden.
- ii. Es existieren in der Statistik mehrere Definitionen für die Quartile und Quantile. Allerdings sind die Unterschiede für grosse Datensätze klein.

Die Software *R* kennt keine eigenen Befehle für die Quartile. Wir können allerdings den allgemeineren Befehl `quantile` benutzen (die Quantile werden gleich genauer behandelt). Damit *R* die Quartile nach unserer Definition berechnet, müssen wir die Option `type=2` hinzufügen.

```
# Syntax für das untere Quartil: p=0.25

quantile(methodeA, 0.25, type = 2)

##      25%
## 80.02

quantile(methodeB, 0.25, type = 2)
```

```
##      25%
## 79.96

# Syntax f\'ur das obere Quartil: p=0.75

quantile(methodeA, 0.75, type = 2)

##      75%
## 80.04
```

Quartilsdifferenz

Die Quartilsdifferenz

$$\text{oberes Quartil} - \text{unteres Quartil}$$

ist ein Streuungsmass für die Daten. Es misst die Länge des Intervalls, das etwa die Hälfte der mittleren Beobachtungen enthält. Je kleiner dieses Mass, umso näher liegt die Hälfte aller Werte beim Median und umso kleiner ist die Streuung. Dieses Streuungsmass ist robust. So ist die Quartilsdifferenz bei der Methode A

$$80.04 - 80.02 = 0.02$$

```
IQR(methodeA, type = 2)

## [1] 0.02
```

Rund die Hälfte aller Messwerte liegt also in einem Bereich der Länge 0.02.

Quantile

Mit der *Quantile* können wir das Konzept der Quartile auf jede andere Prozentzahl verallgemeinern. So ist das 10 %-Quantil derjenige Wert, wo 10 % der Werte kleiner oder gleich und 90 % der Werte grösser oder gleich diesem Wert sind.

Das *empirische α -Quantil* ist anschaulich gesprochen der Wert, bei dem $\alpha \times 100\%$ der Datenpunkte kleiner oder gleich und $(1 - \alpha) \times 100\%$ der Punkte grösser oder gleich sind.

Empirische α -Quantile

$$\frac{1}{2}(x_{(\alpha n)} + x_{(\alpha n + 1)}) \text{ , falls } \alpha \cdot n \text{ eine natürliche Zahl ist,}$$

$$x_{(k)} \text{ wobei } k \text{ die Zahl } \alpha \cdot n \text{ aufgerundet ist , falls } \alpha \cdot n \notin \mathbb{N} .$$

Bemerkungen:

- i. Der empirische Median ist das empirische 50 %-Quantil; das empirische 25 %-Quantil ist das untere Quartil und das empirische 75 %-Quantil das obere Quartil.

Beispiel 2.1.6 Messung der Schmelzwärme von Eis mit Methode A

Wir bestimmen Median, unteres und oberes Quartil mit Hilfe der Definition oben.

Es sind $n = 13$ Messwerte, die wir zuerst der Grösse nach ordnen: der kleinste Wert ist $x_{(1)} = 79.97$, der drittgrösste Wert $x_{(3)} = 80.00$, der grösste Wert $x_{(13)} = 80.05$. Wir wollen das 25 %-Quantil, den Median und das 75 %-Quantil bestimmen. Im Fall vom 25 %-Quantil ist dann $\alpha = 0.25$, also

$$\alpha \cdot n = 0.25 \cdot 13 = 3.25$$

was keine natürliche Zahl ist; folglich runden wir 3.25 auf 4 auf und erhalten für das 25 %-Quantil $x_{(4)} = 80.02$. Im Fall vom Median ist $\alpha = 0.5$, also

$$\alpha \cdot n = 0.5 \cdot 13 = 6.5$$

was keine natürliche Zahl ist; folglich runden wir 6.5 auf 7 auf und erhalten für den Median $x_{(7)} = 80.03$. Im Fall vom 75 %-Quantil ist $\alpha = 0.75$, also

$$\alpha \cdot n = 0.75 \cdot 13 = 9.75$$

was keine natürliche Zahl ist; folglich runden wir 9.75 auf 10 auf und erhalten für das 75 %-Quantil den Beobachtungswert $x_{(10)} = 80.04$.

□

Das 10 %- und 70 %-Quantil der Methode A berechnen wir wie folgt:

```
quantile(methodeA, 0.1, type = 2)
```

```
##      10%
## 79.98
```

```
quantile(methodeA, 0.7, type = 2)
```

```
##      70%  
## 80.04
```

Rund 10 % der Messwerte sind kleiner oder gleich 79.98 . Entsprechend sind rund 70 % der Messwerte kleiner oder gleich 80.04.

Beispiel 2.1.7

In einer Schulklasse mit 24 SchülerInnen gab es an einer Prüfung folgende Noten:

4.2, 2.3, 5.6, 4.5, 4.8, 3.9, 5.9, 2.4, 5.9, 6, 4, 3.7, 5, 5.2, 4.5, 3.6, 5, 6, 2.8, 3.3, 5.5, 4.2, 4.9, 5.1

Wir berechnen nun mit R verschiedene Quantile:

```
noten.1 <- c(4.2, 2.3, 5.6, 4.5, 4.8, 3.9, 5.9, 2.4, 5.9,  
             6, 4, 3.7, 5, 5.2, 4.5, 3.6, 5, 6, 2.8, 3.3, 5.5, 4.2,  
             4.9, 5.1)
```

```
quantile(noten.1, seq(0.2, 1, 0.2), type = 2)
```

```
##    20%   40%   60%   80%  100%  
##    3.6   4.2   5.0   5.6   6.0
```

Rund 20 % der SchülerInnen haben also eine 3.6 oder waren schlechter. Genau 20 % der SchülerInnen ist nicht möglich, da dies 4.8 SchülerInnen entsprechen würde. Das 60 %-Quantil besagt, dass 60 Prozent der SchülerInnen eine 5 haben oder schlechter waren. Folglich haben 40 % eine 5 oder sind besser.

□

2.1.5. Graphische Methoden

Histogramm

Einen graphischen Überblick über die auftretenden Werte erhalten wir mit einem sogenannten *Histogramm*. Histogramme helfen uns bei der Frage, in welchem Wertebereich besonders viele Datenpunkte liegen. Ist die Datenmenge gross, so macht es keinen Sinn, alle Werte einzeln zu betrachten. Wir bilden sogenannte *Klassen*, die

jeweils einen Ausschnitt des Beobachtungsbereiches darstellen. Um ein Histogramm zu zeichnen, bildet man Klassen (einfachheitshalber mit konstanter Breite) und zählt, wie viele Beobachtungen in jede Klasse fallen. Es gibt verschiedene Arten von Histogrammen; wir behandeln hier nur die gebräuchlichste.

Beispiel 2.1.8

In Abbildung 2.1 sehen wir ein Histogramm von dem Ergebnis eines IQ-Testes von 200 Personen.

- Die Breite der Klassen wurde mit 10 IQ-Punkten festgelegt und ist für jede Klasse gleich.
- Die Höhe der Balken gibt die Anzahl Personen an, die in diese Klasse fallen. Zum Beispiel fallen ca. 14 Personen in die Klasse zwischen 120 und 130 IQ-Punkten.

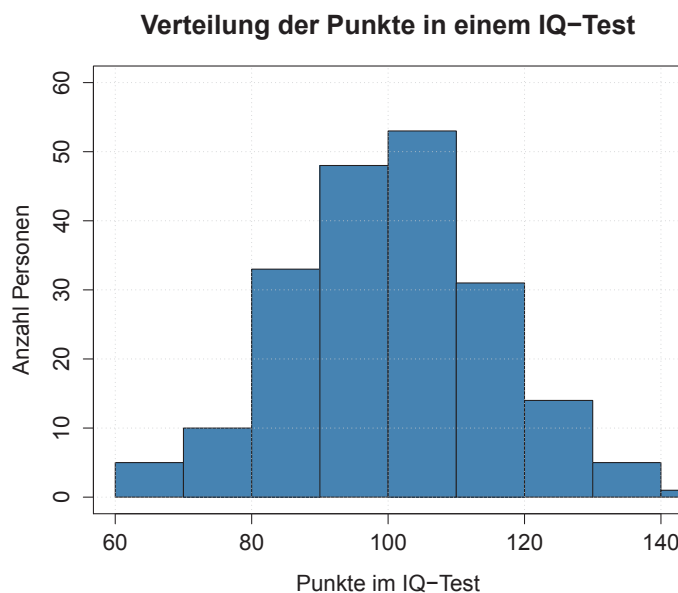


Abbildung 2.1.: Histogramm von dem IQ-Test Ergebnis von 200 Personen.

□

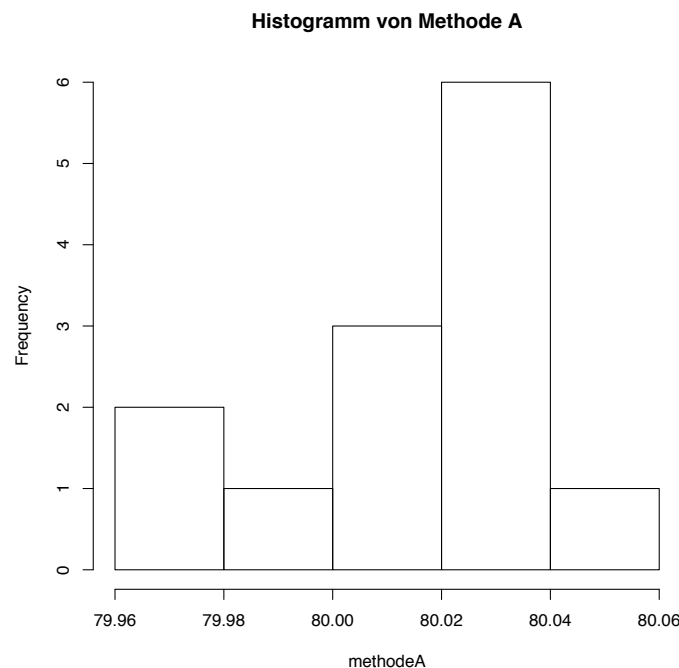
Schrittweise Konstruktion eines Histogramms:

- Wir teilen die Datenmenge in Klassen ein. Für die Festlegung der Anzahl der Klassen bzw. Rechtecke existieren verschiedene Faustregeln: bei weniger als 50 Messungen ist die Klassenzahl 5 bis 7, bei mehr als 250 Messungen wählt man

10 bis 20 Klassen.² Im einfachsten Fall wird die gleiche Breite für alle Klassen gewählt, was aber nicht unbedingt der Fall sein muss.

- Dann zeichnen wir für jede Klasse einen Balken, dessen Höhe proportional zur Anzahl Beobachtungen in dieser Klasse ist.
- Dividieren wir die Anzahl Beobachtungen in einer Klasse durch die Gesamtzahl der Beobachtungen, so erhalten wir den prozentualen Anteil einer Klasse zur Gesamtbeobachtung.

```
hist(methodeA, main = "Histogramm von Methode A")
```



Bemerkungen:

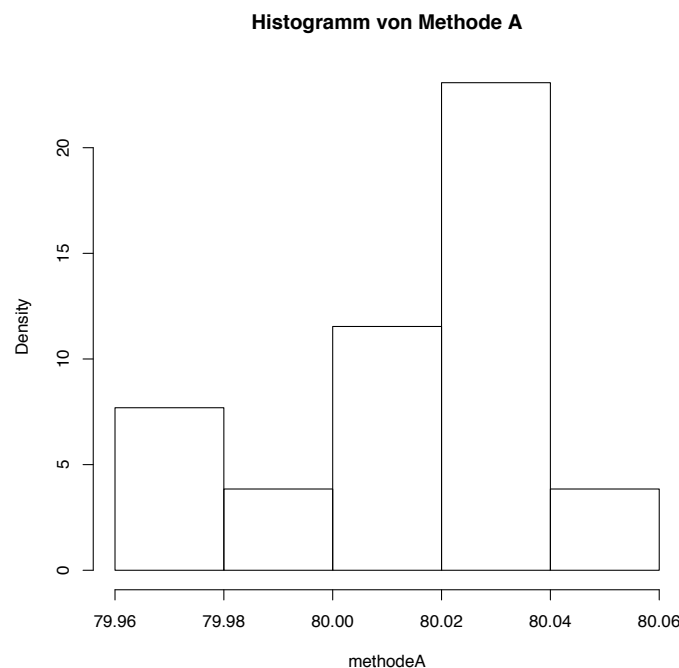
- Da die Methode A nur aus 13 Messungen besteht, wählt man 5 Balken (Sturges-Regel: $k = 1 + \log_2 13 \approx 5$).
- Bedeutung der Anzahlen (Frequency): in der 1. Klasse 79.96-79.98 sind die Beobachtungen mit den Werten 79.97 und 79.98 berücksichtigt; in der 2. Klasse 79.99 und 80.00; usw. Der linke Rand wird also nicht berücksichtigt, der rechte dagegen schon. Man hätte dies auch umgekehrt machen können, und das Histogramm würde etwas anders aussehen. Bei grossen Datensätzen spielen solche Überlegungen kaum eine Rolle.

²Gegebenenfalls kann man die Anzahl der Klassen k auch nach der Sturges-Regel berechnen: $k = 1 + \log_2 n = 1 + 3.3 \cdot \log_{10} n$, wobei n die Anzahl Messungen ist.

- iii. Mit dem R-Befehl lassen sich auch die Anzahl und die Breiten der Klassen festlegen, Überschriften ändern, usw. (siehe Übungen).

Im Histogramm oben entspricht die Höhe der Balken gerade der Anzahl der Beobachtungen in einer Klasse. Oft ist es besser und übersichtlicher, wenn wir die Balkenhöhe so wählen, dass die Balkenfläche dem prozentualen Anteil der jeweiligen Beobachtungen an der Gesamtanzahl Beobachtungen entspricht. Die Gesamtfläche aller Balken muss dann gleich eins sein.

```
hist(methodeA, freq = F, main = "Histogramm von Methode A")
```



Auf der vertikalen Achse sind nun die Dichten angegeben. Wir können also herauslesen, dass sich über $(80.04 - 80.02) \cdot 20 = 0.4$, also über 40 % der Daten zwischen 80.02 und 80.04 befinden. Die Balkenhöhe ermittelt sich, indem man die Anzahl Beobachtungen in einem Balken mit $\frac{1}{n}$ multipliziert, wobei n die Gesamtanzahl Beobachtungen bezeichnet, und diese Zahl durch die Balkenbreite dividiert. Diese Darstellung hat den Vorteil, dass man Messungen mit unterschiedlichen Umfängen besser miteinander vergleichen kann. Würde man also mit Methode A nun eine Messung mit 30 Beobachtungen durchführen, liessen sich mit Dichten besser die Verteilungen von Messwerten auf die jeweiligen Klassen vergleichen.

Boxplot

Der *Boxplot* (siehe Abbildung 2.2) besteht aus

Kapitel 2. Deskriptive Statistik

- einem Rechteck, dessen Höhe vom empirischen 25 %- und vom 75 %-Quantil begrenzt wird,
- Linien, die von diesem Rechteck bis zum kleinsten- bzw. grössten „normalen“ Wert führen (per Definition ist ein „normaler“ Wert höchstens 1.5 mal die Quartilsdifferenz von einem der beiden Quartile entfernt),
- einem horizontalen Strich für den Median,
- kleinen Kreisen, die Ausreisser markieren.

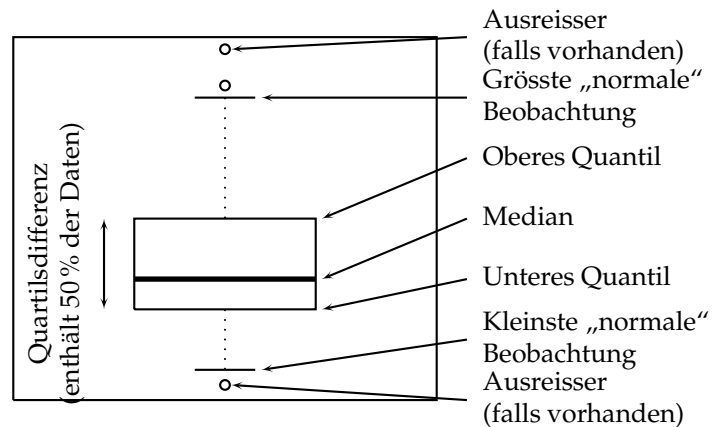
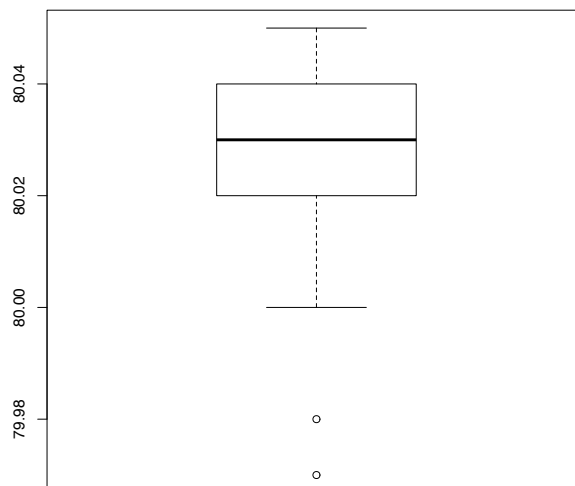


Abbildung 2.2.: Boxplot

boxplot (methodeA)



Bemerkungen:

- i. Die Hälfte der Beobachtungen befindet sich zwischen dem oberen Quartil 80.04 und dem unteren Quartil 80.02, mit Quartilsdifferenz 0.02
- ii. Der Median liegt bei 80.03.
- iii. Der „normale“ Bereich der Werte liegt zwischen 80.00 und 80.05.
- iv. Wir haben zwei Ausreisser 79.97 und 79.98.
- v. Die Punkte a) und b) hatten wir schon bei den Quantilen berechnet. Der Boxplot stellt somit unsere Berechnungen graphisch dar.

Der Boxplot ist vor allem dann geeignet, wenn man die Verteilungen der Daten in verschiedenen Gruppen (die im Allgemeinen verschiedenen Versuchsbedingungen entsprechen) vergleichen will.

Beispiel 2.1.9

Bei unserem Einführungsbeispiel der Schmelzwärme haben wir zwei Methoden zu deren Bestimmung verwendet. Somit können wir die Boxplots auch gegenüberstellen und die Methoden miteinander vergleichen (siehe Abbildung 2.3).

Bemerkungen:

- i. Methode A liefert die grösseren Werte als Methode B, da der Median von A grösser ist.
- ii. Die Daten von Methode A haben weniger Streuung als die Daten von Methode B, da das Rechteck weniger hoch ist (Quartilsdifferenz!).

□

Empirische kumulative Verteilungsfunktion

Eine weitere graphische Darstellung der Daten ist die *empirische kumulative Verteilungsfunktion*. Diese Darstellung hat den Vorteil gegenüber einem Histogramm, dass man den Median sehr leicht ablesen kann.

Die *empirische kumulative Verteilungsfunktion* $F_n(\cdot)$ ist eine Treppenfunktion, die wie folgt erzeugt wird: links von $x_{(1)}$ ist die Funktion gleich null und bei jedem $x_{(i)}$ wird ein Sprung der Höhe $1/n$ gemacht (falls ein Wert mehrmals vorkommt, ist der Sprung

```
boxplot(methodeA, methodeB, xaxt = "n", xlab = "Methode")
axis(1, at = c(1, 2), labels = c("A", "B"))
```

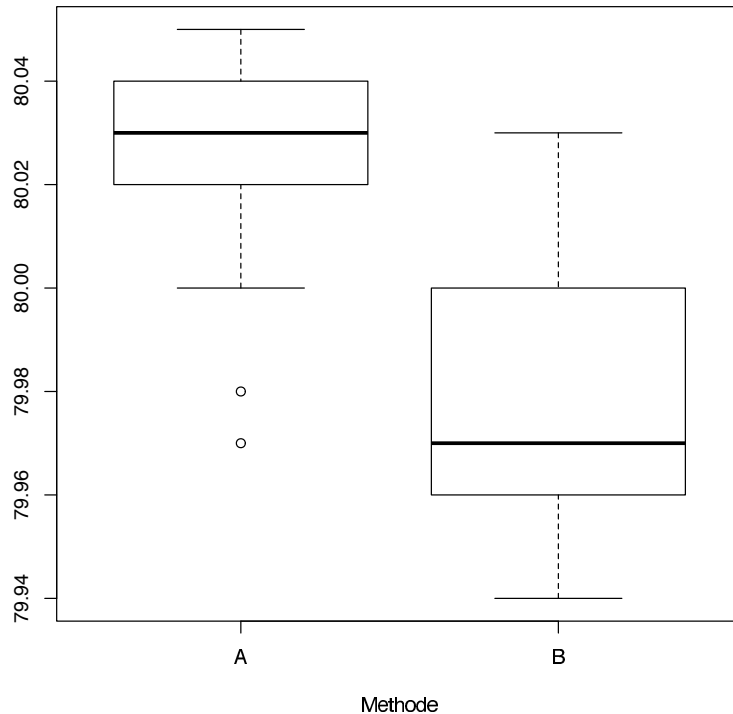


Abbildung 2.3.: Boxplots für die zwei Methoden zur Bestimmung der Schmelzwärme von Eis.

das entsprechende Vielfache von $1/n$). Im folgenden Beispiel wird dieses Vorgehen konkret durchgeführt.

Beispiel 2.1.10 Methode A der Schmelzwärme

In Abbildung 2.4 ist die empirische kumulative Verteilungsfunktion der Methode A aufgezeichnet. Sie entsteht wie folgt:

- Links von 79.97 ist die Funktion 0, da es keinen kleineren Beobachtungswert hat.
- Bei 79.97 macht die Funktion einen Sprung auf $n = 1/13 \approx 0.077$.
- Die Funktion bleibt dann gleich bis 80.00, da es vorher keinen zusätzlichen Beobachtungswert gibt. Bei 80.00 macht die Funktion wieder einen Sprung um 0.077 nach oben, weil es dort einen Messwert hat.

- Bei 80.02 macht die Funktion einen Sprung um $3 \cdot 0.077$ nach oben, da es dort 3 Beobachtungswerte gibt.
- usw.
- Bei 80.05 machen wir unseren letzten Sprung und der Funktionswert wird 1.

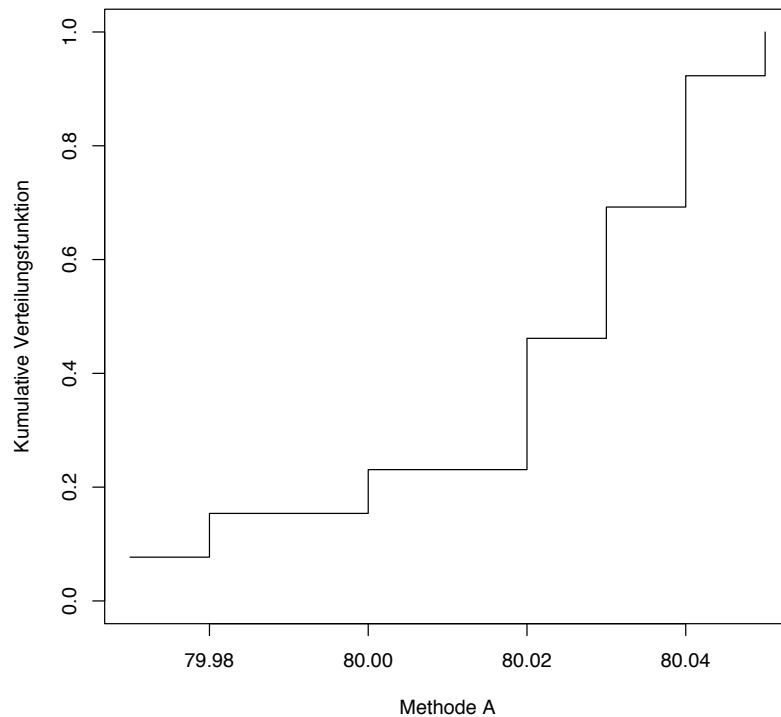


Abbildung 2.4.: Empirische kumulative Verteilungsfunktion der Messungen der Schmelzwärme von Eis mit Methode A.

Was können wir aus der kumulativen Verteilungsfunktion herauslesen?

- Bei 0.5 auf der vertikalen Achse haben wir gerade die Hälfte aller Werte aufsummiert. Zeichnen wir von 0.5 eine horizontale Linie (siehe grüne Linie in Abbildung 2.5, wird die kumulative Verteilungsfunktion bei 80.03 geschnitten. Das entspricht gerade dem Median.
- Dort, wo die kumulative Funktion steil ist, hat es auch viele Beobachtungswerte. Das heisst, die meisten Beobachtungswerte liegen hier zwischen 80.02 und 80.04. Die Werte entsprechen aber gerade dem unteren und oberen Quartil. (Man vergleiche die Funktion mit dem zugehörigen Boxplot.)

□

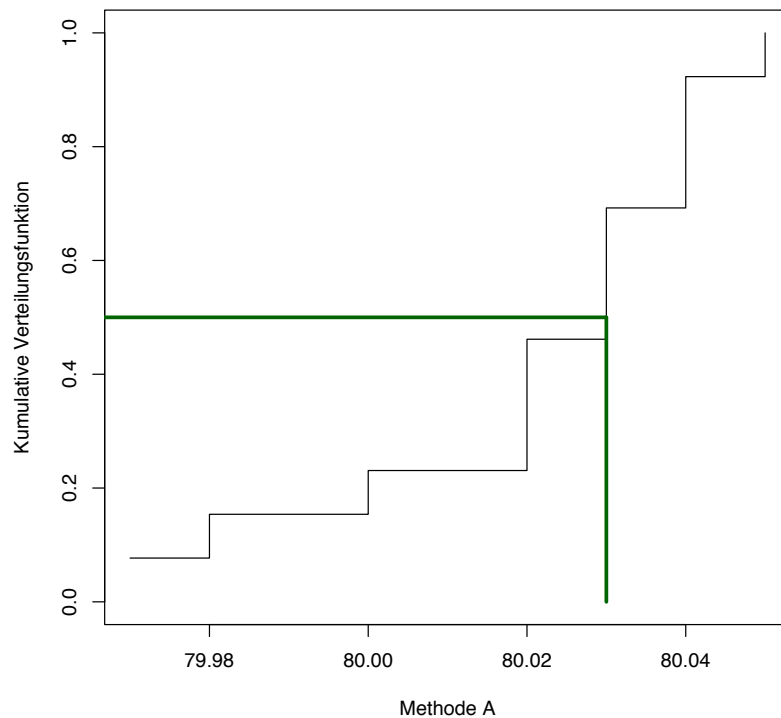


Abbildung 2.5.: Empirische kumulative Verteilungsfunktion der Messungen der Schmelzwärme von Eis mit Methode A.

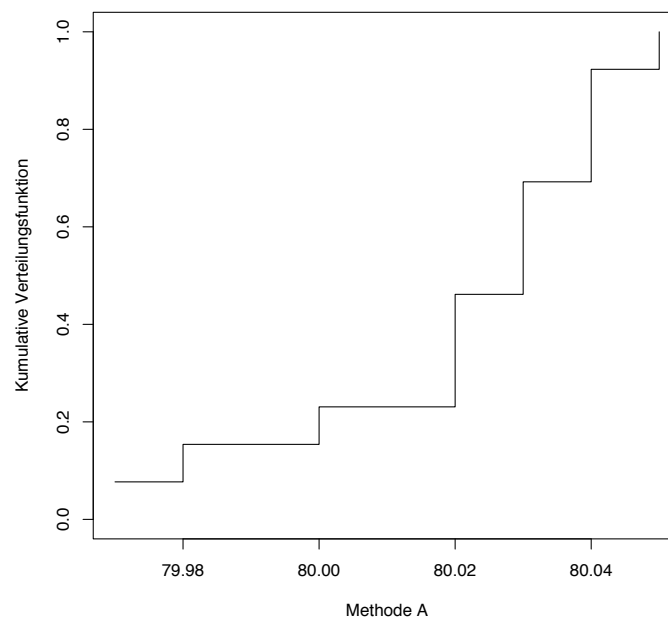
Allgemein

Die **empirische kumulative Verteilungsfunktion** ist definiert als

$$F_n(x) = \frac{1}{n} \text{Anzahl}\{i \mid x_i \leq x\}.$$

Mit R lässt sich die kumulative Verteilungsfunktion in der Abbildung 2.5 folgendermassen aufzeichnen:

```
plot(sort(methodeA), (1:length(methodeA))/length(methodeA),
     type = "s", ylim = c(0, 1), ylab = "Kumulative Verteilungsfunktion",
     xlab = "Methode A")
```



2.2. Deskriptive Statistik zweidimensionaler Daten

Bei zweidimensionalen Daten werden an einem Versuchsobjekt jeweils *zwei* verschiedene Grössen gemessen. So wird beispielsweise an einer Gruppe von Menschen jeweils die Körpergrösse *und* das Körpergewicht gemessen.

Beispiel 2.2.1 Weinkonsum und Mortalität

Wir betrachten als Beispiel einen Datensatz (siehe Tabelle 2.2), der den durchschnittlichen Weinkonsum (in Liter pro Person und Jahr) und die Sterblichkeit (Mortalität) aufgrund von Herz- und Kreislauferkrankungen (Anzahl Todesfälle pro 1000 Personen zwischen 55 und 64 Jahren pro Jahr) in 18 industrialisierten Ländern umfasst (A.S.St.Leger, A.L.Chocrane, and F.Moore, "Factors Associated with Cardiac Mortality in Developed Countries with Particular Reference to the Consumption of Wine." *Lancet*, 1979). Es stellt sich nun die Frage, ob diese Daten suggerieren, dass es einen Zusammenhang zwischen der Sterblichkeitsrate aufgrund von Herzkreislauferkrankung und Weinkonsum gibt.

Ein kurzer Blick auf die Tabelle zeigt, dass ein höherer Weinkonsum eher weniger Todesfälle wegen Herz- und Kreislauferkrankungen zur Folge hat.

□

| Land | Weinkonsum | Mortalität Herzerkrankung |
|--------------------|------------|---------------------------|
| Norwegen | 2.8 | 6.2 |
| Schottland | 3.2 | 9.0 |
| Grossbritannien | 3.2 | 7.1 |
| Irland | 3.4 | 6.8 |
| Finnland | 4.3 | 10.2 |
| Kanada | 4.9 | 7.8 |
| Vereinigte Staaten | 5.1 | 9.3 |
| Niederlande | 5.2 | 5.9 |
| New Zealand | 5.9 | 8.9 |
| Dänemark | 5.9 | 5.5 |
| Schweden | 6.6 | 7.1 |
| Australien | 8.3 | 9.1 |
| Belgien | 12.6 | 5.1 |
| Deutschland | 15.1 | 4.7 |
| Österreich | 25.1 | 4.7 |
| Schweiz | 33.1 | 3.1 |
| Italien | 75.9 | 3.2 |
| Frankreich | 75.9 | 2.1 |

Tabelle 2.2.: Weinkonsumation (Liter pro Person pro Jahr) und Mortalität aufgrund von Herz-Kreislaufkrankung (Todesfälle pro 1000) in 18 Ländern.

2.2.1. Graphische Darstellung: Streudiagramm

Ein wichtiger Schritt in der Untersuchung zweidimensionaler Daten ist die graphische Darstellung. Dies geschieht meist über ein sogenanntes *Streudiagramm* (engl.: *Scatterplot*). Dabei werden jeweils zwei Messungen als Koordinaten von Punkten in einem Koordinatensystem interpretiert und dargestellt.

In unserem Beispiel stellt ein Land eine Versuchseinheit dar, und es wird die Grösse „Weinkonsum“ x_1, \dots, x_{18} und die Grösse „Mortalität“ y_1, \dots, y_{18} gemessen. Wenn wir die Daten in der Form $(x_1, y_1), \dots, (x_{18}, y_{18})$ schreiben, interessiert man sich in erster Linie für die Zusammenhänge und Abhängigkeiten zwischen den Variablen x und y . Die Abhängigkeit zwischen den beiden Messgrössen kann man aus dem *Streudiagramm* ersehen, welches die Daten als Punkte in der Ebene darstellt: Die i -te Beobachtung (i -tes Land) entspricht dem Punkt mit Koordinaten (x_i, y_i) . Die Abbildung 2.6 zeigt das Streudiagramm für die Messgrössen „Weinkonsum“ $(x_1, x_2, \dots, x_{18})$ und „Mortalität“ $(y_1, y_2, \dots, y_{18})$. Man sieht einen klaren monoton fallenden Zusammenhang: Länder mit hohem Weinkonsum haben also eine Tendenz zu einer tieferen Mortalitätsrate wegen Herz- und Kreislaufkrankheiten.

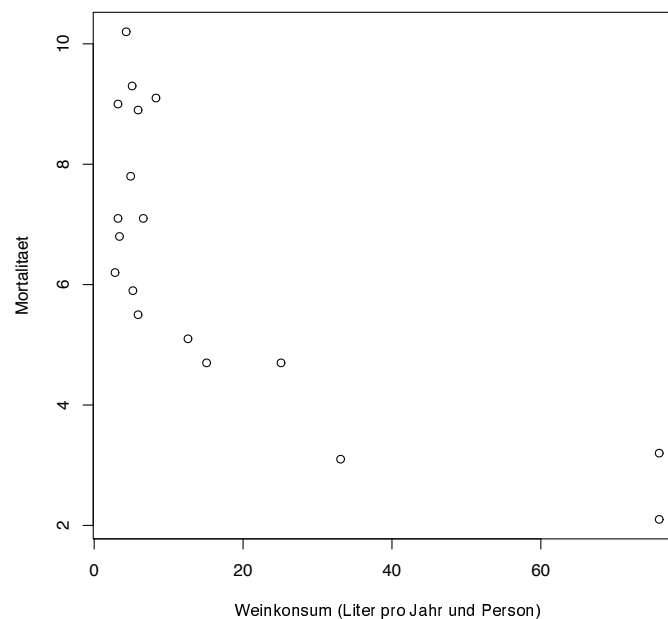


Abbildung 2.6.: Streudiagramm für die Mortalität und den Weinkonsum in 18 industrialisierten Ländern.

Das Streudiagramm in Abbildung 2.6 wurde mit R erstellt.

```
wein <- c(2.8, 3.2, 3.2, 3.4, 4.3, 4.9, 5.1, 5.2, 5.9, 5.9,
          6.6, 8.3, 12.6, 15.1, 25.1, 33.1, 75.9, 75.9)
mort <- c(6.2, 9, 7.1, 6.8, 10.2, 7.8, 9.3, 5.9, 8.9, 5.5,
          7.1, 9.1, 5.1, 4.7, 4.7, 3.1, 3.2, 2.1)

plot(wein, mort, xlab = "Weinkonsum (Liter pro Jahr und Person)",
      ylab = "Mortalitaet")
```

Bemerkungen:

- i. Der Schluss, dass hoher Weinkonsum gesund ist, ist *falsch*. Es *scheint*, dass höherer Weinkonsum zu weniger Toten wegen Herz- und Kreislauferkrankungen führt. Der Einfluss des höheren Weinkonsums auf andere Körperorgane (z.B. Leber) oder auf die Anzahl Verkehrsunfälle, wird hier *nicht* untersucht.
- ii. Aus den Zahlwerten und Scatterplot lässt sich ein Zusammenhang *erahnen*, aber es muss keinen *kausalen* Zusammenhang zwischen den Messreihen haben. Das Streudiagramm kann zufällig so aussehen, also muss der Weinkonsum keinen Einfluss auf die Sterblichkeit haben.

2.2.2. Einfache lineare Regression

Wir haben im vorangehenden Beispiel eine negative (je mehr desto weniger) Abhängigkeit zwischen Mortalität und Weinkonsum festgestellt. Oft ist diese Abhängigkeit sehr einfach, nämlich *linear*.

Beispiel 2.2.2 Zusammenhang Seitenzahl-Preis eines Buches

Wir erklären das Modell der einfachen linearen Regression zunächst mit einem fiktiven Beispiel. Je dicker ein Roman (Hardcover) ist, desto teurer ist er in der Regel. Es gibt also einen Zusammenhang zwischen Seitenzahl x und Buchpreis y . Wir gehen in einen Buchladen und suchen zehn Romane verschiedener Dicken aus. Wir nehmen dabei je ein Buch mit der Seitenzahl 50, 100, 150, ..., 450, 500. Von jedem Buch notieren wir die Seitenzahl und den Buchpreis. Mit diesen Daten erstellen wir Tabelle 2.3. Aus der Tabelle ist tatsächlich ersichtlich, dass dickere Bücher tendenziell mehr

| | Seitenzahl | Buchpreis (SFr) |
|---------|------------|-----------------|
| Buch 1 | 50 | 6.4 |
| Buch 2 | 100 | 9.5 |
| Buch 3 | 150 | 15.6 |
| Buch 4 | 200 | 15.1 |
| Buch 5 | 250 | 17.8 |
| Buch 6 | 300 | 23.4 |
| Buch 7 | 350 | 23.4 |
| Buch 8 | 400 | 22.5 |
| Buch 9 | 450 | 26.1 |
| Buch 10 | 500 | 29.1 |

Tabelle 2.3.: Zusammenhang zwischen Buchpreis und Seitenzahl (fiktiv).

kosten. Wenn wir einen formelmässigen Zusammenhang zwischen Buchpreis und Seitenzahl hätten, könnten wir Vorhersagen über den Preis für Bücher mit Seitenzahlen machen, die wir nicht beobachtet haben. Was würde dann voraussichtlich ein Buch mit 375 Seiten kosten? Oder wir könnten herausfinden, wie teuer ein Buch mit „null“ Seiten wäre. Das wären die Grundkosten des Verlags, die unabhängig von der Seitenzahl anfallen: Einband, administrativer Aufwand für jedes Buch, etc. Wie könnten wir diesen Zusammenhang mit einer Formel beschreiben? Das Streudiagramm in Abbildung 2.7 zeigt diesen Zusammenhang graphisch deutlicher auf.

Auf den ersten Blick scheint eine Gerade recht gut zu den Daten zu passen. Diese Gerade hätte die Form

$$y = a + bx$$

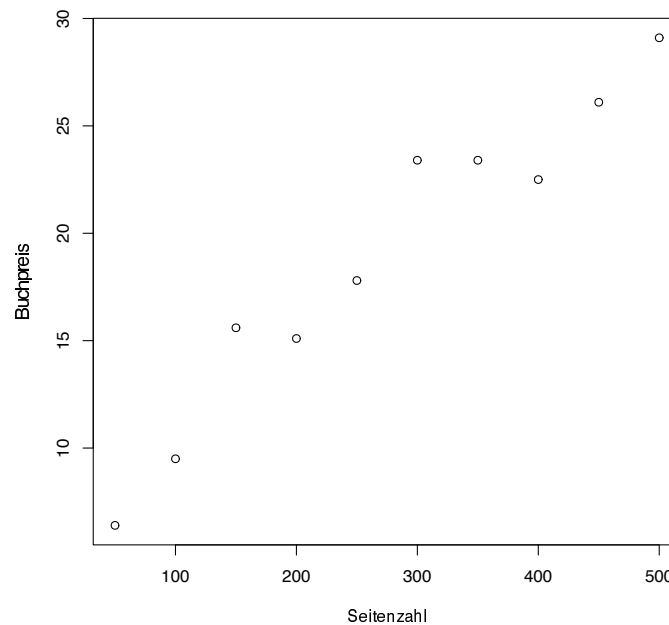


Abbildung 2.7.: Streudiagramm Seitenzahl - Buchpreis

mit y dem Buchpreis und x der Seitenzahl sind. Der Parameter a beschreibt dann die Grundkosten des Verlags und der Parameter b entspricht den Kosten pro Seite.

□

Methode der kleinsten Quadrate

Versuchen wir mit einem Lineal eine Gerade durch *alle* Punkte in Abbildung 2.7 zu legen, so werden wir feststellen, dass das nicht möglich ist (siehe Abbildung 2.8). Die Punkte folgen also nur *ungefähr* einer Geraden.

Da stellt sich uns die Frage: „Wie können wir eine Gerade finden, die *möglichst gut* zu allen Punkten passt?“ Damit stellt sich die nächste Frage: Was heisst „möglichst gut“? Hier gibt es verschiedene Möglichkeiten. Es treten ähnliche Schwierigkeiten auf, wie bei der Bestimmung der Varianz. Die naheliegendste Möglichkeit ist nicht die optimalste.

Wir könnten die Gerade so wählen, dass wir die vertikalen Differenzen zwischen Beobachtung und Gerade (siehe Abbildung 2.9) zusammenzählen und davon ausgehen, dass eine kleine Summe der Abstände eine gute Anpassung bedeutet.

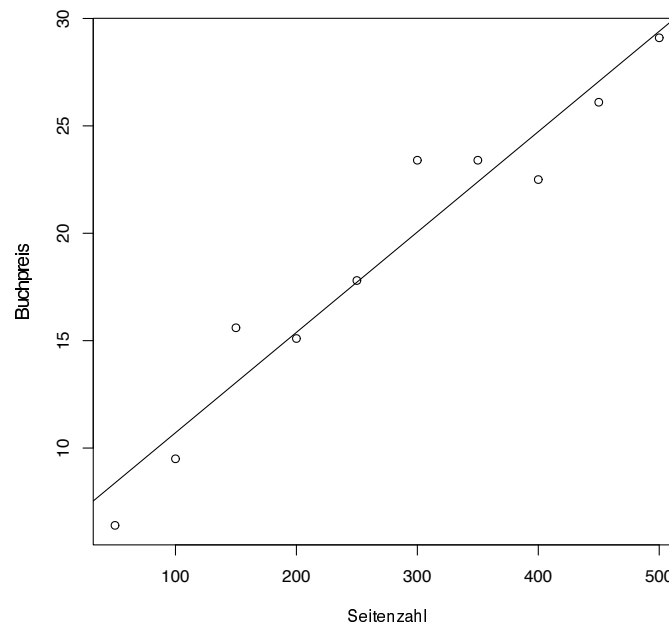


Abbildung 2.8.: Gerade durch das Streudiagramm

Wir bezeichnen die vertikale Differenz zwischen einem Beobachtungspunkt (x_i, y_i) und der Geraden (der Punkt auf der Geraden hat die Koordinaten $(x_i, a + bx_i)$) als **Residuum**:

$$r_i = y_i - (a + bx_i) = y_i - a - bx_i$$

Für unser Beispiel sind die Residuen r_6 und r_8 für *diese* Gerade in Abbildung 2.9 dargestellt. Das Residuum r_6 ist positiv, da der Punkt oberhalb der Geraden liegt. Entsprechend ist $r_8 < 0$.

Wir möchten also die Gerade $y = a + bx$ so bestimmen, dass die Summe

$$r_1 + r_2 + \dots + r_n = \sum_i r_i$$

minimal wird. Diese Methode hat aber eine gravierende Schwäche: Wenn die Hälfte der Punkte weit über der Geraden, die andere Hälfte weit unter der Geraden liegen, so ist die Summe der Abweichungen (Residuen) etwa null. Dabei passt die Gerade gar nicht gut zu den Datenpunkten. Die positiven Abweichungen heben sich nur mit den negativen Abweichungen auf. Wir müssen also das Vorzeichen der Abweichungen eliminieren, bevor wir zusammenzählen. Eine Möglichkeit besteht darin, den Absolutbetrag der Abweichungen aufzusummieren, also $\sum_i |r_i|$, und diese Summe

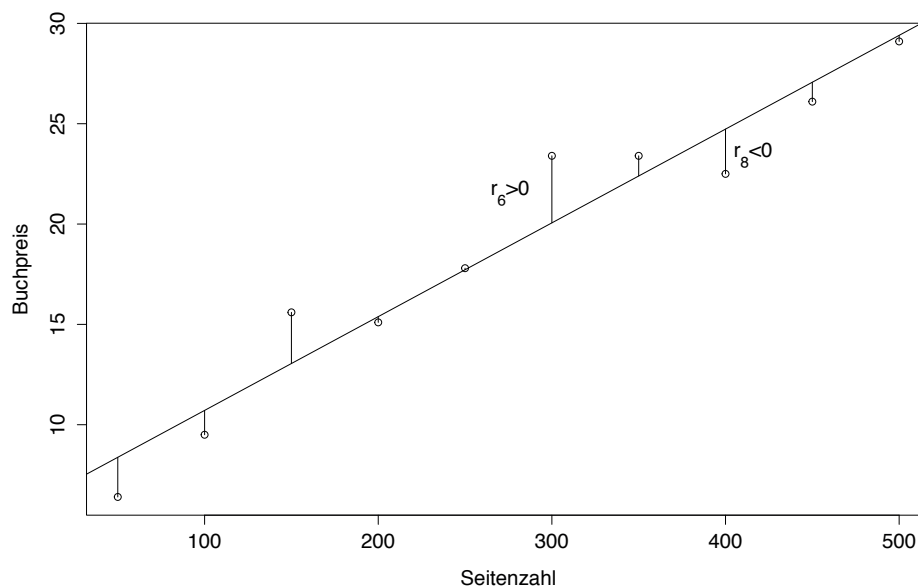


Abbildung 2.9.: Residuen

zu minimieren. Da es sich aber mit Absolutbeträgen nicht besonders bequem rechnen lässt (zum Beispiel, wenn man Ausdrücke mit Absolutbeträgen ableiten möchte), halten wir nach einer anderen Möglichkeit Ausschau. Die besteht darin, die Quadrate der Abweichungen aufzusummieren, also

$$r_1^2 + r_2^2 + \dots + r_n^2 = \sum_i r_i^2$$

Die Parameter a und b sind so zu wählen, dass diese Summe minimal wird. Letztere Methode hat sich durchgesetzt, weil man mit ihr viel leichter rechnen kann, als mit den Absolutbeträgen. Eine Gerade passt (nach unserem Gütekriterium) also dann am besten zu Punkten, wenn die Summe der Quadrate der vertikalen Abweichungen minimal ist. Dieses Vorgehen ist unter dem Namen **Methode der kleinsten Quadrate** bekannt. In unserem Fall erhalten wir mit R die Werte $a = 6.04$ und $b = 0.047$.

```
seitenzahl <- c(seq(50, 500, 50))

buchpreis <- c(6.4, 9.5, 15.6, 15.1, 17.8, 23.4, 23.4, 22.5,
              26.1, 29.1)

lm(buchpreis ~ seitenzahl)

##
## Call:
```

```
## lm(formula = buchpreis ~ seitenzahl)
##
## Coefficients:
## (Intercept)    seitenzahl
##      6.04000      0.04673
```

Die Geradengleichung lautet

$$y = 6.04 + 0.04673x$$

Die Grundkosten des Verlags sind also rund 6 SFr. Pro Seite verlangt der Verlag rund 5 Rappen.

Bemerkungen:

- i. Der Befehl `lm()` steht für „linear model“.
- ii. Mit dem Befehl `lm(y~x)` passt R ein Modell von der Form $y = a + bx$ an die Daten an.
- iii. Diese Gerade wird auch *Regressionsgerade* genannt.

Beispiel 2.2.3

Wie viel würde nach diesem Modell ein Buch von 375 Seiten kosten? Dazu setzen wir $x = 375$ in die Geradengleichung oben ein und erhalten

$$y = 6.04 + 0.04673 \cdot 375 \approx 23.60$$

Das Buch dürfte also etwa CHF. 23.60 kosten. Dieses Modell ist allerdings nur begrenzt gültig. Vor allem bei *Extrapolationen* muss man vorsichtig sein. Wir könnten schon ausrechnen, wie viel ein Buch mit einer Million Seiten kostet, aber dieser Betrag entspricht dann sicher nicht mehr der Realität.

□

Diese Gerade in Abbildung 2.8 auf Seite 34 wird in R wie folgt gezeichnet:

```
seite <- c(seq(50, 500, 50))

preis <- c(6.4, 9.5, 15.6, 15.1, 17.8, 23.4, 23.4, 22.5,
          26.1, 29.1)

plot(seite, preis, xlab = "Seitenzahl", ylab = "Buchpreis")

abline(lm(preis ~ seite))
```

Wie berechnet R die Parameter a und b ?

Die Parameter a und b werden wie folgt bestimmt:

Die Parameter a und b sollen den folgenden Ausdruck minimieren (Methode der Kleinsten-Quadrate)

$$\sum_{i=1}^n (y_i - (a + bx_i))^2$$

Die Lösung dieses Optimierungsproblems ergibt

$$\hat{b} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b} \bar{x}$$

wobei \bar{x} und \bar{y} die entsprechenden Durchschnitte sind. \hat{a} und \hat{b} sind die Schätzer von den Parametern a und b , also die Werte, für welche $\sum_{i=1}^n (y_i - (a + bx_i))^2$ am kleinsten wird.

Bemerkungen:

- i. Wie man auf die Berechnung von a und b kommt, leiten wir hier nicht her. Nur soviel zur Idee: da

$$\sum_i r_i^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2$$

minimal werden muss, muss auch die Ableitung von $\sum_i r_i^2$ nach a und nach b gleich 0 sein. Wir erhalten also ein Gleichungssystem bestehend aus zwei Gleichungen und zwei Unbekannten:

$$\frac{\partial}{\partial a} \sum_{i=1}^n (y_i - (a + bx_i))^2 = \sum_{i=1}^n 2(y_i - a - bx_i) \stackrel{!}{=} 0$$

$$\frac{\partial}{\partial b} \sum_{i=1}^n (y_i - (a + bx_i))^2 = \sum_{i=1}^n 2(y_i - a - bx_i) \cdot x_i \stackrel{!}{=} 0$$

Die algebraischen Umformungen, die zu den Schätzern von a und b führen, sind dann etwas mühsam.

- ii. Die Berechnungen von \hat{a} und \hat{b} werden wir immer mit R berechnen.

Beispiel 2.2.4

Es ist zu vermuten, dass es einen Zusammenhang zwischen der Körpergrösse der Väter und der Körpergrösse der Söhne gibt. Der britische Statistiker Karl Pearson trug dazu um 1900 die Körpergrösse von 10 (in Wahrheit waren 1078) zufällig ausgewählten Männern gegen die Grösse ihrer Väter auf. Dabei erhielt er die Daten von Tabelle 2.4.

| | | | | | | | | | | |
|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Grösse des Vaters | 152 | 157 | 163 | 165 | 168 | 170 | 173 | 178 | 183 | 188 |
| Grösse des Sohnes | 162 | 166 | 168 | 166 | 170 | 170 | 171 | 173 | 178 | 178 |

Tabelle 2.4.: Grössenvergleich von Vätern und Söhnen

Es *scheint* hier tatsächlich einen Zusammenhang zu geben: je grösser der Vater, desto grösser der Sohn. Wenn wir noch das Streudiagramm aufzeichnen (siehe Abbildung 2.10), sehen wir, dass ein (möglicher) linearer Zusammenhang besteht: Die Punktwolke „folgt“ der Geraden $y = 0.445x + 94.7$, wobei wir die Parameter mit der Methode der Kleinsten Quadrate aus den Daten berechnet haben.

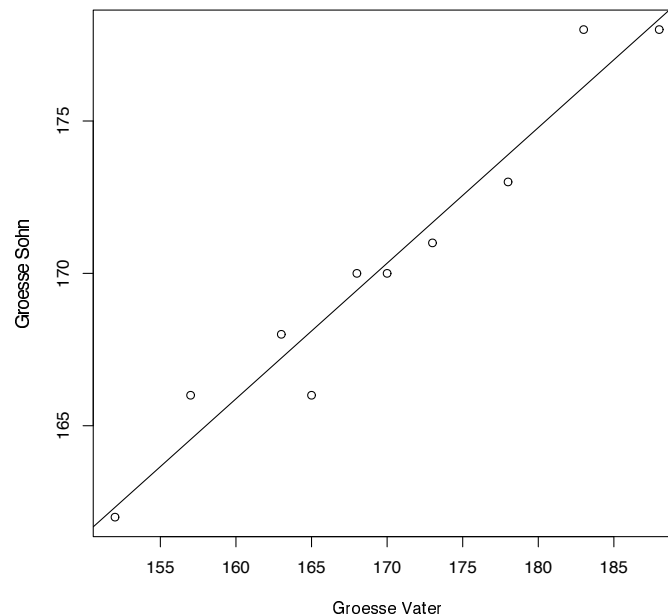


Abbildung 2.10.: Streudiagramm Körpergrössen Väter-Söhne

Wir können also für die in der Tabelle 2.4 nicht vorkommende Grösse von 180 cm des Vaters den zu erwartenden Wert für die Grösse seines Sohnes berechnen.

$$y = 0.445 \cdot 180 + 94.7 \approx 175 \text{ cm}$$

Wir müssen bei dieser Formel allerdings aufpassen, dass wir sie nicht dort anwenden, wo wir sie gar nicht dürfen. So erhalten wir für $x = 0$ einen Wert von 94.7. Was bedeutet dies aber? Wenn der Vater 0 cm gross ist, so wäre der Sohn gemäss dieses Modells ungefähr 95 cm gross, und das macht keinen Sinn.

□

Beispiel 2.2.5

Die folgende Tabelle stellt einen Zusammenhang zwischen den Zahlen der Verkehrstoten her, die es 1988 und 1989 in zwölf Bezirken in den USA geben hat.

| Bezirk | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-------------------|-----|----|-----|-----|-----|-----|----|-----|-----|-----|----|-----|
| Verkehrstote 1988 | 121 | 96 | 85 | 113 | 102 | 118 | 90 | 84 | 107 | 112 | 95 | 101 |
| Verkehrstote 1989 | 104 | 91 | 101 | 110 | 117 | 108 | 96 | 102 | 114 | 96 | 88 | 106 |

Tabelle 2.5.: Verkehrstoten in aufeinanderfolgenden Jahren

Aus der Tabelle ist kein offensichtlicher Zusammenhang ersichtlich. Betrachten wir das Streudiagramm in Abbildung 2.11, so sehen wir, dass kein Zusammenhang besteht. Dies war aber auch zu erwarten, wenn wir vernünftigerweise davon ausgehen können, dass es zwischen den Verkehrstoten der einzelnen Bezirke keinen Zusammenhang gibt. In Abbildung 2.11 ist die Regressionsgerade eingezeichnet. Diese können wir zwar berechnen und einzeichnen. Allerdings macht diese hier keinen Sinn, da es keinen linearen Zusammenhang zwischen den Messgrössen gibt.

Wir werden im nächsten Abschnitt eine Grösse kennenlernen, mit der wir eine Aussage treffen können, wie stark der lineare Zusammenhang zwischen Messgrössen ist.

□

Beispiel 2.2.6

Als weiteres Beispiel betrachten wir wieder die Erhebung, die den Zusammenhang zwischen Weinkonsum und der Sterblichkeit untersucht. Legen wir den Daten ein lineares Modell zu Grunde

$$y = a + bx$$

wobei x den jährlichen Weinkonsum pro Person und y die Mortalität pro 1000 Personen bezeichnet. Dann können wir aufgrund der Datenpunkte die Parameter a und

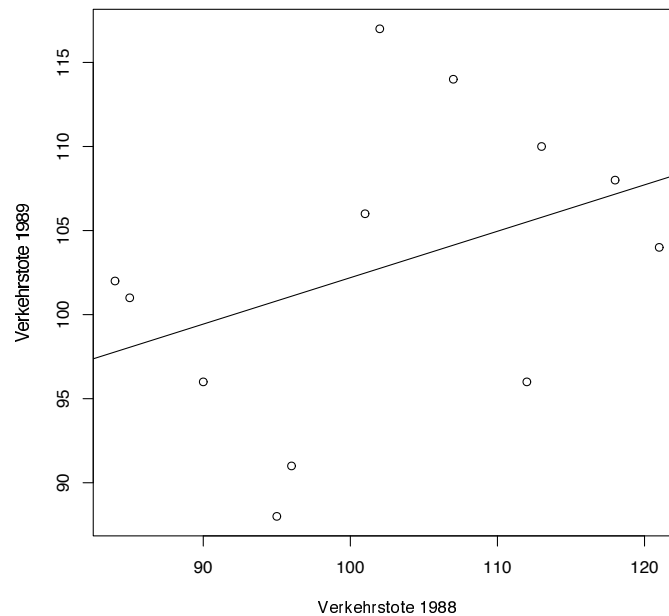


Abbildung 2.11.: Verkehrstote

b mit Hilfe der Methode der Kleinsten Quadrate schätzen und erhalten die Regressionsgerade

$$y = 7.68655 - 0.07608x$$

Betrachten wir allerdings das Streudiagramm mit der Regressionsgerade (siehe Abbildung 2.12), so stellen wir fest, dass der Zusammenhang zwischen den Messgrößen nicht linear ist. Das Streudiagramm deutet eher auf eine Hyperbelstruktur hin.

Die Regressionsgerade sagt hier also wenig über den wahren Zusammenhang aus.

□

Die Regressionsgerade können wir (fast) immer bestimmen. In den letzten beiden Beispielen haben wir aber gesehen, dass die Regressionsgerade sehr wenig über die wirkliche Verteilung der Punkte im Streudiagramm aussagt. Dafür gibt es zwei Gründe

- Die Punkte folgen scheinbar gar keiner Gesetzmässigkeit
- Die Punkte folgen einer nichtlinearen Gesetzmässigkeit

Wie können wir nun aber feststellen, ob ein linearer Zusammenhang der Daten besteht oder nicht? Eine Möglichkeit ist sicher, die Situation graphisch zu betrachten,

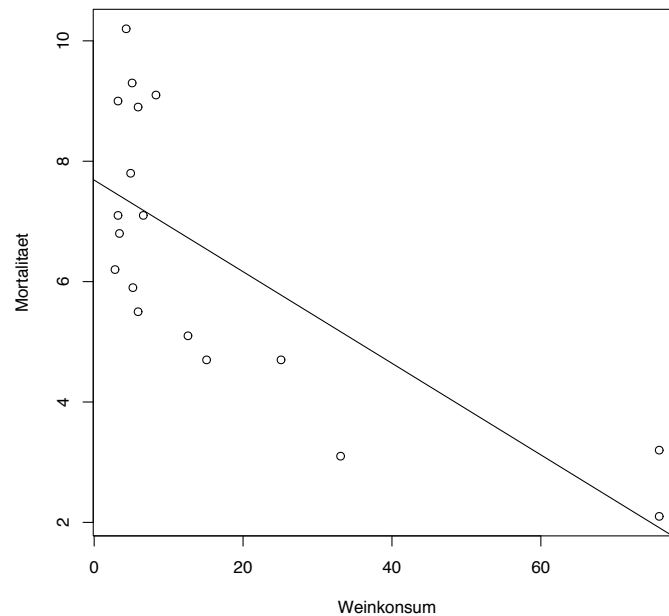


Abbildung 2.12.: Regressionsgerade Weinkonsum-Sterblichkeit

wie wir das eben gemacht haben. Wir können aber auch einen Wert angeben, der den Zusammenhang numerisch beschreibt.

2.2.3. Empirische Korrelation

Für die numerische Zusammenfassung der linearen Abhängigkeit von zwei Größen ist die **empirische Korrelation** r als Kennzahl (oder auch mit $\hat{\rho}$ bezeichnet) am gebräuchlichsten.

Empirische Korrelation

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2) \cdot (\sum_{i=1}^n (y_i - \bar{y})^2)}}$$

Die empirische Korrelation ist eine dimensionslose Zahl zwischen -1 und $+1$ und misst Stärke und Richtung der *linearen Abhängigkeit* zwischen den Daten x und y . Die empirische Korrelation hat folgende Eigenschaften

1. Ist $r = +1$, dann liegen die Punkte auf einer steigenden Geraden ($y = a + bx$ mit $a \in \mathbb{R}$ und ein $b > 0$) und umgekehrt.

2. Ist $r = -1$, dann liegen die Punkte auf einer fallenden Geraden ($y = a + bx$ mit $a \in \mathbb{R}$ und ein $b < 0$) und umgekehrt.
3. Sind x und y unabhängig (d.h. es besteht kein Zusammenhang), so ist $r = 0$.

Die Umkehrung gilt im allgemeinen nicht: $r = 0$ heisst *nicht*, dass x und y unabhängig voneinander sind (siehe Abbildung 2.13 auf Seite 44)

Die ersten beiden Eigenschaften lassen sich einfach nachvollziehen: man setzt im Ausdruck für den Korrelationskoeffizienten $y_i = a + bx_i$ und $\bar{y} = a + b\bar{x}$ ein. Dann ergibt sich

$$\begin{aligned} r &= \frac{\sum_{i=1}^n (x_i - \bar{x})(bx_i + a - (b\bar{x} + a))}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2) \cdot (\sum_{i=1}^n (bx_i + a - (b\bar{x} + a))^2)}} \\ &= \frac{b \cdot \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{|b| \cdot \sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2) \cdot (\sum_{i=1}^n (x_i - \bar{x})^2)}} \\ &= \text{sign}(b), \end{aligned}$$

wobei $\text{sign}(b)$ das Vorzeichen von b bezeichnet, also $\text{sign}(b) = 1$ falls $b > 0$ und $\text{sign}(b) = -1$ falls $b < 0$.

Man sollte jedoch nie r berechnen, ohne einen Blick auf das Streudiagramm zu werfen, da ganz verschiedene Strukturen den gleichen Wert von r ergeben können. Siehe dazu Abbildung 2.13 auf Seite 44.

Beispiel 2.2.7

Für unser Seitenzahl-Preis-Beispiel erhalten wir mit R

```
cor(seitenzahl, buchpreis)

## [1] 0.9681122
```

Der Wert liegt also sehr nahe bei 1 und somit besteht ein enger linearer Zusammenhang. Dazu ist der Wert positiv, was einem „je mehr, desto mehr“, also einem positiven linearen Zusammenhang entspricht.

□

Beispiel 2.2.8

Auch im Beispiel der Körpergrösse von Vater und Sohn erwarten wir einen hohen Korrelationskoeffizienten. Wir erhalten 0.973.



Beispiel 2.2.9

Bei den Verkehrsunfällen haben wir keinen Zusammenhang und erwarten einen Korrelationskoeffizienten nahe null. Er beträgt 0.386.



Beispiel 2.2.10

Auch beim Weinkonsum erwarten wir einen negativen Korrelationskoeffizienten, da mit steigendem Weinkonsum die Mortalität sinkt und der nahe bei null liegt. Er beträgt -0.746 . Ohne die Daten in einem Streudiagramm darzustellen, würde man aufgrund dieses Wertes fälschlicherweise auf einen starken negativen linearen Zusammenhang schliessen.



Beispiel 2.2.11

In Abbildung 2.13 sind 21 verschiedene Datensätze dargestellt, die je aus gleich vielen Beobachtungspaaren (x_i, y_i) mit den entsprechenden Punkten im Streudiagramm bestehen. Über jedem Datensatz steht jeweils die zugehörige empirische Korrelation.

Bei perfektem linearen Zusammenhang ist die empirische Korrelation $+1$ oder -1 (je nachdem ob die Steigung positiv oder negativ; siehe zweite Zeile in Abbildung 2.13). Je mehr die Punkte um den linearen Zusammenhang streuen, desto kleiner wird der Betrag der empirischen Korrelation (siehe erste Zeile).

Da die empirische Korrelation nur den *linearen* Zusammenhang misst, kann es einen (nichtlinearen) Zusammenhang zwischen den beiden Variablen x und y geben, auch wenn die empirische Korrelation null ist (siehe unterste Zeile in Abbildung 2.13).



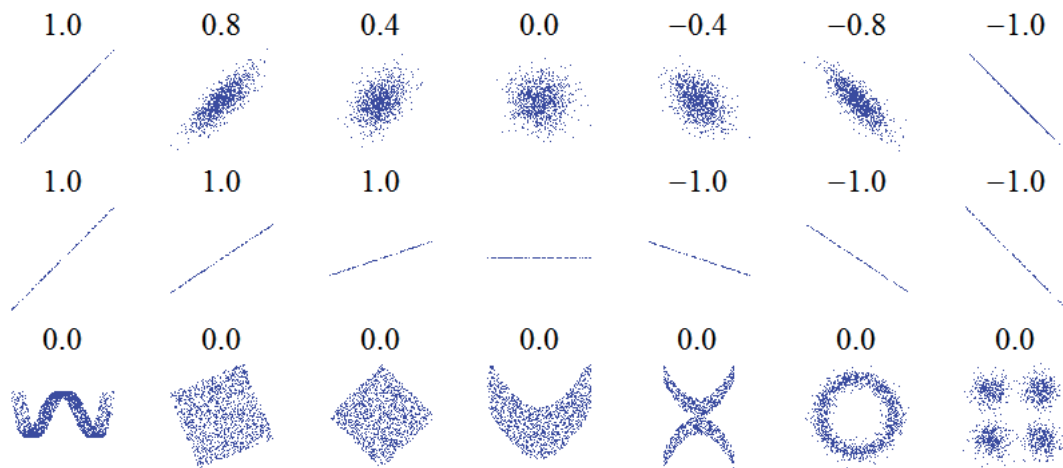


Abbildung 2.13.: 21 verschiedene Datensätze und deren empirische Korrelationskoeffizienten.

Lernziele

- ☐ Sie kennen Methoden der deskriptiven Statistik, können diese interpretieren und folgende Größen ausrechnen: arithmetisches Mittel, Standardabweichung, Varianz, Quantil, Median und Korrelation.
- ☐ Sie verstehen die Grundidee der einfachen linearen Regression: wie die Form des Modells ist, wie man die Koeffizienten interpretiert und wie man man die Koeffizienten schätzt.
- ☐ Sie können Daten mit folgenden graphischen Methoden darstellen: Histogramm, Boxplot, empirische kumulative Verteilungsfunktion, Streudiagramm.

Kapitel 3.

Modelle für Zähldaten

Everybody speaks of probability,
but no one is able to say what
it is, in a way which is
satisfactory for others.

(Garrett Birkhoff)

3.1. Einführung

In diesem Kapitel werden wir uns mit Modellen von Zähldaten beschäftigen. Von Zähldaten sprechen wir, wenn der Ausgang einer Messung eine *natürliche* Zahl ergibt. Als Beispiel betrachten wir folgende Fragestellungen:

- Wir werfen einen Würfel 300 mal. Wie viele Male kommt die Augenzahl 4 vor?
- Wir wählen zufällig 100 männliche Personen aus. Wie viele Personen haben eine Körperlänge von mehr als 173 cm?

In diesen Fällen wird die Messung einmal durchgeführt. Führt man die Messung ein zweites Mal durch, so werden die Resultate verschieden zu jener der ersten Messung sein. Sind im zweiten Beispiel das erste Mal zufällig sehr viele Asiaten dabei und bei einer zweiten Messung zufällig viele Basketballspieler, so wird es sicherlich Unterschiede in der gemessenen Anzahl von Personen mit Körpergrößen von mehr als 173 cm geben. Wird aber die Messung aus dem zweiten Beispiel 1000 Mal wiederholt, dürfte sich herauskristallisieren, was die zu *erwartende* Anzahl Personen mit einer Körpergröße von mehr als 173 cm ist.

Können wir aber theoretisch vorhersagen, wie gross der Wert einer zukünftigen Messung wahrscheinlich sein wird? Um eine solche Vorhersage machen zu können, brauchen wir *Modelle* für die Messungen. Nehmen wir an, wir würfeln mit einem „fairen“ Würfel. Dann gehen wir davon aus, dass die Wahrscheinlichkeit eine 4 zu würfeln, genau $\frac{1}{6}$ ist. Dies ist allerdings nur ein *Modell*. Wenn wir den Würfel 1.2 Million mal werfen, dann wird die Anzahl der auftretenden 4 wohl nie *genau* 200 000 sein. Es ist

durchaus möglich, wenn auch sehr unwahrscheinlich, dass wir in all diesen Würfeln nie eine 4 würfeln. Wir haben also eine - in diesem Fall vernünftige - Idealisierung gemacht. Wir sprechen auch von einer *Modellierung* der Problemstellung. Die Vorhersagen, die wir machen können, sind aber mit Unsicherheiten behaftet, und wir können nur sagen, mit welcher *Wahrscheinlichkeit* ein Resultat eintritt.

Es folgt noch ein physikalisches Beispiel dieser Überlegungen, das uns auch als Standardbeispiel dienen soll.

Der Alpha-Zerfall

Ein radioaktiver Atomkern, der beim Zerfall Helium-4-Atomkerne aussendet, wird als *Alphastrahler* bezeichnet. Der vom zerfallenden Atomkern emittierte Helium-4-Atomkern wird *Alphateilchen* genannt. Diese Alphateilchen kann man messen. Die Emission von Alphateilchen von einer radioaktiven Quelle pro Zeiteinheit ist *nicht konstant*, sondern fluktuiert auf *zufällige* Art und Weise. Der Strahler kann also in einem 10-Sekundenintervall 10 Alphateilchen aussenden, dann in den nächsten 10 Sekunden gar keine und dann wieder 100. Ein typischer Alphastrahler ist Americium 241, ein Zerfallsprodukt von Plutonium-241, das in radioaktiven Abfällen häufig anzutreffen ist.

| Anzahl Zerfälle | Beobachtet in Anzahl Experimente |
|-----------------|----------------------------------|
| 0-2 | 18 |
| 3 | 28 |
| 4 | 56 |
| 5 | 105 |
| 6 | 126 |
| 7 | 146 |
| 8 | 164 |
| 9 | 161 |
| 10 | 123 |
| 11 | 101 |
| 12 | 74 |
| 13 | 53 |
| 14 | 23 |
| 15 | 15 |
| 16 | 9 |
| 17+ | 5 |

Tabelle 3.1.: Anzahl Zerfälle in 10 Sekunden und Anzahl Experimente (von insgesamt 1207), in denen diese Anzahl Zerfälle beobachtet wurde

In einem Experiment wurde nun die Anzahl Zerfälle von Americium 241 in einem Intervall von 10 Sekunden gemessen. Das Experiment wurde 1207 Mal wiederholt,

jedes Mal wurde die Anzahl Zerfälle in 10 Sekunden gemessen. In Tabelle 3.1 ist in der ersten Spalte die Anzahl Zerfälle aufgeführt, in der zweiten Spalte, wie oft diese Anzahl Zerfälle in den 1207 Experimenten beobachtet worden ist. So sind in 18 der 1207 Experimente 0, 1 oder 2 Alphateilchen gemessen worden. In 28 der 1207 Experimente wurden 3 Alphateilchen gemessen.

Wie viele Zerfälle sich in einem bestimmten Zeitintervall ereignen werden, kann mit einem physikalischen Gesetz nicht vorausgesagt werden. Wir können aber versuchen, die Wahrscheinlichkeit für jede Anzahl Zerfälle im Zeitintervall von 10 Sekunden anzugeben. Wir sind also bestrebt, ein Wahrscheinlichkeitsmodell für den Alpha-Zerfall zu *konstruieren*.

Es stellt sich nun die Frage, wie wir als erstes die Zerfallswahrscheinlichkeiten aus den Daten *schätzen* können. Wie gross ist also beispielsweise die Wahrscheinlichkeit, dass wir in 10 Sekunden 12 Zerfälle messen? Wir stellen dazu die Häufigkeiten der Anzahl Zerfälle graphisch dar. Dazu betrachten wir das Histogramm unserer Daten (siehe Abbildung 3.1 links). Wir erhalten dabei eine sogenannte *Häufigkeitsverteilung*. Daraus können wir zum Beispiel ablesen, dass in 74 Experimenten 12 Zerfälle beobachtet wurden. Wir sprechen in diesem Fall von *absoluten Häufigkeiten*. Diese Zahl

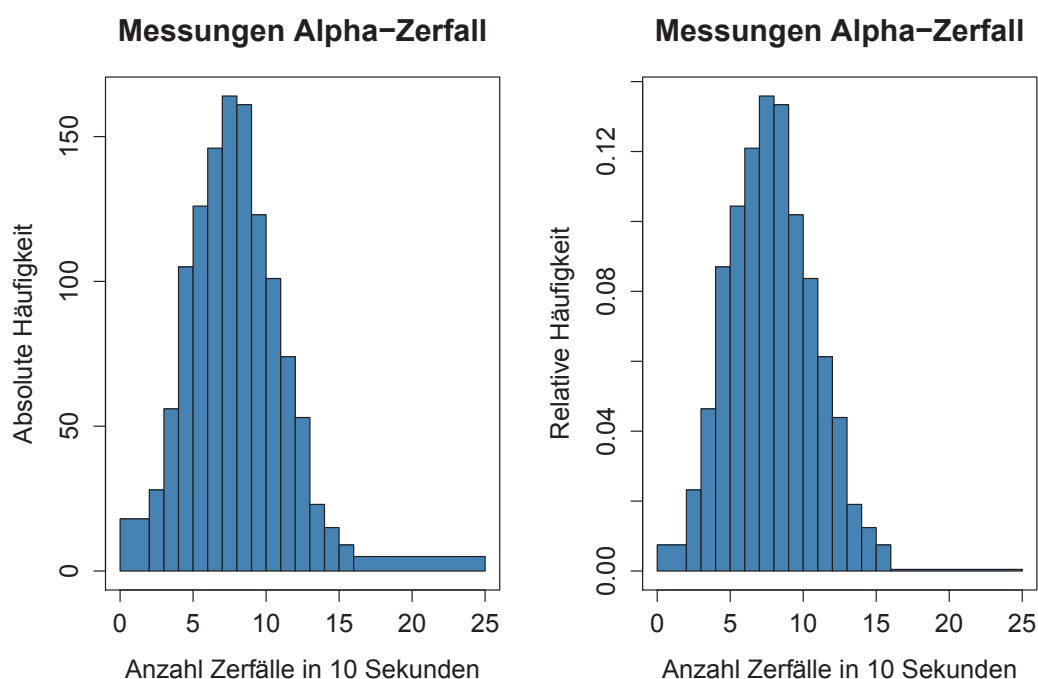


Abbildung 3.1.: Häufigkeitsverteilung des Alphazerfalles

74 in Bezug auf die beobachtete Anzahl Experimente ist ohne die Angabe der Gesamtzahl 1207 beobachteter Experimente allerdings nicht sehr aussagekräftig. Diese

Zahl lässt sich nur schwer mit einer anderen Versuchsreihe vergleichen, zum Beispiel mit einer, bei der 2389 Experimente durchgeführt wurden. Mehr Aufschluss gibt die *relative Häufigkeit* einer bestimmten Anzahl Zerfälle. Dabei dividieren wir die (absolute) Häufigkeit der Experimente, bei denen eine bestimmte Anzahl Zerfälle beobachtet wurde, durch die Gesamtzahl durchgeführter Experimente. Dies ergibt dann den *prozentualen* Anteil der Experimente, bei denen eine bestimmte Anzahl Zerfälle beobachtet wurde. Wie wir im Kapitel über Histogramme gesehen haben, entspricht die Fläche eines Balkens der relativen Häufigkeit von Messwerten im entsprechenden Intervall. So lesen wir aus Abbildung 3.1 rechts, dass in rund 12 % der Experimente 7 Zerfälle in 10 Sekunden stattfanden. Die Gesamtfläche der Balken ergibt in diesem Fall eins. Dies führt uns jetzt natürlich direkt zum Begriff der Zerfallswahrscheinlichkeit. Wir interpretieren also die relativen Häufigkeiten der beobachteten Anzahl Zerfälle pro 10 Sekunden als (geschätzte) *Wahrscheinlichkeiten*: die Wahrscheinlichkeit, dass in 10 Sekunden eine entsprechende Anzahl Zerfälle beobachtet wird. Wahrscheinlichkeiten werden in der Mathematik in der Regel nicht als Prozente angegeben, sondern mit Zahlen zwischen 0 und 1. Die Wahrscheinlichkeit, dass in 10 Sekunden 7 Zerfälle stattfinden, beträgt also 0.12. Je mehr Experimente wir durchführen, desto mehr nähert sich diese geschätzte Wahrscheinlichkeit der wahren Wahrscheinlichkeit an. Weiter stellen wir fest, dass sich alle Zerfälle auf eins aufsummieren.

Als nächstes stellt sich die Frage: wie sind diese Wahrscheinlichkeiten auf die Anzahl Zerfälle verteilt? Gibt es ein mathematisches Modell, das zur beobachteten Verteilung der relativen Häufigkeiten passt, mit dem wir dann auch Vorhersagen machen können?

3.2. Wahrscheinlichkeitsmodelle

Wir betrachten **Zufallsexperimente**, bei denen der Ausgang nicht exakt vorhersagbar ist. Beispiele dafür sind

- Anzahl Zerfälle eines Alphastrahlers
- geworfene Augenzahl bei einem Würfelwurf

Ein **Wahrscheinlichkeitsmodell** beschreibt, welche Ergebnisse in einem solchen Experiment möglich sind und welche Wahrscheinlichkeiten, die verschiedenen Ergebnisse haben. Beim Würfel sind die möglichen Ergebnisse 1, 2, 3, 4, 5, 6 und die Wahrscheinlichkeit eine dieser Zahlen zu werfen, beträgt $\frac{1}{6}$, sofern der Würfel fair ist.

Ein Wahrscheinlichkeitsmodell erlaubt uns dann, gewisse Vorhersagen zu machen, die wir experimentell überprüfen können. Damit können wir uns zum Beispiel bei Glücksspielen eine gute Spielstrategie erarbeiten. Ein Wahrscheinlichkeitsmodell hat die folgenden Komponenten:

- **Grundraum** Ω , bestehend aus den **Elementarereignissen** ω ,

- Ereignisse A, B, C, \dots ,
- Wahrscheinlichkeiten P

Elementarereignisse sind mögliche Ergebnisse oder Ausgänge des Experiments, die zusammen den Grundraum bilden:

$$\Omega = \underbrace{\{\text{mögliche Elementarereignisse } \omega\}}_{\text{mögliche Ausgänge/Resultate}}$$

Beispiel 3.2.1

Beim Würfelwurf ist der Grundraum gegeben durch die möglichen Ergebnisse des Zufallsexperiments

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Das Element $\omega = 2$ ist ein Elementarereignis. Dieses hat die Bedeutung, dass beim Würfeln die Zahl 2 geworfen wurde.

□

Beispiel 3.2.2

Beim Alphazerfall ist der Grundraum gegeben durch

$$\Omega = \{0, 1, 2, 3, 4, \dots\}$$

da hier beliebig viele Zerfälle in einem 10-Sekundenintervall möglich sind. Das Elementarereignis $\omega = 6$ bedeutet, dass in 10 Sekunden 6 Zerfälle gemessen wurden.

□

Bei der Durchführung eines Experiments wird aus der Menge aller Elementarereignisse (Grundraum) ein Elementarereignis zufällig „ausgewählt“.

Beispiel 3.2.3 2-maliges Werfen einer Münze

Wir wählen die Bezeichnungen K für „Kopf“ und Z für „Zahl“ (dies gilt für das gesamte Kapitel). Dann sind alle möglichen Ergebnisse des Experimentes gegeben durch

$$\Omega = \{KK, KZ, ZK, ZZ\}$$

wobei Ω wieder den Grundraum bezeichnet. Ein Elementarereignis ist z.B. $\omega = KZ$.

□

Unter einem **Ereignis** A versteht man eine Teilmenge von Ω :

$$A \subset \Omega$$

„Ein Ereignis A tritt ein“ bedeutet, dass das Ergebnis ω des Experiments zu A gehört.

Beispiel 3.2.4 2-maliges Werfen einer Münze

Wir betrachten das Ereignis A , bei welchem genau einmal K geworfen wird. Dieses Ereignis besteht aus den Elementarereignissen KZ und ZK . Das Ereignis A ist dann gegeben durch die Menge

$$A = \{KZ, ZK\}$$

Würfeln wir ZZ , so trifft das Ereignis A *nicht* ein.

□

Beispiel 3.2.5 Würfeln

Das Ereignis A bezeichne „eine ungerade Augenzahl würfeln“. Dann ist

$$A = \{1, 3, 5\}$$

Das Ereignis A tritt ein, wenn z.B. die Zahl 5 gewürfelt wird.

Wir bezeichnen mit B das Ereignis, eine Zahl kleiner als 7 zu würfeln. Das ist natürlich immer der Fall und somit ist in diesem Fall

$$B = \Omega$$

Wir sprechen vom *sicheren* Ereignis.

Weiter sei C das Ereignis „die Zahl sieben würfeln“. Dies ist unmöglich und wir schreiben

$$C = \emptyset$$

Dabei stellt \emptyset die leere Menge dar, die kein Element enthält. Wir sprechen in diesem Fall vom *unmöglichen* Ereignis.

□

| Name | Symbol | Bedeutung |
|--------------|---------------------------------------|---------------------|
| Vereinigung | $A \cup B$ | A oder B |
| Durchschnitt | $A \cap B$ | A und B |
| Komplement | \overline{A} | nicht A |
| Differenz | $A \setminus B = A \cap \overline{B}$ | A ohne B |

Tabelle 3.2.: Operationen der Mengenlehre

Für den Umgang mit Ereignissen ist es nützlich, sich die Operationen der Mengenlehre und deren Bedeutung in Erinnerung zu rufen. Die Operationen der Mengenlehre (Vereinigung, Durchschnitt, Komplement) werden verwendet, um aus bereits definierten Ereignissen neue Ereignisse zu gewinnen (siehe Tabelle 3.2).

Bemerkungen:

- i. Das „oder“ der Mengenoperation *Vereinigung* ist nicht exklusiv: Ein Element kann also in A *und* in B liegen. Umgangssprachlich wird „oder“ meist als „entweder ... oder ...“ gedeutet. Das ist beim „oder“ der Mengenoperation Vereinigung nicht so.
- ii. Anstelle des Begriffs *Durchschnitt* verwendet man oft auch den Begriff „Schnittmenge“.
- iii. Für das Komplement \overline{A} wird auch die Bezeichnung A^c verwendet.

Die Mengen A und B heissen **disjunkt**, wenn sich A und B gegenseitig ausschliessen und daher nicht zusammen eintreten können. In diesem Fall gilt

$$A \cap B = \emptyset$$

Dieses Ereignis ist also unmöglich.

Ferner gelten die **De Morgan - Regeln**

$$1. \overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$2. \overline{A \cup B} = \overline{A} \cap \overline{B}$$

Alle diese Begriffe, Operationen und Regeln lassen sich einfach mit Venn-Diagrammen illustrieren (siehe Abbildung 3.2).

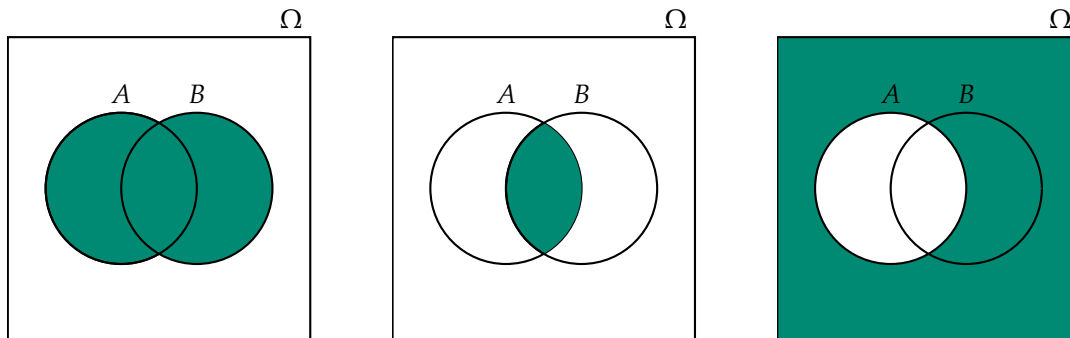


Abbildung 3.2.: Links: $A \cup B$, Mitte: $A \cap B$ und Rechts: $\overline{A \cup B}$

Beispiel 3.2.6

Wir betrachten die Ereignisse

- A : „morgen scheint die Sonne“
- B : „morgen regnet es“

Dann bedeuten folgende Ereignisse:

1. $A \cup B$:

„morgen scheint die Sonne *oder* morgen regnet es“, und dies kann auch bedeuten: „morgen scheint die Sonne *und* morgen regnet es“.

2. $A \cap B$:

„morgen scheint die Sonne *und* morgen regnet es“

3. \overline{A} :

„morgen scheint die Sonne nicht“

□

Die Wahrscheinlichkeitsrechnung wird auf folgenden drei Grundregeln (Axiome) aufgebaut:

Kolmogorov Axiome der Wahrscheinlichkeitsrechnung

Jedem Ereignis A wird eine **Wahrscheinlichkeit** $P(A)$ zugeordnet. Dabei müssen die folgenden drei grundlegenden Regeln - **Axiome der Wahrscheinlichkeitsrechnung** - erfüllt sein:

$$A1: P(A) \geq 0$$

$$A2: P(\Omega) = 1$$

$$A3: P(A \cup B) = P(A) + P(B) \quad \text{falls } A \cap B = \emptyset$$

Bemerkungen:

- i. Die Bezeichnung $P(A)$ steht für die Wahrscheinlichkeit, dass das Ereignis A eintritt. Ist A das Ereignis, eine ungerade Zahl zu würfeln, so gilt bei einem fairen Würfel

$$P(A) = \frac{1}{2}$$

- ii. Der Buchstabe P steht für das englische Wort *probability*.
iii. Wahrscheinlichkeiten sind nie negativ.
iv. Mit $P(\Omega) = 1$ wird festgelegt, dass die Wahrscheinlichkeiten eines Ereignisses zwischen 0 und 1 liegen müssen.

Beispiel 3.2.7

Beim Wurf zweier Münzen ist es plausibel, dass alle 4 Elemente von

$$\Omega = \{KK, KZ, ZK, ZZ\}$$

gleich wahrscheinlich sind. Wegen $P(\Omega) = 1$ müssen sich die Wahrscheinlichkeiten zu eins aufaddieren. Somit gilt

$$P(KK) = P(KZ) = P(ZK) = P(ZZ) = \frac{1}{4}$$

□

Weitere Regeln können aus den drei Axiomen von Kolmogorov hergeleitet werden:

Rechenregeln

Sind A, B und A_1, \dots, A_n Ereignisse, dann gilt

$$P(\overline{A}) = 1 - P(A) \quad \text{für jedes } A$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{für beliebige } A \text{ und } B$$

$$P(A_1 \cup \dots \cup A_n) \leq P(A_1) + \dots + P(A_n) \quad \text{für beliebige } A_1, \dots, A_n$$

$$P(B) \leq P(A) \quad \text{für beliebige } A \text{ und } B \text{ mit } B \subseteq A$$

$$P(A \setminus B) = P(A) - P(B) \quad \text{für beliebige } A \text{ und } B \text{ mit } B \subseteq A$$

Wir können uns die Wahrscheinlichkeiten als Flächen im Venn-Diagramm vorstellen. Dabei ist die Totalfläche von Ω gleich 1 oder $|\Omega| = 1$. Dann erscheinen diese Rechenregeln ganz natürlich. So folgt die erste Regel sofort aus Abbildung 3.2 rechts. Dabei ist $P(A)$ der Flächeninhalt der Fläche A und $P(\overline{A})$ ist der Flächeninhalt der restlichen Fläche in Ω . Es gilt also offensichtlich

$$P(A) + P(\overline{A}) = |\Omega| = 1$$

und damit

$$P(\overline{A}) = 1 - P(A)$$

Überprüfen Sie die restlichen Regeln als Übung.

Grundsätzlich werden in der Wahrscheinlichkeitstheorie die Wahrscheinlichkeiten gewisser Ereignisse A festgelegt (auf Grund von Plausibilitäten, Symmetrieüberlegungen, wissenschaftlichen Theorien, Fachwissen und Daten) und daraus die Wahrscheinlichkeiten von gewissen anderen Ereignissen B aus den obigen Gesetzen hergeleitet.

Die Statistik geht umgekehrt vor: aus Daten, d. h. aus der Information, dass gewisse Ereignisse eingetreten sind, versucht man Rückschlüsse auf ein unbekanntes Wahrscheinlichkeitsmodell (unbekannte Wahrscheinlichkeiten) zu ziehen.

Interpretation von Wahrscheinlichkeiten:

- **Frequentistisch:** Idealisierung der relativen Häufigkeiten bei vielen unabhängigen Wiederholungen

Aus Messungen zum Alphazerfall (Experiment) schliessen wir, dass in 12 % der Beobachtungen 7 Zerfälle in 10 Sekunden vorkommen. Also zum Beispiel in 12 Experimenten von insgesamt 100 wurden 7 Zerfälle beobachtet.

- **Bayes'sch:** Mass für den Glauben, dass ein Ereignis eintreten wird

„Ich bin mir zu 90 % sicher, dass ich die nächste Prüfung bestehen werde.“

Wir behandeln in diesem Kapitel *diskrete* Wahrscheinlichkeitsmodelle, bei denen der Grundraum endlich oder unendlich und diskret ist. Mit dem Begriff „diskret“ meinen wir zum Beispiel folgende Menge

$$\Omega = \{0, 1, \dots, 10\}$$

die *endlich und deshalb diskret* ist, während

$$\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$$

zwar *unendlich*, aber trotzdem *diskret* ist.

Die Menge $\Omega = \mathbb{R}$ ist *nicht* diskret. Sie wird später für Messdaten eine sehr wichtige Rolle spielen.

Im diskreten Fall ist die Wahrscheinlichkeit eines Ereignisses

$$A = \{\omega_1, \omega_2, \dots, \omega_n\}$$

durch die Wahrscheinlichkeiten der zugehörigen Elementarereignisse $P(\omega)$ festgelegt:

$$P(A) = P(\omega_1) + P(\omega_2) + \dots + P(\omega_n) = \sum_{\omega_i \in A} P(\omega_i)$$

Dies folgt aus Axiom A3.

Beispiel 3.2.8

Wir haben einen Würfel, der nicht fair ist. Die Wahrscheinlichkeiten, unterschiedliche Zahlen zu werfen, sind also nicht gleich. In Tabelle 3.3 sind die entsprechenden Wahrscheinlichkeiten für alle Zahlen angegeben.

| | | | | | | |
|-------------|---------------|---------------|----------------|---------------|----------------|----------------|
| ω | 1 | 2 | 3 | 4 | 5 | 6 |
| $P(\omega)$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{4}$ | $\frac{1}{12}$ | $\frac{1}{12}$ |

Tabelle 3.3.: Wahrscheinlichkeiten für einen nicht-fairen Würfel.

Es gilt

$$\begin{aligned} P(\Omega) &= P(1) + P(2) + P(3) + P(4) + P(5) + P(6) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{12} + \frac{1}{4} + \frac{1}{12} + \frac{1}{12} \\ &= 1 \end{aligned}$$

Für das Ereignis $A = \{1, 2, 4\}$ gilt

$$\begin{aligned} P(A) &= P(1) + P(2) + P(4) \\ &= \frac{1}{3} + \frac{1}{6} + \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

□

Beispiel 3.2.9 2-maliges Werfen einer Münze

Es sei das Ereignis A : „genau einmal K werfen“, also

$$A = \{KZ, ZK\}$$

Dann erhalten wir für die Wahrscheinlichkeit $P(A)$, dass das Ereignis A eintritt

$$P(A) = P(KZ) + P(ZK) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Für das Ereignis B : „mindestens einmal Kopf werfen“ erhalten wir die zugehörige Wahrscheinlichkeit, dass dieses Ereignis eintritt

$$P(B) = P(KZ) + P(ZK) + P(KK) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

da $B = \{KZ, ZK, KK\}$. Diese Wahrscheinlichkeit können wir in diesem Fall einfacher mit der sogenannten *Gegenwahrscheinlichkeit* berechnen. Das Komplement \bar{B} von B ist gegeben durch

$$\bar{B} = \{ZZ\}$$

Und damit erhalten wir mit der ersten Rechenregel (siehe oben):

$$P(B) = 1 - P(\bar{B}) = 1 - \frac{1}{4} = \frac{3}{4}$$

□

Laplace-Wahrscheinlichkeit

In vielen Fällen ist es plausibel anzunehmen, dass jedes Elementarereignis die gleiche Wahrscheinlichkeit hat. Ein solches Wahrscheinlichkeitsmodell wird **Modell von Laplace** genannt. Das Standardbeispiel hierfür ist der Wurf eines fairen Würfels. Da wird *jede* Zahl mit der Wahrscheinlichkeit $\frac{1}{6}$ geworfen. In diesen Fällen gibt es eine besonders einfache Möglichkeit, die Wahrscheinlichkeit eines Ereignisses zu berechnen. Ein Ereignis E besteht aus den g verschiedenen Elementarereignissen

$$E = \{\omega_1, \omega_2, \dots, \omega_g\}$$

und der Grundraum Ω aus m Elementarereignissen.

Damit sich die Wahrscheinlichkeiten aller Elementarereignisse zu 1 addieren (siehe oben), haben wir hier

$$P(\omega_k) = \frac{1}{|\Omega|} = \frac{1}{m}$$

Somit gilt für $P(E)$:

$$\begin{aligned} P(E) &= P(\omega_1) + P(\omega_2) + \dots + P(\omega_g) \\ &= \frac{1}{m} + \frac{1}{m} + \dots + \frac{1}{m} \\ &= g \cdot \frac{1}{m} \\ &= \frac{g}{m} \end{aligned}$$

Für ein Ereignis E im Laplace Modell gilt also

$$P(E) = \sum_{\omega_k \in E} P(\omega_k) = \frac{g}{m}$$

Man teilt die Anzahl der „**günstigen**“ Elementarereignisse durch die Anzahl der „**möglichen**“ Elementarereignisse.

Beispiel 3.2.10

Es werden zwei Würfel geworfen. Wie gross ist die Wahrscheinlichkeit, dass die Augensumme 7 ergibt?

Ein Elementarereignis beschreibt die Augenzahlen auf beiden Würfeln. Dieses Ereignis können wir in der Form $(1, 4)$ schreiben, wenn der eine Würfel eine 1 und der andere eine 4 zeigt. Es sind insgesamt 36 Elementarereignisse möglich:

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$$

Wir bezeichnen mit E das Ereignis, dass die Augensumme 7 erreicht wird. Es gibt davon 6 Elementarereignisse:

$$E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

Da alle Elementarereignisse gleich wahrscheinlich sind, ist die Wahrscheinlichkeit für das Ereignis E :

$$P(E) = \frac{|E|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

□

3.3. Der Begriff der Unabhängigkeit

Wenn wir die Wahrscheinlichkeiten $P(A)$ und $P(B)$ kennen, so lässt sich die Wahrscheinlichkeit $P(A \cap B)$ im Allgemeinen nicht aus $P(A)$ und $P(B)$ berechnen: Es sind alle Werte zwischen 0 und dem Minimum von $P(A)$ und $P(B)$ möglich. Ein wichtiger Spezialfall liegt vor, wenn die Berechnung von $P(A \cap B)$ aus $P(A)$ und $P(B)$ mit Hilfe folgender Produktformel möglich ist.

Die Ereignisse A und B heißen **stochastisch unabhängig**, wenn

$$P(A \cap B) = P(A)P(B)$$

Beispiel 3.3.1

Es sei A : „K im 1. Wurf“ und B : „K im 2. Wurf“. Dann gilt

$$P(A) = P(KK) + P(KZ) = \frac{1}{2}$$

und analog

$$P(B) = P(KK) + P(ZK) = \frac{1}{2}$$

Mit $A \cap B$ wird das Ereignis „K im 1. und im 2. Wurf“ beschrieben. Es gilt dann für die Wahrscheinlichkeit dieses Ereignisses

$$P(A \cap B) = P(KK) = \frac{1}{4}$$

Da

$$P(A)P(B) = \frac{1}{4}$$

gilt also

$$P(A \cap B) = P(A)P(B)$$

Somit erfüllen die Ereignisse A und B die Bedingung für stochastische Unabhängigkeit. Die Ereignisse A und B sind also unabhängig.

□

Viel wichtiger ist jedoch der umgekehrte Schluss. Wenn zwischen den Ereignissen A und B kein *kausaler Zusammenhang* besteht - d.h. es gibt keine gemeinsamen Ursachen oder Ausschlüsse - dann nehmen wir stochastische Unabhängigkeit an, und damit *muss* obige Produktformel gelten. In diesem Fall kann also $P(A \cap B)$ aus $P(A)$ und $P(B)$ berechnet werden.

Die stochastische Unabhängigkeit soll durch die folgenden Beispiele veranschaulicht werden.

Beispiel 3.3.2

Wir werfen zuerst einen Würfel und ziehen nachher eine Karte aus einem vollständigen Stapel Jasskarten. Wie gross ist die Wahrscheinlichkeit, zuerst eine 3 zu werfen und nachher einen König zu ziehen?

Dazu bezeichnen wir dieses Elementarereignis mit 3K. Die Wahrscheinlichkeit, eine 3 zu werfen ist $1/6$, also

$$P(3) = \frac{1}{6}$$

Die Wahrscheinlichkeit, einen König zu ziehen ist $1/9$:

$$P(K) = \frac{1}{9}$$

Der Würfelwurf hat auf das Ziehen einer Jasskarte keinen Einfluss. Zur Berechnungen der Wahrscheinlichkeit, dass beide Ereignisse eintreffen, müssen wir $1/9$ von einem Sechstel nehmen. Dies entspricht einer Multiplikation der beiden Wahrscheinlichkeiten und es gilt

$$P(3K) = P(3) \cdot P(K) = \frac{1}{6} \cdot \frac{1}{9} = \frac{1}{54}$$

Die Produktformel gilt, da die beiden Ereignisse unabhängig sind.

□

Beispiel 3.3.3

Es sei A : „ K im 1. Wurf“ und B : „ K im 2. Wurf“.

Es ist plausibel, dass es keinen kausalen Zusammenhang zwischen dem Ergebnis des ersten und des zweiten Wurfs gibt. Mit anderen Worten hat das Ergebnis des 1. Wurfs (Ereignis A) keinen Einfluss auf das Ergebnis des 2. Wurfs (Ereignis B). Die Ereignisse A und B sind also unabhängig. Deshalb kann man $P(A \cap B)$ wie folgt berechnen:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

□

Beispiel 3.3.4

Angenommen, wir werfen zwei Würfel. Sei E das Ereignis, dass der erste Würfel die Augenzahl 4 zeigt und F das Ereignis, dass die *Augensumme* beider Würfel 6 ist. Sind die Ereignisse E und F unabhängig voneinander? Um dies festzustellen, müssen wir überprüfen, ob die Produktregel

$$P(E \cap F) = P(E) \cdot P(F)$$

gilt oder nicht.

Das Ereignis $E \cap F$ besteht nur aus dem Elementarereignis $(4, 2)$. Damit gilt für dessen Wahrscheinlichkeit

$$P(E \cap F) = P((4, 2)) = \frac{1}{36}$$

Es gilt natürlich $P(E) = 1/6$. Das Ereignis F besteht aus den Elementen

$$F = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$$

und damit gilt

$$P(F) = \frac{5}{36}$$

Nun können wir das Produkt dieser Wahrscheinlichkeiten berechnen und damit die Produktformel überprüfen:

$$P(E) \cdot P(F) = \frac{1}{6} \cdot \frac{5}{36} = \frac{5}{216} \neq \frac{1}{36} = P(E \cap F)$$

Deshalb sind E und F nicht unabhängig.

Wir können die Abhängigkeit der beiden Ereignisse E und F auch ohne Rechnung leicht einsehen. Werfen wir den ersten Würfel und erhalten eine 6, dann haben wir keine Möglichkeit mehr in zwei Würfeln die Augensumme 6 zu erreichen : Die Wahrscheinlichkeit ist 0. Werfen wir zuerst irgendeine andere Zahl, so haben wir noch die Möglichkeit, auf die Augensumme 6 zu kommen: Die Wahrscheinlichkeit ist ungleich 0. Damit hängt die Wahrscheinlichkeit, die Augensumme 6 zu erzielen von der Augenzahl des ersten Wurfes ab. Also können E und F nicht unabhängig voneinander sein.

□

Das vorherige Beispiel führt uns direkt zum Begriff der *bedingten Wahrscheinlichkeit*. Die Wahrscheinlichkeit, in zwei Würfeln die Augensumme 6 zu erreichen, *gegeben* im ersten Wurf wurde eine 6 gewürfelt, ist null.

Haben wir drei Ereignisse A_1 , A_2 und A_3 , so sind diese Ereignisse unabhängig, wenn

$$P(A_1 \cap A_2) = P(A_1)P(A_2), P(A_1 \cap A_3) = P(A_1)P(A_3), P(A_2 \cap A_3) = P(A_2)P(A_3)$$

und

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

Für n Ereignisse A_1, \dots, A_n wird die Unabhängigkeit analog definiert. Alle möglichen Schnittmengen von Ereignissen A_1, \dots, A_n müssen die Produktformel erfüllen.

3.4. Wahrscheinlichkeitsbäume

Ein Zufallsprozess, der aus mehreren nacheinander ablaufenden Zufallsexperimenten besteht, wird *mehrstufiges Zufallsexperiment* genannt. Die Beispiele aus dem vorhergehenden Kapitel sind solche mehrstufigen Zufallsexperimente. Wird eine Münze zweimal geworfen, handelt es sich um einen Zufallsprozess, bei dem nacheinander zwei einfache Zufallsexperimente ablaufen (*2-stufiges Zufallsexperiment*). Wird zuerst ein Würfel geworfen und danach eine Jasskarte aus einem Stapel gezogen, so stellt auch dieser Zufallsprozess ein 2-stufiges Zufallsexperiment dar.

Ein äusserst anschauliches Hilfsmittel bei der Berechnung von Wahrscheinlichkeiten bei *mehrstufigen Zufallsexperimenten* ist der sogenannte *Wahrscheinlichkeitsbaum* (häufig auch als *Ereignisbaum* oder *Baumdiagramm* bezeichnet). Er besteht aus einer Wurzel (Ausgangspunkt) und mehreren *Verzweigungspunkten* und einer Vielzahl von *Zweigen*. In Abbildung 3.3 ist ein solcher Wahrscheinlichkeitsbaum dargestellt. Die

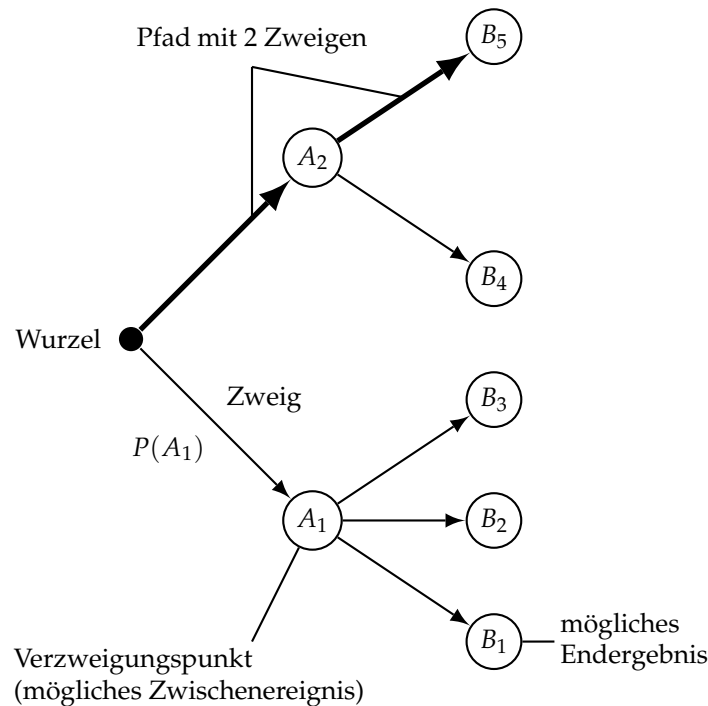


Abbildung 3.3.: Wahrscheinlichkeitsbaum mit den Verzweigungspunkten A_1 und A_2 (mögliche Ergebnisse der 1. Stufe) und den möglichen Endergebnissen B_1 bis B_5 .

Verzweigungspunkte A_1 und A_2 charakterisieren dabei die möglichen Zwischenergebnisse nach der 1. Stufe. Die von diesen Verzweigungspunkten ausgehenden Zweige führen zu den jeweils möglichen Ergebnissen der nachfolgenden 2. Stufe : Hier sind es die möglichen *Endergebnisse* B_1 bis B_5 . Die Wahrscheinlichkeit eines bestimmten Ereignisses schreibt man an den betreffenden Zweig. So ist z.B. $P(A_1)$ die Wahrscheinlichkeit des Ereignisses A_1 . Ein mögliches *Endergebnis* erreicht man dann immer von der Wurzel ausgehend längs eines bestimmten Pfades. Dieser besteht meistens aus mehreren Zweigen, wie z.B. der in Abbildung dick eingezeichnete Pfad, der von der Wurzel über das Zwischenergebnis A_2 zum Endergebnis B_5 führt. Für die Berechnung der Wahrscheinlichkeit eines bestimmten Endergebnisses gelten dann die folgenden Regeln:

Regeln für die Berechnung von Wahrscheinlichkeiten bei mehrstufigen Zufallsexperimenten mit Hilfe des Wahrscheinlichkeitsbaumes

Bei einem in mehreren Stufen ablaufenden Zufallsexperiment lässt sich jedes mögliche Endergebnis durch einen bestimmten Pfad im zugehörigen *Wahrscheinlichkeitsbaum* anschaulich darstellen. Die Berechnung der Wahrscheinlichkeiten für die möglichen Endergebnisse erfolgt dabei mit Hilfe der folgenden Regeln:

1. Die Wahrscheinlichkeiten längs eines *Pfades* werden miteinander *multipliziert*.
2. Führen *mehrere* Pfade zum *gleichen* Endergebnis, so *addieren* sich ihre Wahrscheinlichkeiten.

Beispiel 3.4.1 2-maliges Werfen einer Münze

Es sei A : „K im 1. Wurf“ und B : „K im 2. Wurf“. Wir möchten mit Hilfe des Wahrscheinlichkeitsbaumes die Wahrscheinlichkeit des Ereignisses $E = A \cap B$ berechnen.

Die Wahrscheinlichkeit für „K im 1. Wurf“ beträgt $\frac{1}{2}$, die Wahrscheinlichkeit für „K im 2. Wurf“ beträgt wieder $\frac{1}{2}$. Somit ist die Wahrscheinlichkeit des Endergebnisses E (K im 1. Wurf *und* K im 2. Wurf) gegeben durch die Multiplikation der Wahrscheinlichkeiten längs des Pfades, also

$$P(E) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Wir lesen aus dem Wahrscheinlichkeitsbaum (siehe Abbildung 3.4), dass es für die Wahrscheinlichkeit von „K im 2. Wurf“ keine Rolle spielt, ob wir im 1. Wurf „Z“ oder „K“ erhalten haben. Die Ereignisse „K im 1. Wurf“ und „K im 2. Wurf“ sind stochastisch unabhängig.

□

Beispiel 3.4.2 Würfel und Jasskarte

Betrachten wir das Beispiel, in welchem zuerst ein (fairer) Würfel geworfen und dann eine Jasskarte gezogen wird. Uns interessiert die Wahrscheinlichkeit, zuerst eine 3 zu werfen und nachher einen König zu ziehen. Wir können dazu den in Abbildung 3.5 dargestellten Wahrscheinlichkeitsbaum zur Berechnung verwenden. In diesem Fall haben wir nicht nur ein Endergebnis, sondern deren vier: auf der 2. Stufe sind „König Schellen“, „König Schilten“, „König Eichel“ und „König Rosen“ Elemente des

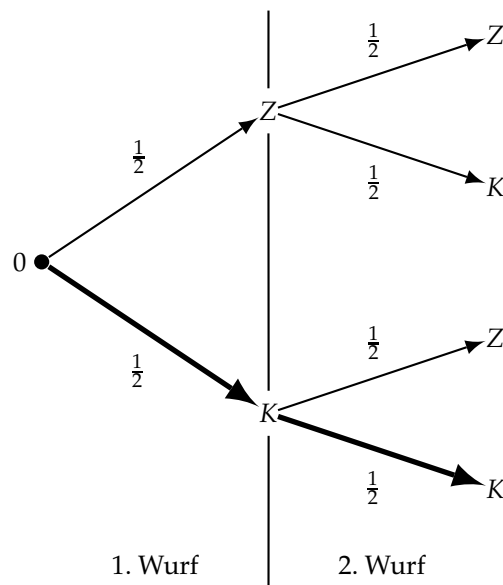


Abbildung 3.4.: Wahrscheinlichkeitsbaum für das zweimalige Werfen einer Münze.

Ereignisses, zuerst eine 3 zu werfen und nachher einen König zu ziehen. Daher addieren wir die Wahrscheinlichkeiten aller Pfade, die zum selben Endergebnis führen. Dies ergibt dann die Wahrscheinlichkeit $4 \cdot \frac{1}{6} \cdot \frac{1}{36} = \frac{1}{54}$.

Würden wir den Wahrscheinlichkeitsbaum vervollständigen, wäre ersichtlich, dass die Wahrscheinlichkeit, auf der 2. Stufe einen König zu ziehen, unabhängig vom Zwischenergebnis der 1. Stufe ist. Dies zeigt wiederum, dass die Augenzahl auf der 1. Stufe und die gezogene Jasskarte unabhängige Ereignisse sind.

□

Beispiel 3.4.3

Betrachten wir wiederum das Beispiel, in dem wir zwei Würfel werfen. Uns interessiert das Endergebnis, wo die *Augensumme* der beiden Würfel 6 beträgt. Wir bezeichnen mit F das Ereignis „Augenzahl des 2. Würfels, so dass Augensumme der beiden Würfel gleich 6 ist“. Der entsprechende Wahrscheinlichkeitsbaum ist in der Abbildung 3.6 dargestellt.

Nun sehen wir, dass die Wahrscheinlichkeiten für die Zweige auf der 2. Stufe vom Ausgang der vorangehenden Stufe abhängen. Die Wahrscheinlichkeit des Ereignisses F am Zweig nach dem Zwischenergebnis „6 beim 1. Würfel“ beträgt 0, während die Wahrscheinlichkeit für F an den übrigen Zweigen $\frac{1}{6}$ beträgt. Es handelt sich hier um *bedingte* Wahrscheinlichkeiten, auf die wir nun im folgenden Kapitel eingehen.

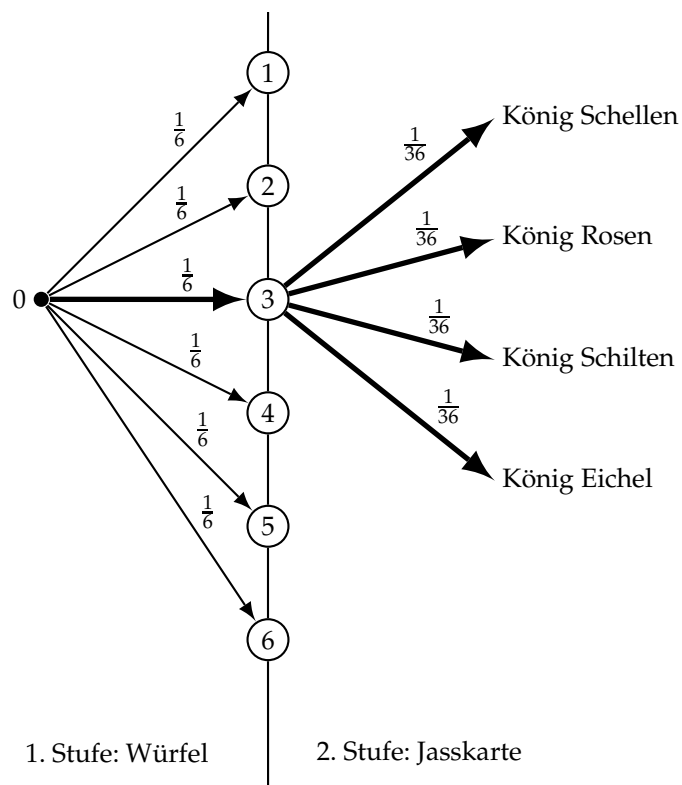


Abbildung 3.5.: Wahrscheinlichkeitsbaum für das Werfen einer (fairen) Münze und dann das Ziehen einer Jasskarte.

□

3.5. Bedingte Wahrscheinlichkeit

Wir wollen zuerst den Begriff der bedingten Wahrscheinlichkeit an einem kleinen Beispiel kennenlernen. Dazu betrachten wir einen fairen Würfel, bei dem die ungeraden Zahlen blau und die geraden Zahlen rot gefärbt sind. Wir werfen nun den Würfel und sehen, dass eine blaue Seite oben liegt. Damit wissen wir, dass eine ungerade Zahl geworfen wurde. Wie gross ist nun die Wahrscheinlichkeit, dass die Zahl 3 geworfen wurde? Es gibt die drei ungeraden Zahlen 1, 3 und 5 auf dem Würfel. Die 3 ist eine davon. Wenn wir davon ausgehen, dass jede Zahl auf dem Würfel gleich wahrscheinlich ist, ist das Ergebnis $\frac{1}{3}$. Natürlich können wir auch nach der Wahrscheinlichkeit fragen, die Zahl 2 zu werfen, wenn eine blaue Seite oben ist. Diese Wahrscheinlichkeit ist natürlich 0, da wir ja schon wissen, dass eine ungerade Zahl geworfen wurde. Der

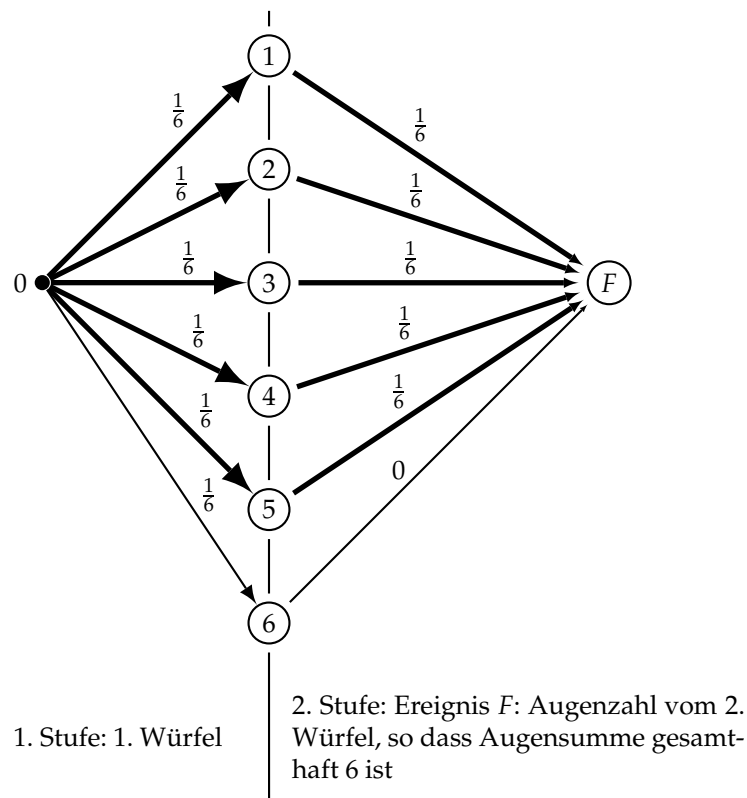


Abbildung 3.6.: Wahrscheinlichkeitsbaum für die Augensumme 6 in zwei Würfeln.

entsprechende Wahrscheinlichkeitsbaum ist in Abbildung 3.7 dargestellt. Wir bemerken bei der Bestimmung der Wahrscheinlichkeiten für die Zweige des Wahrscheinlichkeitsbaumes, dass von der 2. Stufe ab alle Ereignisse vom Ausgang der *vorangegangenen* 1. Stufe abhängen.

Fragen dieser Art kann man auch in komplizierteren Sachverhalten mit dem Begriff der **bedingten Wahrscheinlichkeit** beantworten.

Die **bedingte Wahrscheinlichkeit** ist die Wahrscheinlichkeit, dass das Ereignis A eintritt, wenn wir schon wissen, dass B eingetreten ist. Diese Wahrscheinlichkeit wird mit

$$P(A|B)$$

bezeichnet. Der Längsstrich wird als „unter der Bedingung“ gelesen.

Die bedingte Wahrscheinlichkeit $P(A|B)$ wird definiert durch

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

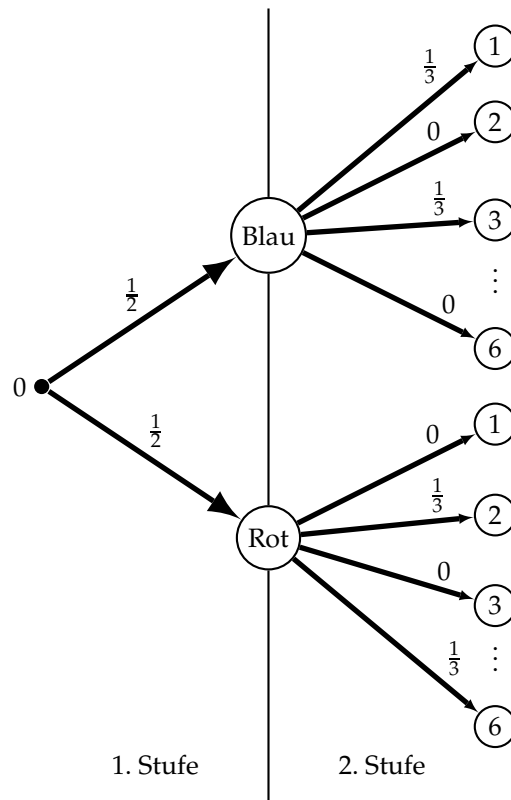


Abbildung 3.7.: Wahrscheinlichkeitsbaum für Wurf eines Würfels, wobei Seiten mit ungerader Augenzahl blau, Seiten mit gerader Augenzahl rot gefärbt sind.

Die Interpretation ist folgendermassen: $P(A|B)$ ist die Wahrscheinlichkeit für das Ereignis A , wenn wir wissen, dass das Ereignis B schon eingetroffen ist.

Wie ist diese Formel zu verstehen? Da wir wissen, dass B schon eingetreten ist, haben wir einen neuen Grundraum $\Omega' = B$. Damit müssen wir von A nur noch denjenigen Teil anschauen, der sich in B abspielt (daher $A \cap B$). Dies müssen wir jetzt noch in Relation zur Wahrscheinlichkeit von B bringen. Das lässt sich anschaulich einfach im Venn-Diagramm in Abbildung 3.8 verdeutlichen.

Wenn wir die Wahrscheinlichkeiten wieder als Flächen denken ($|\Omega| = 1$), dann ist die Wahrscheinlichkeit $P(A \cap B)$ der Flächeninhalt der dunkel gefärbten Fläche, während $P(B)$ der Flächeninhalt der gesamten gefärbten Fläche B ist. Die bedingte Wahrscheinlichkeit ist dann gerade der Anteil der dunkelgefärbten Fläche zur gesamten gefärbten Fläche, also

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

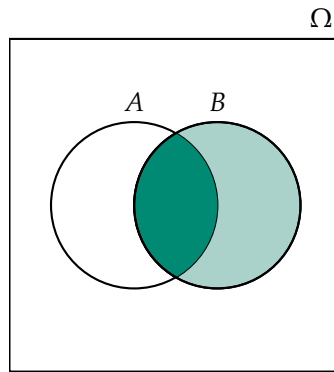


Abbildung 3.8.: Hilfsillustration für bedingte Wahrscheinlichkeiten.

Bemerkungen:

- i. In der Definition für die bedingte Wahrscheinlichkeit wird stillschweigend davon ausgegangen, dass $P(B) > 0$ gilt. Das muss auch so sein, da wir sonst eine Division durch 0 hätten.

Beispiel 3.5.1

Wir kommen nochmals auf das Beispiel vom Anfang dieses Unterkapitels zurück und betrachten es mit den Begriffen der bedingten Wahrscheinlichkeit.

Sei A : „Die 3 wird geworfen“ und B : „Eine ungerade Zahl wird geworfen“ (blau gefärbte Würfelseite). Für die bedingte Wahrscheinlichkeit $P(A|B)$ gilt dann

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

Dies ist die Wahrscheinlichkeit, dass eine 3 geworfen wird, *wenn* wir schon wissen, dass eine ungerade Zahl geworfen wurde. Die Wahrscheinlichkeiten $P(A|B)$ stehen an den Zweigen im Wahrscheinlichkeitsbaum, die vom Ereignis „B“ ausgehen und zu den ungeraden Augenzahlen führen (Abbildung 3.7).

□

Bedingte Wahrscheinlichkeiten sind Wahrscheinlichkeiten für spezielle Situationen. Es gelten wieder entsprechende Rechenregeln.

Rechenregeln

| | |
|---|--|
| $0 \leq P(A B) \leq 1$ | für jedes Ereignis A |
| $P(B B) = 1$ | für jedes Ereignis B |
| $P(A_1 \cup A_2 B) = P(A_1 B) + P(A_2 B)$ | für A_1, A_2 disjunkt (d.h. $A_1 \cap A_2 = \emptyset$) |
| $P(\bar{A} B) = 1 - P(A B)$ | für jedes Ereignis A |

Im Allgemeinen ist $P(A|B)$ **nicht** gleich wie $P(B|A)$, da diese Wahrscheinlichkeiten von völlig verschiedenen Ereignissen sind ¹. So ist $P(A|B)$ im Beispiel vorher die Wahrscheinlichkeit, eine 3 zu werfen, wenn wir schon wissen, dass eine ungerade Zahl geworfen wurde. Auf der anderen Seite ist $P(B|A)$ hingegen die Wahrscheinlichkeit, dass wir eine ungerade Zahl werfen, wenn wir wissen, dass schon die Zahl 3 geworfen wurde. Diese Wahrscheinlichkeit ist natürlich 1, da 3 sowieso ungerade ist.

Wir wollen die bedingte Wahrscheinlichkeit nun an einem realistischen Beispiel untersuchen, das ein überraschendes Ergebnis hat.

Beispiel 3.5.2

Ein medizinischer Test auf eine (tödliche) Krankheit soll feststellen, ob eine Person an dieser Krankheit erkrankt ist oder nicht. Natürlich ist dieser Test nicht ganz genau. Manchmal zeigt er die Krankheit an, obwohl die Person gesund ist, oder er zeigt die Krankheit nicht an, obwohl die Person krank ist. Uns interessiert folgende Frage: Sie gehen zum Arzt und machen diesen Test auf die Krankheit. Der Test ist positiv: Sie haben gemäss dem Test die Krankheit. Wie gross ist die Wahrscheinlichkeit, dass Sie auch wirklich diese Krankheit haben?

Um diese Frage zu beantworten, führen wir folgende Bezeichnungen ein:

- D : Krankheit ist vorhanden; \bar{D} : Krankheit ist nicht vorhanden
- $+$: Test zeigt Krankheit an; $-$: Test zeigt Krankheit nicht an

Die Wahrscheinlichkeiten in Tabelle 3.4 sind durch Versuche bekannt. Z.B. ist die Wahrscheinlichkeit, dass die Krankheit vorhanden ist *und* der Test positiv ausfällt

$$P(D \cap +) = 0.009$$

Diese Wahrscheinlichkeit ist recht klein. Der Grund dafür liegt darin, dass nur ein kleiner Prozentsatz der Bevölkerung diese Krankheit hat.

Hier können wir verschiedene bedingte Wahrscheinlichkeiten berechnen:

- $P(+|D)$: W'keit, dass ein Kranker auch wirklich positiv getestet wird

¹Diese Verwechslung wird vor Gericht (und nicht nur da) so oft gemacht, dass er schon einen eigenen Namen hat: Prosecutor's fallacy.

| | D | D^c |
|-----|-------|-------|
| $+$ | 0.009 | 0.099 |
| $-$ | 0.001 | 0.891 |

Tabelle 3.4.: Wahrscheinlichkeit für eine Krankheit.

- $P(+|\overline{D})$: W'keit, dass ein Gesunder fälschlicherweise positiv getestet wird
- $P(-|D)$: W'keit, dass ein Kranker fälschlicherweise negativ getestet wird
- $P(-|\overline{D})$: W'keit, dass ein Gesunder richtigerweise negativ getestet wird
- $P(D|+)$: W'keit, dass ein positiv Getesteter auch wirklich krank ist
- $P(D|-)$: W'keit, dass ein negativ Getesteter fälschlicherweise krank ist
- $P(\overline{D}|+)$: W'keit, dass ein positiv Getesteter fälschlicherweise gesund ist
- $P(\overline{D}|-)$: W'keit, dass ein negativ Getesteter wirklich auch gesund ist

Wir berechnen zuerst die Wahrscheinlichkeit $P(+|D)$ wie folgt:

$$P(+|D) = \frac{P(+ \cap D)}{P(D)} = \frac{0.009}{0.009 + 0.001} = 0.9$$

Dabei haben wir für $P(D)$ folgende Tatsache benützt

$$P(D) = P(D \cap +) + P(D \cap -) = 0.009 + 0.001$$

Dies ist die Summe der Einträge in der Tabelle in der Spalte unter D . Die Kranken sind entweder positiv oder negativ getestet.

Die bedingte Wahrscheinlichkeit $P(-|\overline{D})$ ist die Wahrscheinlichkeit, dass eine gesunde Person auch wirklich negativ (gemäss Test ist Krankheit nicht vorhanden) getestet wird:

$$P(-|\overline{D}) = \frac{P(- \cap \overline{D})}{P(\overline{D})} = \frac{0.891}{0.891 + 0.099} = 0.9$$

Dieser Test scheint recht genau zu sein. Kranke Personen werden zu 90 % als positiv eingestuft, und gesunde Personen werden zu 90 % als negativ eingestuft.

Wir können die Fragestellung aber auch umkehren. Angenommen, Sie gehen zu einem Test, und dieser wird als positiv eingestuft. Wie gross ist die Wahrscheinlichkeit, dass Sie die Krankheit wirklich haben? Die meisten Leute würden 0.9 antworten. Müssen Sie sich also grosse Sorgen machen und das Testament schreiben oder einer Sterbehilfeorganisation beitreten?

Schauen wir uns nun die *richtige* Antwort an und dies ist die bedingte Wahrscheinlichkeit $P(D|+)$:

$$P(D|+) = \frac{P(+ \cap D)}{P(+)} = \frac{0.009}{0.009 + 0.099} = 0.08$$

Was bedeutet nun dieses Resultat? Die bedingte Wahrscheinlichkeit $P(D|+)$ ist die Wahrscheinlichkeit, dass man bei einem positiven Test auch wirklich krank ist. Diese beträgt aber nur 8 %. Man hat bei einem positiven Test also nur zu 8 % auch wirklich die Krankheit. Ein positiver Test sagt hier also sehr wenig darüber aus, ob man die Krankheit hat oder nicht.

□

Die Lektion ist: Bei bedingten Wahrscheinlichkeiten darf man seiner Intuition nicht trauen, sondern muss die entsprechenden Wahrscheinlichkeiten wirklich ausrechnen.

Bayes Theorem

Zwischen $P(A|B)$ und $P(B|A)$ gibt es folgenden Zusammenhang:

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

Die letzte Gleichung erhalten wir wegen

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

Beispiel 3.5.3

Das Bayes Theorem liefert die gleiche Lösung wie unsere obige Rechnung:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.9 \cdot (0.009 + 0.001)}{0.009 + 0.099} = \frac{0.009}{0.009 + 0.099} = 0.08.$$

□

Ein weiterer nützlicher Begriff ist die *totale Wahrscheinlichkeit*. Dabei wird eine Menge A zunächst in Mengen A_1, \dots, A_k unterteilt, die miteinander keine Schnittmenge haben und zusammen (Vereinigung) die ganze Menge A bilden. Eine solche Aufteilung nennen wir eine **Partitionierung**.

Beispiel 3.5.4

Für den Würfelwurf ist folgende Partitionierung möglich:

$$A_1 = \{1\}, \quad A_2 = \{2, 4\}, \quad A_3 = \{3, 5, 6\}$$

Es gilt also

$$A_1 \cap A_2 = \emptyset; \quad A_1 \cap A_3 = \emptyset; \quad A_2 \cap A_3 = \emptyset$$

und

$$A_1 \cup A_2 \cup A_3 = A$$

□

Es gilt dann das

Gesetz der totalen Wahrscheinlichkeit

Für die Partitionierung A_1, \dots, A_k und für jedes beliebige Ereignis B gilt:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_k)P(A_k) = \sum_{i=1}^k P(B|A_i)P(A_i)$$

Das Gesetz der totalen Wahrscheinlichkeit lässt sich mit Hilfe eines Wahrscheinlichkeitsbaumes sehr gut verstehen (Abbildung 3.9): Ein Ereignis B trete stets in Verbindung mit genau einem der sich paarweise ausschliessenden Ereignisse A_i ($i = 1, 2, \dots, n$) auf, d.h. die Ereignisse A_i sind die möglichen Zwischenstationen auf dem Weg zum Ereignis B .

Beispiel 3.5.5

Ich teile meine Emails in drei Kategorien ein:

$$A_1 : \text{„spam“}, \quad A_2 : \text{„niedrige Priorität“}, \quad A_3 : \text{„hohe Priorität“}$$

Aus früheren Beobachtungen weiss ich, dass

$$P(A_1) = 0.7, \quad P(A_2) = 0.2, \quad \text{und} \quad P(A_3) = 0.1$$

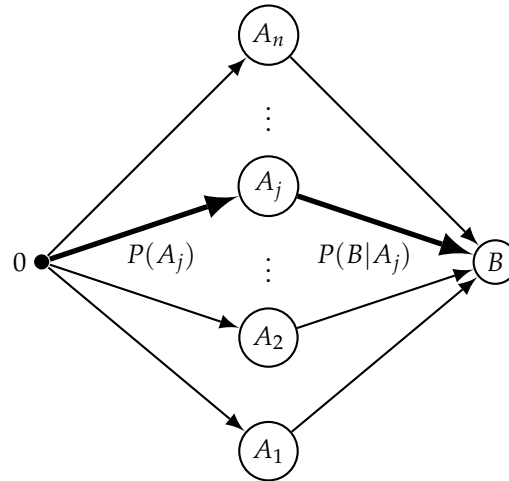


Abbildung 3.9.: Um die Wahrscheinlichkeit $P(B)$ zu berechnen, wird über alle $P(B|A_j) \cdot P(A_j)$ summiert.

Es gilt

$$P(A_1) + P(A_2) + P(A_3) = 1$$

wie es bei einer Partitionierung auch sein sollte.

Sei B das Ereignis, dass das Wort „free“ in der Email auftaucht. Dieses Wort kommt sehr oft in Spam-Mails vor, aber auch in den anderen. Von früheren Beobachtungen weiss ich, dass

$$P(B|A_1) = 0.9, \quad P(B|A_2) = 0.01, \quad \text{und} \quad P(B|A_3) = 0.01$$

Hier ergibt die Summe nicht 1. Dies sind die Wahrscheinlichkeiten, mit denen das Wort „free“ in den drei Mailkategorien vorkommt. Angenommen, ich erhalte eine Email, die das Wort „free“ enthält. Wie gross ist die Wahrscheinlichkeit, dass es sich um Spam handelt? Das Bayes Theorem zusammen mit dem Gesetz der totalen Wahrscheinlichkeit liefert die Lösung:

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{0.9 \cdot 0.7}{(0.9 \cdot 0.7) + (0.01 \cdot 0.2) + (0.01 \cdot 0.1)} \\ &= 0.995 \end{aligned}$$

Viele Spamfilter basieren tatsächlich auf diesem Prinzip: Die Mails werden nach Worten wie „free“, „credit“, etc. durchsucht, die häufig in Spam-Mails vorkommen, in den anderen aber eher nicht.

□

3.6. Zufallsvariable

Der Begriff der *Zufallsvariable* spielt eine ganz zentrale Rolle in der Statistik. Wir beginnen mit einem einfachen Beispiel. Ein Pack Jasskarten besteht aus 36 verschiedenen Karten. Damit wir beim Jassen einen Sieger ermitteln können, werden den Jasskarten Zahlenwerte zugeordnet. So hat ein König den Wert 4. Ohne diese Werte wären verschiedene Stiche beim Jass sehr schwierig miteinander zu vergleichen. Wir haben also eine *Funktion*, die jeder Jasskarte einen Zahlwert zuordnet. Dieselbe Situation haben wir in der Stochastik. Oft wird ein Zufallsexperiment mit Zahlenwerten verknüpft. Zu jedem Elementarereignis ω gehört demnach ein Zahlenwert $X(\omega) = x$. Dabei ist X eine *Funktion*, die jedem Elementarereignis ω den Zahlenwert x zuordnet.

Beispiel 3.6.1

Wir ziehen Karten aus einem Stapel Jasskarten. Dabei wird jeder Karte ein Wert zugeordnet.

$$\begin{array}{ll} \omega = \text{As} & \mapsto X(\omega) = 11 \\ \omega = \text{König} & \mapsto X(\omega) = 4 \\ \vdots & \vdots \\ \omega = \text{Sechs} & \mapsto X(\omega) = 0 \end{array}$$

□

Wie man in diesem Beispiel sieht, ist X eine Funktion auf dem Grundraum Ω . Diese Funktion wird *Zufallsvariable* genannt. Sie ordnet jedem Element des Grundraumes eine Zahl zu. Der Vorteil in diesem Vorgehen liegt darin, dass wir mit den Werten der Zufallsvariable rechnen können.

So ist es im Beispiel oben möglich, mit den Zahlenwerten $X(\omega)$ den „Durchschnitt“ (den sogenannten *Erwartungswert*, der später in diesem Kapitel behandelt wird) der gezogenen Karten zu berechnen. Für die Elementarereignisse „As“, „König“ etc. macht das Wort „Durchschnitt“ keinen Sinn.

Wir halten fest:

Zufallsvariable

Eine **Zufallsvariable** X ist eine **Funktion**:

$$\begin{aligned} X : \quad \Omega &\rightarrow \mathbb{R} \\ \omega &\mapsto X(\omega), \end{aligned}$$

Die Notation X (oder auch Y, Z, \dots) ist eher ungewohnt für die Bezeichnung einer Funktion, ist aber üblich in der Wahrscheinlichkeitsrechnung.

Konventionen

Eine Zufallsvariable wird mit einem **Grossbuchstaben** X (oder Y, Z) bezeichnet. Der entsprechende **Kleinbuchstabe** x (oder y, z) stellt einen konkreten Wert dar, den die Zufallsvariable annehmen kann. Für das Ereignis, bei dem die Zufallsvariable X den Wert x annimmt, schreiben wir $X = x$.

Bei einer Zufallsvariable ist nicht die Funktion $X(\cdot)$ zufällig, sondern nur das Argument ω : Je nach Ausgang des Zufallsexperiments ω erhalten wir einen anderen Wert $x = X(\omega)$. Wir nennen dann x eine **Realisierung** der Zufallsvariablen X . Falls das Experiment zweimal durchgeführt wird, und wir zweimal das gleiche Ergebnis ω erhalten, dann sind auch die realisierten Werte von X gleich.

Beispiel 3.6.2

Im Jasskartenbeispiel entspricht die Realisierung $X = 11$ dem Ziehen eines Asses.

□

In diesem Kapitel sind die Zahlen, die X annehmen kann, *diskret*. D.h. die Anzahl dieser Werte ist endlich wie im Jasskartenbeispiel

$$\{0, 2, 3, 4, 10, 11\}$$

oder sie ist eine unendliche Liste wie

$$\{2.5, 4.5, 6.5, 8.5, \dots\}$$

Wir sagen in diesem Fall, dass die Zufallsvariable X *diskret* ist. Insbesondere sind Anzahlen stets diskret, während Messungen besser als kontinuierliche Zahlen, also mit \mathbb{R} , modelliert werden, obwohl man praktisch nur mit endlicher Genauigkeit messen kann.

Wahrscheinlichkeitsverteilung einer Zufallsvariablen

Wir haben schon gesehen, wie wir die Wahrscheinlichkeit $P(E)$ eines Ereignisses E berechnen. Entsprechend können wir auch die Wahrscheinlichkeit einer allgemeinen Realisierung x einer Zufallsvariable X definieren. Dazu zunächst folgendes Beispiel.

Beispiel 3.6.3

Wir definieren die Zufallsvariable X als den Wert einer gezogenen Jasskarte und fragen nun, wie gross die Wahrscheinlichkeit ist, dass die gezogene Karte den Wert 4 hat. Die Realisierung ist $X = 4$. Die zugehörige Wahrscheinlichkeit bezeichnen wir mit

$$P(X = 4)$$

Die Realisation $X = 4$ entspricht dem Ziehen eines Königs. D.h., wir suchen die Wahrscheinlichkeit, dass ein König gezogen wird:

$$\begin{aligned} P(X = 4) &= P(\{\omega \mid \omega = \text{ein König}\}) \\ &= P(\text{Eicheln-König}) + P(\text{Rosen-König}) + \\ &\quad + P(\text{Schellen-König}) + P(\text{Schilten-König}) \\ &= \frac{4}{36} = \frac{1}{9} \end{aligned}$$

Das Vorgehen hier ist ausführlicher ausgeführt als unbedingt notwendig, aber es lässt sich auf nicht so einfache Beispiele verallgemeinern.

□

Die Wahrscheinlichkeit, dass ein König gezogen wird, ist also gleich der Summe der Wahrscheinlichkeiten, die verschiedenen Könige zu ziehen. Diese Überlegung können wir verallgemeinern.

Die Werte einer Zufallsvariablen X (die möglichen Realisierungen von X) treten mit gewissen Wahrscheinlichkeiten auf. Die Wahrscheinlichkeit, dass X den Wert x annimmt, berechnet sich wie folgt:

$$P(X = x) = P(\{\omega \mid X(\omega) = x\}) = \sum_{\omega; X(\omega)=x} P(\omega)$$

Im Jasskartenbeispiel ist $x = 4$ und ω alle möglichen Könige, deren entsprechende Wahrscheinlichkeiten aufaddiert werden.

Wir haben im Beispiel vorher die Wahrscheinlichkeit *einer* Realisierung berechnet. Natürlich können wir die Wahrscheinlichkeiten *aller* Realisierungen berechnen. Dies führt uns auf den *sehr wichtigen* Begriff einer Wahrscheinlichkeitsverteilung.

Wahrscheinlichkeitsverteilung

Berechnen wir für *jede* Realisierung einer Zufallsvariable die zugehörige Wahrscheinlichkeit, so sprechen wir von einer *Wahrscheinlichkeitsverteilung* dieser Zufallsvariablen.

Beispiel 3.6.4

Die Zufallsvariable X ist wieder der Wert einer gezogenen Jasskarte. Die Wahrscheinlichkeit

$$P(X = 4) = \frac{1}{9}$$

haben wir schon berechnet.

Die Wahrscheinlichkeit $P(X = 0)$ können wir mit der Laplace-Wahrscheinlichkeit berechnen. Es hat unter den 36 Karten genau 16 „leere“ Karten. Somit gilt für die Wahrscheinlichkeit $P(X = 0)$:

$$P(X = 0) = \frac{16}{36} = \frac{4}{9}$$

Die Realisierung $X = 2$ entspricht dem Ziehen eines Unders. Da es 4 von denen gibt, gilt für die Wahrscheinlichkeit $P(X = 2)$:

$$P(X = 2) = \frac{4}{36} = \frac{1}{9}$$

Die Wahrscheinlichkeiten für die anderen Realisierungen berechnen wir analog. Wir haben dann *jeder* Realisierung einen Wahrscheinlichkeitswert zugeordnet. Wir sprechen dann eben von einer *Wahrscheinlichkeitsverteilung*. Die Wahrscheinlichkeitsverteilung von X ist in der Tabelle 3.5 aufgeführt. Die Werte für $P(X = 1)$ oder $P(X =$

| x | 0 | 2 | 3 | 4 | 10 | 11 |
|------------|-----|-----|-----|-----|-----|-----|
| $P(X = x)$ | 4/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

Tabelle 3.5.: Wahrscheinlichkeitsverteilung von gezogenen Jasskarten.

178) sind in Tabelle 3.5 nicht aufgeführt. Der Grund dafür ist natürlich, dass diese Werte nicht gezogen werden können. Wir können ihnen aber trotzdem eine Wahrscheinlichkeit zuordnen, nämlich die Zahl 0:

$$P(X = 1) = 0 \quad \text{oder} \quad P(X = 178) = 0$$

Addieren wir alle Werte in unserer Wahrscheinlichkeitsverteilung, so müssen wir 1 erhalten, da wir eine Realisierung ziehen *müssen*.

$$P(X = 0) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 10) + P(X = 11) = 1$$

□

Wahrscheinlichkeitsverteilung

Die „Liste“ von $P(X = x)$ für alle möglichen Werte x heisst diskrete (**Wahrscheinlichkeits-)** **Verteilung** der diskreten Zufallsvariablen X . Dabei gilt immer

$$\sum_{\text{alle möglichen } x} P(X = x) = 1.$$

Wir werden später auch nichtdiskrete Verteilungen kennenlernen.

3.7. Binomialverteilung

In diesem Kapitel betrachten wir die Situation, wo es um das Zählen der Anzahl Erfolge (oder Misserfolge) geht. Solche Anwendungen treten z.B. auf bei der Qualitätskontrolle, Erfolg/Misserfolg bei Behandlungen (medizinisch, biologisch) oder auch bei Glücksspielen. Wir wollen an einem Beispiel auf die Problemstellung eingehen.

Beispiel 3.7.1

Nehmen wir an, Sie werfen eine faire Münze viermal hintereinander. Die Versuchsreihe hat also einen Umfang von $n = 4$. Die Wahrscheinlichkeit, „Kopf“ (K) zu werfen, ist $\pi = 0.5$; entsprechend ist $1 - \pi = 1 - 0.5$ die Wahrscheinlichkeit, „Zahl“ (Z) zu werfen. Wir belassen die Schreibweise $1 - 0.5$, um das Beispiel später zu verallgemeinern.

Mit X bezeichnen wir die Anzahl Würfe mit „Kopf“. Die möglichen Ausgänge dieses Experiments können wir als eine Liste mit „Wörtern“ bestehend aus 4 Buchstaben darstellen (siehe Tabelle 3.6).

Dabei ist z.B.

$$X(ZZZZ) = 0$$

da kein K in $ZZZZ$ vorkommt. Entsprechend ist

$$X(KKZK) = 3$$

| Elementarereignis | x | Elementarereignis | x |
|-------------------|-----|-------------------|-----|
| ZZZZ | 0 | KZZZ | 1 |
| ZZZK | 1 | KZZK | 2 |
| ZZKZ | 1 | KZKZ | 2 |
| ZZKK | 2 | KZKK | 3 |
| ZKZZ | 1 | KKZZ | 2 |
| ZKZK | 2 | KKZK | 3 |
| ZKKZ | 2 | KKKZ | 3 |
| ZKKK | 3 | KKKK | 4 |

Tabelle 3.6.: Wir werfen eine faire Münze 4 mal und die Zufallsvariable X soll die Anzahl Würfe mit „Kopf“ sein.

da 3 K 's in $KKZK$ vorkommen. Das Ereignis mit $X = 2$ besteht aus allen Elementarereignissen, bei denen zwei K 's (2 von 4 Würfeln sind „Kopf“) in den Worten vorkommen. Dieses Ereignis besteht aus folgenden 6 Elementarereignissen

$$\{ZZKK, ZKZK, ZKKZ, KZZK, KZKZ, KKZZ\}$$

Wir nehmen an, dass jeder Münzwurf unabhängig vom vorhergehenden Münzwurf ist. Dann gilt für die Wahrscheinlichkeit, dass das Elementarereignis $KKZZ$ eintritt

$$P(KKZZ) = P(K) \cdot P(K) \cdot P(Z) \cdot P(Z) = (0.5)^2 \cdot (1 - 0.5)^2 = 0.0625$$

Jedes der 6 Elementarereignisse vom Ereignis $X = 2$ kommt mit derselben Wahrscheinlichkeit vor. Darum gilt

$$P(X = 2) = 6 \cdot 0.0625 = 0.375$$

Die Wahrscheinlichkeit, dass wir bei 4 Würfeln zweimal „Kopf“ werfen, ist 0.375 oder 37.5 %.

Führen wir dieselbe Überlegung für $X = 3$ durch, so gibt es 4 Elementarereignisse mit der Wahrscheinlichkeit

$$P(KKKZ) = P(K) \cdot P(K) \cdot P(K) \cdot P(Z) = (0.5)^3 \cdot (1 - 0.5)^1 = 0.0625$$

Also

$$P(X = 3) = 4 \cdot 0.0625 = 0.25$$

Werfen wir allgemein x -mal K , so werfen wir $(4 - x)$ -mal Z . Somit gilt analog zu oben:

$$P(X = x) = (\text{Anzahl Elementarereignisse mit } x \text{ mal Kopf}) \times (0.5)^x \cdot (1 - 0.5)^{4-x}$$

Jetzt stellt sich bloss noch die Frage, wie man im Allgemeinen die Anzahl Elementarereignisse mit x mal Kopf bei 4 Münzwürfen berechnen kann. Ist $x = 2$, so ist dies die Anzahl der Möglichkeiten (Anzahl Elementarereignisse), bei denen zweimal K vorkommt. Dies berechnen wir mit dem sogenannten Binomialkoeffizienten

$$\binom{4}{x} = \frac{4!}{x!(4-x)!}$$

wobei $4! = 4 \cdot 3 \cdot 2 \cdot 1$ („4 Fakultät“).

Mit R berechnen wir den Binomialkoeffizienten mit Hilfe des Befehls

```
choose(4, 2)
```

```
## [1] 6
```

Somit lautet die Wahrscheinlichkeit $P(X = x)$ analog zu dem Beispiel oben:

$$P(X = x) = \binom{4}{x} (0.5)^x \cdot (1 - 0.5)^{4-x}$$

Im Falle von $x = 2$ haben wir $\binom{4}{2} = 4! / (2! \cdot 2!) = 6$.

$$P(X = 2) = \binom{4}{2} (0.5)^2 \cdot (1 - 0.5)^2 = \frac{4!}{2!2!} (0.5)^2 \cdot (1 - 0.5)^2 = 6 \cdot 0.5^4 = 0.375$$

Und für $X = 3$

$$P(X = 3) = \binom{4}{3} (0.5)^3 \cdot (1 - 0.5)^1 = \frac{4!}{3!1!} (0.5)^3 \cdot (1 - 0.5)^1 = 4 \cdot 0.5^4 = 0.25$$

Führen wir unser Experiment oftmals durch, so können wir in einem Viertel der Experimente dreimal K erwarten.

□

Diese Überlegungen verallgemeinern wir nun.

Bernoulli-Verteilung

Hier interessiert uns jetzt nur, ob ein Ereignis eintritt oder nicht. Wir haben es also ausschliesslich mit einer Ja-Nein-Entscheidung zu tun.

Beispiel 3.7.2

- Bei *einem* Münzwurf ist die Frage zum Beispiel, ob das Ergebnis „Kopf“ eintritt. Falls die Münze fair ist, so trifft dieses Ereignis mit der Wahrscheinlichkeit $\pi = 1/2$ ein. Die Wahrscheinlichkeit, „Zahl“ zu werfen, ist dann $1 - \pi = 1/2$.
- Beim Wurf eines fairen Würfels betrachten wir das Ereignis, die Zahl 4 zu werfen. Die Wahrscheinlichkeit, eine 4 zu werfen, ist $\pi = 1/6$. Die Wahrscheinlichkeit, keine 4 zu werfen, ist dann $1 - \pi = 5/6$.

□

Wir können dies mit Hilfe einer Zufallsvariablen formulieren. Dazu setzen wir die Zahl 1 für „Ereignis tritt ein“ (oder „Erfolg haben“) und die Zahl 0 für „Ereignis tritt nicht ein“ (oder „kein Erfolg haben“). Damit haben wir eine Zufallsvariable definiert.

Wir können unsere Wahrscheinlichkeiten im Würfelwurfbeispiel mit Hilfe der Zufallsvariablen formulieren.

$$P(X = 1) = \frac{1}{6} \quad \text{und} \quad P(X = 0) = \frac{5}{6}$$

Wir können dies verallgemeinern.

Die Verteilung einer Zufallsvariable X mit Werten in $W = \{0, 1\}$ kann durch einen einzelnen Parameter π beschrieben werden:

$$P(X = 1) = \pi, \quad P(X = 0) = 1 - \pi, \quad 0 \leq \pi \leq 1.$$

Diese Verteilung heisst **Bernoulli(π)-Verteilung**.

Die Bernoulli-Verteilung ist für sich allein nicht sehr interessant.

Binomialverteilung

Interessanter wird es, wenn wir das Bernoulli-Experiment n Mal wiederholen, also z.B. die Münze n -mal werfen. Der Grundraum Ω besteht dann aus allen „Wörtern“ der Länge n , welche man mit den Buchstaben K (für „Kopf“) und Z (für „Zahl“) schreiben kann. Der Grundraum Ω hat dann 2^n Elemente. Warum ist das so?

Nehmen wir $n = 4$. So gibt es für den 1. Wurf zwei Möglichkeiten, nämlich K und Z. Für jeden weiteren Wurf gibt es natürlich wieder je zwei Möglichkeiten. Die gesamte Anzahl Möglichkeiten besteht dann aus dem Produkt der Möglichkeiten pro Wurf:

$$2 \cdot 2 \cdot 2 \cdot 2 = 2^4 = 16$$

Für ein allgemeines n geht die Überlegung analog.

Wir betrachten die Zufallsvariablen

$$X_i = \begin{cases} 1 & \text{falls } K \text{ im } i\text{-ten Wurf} \\ 0 & \text{falls } Z \text{ im } i\text{-ten Wurf} \end{cases}$$

$$X = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i : \text{Gesamtzahl von Würfeln mit } K$$

Mit dem Wort $KKZK$ wollen wir die Verwendung dieser Zufallsvariablen veranschaulichen. Es ist dann

$$X_1(KKZK) = 1$$

da beim 1. Wurf K geworfen wurde. Entsprechend gilt

$$X_2(KKZK) = 1; \quad X_3(KKZK) = 0; \quad X_4(KKZK) = 1$$

Für die Zufallsvariable X erhalten wir

$$\begin{aligned} X(KKZK) &= X_1(KKZK) + X_2(KKZK) + X_3(KKZK) + X_4(KKZK) \\ &= 1 + 1 + 0 + 1 \\ &= 3 \end{aligned}$$

Es kommt in diesem Wort also dreimal K vor.

Um die Verteilung von X zu bestimmen, müssen wir Wahrscheinlichkeiten auf Ω festlegen. Dazu sollen die Ereignisse $X_i = 1$ (also „ K im i -ten Wurf“) alle die gleiche Wahrscheinlichkeit π haben und unabhängig sein.

Bemerkungen:

- i. Wir müssen hier jeweils klar zwischen den beiden Zufallsvariablen X_i und X unterscheiden. Die Zufallsvariable X_i kann nur die Werte 0 und 1 annehmen, während X Werte zwischen 0 und n annehmen kann. So gilt für $n = 5$ und das Wort $KKZZK$

$$X_4 = 0 \quad \text{und} \quad X = 3$$

das heisst, im 4. Wurf wurde Z geworfen und insgesamt trifft dreimal K ein.

Um $P(X = x)$ zu berechnen, brauchen wir zwei Überlegungen

- Wie gross ist die Wahrscheinlichkeit, dass in einem speziellen Fall (Wort) x mal die 1 und $n - x$ mal die 0 vorkommt?

Nehmen wir wieder $n = 5$ und das Wort $KKKZZ$. Dann ist

$$P(KKKZZ) = \pi^3(1 - \pi)^2$$

- Auf wie viele Arten kann man x K's auf n Plätze anordnen?

Nehmen wir wieder $n = 5$ und $x = 3$:

KKKZZ, KKZKZ, KZKKZ, ZKKKZ, ZKKZK,
ZKZKK, ZZKKK, KZZKK, KKZZK, KZKZK

Dies sind 10 Anordnungen. Allerdings ist diese Möglichkeit des Abzählens nur für sehr kleine n durchführbar. Es gibt aber eine allgemeine Formel. Die Antwort ist gegeben durch den Binomialkoeffizienten²

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

So ist für $n = 5$ und $x = 3$

$$\binom{5}{3} = \frac{5!}{3! \cdot 2!} = \frac{120}{6 \cdot 2} = 10$$

So ist dann für $x = 3$ und $n = 5$ der Ausdruck $P(X = 3)$ die Wahrscheinlichkeit, dass (in diesem Fall) genau dreimal K geworfen wird. Sie berechnet sich durch

$$P(X = 3) = \binom{5}{3} \pi^3 (1 - \pi)^2$$

Bestimmen wir für *alle* Realisierungen von X , also für *jede* mögliche Anzahl K im Wort der Länge n , die Wahrscheinlichkeit $P(X = x)$, so erhalten wir die allgemeine *Binomialverteilung*. Die möglichen Realisierungen sind

$$\{0, 1, 2, \dots, n\}$$

Binomial(n, π)-Verteilung

Eine Zufallsvariable X mit Werten in $\{0, 1, \dots, n\}$ heisst Binomial(n, π)-verteilt, falls

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, \quad x = 0, 1, \dots, n$$

Dabei ist $0 \leq \pi \leq 1$ der Erfolgsparameter der Verteilung.

²Der Binomialkoeffizient wird im Anhang [A.1](#) hergeleitet.

Bemerkungen:

- i. Damit sind wir in der Lage, für jeden Wert von $X = x$ die zugehörige Wahrscheinlichkeit für gegebene n und π zu berechnen.

Wir können diese Wahrscheinlichkeit auch mit \mathbb{R} berechnen. Ist beispielsweise $x = 3$, $n = 5$ und $\pi = 0.5$, so berechnen wir $P(X = 3)$ durch

```
dbinom(3, 5, 0.5)
```

```
## [1] 0.3125
```

Wir können auch alle Wahrscheinlichkeiten der Verteilung auf einmal berechnen

```
dbinom(0:5, 5, 0.5)
```

```
## [1] 0.03125 0.15625 0.31250 0.31250 0.15625 0.03125
```

Analog wie in unserem Beispiel ist die Binomialverteilung angebracht für die Zufallsvariable „Anzahl Erfolge/Misserfolge“ (Eintreten eines bestimmten Ereignis) bei n **unabhängigen** Versuchen. Die Bedingung „unabhängig“ ist wesentlich für die Korrektheit der Binomialverteilung.

Konvention

Wenn man notieren will, dass die Zufallsvariable X einer gewissen Wahrscheinlichkeitsverteilung F folgt, schreibt man abgekürzt: $X \sim F$. Dabei kann F von Parametern abhängen, also z.B. $X \sim F(\theta)$. Wenn also X einer Binomialverteilung mit Parametern n und π folgt, schreibt man abgekürzt $X \sim \text{Binomial}(n, \pi)$ oder einfach nur $X \sim \text{Bin}(n, \pi)$.

Beispiel 3.7.3

Bei einer Losbude steht: „Jedes 5. Los gewinnt!“, d.h., die Gewinnwahrscheinlichkeit ist bei jedem Los $\pi = 0.2$. Nehmen wir weiter an, dass das Ziehen von einem Los keinen Einfluss auf das Ziehen des nächsten Loses hat (z.B. gibt es eine riesige Anzahl Lose und die Lostrommel wird nach jedem Verkauf eines Loses gut gemischt). Wir kaufen 100 Lose und bezeichnen mit X die Anzahl Gewinne unter den 100 Losen. Dann ist X $\text{Binomial}(n = 100, \pi = 0.2)$ verteilt. Abgekürzt: $X \sim \text{Bin}(100, 0.2)$.

Welche Aussagen können wir machen? Zuerst betrachten wir den Plot der Wahrscheinlichkeitsverteilung (\mathbb{R} -Befehl weiter unten). In Abbildung 3.10 sehen wir, dass

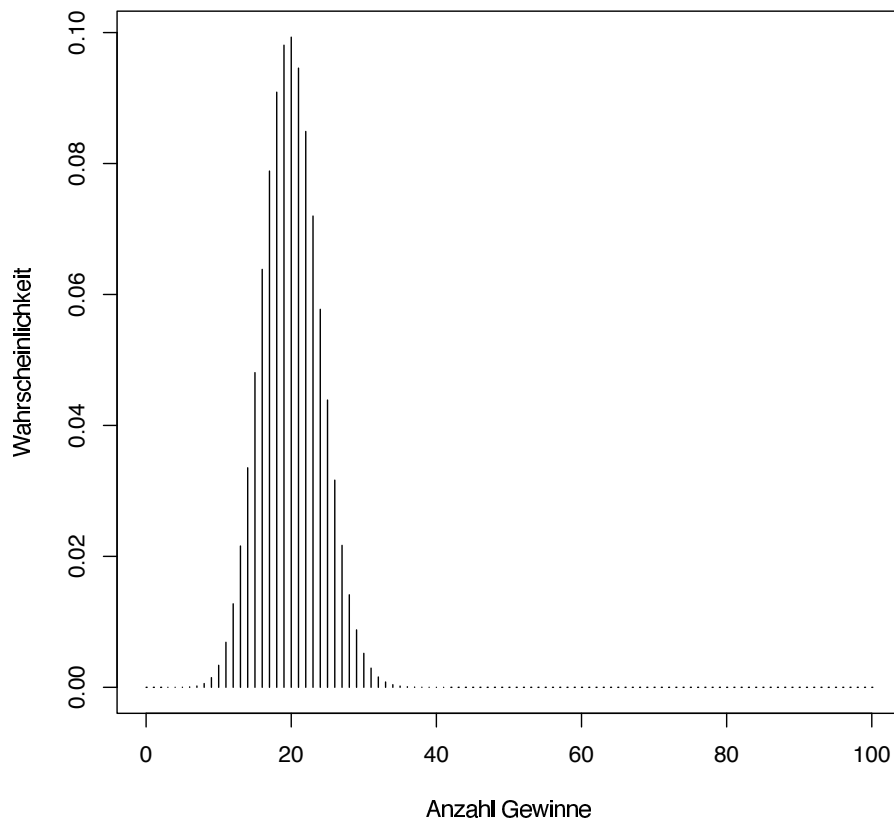


Abbildung 3.10.: Wahrscheinlichkeiten für die Anzahl Gewinne in 100 Losen

es am wahrscheinlichsten ist, etwa 20 Gewinne auf 100 Lose zu ziehen. Das war aber auch zu erwarten, da die Erfolgswahrscheinlichkeit 0.2 ist.

Für $X = 20$ erhalten wir die Wahrscheinlichkeit

$$P(X = 20) = \binom{100}{20} \cdot 0.2^{20} \cdot 0.8^{80} = 0.0993$$

Diesen Wert haben wir natürlich mit R berechnet. Also mit einer Wahrscheinlichkeit von 0.1 gewinnen wir 20-mal beim Kauf von 100 Losen.

Aus [Abbildung 3.10](#) ist auch ersichtlich, dass mit einer Anzahl Gewinne zwischen 15 und 25 zu rechnen ist (was das genau bedeutet, werden wir im nächsten Kapitel

sehen). Wir wollen demnach die Wahrscheinlichkeit von $15 \leq X \leq 25$ berechnen:

$$\begin{aligned} P(15 \leq X \leq 25) &= P(X = 15) + P(X = 16) + \dots + P(X = 25) \\ &= \binom{100}{15} \cdot 0.2^{15} \cdot 0.8^{85} + \binom{100}{16} \cdot 0.2^{16} \cdot 0.8^{84} + \dots + \binom{100}{25} \cdot 0.2^{25} \cdot 0.8^{75} \\ &= 0.83 \end{aligned}$$

Auch dies wurde mit R berechnet:

```
sum(dbinom(15:25, 100, 0.2))

## [1] 0.8320809
```

Mit über 80 % Wahrscheinlichkeit haben wir zwischen 15 und 25 Gewinnen beim Kauf von 100 Losen.

In Abbildung 3.10 sehen wir auch, dass die Wahrscheinlichkeit, 35 und mehr Gewinne zu ziehen, sehr klein ist. Die Rechnung bestätigt diesen Eindruck

$$P(35 \leq X \leq 100) = 0.00034$$

Wir wären ein ziemlich grosser Glückspilz, wenn wir 35 oder mehr Gewinne hätten. Andererseits ist man ein Pechvogel, wenn man weniger als 10 Gewinne zieht

$$P(0 \leq X \leq 9) = 0.0023$$

□

Eigenschaften der Binomialverteilung (siehe Abb. 3.11):

- $P(X = x)$ ist maximal, wenn x gleich dem ganzzahligen Teil von $(n + 1)\pi$ ist, und auf beiden Seiten von diesem Wert nehmen die Wahrscheinlichkeiten monoton ab.
- Wenn $n\pi(1 - \pi)$ nicht allzu klein ist, ist die Verteilung praktisch symmetrisch und hat die Form einer Glocke.
- Wenn n gross ist, sind die meisten Wahrscheinlichkeiten $P(X = x)$ verschwindend klein, d.h. grosse Abweichungen von $(n + 1)\pi$ sind extrem unwahrscheinlich.

Mit R können die sogenannten *Stabdiagramme* aus Abbildung 3.11 wie folgt erzeugt werden:

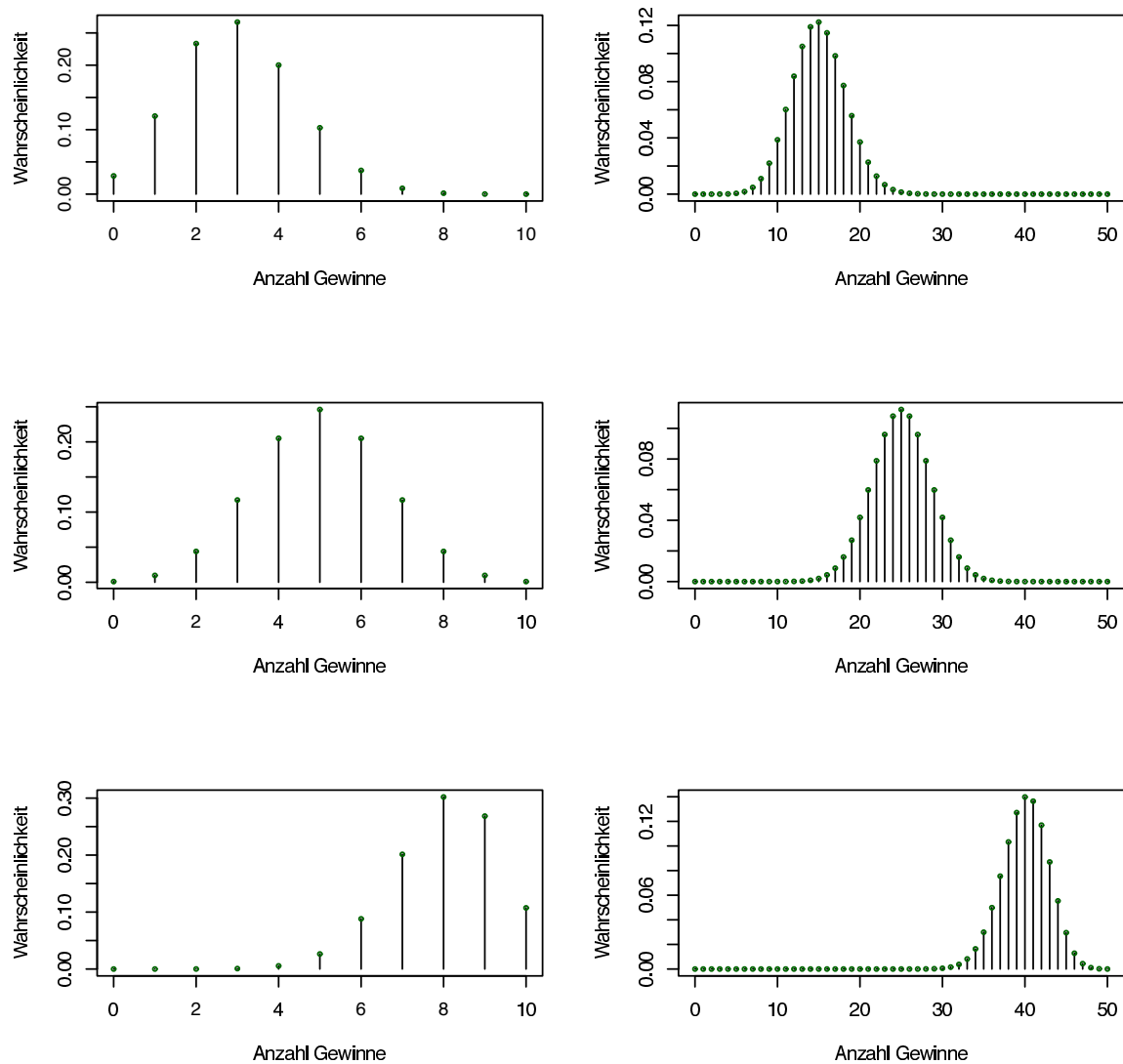


Abbildung 3.11.: Die Binomialwahrscheinlichkeiten $P(X = x)$ wird als Funktion von x für verschiedene Werte von n und π in Form eines Stabsdiagramms dargestellt. Links ist $n = 10$ und rechts ist $n = 50$ für jeweils $\pi = 0.3, 0.5, 0.8$ (von oben nach unten).

```
plot(0:10, dbinom(0:10, size = 10, prob = 0.3), ylim = c(0,
  0.3), xlim = c(0, 100), type = "h", xlab = "Anzahl Gewinne",
  ylab = "Wahrscheinlichkeit")

points(0:10, dbinom(0:10, size = 10, prob = 0.3), cex = 0.5,
  col = "darkgreen")
```

Beispiel 3.7.4

Ein Analog-Digital-Wandler oder *A/D-Wandler* ist ein elektronisches Gerät oder Bauteil zur Umsetzung analoger Eingangssignale in digitale Daten bzw. in einen Datenstrom, der dann weiterverarbeitet oder gespeichert werden kann. Analog-Digital-Wandler sind elementare Bestandteile fast aller Geräte der modernen Kommunikations- und Unterhaltungselektronik wie beispielsweise Mobiltelefonen, Digitalkameras, oder Camcordern. Zudem werden sie in der Messwerterfassung in industriellen Anwendungen, Maschinen und in technischen Alltagsgegenständen wie Autos oder Haushaltsgeräten eingesetzt.

Nach der A/D-Wandlung eines analogen Signals stehen die Daten als binäre Rechteckfolgen direkt für eine weitere Verarbeitung, Speicherung oder Übertragung zur Verfügung. Während in Verarbeitungskomponenten im allgemeinen alle Bits parallel vorliegen, wird insbesondere bei Übertragungen digitaler Signale eine serielle Darstellung der Bits hintereinander in verschachtelter Form bevorzugt.

Die Übertragung von digitalen Signalen ist häufig fehleranfällig. Wir nehmen an, dass die n Übertragungsversuche unabhängig voneinander sind. Wir bezeichnen mit X die Anzahl fehlerhaft übertragener Bits. Wir testen ein neues Material für den Übertragungskanal und stellen fest, dass von 100 übertragenen Bits 6 fehlerhaft empfangen werden. Ein vernünftiges Modell ist dann:

$$X \sim \text{Binomial}(100, \pi),$$

wobei π unbekannt ist. Effektiv beobachtet wurden $x = 6$ fehlerhaft übertragene Bits d.h. $X = x = 6$ wurde tatsächlich **realisiert**. Fortsetzung folgt.

□

3.8. Kennzahlen einer Verteilung

Eine beliebige (diskrete) Verteilung kann vereinfachend zusammengefasst werden durch 2 Kennzahlen, den *Erwartungswert* $E(X)$ und die *Standardabweichung* $\sigma(X)$.

Der **Erwartungswert** einer diskreten Zufallsvariable X beschreibt die mittlere Lage der Verteilung und ist wie folgt definiert:

$$E(X) = \sum_x xP(X = x)$$

Die **Standardabweichung** beschreibt die Streuung der Verteilung. Rechnerisch

ist das Quadrat der Standardabweichung, die sogenannte **Varianz** bequemer:

$$\begin{aligned}\text{Var}(X) &= \sum_x (x - E(X))^2 P(X = x) \\ \sigma(X) &= \sqrt{\text{Var}(X)}.\end{aligned}$$

Der Erwartungswert einer diskreten Zufallsvariable ist das gewichtete arithmetische Mittel von allen möglichen Werten, wobei die Werte mit ihrer Wahrscheinlichkeit gewichtet werden.

Ebenso wird bei der Varianz das Quadrat der Abweichung eines Wertes der Zufallsvariable vom Erwartungswert mit der Wahrscheinlichkeit des Wertes gewichtet. Die Standardabweichung hat dieselbe Einheit wie X , während die Einheit der Varianz deren Quadrat ist: Wird z. B. X in Metern (m) gemessen, so besitzt $\text{Var}(X)$ die Dimension Quadratmeter (m^2) und $\sigma(X)$ wiederum die Dimension Meter (m).

Beispiel 3.8.1

Wir betrachten wiederum das Jasskartenbeispiel mit der Verteilung

| x | 0 | 2 | 3 | 4 | 10 | 11 |
|------------|-----|-----|-----|-----|-----|-----|
| $P(X = x)$ | 4/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

Wir ziehen aus dem Stapel eine Karte. Welches ist der durchschnittliche Wert der Karte, die wir ziehen?

Dazu berechnen wir den Erwartungswert $E(X)$:

$$E(X) = 0 \cdot \frac{4}{9} + 2 \cdot \frac{1}{9} + 3 \cdot \frac{1}{9} + 4 \cdot \frac{1}{9} + 10 \cdot \frac{1}{9} + 11 \cdot \frac{1}{9} = 3.33$$

Dies ist der durchschnittliche Wert, den wir erwarten können, wenn wir sehr oft eine Karte ziehen und diese wieder in den Stapel zurücklegen.

Wir berechnen noch die Varianz und die Standardabweichung:

$$\begin{aligned}\text{Var}(X) &= (0 - 3.33)^2 \cdot \frac{4}{9} + (2 - 3.33)^2 \cdot \frac{1}{9} + (3 - 3.33)^2 \cdot \frac{1}{9} \\ &\quad + (4 - 3.33)^2 \cdot \frac{1}{9} + (10 - 3.33)^2 \cdot \frac{1}{9} + (11 - 3.33)^2 \cdot \frac{1}{9} \\ &= 16.67\end{aligned}$$

und

$$\sigma(X) = \sqrt{16.67} = 4.08$$

Das ist die „mittlere“ Abweichung vom Mittelwert, wobei diese vor allem nach oben möglich ist.

□

Beispiel 3.8.2

Sei $X \sim \text{Bernoulli}(\pi)$. Dann gilt

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = \pi$$

$$\begin{aligned} \text{Var}(X) &= (0 - E(X))^2 P(X = 0) + (1 - E(X))^2 P(X = 1) \\ &= (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi \\ &= \pi(1 - \pi) \end{aligned}$$

$$\sigma(X) = \sqrt{\pi(1 - \pi)}$$

□

Rechenregeln für Erwartungswerte und Varianz

$$E(a + bX) = a + b \cdot E(X) \quad a, b \in \mathbb{R}$$

$$\text{Var}(X) = E(X^2) - E(X)^2$$

$$\text{Var}(a + bX) = b^2 \text{Var}(X) \quad a, b \in \mathbb{R}$$

$$\text{Var}(a) = 0 \quad a \in \mathbb{R}$$

Für die Binomial(n, π)-Verteilung erhält man mit einigen Rechnungen³

$$E(X) = \sum_{x \in W_X} x P(X = x) = \sum_{x \in W_X}^n x \cdot \frac{n!}{x!(n-x)!} \pi^x (1 - \pi)^{n-x} = n\pi$$

und

$$\text{Var}(X) = \sum_{x \in W_X} (x - E(X))^2 P(X = x) = E(X^2) - E(X)^2 = n\pi(1 - \pi)$$

und somit

$$\sigma(X) = \sqrt{n\pi(1 - \pi)}$$

³Die Rechnungen werden im Anhang A.2 ausgeführt.

Beachten Sie, dass aufgrund von $\text{Bernoulli}(\pi) = \text{Binomial}(1, \pi)$ diese Formeln mit dem Resultat des obigen Beispiels übereinstimmen.

Die Kennzahlen fassen also sehr gut zusammen, was wir in der Abbildung 3.11 gesehen haben: Die Verteilung ist um den Erwartungswert konzentriert, die Streuung wächst mit n , aber langsamer als n . Für festes n ist die Streuung maximal, wenn $\pi = 1/2$.

Beispiel 3.8.3

Wir sind wieder bei der Losbude, bei der wir (nach dem vierten Bier) 100 Lose gekauft hatten. Um die Freundin zu beeindrucken, kramen wir unser Statistikwissen hervor und berechnen im Kopf den Erwartungswert und die Standardabweichung der Anzahl Gewinne unter 100 Losen.

$$E(X) = n \cdot \pi = 100 \cdot 0.2 = 20$$

$$\sigma(X) = \sqrt{n\pi(1 - \pi)} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4$$

Wir können also in 100 Losen im Mittel 20 Gewinne erwarten. Die mittlere Abweichung von diesen 20 Gewinnen ist 4. Das stimmt auch in etwa mit unserer Einschätzung aus Abbildung 3.10 auf Seite 85 überein. Wir erinnern uns, dass Beobachtungen typischerweise ein bis zwei Standardabweichungen vom Erwartungswert entfernt liegen und prophezeien der Freundin mit Stolz geschwellter Brust, dass wir wohl zwischen 16 und 24 Gewinnen zu erwarten haben. Sie solle sich schon mal einen Teddybär aussuchen.

Hundert nervenaufreibende Öffnungsversuche später stehen wir mit nur 8 Gewinnen und hängenden Schultern da. Kann das Zufall sein? Ja, aber wir beobachten ein sehr unwahrscheinliches Ereignis (später werden wir solche Überlegungen mit einem statistischen Test präzise formulieren können). Wir beschliessen, den hühnenhaften und grimmig drein schauenden Losbudenbesitzer nicht mit unserer Erkenntnis zu verärgern und denken uns: „Pech im Spiel, Glück in der Liebe!“

□

3.9. Kumulative Verteilungsfunktion

Manchmal ist es für Rechnungen nützlicher, statt der „Liste“ $P(X = x)$ (für alle x) die sukzessiven Summen

$$\sum_{y \leq x} P(X = y) = P(X \leq x)$$

anzugeben. Dabei läuft x ebenfalls über alle Werte von X .

Kumulative Verteilungsfunktion

Man kann in dieser Definition aber auch beliebige reelle Werte x betrachten und erhält dann die sogenannte **kumulative Verteilungsfunktion**

$$F(x) = P(X \leq x) = \sum_{y \leq x} P(X = y),$$

Diese springt an den Stellen, wo X realisiert wird und ist dazwischen konstant. Siehe auch Abbildung 3.12.

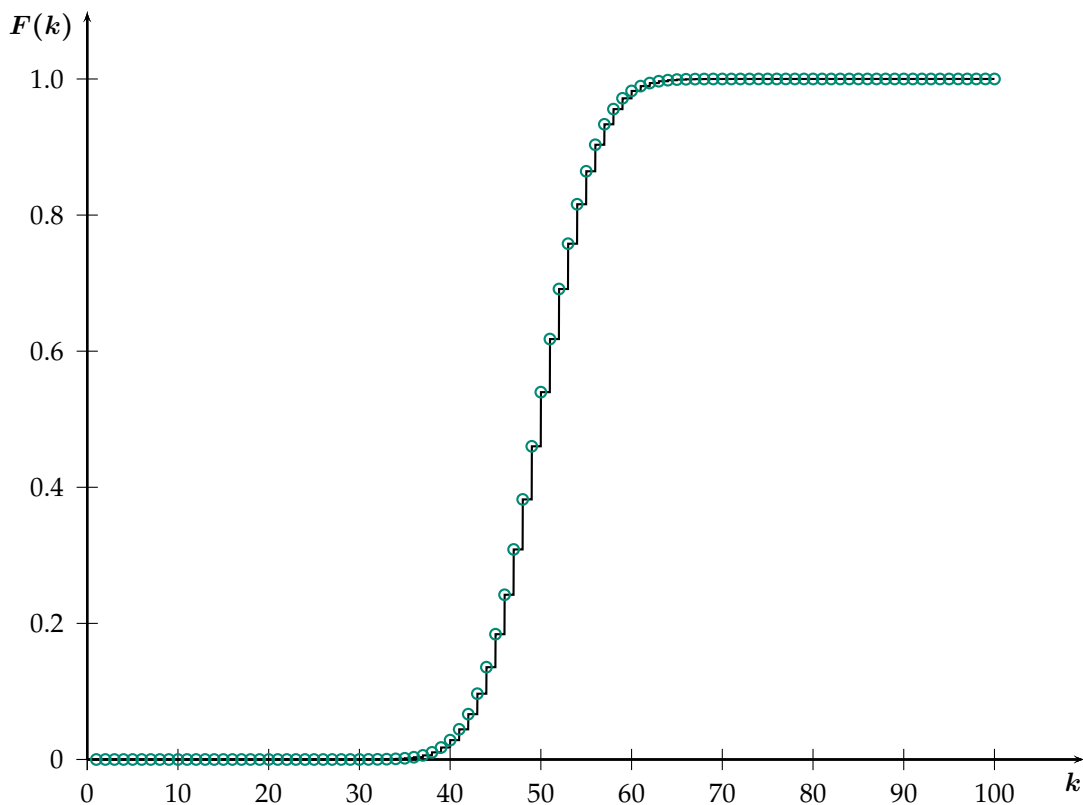


Abbildung 3.12.: Kumulative Verteilungsfunktion $F(\cdot)$ für $X \sim \text{Binomial}(100, 0.5)$. Die Kreise zeigen an, dass an den Sprungstellen der obere Wert gilt.

Angenommen die Zufallsvariable X folgt der Verteilung $\text{Binomial}(100, 0.5)$, dann lässt sich die Wahrscheinlichkeit $P(X \leq 50)$ mit R wie folgt berechnen:

```
pbinom(50, 100, 0.5)
```

```
## [1] 0.5397946
```

Die kumulative Verteilungsfunktion in Abbildung 3.12 kann mit R wie folgt erzeugt werden

```
plot(pbinom(0:100, size = 101, prob = 0.5), type = "s",
     ylim = c(0, 1), xlab = "k", ylab = "F(k)",
     main = "Binom(100,0.5)")

points(pbinom(0:100, size = 101, prob = 0.5), cex = 0.5,
       col = "forest green")
```

Die Wichtigkeit der kumulativen Verteilungsfunktion wird erst im nächsten Kapitel klar.

3.10. Poisson-Verteilung

Bei der Binomial(n, π)-Verteilung nimmt X den beschränkten Wertebereich

$$W = \{0, 1, \dots, n\}$$

an. Oft ist aber nicht von vorneherein klar, wie viele Versuche/Beobachtungen eintreten. So ist die Anzahl Verkehrsunfälle pro Jahr in der Schweiz zumindest theoretisch jede Zahl möglich. Da wir hier das n nicht kennen, können wir auch nicht die Binomialverteilung anwenden.

Für solche Fälle gibt es aber trotzdem eine Verteilung, die sogenannte *Poisson-Verteilung*⁴

Poisson-Verteilung

Eine Zufallsvariable X mit Werten in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ heisst Poisson(λ)-verteilt, falls

$$P(X = x) = \exp(-\lambda) \frac{\lambda^x}{x!} \quad (x = 0, 1, 2, \dots),$$

wobei $\lambda > 0$ ein Parameter der Verteilung ist.

Die Poisson-Verteilung ist die Standardverteilung für unbeschränkte **Zähldaten**.

Die Poisson(λ)-Verteilung kann bei folgenden Anwendungen als Modell gebraucht werden. Bei jeder ist, zumindest theoretisch, eine beliebig grosse Anzahl der betrachteten Ereignisse möglich.

Beispiel 3.10.1

- Anzahl Schadensmeldungen eines Versicherten pro Jahr,

⁴Eine Herleitung der Poisson-Verteilung findet sich im Anhang A.3.

- Anzahl spontaner Ereignisse in einer Nervenzelle während einer Sekunde via Transmitterfreisetzung an einer Synapse.
- Anzahl Autounfälle pro Tag.

□

Die Kennzahlen sind wie folgt für $X \sim \text{Poisson}(\lambda)$ ⁵:

$$E(X) = \sum_{x \in W_X} x P(X = x) = \sum_{x \in W_X} x \cdot \frac{\lambda^x}{x!} e^{-\lambda} = \lambda$$

und

$$\text{Var}(X) = \sum_{x \in W_X} (x - E(X))^2 P(X = x) = \sum_{x \in W_X} (x - \lambda)^2 P(X = x) = \lambda$$

und somit

$$\sigma(X) = \sqrt{\lambda}$$

Aus dem Erwartungswert $E(X)$ wird auch die Bedeutung des Parameters λ klar: Er ist der zu erwartende Wert von X . Meist ist dieser allerdings unbekannt, und wir müssen ihn *schätzen*. Als Schätzwert nehmen wir den arithmetischen Mittelwert von den beobachteten Werten. Diese Methode wird im nächsten Kapitel behandelt und begründet.

Wir möchten an dieser Stelle bereits bemerken, dass es einen fundamentalen Unterschied zwischen dem *Erwartungswert* und dem (arithmetischen) *Mittelwert* gibt. Der Erwartungswert ist eine theoretische Grösse eines Wahrscheinlichkeitsmodells, während der (arithmetische) Mittelwert aufgrund von praktischen Messungen ermittelt wird.

Möchten Sie zehn zufällige Realisierungen der Zufallsvariablen $X \sim \text{Pois}(\lambda = 5)$ generieren, dann lässt sich dies in R folgendermassen bewerkstelligen

```
rpois(10, lambda = 5)
```

```
##      [1] 4 3 3 6 8 4 4 4 4 3
```

Beispiel 3.10.2

In einem Callcenter treffen erfahrungsgemäss durchschnittlich 350 Telefonanrufe in der Stunde ein. Allerdings ist es möglich, dass manchmal gar kein Anruf kommt und in der nächsten Stunde 10 000. Die Anzahl der möglichen Anrufe ist, theoretisch, nach

⁵Die Rechnungen sind im Anhang [A.4](#) ausgeführt.

oben offen. Wir können hier $\lambda = 350$ annehmen. Dann können wir nach der Wahrscheinlichkeit fragen, dass genau 350 Anrufe eintreffen. Diese ist (mit R: `dpois()`)

$$P(X = 350) = 0.021$$

Diese Wahrscheinlichkeit ist an sich sehr klein, obwohl es sich um die Wahrscheinlichkeit des Erwartungswertes handelt. Allerdings ist es sehr unwahrscheinlich, dass *genau* 350 Anrufe eintreffen. Viel grösser wird die Wahrscheinlichkeit sein, dass *ungefähr* 350 Anrufe eintreffen, sagen wir mal zwischen 340 und 360:

$$P(340 \leq X \leq 360) = 0.425$$

Somit treffen in 42.5 % aller Stunden zwischen 340 und 360 Anrufe ein.

□

3.10.1. Poisson-Approximation der Binomialverteilung

Betrachte $X \sim \text{Binomial}(n, \pi)$ und $Y \sim \text{Poisson}(\lambda)$.

Falls n gross und π klein mit $\lambda = n\pi$, dann:

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \approx P(Y = x) = \exp(-\lambda) \frac{\lambda^x}{x!} \quad (x = 0, 1, \dots, n).$$

Das heisst: für grosse n und kleine π : $\text{Binomial}(n, \pi) \approx \text{Poisson}(\lambda)$ mit $\lambda = n\pi$. Mit anderen Worten: die Poisson-Verteilung kann interpretiert werden als Verteilung für *seltenere Ereignisse bei vielen unabhängigen Versuchen*⁶ (selten für einen einzelnen Fall, die Gesamt-Anzahl kann trotzdem gross sein).

Beispiel 3.10.3 Alpha-Zerfall

Ein radioaktiver Zerfall ist im Prinzip ein Ereignis mit zwei möglichen Ausgängen: Zerfall (Erfolg) oder Nicht-Zerfall (Misserfolg). Da jedoch die Gesamtzahl Atome begrenzt ist, kommt als Wahrscheinlichkeitsmodell die Binomialverteilung in Frage. Da in Bezug auf die Gesamtzahl Atome in einer radioaktiven Substanz (bestehend aus einer sehr grossen Anzahl Atome, also grosses n) ein Alpha-Zerfall ein seltenes Ereignis darstellt (kleines π) und jeder Zerfall unabhängig vom vorhergehenden ist, bietet sich die Poisson-Verteilung zur Beschreibung der Verteilung der Zerfallswahrscheinlichkeiten an. Wir betrachten also die in den 1207 Experimenten beobachtete Anzahl

⁶Die Begründung dafür finden Sie im Anhang A.5.

Zerfälle in 10 Sekunden als Realisierungen von Poisson-verteilten Zufallsvariablen. Jede Realisierung hat die Wahrscheinlichkeit:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Im nächsten Kapitel werden wir uns damit beschäftigen, den Parameter λ aus den Daten zu schätzen.

□

3.10.2. Summen von Poisson-verteilten Zufallsvariablen

Die Poisson-Verteilung hat die folgende Additionseigenschaft:

Wenn $X \sim \text{Poisson}(\lambda_X)$ und $Y \sim \text{Poisson}(\lambda_Y)$ unabhängig sind, dann ist $X + Y \sim \text{Poisson}(\lambda_X + \lambda_Y)$.

Wenn zum Beispiel die Anzahl spontaner Ereignisse in einer Nervenzelle in zwei disjunkten Zeitintervallen Poisson-verteilt und unabhängig sind, dann ist auch das Total wieder Poisson-verteilt. Wir erhalten eine Poisson-Verteilung für alle Intervalle. Weil sich bei der Addition der Zufallsvariablen die Parameter der Poisson-Verteilung addieren, ist üblicherweise λ proportional zur Länge des betrachteten Zeitintervalls.

Oft entstehen Poisson-verteilte Daten von einem sogenannten Poisson-Prozess auf einer Menge S (z.B. Zeit, Fläche oder Raum) mit Parameter λ . Das Modell sagt folgendes aus: Wenn S_1, S_2, \dots, S_n nicht überlappende Teilmengen von S sind (z.B. Zeitintervalle, Teile der Gesamtfläche oder Teile vom Raum), dann sind die Anzahl Ereignisse N_1, N_2, \dots, N_n in jeder Teilmenge unabhängige Zufallsvariablen, die jeweils einer Poisson-Verteilung mit den Parametern $\lambda \cdot |S_1|, \lambda \cdot |S_2|, \dots, \lambda \cdot |S_n|$ folgen. Dabei ist $|S_i|$ die „Grösse“ von der Teilmenge S_i (z.B. Zeitdauer in Sekunden, Fläche in m^2 oder Volumen in m^3).

Beispiel 3.10.4 Telefonanrufe

Angenommen ein Büro erhält Telefonanrufe als ein Poisson-Prozess mit $\lambda = 0.5$ pro Minute⁷. Die Anzahl Anrufe in einem 5 Minuten Intervall folgt dann einer Poisson-Verteilung mit Parameter $\rho = 5 \cdot \lambda = 2.5$. Die Wahrscheinlichkeit, dass es in einem

⁷Beachten Sie, dass hier „pro Minute“ steht. Der Parameter des Poisson-Prozesses misst immer eine Anzahl geteilt durch etwas (z.B. Zeitdauer, Fläche, Volumen, etc.)

5 Minuten Intervall keine Anrufe gibt, ist daher $\exp(-2.5) = 0.082$. Die Wahrscheinlichkeit, dass es genau einen Anruf gibt ist $2.5 \cdot \exp(-2.5)$.

□

Beispiel 3.10.5 Alpha Zerfall

Wir betrachten eine radioaktive Substanz, die aus N Atomen besteht. Wenn die Anzahl Zerfälle der Substanz pro Zeiteinheit Poisson-verteilt ist mit Parameter λ , so ist die Anzahl Zerfälle pro Zeiteinheit und pro Atom ebenfalls Poisson-verteilt, allerdings mit Parameter λ/N .

□

Lernziele

- Sie kennen die drei Axiome der Wahrscheinlichkeitsrechnung.
- Sie kennen den Begriff der Unabhängigkeit und können einfache Rechenaufgaben lösen.
- Sie kennen den Begriff der bedingten Wahrscheinlichkeit und können damit umgehen. Insbesondere kennen Sie den Satz von Bayes und den Satz der totalen Wahrscheinlichkeit. Sie wissen, wie man $P(A|B)$ und $P(B|A)$ mit Formeln in einen Zusammenhang bringen kann und können einfache Aufgaben damit lösen.
- Sie kennen den Begriff der Zufallsvariable, der Wahrscheinlichkeitsverteilung und kumulativen Verteilungsfunktion.
- Sie kennen die Binomial- und die Poissonverteilung.
- Sie kennen die Begriffe Erwartungswert, Varianz und Standardabweichung. Sie können diese Werte für einfache Verteilungen berechnen.

Kapitel 4.

Statistik für Zählraten

What the use of P implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred.

(Jeffreys)

4.1. Drei Grundfragestellungen der Statistik

Wir wollen die drei Grundfragestellungen der Statistik an einem einfachen Beispiel kurz anschauen.

Beispiel 4.1.1 Münzenwurf

Wir nehmen eine Münze, für die die Wurfwahrscheinlichkeit π für „Kopf“ (K) *unbekannt* ist. Weiter gehen wir davon aus, dass es tatsächlich einen wahren Wert für die Wurfwahrscheinlichkeit π gibt, aber uns einfach nicht bekannt ist. Werfen wir die Münze 50 mal, so ist die Zufallsvariable X für die Anzahl Würfe mit „Kopf“ $\text{Bin}(50, \pi)$ -verteilt. Uns interessiert nun der unbekannte Parameter π .

1. Wie können wir π schätzen?

Wir werfen 35 mal „Kopf“, und ein vernünftiger Schätzwert ist wohl

$$\hat{\pi} = \frac{35}{50} = 0.7$$

Den Hut $\hat{\pi}$ über dem π verwenden wir, um diesen Näherungswert vom wahren Wert π zu unterscheiden. Würden wir die Münze weitere 50 mal werfen, ergäbe sich für diesen Schätzwert ein anderes Resultat.

2. Wir gehen davon aus, dass die Münze fair ist. Sind dann diese 35 K auf 50 Würfe nicht sehr unwahrscheinlich? Es müssten durchschnittlich etwa 25 K sein.

Wir werden später sehen, dass diese Beobachtung *unter der Annahme, dass die Münze fair ist*, sehr unwahrscheinlich ist. Wir können Zufall hier also ausschliessen und davon ausgehen, dass die Münze *nicht* fair ist.

3. für welche Werte von π ist unsere Beobachtung noch plausibel?

Der entscheidende Punkt bei all diesen Fragen ist, dass wir den wahren Wert des Parameters π *nicht* kennen. Die Aufgabe der Statistik ist es nun, auf die obigen Fragen so gut wie möglich eine mathematisch fundierte Antwort zu geben.

□

Die Statistik fasst Beobachtungen als **Realisierungen** von Zufallsvariablen auf. Das wäre im Beispiel oben die Anzahl der geworfenen K's. Sie untersucht dann, was für Schlüsse man aus den Beobachtungen über die zu Grunde liegende Verteilung ziehen kann. Meist zieht man nur Verteilungen in Betracht, die bis auf einen (oder wenige) Parameter bestimmt sind und möchte dann Fragen über diesen Parameter beantworten. Im Beispiel oben ist die Verteilung gegeben durch $\text{Bin}(50, \pi)$ und der unbekannte Parameter ist π .

Man unterscheidet die folgenden drei Grundfragestellungen, die den drei Fragestellungen im Beispiel oben entsprechen:

1. **Grundfragestellung:** Welches ist der zu den Beobachtungen plausibelste Parameterwert? Die Antwort auf diese Grundfrage heisst **(Punkt-) Schätzung**.
2. **Grundfragestellung:** Sind die Beobachtungen kompatibel (statistisch vereinbar) mit einem vorgegebenen Parameterwert? Die Antwort auf diese Grundfrage heisst **statistischer Test**.
3. **Grundfragestellung:** Welche Parameterwerte sind mit den Beobachtungen kompatibel (statistisch vereinbar)? Die Antwort auf diese Grundfrage heisst **Konfidenz-** oder **Vertrauensintervall**. Das Konfidenzintervall ist allgemeiner und informativer als ein statistischer Test.

Beispiel 4.1.2

Sei $x = 6$ die effektive Anzahl fehlerhaft übertragener Bits bei der Übertragung von 100 digitalen Signalen (siehe Beispiel 3.7.4 auf Seite 88). Wir fassen $x = 6$ als **Realisierung** einer Zufallsvariablen X (Anzahl fehlerhaften Bit) auf. Wir nehmen an, dass $X \sim \text{Bin}(100, \pi)$ ist.

Wir möchten jetzt von der Beobachtung $x = 6$ Schlüsse auf den *unbekannten* Parameter π ziehen. Die Grundfragen lauten hier:

1. Welches ist der plausibelste Wert π zu der Beobachtung $x = 6$?
2. Ist die Beobachtung $x = 6$ kompatibel mit $\pi = 0.1$ (üblicher Übertragungsfehler von Übertragungskanälen) oder mit $\pi < 0.1$?

Für $\pi = 0.1$ müssten an sich 10 fehlerhafte Bits übertragen werden, es sind aber nur deren 6. Können wir dies dem Zufall zuschreiben, oder müssen wir annehmen, dass der wahre Wert von π kleiner als 0.1 ist?

3. Welcher Bereich (Intervall) für den Parameter π ist mit der Beobachtung $x = 6$ kompatibel?

Wie sieht also der Bereich für π aus, wo wir die Beobachtung $x = 6$ als zufällig annehmen können?

Wir werden alle diese Frage in diesem Kapitel beantworten.

□

4.2. (Punkt-)Schätzung bei Binomialverteilung und Poisson-Verteilung

Es gibt zwei verbreitete Methoden, um Parameter einer Wahrscheinlichkeitsverteilung zu schätzen: die **Momentenmethode** und die **Maximum-Likelihood-Schätzmethode**.

Konvention:

Um einen Schätzwert zu kennzeichnen, wird ein Hut ($\hat{\cdot}$) auf die Variable gesetzt. Z.B. ist $\hat{\pi}$ ein Schätzwert des Parameters π ; $\hat{E}(X)$ ist ein Schätzwert des wahren Erwartungswertes $E(X)$; \hat{y} ist ein Schätzwert der Variable y .

Bemerkungen:

- i. Wir möchten an dieser Stelle noch den Begriff „Schätzung“ klären. Es handelt sich hier *nicht* um eine gefühlsmässige Angabe unsererseits, sondern wir *berechnen* diesen Wert aus unvollständigen Informationen, und zwar mit Hilfe von klar definierten statistischen Methoden. Der geschätzte Wert variiert bei jedem neuen Versuch, der genaue wahre Wert bleibt uns in der Realität immer verborgen.

4.2.1. Momentenmethode bei Binomialverteilung

Wir betrachten folgende Situation: Gegeben ist eine Beobachtung x , welche als Realisierung von $X \sim \text{Bin}(n, \pi)$ aufgefasst wird. Wir möchten Schlüsse über den *unbekannten* Parameter π ziehen.

Als Beispiel dazu dient wieder der Münzwurf aus dem letzten Unterkapitel, wo $X \sim \text{Bin}(50, \pi)$. Eine Schätzung für π kann pragmatisch hergeleitet werden. Für eine Binomialverteilung $\text{Bin}(n, \pi)$ hatten wir den Erwartungswert $E(X)$ bereits berechnet (siehe Unterkapitel 3.8)

$$E(X) = n\pi$$

Damit können wir den Parameter π folgendermassen ausdrücken:

$$\pi = \frac{E(X)}{n}$$

Der Wert n (Anzahl unabhängiger Versuche) wird als bekannt vorausgesetzt: die einzige Unbekannte ist dann $E(X)$. Für eine erste pragmatisch motivierte Schätzung ersetzen wir den Erwartungswert durch die Beobachtung:

$$\hat{E}(X) = x = \text{beobachtete Anzahl Gewinne}$$

Man schätzt den Erwartungswert also durch die Beobachtung.

Momentenmethode

Somit ergibt sich aufgrund der **Momentenmethode** die relative Häufigkeit

$$\hat{\pi} = \frac{x}{n}$$

als Schätzung der Erfolgswahrscheinlichkeit.

Beispiel 4.2.1

Wir haben eine Münze, bei der wir uns fragen, ob sie fair ist oder systematisch eher „Kopf“ ergibt. Um dies zu beantworten, wird die Münze 100-mal geworfen, und wir erhalten 58-mal „Kopf“. Dann betrachten wir die Zufallsvariable $X = \text{Anzahl „Kopf“ (K)}$ bei 100 Würfeln. Es ist vernünftig, das Modell

$$X \sim \text{Bin}(100, \pi)$$

zu benutzen. Beobachtet (realisiert) wurde $x = 58$. Die Wahrscheinlichkeit, dass die Münze bei einem Wurf „Kopf“ zeigt, ist gemäss der Momentenmethode also

$$P(K) = \hat{\pi} = \frac{58}{100} = 0.58$$

□

Beispiel 4.2.2

Im Beispiel Übertragung digitaler Signale ist die geschätzte Erfolgswahrscheinlichkeit für das Verfahren also gleich

$$\hat{\pi} = \frac{6}{100} = 0.06$$

Es leuchtet unmittelbar ein, dass bei einer wiederholten Übertragung von 100 digitalen Signalen nicht wiederum genau 6 Bits fehlerhaft übertragen werden. Die wahre Fehlerrate π ist unbekannt und kann nur geschätzt werden, wobei die geschätzte Fehlerrate $\hat{\pi}$ variiert. Idealerweise gibt man einen Bereich an, in dem die wahre Wahrscheinlichkeit für eine fehlerhafte Übertragung mit grosser Wahrscheinlichkeit liegt. Dazu später mehr im Kapitel über Vertrauensintervalle.

□

In der Regel liegen mehrere Beobachtungen aus verschiedenen Zufallsexperimenten vor. Wie wir in diesem Fall eine Schätzung für den unbekannten Parameter π machen, zeigt folgendes Beispiel:

Beispiel 4.2.3

Wir werfen mit einer Münze und machen dabei zwei Versuche. Im 1. Versuch werfen wir 50-mal und erreichen 30-mal K ; im 2. Versuch sind es 115-mal K auf 160 Würfe. Das heisst wir haben die Schätzungen

$$\hat{\pi}_1 = \frac{30}{50} = 0.6 \quad \text{und} \quad \hat{\pi}_2 = \frac{115}{160} = 0.72$$

Wir können dies auch als grossen Versuch mit 210 Würfeln und 145 Erfolgen K ansehen. Dann ist der geschätzte Parameter

$$\hat{\pi} = \frac{30 + 115}{50 + 160} = \frac{145}{210} = 0.69$$

□

Wir haben m Beobachtungen

$$X_1 = x_1, X_2 = x_2, \dots, X_m = x_m$$

mit den zugehörigen Anzahl Versuchen

$$n_1, n_2, \dots, n_m$$

Wir nehmen an, dass diese Beobachtungen unabhängig voneinander sind. Wir können also davon ausgehen, dass alle Beobachtungen mit dem gleichen unbekannten Parameter π binomialverteilt sind. Deshalb können wir dies als ein grosses Zufallsexperiment auffassen mit

$$X = x = x_1 + x_2 + \dots + x_m \quad \text{und} \quad n = n_1 + n_2 + \dots + n_m$$

Die Schätzung des Parameters π aufgrund der Momentenmethode ist dann gegeben durch

$$\hat{\pi} = \frac{x_1 + x_2 + \dots + x_m}{n_1 + n_2 + \dots + n_m}$$

Dies ist die gesamte Anzahl Erfolge durch die gesamte Anzahl Versuche.

4.2.2. Maximum-Likelihood bei Binomialverteilung

Die bei weitem wichtigste Schätzungsmethode ist die *Maximum-Likelihood-Methode*. Am einfachsten lässt sich die Maximum-Likelihood Methode an einem Beispiel erklären.

Beispiel 4.2.4

Die Anzahl „Kopf“ bei n Münzwürfen ist folgendermassen verteilt:

$$X \sim \text{Bin}(n, \pi)$$

In unserem Beispiel ist $n = 100$, und die Zufallsvariable X hat den Wert 58 angenommen. Die Aufgabe besteht nun darin, einen Wert für π zu finden, der möglichst gut zu unserer Beobachtung $x = 58$ passt. Welches Kriterium könnte man verwenden, um zu zeigen, dass ein Wert π_1 besser zu der Beobachtung $x = 58$ passt als π_2 ?

Eine Möglichkeit ist die folgende: Wir berechnen die Wahrscheinlichkeit, genau 58 mal Kopf bei 100 Münzwürfen zu erzielen. Einmal verwenden wir dabei z.B. $\pi_1 = 0.5$ und das andere mal $\pi_2 = 0.6$ und berechnen die zugehörigen Wahrscheinlichkeiten

$$P_{0.5}(X = 58) = 0.0223 \quad \text{und} \quad P_{0.6}(X = 58) = 0.074$$

Anschliessend wählen wir das π aus, das zu der grösseren Wahrscheinlichkeit für 58 mal Kopf führt. Dies wäre hier die Schätzung

$$\hat{\pi} = 0.6$$



Um den *wahrscheinlichsten* Wert für π zu erhalten, müssten wir natürlich nicht nur zwei Werte von π vergleichen, sondern alle, die denkbar sind. Mit anderen Worten: Wir müssen π so wählen, dass der Ausdruck

$$P[X = 58] = \binom{100}{58} \pi^{58} (1 - \pi)^{42}$$

maximal wird.

Dieses Problem lässt sich an sich leicht mit der Methoden aus dem Grundlagenmodul Mathematik lösen

- Wir müssen diesen Ausdruck nach π ableiten
- Diese Ableitung nach π gleich null setzen
- Diese Gleichung nach π auflösen

$$\frac{d}{d\pi} \left(\binom{n}{x} \pi^x (1 - \pi)^{n-x} \right) = 0 \quad \Rightarrow \quad \pi = \dots$$

Oft ist der Ausdruck, den man ableiten muss, aber recht kompliziert (wie auch in unserem Beispiel). In vielen Fällen kann man dann einen Trick anwenden: Jedes Extremum der Funktion $f(x)$ ist auch ein Extremum der Funktion $\log(f(x))$ und umgekehrt. D.h., anstatt $P[X = x]$ bzgl. π zu maximieren, können wir auch $\log(P[X = x])$ bzgl. π maximieren, falls das leichter geht. Das Ergebnis ist für beide Methoden völlig identisch.

In unserem Fall ist die zweite Variante tatsächlich etwas einfacher. Trotzdem ist folgende Rechnung etwas aufwendig und wird hier nur vollständigkeitshalber aufgeführt:

$$\begin{aligned} \log(P[X = x]) &= \log \left[\binom{n}{x} \pi^x (1 - \pi)^{n-x} \right] \\ &= \log \left[\binom{n}{x} \right] + \log(\pi^x) + \log((1 - \pi)^{n-x}) \\ &= \log \left[\binom{n}{x} \right] + x \cdot \log(\pi) + (n - x) \cdot \log(1 - \pi) \end{aligned}$$

Durch Ableiten nach π und Nullsetzen erhalten wir:

$$\begin{aligned} \frac{d}{d\pi} \left[\log \left(\binom{n}{x} \right) + x \cdot \log(\pi) + (n - x) \cdot \log(1 - \pi) \right] &= 0 \\ \frac{x}{\pi} + (n - x) \cdot \frac{1}{1 - \pi} \cdot (-1) &= 0 \end{aligned}$$

Beachten Sie dabei, dass $\log\left(\binom{n}{x}\right)$ nicht von π abhängt und deshalb beim Ableiten verschwindet. Ausserdem gilt: $\frac{d}{dx} \log(x) = \frac{1}{x}$.

Wenn wir diese Gleichung nach π auflösen, erhalten wir:

$$\frac{x}{\pi} - (n - x) \cdot \frac{1}{1 - \pi} = 0$$

$$\frac{x}{\pi} = \frac{n - x}{1 - \pi}$$

$$x - \pi x = \pi n - \pi x$$

$$\pi = \frac{x}{n}$$

Man erhält folglich als Resultat für die Schätzung von π

$$\hat{\pi} = \frac{x}{n}$$

Mit den Zahlen in unserem Beispiel erhalten wir

$$\hat{\pi} = \frac{58}{100} = 0.58$$

Im Falle der Binomialverteilung ist das Ergebnis der Maximum-Likelihood-Methode identisch mit dem Ergebnis der Momentenmethode. Im Allgemeinen gilt dies aber nicht, wie wir später im Fall von anderen Verteilungen sehen werden.

Mit obiger Methode wählen wir also das π , mit dem die Beobachtung am wahrscheinlichsten ist. Daher nennt man diese Methode **Maximum-Likelihood-Methode**. Sie ist die mit Abstand gebräuchlichste Methode, um Parameter zu schätzen und oft der Momentenmethode überlegen.

Wir sprechen bei der zu maximierenden Funktion auch von der **Likelihood-Funktion** $L(\pi)$.

$$L(\pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

In der Statistik-Software R lässt sich die Parameterschätzung im Falle einer Binomialverteilung mit dem folgenden Befehl durchführen:

```
binom.test(x = 58, n = 100, p = 0.5, alternative = "two.sided",
           conf.level = 0.95)

##
## Exact binomial test
##
```

```
## data: 58 and 100
## number of successes = 58, number of trials =
## 100, p-value = 0.1332
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.4771192 0.6780145
## sample estimates:
## probability of success
## 0.58
```

Bemerkungen:

- i. In der ersten Zeile der Ausgabe (data) werden die Inputdaten erneut aufgelistet.
- ii. In der zweiten Zeile steht die Interpretation der Inputdaten für den Binomialtest: Wir hatten 58 Erfolge bei 100 Versuchen.
- iii. In den letzten drei Zeilen erscheint noch die Schätzung für die Gewinnwahrscheinlichkeit. Wie wir schon vorher gesehen hatten, ist er sowohl für die Momentenmethode also auch für die Maximum-Likelihood-Methode 0.58.
- iv. Die restlichen Zeilen der Ausgabe werden Sie im Verlauf dieses Kapitels kennenlernen.

In der Regel erfolgt die Schätzung aufgrund von mehreren Beobachtungen.

Beispiel 4.2.5

Wir werfen mit einer Münze und machen dabei zwei Versuche. Im 1. Versuch werfen wir 50-mal und erreichen 30-mal K ; im 2. Versuch sind es 115-mal K auf 160 Würfe. Wir haben also zwei Zufallsvariablen X_1 und X_2 . Es gilt

$$X_1 \sim \text{Bin}(50, \pi) \quad \text{und} \quad X_2 \sim \text{Bin}(160, \pi)$$

Wichtig hier ist, dass der Parameter π bei beiden Versuchen identisch ist. Diesen unbekannten Parameter π wollen wir nun mit der Maximum-Likelihood-Methode schätzen.

Aus den Angaben suchen wir die Wahrscheinlichkeit

$$P[(X_1 = 30) \cap (X_2 = 115)]$$

Da wir die beiden Versuche als unabhängig betrachten, gilt

$$\begin{aligned} P[(X_1 = 30) \cap (X_2 = 115)] &= P[X_1 = 30] \cdot P[X_2 = 115] \\ &= \binom{50}{30} \pi^{30} (1 - \pi)^{20} \cdot \binom{160}{115} \pi^{115} (1 - \pi)^{45} \end{aligned}$$

Diese Funktion ist unsere Likelihood-Funktion $L(\pi)$:

$$L(\pi) = \binom{50}{30} \pi^{30} (1 - \pi)^{20} \cdot \binom{160}{115} \pi^{115} (1 - \pi)^{45}$$

Diese Funktion müssen wir nach π ableiten, gleich 0 setzen und nach π auflösen. Dies führen wir hier nicht aus, das Resultat lautet

$$\hat{\pi} = \frac{145}{210}$$

Im allgemeinen Fall für $X_1 = x_1$ und $X_2 = x_2$ mit n_1 bzw. n_2 Versuchen, erhalten wir mit der Maximum-Likelihood-Methode

$$\hat{\pi} = \frac{x_1 + x_2}{n_1 + n_2}$$

Dies ist dasselbe Resultat wie bei der Momentenmethode, was für andere Verteilungen aber nicht der Fall sein muss.

□

Allgemein

Wir haben m Beobachtungen

$$X_1 = x_1, X_2 = x_2, \dots, X_m = x_m$$

Wir nehmen an, dass diese Beobachtungen unabhängig voneinander sind. In diesem Fall lässt sich die Wahrscheinlichkeit, dass sich diese n Beobachtungen (Ereignisse) ereignen, schreiben als

$$\begin{aligned} P[(X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_m = x_m)] &= P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \dots \cdot P[X_m = x_m] \\ &\equiv \prod_{i=1}^m P[X_i = x_i] \end{aligned}$$

Die Wahrscheinlichkeit, dass die n unabhängigen Zufallsvariablen x_1, x_2, \dots, x_n beobachtet werden, hängt also nur vom Parameter π ab. Die dazugehörige *Likelihood-Funktion* $L(\pi)$ lautet also:

$$\begin{aligned} L(\pi) &= P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \dots \cdot P[X_m = x_m] \\ &= \binom{n_1}{x_1} \pi^{x_1} (1 - \pi)^{n_1 - x_1} \cdot \binom{n_2}{x_2} \pi^{x_2} (1 - \pi)^{n_2 - x_2} \cdot \dots \cdot \binom{n_m}{x_m} \pi^{x_m} (1 - \pi)^{n_m - x_m} \end{aligned}$$

Leiten wir diese Funktion wieder ab, setzen diese gleich 0 und lösen nach π auf, so erhalten wir dasselbe Resultat wie bei der Momentenmethode

$$\hat{\pi} = \frac{x_1 + x_2 + \dots + x_m}{n_1 + n_2 + \dots + n_m}$$

Haben wir eine andere Wahrscheinlichkeitsverteilung mit einem Parameter ϑ , so erhalten wir die entsprechende Likelihood-Funktion $L(\vartheta)$. Wir werden im Abschnitt 4.2.4 die Likelihood-Funktion $L(\lambda)$ für die Poissonverteilung kennenlernen.

4.2.3. Momentenmethode bei Poisson-Verteilung

Wir betrachten Daten x_1, x_2, \dots, x_m , die wir als Realisierungen von

$$X_1, \dots, X_m \sim \text{Poisson}(\lambda)$$

auffassen. Es stellt sich dann die Frage, wie wir λ aus den Daten schätzen können. Betrachten wir dazu ein Beispiel:

Beispiel 4.2.6 Krebserregende Fasern

Ein Hersteller für Isolationsmaterialien misst die Anzahl krebserregender Fasern pro mm^2 in 5 Proben. Die gemessenen Werte x_1, \dots, x_5 seien:

$$4, 1, 2, 3, 6$$

Wir modellieren die Anzahl krebserregender Fasern mit einer Poisson-Verteilung, da die mögliche Anzahl Fasern pro mm^2 nicht nach oben begrenzt ist. Das heisst, wir nehmen an, dass die Beobachtungen x_1, \dots, x_5 Realisierungen von

$$X_1, \dots, X_5 \sim \text{Poisson}(\lambda)$$

sind.

Wie sollen wir λ wählen? Was ist unsere Schätzung $\hat{\lambda}$ basierend auf den beobachteten Daten?

Wie wir in Unterkapitel 3.10 gesehen haben, ist der Erwartungswert der Poisson-Verteilung gegeben durch:

$$E(X) = \lambda$$

Wir schätzen nun den Erwartungswert mit dem empirischen Mittelwert

$$\hat{E}(X) = \bar{x} = \frac{4 + 1 + 2 + 3 + 6}{5} = 3.2$$

Dann haben wir auch eine Schätzung für den Parameterwert λ der Poisson-Verteilung

$$\hat{\lambda} = \bar{x} = 3.2$$

□

Momentenmethode für beliebige Wahrscheinlichkeitsverteilungen

Wir fassen unsere Daten x_1, x_2, \dots, x_n als Realisierungen einer Zufallsvariablen X auf.

Falls die Zufallsvariable X einer bekannten Wahrscheinlichkeitsverteilung mit unbekanntem Parameter ϑ folgt, dann berechnet man zuerst den Erwartungswert $E(X)$ und löst die Gleichung nach dem unbekannten Parameter ϑ der Wahrscheinlichkeitsverteilung auf.

Dann ersetzt man den Erwartungswert durch dessen empirisches Gegenstück, den empirischen Mittelwert, und man erhält einen **Momentenschätzer** für den unbekannten Parameter ϑ .

4.2.4. Maximum-Likelihood bei Poisson-Verteilung

Wir betrachten die Maximum-Likelihood-Schätzung für die Poisson-Verteilung aufgrund des Beispiels mit krebserregenden Fasern.

Beispiel 4.2.7 Krebserregende Fasern

Ein Hersteller für Isolationsmaterialien misst die Anzahl krebserregender Fasern pro mm^2 in 5 Proben. Die gemessenen Werte x_1, \dots, x_5 seien:

$$4, 1, 2, 3, 6$$

Wir modellieren die Anzahl krebserregender Fasern mit einer Poisson-Verteilung. Wir nehmen also an, dass die Beobachtungen x_1, \dots, x_5 Realisierungen von

$$X_1, \dots, X_5 \sim \text{Poisson}(\lambda)$$

sind. Wie sollen wir λ wählen? Was ist unsere Schätzung $\hat{\lambda}$ basierend auf den beobachteten Daten?

Für ein gegebenes λ ist die Wahrscheinlichkeit, genau die gemessenen Werte $x_1 = 4, \dots, x_5 = 6$ zu beobachten, gegeben durch

$$= \prod_{i=1}^5 P[X_i = x_i] = P[(X_1 = 4) \cap (X_2 = 1) \cap \dots \cap (X_5 = 6)]$$

Unter der Annahme, dass die Messungen unabhängig voneinander sind, ist die Wahrscheinlichkeit, genau die gemessenen Werte zu beobachten:

$$\begin{aligned} L(\lambda) &= P(X_1 = 4) \cdot P(X_2 = 1) \cdot P(X_3 = 2) \cdot P(X_4 = 3) \cdot P(X_5 = 6) \\ &= e^{-\lambda} \frac{\lambda^4}{4!} \cdot e^{-\lambda} \frac{\lambda^1}{1!} \cdot e^{-\lambda} \frac{\lambda^2}{2!} \cdot e^{-\lambda} \frac{\lambda^3}{3!} \cdot e^{-\lambda} \frac{\lambda^6}{6!} \\ &= \prod_{i=1}^{n=5} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \end{aligned}$$

wobei L die **Likelihood-Funktion** bezeichnet. $L(\lambda)$ ist in der Abbildung 4.1 aufgetragen.

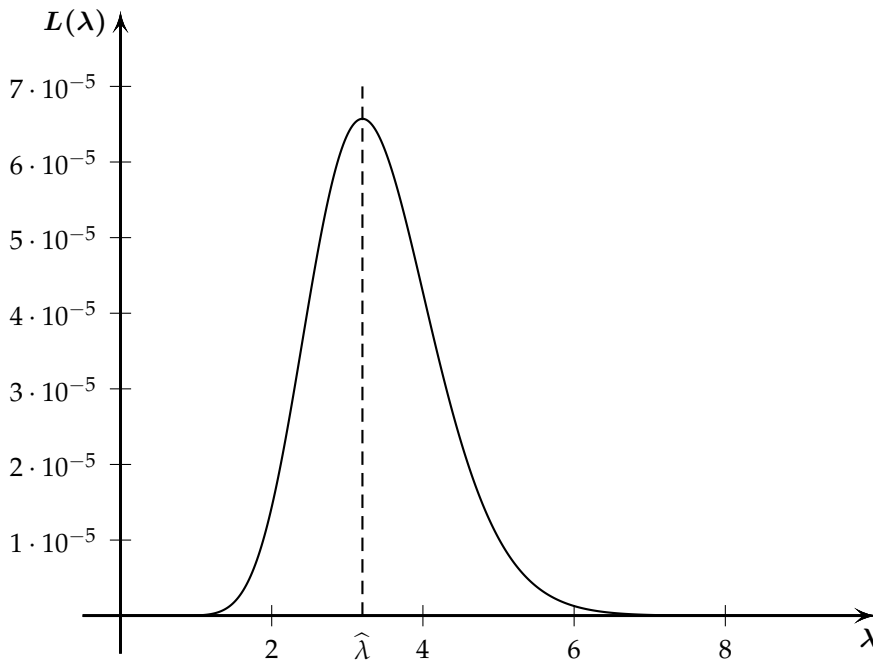


Abbildung 4.1.: Likelihood-Funktion $L(\lambda)$ für die Datenpunkte x_1, \dots, x_5 und $X_i \sim \text{Pois}(\lambda)$.

Für jedes λ können wir also den Wert $L(\lambda)$ berechnen. Wir suchen nun das λ , welches die beobachteten Daten $x_1 = 4, \dots, x_5 = 6$ möglichst plausibel erscheinen lässt. Wir suchen also das λ , welches $L(\lambda)$ *maximiert*, also grösstmöglich macht. Graphisch können wir aus der Graphik herauslesen, dass $L(\lambda)$ ein Maximum an der Stelle $\hat{\lambda} = 3.2$ hat. Schätzen wir λ mit 3.2, also $\hat{\lambda} = 3.2$, so passen die beobachteten Daten nun am besten mit dem Modell $X_i \sim \text{Pois}(\lambda)$ zusammen.

□

Wir können das Maximum der Likelihood-Funktion auch berechnen, indem wir die Likelihood-Funktion nach λ ableiten, gleich 0 setzen und nach λ auflösen. Im folgenden wollen wir genau diese etwas aufwendige Rechnung durchführen, allerdings für den allgemeinen Fall eines Datensatzes x_1, \dots, x_n . Wir fassen die Datenpunkte x_i als Realisierungen der unabhängigen Zufallsvariablen $X_i \sim \text{Pois}(\lambda)$ auf. Die Wahrscheinlichkeit, einen Datenpunkt x_i zu beobachten, ist bei der Poissonverteilung gegeben durch

$$P[X = x_i] = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \quad x_i \in \mathbb{N}_0.$$

Dies führt zur **Likelihood-Funktion**

$$\begin{aligned} L(\lambda) &= e^{-\lambda} \frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \frac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda} \frac{\lambda^{x_n}}{x_n!} \\ &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}. \end{aligned}$$

Es stellt sich auch hier heraus, dass der Logarithmus der Funktion $L(\lambda)$ einfacher abzuleiten ist. Die **log-Likelihood-Funktion** ist somit

$$\begin{aligned} l(\lambda) &= \log(L(\lambda)) \\ &= \sum_{i=1}^n \log \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\ &= \sum_{i=1}^n (x_i \log(\lambda) - \log(x_i!) - \lambda) \\ &= \sum_{i=1}^n (x_i \log(\lambda)) - n\lambda - \sum_{i=1}^n \log(x_i!). \end{aligned}$$

Leitet man $l(\lambda)$ nach λ ab und setzt $l'(\lambda) = 0$, so erhält man die Gleichung

$$\frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0.$$

Dies führt zum Maximum-Likelihood Schätzer

$$\hat{\lambda} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Beispiel 4.2.8 Kanzerogene Fasern

Die log-Likelihood-Funktion $l(\lambda)$ für den Datensatz mit der Anzahl kanzerogener Fasern ist in der Abbildung 4.2 dargestellt. Der Wert für λ , bei dem $l(\lambda)$ maximal wird, liegt nach wie vor bei $\lambda = 3.2$, wie man es aus der Graphik herauslesen kann.

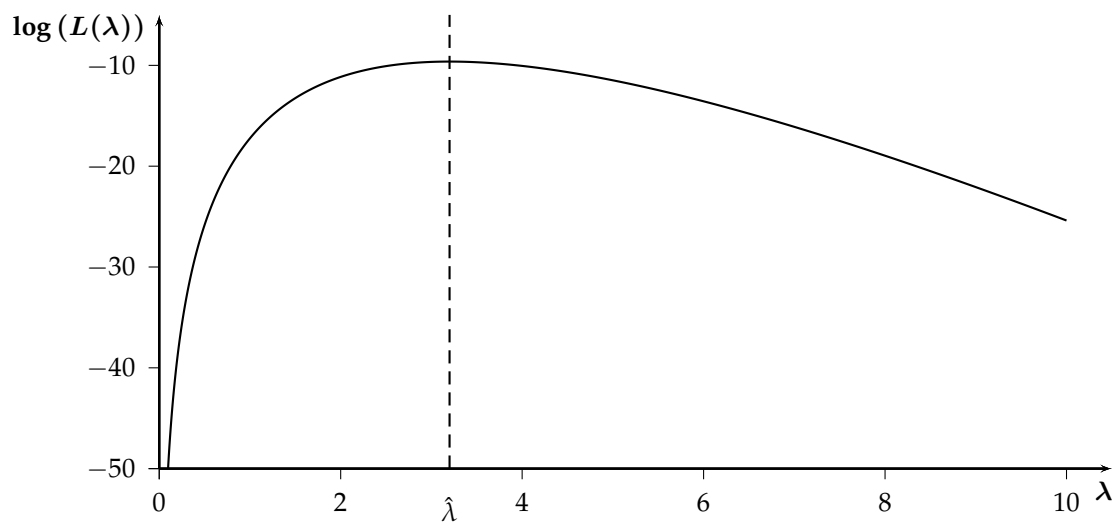


Abbildung 4.2.: Log-Likelihood-Funktion $l(\lambda)$ für die Datenpunkte x_1, \dots, x_5 und $X_i \sim \text{Pois}(\lambda)$.

Bestimmen wir $\hat{\lambda}$ mit Hilfe des Resultats der Maximum-Likelihood-Schätzung, so finden wir

$$\hat{\lambda} = \bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{4 + 1 + 2 + 3 + 6}{5} = 3.2$$

□

Maximum-Likelihood Schätzer

Es liegen m Beobachtungen x_1, x_2, \dots, x_m vor. Sind die Datenpunkte x_i Realisierungen der unabhängigen Zufallsvariablen $X_i \sim \text{Bin}(n, \pi)$, so ergibt der Maximum-Likelihood-Schätzer:

$$\hat{\pi} = \frac{\bar{x}}{n} = \frac{x_1 + x_2 + \dots + x_m}{n + n + \dots + n} = \frac{1}{n \cdot m} \sum_{i=1}^m x_i$$

Sind die Datenpunkte x_i Realisierungen der unabhängigen Zufallsvariablen $X_i \sim \text{Poisson}(\lambda)$, so ergibt der Maximum-Likelihood-Schätzer:

$$\hat{\lambda} = \bar{x} = \frac{x_1 + x_2 + \dots + x_m}{m} = \frac{1}{m} \sum_{i=1}^m x_i$$

4.3. Statistischer Test

4.3.1. Motivation

Um die Grundidee des *statistischen Tests* kennenzulernen, betrachten wir folgendes Beispiel, das als Standardbeispiel dienen soll und an dem Sie sich später bei der Durchführung eines komplizierteren statistischen Test orientieren können.

Beispiel 4.3.1

Bei einem Münzwurf wissen wir in der Regel nicht, ob die Münze *fair* ist oder nicht. Lassen wir uns auf ein Glücksspiel mit einem „Strassenkünstler“ ein, werden wir in der Regel davon ausgehen, dass die Münze unfair ist, aber definitiv wissen wir es nicht. Wie können wir mit statistischen Mitteln feststellen, ob die Münze fair ist oder nicht?

Ist es plausibel, bei einer fairen Münze von 100 Würfeln 58 mal Kopf zu werfen? Dies sind mehr als die zu erwartenden 50 mal. Ist also die wahre, uns unbekannte Wahrscheinlichkeit, „Kopf“ zu werfen, $\pi = 1/2$ oder $\pi > 1/2$? Wir *vermuten* nun, dass die Münze in obigem Versuch nicht fair ist, dass zu viele Male „Kopf“ geworfen wurde. Das Vorgehen in der Statistik ist nun das folgende: wir *nehmen an*, dass die Münze fair ist und versuchen nachzuweisen, dass unter dieser Annahme (der sogenannten *Nullhypothese*), die geworfene Anzahl 58 Köpfe so unwahrscheinlich ist, dass wir Zufall ausschliessen können.

Was dies genau heisst, möchten wir im folgenden erörtern. Zu erwarten sind etwa 50 mal „Kopf“, wenn die Münze fair ist. Wir erachten dann 49 oder 51 beobachtete K als zufällig möglich und würden die Fairness des Würfels in diesem Fall nicht anzweifeln. Wie sieht es aber mit 58 K aus? Liegt dies noch im Bereich, wo wir die Beobachtung dem Zufall zuschreiben können? Oder ist dies so unwahrscheinlich, dass wir nicht mehr annehmen können, dass die Münze fair ist? In diesem Fall würde die Nullhypothese nicht gelten. Wir sagen dann, dass wir unsere Nullhypothese *verwerfen* und von einer unfairen Münze ausgehen. In der Ausdrucksweise der Statistik sprechen wir davon, dass wir die *Alternativhypothese* (Münze ist nicht fair) annehmen.

Wie gross ist nun die Wahrscheinlichkeit, dass unter der Nullhypothese (Münze ist fair) 58 mal in 100 Würfeln „Kopf“ geworfen wird? Für den Münzwurf haben wir die Binomialverteilung $X \sim \text{Bin}(100, 1/2)$. Die Wahrscheinlichkeit, *genau* $x = 58$ mal „Kopf“ zu werfen, ist

$$P(X = 58) = \binom{100}{58} 0.5^{58} \cdot 0.5^{42} = 0.0223$$

Diese Wahrscheinlichkeit ist recht klein. Wie aber sollen wir nun aufgrund dieser Wahrscheinlichkeit entscheiden, ob es sich um eine unfaire Münze handelt? Um zu

entscheiden, ob die Münze fair oder unfair ist, wäre ein allgemeines Kriterium von Vorteil, z.B. ein kritischer Wert c für die Anzahl Würfe mit „Kopf“, ab dem wir der Nullhypothese nicht mehr glauben. Ist 58 grösser als dieser kritische Wert c , so verwerfen wir die Nullhypothese, d.h., die Münze ist unfair. Ist 58 kleiner als c , so behalten wir die Nullhypothese bei und schliessen, dass 58 eine etwas grosse zufällige Schwankung eines an und für sich fairen Würfels war. Uns interessiert also das Ereignis $X \geq c$, da dieses Ereignis gegen die Nullhypothese $\pi = 1/2$ spricht. In der Tabelle 4.1 sind die Wahrscheinlichkeiten für $X \geq c$ mit unterschiedlichen Werten für c aufgeführt, wobei $X \sim \text{Bin}(100, 1/2)$.

| | $c = 53$ | $c = 54$ | $c = 55$ | $c = 56$ | $c = 57$ | $c = 58$ | $c = 59$ | $c = 60$ |
|---------------|----------|----------|----------|----------|----------|----------|----------|----------|
| $P(X \geq c)$ | 0.309 | 0.242 | 0.184 | 0.136 | 0.097 | 0.067 | 0.044 | 0.028 |

Tabelle 4.1.: Zahlen für $X \sim \text{Binomial}(100, 1/2)$.

Der erste Eintrag in der obigen Tabelle ermittelt sich mit R folgendermassen

```
sum(dbinom(53:100, size = 100, prob = 0.5))

## [1] 0.3086497
```

oder eleganter

```
1 - pbinom(52, 100, prob = 0.5)

## [1] 0.3086497
```

Wir müssen nun festlegen, welche Wahrscheinlichkeit wir für das Ereignis $X \geq c$ unter der Nullhypothese noch tolerieren und wann wir die Nullhypothese verwerfen (Zufall ausschliessen) können.

Wenn die Münze fair wäre, würde man das Ereignis „58 mal oder öfter K “ mit einer Wahrscheinlichkeit von nur ca. 6.7% beobachten. D.h., wenn Sie und weitere 999 Studenten jeweils 100 mal diese Münze werfen, würden etwa 67 Studenten das Ereignis „58 mal oder öfter Kopf“ erleben. Ist diese Wahrscheinlichkeit genügend klein, um die Nullhypothese zu verwerfen?

Diese oben berechnete Zahl 6.7% wird in der Literatur auch **P-Wert** genannt (später noch mehr dazu). In diesem Fall ist der P-Wert also 6.7%. In der Statistik hat es sich so ergeben, dass man einer Hypothese nicht mehr glaubt, wenn der P-Wert kleiner als 5% ist. Diese Grenze wird *Signifikanzniveau* genannt und ist natürlich völlig willkürlich. Sie variiert auch von Disziplin zu Disziplin (in Physik sehr klein, in Sozialwissenschaften und Biologie viel grösser). Sie könnte genauso gut 4.7%, 7% oder

irgendeine andere kleine Prozentzahl sein. Wichtig ist nur, dass man einer Hypothese keinen Glauben schenkt, wenn der P -Wert „klein“ ist.

Mit der Festlegung des Signifikanzniveaus entscheiden wir, ob wir eine Beobachtung unter der Annahme der Nullhypothese noch dem Zufall zuordnen können oder nicht. Hätten wir 59 „Köpfe“ geworfen, so ist der P -Wert aus der Tabelle 0.044. Bei einem Signifikanzniveau von 5 % würden wir die Nullhypothese verwerfen und daraus schliessen, dass die Münze nicht fair ist. Dies gilt für alle beobachteten Wurfzahlen von 59 oder grösser. Wir sprechen dann auch vom *Verwerfungsbereich*

$$K = \{59, 60, \dots, 100\}$$

Der kritische Wert c ist also 59. Liegt ein beobachteter Wert im Verwerfungsbereich $K = \{c, \dots, 100\}$, so nehmen wir an, dass die Münze nicht fair ist. Wir sprechen von einem *Testentscheid*. Man spricht auch von einer „statistisch signifikanten“ Beobachtung, wenn wir die Nullhypothese dabei verwerfen. Die Beobachtung ist in diesem Fall ein *deutlicher* Hinweis, dass die Nullhypothese nicht gilt.

□

Bemerkungen:

- i. Im Beispiel oben wird bei 58 K die Nullhypothese nicht verworfen. Das heisst allerdings *nicht*, dass die Nullhypothese gilt. Wir können einfach Zufall nicht ausschliessen. Es ist mehr eine Aussage im Sinne „im Zweifel für den Angeklagten“.
- ii. Die Festlegung des Signifikanzniveaus (die Grenze zwischen Verwerfen und Beibehalten der Nullhypothese) ist willkürlich. Im Beispiel oben wird bei 58 K die Nullhypothese beibehalten, bei 59 K aber verworfen. Die Zahlen 58 und 59 liegen aber nicht so weit auseinander. Aber irgendwo müssen wir die Grenze ziehen.
- iii. Warum legen wir das Signifikanzniveau nicht für die Punktwahrscheinlichkeit $P(X = 58)$, sondern für die Wahrscheinlichkeit des Ereignisses $X \geq 58$, also für $P(X \geq 58)$ fest? Zum einen interessiert uns die Wahrscheinlichkeit, dass eine Beobachtung in den gesamten Bereich $X \geq 58$ fällt. Zum anderen existieren Punktwahrscheinlichkeiten nur für diskrete Zufallsvariablen, bei den stetigen Zufallsvariablen müssten wir sowieso einen Bereich angeben.

4.3.2. Formales Vorgehen

Wir wollen das Vorgehen im Beispiel vorher schematisch darstellen. Dieses Schema werden wir in allen Beispielen für statistische Tests verwenden.

Beispiel 4.3.2

1. Modell:

X : Anzahl Würfe, die Kopf zeigen, wenn man 100 mal wirft.

$$X \sim \text{Bin}(100, \pi)$$

2. Nullhypothese:

$$H_0 : \pi = 0.5$$

Alternative:

$$H_A : \pi > 0.5$$

3. Teststatistik:

T : Anzahl Würfe, die Kopf zeigen, wenn man 100 mal wirft.

Verteilung der Teststatistik unter H_0 :

$$T \sim \text{Bin}(100, 0.5)$$

4. Signifikanzniveau:

$$\alpha = 0.05$$

5. Verwerfungsbereich:

| | $c = 53$ | $c = 54$ | $c = 55$ | $c = 56$ | $c = 57$ | $c = 58$ | $c = 59$ | $c = 60$ |
|---------------|----------|----------|----------|----------|----------|----------|----------|----------|
| $P(X \geq c)$ | 0.309 | 0.242 | 0.184 | 0.136 | 0.097 | 0.067 | 0.044 | 0.028 |

Tabelle 4.2.: Verwerfungsbereich für $X \sim \text{Bin}(100, 1/2)$.

Aus Tabelle 4.2 kann man ablesen, dass $c = 59$ die kleinste Zahl ist, bei der $P[T \geq c]$ kleiner als 5 % ist. Also ist der Verwerfungsbereich

$$K = \{59, 60, \dots, 100\}$$

6. Testentscheid:

Der beobachtete Wert der Teststatistik $t = 58$ liegt nicht im Verwerfungsbereich $K = [59, 100]$. Daher kann die Nullhypothese auf dem Signifikanzniveau 5 % nicht verworfen werden. Das heisst, es gibt nicht genügend statistische Evidenz (auf dem Signifikanzniveau $\alpha = 0.05$) dafür, dass die Münze zu Gunsten von „Kopf“ (K) gefälscht ist.

□

Bemerkungen:

- i. Beachten Sie, dass hier ein kleines t im Fall von $t = 58$ und nicht $T = 58$ verwendet wurde, weil eine konkrete Realisierung der Zufallsvariablen T mit dem Buchstaben t bezeichnet wird.
- ii. Man könnte auf die Idee kommen, das Signifikanzniveau sehr klein zu wählen (z.B. $\alpha = 0.0000000001$), um ganz sicher zu sein, dass die Münze nicht fair ist. Dies hat aber Nachteile, die wir in Abschnitt 4.3.4 auf Seite 125 besprechen werden.
- iii. Im obigen Beispiel sind die Zufallsvariablen des Modells und die Teststatistik identisch, was im Allgemeinen nicht der Fall sein muss. Die Teststatistik ist die Zufallsvariable, die der tatsächlichen Beobachtung einen Wert zuordnet. Im Allgemeinen ist die Teststatistik eine Funktion der Zufallsvariablen des Modells. Zum Beispiel könnte die Zufallsvariable V der Geschwindigkeit eines Teilchens entsprechen, aber beobachtet wird die kinetische Energie E . In diesem Fall ist die kinetische Energie E eine Funktion der Geschwindigkeit V : $E = \frac{1}{2}mV^2$.

Im obigen Beispiel haben wir vermutet, dass das wahre π grösser als ein gegebener Wert ist. Im nächsten Beispiel vermuten wir für das wahre π einen kleineren Wert.

Beispiel 4.3.3

Bei der Übertragung von insgesamt 100 Bits (vgl. Beispiel 3.7.4 auf Seite 88) wurden $x = 6$ Bits fehlerhaft übertragen. Es stellt sich also die Frage, ob wir im Falle des getesteten Materials signifikant weniger fehlerhaft übertragener Bits haben als für das Material mit einer Fehlerrate von 10 %:

1. Modell:

X : Anzahl fehlerhaft übertragener Bits

$$X \sim \text{Bin}(100, \pi)$$

2. Nullhypothese :

$$H_0 : \pi = \pi_0 = 0.1$$

Alternative:

$$H_A : \pi < \pi_0 = 0.1$$

3. Teststatistik:

T : Anzahl fehlerhaft übertragener Bits;

Verteilung der Teststatistik unter H_0

$$H_0 : T \sim \text{Bin}(100, 0.1)$$

4. Signifikanzniveau:

$$\alpha = 0.05$$

5. Verwerfungsbereich:

Mit R kann man die untenstehende Tabelle wie folgt berechnen

```
pbinom(0:10, 100, prob = 0.1)

## [1] 0.0000265614 0.0003216881 0.0019448847
## [4] 0.0078364871 0.0237110827 0.0575768865
## [7] 0.1171556154 0.2060508618 0.3208738884
## [10] 0.4512901654 0.5831555123
```

| | $t = 0$ | $t = 1$ | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ |
|---------------|----------|---------|----------|---------|---------|---------|---------|
| $P(T \leq t)$ | 0.000027 | 0.00032 | 0.001944 | 0.00784 | 0.02371 | 0.0575 | 0.1172 |

Tabelle 4.3.: Verwerfungsbereich für $X \sim \text{Binomial}(100, 0.1)$

Daher (siehe Tabelle 4.3) ist der Verwerfungsbereich für die Teststatistik:

$$K = \{0, 1, 2, 3, 4\}$$

6. Testentscheid:

Der beobachtete Wert $t = 6$ der Teststatistik liegt nicht im Verwerfungsbereich der Teststatistik. Daher kann die Nullhypothese auf dem Signifikanzniveau 5% nicht verworfen werden. D. h., mit dem neuen Material werden in Bezug auf das Vergleichsmaterial nicht statistisch signifikant weniger Bits fehlerhaft übertragen.

Für das Signifikanzniveau $\alpha = 0.05$ ist der Verwerfungsbereich also

$$K = \{0, 1, 2, 3, 4\}$$

Für das Niveau $\alpha = 0.01$ ist der Verwerfungsbereich gegeben durch

$$K = \{0, 1, 2, 3\}$$

Für beide Niveaus wird die Nullhypothese also beibehalten, d.h. das Testmaterial für den Übertragungskanal bietet keinen statistisch signifikanten Vorteil in Bezug auf das Vergleichsmaterial, sowohl auf dem 5 %- als auch auf dem 1 %-Niveau.

Wenn jemand nur an einem Material interessiert ist, deren Fehlerrate kleiner als 5 % ist (weil die Konstruktion von Übertragungskanälen aus diesem Material z. B. erst dann ökonomisch interessant wird), dann wird diese Person im 2. Punkt des Tests folgende Änderung machen:

$$H_0 : \pi = \pi_0 = 0.05$$

$$H_A : \pi < \pi_0 = 0.05$$

Mit Hilfe einer Tabelle erhält man aus

$$P_{\pi_0=0.05}(T \in K) \approx \leq 0.05$$

den Verwerfungsbereich

$$K = \{0, 1\}$$

auf dem 5 %-Niveau. Das Zeichen

$$\approx \leq$$

bedeutet „kleiner gleich, aber so gross wie möglich“. Der beobachtete Wert der Teststatistik liegt in diesem Fall ebenfalls nicht im Verwerfungsbereich. Daher wird H_0 belassen, d.h. eine Fehlerrate von weniger als 5 % bei der Übertragung von digitalen Signalen kann beim neuen Material nicht nachgewiesen werden.

□

Im Beispiel des Münzwurfs machten wir einen Test *nach oben*, d.h. wir vermuteten $\pi > 0.5$. Bei den übertragenen Bits hingegen machten wir einen Test *nach unten*, d.h. wir vermuteten $\pi < 0.1$. Wir sprechen in diesen beiden Fällen von *einseitigen Tests*.

Allgemeines Vorgehen beim einseitigen statistischen Test

Ein statistischer Test für den Parameter π im Modell $X \sim \text{Bin}(n, \pi)$ ist wie folgt aufgebaut.

1. **Modell:** X : Anzahl Treffer bei n Versuchen;

$$X \sim \text{Bin}(n, \pi)$$

2. Spezifiziere die sogenannte **Nullhypothese** H_0 :

$$H_0 : \pi = \pi_0$$

und (anhand der Problemstellung) eine sogenannte **Alternative** H_A :

$$\pi > \pi_0 \quad (\text{einseitig nach oben})$$

$$\pi < \pi_0 \quad (\text{einseitig nach unten})$$

Am häufigsten lautet die Nullhypothese

$$H_0 : \pi_0 = \frac{1}{2}$$

also „reiner Zufall“ oder „kein Effekt“. Meist führt man einen Test durch, weil man glaubt, dass die Alternative richtig ist und man auch Skeptiker davon überzeugen möchte.

3. **Teststatistik:**

T : Anzahl Treffer bei n Versuchen

Verteilung der Teststatistik unter H_0 :

$$T \sim \text{Bin}(n, \pi_0)$$

4. Lege das sogenannte **Signifikanzniveau** α fest. Typischerweise wählt man

$$\alpha = 0.05 \text{ (5 \%)} \quad \text{oder} \quad \alpha = 0.01 \text{ (1 \%)}$$

5. Bestimme den sogenannten **Verwerfungsbereich** K für die Teststatistik.

Qualitativ zeigt K in Richtung der Alternative (Nullhypothese wird verworfen):

$$K = [c_o, n] \quad \text{falls } H_A : \pi > \pi_0$$

$$K = [0, c_u] \quad \text{falls } H_A : \pi < \pi_0$$

Quantitativ wird K so berechnet, dass

$$P_{H_0}(T \in K) = \quad \text{und} \quad \underbrace{P_{\pi_0}}_{\text{von Bin}(n, \pi_0)}(T \in K) \stackrel{\approx}{\leq} \alpha$$

Dabei bedeutet $\stackrel{\approx}{\leq}$ wieder, dass die linke Seite kleiner oder gleich der rechten Seite sein soll, aber so nahe wie möglich.

6. Testentscheid:

Überprüfen Sie nun, ob der beobachtete Wert der Teststatistik in den Verwerfungsbereich K fällt. Falls die Beobachtung in den Verwerfungsbereich fällt, so verwerfen Sie H_0 . H_0 ist in diesem Fall statistisch widerlegt, d.h., die Abweichung von der Nullhypothese ist „signifikant“. Wir können in diesem Fall Zufall ausschließen. Falls die Beobachtung nicht in den Verwerfungsbereich fällt, so belassen Sie H_0 . Dies bedeutet nicht, dass damit H_0 statistisch bewiesen ist.

Diese Art der Test-Entscheidung beruht auf dem Widerspruchs-Prinzip: Hypothesen können nur falsifiziert und nicht verifiziert werden.

Absence of evidence is not evidence of absence.

Falls wir die Nullhypothese nicht verwerfen können, ist das (leider) *kein Nachweis* für die Nullhypothese. Nehmen wir an, dass $\pi_0 = 0.1$ und wir $H_0 : \pi = \pi_0$ nicht verwerfen können. Dann bleibt $\pi = 0.1$ zwar ein plausibler Wert für den Parameter, aber z. B. $\pi = 0.11$ wäre wohl auch noch plausibel.

Oder anders ausgedrückt: Nur weil wir keine Abweichung von π_0 nachweisen können, heisst dies leider noch lange nicht, dass keine Abweichung vorhanden ist! Oder besser in Englisch: „Absence of evidence is not evidence of absence.“

4.3.3. Zweiseitiger statistischer Test

Im Beispiel mit dem Münzwurf wollten wir testen, ob es sich um eine faire Münze handelt. Wir hatten die Vermutung, dass „Kopf“ zu oft oben liegt und haben deshalb einen einseitigen Test ($H_A : \pi > \pi_0$) durchgeführt. Die Münze wäre aber auch unfair, wenn Kopf zu häufig **oder** zu selten oben liegt. In diesem Fall ist ein *zweiseitiger* Test angebracht: $H_A : \pi \neq \pi_0$.

Zweiseitiger statistischer Test

Beim **zweiseitigen Test** auf dem Signifikanzniveau α hat der Verwerfungsbereich die Form (qualitativ zeigt K in Richtung der Alternative):

$$K = [0, c_u] \cup [c_o, n] \quad \text{falls } H_A : \pi \neq \pi_0.$$

Quantitativ werden c_u und c_o beim zweiseitigen statistischen Test so berechnet, dass

$$P_{H_0}(T \leq c_u) \approx \frac{\alpha}{2} \quad \text{und} \quad P_{H_0}(T \geq c_o) \approx \frac{\alpha}{2}$$

Beispiel 4.3.4

Wir haben wiederum eine Münze und wollen prüfen, ob sie fair ist („Kopf“ und „Zahl“ mit Wahrscheinlichkeit 50 %). Wir werfen die Münze 10 mal. Zunächst machen wir einen einseitigen Test:

1. Modell:

X : Anzahl Würfe, die „Kopf“ zeigen, wenn man 10 mal wirft.

$$X \sim \text{Bin}(10, \pi)$$

2. Nullhypothese:

$$H_0 : \pi = 0.5$$

Alternative:

$$H_A : \pi > 0.5$$

3. Teststatistik:

T : Anzahl Würfe, die „Kopf“ zeigen, wenn man 10 mal wirft

Verteilung der Teststatistik unter H_0 :

$$T \sim \text{Bin}(10, 0.5)$$

4. Signifikanzniveau:

$$\alpha = 0.05$$

5. Verwerfungsbereich:

$$K_1 = \{9, 10\}$$

6. Testentscheid:

Weil nicht gesagt wurde, wie oft wir Kopf beobachtet haben, können wir den Testentscheid nicht fällen.

Um zu testen, ob Kopf „zu selten“ vorkommt, hätten wir $H_A : \pi < 0.5$ wählen müssen. In diesem Fall würden wir für den Verwerfungsbereich $K = \{0, 1\}$ erhalten. Für den zweiseitigen Test erhalten wir einen anderen Verwerfungsbereich:

1. Modell: X : Anzahl Würfe, die „Kopf“ zeigen, wenn man 10 mal wirft.

$$X \sim \text{Bin}(10, \pi)$$

2. Nullhypothese:

$$H_0 : \pi = 0.5$$

Alternative:

$$H_A : \pi \neq 0.5$$

3. Teststatistik:

T : Anzahl Würfe, die „Kopf“ zeigen, wenn man 10 mal wirft

Verteilung der Teststatistik unter H_0 :

$$T \sim \text{Bin}(10, 0.5)$$

4. Signifikanzniveau:

$$\alpha = 0.05$$

5. Verwerfungsbereich:

Den zweiseitigen Verwerfungsbereich kann man leicht berechnen: Zunächst bestimmt man die Verwerfungsbereiche ($K_>$ und $K_<$) für die einseitigen Alternativen $H_A : \pi > 0.5$ und $H_A : \pi < 0.5$ mit dem **halben** Signifikanzniveau $\alpha/2$, hier also 0.025. Anschliessend nimmt man die Vereinigung der beiden Verwerfungsbereiche. Mit Tabelle 4.4 ergibt sich $K_> = \{9, 10\}$. Für die andere Alternativhypothese ergibt sich $K_< = \{0, 1\}$. Für den Verwerfungsbereich des zweiseitigen Tests erhal-

| | 0 | 1 | 2 | 8 | 9 | 10 |
|---------------|-------|-------|-------|-------|-------|-------|
| $P(X \geq c)$ | | | | 0.055 | 0.011 | 0.001 |
| $P(X \leq c)$ | 0.001 | 0.011 | 0.055 | | | |

Tabelle 4.4.: Verwerfungsbereich für $X \sim \text{Bin}(10, 0.5)$ für den zweiseitigen Test

ten wir also:

$$K_2 = K_< \cup K_> = \{0, 1\} \cup \{9, 10\}$$

6. Testentscheid:

Weil nicht gesagt wurde, wie oft wir Kopf beobachtet haben, können wir den Testentscheid nicht fällen.



Bemerkungen:

- i. In der Tabelle 4.4 zur Bestimmung des Verwerfungsbereichs sind nur die relevanten Zahlen eingetragen. Können Sie sich erklären, warum die Zahlen in der Tabelle in beiden Zeilen gleich sind? Wäre das auch so, wenn wir $H_0 : \pi = 0.7$ testen würden?

Zusammenfassung Statistischer Test

Die Durchführung eines statistischen Tests kann – zumindest teilweise – „rezeptartig“ erfolgen:

1. Wähle ein geeignetes Modell für die Daten.
2. Lege die Nullhypothese $H_0 : \vartheta = \vartheta_0$ fest (ϑ bezeichnet hier allgemein einen Parameter in einem Modell). Anhand der Problemstellung, spezifiziere die Alternative

$$H_A : \vartheta \neq \vartheta_0 \text{ (zweiseitig)}$$

$$\vartheta > \vartheta_0 \text{ (einseitig nach oben)}$$

$$\vartheta < \vartheta_0 \text{ (einseitig nach unten)}$$

3. Definiere eine Teststatistik T (Grösse, die beobachtet wird) und bestimme, wie T unter der Nullhypothese H_0 verteilt ist
4. Wähle das Signifikanzniveau α , typischerweise $= 0.05$ oder 0.01 .
5. Konstruiere den Verwerfungsbereich K für H_0 , so dass gilt

$$P_{\vartheta_0}(T \in K) \leq \alpha$$

Der Verwerfungsbereich zeigt immer in Richtung der Alternative H_A .

6. Betrachte, ob die Beobachtung x (oder eine Funktion von mehreren Beobachtungen) in den Verwerfungsbereich fällt. Falls ja, so verwerfe H_0 zugunsten von H_A . Falls x nicht in den Verwerfungsbereich fällt, so belassen wir H_0 , was aber noch lange nicht heisst, dass deswegen H_0 statistisch nachgewiesen wurde („absence of evidence is not evidence of absence“).

Oft stellt sich die Frage, ob ein einseitiger oder ein zweiseitiger Test für eine Fragestellung besser geeignet wäre. Im Folgenden zeigen wir, dass beide Tests Vor- und

Nachteile haben.

Ein zweiseitige Test stellt sowohl zu viele Erfolge als auch zu wenige Erfolge fest. Wenn der Test einseitig mit $H_A : \pi > \pi_0$ ist, dann stellt er nur *zu viele* Erfolge fest. Wenn der Test einseitig mit $H_A : \pi < \pi_0$ ist, dann stellt er nur *zu wenige* Erfolge fest. Damit scheint der zweiseitige dem einseitigen Test überlegen zu sein. Für die „Sehschärfe“ in beide Richtungen zahlt man allerdings einen Preis:

- Der **zweiseitige Test** stellt zwar Abweichungen in beide Richtungen von H_0 fest, aber die Abweichung muss sehr deutlich sein, damit er sie erkennt. D.h., die Wahrscheinlichkeit, die Nullhypothese zugunsten der Alternativhypothese zu verwerfen, falls die Alternativhypothese richtig ist, ist beim zweiseitigen Test klein. Die **Macht** (siehe Abschnitt 4.3.4) ist also klein.
- Der **einseitige Test** stellt nur Abweichungen in eine Richtung von H_0 fest, aber die Abweichungen müssen nicht so gross sein, damit er sie erkennt. D.h., die Wahrscheinlichkeit, die Nullhypothese zugunsten der Alternativhypothese zu verwerfen, falls die Alternativhypothese richtig ist, ist beim einseitigen Test gross. Die **Macht** ist also gross.

Es ist gängige Praxis, die Entscheidung bzgl. der Form von H_A zu fällen, *bevor* die Experimente durchgeführt werden.

4.3.4. Fehler 1. und 2. Art beim statistischen Test

Verwerfen wir eine Nullhypothese, so heisst das, dass das beobachtete Ereignis unter dieser Hypothese so unwahrscheinlich ist, dass wir Zufall ausschliessen können. Allerdings ist es immer noch möglich, dass die Nullhypothese *richtig* ist und wir einfach ein sehr unwahrscheinliches Ereignis beobachtet haben. Wir haben dann mit dem Verwerfen der Nullhypothese einen *Fehler* gemacht. In diesem Fall sprechen wir von einem *Fehler 1. Art*. Auf der anderen Seite können wir die Nullhypothese beibehalten, obwohl die Alternativhypothese richtig ist. Dies nennen wir einen *Fehler 2. Art*. Bei einem statistischen Test treten also 2 Arten von Fehlern auf:

Fehler 1. Art: Fälschliches Verwerfen von H_0 , obwohl H_0 richtig ist.

Fehler 2. Art: Fälschliches Beibehalten von H_0 , obwohl die Alternative zutrifft.

Wir werden uns zuerst mit dem Fehler 1. Art auseinandersetzen. Formal lautet die Definition

Fehler 1. Art

Der **Fehler 1. Art** ist definiert als

$$P(\text{Fehler 1. Art}) = P_{H_0}(T \in K) \leq \alpha$$

Nach Konstruktion des Tests ist die Wahrscheinlichkeit für einen Fehler 1. Art höchstens gleich α .

Notation:

Wir werden nun häufig Wahrscheinlichkeiten für verschiedene Werte von π berechnen. Dabei deuten wir mit der Bezeichnung $P_{0.8}$ an, welchen Parameter wir verwendet haben (hier $\pi = 0.8$).

Beispiel 4.3.5

In Abbildung 4.3 ist für das Beispiel 4.3.4 auf Seite 122 der Fehler erster Art dargestellt. Links: Der einseitige nach oben gerichtete Verwerfungsbereich ist gegeben

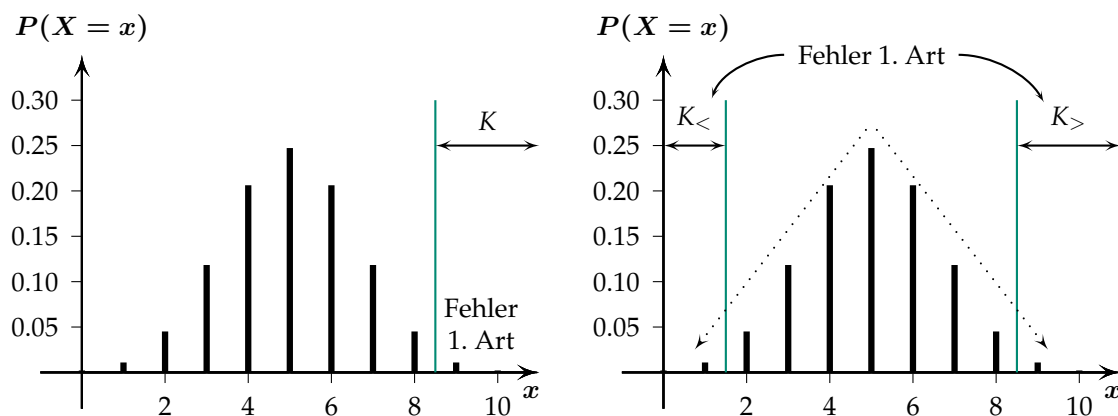


Abbildung 4.3.: Beispiel Fehler 1. Art

durch

$$K = \{9, 10\}$$

Der Fehler 1. Art ist also die Summe der Wahrscheinlichkeiten für die Werte von X im Verwerfungsbereich unter der Nullhypothese $\pi_0 = 0.5$

$$P_{0.5}(X = 9) + P_{0.5}(X = 10) = 0.011$$

In einem von 100 Fällen, also in 1 % der Fälle, wird die Nullhypothese fälschlicherweise verworfen.

Rechts: Der zweiseitige Verwerfungsbereich ist gegeben durch

$$K_{>} = \{9, 10\} \quad \text{und} \quad K_{<} = \{0, 1\}$$

Der Fehler 1. Art ist also die Summe der Wahrscheinlichkeiten für die Werte von X im Verwerfungsbereich unter der Nullhypothese:

$$P_{0.5}(X = 0) + P_{0.5}(X = 1) + P_{0.5}(X = 9) + P_{0.5}(X = 10) = 0.021$$

Die Wahrscheinlichkeit beträgt also etwa 2 %, dass die Nullhypothese fälschlicherweise verworfen wird.

□

Betrachten wir nun im Folgenden den Fehler 2. Art, der gegeben ist durch das fälschliche Beibehalten von H_0 , obwohl die Alternative zutrifft. Formal lautet die Definition

Fehler 2. Art

Der **Fehler 2. Art** ist definiert als

$$P(\text{Fehler 2. Art}) = P_{H_A}(T \in \bar{K})$$

Beispiel 4.3.6

In Abbildung 4.4 ist für das Beispiel 4.3.4 auf Seite 122 der Fehler 2. Art dargestellt.

Die Binomialverteilung $\text{Bin}(10, \pi)$ ist unter der Nullhypothese $H_0 : \pi_0 = 0.5$ in schwarz eingezeichnet und die alternative Hypothese $H_A : \pi_0 = 0.8$ ist in grün dargestellt.

Der Verwerfungsbereich auf dem Signifikanzniveau $\alpha = 0.05$ ist $K = \{9, 10\}$. Dann entspricht die Summe der Wahrscheinlichkeiten für alle Beobachtungen *ausserhalb* des Verwerfungsbereichs unter der Alternativhypothese dem Fehler 2. Art:

$$P_{0.8}(X = 0) + P_{0.8}(X = 1) + \dots + P_{0.8}(X = 8) = P_{0.8}(X \leq 8) = 0.624$$

Das heisst, dass wir in rund 62 % der Fälle die Nullhypothese beibehalten, obwohl die Alternativhypothese $H_A : \pi_0 = 0.8$ richtig ist.

□

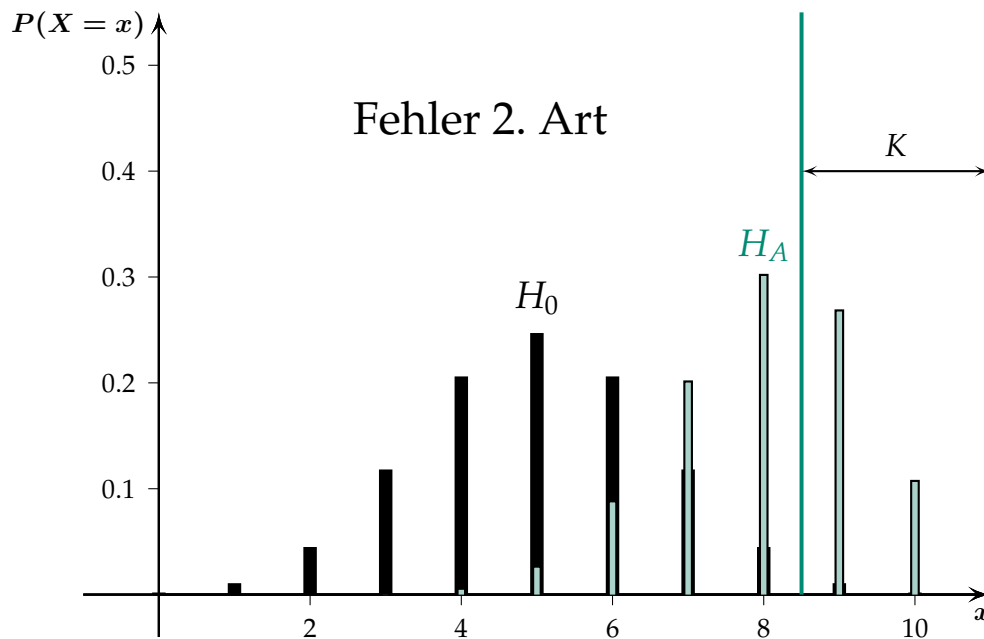


Abbildung 4.4.: Fehler 2. Art für Binomialverteilung

Welche Fehlerart ist wichtiger?

Dem Fehler 1. Art wird traditionell mehr Gewicht gegeben als dem Fehler 2. Art: Wissenschaftler arbeiten genau und haben Angst, einen Humbug zu publizieren, der sich dann als falsch herausstellt. Denn wenn Wissenschaftler einen Effekt (signifikante Abweichung von Nullhypothese) beobachten, möchten sie sicher sein, dass es sich nicht bloss um Zufall handelt. Der Fehler 1. Art soll vermieden werden. Dabei nimmt man in Kauf, dass man manchmal einen wichtigen Effekt verpasst. Der Fehler 2. Art ist also zweitrangig.

Der Fehler 1. Art wird direkt kontrolliert mittels der Konstruktion eines Tests, indem α möglichst klein gehalten wird. Über die Wahrscheinlichkeit eines Fehlers 2. Art haben wir hingegen keine solche Kontrolle. Die beiden Fehlerarten konkurrenzieren sich gegenseitig:

$P(\text{Fehler 2. Art})$ wird grösser falls α kleiner gewählt wird

Die Wahl von α steuert also einen Kompromiss zwischen Fehler 1. und 2. Art. Weil man aber primär einen Fehler 1. Art vermeiden will, wählt man α klein, z.B. $\alpha = 0.05$. Dies ist ersichtlich in Abbildung 4.4: je kleiner α , desto kleiner der Verwerfungsbereich. Die vertikale Linie wandert also nach rechts und somit wird der Fehler 2. Art umso grösser.

4.3.5. Die Macht eines statistischen Tests

Statt der Wahrscheinlichkeit eines Fehlers 2. Art gibt man oft die sogenannte *Macht* eines Tests an. Die Macht gibt die Wahrscheinlichkeit an, H_A zu entdecken, falls H_A richtig ist.

Macht

Die **Macht** ist definiert als

$$\begin{aligned}\text{Macht} &= 1 - P(\text{Fehler 2. Art}) \\ &= P(\text{Verwerfen von } H_0 \text{ falls } H_A \text{ stimmt}) \\ &= P_{H_A}(T \in K)\end{aligned}$$

Für die Berechnung der Macht muss der Parameter π_A der Alternativhypothese H_A bekannt sein.

Beispiel 4.3.7

In Abbildung 4.5 für das Beispiel 4.3.4 auf Seite 122 ist die Macht eines (einseitigen) statistischen Tests dargestellt. Die Binomialverteilung $\text{Binomial}(10, \pi)$ unter der Null-

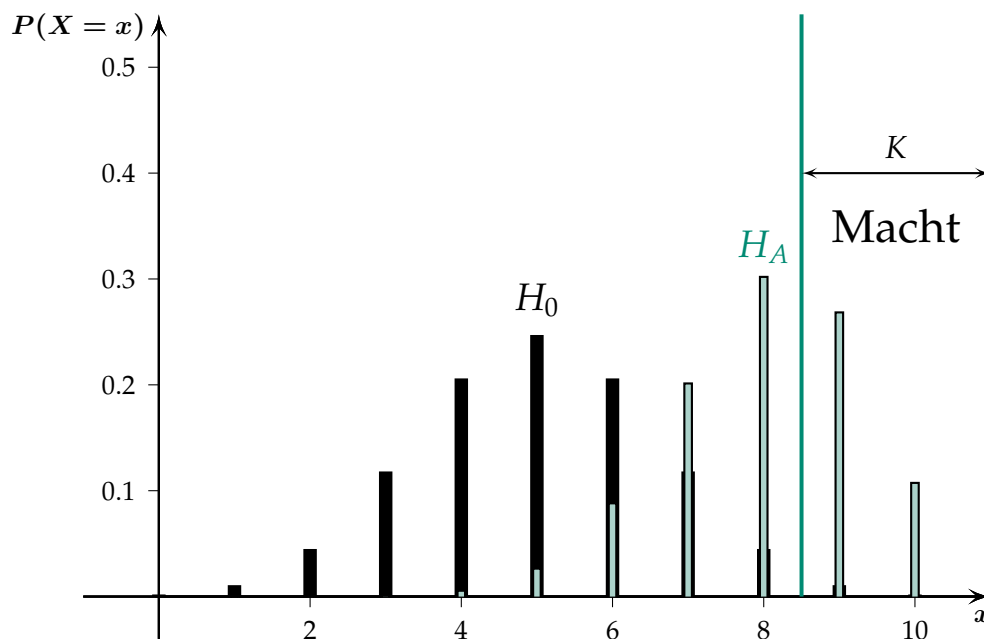


Abbildung 4.5.: Macht des Tests für Binomialverteilung.

hypothese $H_0 : \pi_0 = 0.5$ ist in schwarz eingezeichnet und die alternative Hypothese $H_A : \pi = 0.8$ ist in grün dargestellt. Der Verwerfungsbereich auf dem Signifikanzniveau $\alpha = 0.05$ lautet $K = \{9, 10\}$. Dann entspricht die Summe der Punktwahrscheinlichkeiten unter der Alternativhypothese für alle Beobachtungen, die innerhalb des Verwerfungsbereichs liegen, der *Macht des Tests*. Wenn nun in Tat und Wahrheit die Wahrscheinlichkeit, „Kopf“ zu werfen, durch $\pi = 0.8$ gegeben ist, berechnet sich die Macht M_1 des einseitigen Tests mit $X \sim \text{Bin}(10, 0.8)$ zu:

$$\begin{aligned} M_1 &= P_{0.8}(X \in K_1) = P_{0.8}(X \in \{9, 10\}) \\ &= P_{0.8}(X = 9) + P_{0.8}(X = 10) \\ &= \binom{10}{9} 0.8^9 0.2^1 + \binom{10}{10} 0.8^{10} \\ &= 0.3758096 \end{aligned}$$

Nun berechnen wir die Macht M_2 des zweiseitigen Tests. Wir nehmen wiederum an, dass $X \sim \text{Bin}(10, 0.8)$:

$$\begin{aligned} M_2 &= P_{0.8}(X \in K_2) = P_{0.8}(X \in \{0, 1\} \cup \{9, 10\}) \\ &= P_{0.8}(X = 0) + P_{0.8}(X = 1) + P_{0.8}(X = 9) + P_{0.8}(X = 10) \\ &= \binom{10}{0} 0.2^{10} + \binom{10}{1} 0.8^1 0.2^9 + \binom{10}{9} 0.8^9 0.2^1 + \binom{10}{10} 0.8^{10} \\ &= 0.3758138 \end{aligned}$$

Die Macht des einseitigen Tests ist in diesem Fall also etwa gleich gross wie die des zweiseitigen Tests.

□

Beispiel 4.3.8

Bei der Übertragung von 100 digitalen Bits (siehe Beispiel 3.7.4 auf Seite 88 und Beispiel 4.3.3 auf Seite 117) nehmen wir nun an, dass in Tat und Wahrheit die Wahrscheinlichkeit für eine fehlerhafte Übertragung $H_A : \pi = 0.04$ ist. Die Spezifikationen des Tests lauten also: $H_0 : \pi_0 = 0.1$, $H_A : \pi < 0.1$ und $\alpha = 0.05$. Wie gross ist die Wahrscheinlichkeit, dass unser Test die Richtigkeit der Alternativhypothese $H_A : \pi < 0.1$ nicht entdeckt?

Da der Verwerfungsbereich $K = \{0, 1, 2, 3, 4\}$ ist (siehe Beispiel 4.3.3 auf Seite 117), gilt:

$$\begin{aligned} P(\text{Test behält } H_0 \text{ bei, obschon } \pi = 0.04) &= P_{0.04}(T \geq 5) \\ &= 1 - P_{0.04}(T \leq 4) \\ &= 0.37 \end{aligned}$$

Es tritt also ein Fehler 2. Art mit einer Wahrscheinlichkeit von rund 37 % auf. Diese relative hohe Wahrscheinlichkeit für einen Fehler 2. Art ist natürlich enttäuschend, wenn $\pi = 0.04$ ökonomisch interessant wäre.

Beachten Sie, dass die Wahrscheinlichkeit für einen Fehler 1. Art höchstens 0.05, also klein ist. In diesem Beispiel ist die Macht M

$$M = 1 - 0.37 = 0.63$$

Somit wird die Nullhypothese mit einer Wahrscheinlichkeit von 63 % verworfen, wenn die Alternativhypothese richtig ist.

□

4.3.6. Der P -Wert

Der P -Wert ist ein Wert zwischen 0 und 1, der angibt, wie gut *Nullhypothese* und *Daten* zusammenpassen (0: passt gar nicht; 1: passt sehr gut). Etwas präziser formuliert, definieren wir den P -Wert als die Wahrscheinlichkeit, unter Gültigkeit der Nullhypothese das erhaltene Ergebnis oder ein **extremeres** zu erhalten (siehe Abbildung 4.6). *Extrem* bezieht sich auf die Alternativhypothese: Wenn $H_A : \pi > \pi_0$, dann sind sehr grosse Anzahlen von Treffern extrem. Wenn $H_A : \pi < \pi_0$, dann sind sehr kleine Anzahlen von Treffern extrem. Wenn $H_A : \pi \neq \pi_0$, dann sind sowohl sehr grosse als auch sehr kleine Anzahlen von Treffern extrem.

Mit dem P -Wert wird also angedeutet, wie extrem das Ergebnis ist: Je kleiner der P -Wert, desto mehr spricht das Ergebnis gegen die Nullhypothese. Werte kleiner als eine im voraus festgesetzte Grenze, wie 5 %, 1 % oder 0.1 % sind Anlass, die Nullhypothese abzulehnen.

P -Wert

Der **P -Wert** ist die Wahrscheinlichkeit, unter der Nullhypothese ein mindestens so extremes Ereignis (in Richtung der Alternative) zu beobachten wie das aktuell

beobachtete.

Beispiel 4.3.9

Bei einer Binomialverteilung mit $n = 10$ wollen wir die Nullhypothese

$$H_0 : \pi = \pi_0 = 0.5$$

gegen die Alternative

$$H_A : \pi > 0.5$$

testen (π ist z.B. die Wahrscheinlichkeit für „Kopf“ bei einer Münze). Wir haben also unter H_0 die Zufallsvariable $X \sim \text{Bin}(10, 0.5)$. Die Zufallsvariable X ist dann die Anzahl Würfe mit „Kopf“ bei insgesamt 10 Würfeln.

Die Verteilung von X unter H_0 ist in Abbildung 4.6 dargestellt. Beobachtet wurde $x = 7$. Der P -Wert ist hier die Summe aller Wahrscheinlichkeiten für X grösser gleich 7, d.h.

$$P\text{-Wert} = P_{0.5}(X \geq 7) = 0.172$$

Wenn wir diese Zahl als Signifikanzniveau wählen, verwerfen wir die Nullhypothese gerade (knapp) noch.

Wenn wir zweiseitig testen würden ($H_A : \pi \neq 0.5$), dann müssten wir die Wahrscheinlichkeiten „auf der anderen Seite“ auch noch dazu addieren, d.h. wir hätten dann

$$P\text{-Wert} = P_{0.5}(X \leq 3) + P_{0.5}(X \geq 7) = 0.344$$

als P -Wert.

□

Das Vorgehen ist also praktisch identisch wie beim Testentscheid (verwerfen/beibehalten), ausser dass der P -Wert noch einen konkreten Wert mitliefert.

P -Wert und Statistischer Test

Man kann anhand des P -Werts direkt den Testentscheid ablesen: Wenn der P -Wert kleiner als das Niveau ist, so verwirft man H_0 , ansonsten nicht. Verglichen mit dem reinen Testentscheid enthält der P -Wert aber mehr Information, da man direkt sieht, „wie stark“ die Nullhypothese verworfen wird. Bei einem vorbestimmten Signifikanzniveau α (z.B. $\alpha = 0.05$) gilt aufgrund der Definition des P -Werts:

1. Verwerfe H_0 falls $P\text{-Wert} \leq \alpha$

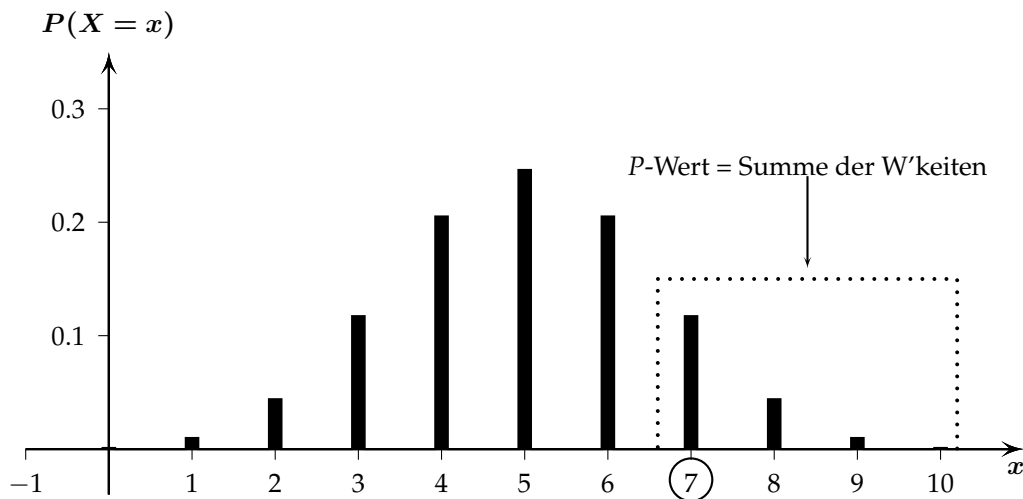


Abbildung 4.6.: P -Wert für Binomialverteilung

2. Belasse H_0 falls $P\text{-Wert} > \alpha$

Viele Computer-Pakete liefern den Testentscheid nur indirekt, indem der P -Wert angegeben wird.

Zusätzlich zu dieser Entscheidungsregel quantifiziert der P -Wert, *wie signifikant* eine Alternative ist (d.h. wie gross die Evidenz ist für das Verwerfen von H_0). Manchmal werden sprachliche Formeln oder Symbole anstelle der P -Werte angegeben:

$P\text{-Wert} \approx 0.05$: schwach signifikant, “.”

$P\text{-Wert} \approx 0.01$: signifikant, “*”

$P\text{-Wert} \approx 0.001$: stark signifikant, “***”

$P\text{-Wert} \leq 10^{-4}$: äusserst signifikant, “****”

Beispiel 4.3.10

Bei der Übertragung von 100 digitalen Signalen betrachten wir die Nullhypothese $\pi_0 = 0.1$ und die Alternative $\pi < 0.1$. Beobachtet wurde $x = 6$, aufgefasst als Realisierung von $X \sim \text{Bin}(100, \pi)$. Der P -Wert ist dann:

$$\begin{aligned} P_{\pi_0=0.1}(X \leq 6) &= P_{\pi_0=0.1}(X = 6) + P_{\pi_0=0.1}(X = 5) + \cdots + P_{\pi_0=0.1}(X = 0) \\ &= 0.12 \end{aligned}$$

Wie wir bereits früher gesehen haben, liefert dies kein Verwerfen von H_0 auf dem Signifikanzniveau $\alpha = 0.05$ (weil der P -Wert grösser als 0.05 ist). Auf dem Signifikanzniveau $\alpha = 0.15$ könnte man H_0 allerdings verwerfen, weil der P -Wert kleiner gleich 0.15 ist.

□

Wir berechnen den P -Wert aus obigem Beispiel nun mit R:

```
binom.test(x = 6, n = 100, p = 0.1, alternative = "less",
  conf.level = 0.95)

##
## Exact binomial test
##
## data: 6 and 100
## number of successes = 6, number of trials = 100,
## p-value = 0.1172
## alternative hypothesis: true probability of success is less than 0.1
## 95 percent confidence interval:
## 0.0000000 0.1149853
## sample estimates:
## probability of success
## 0.06
```

Bemerkungen:

- i. Bei der von uns festgelegten Nullhypothese und Richtung der Alternativhypothese lautet der P -Wert 0.12, wie wir im obigem Beispiel bereits berechnet haben.
- ii. **Achtung:** Der P -Wert ist nicht die Wahrscheinlichkeit, dass die Nullhypothese stimmt. Darüber können wir hier gar keine Aussagen machen, da die Parameter fix und nicht zufällig sind. Bei der Berechnung des P -Werts geht man davon aus, dass H_0 stimmt und schaut, wie extrem dann das beobachtete Ereignis ist.
- iii. Streng genommen handelt es sich beim P -Wert um *keine* Wahrscheinlichkeit. Denn die Wahrscheinlichkeit, aufgefasst als relative Auftretenshäufigkeit, eines Ereignisses macht hier keinen Sinn. Der P -Wert macht keine Aussage über die Auftretenshäufigkeit der Beobachtung oder eines extremeres Ereignisses, sondern bloss wie verträglich dieses Ereignis mit der Nullhypothese ist.

Strikte betrachtet ist der P -Wert einfach eine transformierte Zufallsvariable. Eine alternative und unmissverständlichere Definition des P -Wertes lautet:

Der P -Wert ist das kleinste Signifikanzniveau, bei dem die Nullhypothese H_0 (gerade noch) verworfen wird.

Diese Definition vermeidet es, den P-Wert als Wahrscheinlichkeit zu interpretieren. Nichtsdestotrotz werden wir im Folgenden im Zusammenhang vom P-Wert von einer Wahrscheinlichkeit sprechen.

4.4. Vertrauensintervall

Das *Vertrauensintervall*, das auch *Konfidenzintervall* genannt wird, liefert eine Antwort auf die 3. Grundfragestellung der Statistik vom Unterkapitel 4.1: Welche Werte von π sind mit der Beobachtung x kompatibel (statistisch vereinbar).

Die Fragestellung ist hier also in einem gewissen Sinne umgekehrt zum statistischen Test. Wir suchen die Werte π für die Nullhypothese, bei denen eine Beobachtung *nicht* verworfen wird. Das Vertrauensintervall ist informativer als ein statistischer Test.

Beispiel 4.4.1

Sie kommen an einer Losbude vorbei und ziehen 50 Lose: darunter sind 7 Gewinne. Es stellt sich also die Frage: Welche Werte für die Gewinnwahrscheinlichkeit π sind kompatibel mit Ihrer Beobachtung?

Wie können wir die untere Grenze π_u eines Intervalls von mit der Beobachtung kompatiblen π -Werten ermitteln? π_u soll gerade noch kompatibel mit der Beobachtung $x = 7$ sein. Wie definieren wir den Begriff „kompatibel“? Als Kriterium oder Mass für „Kompatibilität“ könnten wir den einseitigen P-Wert, hier $P_\pi(X \geq 7)$, in Betracht ziehen, wobei wir eine kritische Grenze für diesen Wert brauchen, bei deren Unterschreitung wir die Beobachtung $x = 7$ als nicht mehr kompatibel mit π auffassen. Wir legen die kritische Grenze für diesen einseitigen P-Wert bei $\alpha/2$ fest: Ist $P_\pi(X \geq 7)$ kleiner als $\alpha/2$, würden wir π nicht mehr als kompatibel mit der Beobachtung $x = 7$ bezeichnen.

Zur Ermittlung von π_u lassen wir den Wert von π nach unten „wandern“ und berechnen dabei für jeden Wert von π den P-Wert $P_\pi(X \geq 7)$. Wir starten beim Schätzwert $\hat{\pi} = 0.14$, und verkleinern π so lange, bis der einseitige P-Wert für unsere Beobachtung $x = 7$ den Wert $\alpha/2 = 0.025$ annimmt. Dies ist der Fall bei $\pi = \pi_u = 0.06$.

Formal bedeutet dies (siehe Abbildung 4.7):

$$P_{\pi_u}(X \geq 7) \stackrel{\approx}{\geq} \frac{\alpha}{2} = 0.025$$

Mit R können wir dies (vorläufig) mit folgendem Befehl untersuchen (R macht dies auch automatisch, wie wir unten sehen werden).

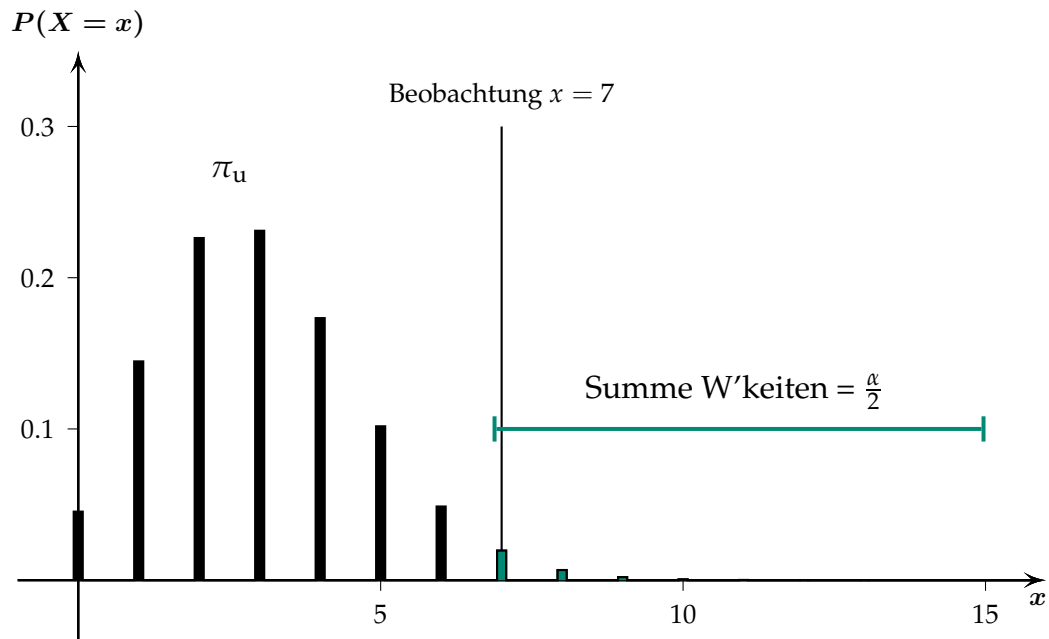


Abbildung 4.7.: Untere Grenze π_u des 95 %-Vertrauensintervalls hat die Eigenschaft, dass $P_{\pi_u}(X \geq 7) \approx \alpha/2$ ist.

```
pi.trial <- seq(0.14, 0, -0.01)
for (i in 1:length(pi.trial)) {
  p <- sum(dbinom(7:50, size = 50, prob = pi.trial[i]))
  cat(pi.trial[i], p, "\n")
}
```

```
## 0.14 0.5616338
## 0.13 0.478908
## 0.12 0.3934784
## 0.11 0.3090581
## 0.1 0.2297732
## 0.09 0.1596321
## 0.08 0.1018718
## 0.07 0.05831364
## 0.06 0.02892409
## 0.05 0.01178645
## 0.04 0.003610064
## 0.03 0.0007037346
## 0.02 6.012668e-05
## 0.01 6.852838e-07
## 0 0
```


Hier sehen wir, dass bei $\pi = 0.05$ der Wert von $P_\pi(X \geq 7)$ die kritische Grenze von $\alpha/2 = 0.025$ unterschreitet, und somit die Nullhypothese $\pi = 0.05$ gerade verworfen wird. Bei $\pi = 0.06$ wird die Nullhypothese gerade *nicht* mehr verworfen. Wir *erlauben* also den Parameter $\pi = 0.06$ als mögliche untere Trefferwahrscheinlichkeit.

Die obere Grenze π_o für unser Intervall ermitteln wir auf die gleiche Art und Weise: wir lassen π von $\hat{\pi} = 0.14$ nach oben „wandern“, bis der einseitige P -Wert $P_\pi(X \leq 7)$ für unsere Beobachtung $x = 7$ den Wert $\alpha/2 = 0.025$ annimmt. Dies ist der Fall bei $\pi = \pi_o = 0.27$. Formal bedeutet dies (siehe Abbildung 4.8):

$$P_{\pi_o}(X \leq 7) \approx \frac{\alpha}{2}$$

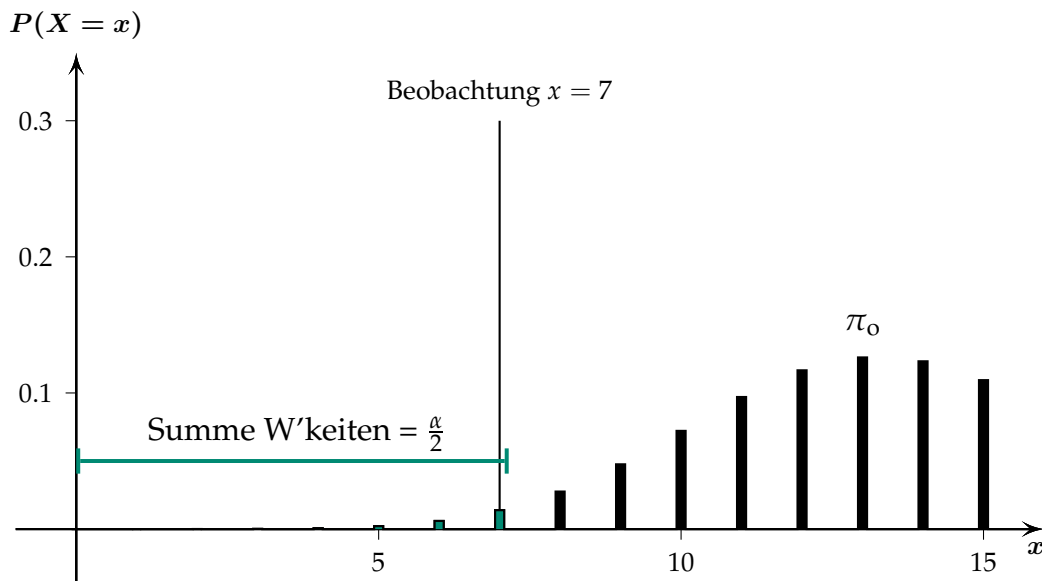


Abbildung 4.8.: Obere Grenze π_o des 95 %-Vertrauensintervalls hat die Eigenschaft, dass $P_{\pi_o}(X \leq 7) \approx \alpha/2$ ist.

Das soeben ermittelte Intervall nennen wir **Vertrauensintervall**. Es lautet

$$I(x=7) = [0.06, 0.27]$$

Liegt also der Wert π für die Nullhypothese in diesem Bereich, so wird für die Beobachtung $x = 7$ die Nullhypothese auf dem Signifikanzniveau α *nicht* verworfen. Falls der wahre Wert von π ausserhalb des soeben konstruierten Vertrauensintervalles liegt, so wird die Nullhypothese mit dem wahren Wert von π mit einer Wahrscheinlichkeit von höchstens α verworfen. Der wahre Wert von π liegt also mit einer Wahrscheinlichkeit von 95% in obigem Vertrauensintervall, weshalb wir dieses Vertrauensintervall auch 95%–**Vertrauensintervall** nennen.

Der R-Befehl zur Bestimmung des Vertrauensintervalls sieht wie folgt aus:

```
binom.test(7, 50)

##
## Exact binomial test
##
## data: 7 and 50
## number of successes = 7, number of trials = 50,
## p-value = 2.099e-07
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.0581917 0.2673960
## sample estimates:
## probability of success
##                                0.14
```

In der 4. und 5. Zeile von unten sind die Angaben zum Vertrauensintervall.



Vertrauensintervall

Ein Vertrauensintervall I zum Niveau $1 - \alpha$ besteht aus allen Parameterwerten, die im Sinne des statistischen Tests zum Signifikanzniveau α mit der Beobachtung verträglich sind (üblicherweise nimmt man den zweiseitigen Test). Mathematisch heisst dies:

$$I = \{\pi_0 \mid \text{Nullhypothese } H_0 : \pi = \pi_0 \text{ wird belassen}\} = [\pi_u, \pi_o]$$

Diese Beziehung stellt eine Wechselwirkung zwischen Tests und Vertrauensintervallen dar.

Die Berechnung des Vertrauensintervalls erfolgt in der Praxis natürlich meist mit dem Computer.

Das Vertrauensintervall $I = I(x)$ hängt von der Beobachtung ab. Wenn man anstelle der Beobachtung die zugehörige Zufallsvariable X einsetzt, so ist $I(X)$ zufällig und hat die Eigenschaft:

$$P_\pi(\pi \in I(X)) \approx 1 - \alpha \quad \text{für jedes } \pi$$

Dies kann so interpretiert werden, dass das Konfidenzintervall I den wahren Parameter π mit Wahrscheinlichkeit $1 - \alpha$ mit einschliesst.

Beispiel 4.4.2

Um das zu veranschaulichen, haben wir mit dem Computer folgende Situation simuliert: Hundert Personen gehen auf den Jahrmarkt. Alle kommen an einer Losbude vorbei, bei der die wahre Gewinnwahrscheinlichkeit $\pi = 0.1$ ist. Allerdings ist die Gewinnwahrscheinlichkeit nirgends angeschrieben, so dass die Personen sie nicht kennen. Jede Person beschliesst nun, 50 Lose zu kaufen und ein 95 %-Vertrauensintervall für die Gewinnwahrscheinlichkeit zu bestimmen.

In Abbildung 4.9 ist das Ergebnis dieser Simulation dargestellt. Auf der x -Achse ist die Gewinnwahrscheinlichkeit aufgetragen. Der wahre Wert ist 0.1. Darüber sind 100 horizontale Striche gezeichnet. Jeder Strich repräsentiert das 95 %-Vertrauensintervall für die Gewinnwahrscheinlichkeit von einer der 100 Personen. Der vertikale Strich zeigt die Position der wahren Gewinnwahrscheinlichkeit. Man sieht, dass 4 Vertrauensintervalle die wahre Gewinnwahrscheinlichkeit nicht enthalten (sie sind rot markiert). Die übrigen 96 Vertrauensintervalle enthalten die wahre Gewinnwahrscheinlichkeit. Die Abbildung zeigt also, dass ein 95 %-Vertrauensintervall mit ca. 95 % Wahrscheinlichkeit den wahren Parameter enthält.

Da ein 95 %-Vertrauensintervall den wahren Parameter mit 95 % Wahrscheinlichkeit enthält, sollten rund 95 Personen von den 100 den wahren Parameter in ihrem Vertrauensintervall haben.

□

Einseitige Vertrauensintervalle

Wollen wir für die Beobachtung $X = x$ ein **einseitiges nach unten gerichtetes** Vertrauensintervall auf dem Signifikanzniveau α bestimmen, so haben wir bloss einer Bedingung Rechnung zu tragen

$$P_{\pi_o}(X \leq x) \stackrel{\approx}{\geq} \alpha$$

Das einseitige nach unten gerichtete Vertrauensintervall hat dann die Form $[0, \pi_o]$. Analog lässt sich das **einseitige nach oben gerichtete** Vertrauensintervall $[\pi_u, 1]$ auf dem Signifikanzniveau α mit der Bedingung

$$P_{\pi_u}(X \geq x) \stackrel{\approx}{\geq} \alpha$$

bestimmen.

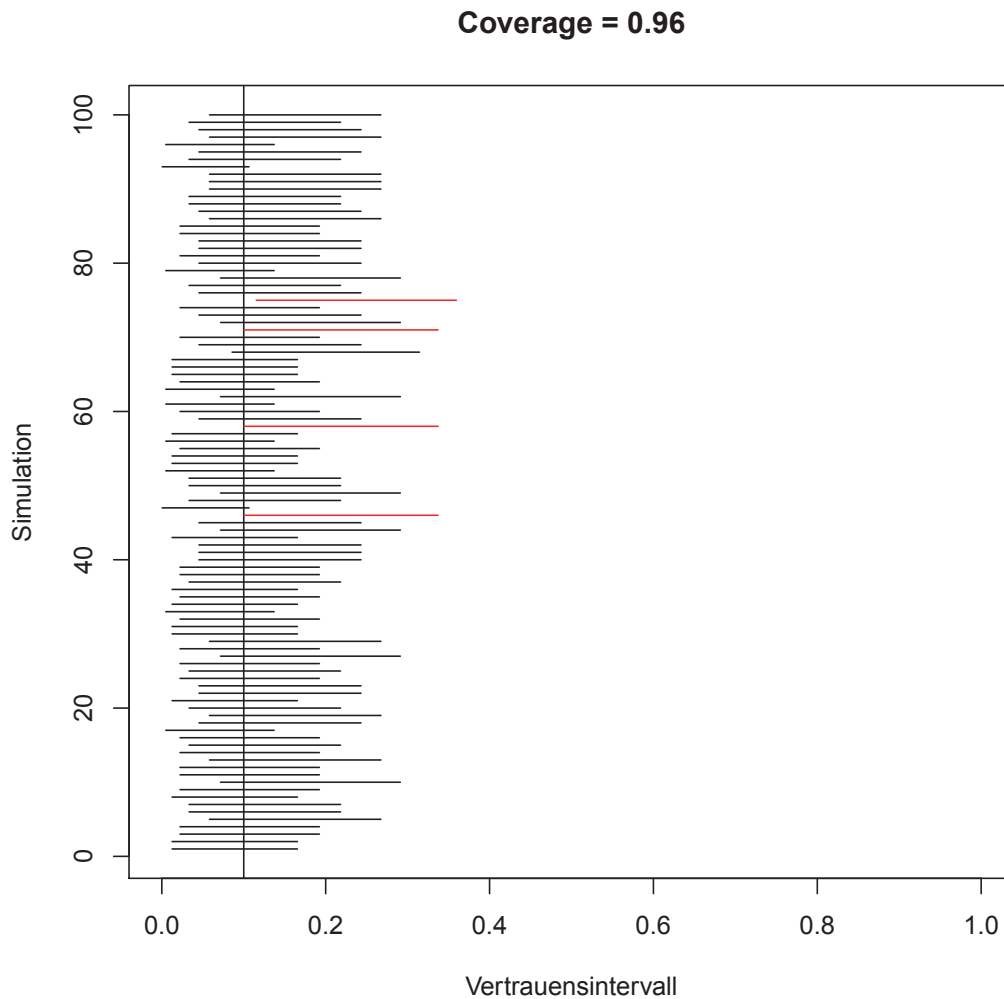


Abbildung 4.9.: Vergleich von 100 95 %-Vertrauensintervallen

Beispiel 4.4.3

Bei der Übermittlung von 100 digitalen Bits erhält man aufgrund der beobachteten 6 fehlerhaften Bits das einseitige nach unten gerichtete Konfidenzintervall zum Niveau $1 - \alpha = 0.95$ mittels eines Computers für die Berechnung (siehe dazu das Beispiel 3.7.4 auf Seite 88):

$$I = [0, 0.11]$$

Das heisst, dass die wahre Fehlerrate π mit einer Wahrscheinlichkeit von 95 % in I liegt.

□

Mit R lässt sich jeweils das Vertrauensintervall einfach bestimmen. Im obigen Beispiel eines einseitigen Vertrauensintervalls findet man:

```
binom.test(x = 6, n = 100, p = 0.1, alternative = "less",
           conf.level = 0.95)

##
## Exact binomial test
##
## data: 6 and 100
## number of successes = 6, number of trials = 100,
## p-value = 0.1172
## alternative hypothesis: true probability of success is less than 0.1
## 95 percent confidence interval:
##  0.0000000 0.1149853
## sample estimates:
## probability of success
##                                0.06
```

Bemerkungen:

- i. Beim einseitigen 95 %-Vertrauensintervall mit Alternativhypothese nach unten interessiert uns, wie wahrscheinlich das Ereignis $X \leq x$ unter der Nullhypothese π_0 ist. Das 95 %-Vertrauensintervall hat also die Form $[0, \pi_o]$, denn der P-Wert $P_\pi(X \leq x)$ für die Beobachtung x kann nur kleiner als das Signifikanzniveau α werden, wenn sich π nach oben bewegt.
- ii. Beim einseitigen 95 %-Vertrauensintervall mit Alternativhypothese nach oben interessiert uns, wie wahrscheinlich das Ereignis $X \geq x$ unter der Nullhypothese π_0 ist. Das 95 %-Vertrauensintervall hat also die Form $[\pi_u, 1]$, denn der P-Wert $P_\pi(X \geq x)$ für die Beobachtung x kann nur kleiner als das Signifikanzniveau α werden, wenn sich π nach unten bewegt.
- iii. In der vierten und fünften Zeile von unten steht das 95 %-Vertrauensintervall für die Gewinnwahrscheinlichkeit: Mit 95 % Wahrscheinlichkeit liegt die Gewinnwahrscheinlichkeit im Bereich $[0, 0.1149853]$.
- iv. Die Alternative mit „less“ bezeichnet eine nach unten gerichtete Alternativhypothese. Mit „greater“ wird eine nach oben gerichtete Alternativhypothese definiert und „two.sided“ bezeichnet eine zweiseitige Alternativhypothese.

Häufig tritt die Situation auf, dass wir eine Nullhypothese nicht verwerfen möchten, sondern dass wir zeigen möchten, dass sie zutrifft. Zum Beispiel in der Qualitätskontrolle möchte man in der Regel ja zeigen, dass die Qualitätsstandards eingehalten werden, respektive dass der Produktionsprozess beim Sollwert verläuft. Dies zu überprüfen, ist im Rahmen des statistischen Tests nicht möglich, da beim statistischen

Test die Nullhypothese verworfen werden soll. In diesem Fall liefert das Vertrauensintervall aber eine Antwort auf die Frage, ob der Produktionsprozess beim Sollwert verläuft. Ist der Sollwert im Vertrauensintervall enthalten, dann ist der Produktionsprozess unter Kontrolle.

Beispiel 4.4.4

Ein Zeltli-Produzent garantiert, dass 10 % der Zeltli grün sind. Er entnimmt einer grossen Lieferung eine Stichprobe von 50 Zeltli und zählt 3 grüne Zeltli. Kann er sicher sein, dass seine Produktangaben korrekt sind?

In diesem Fall kann er dies nicht mit einem statistischen Test überprüfen, denn er will ja die „Nullhypothese“, dass 10 % der Zeltli grün sind, nicht verwerfen, sondern belegen.

Wir bezeichnen mit X die Anzahl grüner Zeltli. X folgt einer Binomialverteilung mit $n = 50$. Welche Gewinnwahrscheinlichkeiten, dass wir ein grünes Zeltli ziehen, sind kompatibel mit der Realisierung $x = 3$? Wir berechnen dazu das zweiseitige 95 %-Vertrauensintervall

```
binom.test(x = 3, n = 50, alternative = "two.sided", conf.level = 0.95)

##
## Exact binomial test
##
## data: 3 and 50
## number of successes = 3, number of trials = 50,
## p-value = 3.708e-11
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.01254859 0.16548195
## sample estimates:
## probability of success
##                0.06
```

Das zweiseitige 95 %-Vertrauensintervall $[0.013, 0.17]$ enthält die Gewinnwahrscheinlichkeit 10 %. Der Zeltli-Hersteller kann also davon ausgehen, dass der Produktionsprozess unter Kontrolle ist.

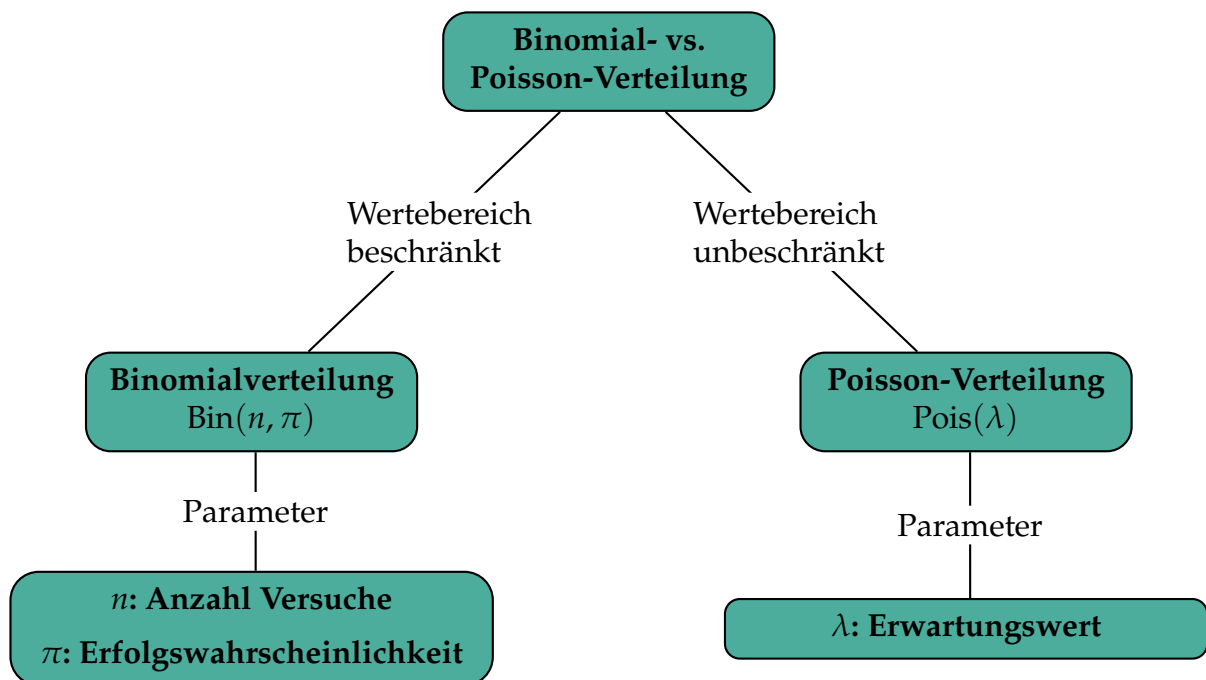
□

Das Vertrauensintervall ist informativer als der statistische Test und der P-Wert.

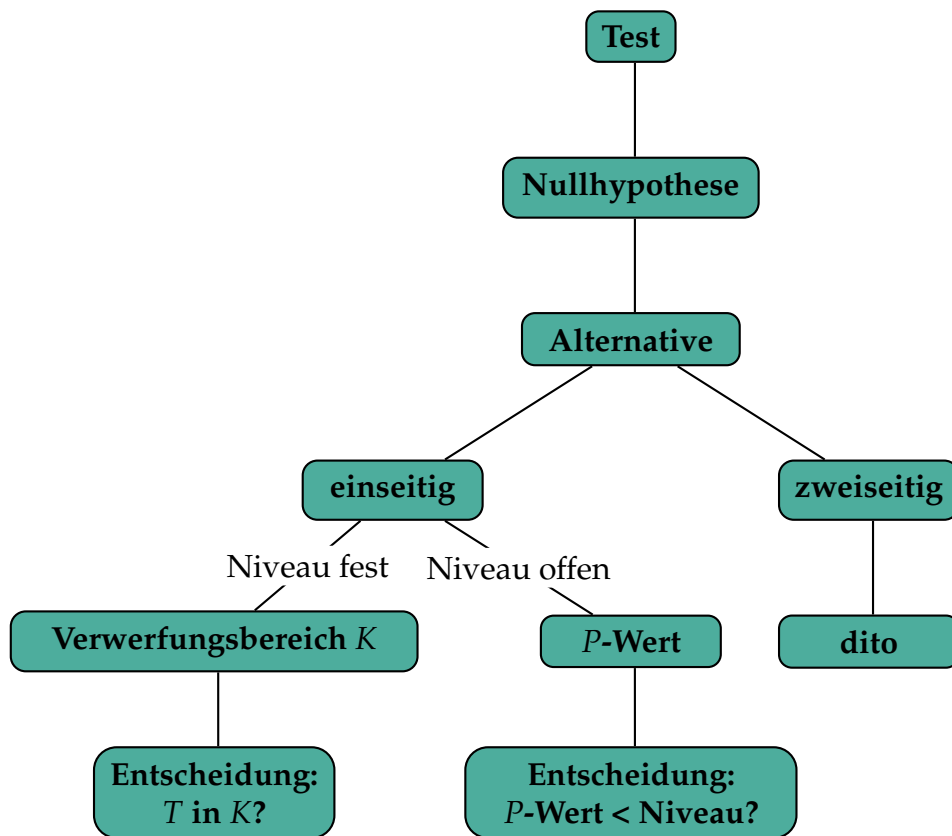
Lernziele

- ☐ Sie kennen die Momentenmethode und die Maximum-Likelihood-Methode zum Schätzen von Parametern und können Sie auf einfache Beispiele anwenden.
- ☐ Sie können die sechsstufige Struktur eines statistischen Tests aufschreiben und haben die Bestandteile verstanden. Insbesondere kennen Sie die Begriffe Fehler 1. Art, Fehler 2. Art und Macht eines Tests.
- ☐ Sie können testen, ob eine vorgegebene Gewinnwahrscheinlichkeit bei einer Binomialverteilung plausibel ist (Binomialtest) und können den Test auf dem Papier durchführen (einseitig und zweiseitig). Sie können entscheiden, ob ein einseitiger oder ein zweiseitiger Test angebracht ist.
- ☐ Sie verstehen die Definition des P -Wertes und können ihn in kleinen Beispielen auf dem Papier berechnen.
- ☐ Sie haben begriffen, was ein Vertrauensintervall ist und können es für eine Gewinnwahrscheinlichkeit berechnen.
- ☐ Sie können einen Binomialtest mit \mathbb{R} durchführen und auswerten.

Binomial- vs. Poisson-Verteilung



Übersicht



Kapitel 5.

Modelle für Messdaten

Everybody believes in the exponential law of errors [i.e., the Normal distribution]: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation.

(E. T. Whittaker and G. Robinson)

5.1. Stetige Zufallsvariablen und Wahrscheinlichkeitsverteilungen

In vielen Anwendungen hat man es nicht mit Zähl-, sondern mit Messdaten zu tun. Diese können grundsätzlich jeden Wert in einem bestimmten Bereich annehmen, wobei die Genauigkeitsangabe des Messwertes durch die Messgenauigkeit vorgegeben wird.

Der Wertebereich W_X einer Zufallsvariablen X ist die Menge aller Werte, die X annehmen kann. Eine Zufallsvariable X heisst *stetig*, wenn deren Wertebereich W_X kontinuierlich ist. Beispiele von kontinuierlichen Wertebereichen sind

$$W_X = \mathbb{R}, \mathbb{R}^+ \quad \text{oder} \quad [0, 1]$$

Konvention bei der Klammernotation:

Bei einer runden Klammer liegt der Wert ausserhalb des Intervalls, bei einer eckigen Klammer liegt der Wert innerhalb des Intervalls. Das Intervall $(a, b]$ beschreibt also alle Punkte x mit $x > a$ und $x \leq b$.

In Unterkapitel 3.6 hatten wir gesehen, dass die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen beschrieben werden kann, indem man die Punktwahrscheinlichkeiten $P(X = x)$ für alle möglichen Werte x im Wertebereich angibt. Für eine stetige Zufallsvariable X gilt jedoch:

$$P(X = x) = 0$$

für alle $x \in W_X$. Dies impliziert, dass wir die Wahrscheinlichkeitsverteilung von X nicht mittels der Angaben von Punktwahrscheinlichkeiten beschreiben können.

Beispiel 5.1.1

Diesen Sachverhalt kann man intuitiv einfach an einem Beispiel verstehen. Angenommen, wir haben eine Zufallsvariable X_0 , die jeden Wert aus der diskreten Menge

$$W_0 = \{0, 1, 2, \dots, 8, 9\}$$

mit gleicher Wahrscheinlichkeit annimmt. Die Wahrscheinlichkeit, dass die Zufallsvariable X_0 einen konkreten Wert x (z. B. $x = 3$) aus dem Bereich W_0 annimmt, ist also

$$P[X_0 = x] = \frac{1}{10}$$

weil W_0 aus zehn Elementen besteht. Jetzt vergrößern wir die diskrete Menge, indem wir jede Zahl auf eine Nachkommastelle genau angeben:

$$W_1 = \{0.0, 0.1, 0.2, \dots, 9.8, 9.9\}$$

Die Zufallsvariable X_1 nimmt jeden Wert aus W_1 mit gleicher Wahrscheinlichkeit an, also ist

$$P[X_1 = x] = \frac{1}{100}$$

weil W_1 aus hundert Elementen besteht. Zum Beispiel haben wir also

$$P[X_1 = 3.0] = \frac{1}{100}$$

Wenn man noch eine Nachkommastelle hinzufügt, erhält man eine Menge aus tausend Elementen und die Wahrscheinlichkeit, ein bestimmtes Element zufällig zu ziehen (z. B. die Zahl 3.00), beträgt nur noch $1/1000$. Wenn wir diese Entwicklung fortsetzen, kommen wir zu folgender Regel: Wenn jede der Zahlen zwischen 0 und kleiner als 10 mit i Nachkommastellen mit gleicher Wahrscheinlichkeit gezogen wird, beträgt die Wahrscheinlichkeit, eine bestimmte Zahl aus dieser Menge zu ziehen,

$$\frac{1}{10^{i+1}}$$

Wenn wir unendlich viele Nachkommastellen zulassen, wird aus der diskreten Menge die kontinuierliche Menge

$$W_\infty = [0, 10]$$

Entsprechend beträgt die Wahrscheinlichkeit, ein bestimmtes Element mit unendlich vielen Nachkommastellen aus dieser Menge zu ziehen,

$$P[X_{\infty} = x] = \frac{1}{\infty} = 0$$

□

Beispiel 5.1.2

Wir messen die Körpergrösse von Personen. Die Wahrscheinlichkeit, eine Körpergrösse von *genau* 182.254 680 895 434... cm zu messen, ist gleich 0:

$$P(X = 182.254\,680\,895\,434\ldots) = 0$$

wobei X die Körpergrösse misst. Was für eine Wahrscheinlichkeit können wir aber im Zusammenhang von Körpergrössen angeben? Nun wir könnten die Wahrscheinlichkeit angeben, dass ein Messwert in einem bestimmten Bereich liegt, wie z.B. zwischen 174 und 175 cm:

$$P(174 < X \leq 175)$$

Diese Wahrscheinlichkeit ist dann nicht mehr 0. Um diese Wahrscheinlichkeit zu berechnen, können wir allerdings nicht einfach die Punktwahrscheinlichkeiten aufaddieren, da diese 0 ergäbe. Wir brauchen also ein neues Konzept, und zwar die sogenannte *Wahrscheinlichkeitsdichte*.

□

Hat man einen Datensatz mit experimentellen Messdaten, so wird man feststellen, dass die relative Häufigkeit von Messpunkten in bestimmten Intervallen grösser ist als in anderen. Die *Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen* X kann also beschrieben werden, indem man die Wahrscheinlichkeiten für alle Intervalle $(a, b]$ mit $a < b$ angibt:

$$P(X \in (a, b]) = P(a < X \leq b)$$

Dazu genügt es, die *kumulative Verteilungsfunktion*

$$F(x) = P(X \leq x)$$

anzugeben, denn es gilt

$$P(a < X \leq b) = F(b) - F(a)$$

Diese kumulative Verteilungsfunktion hat wichtige Eigenschaften (siehe Abbildung 5.1):

1. Da es sich bei $F(x) = P(X \leq x)$ um eine Wahrscheinlichkeit handelt, gilt

$$0 \leq F(x) \leq 1$$

2. Die Wahrscheinlichkeit $P(X \leq -\infty)$, dass ein Messwert kleiner als $-\infty$ ist, ist offensichtlich 0. Und damit ist auch

$$F(-\infty) = 0$$

3. Die Wahrscheinlichkeit $P(X \leq \infty)$, dass ein Messwert kleiner als ∞ ist, ist offensichtlich 1:

$$F(\infty) = 1$$

4. Die Funktion von $F(x)$ ist monoton wachsend. Es gilt also für $a < b$:

$$F(a) \leq F(b)$$

Die Ableitung $F'(x)$ von $F(x)$ ist also immer grösser gleich 0.

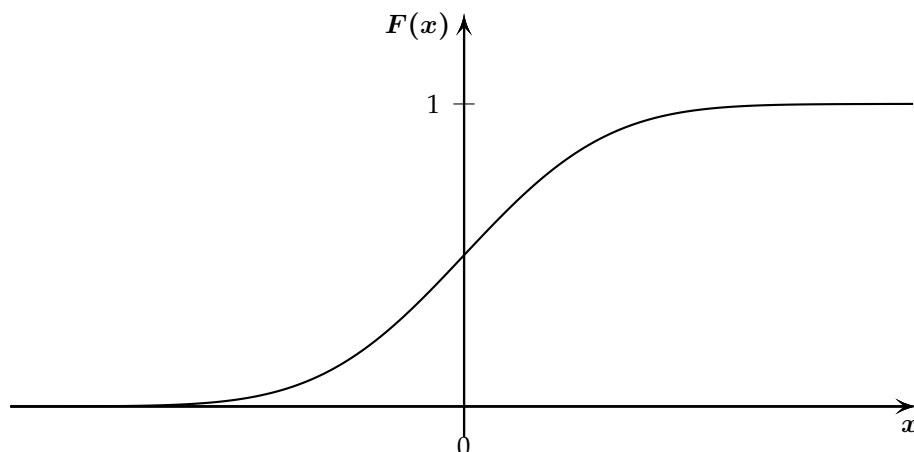


Abbildung 5.1.: Beispiel für eine kumulative Verteilungsfunktion.

Zusammenfassend heisst dies, dass die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen X durch die kumulative Verteilungsfunktion beschrieben werden kann. Weil für stetige Zufallsvariablen

$$P(X = a) = P(X = b) = 0$$

gilt, spielt es keine Rolle, ob wir $<$ oder \leq schreiben. Bei diskreten Zufallsvariablen ist diese Unterscheidung allerdings wichtig.

5.1.1. Wahrscheinlichkeitsdichte

Für *stetige* Zufallsvariablen können wir einen zur Punktwahrscheinlichkeit $P(X = x)$ analogen Begriff mit Hilfe der Ableitung der kumulativen Verteilungsfunktion gewinnen. Um dies zu erreichen, wollen wir die Wahrscheinlichkeit

$$P(a < X \leq b)$$

als Summe von Wahrscheinlichkeiten schreiben. Das Intervall $(a, b]$ unterteilen wir dazu in n -Teilintervalle mit Teilpunkten

$$a = x_0, x_1, x_2, \dots, b = x_n$$

Die einzelnen Teilintervalle haben die gleiche Länge Δx :

$$\Delta x = \frac{b - a}{n}$$

Dann können wir die Wahrscheinlichkeit $P(a < X \leq b)$ in der Form

$$\begin{aligned} P(a < X \leq b) &= P(x_0 < X \leq x_1) + P(x_1 < X \leq x_2) + \dots + P(x_{n-1} < X \leq x_n) \\ &= P(x_0 < X \leq x_0 + \Delta x) + \dots + P(x_{n-1} < X \leq x_{n-1} + \Delta x) \\ &= \sum_{i=0}^{n-1} P(x_i < X \leq x_i + \Delta x) \end{aligned}$$

schreiben. Für die Summanden in der letzten Summe verwenden wir nun die kumulative Verteilungsfunktion F :

$$P(a < X \leq b) = \sum_{i=0}^{n-1} P(x_i < X \leq x_i + \Delta x) = \sum_{i=0}^{n-1} (F(x_i + \Delta x) - F(x_i))$$

Wir erweitern nun die Summanden mit Δx :

$$P(a < X \leq b) = \sum_{i=0}^{n-1} (F(x_i + \Delta x) - F(x_i)) = \sum_{i=0}^{n-1} \frac{F(x_i + \Delta x) - F(x_i)}{\Delta x} \Delta x$$

Aus der Definition der Ableitung folgt für kleine Δx :

$$F'(x_i) \approx \frac{F(x_i + \Delta x) - F(x_i)}{\Delta x}$$

Setzen wir dies in unsere Summe ein, so erhalten wir

$$P(a < X \leq b) \approx \sum_{i=0}^{n-1} F'(x_i) \Delta x$$

Diese Summe ist aber eine Riemann-Summe, die für $\Delta x \rightarrow 0$ gegen ein bestimmtes Integral strebt. Es gilt also

$$P(a < X \leq b) = \int_a^b F'(x) \, dx$$

Dieser Funktion $F'(x)$ nennen wir

Wahrscheinlichkeitsdichte

Die **Wahrscheinlichkeitsdichte** f ist definiert als Ableitung der kumulativen Verteilungsfunktion:

$$f(x) = F'(x)$$

Aus der Dichte kann man die kumulative Verteilungsfunktion zurückgewinnen:

$$F(x) = \int_{-\infty}^x f(y) \, dy$$

weil F eine Stammfunktion von f ist und $F(-\infty) = 0$.

Eigenschaften Wahrscheinlichkeitsdichte

Für eine Wahrscheinlichkeitsdichte $f(x)$ gelten folgende Eigenschaften (siehe Abbildung 5.2):

1. Es gilt

$$f(x) \geq 0$$

für alle x , da $F(x)$ monoton wachsend ist und damit deren Ableitung größer gleich 0 sein muss.

2. Es gilt

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) \, dx$$

Dies entspricht der Fläche zwischen a und b unter $f(x)$.

3. Es gilt

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Dies ist die Wahrscheinlichkeit, dass *irgendein* Wert gemessen wird.

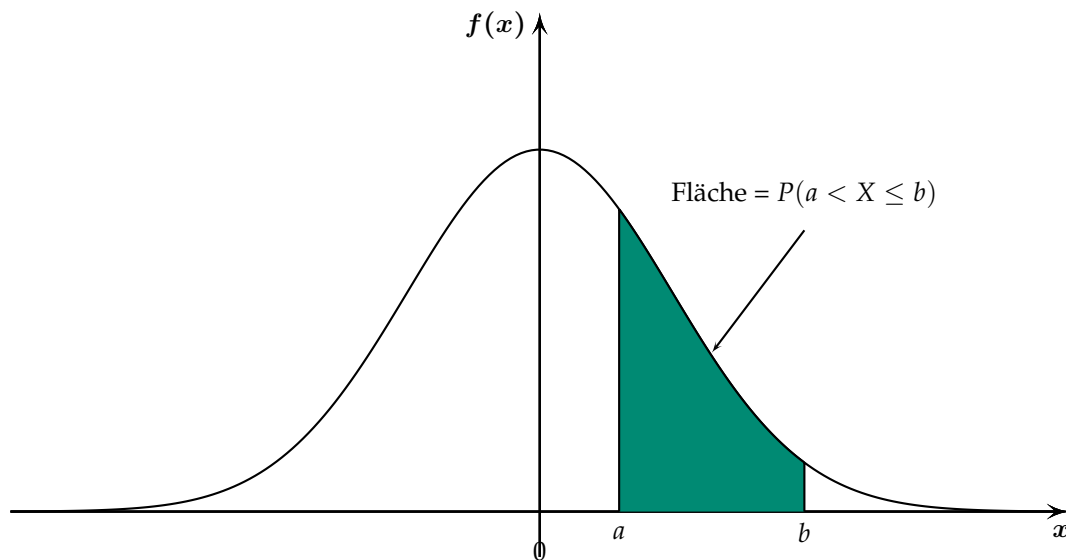


Abbildung 5.2.: Illustration einer Dichte einer Zufallsvariablen und der Wahrscheinlichkeit, in das Intervall $(a, b]$ zu fallen (grüne Fläche).

5.1.2. Kennzahlen von stetigen Verteilungen

Der **Erwartungswert** $E(X)$ und die **Standardabweichung** σ_X einer stetigen Zufallsvariablen X haben dieselbe Bedeutung wie im diskreten Fall im Unterkapitel 3.8: Der Erwartungswert beschreibt die mittlere Lage der Verteilung und die Standardabweichung deren Streuung. Wir erinnern uns, dass der Erwartungswert einer diskreten Zufallsvariablen definiert ist als

$$E(X) = \sum_i x_i P(X = x_i)$$

Die Formeln für den *Erwartungswert* und die *Varianz* einer stetigen Zufallsvariablen ergeben sich, indem wir beim diskreten Fall $P(X = x)$ durch $f(x)dx$ und die Summe durch ein Integral ersetzen:

Erwartungswert und Varianz

Erwartungswert und Varianz sind wie folgt definiert:

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x) dx$$

$$\text{Var}(X) = \sigma_X^2 = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = E(X^2) - E(X)^2$$

Die **Quantile** $q(\alpha)$ für $0 < \alpha < 1$ einer Zufallsvariablen X , bzw. deren Verteilung, sind wie folgt definiert:

$$P(X \leq q(\alpha)) = \alpha$$

Das heisst:

$$F(q(\alpha)) = \alpha \quad \Leftrightarrow \quad q(\alpha) = F^{-1}(\alpha).$$

Dies kann auch so interpretiert werden, dass $q(\alpha)$ der Punkt ist, so dass die Fläche von $-\infty$ bis $q(\alpha)$ unter der Dichte f gleich α ist. Siehe auch Abbildung 5.3. Das 50 %-Quantil heisst der **Median**.

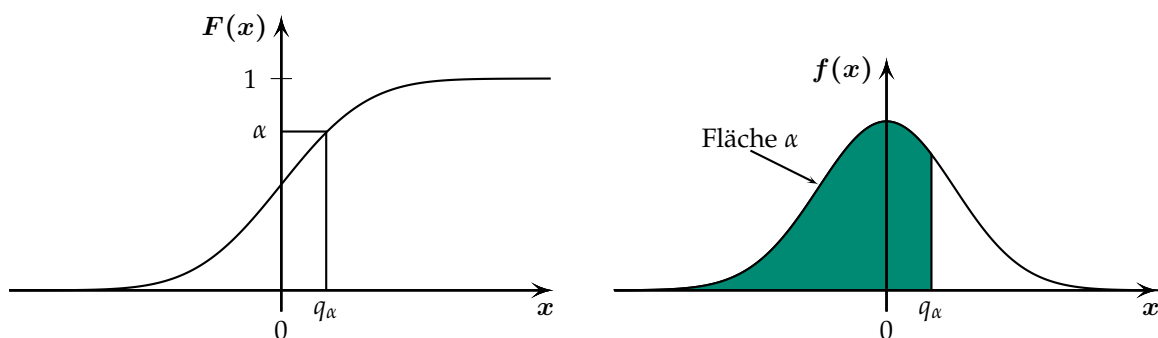


Abbildung 5.3.: Illustration des Quantils q_α anhand der Verteilungsfunktion $F(x)$ (links) und der Dichte $f(x)$ (rechts) für $\alpha = 0.75$.

Beispiel 5.1.3

Wir betrachten wiederum die Verteilung von Körpergrössen. Ist beispielsweise für $\alpha = 0.75$ das zugehörige Quantil gegeben durch $q(\alpha) = 182.5$, so bedeutet dies, dass 75 % der gemessenen Personen kleiner oder gleich 182.5 cm sind.

□

5.2. Wichtige stetige Verteilungen

Im diskreten Fall haben wir die Binomialverteilung und die Poisson-Verteilung als diskrete Verteilungen kennengelernt. In diesem Kapitel werden wir die wichtigsten stetigen Verteilungen kennenlernen.

Wir haben im Unterkapitel 5.1 gesehen, dass wir die Wahrscheinlichkeitsverteilung einer stetigen Zufallsvariablen mit der kumulativen Verteilungsfunktion F oder der Dichte f charakterisieren können.

5.2.1. Uniforme Verteilung

Die Uniforme Verteilung tritt als Formalisierung der völligen „Ignoranz“ auf.

Uniforme Wahrscheinlichkeitsverteilung

Eine Zufallsvariable X mit Wertebereich $W_X = [a, b]$ heisst $\text{Uniform}([a, b])$ verteilt, falls

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Die Dichte ist also konstant auf dem Intervall $[a, b]$ (siehe Abbildung 5.4 links). Das heisst, dass die gleiche Wahrscheinlichkeit vorliegt auf dem ganzen Wertebereich $W_X = [a, b]$, deshalb der Name *uniform* (gleichförmig).

Die zugehörige kumulative Verteilungsfunktion (siehe Abbildung 5.4 rechts) ist

$$F(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b \\ 1 & \text{falls } x > b \end{cases}$$

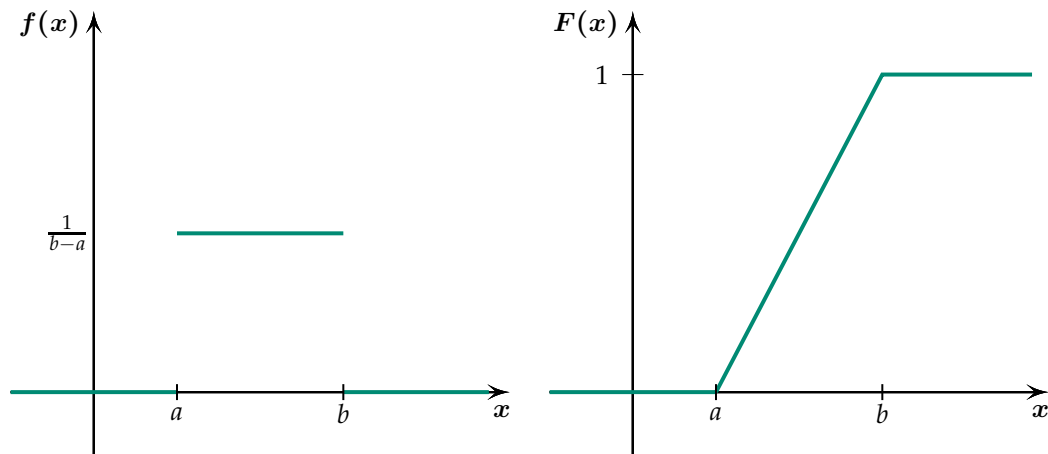


Abbildung 5.4.: Dichte (links) und Verteilungsfunktion (rechts) der uniformen Verteilung.

Für $X \sim \text{Uniform}([a, b])$ sind die Kennzahlen wie folgt:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}$$

$$\sigma_X = \frac{b-a}{\sqrt{12}}$$

Beispiel 5.2.1

Mit R lässt sich der Wert der Wahrscheinlichkeitsdichtefunktion $\text{Uniform}([1, 10])$ an der Stelle $x = 5$ folgendermassen berechnen:

```
dunif(x = 5, min = 1, max = 10)
```

```
## [1] 0.1111111
```

Vorsicht:

Im Gegensatz zu `dbinom()` oder `dpois()` berechnet `dunif()` *nicht* die zugehörige Wahrscheinlichkeit, sondern die Wahrscheinlichkeits*dichte*.

Falls $X \sim \text{Uniform}([1, 10])$, dann entspricht die Wahrscheinlichkeit $P(1 \leq X \leq 5)$ in diesem Fall gerade der Wahrscheinlichkeit $P(X \leq 5)$. Diese berechnen wir mit R folgendermassen:

```
punif(q = 5, min = 1, max = 10)
```

```
## [1] 0.4444444
```

Die Wahrscheinlichkeit $P(1.2 \leq X \leq 4.8)$ berechnet sich wie folgt

```
punif(4.8, 1, 10) - punif(1.2, 1, 10)
```

```
## [1] 0.4
```

Was von grosser Bedeutung ist, ist die Generierung von uniform verteilten Zufallsvariablen. Mit R lassen sich uniform verteilte Zufallsvariablen folgendermassen erzeugen:

```
runif(5, min = 1, max = 10)
```

```
## [1] 7.256078 6.971403 1.733422 4.738717 1.417112
```

□

5.2.2. Exponentialverteilung

Die Exponentialverteilung ist das einfachste Modell für *Wartezeiten auf Ausfälle*, also für die *Lebensdauer*.

Beispiel 5.2.2

Lebenszeit von elektronischen Geräten, wenn Alterungserscheinungen nicht betrachtet werden müssen.

□

Beispiel 5.2.3

Wie lange müssen wir warten, bis der nächste Zerfall eines Alphastrahlers stattfindet? In diesem Fall fassen wir die Zerfallszeit als eine Lebensdauer auf, und zwar als Lebensdauer eines Isotops.

□

Die Poissonverteilung beschreibt die Anzahl Beobachtungen in einem festen Zeitintervall. Mit der Exponentialverteilung ermitteln wir die Wahrscheinlichkeit für eine Lebensdauer.

Notation

Die natürliche Exponentialfunktion e^x schreiben wir oft in der Form:

$$\exp(x) := e^x$$

Exponentialverteilung

Eine Zufallsvariable X mit Wertebereich $W_X = \mathbb{R}^+ = [0, \infty)$ heisst *exponentialverteilt* mit Parameter $\lambda \in \mathbb{R}^+$ falls

$$f(x) = \begin{cases} \lambda \cdot \exp(-\lambda x), & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

Wir schreiben

$$X \sim \text{Exp}(\lambda)$$

Die zugehörige kumulative Verteilungsfunktion ist gegeben durch

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{falls } x < 0 \end{cases}$$

Die Dichte und kumulative Verteilungsfunktion sind für verschiedene λ in Abbildung 5.5 zu sehen.

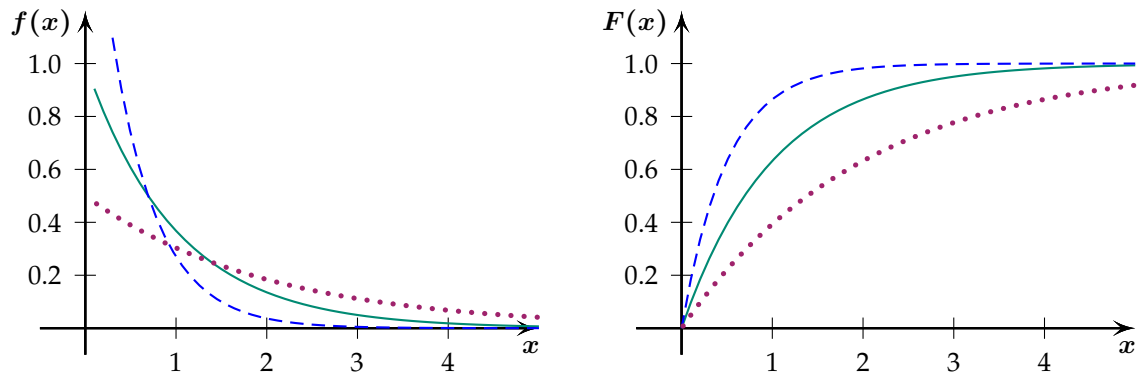


Abbildung 5.5.: Dichte und Verteilungsfunktion der Exponentialverteilung für $\lambda = 1$ (grün), $\lambda = 2$ (blau gestrichelt) und $\lambda = 1/2$ (violett gepunktet).

Für $X \sim \text{Exp}(\lambda)$ sind die Kennzahlen wie folgt:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \cdot \lambda \cdot e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

$$\sigma_X = \frac{1}{\lambda}$$

Beispiel 5.2.4

Wie lange dauert es, bis ein bestimmtes radioaktives Isotop mit einer Wahrscheinlichkeit von $1/2$ zerfallen ist?

Wir bezeichnen mit T die Zerfallszeit. Diese kann auch als Lebensdauer aufgefasst werden. Als Modell für diese zufällige Dauer ist die Exponentialverteilung geeignet, also

$$T \sim \text{Exp}(\lambda)$$

Der Parameter λ hängt vom jeweiligen Isotop ab und muss in der Regel aus Experimenten geschätzt werden. Für welchen Zeitpunkt wird die Wahrscheinlichkeit, dass das Isotop bis dahin zerfällt, resp. „überlebt“, gleich $1/2$? Die Antwort gibt der Median

$$F(t_{1/2}) = 1 - \exp(-\lambda t_{1/2}) = \frac{1}{2} \Rightarrow \exp(-\lambda t_{1/2}) = \frac{1}{2}$$

Diese Gleichung lösen wir nach $t_{1/2}$ auf, indem wir beide Seiten der Gleichung logarithmieren:

$$-\lambda t_{1/2} = \ln\left(\frac{1}{2}\right) \Rightarrow t_{1/2} = \frac{\ln(2)}{\lambda} = \frac{0.693}{\lambda}$$

In einer radioaktiven Sample gibt es allerdings sehr viele aktive Isotope. Wenn die Wahrscheinlichkeit, dass ein einzelnes Isotop bis zum Zeitpunkt $\frac{0.693}{\lambda}$ zerfällt, $1/2$ beträgt, dann ist die relative Häufigkeit der *überlebenden* Isotope im Sample zum Zeitpunkt $\frac{0.693}{\lambda}$ ebenfalls $1/2$. Die relative Häufigkeit der *zerfallenen* aktiven Isotope beträgt demnach ebenfalls $1/2$. Man nennt deshalb

$$t_{1/2} = \frac{\ln(2)}{\lambda}$$

die *Halbwertszeit*. Nach dieser Zeit sind im Mittel die Hälfte aller aktiven Isotope zerfallen.

□

Angenommen $X \sim \text{Exp}(3)$, dann lässt sich die Wahrscheinlichkeit $P(0 \leq X \leq 4)$ mit R wie folgt berechnen:

```
pexp(4, rate = 3)

## [1] 0.9999939
```

Es besteht folgender Zusammenhang zwischen der Exponential- und der Poisson-Verteilung:

Zusammenhang zwischen Poisson-Verteilung und Exponentialverteilung

Wenn die *Zeiten* zwischen den Ausfällen eines Systems $\text{exponential}(\lambda)$ -verteilt sind, dann ist die *Anzahl* Ausfälle in einem Intervall der Länge t $\text{Poisson}(\lambda t)$ -verteilt.

Beispiel 5.2.5

Angenommen, zum Zeitpunkt $t_0 = 0$ ereignet sich in einer radioaktiven Sample ein radioaktiver Zerfall. Wie gross ist die Wahrscheinlichkeit, dass erst nach dem Zeitpunkt t erneut ein Zerfall eintreten kann?

Die Wahrscheinlichkeit, dass sich erst nach der Zeit t wieder ein Zerfall ereignet, ist

$$P(T > t) = P(\text{kein Zerfall in } [0, t])$$

Die Anzahl Zerfälle im Zeitintervall $[0, t]$ folgt einer Poisson-Verteilung mit Parameter λt . Folglich ist

$$P(T > t) = P(\text{kein Zerfall in } [0, t]) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

Also folgt die Lebenszeit T eines Atoms einer Exponentialverteilung mit Parameter λ . Die kumulative Verteilungsfunktion ist gegeben durch

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t} \quad \text{für } t \geq 0$$

□

5.2.3. Normalverteilung (Gauss-Verteilung)

Die *Normalverteilung* (manchmal auch *Gauss-Verteilung* genannt) ist die häufigste Verteilung für Messwerte. Sie tritt in vielen Anwendungen auf und ist die wichtigste Wahrscheinlichkeitsverteilung in der Statistik. Sie hat neben praktischer auch grosse theoretische Bedeutung.

Normalverteilung

Eine Zufallsvariable X mit Wertebereich $W_X = \mathbb{R}$ heisst *normalverteilt* mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 \in \mathbb{R}^+$ falls

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Wir notieren die Verteilung für die Zufallsvariable X folgendermassen

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Die zugehörige kumulative Verteilungsfunktion

$$F(x) = \int_{-\infty}^x f(y) \, dy$$

ist nicht explizit mit Standardfunktionen wie x^2 , $\exp(x)$, $\log(x)$ etc. darstellbar. Diese Integrale werden (von Computersoftware) numerisch berechnet.

Für $X \sim \mathcal{N}(\mu, \sigma^2)$ sind die Kennzahlen wie folgt:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2$$

$$\sigma_X = \sigma$$

Das heisst, dass die Parameter μ und σ^2 eine natürliche Interpretation als Erwartungswert und Varianz der Verteilung haben. Drei Normalverteilungen mit verschiedenen Werten von μ und σ sind in Abbildung 5.6 dargestellt.

Die Dichte der Normalverteilung ist symmetrisch um den Erwartungswert μ . Je grösser σ , desto flacher oder breiter wird die Dichte. Für kleine σ gibt es einen „schmalen und hohen“ Gipfel. Mit μ verschieben wir einfach die Dichte nach links bzw. rechts.

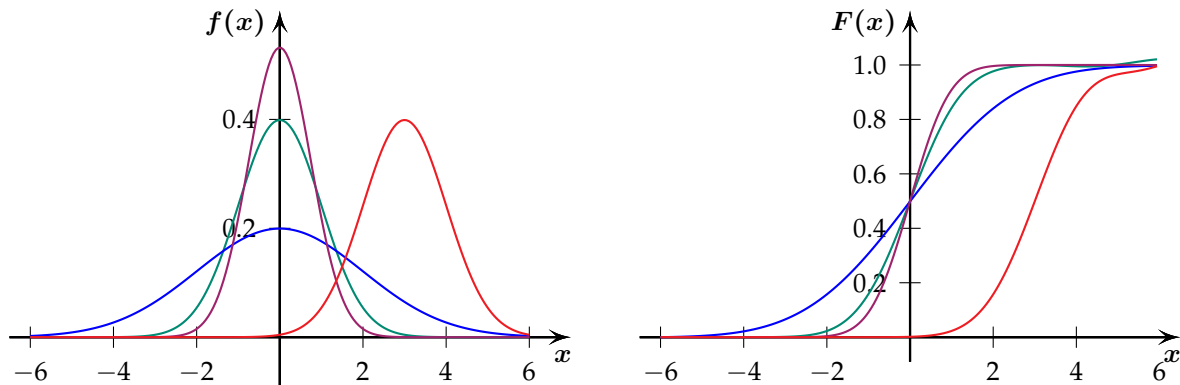


Abbildung 5.6.: Dichten (*links*) und kumulative Verteilungsfunktionen (*rechts*) der Normalverteilungen für $\mu = 0, \sigma = 1$ (grün), $\mu = 0, \sigma = 2$ (blau), $\mu = 0, \sigma = 0.75$ (violett) und $\mu = 3, \sigma = 1$ (rot).

Beispiel 5.2.6

Der Intelligenzquotient (IQ) wird in der Regel mit Intelligenztests ermittelt. Die Ergebnisse von einem IQ Test folgen in etwa einer Normalverteilung mit Mittelwert 100 und Standardabweichung 15. Im Allgemeinen gilt eine Person als hochbegabt, wenn ihr IQ zwei und mehr Standardabweichungen vom Mittelwert nach oben entfernt ist. Wir suchen die Wahrscheinlichkeit, dass jemand hochbegabt ist, also einen IQ von mehr als 130 hat. Dies ist die Wahrscheinlichkeit $P(X > 130)$, wobei

$$X \sim \mathcal{N}(100, 15^2)$$

Mit R suchen wir also die Wahrscheinlichkeit

$$1 - P(X \leq 130)$$

```
1 - pnorm(130, mean = 100, sd = 15)

## [1] 0.02275013
```

Also rund 2% der Bevölkerung ist hochbegabt.

Wir können uns auch fragen, wie viel Prozent der Bevölkerung innerhalb einer Standardabweichung vom Mittelwert liegen. Wir suchen also die Wahrscheinlichkeit

$$P(85 \leq X \leq 115)$$

```
pnorm(115, mean = 100, sd = 15) - pnorm(85, mean = 100,
sd = 15)

## [1] 0.6826895
```

D.h., etwa $\frac{2}{3}$ der Bevölkerung haben einen IQ zwischen 85 und 115.

□

Das letzte Resultat aus dem Beispiel oben gilt für alle Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$. Die Wahrscheinlichkeit, dass eine Beobachtung höchstens eine Standardabweichung vom Erwartungswert abweicht, ist etwa $\frac{2}{3}$:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx \frac{2}{3}$$

Wir können auch noch die Wahrscheinlichkeit berechnen, dass eine Beobachtung höchstens zwei Standardabweichungen vom Erwartungswert abweicht:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$$

Da diese Wahrscheinlichkeiten durch Integrale berechnet werden, können wir sie als Flächen interpretieren. Die Fläche der Normalverteilung über dem Intervall $[\mu - \sigma, \mu + \sigma]$ ist ca. $\frac{2}{3}$. Die Fläche über dem Intervall $[\mu - 2\sigma, \mu + 2\sigma]$ ist ca. 0.95, siehe auch Abbildung 5.7.

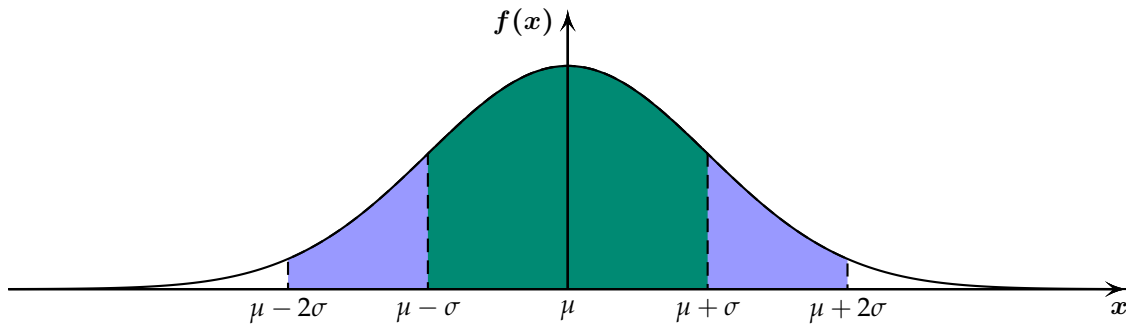


Abbildung 5.7.: Dichte der Normalverteilung. Ca. 66 % der Fläche befindet sich im Intervall $[\mu - \sigma, \mu + \sigma]$, ca. 95 % der Fläche im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$.

Die Standardnormalverteilung

Die Normalverteilung $\mathcal{N}(0, 1)$ trägt einen speziellen Namen.

Standardnormalverteilung

Die Normalverteilung mit $\mu = 0$ und $\sigma^2 = 1$ heisst **Standardnormalverteilung**. Deren Dichte und kumulative Verteilungsfunktion werden wie folgt bezeichnet:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

$$\Phi(x) = \int_{-\infty}^x \varphi(y) \, dy.$$

Wie die Werte von Φ mit R ermittelt werden können, betrachten wir im nächsten Beispiel.

Beispiel 5.2.7

Sei $Z \sim \mathcal{N}(0, 1)$.

- Was ist $P[Z \leq 1.13]$?

```
pnorm(1.13)
```

```
## [1] 0.8707619
```

- Für welchen Wert von z ist $\Phi(z) = 0.7910$?

Anders gefragt: Was ist $\Phi^{-1}(0.7910)$?

```
qnorm(0.791)
```

```
## [1] 0.8098959
```

Dies ist gerade das 0.7910-Quantil.

- Was ist $P[Z \leq -0.2]$?

Weil die Standardnormalverteilung um null herum symmetrisch ist, ist die Fläche links von -0.2 wegen der Symmetrie genau so gross wie die Fläche rechts von 0.2 . D.h.:

$$P[Z \leq -0.2] = P[Z \geq 0.2] = 1 - P[Z \leq 0.2]$$

Dies überprüfen wir leicht mit R

```
pnorm(-0.2)
```

```
## [1] 0.4207403
```

```
1 - pnorm(0.2)
```

```
## [1] 0.4207403
```

□

Wir werden sehen, dass eine Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ immer in eine Standard-Normalverteilung transformiert werden kann. Deshalb genügen an sich die Werte von Φ , um Wahrscheinlichkeiten und Quantile einer allgemeinen $\mathcal{N}(\mu, \sigma^2)$ -Verteilung zu berechnen.

5.3. Funktionen einer Zufallsvariable

Wenn $g : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion von \mathbb{R} nach \mathbb{R} und X eine Zufallsvariable ist, dann ist die Zusammensetzung

$$Y = g(X)$$

eine neue Zufallsvariable. Die Zusammensetzung bedeutet einfach, dass zu jeder Realisierung x von X die Realisierung $y = g(x)$ von Y gehört. Solche Transformationen treten häufig auf. Wir beginnen mit dem wichtigsten Spezialfall:

5.3.1. Lineare Transformationen von Zufallsvariablen

Wir betrachten hier zuerst den Fall einer **linearen Transformation**

$$y = g(x) = a + bx \quad (a, b \in \mathbb{R})$$

Beispiel 5.3.1

Wir haben in einer Messreihe Temperaturen in Grad Celsius gemessen und wollen diese nun in Grad Fahrenheit umformen. Die Temperatur T_C in Grad Celsius kann man folgendermassen in die Temperatur T_F in Grad Fahrenheit umrechnen:

$$T_F = \frac{9}{5} \cdot T_C + 32$$

□

Eigenschaften von linearen Transformationen einer Zufallsvariablen

Für

$$Y = a + bX$$

gelten dann folgende Beziehungen:

- (i) $E(Y) = E(a + bX) = a + b E(X)$
- (ii) $\text{Var}(Y) = \text{Var}(a + bX) = b^2 \text{Var}(X), \quad \sigma_Y = |b| \sigma_X$
- (iii) α – Quantil von $Y = q_Y(\alpha) = a + b q_X(\alpha)$
- (iv) $f_Y(y) = \frac{1}{b} f_X\left(\frac{y-a}{b}\right)$

Die Richtigkeit der ersten Gleichung können wir uns leicht überlegen. Für den Spezialfall

$$Y = a + X$$

werden die Beobachtungen X alle um a verschoben und damit verschiebt sich auch der Erwartungswert um a . Für den anderen Spezialfall

$$Y = bX$$

werden alle Beobachtungen X mit b multipliziert. Somit wird auch der Erwartungswert von X mit b multipliziert.

Kapitel 5. Modelle für Messdaten

Der Vollständigkeit halber folgen noch die Herleitungen der 4 Regeln.

Die erste Gleichung folgt aus

$$\begin{aligned} E(Y) &= E(a + bX) \\ &= \int_{-\infty}^{\infty} (a + bx) f_X(x) dx \\ &= a \int_{-\infty}^{\infty} f_X(x) dx + b \int_{-\infty}^{\infty} x f_X(x) dx \\ &= a \cdot 1 + b \cdot E(X) \\ &= a + b E(X) \end{aligned}$$

Die zweite Gleichung folgt aus der Definition der Varianz und aus der ersten Gleichung:

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(a + bX) \\ &= E((a + bX - E(a + bX))^2) \\ &= E((a + bX - (a + bE(X)))^2) = E(b^2(X - E(X))^2) \\ &= b^2 E((X - E(X))^2) = b^2 \text{Var}(X) \end{aligned}$$

Durch Ziehen der Wurzel erhält man für die Standardabweichung $\sigma_Y = |b| \cdot \sigma_X$.

Die dritte Gleichung folgt aus

$$\begin{aligned} F_Y(q_Y(\alpha)) &= P(Y \leq q_Y(\alpha)) \\ &= P(a + bX \leq q_Y(\alpha)) \\ &= P\left(X \leq \frac{q_Y(\alpha) - a}{b}\right) \\ &= F_X\left(\frac{q_Y(\alpha) - a}{b}\right) \end{aligned}$$

Aus

$$F_Y(q_Y(\alpha)) = \alpha = F_X\left(\frac{q_Y(\alpha) - a}{b}\right)$$

folgt $\frac{q_Y(\alpha) - a}{b} = q_X(\alpha)$ und somit die dritte Gleichung. Die vierte Beziehung erhält man aus

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X\left(\frac{y - a}{b}\right) \\ &= \frac{1}{b} f_X\left(\frac{y - a}{b}\right) \end{aligned}$$

Wir haben dabei folgende Gleichheit benützt

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P(aX + b \leq y) \\&= P\left(X \leq \frac{y-b}{a}\right) \\&= F_X\left(\frac{y-b}{a}\right)\end{aligned}$$

Beispiel 5.3.2

Wir haben eine Temperatur in Grad Celsius gemessen und kennen die Standardabweichung des Messfehlers auf dieser Skala: $\sigma_C = 1/3$ Grad Celsius. Für einen Bericht, der im englischsprachigen Raum gelesen werden soll, wollen wir die Temperatur aber nicht in Grad Celsius, sondern in Grad Fahrenheit angeben. Wie gross ist die Standardabweichung σ_F des Messfehlers, wenn wir die Temperatur in Grad Fahrenheit angeben?

Wie wir schon gesehen haben, können wir die Temperatur T_C in Grad Celsius in die Temperatur T_F in Grad Fahrenheit umrechnen:

$$T_F = \frac{9}{5} \cdot T_C + 32$$

Daher ist die Standardabweichung in Grad Fahrenheit

$$\sigma_F = \frac{9}{5}\sigma_C = \frac{9}{5} \cdot \frac{1}{3} = \frac{3}{5}$$

□

Beispiel 5.3.3 Fahrenheit für Fortgeschrittene

Wenn die Messungen von T_C einer Normalverteilung mit Mittelwert μ_C und Standardabweichung σ_C folgen, also $T_C \sim \mathcal{N}(\mu_C, \sigma_C^2)$, dann ist

$$T_F = a + bT_C \sim \mathcal{N}(a + b\mu_C, b^2\sigma_C^2)$$

mit $a = 32$ und $b = 9/5$. Denn nach der Regel (iv) der Regeln für lineare Transformationen gilt

$$\begin{aligned} f_{T_F}(T_F) &= \frac{1}{\sqrt{2\pi}\sigma_C b} \exp\left(-\frac{\left(\frac{T_F-a}{b} - \mu_C\right)^2}{2\sigma_C^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_C b} \exp\left(-\frac{(T_F - (a + b\mu_C))^2}{2\sigma_C^2 b^2}\right). \end{aligned}$$

Eine linear transformierte Normalverteilung ist also wiederum eine Normalverteilung. Diese Eigenschaft, dass man mit linearen Transformationen innerhalb der Verteilungsklasse bleibt, ist eine spezielle Eigenschaft der Normalverteilung und im allgemeinen nicht richtig.

□

Standardisieren einer Zufallsvariablen

Wir können X immer linear transformieren, so dass die transformierte Zufallsvariable Erwartungswert = 0 und Varianz = 1 hat. Dies geschieht wie folgt: betrachte die lineare Transformation

$$g(x) = \frac{x - E(X)}{\sigma_X} = -\frac{E(X)}{\sigma_X} + \frac{1}{\sigma_X}x = a + bx$$

mit

$$a = -\frac{E(X)}{\sigma_X} \quad \text{und} \quad b = \frac{1}{\sigma_X}$$

Damit bilden wir die transformierte Zufallsvariable

$$Z = g(X) = \frac{X - E(X)}{\sigma_X}$$

Mit Hilfe der Regeln für lineare Transformationen gilt dann:

$$E(Z) = a + b E(X) = 0, \quad \text{Var}(Z) = b^2 \text{Var}(X) = 1$$

Standardisieren einer normalverteilten Zufallsvariablen

Falls $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist die standardisierte Zufallsvariable wieder normalverteilt, hat nun aber Erwartungswert null und Varianz eins. Man erhält also die

Standardnormalverteilung:

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Damit lassen sich Wahrscheinlichkeiten für beliebige Normalverteilungen mit Hilfe der Standardnormalverteilung berechnen. Es reicht daher, wenn man nur die Werte der Standardnormalverteilung mit R berechnen kann.

Beispiel 5.3.4

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = 2$ und $\sigma^2 = 4$. Berechnen Sie $P(X \leq 5)$.

$$P(X \leq 5) = P\left(\frac{X - \mu}{\sigma} \leq \frac{5 - \mu}{\sigma}\right) = P\left(Z \leq \frac{5 - 2}{2}\right) = P(Z \leq 1.5) = \Phi(1.5) = 0.93$$

□

Beispiel 5.3.5

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = 2$ und $\sigma^2 = 4$. Berechnen Sie $P(X > 5)$.

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \Phi(1.5) = 1 - 0.93 = 0.07$$

□

Beispiel 5.3.6

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = 2$ und $\sigma^2 = 4$. Wie gross ist das 90 % Quantil γ von X ?

$$\begin{aligned} P(X \leq \gamma) &= 0.9 \\ \Rightarrow P\left(Z \leq \frac{\gamma - \mu}{\sigma}\right) &= 0.9 \\ \Rightarrow \Phi\left(\frac{\gamma - \mu}{\sigma}\right) &= 0.9 \end{aligned}$$

Mit Hilfe einer Statistik-Software finden wir, dass

$$\frac{\gamma - \mu}{\sigma} = \Phi^{-1}(0.9) = 1.28$$

Auflösen nach γ ergibt:

$$\gamma = \mu + 1.28\sigma = 2 + 1.28 \cdot 2 = 4.56$$

□

Beispiel 5.3.7

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = 2$ und $\sigma^2 = 4$. Berechnen Sie $P(|X| \leq 2)$.

$$\begin{aligned} P(|X| \leq 2) &= P(-2 \leq X \leq 2) = P(X \leq 2) - P(X \leq -2) \\ &= P\left(Z \leq \frac{2-2}{2}\right) - P\left(Z \leq \frac{-2-2}{2}\right) = P(Z \leq 0) - P(Z \leq -2) \\ &= \Phi(0) - \Phi(-2) = \Phi(0) - (1 - \Phi(2)) = 0.5 - (1 - 0.97) \\ &= 0.5 - 0.03 = 0.47 \end{aligned}$$

□

Bemerkungen:

- i. Wir hätten alle diese Aufgaben auch ohne Standardisierung mit dem R-Befehl

```
pnorm(..., mean = 2, sd = 2)    #bzw.
qnorm(..., mean = 2, sd = 2)
```

lösen können. Da stellt sich natürlich die Frage, *warum* man Zufallsvariablen standardisiert. Vor dem Computerzeitalter wurden die Werte für Φ in Tabellen aufgeführt. Damals musste man normalverteilte Zufallsvariablen standardisieren, da es nicht möglich war, für sämtliche Parameter μ und σ^2 entsprechende Tabellen anzufertigen.

- ii. Wir werden die Standardisierung von Zufallsvariablen allerdings später noch bei der sogenannten t -Verteilung benötigen.

5.3.2. Nichtlineare Transformationen von Zufallsvariablen

Wir beschränken uns einfachheitshalber zuerst auf den Fall einer **quadratischen Transformation**

$$y = g(x) = x^2$$

Beispiel 5.3.8

Es bezeichne $X \sim \text{Uniform}([a, b])$ eine uniformverteilte Zufallsvariable. Wir definieren die Zufallsvariable Y durch die quadratische Transformation der Zufallsvariablen X

$$Y = X^2$$

Wie lautet also der Erwartungswert $E[Y]$? Wir könnten nun versucht sein, zuerst die Dichte von Y , also $f_Y(Y)$ zu bestimmen und dann aufgrund von $f_Y(y)$ den Erwartungswert $E[Y]$ zu berechnen. Es stellt sich allerdings heraus¹, dass die Berechnung des Erwartungswertes von Y viel einfacher geht:

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2 + ab + b^2}{3}$$

□

Beispiel 5.3.9 Boltzmann-Verteilung - für Fortgeschrittene

Wir bezeichnen mit V den Geschwindigkeitsbetrag eines Gasmoleküls. Gemäss der kinetischen Gastheorie ist der Betrag der Geschwindigkeit eines Gasmoleküls zufällig und folgt der Wahrscheinlichkeitsverteilung mit der Wahrscheinlichkeitsdichte

$$f_V(v) = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} v^2 \exp \left(-\frac{mv^2}{2k_B T} \right),$$

wobei k_B die Boltzmann-Konstante und T die Temperatur in Kelvin, m die Masse des Gasmoleküls bezeichnet. Aufgrund dieser Wahrscheinlichkeitsdichte können wir den Erwartungswert des Geschwindigkeitsbetrags $E(V)$ eines Gasmoleküls in einem Gas bestimmen. Wir interessieren uns aber für die mittlere kinetische Energie $Y = \frac{1}{2}mV^2$.

Wir könnten nun wiederum zuerst die Dichte von Y , also $f_Y(Y)$ zu bestimmen versuchen und dann aufgrund von $f_Y(y)$ den Erwartungswert der kinetischen Energie

¹Dies kann in einer nicht besonders aufwendigen Rechnung bewiesen werden, siehe Anhang.

$E(Y)$ berechnen. Die Berechnung des Erwartungswertes der kinetischen Energie gestaltet sich aber analog zum vorhergehenden Beispiel direkter:

$$E(Y) = \int_0^{\infty} \frac{1}{2} m v^2 f_V(v) dv = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} \int_0^{\infty} v^2 \exp \left(-\frac{m v^2}{2 k_B T} \right) dv = \frac{3}{2} k_B T.$$

□

Allgemeine Transformationen von Zufallsvariablen

Wir betrachten den Fall einer **allgemeinen Transformation** der Zufallsvariablen X

$$Y = g(X)$$

Die kumulative Verteilungsfunktion und die Dichte von Y sind durch die Verteilungsfunktion und die Dichte von X bestimmt. Für den Erwartungswert gilt stets die folgende Formel

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

5.4. Funktionen von mehreren Zufallsvariablen

Im letzten Unterkapitel haben wir untersucht, wie die Funktion *einer* Zufallsvariable verteilt ist. In den meisten Anwendungen hat man es aber nicht mit einer, sondern mit *mehreren* Zufallsvariablen zu tun. Üblicherweise misst man die *gleiche* Grösse mehrmals; man hat zum Beispiel mehrere Individuen, oder man wiederholt die Messungen. In diesem Abschnitt untersuchen wir, wie eine Funktion mehrerer Zufallsvariablen verteilt ist.

Die Messungen x_1, x_2, \dots, x_n fassen wir als Realisierungen der Zufallsvariablen

$$X_1, \dots, X_n$$

auf. Diese Auffassung, dass X_i die i -te Wiederholung von unserem Zufallsexperiment

ist, ist oft bequemer als die Interpretation, dass die Messungen n unabhängige Realisierungen einer Zufallsvariablen X sind.

Beispiel 5.4.1

Wir machen 20 Messungen der Wasserverschmutzung in einem See. Wir haben also Messungen

$$x_1, x_2, \dots, x_{20}$$

die Realisierungen der Zufallsvariablen

$$X_1, X_2, \dots, X_{20}$$

darstellen. Wir gehen davon aus, dass diese 20 Zufallsvariablen Wahrscheinlichkeitsverteilungen haben, die gleich sind, da die Wasserproben alle aus demselben See stammen und mit der identischen Methode durchgeführt werden. Uns interessiert nun der Durchschnitt dieser Messungen und die Verteilung der zugehörigen Zufallsvariablen. Dazu benötigen wir eine Theorie für Funktionen von mehreren Zufallsvariablen.

□

Funktionen der Messwerte x_1, x_2, \dots, x_n haben die Form:

$$y = g(x_1, \dots, x_n)$$

Diese Funktion hat als Input n unabhängige Variablen und eine reelle Zahl als Output. Wenn x_1, x_2, \dots, x_n Realisierungen der Zufallsvariablen X_1, \dots, X_n sind, dann ist y eine Realisierung der Zufallsvariablen

$$Y = g(X_1, \dots, X_n)$$

Wir betrachten hier vor allem die Spezialfälle **Summe**

$$g(X_1, \dots, X_n) = S_n = X_1 + \dots + X_n = \sum_{i=1}^n X_i$$

und **arithmetisches Mittel**

$$g(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} S_n$$

Das arithmetische Mittel der Daten \bar{x}_n ist also eine Realisierung der Zufallsvariablen \bar{X}_n .

Wir sind an der Verteilung der Zufallsvariablen \bar{X}_n interessiert: Die Kenntnis dieser Verteilung wird uns erlauben, Statistik aufgrund von arithmetischen Mitteln von Daten zu machen.

5.4.1. Unabhängigkeit und i.i.d. Annahme

Oft treffen wir die Annahme, dass die Zufallsvariablen X_1, \dots, X_n **unabhängig** voneinander sind. Anschaulich heisst das, es gibt keine gemeinsamen Faktoren, die den Ausgang der verschiedenen Messungen beeinflussen und keine „carry over“ Phänomene von einer Messung zur nächsten.

Wenn die Zufallsvariablen X_1, \dots, X_n unabhängig sind und alle **dieselbe** Verteilung haben, dann schreiben wir das kurz als

$$X_1, \dots, X_n \text{ i.i.d.}$$

Die Abkürzung i.i.d. steht für:

independent, **i**dentically **d**istributed

Wir werden meistens mit dieser i.i.d. Annahme arbeiten, da wir oft von n Durchführungen des gleichen Experimentes ausgehen. Welche Verteilung die X_i 's haben, lassen wir offen. Oft handelt es sich um eine Normalverteilung, dies muss aber nicht so sein. Die Unabhängigkeit spielt eine Rolle bei den **Regeln für Erwartungswerte und Varianzen** von Summen. Die Beziehung

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

gilt immer,

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$$

jedoch nur, wenn X_1 und X_2 unabhängig sind.

5.4.2. Kennzahlen von S_n und \bar{X}_n

Beispiel 5.4.2 Relative Häufigkeit von fairen Münzwürfen

Wirft man eine faire Münze bloss ein paar wenige Male, so ist das Verhältnis von Würfeln mit Kopf zur Anzahl Würfe mit Zahl noch nicht genau $1/2$. Es ist aber eine vernünftige Annahme, dass dieses Verhältnis für sehr viele Würfe mit einer fairen Münze von Würfeln etwa $1/2$ ergibt.

Der südafrikanische Mathematiker John Kerich hat diese Vermutung als Kriegsgefangener während des zweiten Weltkrieges getestet. Er hat 10'000 mal eine Münze geworfen und beobachtete dabei 5067 mal Kopf. Die aufeinanderfolgenden Münzwürfe

sind modelliert als unabhängige Zufallsvariablen X_i . Die Zufallsvariable nimmt entweder den Wert 0 oder 1 an, je nachdem, ob der i -te Wurf Zahl oder Kopf ergibt. Die relative Häufigkeit von Kopf in n Versuchen ist dann

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

In John Kerichs Zufallsexperiment ergab sich für \bar{x}_{10000} also der Wert 0.5067

$$\bar{x}_{10000} = 0.5067$$

Hätte er die Münze bloss $n = 10$ geworfen, wäre ein Wert von $\bar{x}_{10} = 0.7$ oder ein Wert von $\bar{x}_{10} = 0.3$ nicht sehr überraschend gewesen, selbst wenn die Münze fair ist. Allerdings würde man die Fairness der Münze stark anzweifeln, wenn von 10'000 Würfeln bloss 3000 mal Kopf geworfen würde.

□

Wir nehmen in diesem Abschnitt an, dass

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{kumulative Verteilungsfkt. } F$$

Wegen dem zweiten „i“ in i.i.d. hat jedes X_i dieselbe Verteilung und dieselben Kennzahlen:

$$E(X_i) = \mu \quad \text{und} \quad \text{Var}(X_i) = \sigma_X^2$$

Die Kennzahlen von S_n und \bar{X}_n folgen dann aus den allgemeinen Regeln für Erwartungswert und Varianz von Summen:

Kennzahlen von S_n

$$E(S_n) = E(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n E(X_i) = n\mu$$

$$\text{Var}(S_n) = \sum_{i=1}^n \text{Var}(X_i) = n \text{Var}(X_i)$$

$$\sigma(S_n) = \sqrt{n}\sigma_X$$

Kennzahlen von \bar{X}_n

$$E(\bar{X}_n) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} n E(X_i) = \mu$$

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} n \sigma_X^2 = \frac{\sigma_X^2}{n}$$

$$\sigma(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}}$$

Die Standardabweichung von \bar{X}_n heisst auch der **Standardfehler** des arithmetischen Mittels.

Die Standardabweichung der Summe wächst also mit wachsendem n , aber langsamer als die Anzahl Beobachtungen n . D.h. auf einer relativen Skala haben wir eine kleinere Streuung für wachsendes n .

Der Erwartungswert von \bar{X}_n ist also gleich demjenigen einer einzelnen Zufallsvariablen X_i , die *Streuung nimmt jedoch ab mit wachsendem n* .

Gesetz der grossen Zahlen

Für $n \rightarrow \infty$ geht die Streuung gegen null. Es gilt das **Gesetz der grossen Zahlen**: Falls X_1, \dots, X_n i.i.d., dann

$$\bar{X}_n \longrightarrow \mu \quad \text{für } n \rightarrow \infty$$

Standardfehler

Die Standardabweichung des arithmetischen Mittels (**Standardfehler**) ist jedoch *nicht* proportional zu $1/n$, sondern nimmt nur ab mit dem Faktor $1/\sqrt{n}$

$$\sigma_{\bar{X}_n} = \frac{1}{\sqrt{n}} \sigma_X$$

Um den **Standardfehler** zu halbieren, braucht man also *viermal* so viele Beobachtungen. Dies nennt man auch das **\sqrt{n} -Gesetz**.

Beispiel 5.4.3

Im Falle einer fairen Münze können wir X_i auffassen als

$$X_i \sim \text{Bernoulli}(\pi = 0.5)$$

mit

$$E(X_i) = \pi = \frac{1}{2}, \quad \text{Var}(X_i) = \pi(1 - \pi) = \frac{1}{4}, \quad \sigma_{X_i} = \sigma_X = \sqrt{\pi(1 - \pi)} = \frac{1}{2}$$

Im Falle von John Kerichs Experiment ergibt sich für die Kennzahl S_n mit $n = 10000$

$$E[S_{n=10000}] = n E(X_i) = 5000$$

$$\text{Var}[S_{n=10000}] = n \text{Var}(X_i) = 10'000 \cdot \frac{1}{4} = 2500$$

$$\sigma(S_n) = 50$$

Die Kennzahlen von der relativen Häufigkeit berechnen sich zu:

$$E[\bar{X}_{n=10000}] = E(X_i) = 0.5$$

$$\text{Var}[\bar{X}_{n=10000}] = \frac{1}{n} \text{Var}(\bar{X}_i) = \frac{1}{10'000} \cdot \frac{1}{4} = 0.000025$$

$$\sigma(\bar{X}_n) = \frac{\sigma_X}{\sqrt{n}} = \frac{0.5}{100} = 0.005$$

Die von John Kerich beobachtete relative Häufigkeit von Kopf ist 0.5067 und ist somit etwas ausserhalb von einer Standardweichung vom erwarteten Mittelwert oder vom Standardfehler von 0.005.

Würde John Kerich sein Experiment mit $n = 10'000$ mehrere Mal wiederholen, so würde er für die Standardabweichung des Mittelwertes also etwa 0.005 erwarten.

□

5.4.3. Verteilungen von S_n und \bar{X}_n

Die Verteilungen von S_n und \bar{X}_n sind im allgemeinen schwierig anzugeben. Es gibt folgende Ausnahmen, unter der Annahme, dass die X_1, \dots, X_n i.i.d. sind:

1. Wenn $X_i \in \{0, 1\}$, dann ist

$$S_n \sim \text{Bin}(n, \pi) \quad \text{mit} \quad \pi = P(X_i = 1)$$

2. Wenn $X_i \sim \text{Pois}(\lambda)$, dann ist

$$S_n \sim \text{Pois}(n\lambda)$$

3. Wenn $X_i \sim \mathcal{N}(\mu, \sigma^2)$, dann ist

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2) \quad \text{und} \quad \bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right)$$

Falls die einzelnen X_i 's nicht normalverteilt sind, so gilt erstaunlicherweise die obige Verteilungsformel c) immer noch approximativ. Dies liefert der berühmte Zentrale Grenzwertsatz².

Zentraler Grenzwertsatz

Falls X_1, \dots, X_n i.i.d. mit irgendeiner Verteilung mit Erwartungswert μ und Varianz σ^2 , dann gilt

$$S_n \approx \mathcal{N}(n\mu, n\sigma_X^2)$$

$$\bar{X}_n \approx \mathcal{N}(\mu, \sigma_X^2/n)$$

wobei die Approximation im allgemeinen besser wird mit grösserem n . Über-

²Einen Beweis für diesen sehr bedeutenden Satz finden Sie in Kapitel A.8 des Anhangs.

dies ist auch die Approximation besser, je näher die Verteilung von X_i bei der Normalverteilung $\mathcal{N}(\mu, \sigma_X^2)$ ist.

Selbst wenn wir die Verteilung der X_i nicht kennen, so haben wir eine Ahnung über die approximative Verteilung von S_n und X_n ! Der Zentrale Grenzwertsatz (ZGWS) ist mitunter ein Grund für die Wichtigkeit der Normalverteilung.

Wir haben schon bei der Binomialverteilung gesehen, dass diese für grosse n „glockenförmig“ aussieht. Dasselbe gilt für die Poissonverteilung für grösser werdendes λ . Man kann daher die Normalverteilung verwenden, um die Binomialverteilung mit grossem n zu approximieren (denn die Binomialverteilung ist eine i.i.d. Summe von Bernoulliverteilungen). Man spricht dann von der sogenannten **Normalapproximation** der Binomialverteilung.

Wenn

$$X \sim \text{Bin}(n, \pi)$$

dann haben wir

$$E(X) = n\pi \quad \text{und} \quad \text{Var}(X) = n\pi(1 - \pi)$$

Für grosse n können wir also X gemäss dem ZGWS approximativ als Normalverteilung mit Erwartungswert $E(X) = n\pi$ und Varianz $\sigma^2 = n\pi(1 - \pi)$ behandeln. D.h. es gilt dann

$$X \sim \text{Bin}(n, \pi) \approx \mathcal{N}(n\pi, n\pi(1 - \pi))$$

und damit

$$P(X \leq x) \approx \Phi\left(\frac{x - n\pi}{\sqrt{n\pi(1 - \pi)}}\right)$$

Beispiel 5.4.4

Wie gross ist die Wahrscheinlichkeit, dass bei 10 000 Würfeln mit einer Münze maximal 5100 mal Kopf erscheint?

Die Anzahl Würfe, bei denen Kopf erscheint, ist $\text{Bin}(10\,000, 0.5)$ -verteilt. Diese Verteilung approximieren wir mit einer Normalverteilung, d.h.,

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

mit

$$\mu = 10\,000 \cdot 0.5 = 5000 \quad \text{und} \quad \sigma^2 = 10000 \cdot 0.5 \cdot (1 - 0.5) = 2500$$

Und wir erhalten

$$X \sim \text{Bin}(10\,000, 0.5) \approx \mathcal{N}(5000, 2500)$$

Von Interesse ist

$$P(X \leq 5100) = \Phi\left(\frac{5100 - 5000}{\sqrt{2500}}\right) = \Phi(2) \approx 0.98$$

Mit R erfolgt die Berechnung folgendermassen

```
pnorm(5100, 5000, sqrt(2500))  
  
## [1] 0.9772499
```

Vergleichen wir das vorhergehende Resultat mit der Verteilung

$$X \sim \text{Bin}(10\,000, 0.5)$$

dann erhalten wir

$$P(X \leq 5100) \approx 0.98$$

```
pbinom(5100, 10000, 0.5)  
  
## [1] 0.9777871
```

Die Übereinstimmung ist also sehr gut.



Beispiel 5.4.5

Wir ziehen $n = 10$ Zufallszahlen X_i . Die zehn Zufallsvariablen sind unabhängig und es gilt für jedes i

$$X_i \sim \text{Uniform}([0, 1])$$

Wie gross ist die Wahrscheinlichkeit, dass die Summe der Zufallszahlen

$$S_{10} = \sum_{i=1}^{10} X_i$$

grösser als sechs ist?

D. h., wir suchen

$$P[S_{10} > 6]$$

Aus Abschnitt 5.2.1 wissen wir, wie man Erwartungswert und Varianz von jedem X_i berechnet:

$$E(X_i) = \frac{1+0}{2} = 0.5 \quad \text{und} \quad \text{Var}(X_i) = \frac{(1-0)^2}{12} = \frac{1}{12}$$

Aus dem Zentralen Grenzwertsatz folgt:

$$S_n \approx \mathcal{N}(n E(X_i), n \text{Var}(X_i)) = \mathcal{N}\left(5, \frac{10}{12}\right) = \mathcal{N}(5, 0.83)$$

Damit kommen wir zu folgender Lösung:

$$\begin{aligned} P(S_n > 6) &= 1 - P(S_n \leq 6) = 1 - P\left(\frac{S_n - 5}{\sqrt{0.83}} \leq \frac{6 - 5}{\sqrt{0.83}}\right) = 1 - P(Z \leq 1.1) \\ &= 1 - \Phi(1.1) = 1 - 0.86 = 0.14 \end{aligned}$$

□

Für eine exakte Formulierung des Zentralen Grenzwertsatzes betrachtet man die standardisierte Zufallsvariable

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma_X}$$

Diese ist ungefähr $\mathcal{N}(0, 1)$ verteilt, was bedeutet, dass für alle x gilt

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \Phi(x)$$

Der Zentrale Grenzwertsatz gilt auch für diskrete Zufallsvariablen X_i .

5.4.4. Fehlerrechnung bei Messreihen

Messungen physikalischer Größen sind grundsätzlich fehlerbehaftet, d. h., man erhält Messwerte, die vom wahren Wert mehr oder weniger abweichen. Je nach Ursache der Abweichung unterscheidet man zwischen *systematischen* und *zufälligen* Fehlern. Systematische Fehler rühren von der Unvollkommenheit der Messgeräte her, wie zum Beispiel Funktionsfehler und Eichfehler, oder von der Unvollkommenheit der Messverfahren.

Beispiel 5.4.6 Systematische Fehler

Beispiele von systematischen Messfehlern:

1. Bei Kurzschluss der Eingänge zeigt ein Voltmeter nicht mehr 0 V an (Nullpunktsfehler).

2. Durch den Innenwiderstand eines Voltmeters sind gemessene Strom- oder Spannungswerte stets zu klein.
3. Bei einem Pendelversuch wird durch die Luft- und Lagerreibung die Schwingung gedämpft, wodurch die Frequenz der Schwingung verringert wird.

□

Systematische Fehler sollten nach Möglichkeit vermieden oder klein gehalten werden. Sie sind jedoch nicht Gegenstand einer Fehlerrechnung.

Zufällige Fehler besitzen beiderlei Vorzeichen (im Gegensatz zu systematischen Fehlern) und entstehen vor allem durch die Naturgesetze selber wie zum Beispiel aufgrund der statistischen Natur von Kernzerfällen, durch Ungeschicklichkeit beim Messen oder durch statistisch schwankende äussere und innere Einflüsse. Beispiele für solche schwankenden Einflüsse sind Schwankungen der Umgebungsparameter wie Druck, Temperatur, Luftfeuchtigkeit oder Schwankungen der Ausgangsparameter.

Wird die Messung im Rahmen einer Messreihe mehrfach durchgeführt, so streuen die Messwerte um den arithmetischen Mittelwert

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

Je grösser die Messreihe ist (also je grösser n), um so näher liegt der arithmetische Mittelwert am wahren Wert und umso kleiner wird der Fehler des Mittelwertes. Der Fehler des Mittelwertes ist durch den *Standardfehler* gegeben

$$s_{\bar{x}_n} = \frac{s_x}{\sqrt{n}}$$

wobei die *empirische Standardabweichung* der Messreihe

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

unabhängig von der Anzahl Datenpunkte ist.

Absoluter und relativer Fehler

Der Standardfehler der Messreihe entspricht dem **absoluten Fehler**, den man im Zusammenhang mit dem arithmetischen Mittelwert folgendermassen angibt:

$$\bar{x}_n \pm s_{\bar{x}_n} = \bar{x}_n \pm \frac{s_x}{\sqrt{n}}$$

Der **relative Fehler** wird folgendermassen angegeben:

$$\bar{x}_n \pm \frac{s_{\bar{x}_n}}{\bar{x}_n}$$

Bemerkungen:

- i. *Darstellung von Messwerten:* Mittelwert und Fehler sollen mit gleich vielen Dezimalstellen geschrieben werden, wobei der Mittelwert genauso viele signifikante Stellen hat wie das ungenaueste Messergebnis in der Messreihe.
- ii. *Signifikante Stellen:* Die folgenden Ausdrücke haben zwei signifikante Stellen:

$$0.0012, \quad 0.012, \quad 0.12, \quad 1.2, \quad 12$$

- iii. Sowohl zum arithmetischen Mittelwert als auch zum Fehler des Mittelwerts gehören Einheiten. Die Einheit ist für beide dieselbe wie für die Messgrösse.

Beispiel 5.4.7 Umlaufzeit von Plattentellern

Umlaufzeit eines Plattentellers mit **absolutem Fehler** des Mittelwertes:

$$T = (1.817 \pm 0.012) \text{ s}$$

Umlaufzeit eines Plattentellers mit **relativem Fehler** des Mittelwertes:

$$T = 1.817 \text{ s} \pm \frac{0.012 \text{ s}}{1.817 \text{ s}} = 1.817 \text{ s} \pm 0.66 \%$$

□

Beispiel 5.4.8 Fallzeiten

Wir haben 100 mal die Fallzeit eines Körpers gemessen. Diese Messungen sind in der Abbildung 5.8 graphisch dargestellt. Der arithmetische Mittelwert der 100 Messungen ergibt

$$\bar{t}_{100} = 4.2 \text{ s}$$

die empirische Standardabweichung

$$s_t = 0.71 \text{ s}$$

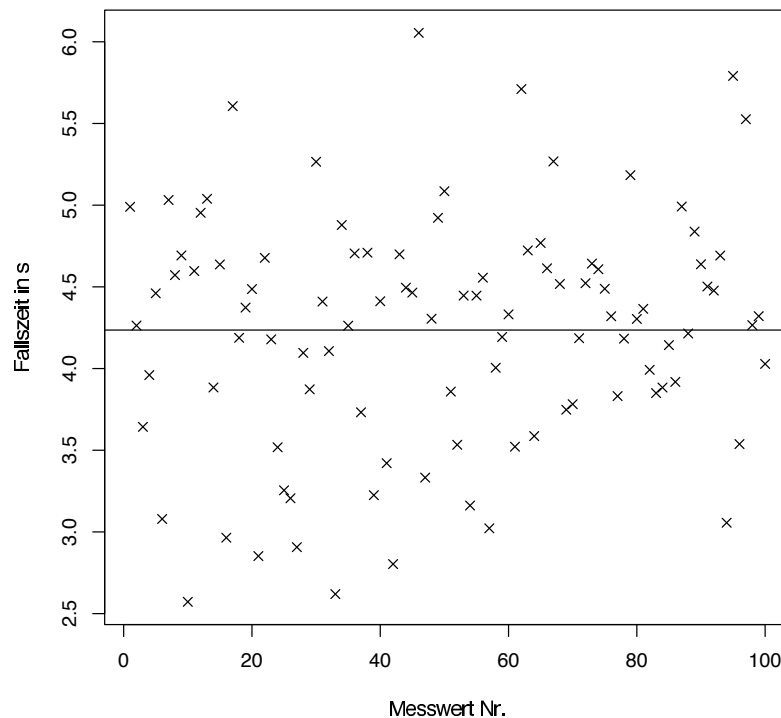


Abbildung 5.8.: 100 Messungen der Fallzeit eines Körpers.

Der Fehler, der uns allerdings interessiert, ist der statistische Fehler des Mittelwertes, also der Standardfehler

$$s_{\bar{t}_n} = \frac{s_t}{\sqrt{n}}$$

Der Fehler des Mittelwerts oder Standardfehler ergibt für diese Messreihe 0.071 s. Wir schreiben also für das Ergebnis unserer Messung mit absolutem Fehler

$$T = (4.2 \pm 0.1) \text{ s}$$

und mit relativem Fehler

$$T = 4.2 \text{ s} \pm 1.7 \%$$

Man beachte, dass im Falle des absoluten Fehlers eine Nachkommastelle angegeben wurde und im Falle des relativen Fehlers zwei signifikante Stellen. Würde man nur die ersten 10 Zeitmessungen zur Mittelwertbildung heranziehen, erhielte man nach korrekter Rundung folgendes Ergebnis für den arithmetischen Mittelwert: 4.1 s. Die empirische Standardabweichung ergibt allerdings in diesem Fall 0.82 s. Den Fehler des Mittelwerts oder Standardfehler ermitteln wir zu 0.26 s und das Endergebnis schreiben wir im Falle von 10 Datenpunkten mit absolutem Fehler

$$T = (4.1 \pm 0.3) \text{ s}$$

und mit relativem Fehler

$$T = 4.1 \text{ s} \pm 6.3 \%$$

Es ist ersichtlich, dass der Fehler des Mittelwertes in diesem Fall grösser ist und somit das Messergebnis insgesamt schlechter zu bewerten ist.

□

Lernziele

- Sie verstehen das Konzept der Wahrscheinlichkeitsdichte und der kumulativen Verteilungsfunktion; Sie kennen die uniforme Verteilung, die Exponentialverteilung und die Normalverteilung. Sie können die Normalverteilung standardisieren und mit dem Computer für beliebige Normalverteilungen Wahrscheinlichkeiten ausrechnen.
- Sie können den Erwartungswert, Varianz und Quantil für linear transformierte Zufallsvariablen berechnen. Sie können den Erwartungswert und die Varianz für Summen von unabhängigen Zufallsvariablen berechnen.
- Sie kennen das Gesetz der grossen Zahlen und das \sqrt{n} -Gesetz. Sie kennen den Zentralen Grenzwertsatz und können ihn auf einfache Sachverhalte anwenden.
- Sie kennen den Unterschied zwischen Standardabweichung einer Einzelbeobachtung X_i und dem Standardfehler.

Übersicht Verteilungen

| Diskrete Verteilungen | | | | |
|---|--|---|-----------|-----------------|
| Verteilung | kum. Verteilungsfunktion | Wahrscheinlichkeitsfunktion | $E(X)$ | $\text{Var}(X)$ |
| Binomialverteilung $\text{Bin}(n, \pi)$ | $F(x) = \sum_{i=0}^x \binom{n}{i} \pi^i (1 - \pi)^{n-i}$ | $P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$ | $n\pi$ | $n\pi(1 - \pi)$ |
| Poisson-Verteilung $\text{Pois}(\lambda)$ | $F(x) = \sum_{i=0}^x e^{-\lambda} \frac{\lambda^i}{i!}$ | $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$ | λ | λ |

| Stetige Verteilungen | | | | |
|---|--|--|---------------------|-----------------------|
| Verteilung | kum. Verteilungsfunktion | Wahrscheinlichkeitsdichte | | |
| Uniforme Verteilung $\text{Unif}(a, b)$ | $F(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b \\ 1 & \text{falls } x > b \end{cases}$ | $f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |
| Exponentialverteilung $\text{Exp}(\lambda)$ | $F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$ | $f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ | $F(x) = \int_{-\infty}^x f(y) dy$ | $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | μ | σ^2 |

Kapitel 6.

Statistik für Messdaten

We must be careful not to confuse data with the abstractions we use to analyze them.

(William James)

6.1. Überprüfen der (Normal-) Verteilungsannahme

In Kapitel 5 sind wir davon ausgegangen, dass wir die Wahrscheinlichkeitsverteilung einer Zufallsvariablen, wie zum Beispiel $\mathcal{N}(3, 2)$, kennen. Aufgrund von dieser Wahrscheinlichkeitsverteilung haben wir dann Kennzahlen und diverse Wahrscheinlichkeiten berechnet.

In der Praxis allerdings müssen wir uns oft basierend auf (wenigen) Daten für eine Verteilungsfamilie, wie zum Beispiel die Normalverteilung, entscheiden, mit der wir ein Zufallsereignis modellieren wollen. Nehmen wir also an, dass wir einen Datensatz x_1, \dots, x_n mit n Beobachtungen haben. Die Wahl einer Verteilungsfamilie kann einerseits durch Erfahrung („was sich bisher bewährt hat“) oder aber auch durch physikalische Argumente geschehen. Ob eine Verteilungsfamilie zu einem konkreten Datensatz passt, kann man qualitativ gut mit graphischen Methoden überprüfen. So könnten wir beispielsweise schauen, wie gut das (normierte) Histogramm der Daten zur Dichte unserer Modellverteilung passt, z.B. zu einer bestimmten Normalverteilung. Es zeigt sich aber, dass man Abweichungen besser durch den Vergleich der Quantilen erkennen kann.

Q-Q-Plots sind graphische Darstellungen, um zu überprüfen, wie gut eine Verteilungsfamilie zu einem Datensatz passt. Nun gibt es auch quantitative Methoden beruhend auf Teststatistiken, um festzustellen, wie gut eine (vermutete) Verteilungsfamilie zu einem Datensatz passt. Wir verweisen zum Beispiel auf den Chi-Quadrat-Test, der im Kapitel B.3 des Anhangs besprochen wird.

6.1.1. Q-Q-Plot

Die Idee des *Q-Q-Plot* (Quantil-Quantil Plot) besteht darin, die empirischen Quantile gegen die theoretischen Quantile der vermuteten Modell-Verteilung zu plotten.

QQ-Plot

(i) Berechne für

$$\alpha_k = \frac{k - 0.5}{n} \quad \text{mit } k = 1, \dots, n$$

also für

$$\alpha_1 = \frac{0.5}{n}, \dots, \alpha_n = \frac{n - 0.5}{n}$$

die theoretischen Quantile der Modell-Verteilung

$$q(\alpha_k) = F^{-1}(\alpha_k)$$

wobei n die Anzahl Datenpunkte bezeichnet.

(ii) Bestimme die empirischen α_k -Quantile, welche den geordneten Beobachtungen

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

entsprechen. ($x_{(\alpha_k \cdot n + 0.5)} = x_{(k)}$)

(iii) Zeichne die theoretischen Quantile $q(\alpha_k)$ auf der x -Achse gegen die empirischen Quantile $x_{(k)}$ auf der y -Achse auf.

Wenn die Beobachtungen gemäss der Modell-Verteilung erzeugt wurden, sollten diese Punkte ungefähr auf der Winkelhalbierenden $y = x$ liegen.

Bemerkungen:

- i. Wir erinnern daran, dass wir mit $x_{(1)}$ die kleinste Beobachtung im Datensatz bezeichnen, mit $x_{(n)}$ die grösste.
- ii. Warum berechnen wir die theoretischen Quantile zu den Werten $\alpha_k = \frac{k-0.5}{n}$? Nun wir hatten das empirische α -Quantil als die Beobachtung $x_{(\alpha \cdot n + 0.5)}$ definiert, wobei wir den Wert $\alpha \cdot n + 0.5$ runden (α und n sind gegeben). Betrachten wir nun aber die Beobachtung $x_{(k)}$, also die k -grösste Beobachtung, dann entspricht k dem gerundeten Wert von $(\alpha \cdot n + 0.5)$. Also ist $x_{(k)}$ das α_k -Quantil mit $\alpha_k = \frac{k-0.5}{n}$.

Beispiel 6.1.1 Betondruckfestigkeit

Es wurden Messungen der Betondruckfestigkeit an $n = 20$ verschiedenen Proben durchgeführt. Wir wollen schauen, wie gut die Daten mit einer Normalverteilung beschrieben werden können.

Die Werte sind in der Tabelle 6.1 der Grösse nach aufgeführt. Für jeden Messwert wird $\alpha_k = \frac{k-0.5}{n}$ berechnet. Der erste Wert entspricht also dem empirischen 2.5-Prozent-Quantil, die $k = 11$ -te Messung entspricht in etwa dem Median und die $k = 16$ -te Messung entspricht dem 75 % Quantil.

| k | $x_{(k)}$ | $\alpha_k = (k - 0.5) / n$ | q_{α_k} für $\mathcal{N}(32.7, 4.15^2)$ | $\Phi^{-1}(\alpha_k)$ |
|-----|-----------|----------------------------|--|-----------------------|
| 1 | 24.4 | 0.0250 | 24.5 | -1.96 |
| 2 | 27.6 | 0.075 | 26.7 | -1.44 |
| 3 | 27.8 | 0.125 | 27.9 | -1.15 |
| 4 | 27.9 | 0.175 | 28.8 | -0.935 |
| 5 | 28.5 | 0.225 | 29.5 | -0.755 |
| 6 | 30.1 | 0.275 | 30.2 | -0.600 |
| 7 | 30.3 | 0.325 | 30.8 | -0.453 |
| 8 | 31.7 | 0.375 | 31.3 | -0.319 |
| 9 | 32.2 | 0.425 | 31.9 | -0.189 |
| 10 | 32.8 | 0.475 | 32.4 | -0.0627 |
| 11 | 33.3 | 0.525 | 32.9 | 0.0627 |
| 12 | 33.5 | 0.575 | 33.4 | 0.189 |
| 13 | 34.1 | 0.625 | 34.0 | 0.319 |
| 14 | 34.6 | 0.675 | 34.5 | 0.454 |
| 15 | 35.8 | 0.725 | 35.1 | 0.598 |
| 16 | 35.9 | 0.775 | 36.0 | 0.755 |
| 17 | 36.8 | 0.825 | 36.5 | 0.935 |
| 18 | 37.1 | 0.875 | 37.4 | 1.15 |
| 19 | 39.2 | 0.925 | 38.6 | 1.44 |
| 20 | 39.7 | 0.975 | 40.8 | 1.96 |

Tabelle 6.1.: 20 Messungen der Betondruckfestigkeit.

Nun vermuten wir, dass diese Messungen normalverteilt sind. Wir schätzen¹ also den Parameter μ der Normalverteilung durch den empirischen Mittelwert und den Parameter σ durch die empirische Standardabweichung und finden

$$\hat{\mu} = 32.7 \quad \text{und} \quad \hat{\sigma} = 4.15$$

Wir berechnen als nächstes für jedes α_k das entsprechende α_k -Quantil, also $\Phi^{-1}(\alpha_k)$, der Normalverteilung

$$\mathcal{N}(32.7, 4.15^2)$$

Sollte unser Datensatz tatsächlich normalverteilt sein, dann erwarten wir, dass die empirischen Quantilen unseres Datensatzes und die Quantilen der Normalverteilung in etwa gleich gross sind, also auf der Winkelhalbierenden $y = x$ liegen.

Mit R erstellt man einen Q-Q-Plot folgendermassen

```
x <- c(24.4, 27.6, 27.8, 27.9, 28.5, 30.1, 30.3, 31.7, 32.2,
      32.8, 33.3, 33.5, 34.1, 34.6, 35.8, 35.9, 36.8, 37.1,
      39.2, 39.7)

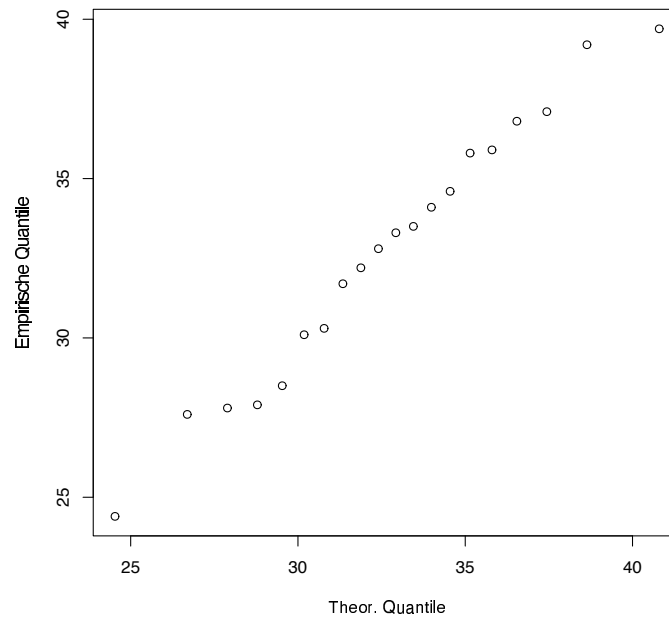
alphak <- (seq(1, length(x), by = 1) - 0.5) / length(x)

quantile.theor <- qnorm(alphak, mean = mean(x), sd = sd(x))

quantile.empir <- sort(x)

qqplot(quantile.theor, quantile.empir, xlab = "Theor. Quantile",
       ylab = "Empirische Quantile")
```

¹Dieses plausible Vorgehen für die Parameterschätzung werden wir im nächsten Kapitel 6.2 ausführlich begründen.



Die Idee des QQ-Plots ist nochmals in Abbildung 6.1 dargestellt. Es werden dabei verschiedene Quantile des Modells und des Datensatzes betrachtet. Stimmen diese mehr oder weniger überein, so können wir annehmen, dass unser Datensatz die vermutete Modellverteilung hat.

□

6.1.2. Normal-Plot

Meist will man nicht eine spezifische Verteilung, sondern eine ganze Klasse von Verteilungen prüfen. Zum Beispiel möchte man überprüfen, ob ein Datensatz einer Normalverteilung mit beliebigem μ und σ folgt. Ein Spezialfall von einem Q-Q-Plot ist der **Normal-Plot**: ein Q-Q Plot, bei dem die Modell-Verteilung gleich der Standardnormalverteilung $\mathcal{N}(0, 1)$ ist, heisst Normal-Plot.

Falls die Daten Realisierungen von $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ sind und wir die Zufallsvariable X auffassen als $X = \mu_X + \sigma_X \cdot Y$ mit $Y \sim \mathcal{N}(0, 1)$, so gilt für die Quantile von X :

$$\underbrace{q_X(\alpha)}_{y(x)} = \underbrace{\mu_X}_a + \underbrace{\sigma_X \cdot q_Y(\alpha)}_{b \cdot x},$$

wobei $q_Y(\alpha) = \Phi^{-1}(\alpha)$ die α -Quantile der Standardnormalverteilung (Modellverteilung) ist (siehe Regel (iii) der linearen Transformationen in Kapitel 5.3). Wenn man

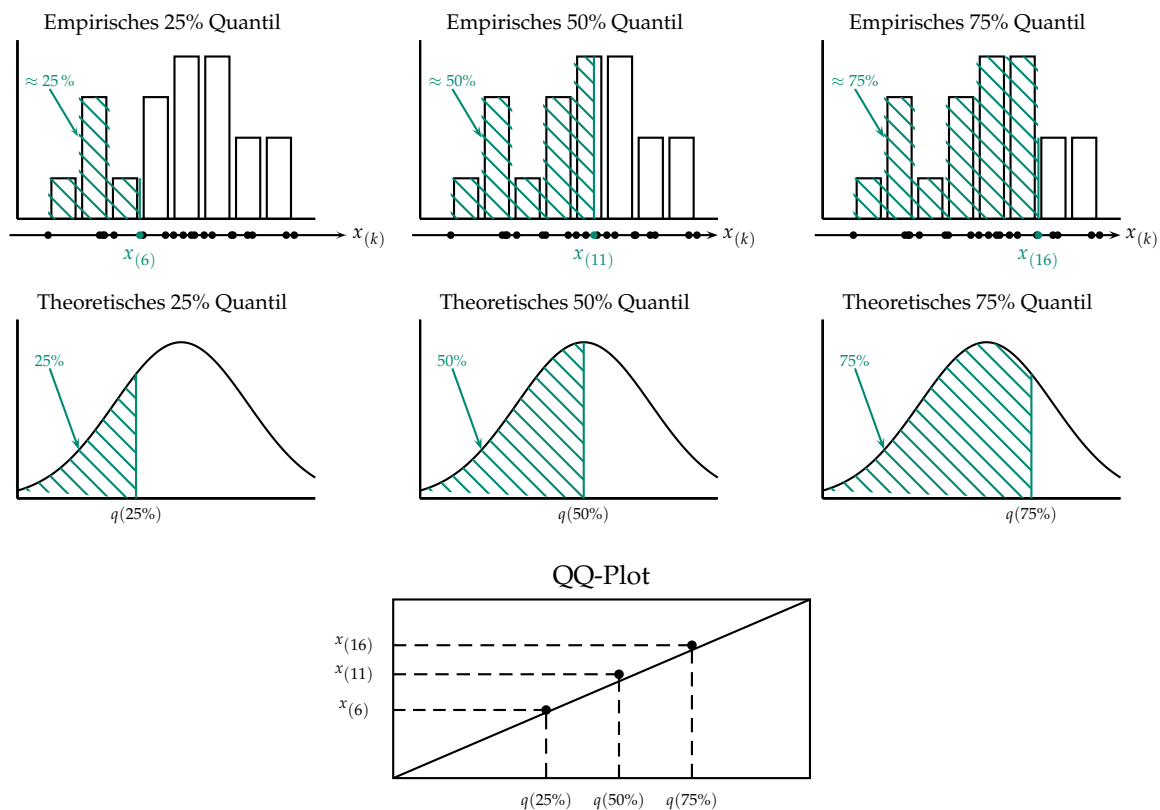


Abbildung 6.1.: Empirisches und theoretisches Quantil mit QQ-Plot vom Datensatz der Betondruckfestigkeit.

also einen Normal-Plot macht, so sollten die Punkte im Normal-Plot ungefähr auf der Geraden mit Achsenabschnitt μ und Steigung σ liegen.

Abbildung 6.2 zeigt zwei Normal-Plots: einmal ist die datengenerierende Verteilung eine Normalverteilung. Die Punkte liegen in etwa auf einer Geraden. Das andere mal sind die Daten von einer sehr langschwänzigen Verteilung erzeugt worden. Hier liegen die Punkte gar nicht auf einer Geraden, sondern an der rechten Extremität der Verteilung weit oberhalb der Geraden und auf der linken Extremität weit unterhalb der Geraden.

Natürlich liegen die Punkte in einem Normal-Plot nicht perfekt auf einer Geraden. Damit Sie ein Gefühl dafür bekommen, was eine akzeptable Abweichung von der Geraden ist, sind in Abbildung 6.3 die Normal-Plots von neun Datensätzen (je 50 Beobachtungen) gezeigt, die alle von einer Standardnormalverteilung simuliert wurden. Abweichungen in dem Mass, wie sie dort zu sehen sind, kann man also erwarten, wenn die Daten tatsächlich von einer Normalverteilung stammen.

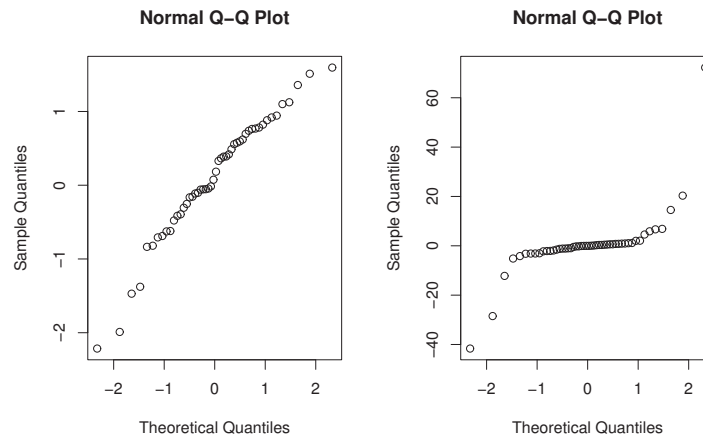


Abbildung 6.2.: Links: Normal-Plot für 50 Realisierungen von $\mathcal{N}(0,1)$. Die Punkte liegen auf einer Geraden. Rechts: Normal-Plot für 50 Realisierungen von einer Verteilung, die sehr langschwänzig ist. Die Punkte liegen nicht auf einer Geraden.

6.2. Parameterschätzung für stetige Wahrscheinlichkeitsverteilungen

6.2.1. Momentenmethode

Im letzten Kapitel haben wir gesehen, wie wir überprüfen können, ob eine Verteilungsfamilie zu einem Datensatz passt. Nun möchten wir die Parameter einer Verteilung schätzen. Dafür rufen wir uns die Momentenmethode aus Kapitel 4 in Erinnerung:

- Wir fassen unsere Daten x_1, x_2, \dots, x_n als Realisierungen von Zufallsvariablen X_1, X_2, \dots, X_n auf mit bekannter Verteilung.
- Wir berechnen den Erwartungswert $E(X)$ und lösen die Gleichung nach dem unbekannten Parameter auf, welchen wir schätzen wollen.
- Wir ersetzen den Erwartungswert durch dessen Gegenstück, den empirischen Mittelwert und erhalten so eine Schätzung für den unbekannten Parameter. Ein Momentenschätzer für die Standardabweichung ist die empirische Standardabweichung.

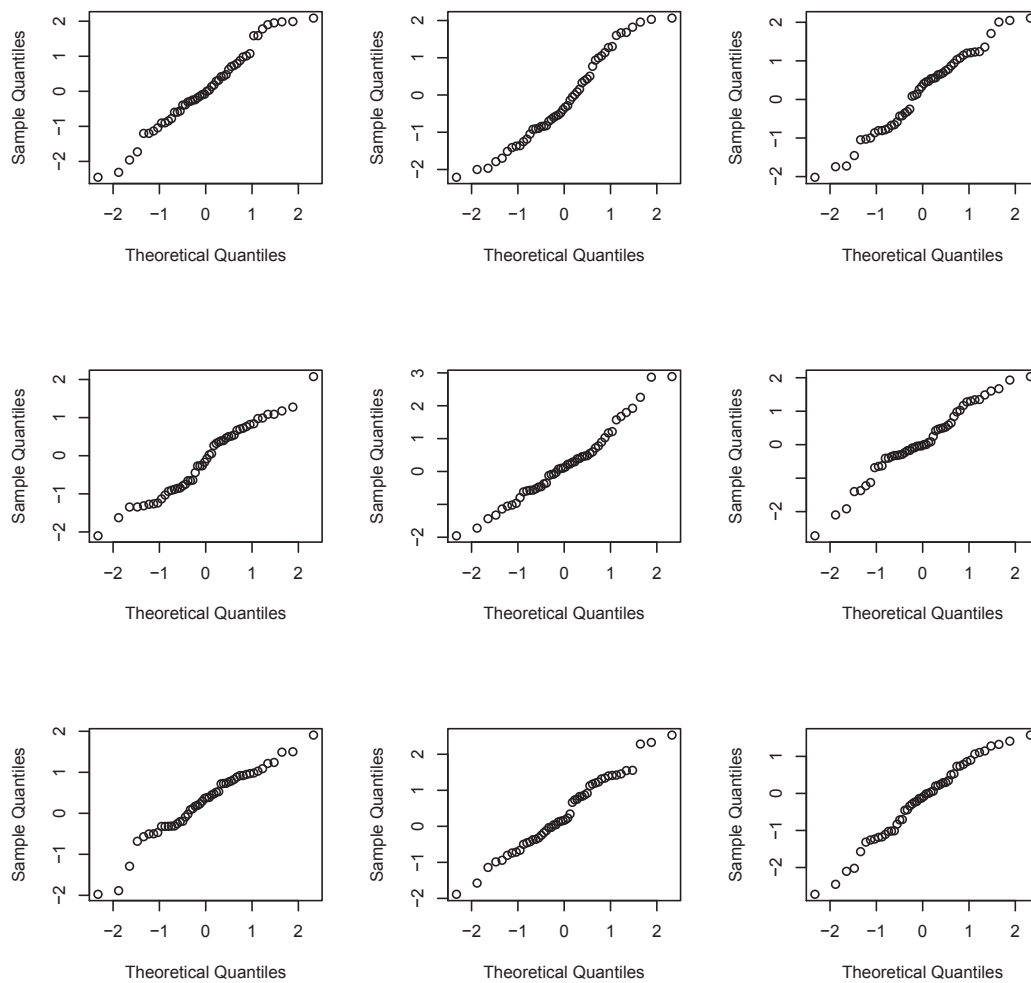


Abbildung 6.3.: Normal-Plots von neun Datensätzen (je 50 Beobachtungen), die alle von einer Standard-Normalverteilung simuliert wurden. Abweichungen von einer Geraden in diesem Ausmass sind also zu erwarten, wenn die Daten wirklich normal verteilt sind. Falls die Abweichung von einer Geraden deutlich grösser ist, sind die Daten wohl nicht normalverteilt.

Beispiel 6.2.1 Betondruckfestigkeit

Bei der Messung der Betondruckfestigkeit hatten wir die Vermutung, dass die Daten normalverteilt sind. Mit einem Q-Q-Plot konnten wir das auch bestätigen. Somit folgt die Zufallsvariable X für die Betondruckfestigkeit einer Normalverteilung, also $X \sim \mathcal{N}(\mu, \sigma^2)$ mit den Parametern μ und σ . Wir möchten nun mit der Momentenmethode

einen Schätzer für μ finden. In diesem Fall gilt, dass $E(X) = \mu$ und wir müssen den Erwartungswert bloss noch durch den empirischen Mittelwert ersetzen:

$$\mu = E(X) \Rightarrow \hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{653.3}{20} = 32.7$$

Für die Schätzung der Standardabweichung σ mit Hilfe der Momentenmethode benötigen wir folgende Beziehungen:

$$\begin{aligned}\sigma^2 &= E(X^2) - E(X)^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

Schätzen wir $E(X)$ durch \bar{X}_n und $E(X^2)$ durch $\frac{1}{n} \sum_{i=1}^n X_i^2$, so ergibt sich folgendes Gleichungssystem

$$\begin{aligned}\hat{\mu} &= \bar{X}_n \\ \hat{\mu}^2 + \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2\end{aligned}$$

dessen Lösung gegeben ist durch

$$\begin{aligned}\hat{\mu} &= \bar{X}_n \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^n X_i^2 - n \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right)^2}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}\end{aligned}$$

Somit ergibt sich für unseren Datensatz folgende Schätzung für die Standardabweichung

$$\begin{aligned}\hat{\sigma} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2} \\ &= \sqrt{\frac{1}{20} \sum_{i=1}^{20} (x_i - 32.7)^2} \\ &= 16.4\end{aligned}$$

□

Der Momentenmethode-Schätzer ist nicht erwartungstreu, denn der Faktor $\frac{1}{n}$ müsste durch $\frac{1}{n-1}$ ersetzt werden.

Beispiel 6.2.2 Exponentialverteilung

Wir messen die Lebenserwartung eines elektronischen Moduls, welches in der Motorsteuerung eines Autos verwendet wird. Der Test ergab die folgenden Messungen (in Stunden):

$$x_1 = 11.96, x_2 = 5.03, x_3 = 67.40, x_4 = 16.07, x_5 = 31.50, x_6 = 7.73, x_7 = 11.10, x_8 = 22.38$$

Die Lebenserwartung ist in der Regel exponentialverteilt mit Parameter λ . Der Momentenschätzer von λ lautet:

$$E(X) = \frac{1}{\lambda} \Rightarrow \hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{21.6} = 0.0462$$

□

6.2.2. Maximum-Likelihood-Methode

Wir haben n Beobachtungen

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$$

Wir nehmen an, dass diese Beobachtungen i.i.d. sind. Im Fall von einer diskreten Wahrscheinlichkeitsverteilung lässt sich die (Punkt-) Wahrscheinlichkeit, dass sich diese n Beobachtungen (Ereignisse) ereignet haben, schreiben als

$$\begin{aligned} P[(X_1 = x_1) \cap (X_2 = x_2) \cap \dots \cap (X_n = x_n)] &= P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \dots \cdot P[X_n = x_n] \\ &\equiv \prod_{i=1}^n P[X_i = x_i] \end{aligned}$$

Die Wahrscheinlichkeit, dass die n unabhängigen Zufallsvariablen x_1, x_2, \dots, x_n beobachtet werden, hängt vom Parameter θ ab, welchen wir schätzen möchten. Die dazugehörige *Likelihood-Funktion* $L(\theta)$ lautet also:

$$\begin{aligned} L(\theta) &= P[X_1 = x_1 | \theta] \cdot P[X_2 = x_2 | \theta] \cdot \dots \cdot P[X_n = x_n | \theta] \\ &= \prod_{i=1}^n P[X_i = x_i | \theta] \end{aligned}$$

wobei $P[X_i = x_i|\theta]$ die Punktwahrscheinlichkeit bezeichnet, dass der Wert x_i beobachtet wurde, gegeben der Parameterwert θ . Die Idee der Maximum-Likelihood-Methode ist nun, den Parameter θ so zu schätzen, dass die beobachteten Ereignisse am wahrscheinlichsten sind. Wir suchen also das Maximum der Maximum-Likelihood-Funktion $L(\theta)$.

Im folgenden betrachten wir nun kontinuierliche Wahrscheinlichkeitsverteilungen mit der Wahrscheinlichkeitsdichtefunktion $f(x;\theta)$. Unter der Annahme von $f(x;\theta)$, einschliesslich eines Wertes von θ und der Unabhängigkeit unserer Beobachtungen, ist die Wahrscheinlichkeit, dass die erste Beobachtung x_1 im Intervall $[x_1, x_1 + dx_1]$ liegt, $f(x_1;\theta) dx_1$, die Wahrscheinlichkeit, dass die zweite Beobachtung x_2 im Intervall $[x_2, x_2 + dx_2]$ liegt, ist $f(x_2;\theta) dx_2$, etc.. Somit ist die Wahrscheinlichkeit, dass jede Beobachtung x_i im Intervall $[x_i, x_i + dx_i]$ liegt, gegeben durch

$$\prod_{i=1}^n f(x_i;\theta) dx_i$$

Falls die vermutete Wahrscheinlichkeitsdichtefunktion $f(x_i;\theta)$ und der Parameterwert von θ korrekt sind, erwartet man eine hohe Wahrscheinlichkeit für die Daten, die tatsächlich beobachtet wurden. Falls der Parameterwert θ weit weg vom wahren Parameterwert liegt, wird die obige Funktion einen kleinen Wert ergeben. Da die (infinitesimalen) Intervallbreiten dx_i nicht vom Parameterwert θ abhängen, gelten diese Überlegungen ebenfalls für die folgende Funktion

$$L(\theta) = \prod_{i=1}^n f(x_i;\theta)$$

die die **Likelihood-Funktion** für kontinuierliche Wahrscheinlichkeitsverteilungen darstellt. Um den Parameterwert von θ zu schätzen, werden wir wiederum $L(\theta)$ maximieren.

Beispiel 6.2.3 Normalverteilung; σ^2 bekannt

Sei X normalverteilt mit unbekanntem μ und bekanntem σ^2 . Die Likelihood-Funktion für den Datensatz x_1, x_2, \dots, x_n ist gegeben durch:

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Es ist einfacher, die Log-Likelihood-Funktion zu maximieren:

$$\ln(L(\mu)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Die Ableitung ergibt:

$$\frac{d \ln(L(\mu))}{d\mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)$$

Setzen wir diese Gleichung gleich 0 und lösen nach μ auf, ergibt dies:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

Somit ist der empirische Mittelwert der Likelihood-Schätzer für μ gegeben durch den empirischen Mittelwert. Dieser Schätzer ist identisch mit dem Momentenschätzer.

□

Beispiel 6.2.4 Normalverteilt; σ^2 unbekannt

Sei X normalverteilt mit Erwartungswert μ und Varianz σ^2 , wobei der Erwartungswert und die Varianz unbekannt sind. Die Likelihood-Funktion für den Datensatz x_1, x_2, \dots, x_n ist gegeben durch:

$$L(\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

Die Log-Likelihood-Funktion ist dann:

$$\ln(L(\mu, \sigma^2)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Wir suchen nun die Schätzer für μ und σ^2 , welche die Log-Likelihood-Funktion maximieren. Daher leiten wir die Log-Likelihood-Funktion partiell nach beiden Variablen ab und setzen die Ableitungen gleich 0:

$$\begin{aligned} \frac{\partial \ln(L(\mu, \sigma^2))}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \stackrel{!}{=} 0 \\ \frac{\partial \ln(L(\mu, \sigma^2))}{\partial (\sigma^2)} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \stackrel{!}{=} 0 \end{aligned}$$

Die Lösung dieses Gleichungssystems sind unsere Maximum-Likelihood-Schätzer

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Die Maximum-Likelihood-Schätzung von σ^2 ist identisch mit der Schätzung der Momentenmethode. Die Maximum-Likelihood-Schätzung von σ^2 ist folglich nicht erwartungstreu.

□

Beispiel 6.2.5 Exponentialverteilung

Sei X exponentialverteilt mit Parameter λ . Die Likelihood-Funktion für den Datensatz x_1, x_2, \dots, x_n ist gegeben durch:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

Die Log-Likelihood-Funktion ist

$$\ln(L(\lambda)) = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i$$

Wir leiten die Log-Likelihood-Funktion nach λ ab und setzen die Ableitung gleich 0

$$\frac{d \ln(L(\lambda))}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i \stackrel{!}{=} 0$$

Der Maximum-Likelihood-Schätzer $\hat{\lambda}$ ist die Lösung der oberen Gleichung:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$$

□

6.3. Statistische Tests und Vertrauensintervall für eine Stichprobe bei normalverteilten Daten

Wir betrachten Daten x_1, \dots, x_n , welche als Realisierungen von

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2)$$

aufgefasst werden. Als Beispiel kann man sich 10 Messungen einer Schadstoffkonzentration vorstellen. Zwei Kennzahlen der Zufallsvariablen X_i sind:

$$E(X_i) = \mu \quad \text{und} \quad \text{Var}(X_i) = \sigma_X^2$$

Typischerweise sind diese (und andere) Kennzahlen unbekannt und man möchte Rückschlüsse aus den Daten darüber machen. Die (Punkt-) Schätzungen für den Erwartungswert und die Varianz sind:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Beachten Sie, dass die Schätzer hier als Funktionen der Zufallsvariablen X_1, \dots, X_n geschrieben sind: insbesondere sind $\hat{\mu}$ und $\hat{\sigma}_X^2$ selbst wieder Zufallsvariablen - die Verteilungseigenschaften von $\hat{\mu}$ wurden im Kapitel 5.4 diskutiert.

Wann immer wir eine Messreihe neu erheben, werden wir für die betreffende Messreihe unterschiedliche Werte (Realisierungen) von $\hat{\mu}$ und $\hat{\sigma}_X^2$ ermitteln. Es ist also vernünftig, für den Schätzer $\hat{\mu}$ ein Intervall anzugeben, in welches der wahre Wert von μ mit einer bestimmten Wahrscheinlichkeit fällt. Es handelt sich also darum, für einen geschätzten Parameter ein *Vertrauensintervall* anzugeben. Weiter kann man sich fragen, ob eine Realisierung von μ kompatibel ist mit einem vermuteten μ_0 . Die entsprechende Fragestellung wird durch einen *statistischen Test* beantwortet.

6.3.1. Der z-Test (σ_X bekannt)

Wir nehmen an, dass die Daten x_1, \dots, x_n Realisierungen von

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2)$$

sind. Überdies machen wir die Annahme, dass σ_X^2 bekannt ist. Der z-Test für den Parameter μ erfolgt dann wie folgt.

1. **Modell:** X_i ist eine kontinuierliche Messgröße;

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2), \quad \sigma_X \text{ bekannt}$$

2. **Nullhypothese:**

$$H_0: \mu = \mu_0$$

Alternative:

$$H_A: \mu \neq \mu_0 \quad (\text{oder „<“ oder „>“})$$

3. Teststatistik:

$$Z = \frac{(\bar{X}_n - \mu_0)}{\sigma_{\bar{X}_n}} = \frac{(\bar{X}_n - \mu_0)}{\sigma_X / \sqrt{n}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{Standardfehler}}$$

Verteilung der Teststatistik unter H_0 :

$$Z \sim \mathcal{N}(0, 1)$$

4. Signifikanzniveau:

$$\alpha$$

5. Verwerfungsbereich für die Teststatistik:

$$K = (-\infty, z_{\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, \infty) \text{ bei } H_A : \mu \neq \mu_0,$$

$$K = (-\infty, z_{\alpha}] \text{ bei } H_A : \mu < \mu_0,$$

$$K = [z_{1-\alpha}, \infty) \text{ bei } H_A : \mu > \mu_0,$$

wobei

$$z_{\alpha/2} = \Phi^{-1}(\alpha/2)$$

6. Testentscheid:

Überprüfe, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich liegt.

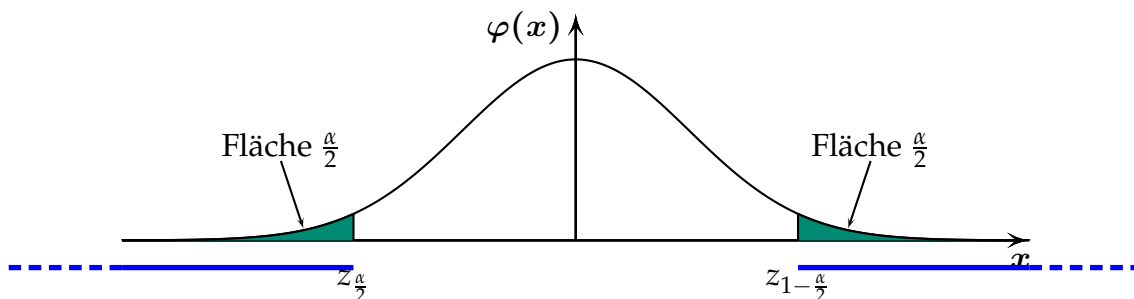


Abbildung 6.4.: Dichtefunktion der Teststatistik Z mit Verwerfungsbereich (blau) des zweiseitigen Z -Tests zum Niveau α . Beachte $z_{\alpha/2} = -z_{1-\alpha/2}$, wobei $z_{\alpha/2} = \Phi^{-1}(\alpha/2)$ ist.

Im Unterschied zu den Tests im Unterkapitel 4.3 basiert der z -Test auf *mehreren* Beobachtungen. Denn die realisierte Teststatistik z fasst die Beobachtungen in Form des

arithmetischen Mittelwertes zusammen. Ansonsten folgen die Konzepte der Vorgehensweise wie in Kapitel 4.3 behandelt.

Beispiel 6.3.1

Bei Methode A (vgl. Abschnitt 2.2.1) scheint die Schmelzwärme grösser als 80.00 zu sein. Angenommen, wir wissen aus vorhergehenden Studien, dass die Standardabweichung unseres Messinstruments $\sigma_X = 0.01$ ist. Ist es plausibel, dass die Schmelzwärme genau 80.00 cal/g ist? Wir führen dazu einen z-Test durch:

1. **Modell:** X_i ist eine kontinuierliche Messgrösse;

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2), \quad \sigma_X = 0.01 \text{ bekannt}, \quad n = 13$$

2. **Nullhypothese:**

$$H_0 : \mu = \mu_0 = 80.00$$

Alternative:

$$H_A : \mu \neq \mu_0$$

3. **Teststatistik:**

$$Z = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_X}$$

Verteilung der Teststatistik unter H_0 :

$$Z \sim \mathcal{N}(0, 1)$$

4. **Signifikanzniveau:**

$$\alpha = 0.05$$

5. **Verwerfungsbereich für die Teststatistik:**

$$K = (-\infty, z_{\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, \infty) \quad \text{bei} \quad H_A : \mu \neq \mu_0,$$

Mit dem Computer erhalten wir mit $\alpha = 0.05$:

$$z_{\frac{\alpha}{2}} = \Phi^{-1}\left(\frac{\alpha}{2}\right) = \Phi^{-1}(0.025) = -1.96$$

Damit ergibt sich für den Verwerfungsbereich der Teststatistik

$$K = (-\infty, -1.96] \cup [1.96, \infty)$$

6. Testentscheid:

Aus den $n = 13$ Daten errechnen wir

$$\bar{X}_n = 80.02$$

Damit ergibt sich als Wert für die Teststatistik

$$z = \frac{\sqrt{13}(80.02 - 80.00)}{0.01} = 7.211$$

Der beobachtete Wert liegt im Verwerfungsbereich der Teststatistik. Daher wird die Nullhypothese auf dem 5 % Signifikanzniveau verworfen.

Bemerkungen:

- i. Hier ist der Umweg über die Standardisierung wegen der technischen Hilfsmittel (Computersoftware) an sich unnötig.

- 1. **Modell:** X_i ist eine kontinuierliche Messgröße;

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2), \quad \sigma_X = 0.01 \text{ bekannt, } n = 13$$

- 2. **Nullhypothese:**

$$H_0: \mu = \mu_0 = 80.00$$

Alternative:

$$H_A: \mu \neq \mu_0$$

- 3. **Teststatistik:**

Der Mittelwert der Messungen

$$T: \bar{X}_n$$

Verteilung der Teststatistik unter H_0 :

$$T \sim \mathcal{N}\left(\mu_0, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(80, \frac{0.01^2}{13}\right)$$

- 4. **Signifikanzniveau:**

$$\alpha = 0.05$$

- 5. **Verwerfungsbereich für die Teststatistik:**

$$K = (-\infty, c_u] \cup [c_o, \infty) \quad \text{bei} \quad H_A: \mu \neq \mu_0,$$

Mit R erhalten wir mit $\alpha = 0.05$:

```
qnorm(0.025, 80, 0.01/sqrt(13))
## [1] 79.99456
qnorm(0.975, 80, 0.01/sqrt(13))
## [1] 80.00544
```

Damit erhalten wir den Verwerfungsbereich der Teststatistik:

$$K = (-\infty, 80.00] \cup [80.01, \infty)$$

6. Testentscheid:

Aus den $n = 13$ Daten errechnen wir

$$\bar{X}_n = 80.02$$

Der beobachtete Wert liegt im Verwerfungsbereich der Teststatistik. Daher wird die Nullhypothese auf dem 5 % Signifikanzniveau verworfen.

□

Die Begriffe *Fehler 1. Art*, *Fehler 2. Art* und *Macht*, die wir bereits im Kapitel 4.3 beim Binomialtest kennengelernt haben, werden auch für alle anderen Tests verwendet. Wir wiederholen hier nochmals die Definitionen:

Fehler 1. und 2. Art und Macht

Fehler 1. Art = fälschliches Verwerfen von H_0 , obschon H_0 richtig ist

Fehler 2. Art(μ) = fälschliches Beibehalten von H_0 falls $\mu \in H_A$ richtig ist

Macht(μ) = $1 - P(\text{Fehler 2. Art}(\mu)) = P(\text{Verwerfen von } H_0 \text{ falls } \mu \in H_A \text{ stimmt})$

Die Wahrscheinlichkeit für einen Fehler 1. Art ist gerade gleich α . Für $\mu \in H_A$ kann man die Macht(μ) als die Wahrscheinlichkeit interpretieren, dass man richtigerweise H_A entdeckt, falls $\mu \in H_A$ stimmt. Für eine Teststatistik T und einen dazugehörigen Verwerfungsbereich K gilt dann:

$$P_{\mu_0}(T \in K) = \alpha$$

$$P_{\mu}(T \in K) = \text{Macht}(\mu)$$

6.3.2. Der t -Test (σ_X unbekannt)

Wie vorhin nehmen wir an, dass die Daten Realisierungen von

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2)$$

sind. In der Praxis ist die Annahme, dass σ_X bekannt ist, oftmals unrealistisch. In solchen Fällen kann die Teststatistik z nicht berechnet werden, weil sie auf σ_X basiert.

Allerdings können wir stattdessen die Schätzung

$$\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

benutzen. Dies führt aber zu einer zusätzlichen Unsicherheit, was zur Folge hat, dass sich die Verteilung der Teststatistik ändert.

t -Verteilung

Die Teststatistik beim t -Test ist gegeben durch

$$T = \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_X / \sqrt{n}} = \frac{\text{beobachtet} - \text{erwartet}}{\text{geschätzter Standardfehler}}$$

Deren Verteilung unter der Nullhypothese $H_0 : \mu = \mu_0$ ist eine sogenannte **t -Verteilung** mit $n - 1$ Freiheitsgraden, die wir mit t_{n-1} bezeichnen.

Die t_n -Verteilung² ist eine symmetrische Verteilung um 0, welche langschwänziger ist als die Standardnormalverteilung $\mathcal{N}(0, 1)$ (siehe Abbildung 6.5). Für

$$T \sim t_n$$

gilt:

$$\begin{aligned} E(T) &= 0 \\ \text{Var}(T) &= \frac{n}{n-2} \end{aligned}$$

Für grosse n ist t_n ähnlich zu $\mathcal{N}(0, 1)$. Insbesondere strebt die t_n -Verteilung für $n \rightarrow \infty$ gegen die Standardnormalverteilung $\mathcal{N}(0, 1)$.

Beispiel 6.3.2

Der folgende Datensatz besteht aus normalverteilten Datenpunkten x_1, \dots, x_{20}

²Eine genaue Herleitung der t -Verteilung wird im Kapitel B.2 des Anhangs gegeben.

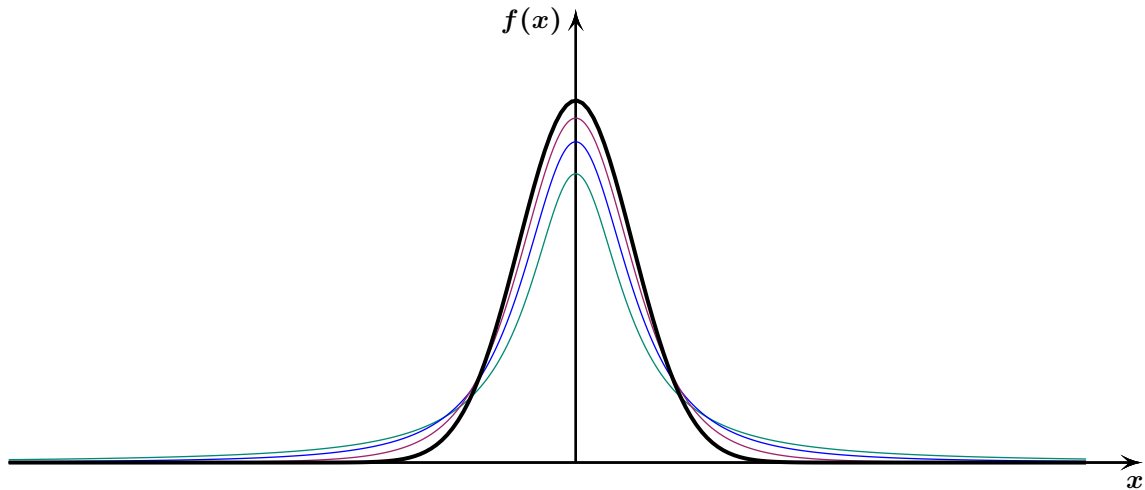


Abbildung 6.5.: Dichten der t -Verteilung mit 1 (grün), 2 (blau) und 5 (violett) Freiheitsgraden. Die schwarze Kurve ist die Dichte der Standardnormalverteilung.

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5.9 | 3.4 | 6.6 | 6.3 | 4.2 | 2.0 | 6.0 | 4.8 | 4.2 | 2.1 |
| 8.7 | 4.4 | 5.1 | 2.7 | 8.5 | 5.8 | 4.9 | 5.3 | 5.5 | 7.9 |

Wir vermuten, dass unsere Daten x_1, x_2, \dots, x_{20} Realisierungen von

$$X_i \sim \mathcal{N}(5, \sigma_X)$$

sind, wobei wir σ_X nicht kennen. Wir müssen σ_X also aus den Daten schätzen. Die Nullhypothese lautet in diesem Fall $\mu_0 = 5$. Wie gross ist die Wahrscheinlichkeit, dass der Mittelwert eines Datensatzes mit Verteilung

$$X_i \sim \mathcal{N}(5, \hat{\sigma}_X)$$

kleiner ist als der Mittelwert unseres Datensatzes, wobei $\hat{\sigma}_X$ die aus unseren Daten geschätzte Standardabweichung ist? Wir möchten also die Wahrscheinlichkeit

$$P(\bar{X}_{20} \leq \bar{x}_{20})$$

bestimmen, wobei

$$\bar{X}_{20} \sim \mathcal{N}(5, \hat{\sigma}_X / \sqrt{20})$$

ist. \bar{x}_{20} bezeichnet den arithmetischen Mittelwert unseres Datensatzes und $\hat{\sigma}_X$ ist die geschätzte empirische Standardabweichung des Datensatzes.

Um diese Wahrscheinlichkeit zu berechnen, standardisieren wir den Mittelwert, indem wir von \bar{X}_n den erwarteten Mittelwert $\mu_0 = 5$ subtrahieren, und dann durch den geschätzten Standardfehler $\hat{\sigma}_X/\sqrt{n}$ dividieren. Wir berechnen die Wahrscheinlichkeit

$$P(\bar{X}_{20} \leq \bar{x}_{20})$$

folglich mittels

$$\begin{aligned} P\left(\frac{\bar{X}_n - 5}{\hat{\sigma}_X/\sqrt{n}} \leq \frac{\bar{x}_{20} - 5}{\hat{\sigma}_X/\sqrt{n}}\right) &= P\left(T \leq \frac{5.2 - 5}{1.9/\sqrt{20}}\right) \\ &= P(T \leq 0.51) \\ &= 0.69 \end{aligned}$$

Dabei haben wir benützt, dass T einer t_{20-1} -Verteilung folgt.

```
x <- c(5.9, 3.4, 6.6, 6.3, 4.2, 2, 6, 4.8, 4.2, 2.1, 8.7,
      4.4, 5.1, 2.7, 8.5, 5.8, 4.9, 5.3, 5.5, 7.9)

mean.x <- mean(x)
mean.x

## [1] 5.215

sigma.x <- sd(x)
sigma.x

## [1] 1.883802

t <- (mean.x - 5)/(sigma.x/sqrt(20))
t

## [1] 0.5104088

pt(t, df = 19)

## [1] 0.6921781
```

□

Den Verwerfungsbereich beim Test erhalten wir, indem wir einen Bereich der Wahrscheinlichkeit α bei der t_{n-1} -Verteilung abschneiden (je nach Alternative auf einer Seite, oder je die Hälfte auf beiden Seiten). Dazu brauchen wir die Quantile $t_{n,\alpha}$, welche mittels Computer berechnet werden können.

Zusammenfassend erfolgt der t -Test gemäss den folgenden sechs Schritten:

1. **Modell:** X_i ist eine kontinuierliche Messgrösse;

$$X_1, \dots, X_n \text{ i.i.d. } \mathcal{N}(\mu, \sigma_X^2), \quad \sigma_X \text{ wird durch } \hat{\sigma}_X \text{ geschätzt}$$

2. **Nullhypothese:**

$$H_0 : \mu = \mu_0$$

Alternative:

$$H_A : \mu \neq \mu_0 \quad (\text{oder „<“ oder „>“})$$

3. **Teststatistik:**

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} = \frac{\text{beobachtet} - \text{erwartet}}{\text{geschätzter Standardfehler}}$$

Verteilung der Teststatistik unter H_0 :

$$T \sim t_{n-1}$$

4. **Signifikanzniveau:**

$$\alpha$$

5. **Verwerfungsbereich für die Teststatistik:**

$$K = (-\infty, t_{n-1; \frac{\alpha}{2}}] \cup [t_{n-1; 1-\frac{\alpha}{2}}, \infty) \quad \text{bei } H_A : \mu \neq \mu_0,$$

$$K = (-\infty, t_{n-1; \alpha}] \quad \text{bei } H_A : \mu < \mu_0,$$

$$K = [t_{n-1; 1-\alpha}, \infty) \quad \text{bei } H_A : \mu > \mu_0.$$

6. **Testentscheid:**

Überprüfe, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich liegt.

Da das Quantil der t -Verteilung grösser ist als das Quantil der Normalverteilung, erhält man einen etwas kleineren Verwerfungsbereich als beim z -Test. Für grosse n ist der Unterschied allerdings minim (da $t_{n-1} \approx \mathcal{N}(0, 1)$ falls n gross).

Beispiel 6.3.3

Wir berechnen nun nochmals das Beispiel 6.3.1. Diesmal schätzen wir allerdings die Standardabweichung σ_X aus den Daten. Wir führen also einen t -Test auf dem 5% Signifikanzniveau durch:

1. **Modell:** X_i ist eine kontinuierliche Messgröße;

$$X_1, \dots, X_n \text{ i.i.d. } \mathcal{N}(\mu, \sigma_X^2), \quad \sigma_X \text{ wird geschätzt, } \widehat{\sigma}_X = 0.024$$

2. **Nullhypothese:**

$$H_0: \mu = \mu_0 = 80.00$$

Alternative:

$$H_A: \mu \neq \mu_0$$

3. **Teststatistik:**

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\widehat{\sigma}_X}$$

Verteilung der Teststatistik unter H_0 : $T \sim t_{n-1}$

4. **Signifikanzniveau:**

$$\alpha = 0.05$$

5. **Verwerfungsbereich für die Teststatistik:**

$$K = (-\infty, t_{n-1; \frac{\alpha}{2}}] \cup [t_{n-1; 1-\frac{\alpha}{2}}, \infty) \quad \text{bei } H_A: \mu \neq \mu_0,$$

Den Wert

$$t_{n-1; 1-\frac{\alpha}{2}} = t_{12; 0.975} = 2.179$$

ermitteln wir mit Hilfe von R, wobei $\alpha = 0.05$ und $n = 13$, wie folgt

```
qt(0.975, 12)
## [1] 2.178813
```

Der Verwerfungsbereich der Teststatistik ist also:

$$K = (-\infty, -2.179] \cup [2.179, \infty)$$

6. **Testentscheid:**

Aus den $n = 13$ Daten haben wir

$$\bar{x} = 80.02 \quad \text{und} \quad \widehat{\sigma}_X = 0.024$$

errechnet. Damit ergibt sich für die Teststatistik der Wert

$$t = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\widehat{\sigma}_X} = \frac{\sqrt{13}(80.02 - 80.00)}{0.024} = 3.00$$

Der beobachtete Wert der Teststatistik liegt im Verwerfungsbereich. Daher wird die Nullhypothese auf dem 5 % Niveau verworfen.



Wir können obiges Beispiel mit R direkt berechnen, und zwar mit Hilfe des Befehls

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04,
       79.97, 80.05, 80.03, 80.02, 80, 80.02)

t.test(x, alternative = "two.sided", mu = 80, conf.level = 0.95)

##
## One Sample t-test
##
## data: x
## t = 3.1246, df = 12, p-value = 0.008779
## alternative hypothesis: true mean is not equal to 80
## 95 percent confidence interval:
## 80.00629 80.03525
## sample estimates:
## mean of x
## 80.02077
```

Bemerkungen:

- i. Der beobachtete Wert der Teststatistik ist 3.12 und folgt unter der Nullhypothese einer t -Verteilung mit $df = 12$ Freiheitsgraden.
- ii. Der P -Wert mit einer zweiseitigen Alternative ist 0.008 779 und ist somit auf dem 5 % Signifikanzniveau signifikant.
- iii. Der P -Wert bei 2-seitiger Alternative $H_A: \mu \neq \mu_0$ kann wie folgt berechnet werden (der beobachtete Wert der Teststatistik ist $t = \frac{\sqrt{n}|\bar{x}_n - \mu_0|}{\hat{\sigma}_X}$):

$$\begin{aligned} P\text{-Wert} &= P(|T| > |t|) \\ &= P(T < -|t|) + P(T > |t|) \\ &= 2 \cdot P(T > |t|) \end{aligned}$$

- iv. Der beobachtete Mittelwert der Daten ist 80.02. Ein 95 %-Vertrauensintervall für den wahren Mittelwert der Messungen ist [80.006, 80.035].

6.3.3. Vertrauensintervall für μ

Analog zum Vertrauensintervall bei Zählraten im Kapitel 4.4 besteht das Vertrauensintervall bei Messdaten aus denjenigen Werten μ , bei denen der entsprechende Test nicht verwirft. Das sind also alle Parameterwerte des Zufallsmodells, bei denen die Daten recht wahrscheinlich oder plausibel sind.

Wir nehmen wiederum an, dass die Daten Realisierungen von

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2)$$

sind. Bei einem zweiseitigen t -Test hat der Verwerfungsbereich die Form

$$K = (-\infty, -t_{n-1;1-\frac{\alpha}{2}}] \cup [t_{n-1;1-\frac{\alpha}{2}}, \infty)$$

Der t -Test verwirft H_0 nicht, wenn der Wert der Teststatistik nicht im Verwerfungsbereich der Teststatistik ist. Wenn H_0 nicht verworfen wird, muss also gelten:

$$-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\hat{\sigma}_X} \quad \text{und} \quad t_{n-1;1-\frac{\alpha}{2}} \geq \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\hat{\sigma}_X}$$

Um das zweiseitige Vertrauensintervall von μ zu finden, müssen wir alle Werte von μ_0 finden, die obige Gleichungen erfüllen. Am einfachsten geht das, wenn wir beide Gleichungen nach μ_0 auflösen:

$$\mu_0 \leq \bar{x}_n + \frac{\hat{\sigma}_X \cdot t_{n-1;1-\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{und} \quad \mu_0 \geq \bar{x}_n - \frac{\hat{\sigma}_X \cdot t_{n-1;1-\frac{\alpha}{2}}}{\sqrt{n}}$$

Zweiseitiges Vertrauensintervall

Dies führt dann auf die folgenden **zweiseitigen Vertrauensintervalle** (die dazugehörigen Tests sind zweiseitig mit Alternative $H_A: \mu \neq \mu_0$) zum Niveau $1 - \alpha$:

$$\left[\bar{x}_n - t_{n-1,1-\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}}, \bar{x}_n + t_{n-1,1-\alpha/2} \frac{\hat{\sigma}_X}{\sqrt{n}} \right]$$

Analog kann man auch **einseitige Vertrauensintervalle** konstruieren. Sie enthalten alle Parameter, bei denen ein einseitiger Test nicht verwerfen würde. Beim t -Test sehen die einseitigen $(1 - \alpha)$ -Vertrauensintervalle so aus:

$$\text{Falls } H_A: \mu < \mu_0 : (-\infty; \bar{x}_n + t_{n-1,1-\alpha} \cdot \frac{\hat{\sigma}_X}{\sqrt{n}}]$$

$$\text{Falls } H_A: \mu > \mu_0 : [\bar{x}_n - t_{n-1,1-\alpha} \cdot \frac{\hat{\sigma}_X}{\sqrt{n}}; \infty)$$

Beispiel 6.3.4 Methode A zur Bestimmung der Schmelzwärme

Wir haben

$$n - 1 = 13 - 1 = 12$$

Freiheitsgrade und

$$t_{12,0.975} = 2.18$$

Die mittlere mit Methode A gemessene Schmelzwärme ist

$$\bar{x}_n = 80.02$$

und die Standardabweichung lautet

$$\hat{\sigma}_X = 0.024$$

Dann ist das zweiseitige Konfidenzintervall für die mit Methode A gemessene Schmelzwärme also gegeben durch

$$I = 80.02 \pm 2.18 \cdot 0.024 / \sqrt{13} = [80.01, 80.03]$$

Insbesondere liegt 80.00 nicht im Intervall I . Der Wert $\mu = 80.00$ ist folglich nicht mit den Daten kompatibel, was wir bereits mit Hilfe des t -Tests ermittelt hatten. Dieses Intervall wurde schon oben mit dem R-Befehl `t.test(...)` bestimmt.

□

6.4. Statistische Tests bei nicht-normalverteilten Daten

Der z - und t -Test sind optimal, falls die Daten Realisierungen von normalverteilten Zufallsvariablen sind, also

$$X_1, \dots, X_n \text{ i.i.d. } \sim \mathcal{N}(\mu, \sigma_X^2)$$

Optimalität bedeutet hier, dass der Test die grösste Macht hat.

Wir betrachten in diesem Kapitel die allgemeinere Situation, in der die Daten Realisierungen sind von

$$X_1, \dots, X_n \text{ i.i.d.}$$

wobei X_i einer beliebigen Verteilung folgen kann. Wir bezeichnen mit μ einen Lageparameter der Verteilung (z.B. $\mu = \text{Median der Verteilung von } X_i$). Die Nullhypothese ist von der Form $H_0: \mu = \mu_0$.

6.4.1. Der Vorzeichentest

Der Vorzeichentest testet Hypothesen über den Median der Verteilung von X_i , den wir hier mit μ bezeichnen. Im Falle einer symmetrischen Verteilung ist $\mu = E(X_i)$. Beim Vorzeichentest betrachten wir die Anzahl Beobachtungen, die grösser oder kleiner als μ sind.

Beispiel 6.4.1

Wir betrachten nochmal die Daten zur Messung der Schmelzwärme mit der Methode A (siehe Tabelle 6.2).

| | | | | | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Methode A | 79.98 | 80.04 | 80.02 | 80.04 | 80.03 | 80.03 | 80.04 | 79.97 | 80.05 |
| Methode A | 80.03 | 80.02 | 80.00 | 80.02 | | | | | |

Tabelle 6.2.: Messung der Schmelzwärme mit der Methode A.

Wir berechnen nun die Unterschiede zum vermuteten Median $\mu = 80.00$ (siehe Tabelle 6.3).

Bei der Testdurchführung untersuchen wir die Anzahl Erfolge „+“ gegen??ber der Anzahl Misserfolgen „−“. Ist $\mu = 80.00$ tatsächlich der Median, so ist die Wahrscheinlichkeit, dass ein Messwert ein positives Vorzeichen hat, $\pi = 0.5$. Somit haben wir es hier mit einem Binomialtest mit $n = 13$ und $\pi_0 = 0.5$ zu tun.

```
binom.test(11, 13, p = 0.5)

##
##  Exact binomial test
##
## data:  11 and 13
## number of successes = 11, number of trials = 13,
## p-value = 0.02246
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##   0.5455289 0.9807933
## sample estimates:
## probability of success
##                0.8461538
```

Die Nullhypothese ist $H_0: \mu = \mu_0 = 80.00$. Die Teststatistik V ist die Anzahl positiver Vorzeichen von $X_i - \mu_0$. Die realisierte Teststatistik ist dann $v = 11$ und der P -Wert des Vorzeichentests bei zweiseitiger Alternative $H_A: \mu \neq \mu_0 = 80.00$ ist 0.02246 (beim t -Test war der P -Wert 0.0088). Wir m??ssen die Nullhypothese also wieder *verwerfen*.

| k | x_k | $x_k - \mu_0$ | Vorzeichen |
|-----|-------|---------------|------------|
| 1 | 79.98 | -0.02 | - |
| 2 | 80.04 | 0.04 | + |
| 3 | 80.02 | 0.02 | + |
| 4 | 80.04 | 0.04 | + |
| 5 | 80.03 | 0.03 | + |
| 6 | 80.03 | 0.03 | + |
| 7 | 80.04 | 0.04 | + |
| 8 | 79.97 | -0.03 | - |
| 9 | 80.05 | 0.05 | + |
| 10 | 80.03 | 0.03 | + |
| 11 | 80.02 | 0.02 | + |
| 12 | 80.00 | 0.00 | + |
| 13 | 80.02 | 0.02 | + |

Tabelle 6.3.: Vorzeichen der Schmelzwärme Messung mit der Methode A in Bezug zum Median.

□

Wenn μ der Median der Verteilung von X ist, dann ist die Wahrscheinlichkeit, dass eine Realisierung von X grösser als μ ist, genauso gross wie die Wahrscheinlichkeit, dass eine Realisierung von X kleiner als μ ist, also $P(X > \mu) = 0.5$.

Zweiseitiger Vorzeichentest

Der (zweiseitige) Vorzeichen-Test ist folgendermassen aufgebaut:

1. **Modell:**

$$X_1, \dots, X_n \text{ i.i.d.}$$

wobei X_i eine beliebige Verteilung hat.

2. **Nullhypothese:**

$$H_0 : \mu = \mu_0 \quad (\mu \text{ ist der Median})$$

Alternative:

$$H_A : \mu \neq \mu_0 \quad (\text{oder einseitige Variante})$$

3. **Teststatistik:**

$$V : \text{Anzahl } X_i\text{'s mit } (X_i \geq \mu_0)$$

Verteilung der Teststatistik unter H_0 :

$$V \sim \text{Bin}(n, \pi_0) \quad \text{mit} \quad \pi_0 = 0.5$$

4. **Signifikanzniveau:**

$$\alpha$$

5. **Verwerfungsbereich für die Teststatistik:**

$$K = [0, c_u] \cup [c_o, n] \text{ falls } H_A : \mu \neq \mu_0$$

Die Grenzen c_u und c_o müssen mit der Binomialverteilung oder der Normalapproximation berechnet werden.

6. **Testentscheid:**

Entscheide, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich der Teststatistik liegt.

Bemerkungen:

- i. Der Vorzeichentest ist also nichts anderes als ein Binomialtest. Wenn wir $\mu_0 = 0$ wählen, entspricht die Teststatistik gerade der Anzahl „+“ im Datensatz, daher der Name „Vorzeichentest“. Wenn Sie den Binomialtest verstanden haben, müssen Sie für den Vorzeichentest also gar nichts Neues lernen.
- ii. Beim einseitigen Vorzeichentest mit Alternativhypothese $\mu > \mu_0$ ist die Teststatistik die Anzahl positiver Vorzeichen von $X_i - \mu$. Dementsprechend zeigt der

Verwerfungsbereich nach oben.

Alternativ könnten wir aber auch die negativen Vorzeichen von $X_i - \mu$ als Teststatistik definieren und die Alternativhypothese anpassen zu: $\mu < \mu_0$ und damit auch den nach unten zeigenden Verwerfungsbereich.

- iii. Beim einseitigen Vorzeichentest mit Alternativhypothese $\mu < \mu_0$ ist die Teststatistik die Anzahl positiver Vorzeichen von $X_i - \mu$. Dementsprechend zeigt der Verwerfungsbereich nach unten.

Alternativ könnten wir aber auch die negativen Vorzeichen von $X_i - \mu$ als Teststatistik definieren und die Alternativhypothese anpassen zu: $\mu > \mu_0$ und damit auch den jetzt nach oben zeigenden Verwerfungsbereich.

Der Vorzeichentest ist immer angebracht, falls die Daten Realisierungen von

$$X_1, \dots, X_n \text{ i.i.d.}$$

sind. Die Wahrscheinlichkeit für einen Fehler 1. Art wird kontrolliert durch α bei beliebiger Verteilung der X_i 's.

Vom Standpunkt der Macht gibt es keine eindeutige Antwort, ob der Vorzeichen- oder der t -Test besser ist. Wenn die Verteilung der X_i langschwänzig ist, kann der Vorzeichentest grössere Macht haben. Weil der Vorzeichentest die Information, um wie viel die X_i von dem Wert μ_0 abweichen (siehe die Definition der Teststatistik V oben), nicht ausnützt, kann die Macht aber auch wesentlich schlechter sein als beim t -Test.

6.4.2. Der Wilcoxon-Test

Der Wilcoxon-Test ist ein Kompromiss, der weniger voraussetzt als der t -Test, die Information der Daten aber besser ausnützt als der Vorzeichentest. Der Wilcoxon-Test setzt bloss voraus, dass die Verteilung unter der Nullhypothese *symmetrisch* ist bezüglich μ_0 .

Als erstes bildet man die Ränge der Differenzen bezüglich des Absolutwertes:

$$\text{Rang}(|X_i - \mu_0|) = k$$

wobei $|X_i - \mu_0|$ den k -ten kleinsten Wert hat unter

$$|X_1 - \mu_0|, \dots, |X_n - \mu_0|$$

Wenn einzelne Werte zusammenfallen, teilt man die Ränge auf durch Mittlung. Falls für ein X_i

$$|X_i - \mu_0| = 0$$

ist, dann lässt man diesen Datenpunkt weg. Ferner sei V_i der Indikator dafür, ob $X_i - \mu_0$ positiv ist, d.h. $V_i = 1$ falls $X_i > \mu_0$ ist und $V_i = 0$ sonst. Dann verwirft man die Nullhypothese, falls

$$W = \sum_{i=1}^n \text{Rang}(|X_i - \mu_0|) V_i$$

entweder zu gross oder zu klein ist, je nach Spezifikation der Alternative. Wir gehen an dieser Stelle nicht darauf ein, wie die Verteilung von W beschaffen ist. Wir werden den Testentscheid direkt am P -Werts des Computer-Outputs ablesen.

Man kann zeigen, dass dieser Test das Niveau exakt einhält: Die Wahrscheinlichkeit für einen Fehler 1. Art ist gleich α , wenn die X_i i.i.d. sind und eine um μ_0 symmetrische Dichte haben. Beim t -Test wird das Niveau auch ungefähr eingehalten bei vielen nicht-normalen Verteilungen (wegen dem zentralen Grenzwertsatz), aber die Wahrscheinlichkeit eines Fehlers 2. Art ist unter Umständen beim t -Test viel grösser als beim Wilcoxon-Test.

Beispiel 6.4.2

Für die Schmelzwärme der Methode A sind die Ränge und die V_i 's in Tabelle 6.4 dargestellt.

| k | x_k | $ x_k - \mu_0 $ | $\text{Rang}(x_k - \mu_0)$ | V_k |
|-----|-------|-----------------|------------------------------|-------|
| 1 | 79.98 | 0.02 | 2.5 | 0 |
| 2 | 80.04 | 0.04 | 10 | 1 |
| 3 | 80.02 | 0.02 | 2.5 | 1 |
| 4 | 80.04 | 0.04 | 10 | 1 |
| 5 | 80.03 | 0.03 | 6.5 | 1 |
| 6 | 80.03 | 0.03 | 6.5 | 1 |
| 7 | 80.04 | 0.04 | 10 | 1 |
| 8 | 79.97 | 0.03 | 6.5 | 0 |
| 9 | 80.05 | 0.05 | 12 | 1 |
| 10 | 80.03 | 0.03 | 6.5 | 1 |
| 11 | 80.02 | 0.02 | 2.5 | 1 |
| 12 | 80.00 | 0.00 | | 1 |
| 13 | 80.02 | 0.02 | 2.5 | 1 |

Tabelle 6.4.: Daten und entsprechende Ränge im Beispiel der Schmelzwärme Messung mit der Methode A.

Betrachten wir den Wert $x_1 = 79.98$. $(|x_1 - \mu_0|) = |79.98 - 80.00|$ ist der kleinste Wert, da der Fall $(|x_{12} - \mu_0|) = |80.00 - 80.00| = 0$ weggelassen wird. Folglich ist der Rang von $x_1 = 79.98$ 1. Nun gibt es drei weitere Werte mit einem Absolutbetrag der Differenz von 0.2. Somit teilen wir die Summe der Ränge 1, 2, 3 und 4 durch vier, was den Wert 2.5 ergibt.

Die Rangsumme lautet dann

$$W = 0 + 10 + 2.5 + 10 + 6.5 + 6.5 + 10 + 0 + 12 + 6.5 + 2.5 + 0 + 2.5 = 69$$

Die Nullhypothese lautet $H_0: \mu = \mu_0 = 80.00$. Der Wilcoxon-Test ergibt dann bei zweiseitiger Alternative $H_A: \mu \neq \mu_0 = 80.00$ einen P -Wert von 0.0195. Mit R erfolgt der Wilcoxon-Test folgendermassen:

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04,
       79.97, 80.05, 80.03, 80.02, 80, 80.02)

wilcox.test(x, mu = 80, alternativ = "two.sided")

##
##  Wilcoxon signed rank test with continuity
##  correction
##
## data:  x
## V = 69, p-value = 0.0195
## alternative hypothesis: true location is not equal to 80
```

□

Wilcoxon-Test

Der Wilcoxon-Test ist in den allermeisten Fällen dem t -Test oder Vorzeichentest vorzuziehen: er hat in vielen Situationen oftmals wesentlich grössere Macht und selbst in den ungünstigsten Fällen ist er nie viel schlechter. Wenn man trotzdem den t -Test verwendet, dann sollte man die Daten auch graphisch ansehen, damit wenigstens grobe Abweichungen von der Normalverteilung entdeckt werden. Insbesondere sollte der Normal-Plot (siehe Abschnitt 6.1) angeschaut werden.

6.5. Statistische Tests bei zwei Stichproben

Wir besprechen in diesem Kapitel statistische Methoden, um einen Vergleich zwischen zwei Gruppen, Versuchsbedingungen oder Behandlungen hinsichtlich der La-

ge der Verteilung anzustellen.

6.5.1. Gepaarte Stichproben

Struktur der Daten

Wann immer möglich sollte man eine Versuchseinheit beiden Versuchsbedingungen unterwerfen. Es liegt eine **gepaarte Stichprobe** vor, wenn

- beide Versuchsbedingungen an derselben Versuchseinheit eingesetzt werden
- oder wenn jeder Versuchseinheit aus der einen Gruppe genau eine Versuchseinheit aus der anderen Gruppe zugeordnet werden kann.

Die Daten sind dann von der folgenden Struktur:

$$\begin{aligned}x_1, \dots, x_n &\text{ unter Versuchsbedingung 1} \\ y_1, \dots, y_n &\text{ unter Versuchsbedingung 2}\end{aligned}$$

Notwendigerweise ist dann die Stichprobengrösse n für beide Versuchsbedingungen dieselbe. Zudem sind x_i und y_i abhängig, weil die Werte von der gleichen Versuchseinheit kommen.

Beispiel 6.5.1

Wir testen den Muskelzuwachs durch ein Krafttraining. Dazu messen wir die Kraft von 10 Testpersonen zu Beginn des Trainings. Anschliessend durchlaufen alle Testpersonen ein 6-wöchiges Trainingsprogramm. Dann wird die Kraft erneut gemessen. Für jede Testperson gibt es also zwei Messungen: Vorher und nachher, die Zuordnung ist eindeutig. Somit handelt es sich um gepaarte Stichproben.

□

Beispiel 6.5.2

Die Wirksamkeit von Augentropfen zur Reduktion des Augeninnendrucks soll untersucht werden. Wir haben 12 Patienten. Bei jedem Patienten wählen wir zufällig ein Auge aus. In das eine Auge kommen die Augentropfen mit dem Wirkstoff. In das andere Auge kommen Tropfen ohne Wirkstoff (Placebo). Für jede Testperson haben wir also zwei Messungen: Eine für das rechte und eine für das linke Auge; die Zuordnung ist eindeutig. Somit handelt es sich um gepaarte Stichproben.

□

Beispiel 6.5.3

Wir haben eine Gruppe von 15 eineiigen Zwillingen, die sich für eine Studie für ein Haarwuchsmittel gemeldet haben. Bei jedem Zwillingenspaar wird eine Person zufällig ausgewählt und erhält das Medikament. Die andere Person des Zwillingenspaares erhält ein Placebo. Nach drei Wochen misst man den Haarwuchs. Zu jeder Person aus der Gruppe mit Haarwuchsmittel kann man eindeutig eine Person aus der Gruppe ohne Haarwuchsmittel zuordnen. Somit handelt es sich um gepaarte Stichproben.

□

Statistischer Test für gepaarte Stichproben

Bei der Analyse von gepaarten Vergleichen arbeitet man mit den Differenzen innerhalb der Paare,

$$d_i = x_i - y_i \quad (i = 1, \dots, n),$$

welche wir als Realisierungen von i.i.d. Zufallsvariablen D_1, \dots, D_n auffassen. Kein Unterschied zwischen den beiden Versuchsbedingungen heisst dann einfach

$$E[D_i] = 0$$

(oder auch $\text{Median}(D_i) = 0$, je nach Test). Statistische Tests dafür sind in Unterkapitel 6.3 beschrieben: Falls die Daten normalverteilt sind, eignet sich ein t -Test. Einen t -Test für gepaarte Stichproben kann man leicht durchführen, indem man das Argument `paired = TRUE` im R-Befehl `t.test()` verwendet.

```
vorher <- c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)

nachher <- c(27, 29, 37, 56, 46, 82, 57, 80, 61, 59, 43)

t.test(nachher, vorher, alternative = "two.sided", mu = 0,
       paired = TRUE, conf.level = 0.95)

##
## Paired t-test
##
## data:  nachher and vorher
## t = 4.2716, df = 10, p-value = 0.001633
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  4.91431 15.63114
## sample estimates:
## mean of the differences
##                10.27273
```

Bemerkungen:

- i. Man hätte auch direkt die Differenzen d_i von den gepaarten Stichproben berechnen und einen t -Test für eine Stichprobe durchführen können.
- ii. Der Unterschied der Gruppenmittelwerte hat bei einer zweiseitigen Alternative (siehe Zeile `alternative hypothesis: ...`) einen P -Wert von 0.0016 und ist somit auf dem 5 % Signifikanzniveau signifikant.
- iii. Der Wert der Teststatistik ist 4.27 und folgt unter der Nullhypothese einer t -Verteilung mit $df = 10$ Freiheitsgraden.
- iv. Der Unterschied `nachher-vorher` erscheint im Funktionsaufruf, indem sich das erste Argument auf das „nachher“ und das zweite Argument auf das „vorher“ bezieht.
- v. Das 95 %-Vertrauensintervall für diese Differenz ist gegeben durch

$$[4.91, 15.63]$$

Alternativ kommt ein Vorzeichentest oder ein Wilcoxon-Test in Frage. Dabei ist zu beachten, dass die vorausgesetzte Symmetrie für die Verteilung von D_i beim Wilcoxon-Test immer gilt unter der Nullhypothese, dass X_i und Y_i dieselbe Verteilung haben.

6.5.2. Ungepaarte Stichproben

Oft ist es nicht möglich, jeder Behandlungseinheit aus der einen Gruppe eine Behandlungseinheit aus der zweiten Gruppe eindeutig zuzuordnen. In diesem Fall ist eine Paarung nicht möglich und man spricht von einer ungepaarten Stichprobe. Auch hier muss die Zuordnung zur Behandlungsgruppe durch das Los erfolgen um systematische Fehler zu vermeiden.

Struktur der Daten

Bei ungepaarten Stichproben hat man Daten x_1, \dots, x_n und y_1, \dots, y_m (siehe Abschnitt 6.5.2), welche wir als Realisierungen der folgenden Zufallsvariablen auffassen:

$$X_1, \dots, X_n \text{ i.i.d.}$$

$$Y_1, \dots, Y_m \text{ i.i.d.}$$

wobei auch alle X_i 's von allen Y_j 's unabhängig sind. Bei einer solchen zufälligen Zuordnung von Versuchseinheiten zu einer von zwei verschiedenen Versuchsbedingungen spricht man von einer ungepaarten Stichprobe. Im allgemeinen ist in einer ungepaarten Stichprobe $m \neq n$, aber nicht notwendigerweise. Entscheidend ist, dass x_i

und y_i zu verschiedenen Versuchseinheiten gehören und als unabhängig angenommen werden können.

Beispiel 6.5.4

Datensatz zu latenter Schmelzwärme von Eis im Kapitel 2.2. Wir haben die Schmelzwärme mit zwei verschiedenen Methoden hintereinander gemessen. Jede Messung ist entweder mit Methode A oder mit Methode B, aber nicht mit beiden gleichzeitig gemacht worden. Es gibt also keinen eindeutigen Zusammenhang zwischen den Messungen der Methode A und den Messungen der Methode B. Daher sind die beiden Stichproben ungepaart.

□

Beispiel 6.5.5

Zufällige Zuordnung von 100 Testpatienten zu einer Gruppe der Grösse 50 mit Medikamentenbehandlung und zu einer anderen Gruppe der Grösse 50 mit Placebo-Behandlung. Es gibt keine eindeutige Zuordnung von einem Patienten aus der Medikamentengruppe zu einem Patienten in der Placebo-Gruppe. Daher handelt es sich um ungepaarte Stichproben, obwohl beide Gruppen gleich gross sind.

□

Zwei-Stichproben t -Test für ungepaarte Stichproben

Nehmen wir an, wir haben einen ungepaarten Datensatz mit

$$\begin{aligned} X_1, \dots, X_n \text{ i.i.d. } &\sim \mathcal{N}(\mu_X, \sigma_X^2) \\ Y_1, \dots, Y_m \text{ i.i.d. } &\sim \mathcal{N}(\mu_Y, \sigma_Y^2) \end{aligned}$$

Als Beispiel betrachten wir den Datensatz mit den Wärmekapazitäten und analysieren mit R, ob es einen signifikanten Unterschied zwischen den beiden Messmethoden A und B gibt.

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04,
       79.97, 80.05, 80.03, 80.02, 80, 80.02)

y <- c(80.02, 79.94, 79.98, 79.97, 80.03, 79.95, 79.97)

t.test(x, y, alternative = "two.sided", mu = 0, paired = FALSE,
       conf.level = 0.95)

##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 2.8399, df = 9.3725, p-value = 0.01866
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.008490037 0.073048425
## sample estimates:
## mean of x mean of y
##  80.02077  79.98000
```

Bemerkungen:

- i. Die ersten beiden Argumente enthalten die Daten der beiden Stichproben. Das Argument `alternative` gibt an, ob die Alternative einseitig (und wenn ja in welche Richtung mit `alternative = „greater“` und `alternative = „less“`) oder zweiseitig (mit `alternative = „two.sided“`) ist. Das Argument `mu` gibt an, welcher Unterschied in den Mittelwerten der beiden Gruppen in der Nullhypothese getestet werden soll.
- ii. Wenn man testen will, ob die beiden Gruppenmittelwerte gleich sind, ist `mu=0` die richtige Wahl.
- iii. `paired = FALSE` gibt an, dass es sich um zwei ungepaarte Stichproben handelt. Mit `conf.level = 0.95` wird ein 95%-Vertrauensintervall des Unterschieds zwischen den beiden Gruppenmittelwerten ausgegeben.
- iv. In der Zeile `t=...` steht zunächst der beobachtete Wert der Teststatistik: $t = 2.8399$. Unter der Nullhypothese folgt die Teststatistik einer t -Verteilung mit $df = 19$ Freiheitsgraden. Das ergibt bei einer zweiseitigen Alternative (siehe Zeile `alternative hypothesis: ...`) einen P -Wert von 0.01866. Der Unterschied ist also auf dem 5% Signifikanzniveau signifikant, weil der P -Wert kleiner als 5% ist.
- v. Der Computer berechnet auch das 95%-Vertrauensintervall des Unterschieds in den Gruppenmittelwerten: Mit 95 % Wahrscheinlichkeit ist der Gruppenmittelwert von x um eine Zahl im Bereich $[0.0085, 0.0730]$ grösser als der Gruppenmit-

telwert von y . Die Null ist nicht enthalten, also ist der Unterschied der Mittelwerte signifikant.

- vi. In der letzten Zeile werden schliesslich noch die Mittelwerte der beiden Gruppen angegeben. Beachten Sie, dass kein Verwerfungsbereich ausgegeben wird.

Zwei-Stichproben Wilcoxon-Test (Mann-Whitney-Test)

Die Voraussetzungen für den Zwei-Stichproben Wilcoxon-Test, manchmal auch Mann-Whitney Test genannt, bezüglich

$$X_1, \dots, X_n \text{ i.i.d.}$$

$$Y_1, \dots, Y_m \text{ i.i.d.}$$

sind wie folgt:

$$X_1, \dots, X_n \text{ i.i.d.} \sim F_X,$$

$$Y_1, \dots, Y_m \text{ i.i.d.} \sim F_Y,$$

$$F_X \text{ beliebige stetige Verteilungsfunktion, } F_Y(x) = F_X(x - \delta).$$

Dies bedeutet, dass die Verteilung von Y_j die um δ verschobene Verteilung von X_i ist, denn:

$$P(Y_j \leq x + \delta) = F_Y(x + \delta) = F_X(x + \delta - \delta) = F_X(x) = P(X_i \leq x)$$

Die Berechnung des P -Werts eines Zwei-Stichproben Wilcoxon-Tests kann mittels Computer erfolgen. Aus den gleichen Gründen wie im Fall einer Stichprobe (siehe Kapitel 6.4) ist der Wilcoxon-Test im Allgemeinen dem t -Test vorzuziehen.

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97, 80.05,
      80.03, 80.02, 80, 80.02)

y <- c(80.02, 79.94, 79.98, 79.97, 80.03, 79.95, 79.97)

wilcox.test(x, y, alternative = "two.sided", mu = 0)

##
## Wilcoxon rank sum test with continuity
## correction
##
## data: x and y
## W = 76.5, p-value = 0.01454
## alternative hypothesis: true location shift is not equal to 0
```

Auch der Wilcoxon-Test verwirft die Nullhypothese $\mu = 0$, und zwar aufgrund des P -Wertes 0.01454, der etwas kleiner ist als im Falle des t -Tests.

6.6. Statistische Signifikanz und fachliche Relevanz bei statistischen Tests

Der Begriff der statistischen Signifikanz wird oft missbraucht, um gleichzeitig auch die entsprechende fachliche Relevanz zu untermauern. Diese beiden Begriffe müssen aber nicht unbedingt miteinander einhergehen. Wenn man genügend viele Beobachtungen sammelt, dann wird man jede Nullhypothese verwerfen können (denn diese stimmt in der Praxis nie exakt). Bedeutet dies nun, dass die vorgestellten Konzepte in der Praxis alle nutzlos sind? Die Antwort ist natürlich nein, aber man muss sie richtig verwenden.

Hierzu müssen wir das beste aus beiden Bereichen miteinander kombinieren: Entsprechendes Fachwissen und der statistische Output. Wir müssen zuerst basierend auf Fachwissen definieren, was ein relevanter Effekt oder Unterschied ist (die Statistik kann uns hier nicht helfen). Wenn wir dies gemacht haben, können wir die Statistik ins Spiel bringen.

Betrachten wir als Beispiel die Produktion von Schrauben: Nehmen wir an, dass Abweichungen bis 0.5 mm von der Solllänge 100 mm keine Rolle spielen, also nicht relevant sind. Wir haben also einen „irrelevanten Bereich“, der von 99.5 mm bis 100.5 mm geht. Ausserhalb sprechen wir vom *Relevanzbereich*. Die Idee besteht nun darin, zu schauen, wie das Vertrauensintervall für μ liegt. Angenommen wir haben in einer Stichprobe das Vertrauensintervall für das wahre μ zu $[99.73, 100.31]$ bestimmt, was vollständig im irrelevanten Bereich liegt. Wir würden daher die Abweichung als statistisch signifikant, aber als nichtrelevant taxieren. Andere mögliche Fälle und deren Interpretation sind in Abbildung 6.6 dargestellt.

Lernziele

- ☐ Sie können den z -, t - und Vorzeichentest für eine Stichprobe auf Papier und Computer durchführen (einseitig und zweiseitig). Sie können entscheiden, ob ein einseitiger oder ein zweiseitiger Test angebracht ist. Sie können den Computeroutput eines t -Tests interpretieren (bei der Software R).
- ☐ Sie können den t -Test für zwei gepaarte Stichproben auf Papier und Computer durchführen (einseitig, zweiseitig; gepaart). Sie können entscheiden, ob ein einseitiger/zweiseitiger und gepaarter/ungepaarter Test angebracht ist. Sie können Computeroutputs eines t -Tests interpretieren (bei der Software R).
- ☐ Sie wissen, dass es den Wilcoxon-Test als Alternative zum t -Test gibt. Sie können den Wilcoxon-Test für eine oder zwei Stichproben mit dem Computer durchführen und den Computeroutput eines Wilcoxon-Tests interpretieren (bei der Software R).

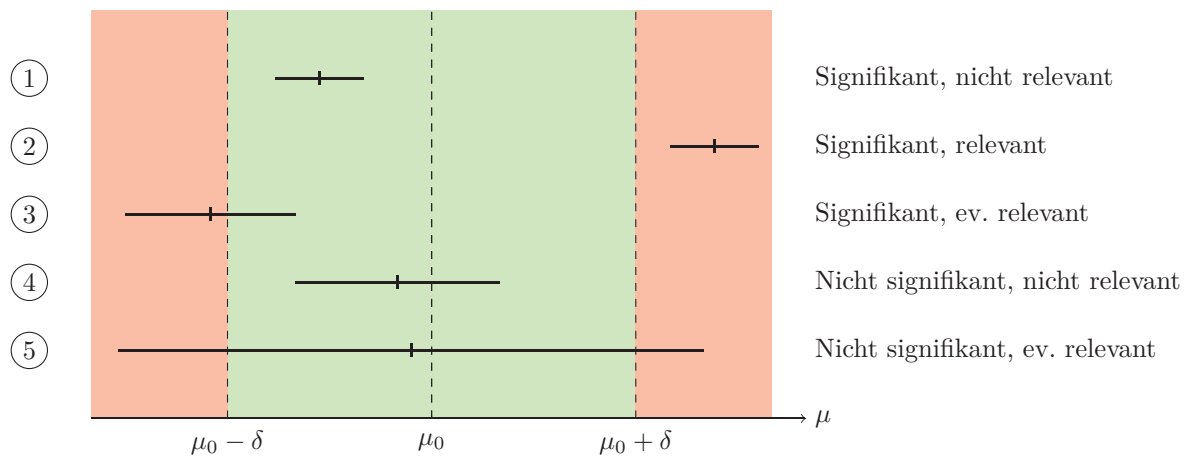


Abbildung 6.6.: Verschiedene Fälle (1 bis 5) von statistischer Signifikanz und fachlicher Relevanz. Die Vertrauensintervalle für μ sind durch Striche dargestellt. Der „irrelevante Bereich“ geht von $\mu_0 - \delta$ bis zu $\mu_0 + \delta$ (grün), wobei das δ durch entsprechendes Fachwissen definiert wurde.

Kapitel 7.

Stochastische Prozesse

The true logic of this world is in the calculus of probabilities.

(James Clerk Maxwell)

7.1. Vom Random Walk zum Thermischen Rauschen von elektrischen Widerständen

7.1.1. Random Walk

Ein *Random Walk*, auch Zufallsbewegung oder Irrfahrt genannt, ist ein mathematisches Modell für eine Bewegung, bei der die einzelnen Schritte zufällig erfolgen. Es handelt sich um einen stochastischen Prozess in diskreter Zeit mit unabhängigen und identisch verteilten Zuwächsen.

Beispiel 7.1.1 Aktienkurs

Aktienkurse können nicht vorhergesagt werden. Durch Random-Walks können die zufälligen Kursanstiege oder Kurszerfälle von Tag zu Tag (diskrete Zeit) modelliert werden.

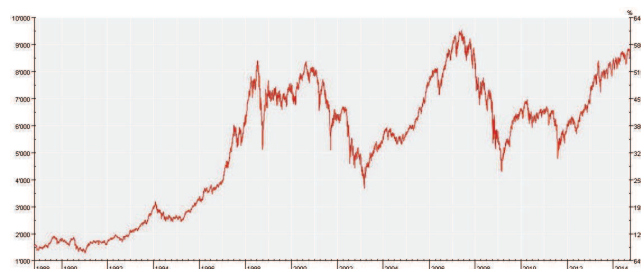


Abbildung 7.1.: Aktienkurs : SMI-Entwicklung 1989-2014.

□

Beispiel 7.1.2 Weisses Rauschen

In einem Leiter, an dem keine Spannung angelegt ist, bewegen sich die Elektronen trotzdem, und zwar in zufälliger Art und Weise aufgrund der Wärme. Dies führt zu einer kleinen Spannung im Leiter, die dauernd zufällig ändert. Wir sprechen in diesem Fall von *weissem Rauschen*.



Abbildung 7.2.: Weisses Rauschen.

Auch diesen Prozess können wir mit Hilfe eines Random-Walks modellieren.

□

Beispiel 7.1.3 Galton-Brett

Ein weiteres Beispiel für einen Random Walk stellt das sogenannte Galtonsche Nagelbrett dar. Die Sprungwahrscheinlichkeit der Kugel auf einem Nagel nach links oder nach rechts ist die gleiche, nämlich 50 %. Die Verteilung der Kugeln nach N Sprüngen folgt einer Binomialverteilung.

□

Um den Begriff des *Random Walks* besser zu verstehen, stellen wir uns einen Betrunkenen vor, der aus einer Bar in eine Gasse tritt und versucht, nach Hause zu gehen. Da er etwas durcheinander ist, geht nicht jeder seiner Schritte in die richtige Richtung. Wir gehen davon aus, dass jeder seiner Schritte die Grösse Δx hat und mit Wahrscheinlichkeit p nach rechts geht und mit Wahrscheinlichkeit $q = 1 - p$ nach links. Wenn $p > q$ ist, geht er häufiger nach rechts als nach links. Der Weg des Betrunkenen

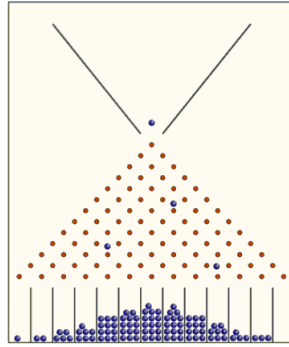


Abbildung 7.3.: Galton-Nagelbrett

ist ein Beispiel für einen Zufallspfad („random walk“) auf einem eindimensionalen Gitter mit der Gitterkonstanten Δx (siehe Abbildung 7.4).

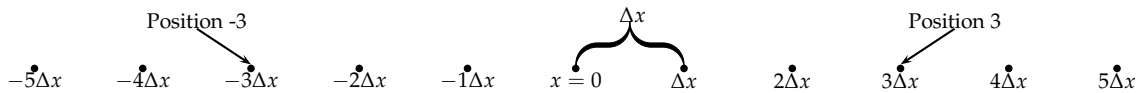


Abbildung 7.4.: Schritte auf einem eindimensionalen Gitter bei einem Random-Walk.

Wir nehmen an, dass unser Betrunkener am Ort $x = 0$ seinen nächtlichen Spaziergang beginnt. Bezeichnen wir mit X die Anzahl Schritte, die unser Betrunkener nach rechts geht, dann ist $X \sim \text{Bin}(N, p)$, wobei N die Gesamtzahl Schritte ist.

Nun möchten wir wissen, wie gross die Wahrscheinlichkeit ist, dass sich unser Betrunkener nach N Schritten an der Position m (in Einheiten von Δx) auf dem Gitter befindet. Man beachte, dass m auch negativ sein kann. Wir bezeichnen mit M_N die Zufallsvariable für die Position unseres Spaziergängers nach N Schritten auf dem Gitter (in Einheiten von Δx). Wir fragen also, wie gross die Wahrscheinlichkeit $P(M_N = m)$ ist, dass der Pfad nach N Schritten an der Position m ist. Weiter bezeichnen wir mit S_r die Anzahl Schritte nach rechts und mit S_l die Anzahl Schritte nach links. Nach unseren Voraussetzungen muss dann gelten

$$S_r + S_l = N \quad \text{und} \quad S_r - S_l = m$$

Dies ist ein lineares Gleichungssystem mit zwei Gleichungen und den beiden Unbekannten S_r und S_l . Wir lösen nach S_r und S_l auf, indem wir die beiden obigen Gleichungen addieren und subtrahieren:

$$S_r = \frac{N + m}{2} \quad \text{und} \quad S_l = \frac{N - m}{2}$$

Um nach N Schritten bei m zu sein, müssen wir also $(N + m)/2$ von den N Schritten nach rechts und $(N - m)/2$ Schritte nach links gehen. $m + N$ muss eine gerade Zahl

sein: wenn also N gerade ist, ist auch m gerade; ist N ungerade, so ist es m . Die Anzahl Schritte nach rechts bestimmt die Position M_N eindeutig. Wir können die Wahrscheinlichkeit, dass sich der Betrunkene nach insgesamt N Schritten bei $M_N = m$ befindet, berechnen, indem wir die Zufallsvariable X (Anzahl Schritte nach rechts) durch $(N + m)/2$ ersetzen:

$$P(M_N = m) = \binom{N}{\frac{N+m}{2}} p^{(N+m)/2} q^{(N-m)/2}$$

Wenn N gross ist (dies ist bei einem betrunkenen Bargänger normalerweise der Fall), erhalten wir unter Verwendung des Zentralen Grenzwertsatzes eine Normalverteilung. Denn eine binomialverteilte Zufallsvariable kann als die Summe von N unabhängigen Bernoulli-verteilten Zufallsvariablen aufgefasst werden, und für grosse N kann die Verteilung der Summe von beliebig verteilten Zufallsvariablen als Normalverteilung angenähert werden (siehe Abschnitt 5.4.3 zur Normalapproximation). Folglich ist die Wahrscheinlichkeit, dass sich der Bargänger nach N Schritten bei $M_N = m$ befindet:

$$P(M_N = m) = \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(m-\mu)^2}{2\sigma^2}} \quad (7.1)$$

Der Vorfaktor 2 ist so gewählt, dass die Summe über alle m -Werte eins ergibt, die Punktwahrscheinlichkeiten also normiert sind. Hierbei muss man beachten, dass für (un)gerade N nur (un)gerade Werte von m auftauchen, so dass der Abstand zweier benachbarter Werte von m den Betrag 2 hat:

$$\Delta m = \pm 2$$

Diese Schrittgrösse steht im Zähler des Vorfaktors, so dass die Summe von $P(M_N = m)$ über alle m mit $\Delta m = \pm 2$ eins ergibt. Es handelt sich bei $P(M_N = m)$ immer noch um eine diskrete Wahrscheinlichkeitsverteilung: $P(M_N = m)$ ist die Punktwahrscheinlichkeit für die diskrete Zufallsvariable M_N .

Wie bestimmen wir nun die Werte der Parameter μ und σ in der Gleichung (7.1)? Als Erwartungswert der Normalverteilung für die Zufallsvariable M_N nehmen wir den Erwartungswert der Binomialverteilung: N mal die mittlere Positionsänderung eines Schrittes (in Einheiten von Δx),

$$\begin{aligned} \mu = E(M_N) &= N \cdot E(M_1) = N \left(1 \cdot P(M_1 = +1) - 1 \cdot P(M_1 = -1) \right) \\ &= N(p - q) \\ &= N(p - (1 - p)) = N(2p - 1) \end{aligned}$$

wobei M_1 Bernoulli-verteilt ist. Die Varianz der Normalverteilung ist N mal die Varianz eines Schrittes

$$\begin{aligned}\sigma^2(M_N) &= \text{Var}(M_N) \\ &= N \cdot \text{Var}(M_1) \\ &= N \left((1 - E(M_1))^2 P(M_1 = +1) + (-1 - E(M_N))^2 P(M_1 = -1) \right) \\ &= N \left((1 - (p - q))^2 p + (-1 - (p - q))^2 q \right) \\ &= 4Npq\end{aligned}$$

Also ist die Wahrscheinlichkeit, dass sich unser nächtlicher Spaziergänger nach N Schritten an der Position m befindet

$$\begin{aligned}P(M_N = m) &= \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(m-\mu)^2}{2\sigma^2}} \\ &= \frac{2}{\sqrt{8\pi Npq}} e^{-\frac{(m-N(p-q))^2}{8Npq}}\end{aligned}$$

Wir können uns nun fragen, wie weit unser Betrunkener nach N Schritten im Durchschnitt kommt. Die Antwort ist durch den Erwartungswert der Position, also

$$\mu = E[M_N] = N(2p - 1)$$

gegeben. Falls unser Spaziergänger so betrunken ist, dass jeder seiner Schritte mit gleicher Wahrscheinlichkeit nach rechts oder links geht, also $p = 1/2$ ist, dann ist seine erwartete Position null. Dies hat einerseits damit zu tun, dass der Barbesucher mit grösster Wahrscheinlichkeit wieder bei der Bar landet, andererseits aber, wenn er mal vorangekommen ist, gleich oft – über viele Barbesuche gemittelt – eine gewisse Distanz nach rechts wie nach links zurückgelegt hat. Diese Abstände heben sich dann auf.

Wollen wir das Vorankommen unseres Barbesuchers beschreiben, müssen wir die Vorzeichen der zurückgelegten Distanz „neutralisieren“. Wir könnten die mittlere zurückgelegte Distanz des Spaziergängers von der Startposition mit dem Erwartungswert des Absolutwertes von M_N , also $E[|M_N|]$, messen. Es ist allerdings vorteilhafter, das mittlere Quadrat der Verschiebung zu bestimmen, also $E[(M_N - M_0)^2]$, wobei $M_N - M_0$ die Verschiebung von der Anfangsposition M_0 bezeichnet. Im Fall von $p = 1/2$ und $M_0 = \mu = 0$ ist dies aber identisch mit

$$E[(M_N - M_0)^2] = E[(M_N - 0)^2] = E[(M_N - \mu)^2] = \sigma^2 = 4Npq.$$

Also für unseren betrunkenen Spaziergänger mit $p = 1/2$ ist $E[(M_N - M_0)^2] = N$.

Um die Einheit von Distanz zu bekommen, kann man die Wurzel vom **mittleren Verschiebungsquadrat**

$$\sqrt{\langle (M_N - M_0)^2 \rangle} \equiv \sqrt{E[(M_N - M_0)^2]}$$

betrachten, wobei die Notation $\langle \cdot \rangle$ eine Mittelung oder die Erwartungswertoperation bezeichnet. Im Falle von $p = q = 1/2$ und $M_0 = 0$ finden wir

$$\sqrt{\langle (M_N - M_0)^2 \rangle} = \sqrt{N}$$

Bei $N = 1000$ Schritten der Schrittweite 1 m hätte unser Betrunkener eine **quadratisch gemittelte Distanz** von 34 m zurückgelegt.

7.1.2. Brownsche Bewegung

Der Begriff **Brownsche Bewegung** bezieht sich auf die 1827 vom Biologen Robert Brown durchgeführten Arbeiten, in welchen dieser die zufällige Bewegung von in Wasser schwimmenden Pollen beobachtete. Einstein lieferte 1905 eine Erklärung dafür: die Zitterbewegung der Pollen wird durch fortwährende Stösse mit sich zufällig bewegendem Wassermolekülen verursacht (siehe Abbildung 7.5). Dies war in jener Zeit tatsächlich noch ein gewichtiges Argument für die Existenz von Atomen und Molekülen, die im 19. Jahrhundert noch heftig umstritten gewesen ist. Und gleichzeitig passte Einsteins Beschreibung zur molekularen Theorie der Wärme. Je wärmer beispielsweise Wasser ist, um so grösser ist die mittlere Geschwindigkeit, mit der die Wassermoleküle ungeordnet umherflitzen und damit Stösse verursachen.

Die mathematische Theorie der Brownschen Bewegung wurde von Louis Bachelier im Jahre 1900 in seiner Doktorarbeit „Theorie de la speculation“ entwickelt, die Random Walks mit der Entwicklung von Börsenkursen in Verbindung brachte. Der Durchbruch kam jedoch, als Albert Einstein 1905 ohne Kenntnis von Bacheliers Arbeiten und unabhängig von ihm Marian Smoluchowski (1906), den stochastischen Prozess der Brownschen Bewegung (auch Wiener-Prozess genannt) in seiner heutigen Gestalt definierte. Einen Beweis für die wahrscheinlichkeitstheoretische Existenz des Prozesses blieb Einstein allerdings schuldig. Dieser gelang erst 1923 dem US-amerikanischen Mathematiker Norbert Wiener.

Brownsche Bewegung ergibt sich aus einem **Random Walk**, wobei die Zeitvariable kontinuierlich gemacht wird und die Schrittweiten als normalverteilte Zufallsvariablen aufgefasst werden. Anstelle eines Betrunkenen betrachten wir dazu nun Partikel in Wasser. Geben wir zum Beispiel einen Tintentropf in Wasser, so diffundieren die Tintenmoleküle im Wasser und verteilen sich gleichmässig.

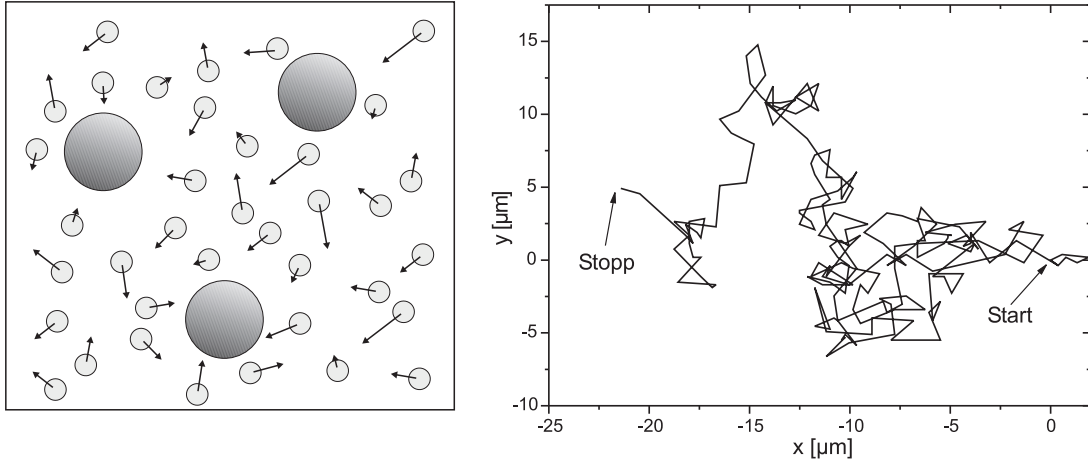


Abbildung 7.5.: *Links*: Modell der Brownschen Bewegung. Die Moleküle des umgebenden Mediums stossen aufgrund ihrer thermischen Energie mit den suspendierten Partikeln zusammen, wodurch sich diese auf völlig unregelmässigen Bahnen bewegen. *Rechts*: Gemessene Bahn eines einzelnen Partikels.

Nehmen wir an, in der Flüssigkeit befinden sich n Tintenmoleküle zum Zeitpunkt $t = 0$ an der Stelle x_0 , dann kann die zeitabhängige Teilchendichte als

$$n(x, t) = \frac{n}{\sqrt{4\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}} \quad (7.2)$$

beschrieben werden, wobei D die von der Temperatur T des Wassers abhängige Diffusionskonstante bezeichnet (siehe Abbildung 7.6). Aufgrund der Diffusion der Tintenmoleküle wird die Teilchendichtekurve mit grösser werdendem t immer breiter.

Wir können den Zusammenhang des makroskopisch beschriebenen Phänomens der *Diffusion* mit dem mikroskopischen Phänomen der *Brownschen Bewegung* eines Partikels verstehen, indem wir zu einer Kontinuumsbeschreibung des Random Walks übergehen. Wir fassen also die Bewegung des Tintenmoleküls im Wasser als einen Random Walk auf (in Analogie zum Barbesucher), lassen nun aber die Schrittlänge Δx sowie den zeitlichen Abstand Δt zwischen zwei Schritten immer kleiner werden. Dies wollen wir an dieser Stelle nicht machen¹, sondern nur das Schlussresultat angeben. Wir erhalten dann folgende Wahrscheinlichkeitsdichte für die Position x des Brownschen Partikels zur Zeit t

$$f(x; t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}} \quad (7.3)$$

¹Die (kurze) Herleitung wird in Kapitel C.1 des Anhangs besprochen.

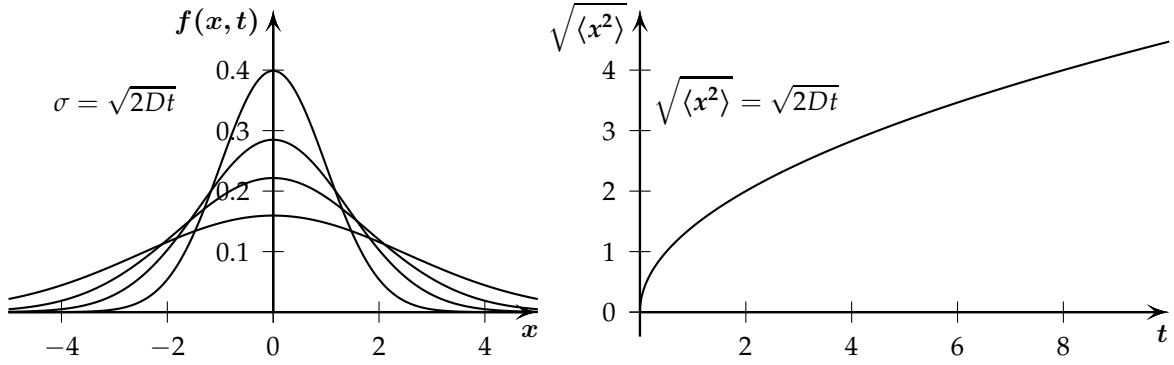


Abbildung 7.6.: *Links:* Zeitabhängige Wahrscheinlichkeitsdichte mit dem Mittelwert $\langle x \rangle = x_0 = 0$ und der Varianz $\sigma^2 = \langle x^2 \rangle = 2Dt$. Da die Varianz zeitabhängig ist, wird die Verteilung mit zunehmender Zeit immer breiter. *Rechts:* Mittleres Verschiebungsquadrat $\langle x^2 \rangle$ als Funktion der Zeit.

Die Position des Brownschen Partikels zur Zeit $t = 0$ bezeichnen wir mit x_0 . In unserer Herleitung aus dem Random Walk haben wir stets $x_0 = 0$ angenommen. Falls dies nicht der Fall ist, ersetzen wir x einfach durch $x - x_0$. Der Parameter D wird **Diffusionskoeffizient** und v wird **mittlere Driftgeschwindigkeit** genannt.

Gleichung (7.3) ist natürlich eine zeitabhängige Normalverteilung mit Mittelwert vt und Varianz $2Dt$ (siehe Abbildung 7.6). Die mittlere Verschiebung $E[X]$ des Brownschen Partikels ergibt

$$\langle x \rangle \equiv E[X] = \int_{-\infty}^{\infty} x f(x, t) dx = vt$$

Im folgenden nehmen wir einfachheitshalber an, dass die Driftgeschwindigkeit null ist. Die mittlere Verschiebung $\langle x \rangle$ ist in diesem Fall also null (oder x_0). Es macht daher mehr Sinn, das mittlere Verschiebungsquadrat $E[X^2]$ als Mass für den mittleren zurückgelegten Abstand zu betrachten:

$$\langle x^2 \rangle \equiv E[X^2] = \int_{-\infty}^{\infty} x^2 f(x, t) dx = 2Dt$$

was der Varianz σ^2 der zeitabhängigen Wahrscheinlichkeitsdichte $f(x, t)$ entspricht. Damit können wir das wichtige Ergebnis unserer Untersuchung wie folgt formulieren:

Einstein-Smoluchowski-Gleichung

Der **quadratisch gemittelte Abstand** ($\equiv \sqrt{\langle x^2 \rangle}$) eines Partikels vom Ursprungs-ort, nimmt mit der Quadratwurzel der Zeit t zu:

$$\sqrt{\langle x^2 \rangle} = \sqrt{2Dt} \quad \text{Einstein-Smoluchowski-Gleichung}$$

Der quadratisch gemittelte Abstand eines Teilchens als Funktion der Zeit t ist in Abbildung 7.6 dargestellt.

Bisher haben wir die Brownsche Bewegung nur in einer Dimension untersucht. Unser Ergebnis lässt sich aber sehr einfach auf mehrere Dimensionen übertragen. Findet die Brownsche Bewegung in zwei Dimensionen statt, so gilt für das mittlere Verschiebungsquadrat $\langle r^2 \rangle$:

$$\langle r^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle$$

Da die Brownsche Bewegung isotrop (alle Richtungen sind gleichberechtigt) ist, liefert jeder Summand den Beitrag $2Dt$ und somit ist

$$\langle r^2 \rangle = 4Dt$$

bzw. im Dreidimensionalen:

$$\langle r^2 \rangle = 6Dt$$

Der Parameter D wird als **Diffusionskoeffizient** bezeichnet und ist ein Mass für die Beweglichkeit des Partikels im umgebenden Medium. Nach Einstein und Stokes ist der Diffusionskoeffizient gegeben durch

$$D = \frac{kT}{6\pi\eta a}$$

wobei k die Boltzmann-Konstante, T die Temperatur der Flüssigkeit, η die Viskositätskonstante und a den Radius des kugelförmigen Brownschen Partikels bezeichnet. Die Einstein-Smoluchowski-Gleichung für ein Brownsches Teilchen mit Radius a in einer Flüssigkeit mit Temperatur T und Viskosität η in drei Dimensionen lautet also

$$\sqrt{\langle r^2 \rangle} = \sqrt{6Dt} = \sqrt{\frac{kTt}{\pi\eta a}}$$

Nehmen wir an, in der Flüssigkeit befinden sich n Brownsche Partikel, dann kann die zeitabhängige Teilchendichte als

$$n(x, t) = n \cdot f(x; t) = \frac{n}{\sqrt{4\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}} \quad (7.4)$$

geschrieben werden, wobei $f(x;t)$ die in Gleichung (7.3) gegebene Wahrscheinlichkeitsdichte ist. Dies ist also die Teilchendichtefunktion, die die Diffusionsgleichung löst².

Ein weiteres Beispiel von Brown'scher Bewegung ist das **thermische Rauschen** oder **Widerstandsrauschen**. Thermisches Rauschen kommt in jedem elektrischen Leiter vor und wird durch die ungeordnete Wärmebewegung der Ladungsträger hervorgerufen (Brownsche Bewegung). In einem Ohmschen Widerstand, wie er in Abbildung 7.7 gezeigt wird, tritt an den Anschlüssen durch eine zufällige Ansammlung von Elektronen sporadisch eine Rauschspannung auf, selbst wenn kein Strom durch den Leiter fließt.

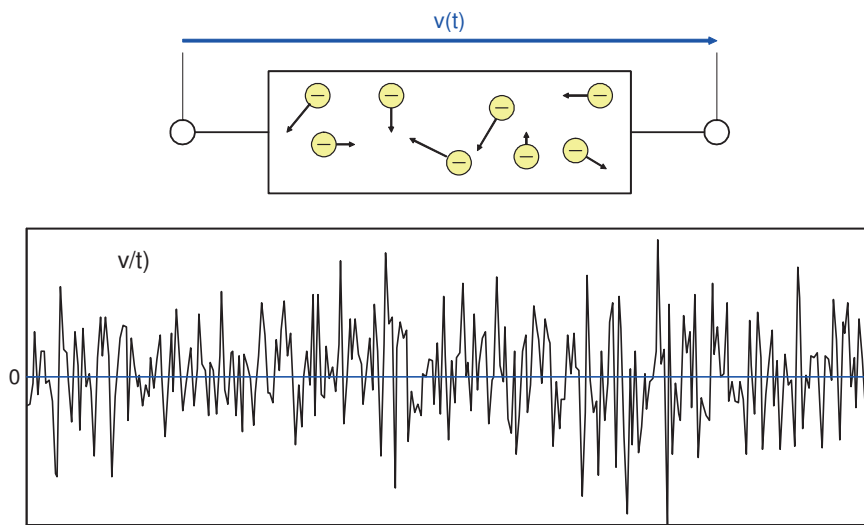


Abbildung 7.7.: (oben): Die thermische Bewegung der Elektronen in einem Widerstand erzeugt eine zufällige Rauschspannung an den Klemmen. (unten): Zufälliger oder „stochastischer“ Verlauf einer Rauschspannung über die Zeit.

Ein typischer Verlauf einer Rauschspannung $v(t)$ ist in Abbildung 7.7 aufgetragen. Die auftretenden Spannungen liegen unter üblichen Bedingungen in der Größenordnung von Mikrovolt. Wir haben es mit einem Zufallssignal (engl. random signal) oder einem „stochastischen Signal“ zu tun, da wir aus der Vergangenheit des Signals den zukünftigen Verlauf nicht vorhersagen können, und sich ein einmal aufgetretener Signalverlauf in einem wiederholten Experiment nicht wiederholen lässt. Stochastische Signale können nur mit statistischen Größen oder Mittelwerten beschrieben werden. Bei einem thermischen Rauschsignal ist der lineare Mittelwert der Spannung null, da ja im Mittel kein Strom fließen kann.

²Eine etwas ausführlichere Diskussion befindet sich in Kapitel C.1 des Anhangs.

7.2. Charakteristische Grössen von stochastischen Signalen

Ein Signal wird **stochastisch** genannt, wenn es zufällige Werte annimmt. Wir betrachten im folgenden zwei stochastische Signale $S_1(t)$ und $S_2(t)$, dargestellt in Abbildung 7.8.

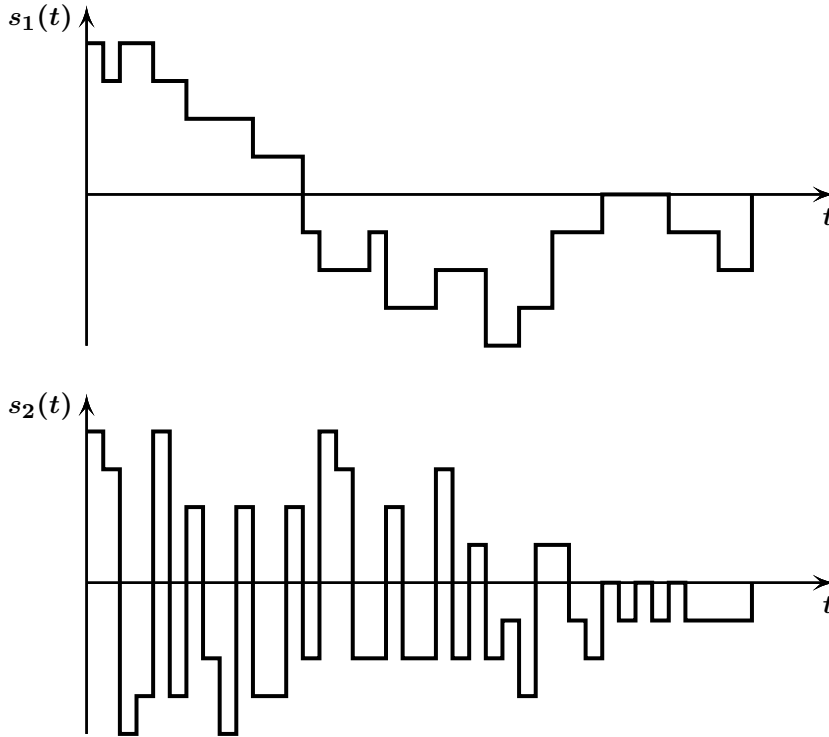


Abbildung 7.8.: Zwei diskrete stochastische Signale $S_1(t)$ und $S_2(t)$.

Der Einfachheit halber nehmen wir an, dass beide Signale nur eine endliche Anzahl von Amplitudenniveaus annehmen können, nämlich

$$\{0, \pm 1, \pm 2, \pm 3, \pm 4\}$$

Es handelt sich also bezüglich Amplitude um einen sogenannten **diskreten** Zufallsprozess. Weiterhin können die Signale nur bei Vielfachen von T ihre Amplitude ändern, d.h. dieser Zufallsprozess ist auch **zeitdiskret**.

Im Gegensatz zu deterministischen Signalen kann im Falle von stochastischen Signalen aufgrund der Kenntnis eines Signalwertes zu einer Zeit t_0 nicht mit Sicherheit auf den Signalwert zur Zeit $t_0 + \Delta T$, $\Delta T > 0$ geschlossen werden. Deswegen müssen andere Beschreibungsmethoden angewandt werden. Eine Möglichkeit besteht darin, dass man angibt, wie häufig im Mittel ein bestimmter Amplitudenwert auftritt. In

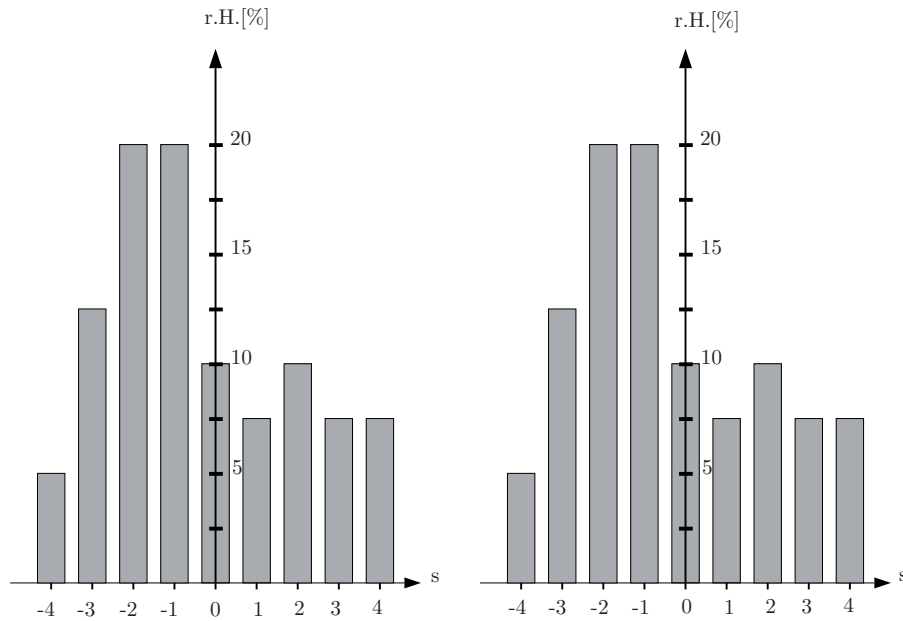


Abbildung 7.9.: Relative Häufigkeiten der Signalwerte (Wahrscheinlichkeitsverteilung) von: a) $S_1(t)$, b) $S_2(t)$.

Abbildung 7.9 sind die relativen Häufigkeiten des Auftretens der Signalwerte s von $S_1(t)$ und $S_2(t)$ dargestellt.

Stellen wir uns die Frage, wie viel Prozent der Signalwerte des Signals $S_1(t)$ liegen unterhalb eines Schwellenwertes z , so führt dies zur kumulativen Verteilungsfunktion, dargestellt in Abbildung 7.10.

In den Abbildungen 7.9 sind die relativen Häufigkeiten der Amplituden in einem endlichen Zeitintervall des Signals dargestellt. Für eine sehr grosse Anzahl Amplitudenwerte bzw. ein unbeschränktes Zeitintervall können wir diese relativen Häufigkeiten im Grenzfall als Auftrittswahrscheinlichkeiten auffassen.

Die Darstellung der Amplitudenwerte in Form eines Histogramms (siehe Abbildung 7.9) und als kumulative Verteilungsfunktion (siehe Abbildung 7.10) liefern in verschiedener Form die gleiche Information über ein Signal. Die beiden Darstellungen werden als **Wahrscheinlichkeitsverteilungen** bezeichnet und zur Beschreibung von stochastischen Signalen verwendet. Aus Abbildung 7.9 erkennen wir, dass die relativen Häufigkeiten (bzw. Auftrittswahrscheinlichkeiten) der Signalwerte für die Signale $S_1(t)$ und $S_2(t)$ gleich sind. Wenn wir allerdings die beiden Signale $S_1(t)$ und $S_2(t)$ betrachten, so erscheint anschaulich klar, dass der frequenzmässige Aufbau der Signale sehr unterschiedlich ist, da $S_1(t)$ lediglich schwach, $S_2(t)$ hingegen stark oszilliert. Mit der Angabe der Auftrittswahrscheinlichkeiten haben wir die Signale also noch nicht ausreichend charakterisiert.

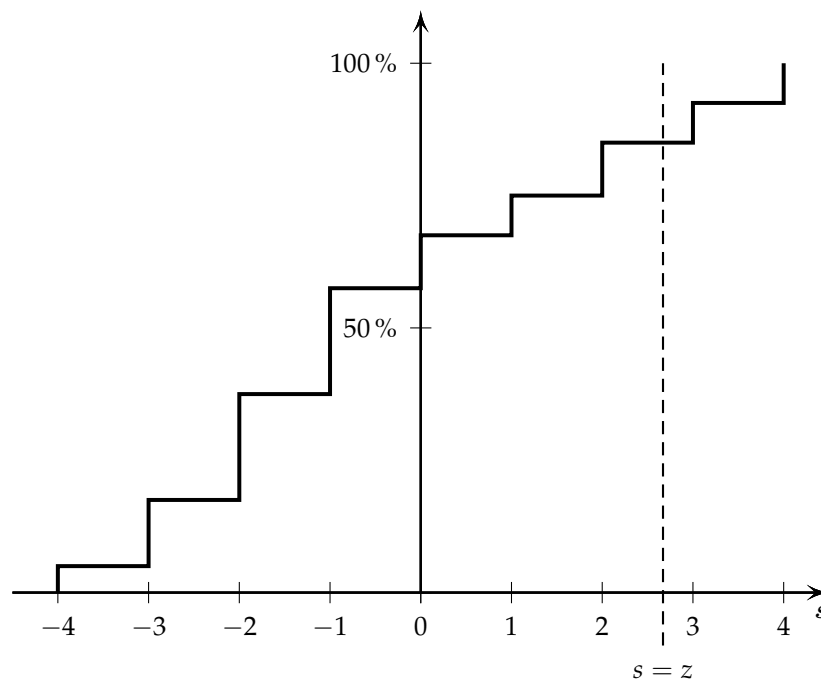


Abbildung 7.10.: Prozentualer Anteil der Signalwerte (kumulative Wahrscheinlichkeitsverteilung) von $S_1(t)$ unterhalb des Schwellenwertes z .

Es fehlt noch ein Mass für den „inneren Zusammenhang“ eines Signals, also für die Stärke der Schwankung. Zwei relativ nahe beieinanderliegende Signalwerte von $S_1(t)$ sind wahrscheinlich ungefähr gleich gross. Diese Aussage ist für das Signal $S_2(t)$ nicht richtig, d.h. benachbarte Signalwerte hängen im Signal $S_1(t)$ offenbar stärker zusammen als im Signal $S_2(t)$. Für diesen „inneren Zusammenhang“ eines Signals existiert ebenfalls ein Mass, nämlich die **Autokorrelationsfunktion**. Benachbarte Signalwerte von $S_1(t)$ sind stärker korreliert als solche von $S_2(t)$. Die Autorrelationsfunktionen werden wir im Abschnitt 7.4.1 näher betrachten.

7.3. Wahrscheinlichkeitsverteilungsfunktion und Wahrscheinlichkeitsdichtefunktion

Wir betrachten **stochastische Signale** als Musterfunktionen von Zufallsprozessen. Es gilt die folgende Definition (Abbildung 7.11):

Stochastischer Prozess

Ein **Zufallsprozess** oder stochastischer Prozess $S(t)$ ist durch ein **Ensemble** von **Musterfunktionen** $\{s_1(t), s_2(t), \dots, s_N(t)\}$ gegeben. Eine **Realisierung** ergibt sich durch die zufällige Auswahl einer Musterfunktion $s_i(t)$ mit $1 \leq i \leq N$ des Ensembles.

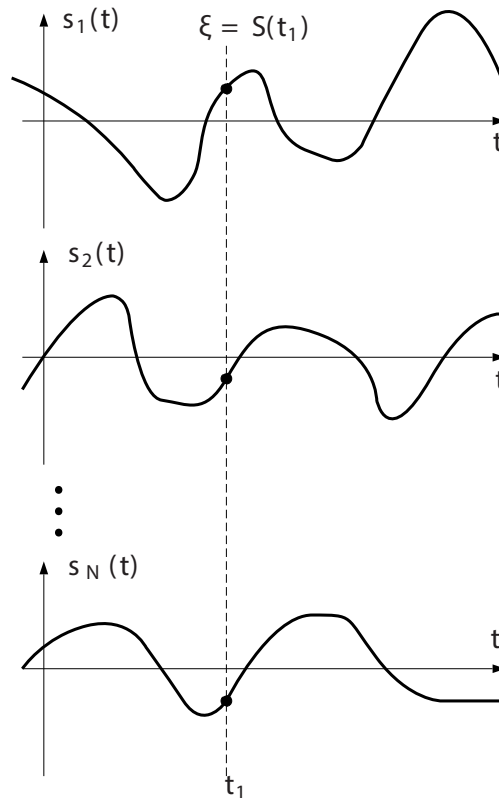


Abbildung 7.11.: Ensemble von Musterfunktionen eines kontinuierlichen Zufallsprozesses.

Der Zufallsprozess $S(t)$ kann nun mit Hilfe von Wahrscheinlichkeitsverteilungsfunktionen beschrieben werden. Dazu betrachten wir den (gemessenen) Funktionswert $s(t_1)$ des Prozesses $S(t)$ zum Zeitpunkt $t = t_1$ als Realisierung der Zufallsvariable $\xi = S(t_1)$ (Abbildung 7.11), wobei es sich hier im Gegensatz zu den in Abbildung 7.8 dargestellten Zufallsprozessen um Musterfunktionen eines **kontinuierlichen** Zufallsprozesses handelt.

Über das Ensemble der Musterfunktionen kann man die Wahrscheinlichkeit ermitteln, mit der die Werte $s(t_1)$ im Bereich $-\infty < \xi \leq s$ anzutreffen sind:

$$P(\xi \leq s) = F(s)$$

wobei $F(s)$ die kumulative Wahrscheinlichkeitsverteilungsfunktion ist. Sie ist analog zu der in Abbildung 7.10 gezeigten Treppenkurve; sie ist in diesem Fall aber eine stetige Funktion. Da ihr Funktionswert eine Wahrscheinlichkeit ist, gilt

$$0 \leq F(s) \leq 1$$

Da das Ereignis $\{\xi \leq s\}$ grösser wird, wenn s grösser wird, ist $F(s)$ monoton wachsend mit s . Die beiden Ereignisse $\xi \leq s$ und $s < \xi \leq (s + \Delta s)$ schliessen sich gegenseitig aus. Deshalb dürfen die zugehörigen Wahrscheinlichkeiten addiert werden:

$$P(\xi \leq s) + P(s < \xi \leq (s + \Delta s)) = P(\xi \leq (s + \Delta s))$$

Damit gilt auch:

$$P(s < \xi \leq (s + \Delta s)) = F(s + \Delta s) - F(s) \geq 0$$

Dividieren wir obigen Ausdruck durch Δs , so führt Grenzwertbildung auf die Definition der **Wahrscheinlichkeitsdichtefunktion** $f(s)$:

$$f(s) := \lim_{\Delta s \rightarrow 0} \frac{P(s < \xi \leq (s + \Delta s))}{\Delta s} = \frac{dF(s)}{ds} \geq 0 \quad (7.5)$$

Man beachte ferner, dass $f(s)$ ggf. dimensionsbehaftet ist mit $[f(s)] = s^{-1}$, also ist z.B. $[f(s)] = \text{Sekunde}^{-1}$, sofern s die Zeit gemessen in Sekunden ist. Die Dichtefunktion hat die wichtige Eigenschaft

$$\int_{-\infty}^{\infty} f(s) ds = 1$$

Sind die zu einem stochastischen Signal gehörenden Funktionen $f(s)$ bzw. $F(s)$ bekannt, so lassen sich systemtechnisch wichtige Grössen ermitteln. Zum Beispiel kann bei bekannter Wahrscheinlichkeitsdichte $f(s)$ der Werte $\{s(t_1)\}$ ausgesagt werden, dass sich diese mit der Wahrscheinlichkeit

$$P_{ab} = \int_a^b f(s) ds$$

im Bereich $a \leq s(t_1) \leq b$ aufhalten. Anschaulich ist also die Fläche unter der Kurve $f(s)$ von $s = a$ bis $s = b$ gleich der Wahrscheinlichkeit, mit der die Zufallsvariable ξ einen Wert zwischen a und b annimmt (siehe Abbildung 7.12).

Bis anhin wurde lediglich der Funktionswert eines stochastischen Signals zum Zeitpunkt $t = t_1$ betrachtet. Allgemein sind Wahrscheinlichkeitsverteilungsfunktion und Wahrscheinlichkeitsdichtefunktion zeitabhängig, d.h. man hat $F(s, t)$ und $f(s, t)$. Bei vielen technisch relevanten Zufallsprozessen sind die statistischen Eigenschaften unabhängig von der Zeit. Diese Eigenschaft nennt man **Stationarität**:

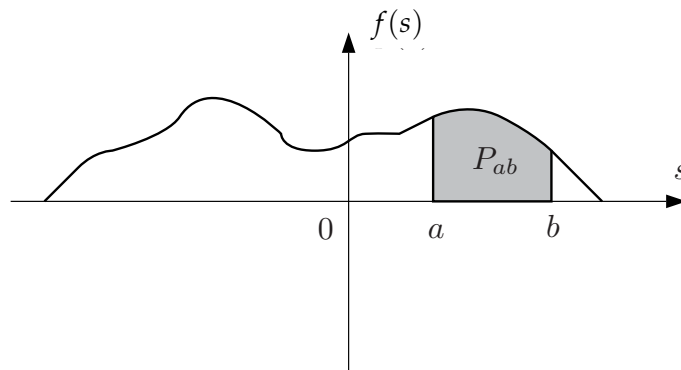


Abbildung 7.12.: Interpretation der Wahrscheinlichkeitsdichtefunktion.

Stationärer stochastischer Prozess

Ein Zufallsprozess heisst **stationär**, wenn die kumulative Wahrscheinlichkeitsverteilungsfunktion oder die Wahrscheinlichkeitsdichtefunktion zeitunabhängig sind:

$$\begin{aligned} F(s, t) &= F(s) \\ f(s, t) &= f(s) \end{aligned}$$

Bemerkungen:

- i. Diese Definition von Stationarität wird auch als starke Stationarität bezeichnet. Ein stochastischer Prozess $X(t)$ heisst **stark stationär**, wenn die Verteilung von $X(t + s)$ nicht von der Verschiebung s abhängt.
- ii. Ein stochastischer Prozess heisst **schwach stationär**, wenn der Erwartungswert konstant ist und die Varianz endlich ist.

Beispiel 7.3.1

Ein Beispiel für einen stationären Zufallsprozess ist das **thermische Rauschen** $V(t)$ in einem elektrischen Widerstand, siehe Abbildung 7.7. Betrachten wir die Verteilung der Rauschamplituden $V(t)$, z.B. die Werte der Spannung v aus Abbildung 7.7, dann stellen wir erstens fest, dass sich die Verteilung zeitlich nicht ändert und zweitens, dass sie einer Normalverteilung folgt:

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(v-\mu)^2}{2\sigma^2}}$$

wobei μ den Erwartungswert und σ^2 die Varianz bzw. σ die Standardabweichung der Zufallsvariablen $V(t) = V$ (Rauschamplitude) bezeichnet. Würden wir die Verteilung der Rauschamplituden in einem elektrischen Widerstand in einem Jahr mit der Verteilung der momentanen Rauschamplituden vergleichen, würden wir keinen Unterschied feststellen können.

Anders verhält es sich bei einem **Diffusionsprozess**: in diesem Fall ändert sich die Wahrscheinlichkeitsdichte für die Position x eines Teilchens in Abhängigkeit der Zeit t gemäss

$$f(x;t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{(x-\mu t)^2}{4Dt}}$$

unter der Annahme, dass es sich zum Zeitpunkt $t = 0$ an der Stelle $x = 0$ befand. In diesem Fall haben wir also einen nicht-stationären stochastischen Prozess.

□

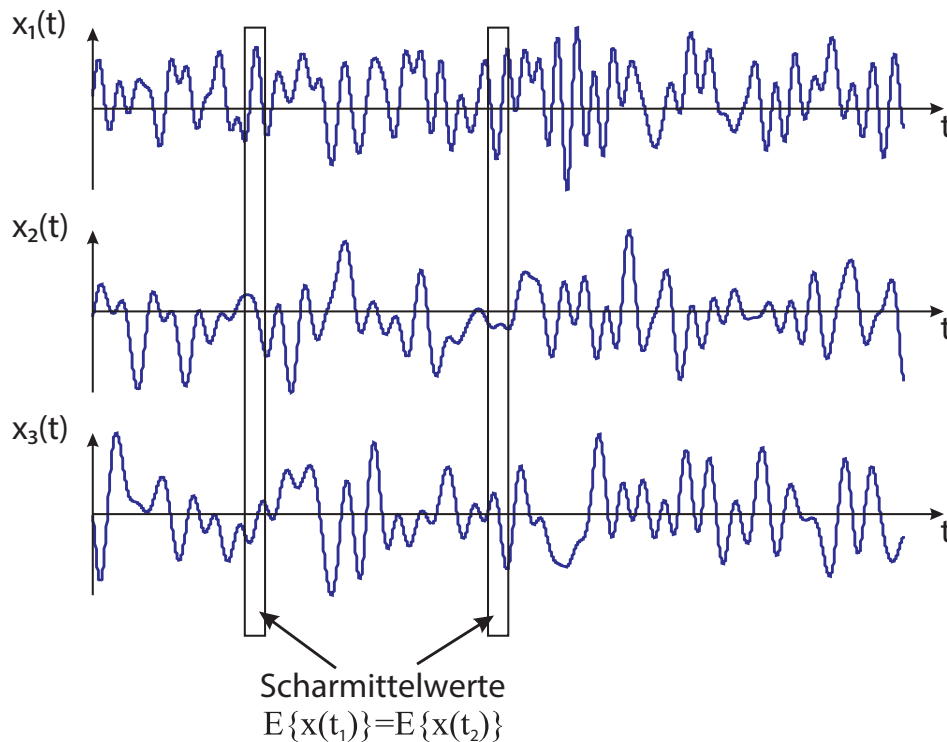


Abbildung 7.13.: Bestimmung des Erwartungswertes über eine Schar von Musterfunktionen $\{x_1(t), x_2(t), \dots, x_N(t)\}$ eines stationären stochastischen Prozesses $X(t)$. In diesem Fall sind die Scharmittelwerte $\mu_X(t) = E[X(t)]$ zeitunabhängig.

Für eine beliebige Wahrscheinlichkeitsdichtefunktion $f(s, t)$ berechnen sich Erwartungswert und Varianz gemäss

$$\mu(t) = \int_{-\infty}^{\infty} s \cdot f(s, t) \, ds \quad (7.6)$$

$$\sigma^2(t) = \int_{-\infty}^{\infty} (s - \mu)^2 f(s, t) \, ds \quad (7.7)$$

Die Berechnungen gemäss (7.6) und (7.7) können als Mittelwertbildung über das Ensemble $\{s_1(t), s_2(t), \dots, s_N(t)\}$ bzw. über die Schar des Zufallsprozesses $S(t)$ angesehen werden. Ist der Zufallsprozess $S(t)$ stationär, so sind die Scharmittelwerte $\mu_S(t) = E[S(t)]$ zeitunabhängig (siehe Abbildung 7.13). Man bezeichnet $\mu_S(t)$ und $\sigma_S^2(t)$ deshalb auch als *Scharmittelwert*, resp. *Scharvarianz*. Sie können berechnet werden, wenn eine mathematische Beschreibung der Wahrscheinlichkeitsdichtefunktion $f(s, t)$ vorliegt. In der Praxis wird jedoch oft nur eine einzige Realisierung des Prozesses beobachtet, und man kennt die Wahrscheinlichkeitsdichtefunktion nicht. In diesem Fall lassen sich μ_S und σ_S^2 nur bestimmen, wenn die betrachteten Zufallsprozesse eine weitere Regularität aufweisen, die als *Ergodizität* bezeichnet wird (siehe Abbildung 7.15). Dazu definieren wir zuerst den Zeitmittelwert:

Zeitmittelwert

Für einen stochastischen Prozess $S(t)$ definieren wir den Zeitmittelwert

$$\langle s(t) \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} s(t) \, dt$$

Geometrisch können wir den Zeitmittelwert wie folgt interpretieren: Die Fläche unter der Kurve des Stochastischen Signals $S(t) > 0$ dividieren wir durch die Länge der Integrationszeit. D. h. wir erhalten die Höhe des Rechtecks mit der gleichen Fläche wie die Fläche unter der Kurve des Stochastischen Prozesses $S(t)$ (siehe Abbildung 7.14).

Beispiel 7.3.2

Sei A eine auf dem Intervall $[0, 1]$ uniform verteilte Zufallsvariable. Wir definieren den stochastischen Prozess

$$S(t) = A$$

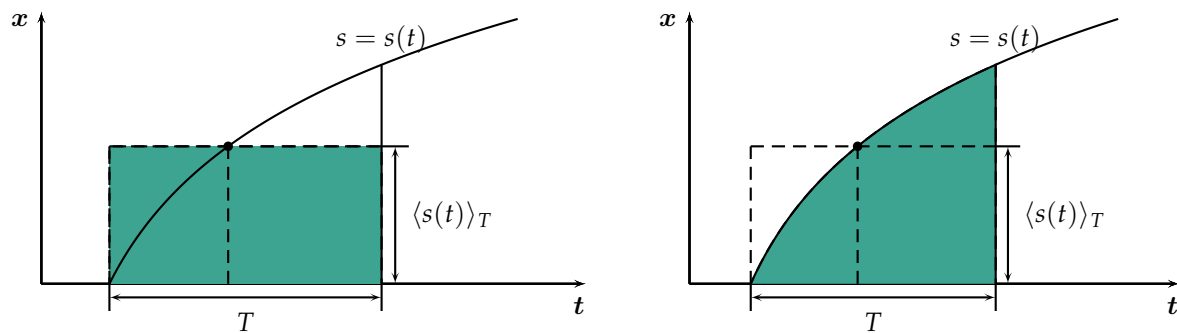


Abbildung 7.14.: Gleich grosse Flächen

Eine Realisierung dieses stochastischen Prozesses besteht dann darin, dass eine auf dem Intervall $[0, 1]$ uniform verteilte Zufallszahl „gezogen“ wird und zu jedem Zeitpunkt diesen Wert besitzt. Es handelt sich also um einen *stationären* stochastischen Prozess. Der Zeitmittelwert ist dann gegeben durch

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} A \, dt = A \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} dt = A$$

Da A gerade die Höhe des Rechtecks ist, und A eine Zufallsvariable ist, ist der Zeitmittelwert für jede Realisierung des Zufallsprozesses $S(t)$ verschieden.

□

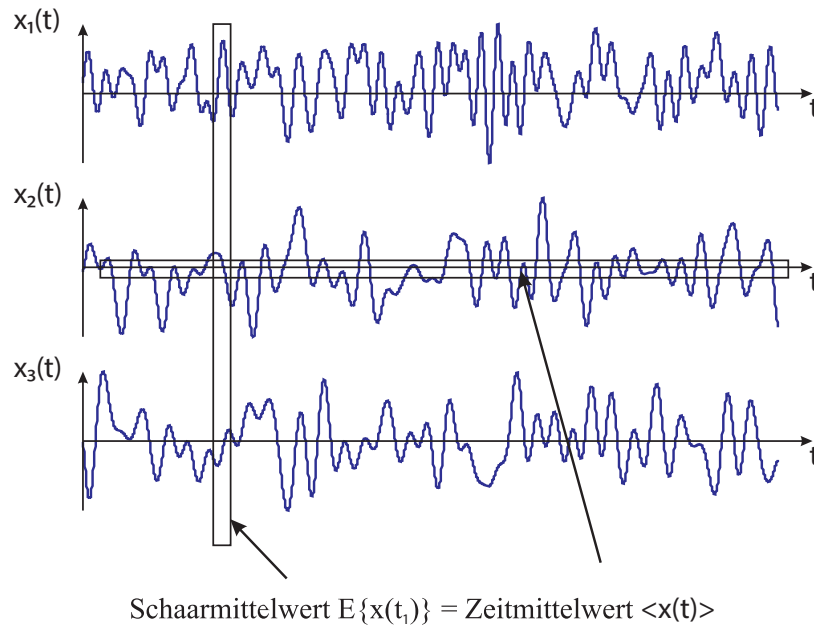


Abbildung 7.15.: Falls die Scharmittelwerte bei einem stationären stochastischen Prozess $X(t)$ mit den Zeitmittelwerten übereinstimmen, handelt es sich um einen **ergodischen** Prozess.

Ergodischer stochastischer Prozess

Ein stationärer Zufallsprozess $S(t)$ heisst **ergodisch**, wenn Zeit- und Scharmittelwerte übereinstimmen: d. h. für **jede** Realisierung $s(t)$ des Zufallsprozesses $S(t)$ muss gelten

$$\mu_S = \int_{s=-\infty}^{\infty} s \cdot f(s) \, ds \stackrel{!}{=} \langle s(t) \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} s(t) \, dt = m_s$$

$$\sigma_S^2 = \int_{s=-\infty}^{\infty} (s - \mu_S)^2 \cdot f(s) \, ds \stackrel{!}{=} \langle (s(t) - m_s)^2 \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} (s(t) - m_s)^2 \, dt$$

wobei $\langle \cdot \rangle$ die zeitliche Mittelwertbildung bezeichnet.

Beispiel 7.3.3

Sei A eine auf dem Intervall $[0, 1]$ uniform verteilte Zufallsvariable. Wir definieren den Stochastischen Prozess

$$S(t) = A$$

Den Zeitmittelwert haben wir bereits zu $\langle S(t) \rangle_T = A$ berechnet. Für den Scharmittelwert erhalten wir

$$\mu_S(t) = E[S(t)] = E[A] = \frac{1-0}{2} = \frac{1}{2}$$

Da Scharmittelwert und Zeitmittelwert nicht identisch sind, ist der Prozess $S(t)$ nicht ergodisch.

□

Beispiel 7.3.4

Wir betrachten den stochastischen Prozess $X(t)$, der gegeben ist durch

$$X(t) = A \cos(\omega t + \Theta)$$

wobei A und ω konstant sind und Θ eine gleichmässig verteilte Zufallsvariable über das Intervall $[-\pi, \pi]$ ist. Der Mittelwert $\mu(t)$ berechnet sich wie folgt

$$\mu(t) = E(X(t)) = \int_{-\infty}^{\infty} A \cos(\omega t + \theta) f_{\Theta}(\theta) d\theta$$

wobei die Dichtefunktion gegeben ist durch

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi} & -\pi \leq \theta \leq \pi \\ 0 & \text{sonst} \end{cases}$$

Somit ergeben sich für den Mittelwert

$$\mu(t) = \frac{A}{2\pi} \int_{-\pi}^{\pi} \cos(\omega t + \theta) d\theta = 0$$

und für die Varianz

$$\sigma(t)^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} (A \cos(\omega t + \theta) - 0)^2 d\theta = \frac{A^2}{2}$$

Da der Mittelwert und die Varianz des betrachteten stochastischen Prozesses $X(t)$ unabhängig von der Zeit t sind, handelt es sich um einen *stationären* Zufallsprozesse.

Der zeitliche Mittelwert von $X(t)$ ergibt

$$\langle x(t) \rangle_{T_0} = \frac{A}{T_0} \int_{-T_0/2}^{T_0/2} \cos(\omega t + \theta) dt = 0$$

wobei $T_0 = 2\pi/\omega$. Eine ähnliche Rechnung ergibt für die Varianz

$$\langle \sigma^2(t) \rangle_{T_0} = \frac{A^2}{2}$$

woraus wir schliessen, dass $X(t)$ nicht nur stationär, sondern auch *ergodisch* ist (Scharmittelwerte sind identisch mit den zeitlichen Mittelwerten). Da $X(t)$ periodisch ist, brauchen wir bei der Berechnung des zeitlichen Mittelwertes bloss über eine Periode zu integrieren.

□

Beispiel 7.3.5

Die Maxwell-Boltzmann-Verteilung ist eine Wahrscheinlichkeitsverteilung der statistischen Physik und spielt in der kinetischen Gastheorie eine wichtige Rolle. Sie beschreibt die statistische Verteilung des Betrags $v = |\vec{v}|$ der Teilchengeschwindigkeiten in einem idealen Gas und ist gegeben durch

$$f(v) = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} v^2 \exp \left(-\frac{mv^2}{2k_B T} \right)$$

wobei k_B die Boltzmann-Konstante und T die Temperatur in Kelvin und m die Masse des Gasmoleküls bezeichnet. Die Geschwindigkeitsverteilung ist dargestellt in der Abbildung 7.16.

Die mittlere Geschwindigkeit der Gasmoleküle ergibt sich zu

$$\mu_v = \int_0^\infty v f(v) dv = \sqrt{\frac{8k_B T}{\pi m}}$$

Die für die statistische Mechanik grundlegende **Ergodenhypothese** besagt, dass der Zeitmittelwert gleich dem Scharmittelwert einer Messgrösse ist. Im Falle der sich zeitlich ändernden Geschwindigkeit $v(t)$ eines Gasmoleküls ergibt der Zeitmittelwert:

$$\langle v(t) \rangle_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{\frac{T}{2}} v(t) dt \stackrel{!}{=} \mu_v = \sqrt{\frac{8k_B T}{\pi m}}$$

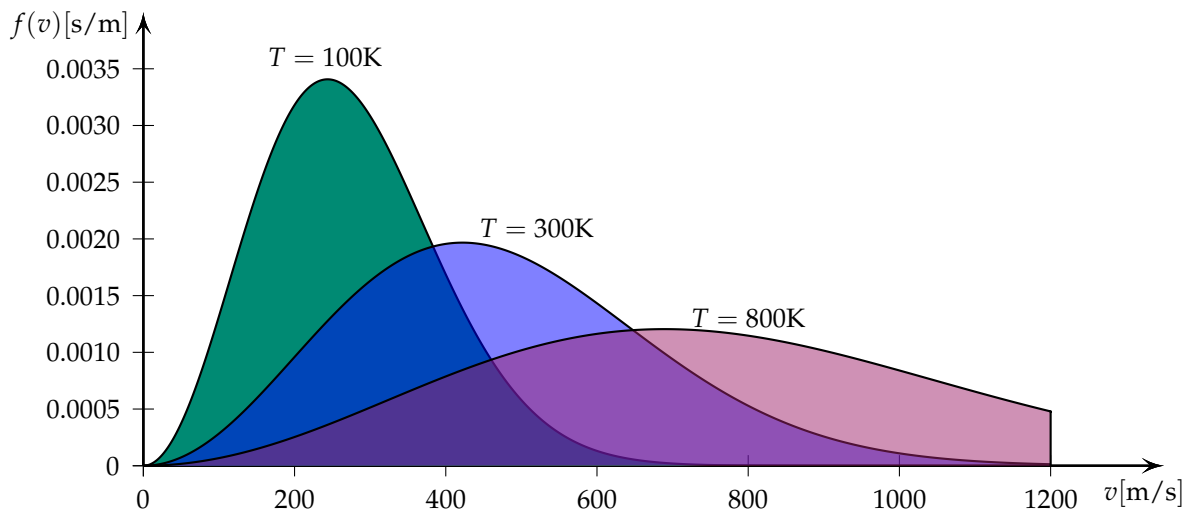


Abbildung 7.16.: Maxwell-Boltzmann Geschwindigkeitsverteilung (Geschwindigkeitsbetrag) für das zweiatomige Stickstoffmolekül (N_2) für drei verschiedene Temperaturen: $T = 100\text{ K}$ (grün), $T = 300\text{ K}$ (blau) und $T = 800\text{ K}$ (violett).

Um die zeitlich gemittelte Geschwindigkeit eines Gasmoleküls zu bestimmen, brauchen wir also bloss über die Geschwindigkeiten aller in einem idealen Gas befindlichen Gasmoleküle zu einem beliebigen Zeitpunkt zu mitteln.

□

7.4. Autokorrelationsfunktion und Kreuzkorrelationsfunktion

7.4.1. Die Autokorrelationsfunktion

Im Unterkapitel 7.3 wurden Aussagen darüber gemacht, wie häufig gewisse Werte eines Signals auftreten, ohne dass in irgendeiner Weise der zeitliche Verlauf des Signals berücksichtigt wurde. Betrachten wir nochmal die beiden stochastischen Signale, die in Abbildung 7.8 dargestellt sind. Wie wir gesehen haben, besitzen die beiden Signale identische Verteilungen der Amplitudenwerte. Allerdings sind die Schwankungen innerhalb eines Zeitintervalls bei Signal $S_1(t)$ viel kleiner als bei Signal $S_2(t)$. Insbesondere sind zwei relativ nahe beieinanderliegende Signalwerte von $S_1(t)$ etwa

gleich gross, während dies für das Signal $S_2(t)$ nicht zutrifft. Benachbarte Signalwerte hängen im Signal $S_1(t)$ offenbar stärker zusammen als im Signal $S_2(t)$. Diesen „inneren Zusammenhang“ messen wir mit Hilfe der Korrelation von Signalausschnitten, die um eine Zeit τ gegeneinander verschoben sind. Da diese Ähnlichkeit grundsätzlich für jede Verschiebung τ existiert, lässt sich die Korrelation durch eine Funktion von τ beschreiben, die wir mit $\varphi(t, t + \tau)$ notieren und als **Autokorrelationsfunktion** bezeichnen.

Autokorrelationsfunktion

In der Signaltheorie^a ist die **Autokorrelationsfunktion** $\varphi(t, t + \tau)$ eines stochastischen Prozesses $S(t)$ definiert als der Erwartungswert des Produkts der beiden Zufallsvariablen $S(t)$ und $S(t + \tau)$

$$\varphi(t, t + \tau) = E[S(t)S(t + \tau)]$$

^aIn der Stochastik ist die **Autokorrelationsfunktion** anders als in der Signaltheorie definiert, nämlich als normierte Form der Autokovarianzfunktion:

$$\varphi(t, t + \tau) = \frac{E[(S(t) - \mu(t))(S(t + \tau) - \mu(t + \tau))]}{\sigma(t) \cdot \sigma(t + \tau)}.$$

Für eine Motivierung dieser Definitionen aus der Signaltheorie, siehe Anhang C.2.

Bemerkungen:

- i. Der Wert der Autokorrelationsfunktion $\varphi(t, t + \tau)$ entspricht also dem Produkt eines Signalwertes $s(t)$ (einer Musterfunktion oder Realisierung von $S(t)$) zum Zeitpunkt t mit dem Signalwert zum Zeitpunkt $t + \tau$. Danach wird der Mittelwert über alle Musterfunktionen gebildet.
- ii. Die Autokorrelationsfunktion gibt an, wie viel Ähnlichkeit das um die Zeit τ verschobene Signal $s(t + \tau)$ mit dem ursprünglichen Signal $s(t)$ gemeinsam hat. Da die unverschobene Folge mit sich selbst am ähnlichsten ist, hat die Autokorrelation für die unverschobene Folge ($\tau = 0$) den höchsten Wert.
- iii. Handelt es sich bei $S(t)$ um einen **stationären** stochastischen Prozess, ist die Autokorrelationsfunktion unabhängig von t , so dass $\varphi(t, t + \tau) = \varphi(\tau)$.
- iv. Die Autokorrelationsfunktion entspricht der Autokovarianzfunktion - beide gemäss den Definitionen in der Signaltheorie -

$$\varphi(t_1, t_2) = E[(X(t_1) - \mu_{t_1})(X(t_2) - \mu_{t_2})]$$

für mittelwertfreie (Erwartungswert gleich null), stationäre Signale $X(t)$.

Beispiel 7.4.1

Wir betrachten den stochastischen Prozess $X(t)$, der gegeben ist durch

$$X(t) = A \cos(\omega t + \Theta),$$

wobei A und ω konstant sind und Θ eine gleichmässig verteilte Zufallsvariable über das Intervall $[-\pi, \pi]$ ist. Die Autokorrelationsfunktion für den stochastischen Prozess $X(t)$ ergibt also

$$\begin{aligned} \varphi(\tau) &= E[A \cos(\omega t + \theta) A \cos(\omega(t + \tau) + \theta)] \\ &= \int_{-\pi}^{\pi} A^2 \cos(\omega t + \theta) \cos(\omega(t + \tau) + \theta) \frac{1}{2\pi} d\theta \\ &= \frac{A^2}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} [\cos \omega \tau + \cos(2\omega t + 2\theta + \omega \tau)] d\theta \\ &= \frac{A^2}{4\pi} \int_{-\pi}^{\pi} [\cos \omega \tau + \cos(2\omega t + 2\theta + \omega \tau)] d\theta \\ &= \frac{A^2}{2} \cos(\omega \tau). \end{aligned}$$

□

Für **ergodische stationäre** Zufallsprozesse gilt

Autokorrelationsfunktion für ergodischen stationären stochastischen Prozess

$$\varphi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t + \tau) dt. \quad (7.8)$$

Bemerkungen:

- i. Man beachte, dass ein ergodischer Prozess $\varphi(\tau)$ einen deterministischen Wert hat, obwohl er über eine Realisierung definiert ist.

- ii. Die innere Übereinstimmung ist für $\tau = 0$ sicher am grössten. Das Maximum von $\varphi(\tau)$ wird **normierte Leistung** des Signals genannt

$$\mathcal{S} := \langle s^2(t) \rangle = \varphi(0).$$

- iii. Aus Gleichung (7.8) folgt sofort, dass die Autokorrelationsfunktion eine gerade Funktion ist:

$$\varphi(\tau) = \varphi(-\tau).$$

Für $\tau > 0$ wird der Wert der Autokorrelationsfunktion kleiner als $\varphi(0)$. Wenn $S(t)$ keinen konstanten oder periodischen Anteil enthält, so geht $\varphi(\tau)$ für $|\tau| \rightarrow \infty$ gegen null, da sich in diesem Fall die positiven und negativen Anteile bei der Integration infolge der Unabhängigkeit von $S(t)$ und $S(t + \tau)$ aufheben:

$$\lim_{\tau \rightarrow \pm\infty} \varphi(\tau) = 0.$$

In der Praxis muss man sich auf begrenzte Zeitintervalle beschränken. Die derart gebildeten Funktionen heissen **Kurzzeit-Autokorrelationsfunktionen** $\varphi_T(\tau)$:

Kurzzeit-Autokorrelationsfunktion

$$\varphi_T(\tau) := \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t)s(t + \tau) dt.$$

Jede Realisierung $s(t)$ eines ergodischen stochastischen Prozesses $S(t)$ wird einen etwas anderen Verlauf nehmen, d. h. die Grösse $\varphi_T(\tau)$ ist selbst eine Zufallsvariable, deren Werte um die exakten Werte von $\varphi(\tau)$ schwanken. Wenn jedoch das Intervall T gross genug gewählt wird, gilt

$$\varphi_T(\tau) \approx \varphi(\tau).$$

Für die Autokorrelationsfunktion eines periodischen Signals mit der Periode T_0 gilt

$$\varphi(\tau) := \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} s(t)s(t + \tau) dt. \quad (7.9)$$

Man beachte, dass im Falle eines periodischen Signals $\varphi(\tau + T_0) = \varphi(\tau)$ wegen $S(t + T_0) = S(t)$.

Beispiel 7.4.2

Wir betrachten den stochastischen Prozess $X(t)$, der gegeben ist durch

$$X(t) = A \cos(\omega t + \Theta),$$

wobei A und ω konstant sind und Θ eine gleichmässig verteilte Zufallsvariable über das Intervall $[-\pi, \pi]$ ist. Wie wir bereits festgestellt haben, handelt es sich bei $X(t)$ um einen ergodischen Prozess. Die Autokorrelationsfunktion für den stochastischen Prozess $X(t)$ ergibt also

$$\begin{aligned} \varphi(\tau) &= \langle x(t)x(t+\tau) \rangle_{T_0} \\ &= \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} A^2 \cos(\omega t + \theta) \cos[\omega(t+\tau) + \theta] dt \\ &= \frac{A^2}{T_0} \int_{-T_0/2}^{T_0/2} \frac{1}{2} [\cos \omega \tau + \cos(2\omega t + 2\theta + \omega \tau)] dt \\ &= \frac{A^2}{2} \cos(\omega \tau). \end{aligned}$$

Mit der Definition $\varphi(\tau) = E[X(t)X(t+\tau)]$ ergab sich derselbe Wert.

□

Wollen wir in der Praxis die Autokorrelationsfunktion eines stationären stochastischen Prozesses $S(t)$ aus einer Realisation mit diskreten Zeitwerten $s(t_1), s(t_2), \dots, s(t_n)$ schätzen, so ist

$$\hat{\varphi}(h) = \frac{1}{n} \sum_{i=1}^{n-h} s(t_i) \cdot s(t_{i+h}).$$

Mit der Software R wird die Autokorrelationsfunktion folgendermassen geschätzt:

$$\hat{\varphi}(h) = \frac{\frac{1}{n} \sum_{i=1}^{n-h} (s(t_i) - \hat{\mu}) \cdot (s(t_{i+h}) - \hat{\mu})}{\frac{1}{n} \sum_{i=1}^n (s(t_i) - \hat{\mu})^2},$$

wobei $\hat{\mu} = \frac{1}{n} \sum s(t_i)$ den geschätzten Scharmittelwert bezeichnet. Die (normierte) Autokorrelationsfunktion kann in einem sogenannten **Korrelogramm** aufgezeichnet werden.

Beispiel 7.4.3

Von 1945 bis 1974 wurden vierteljährlich die Werte ermittelt, welche die Zustimmung für den Präsidenten der USA ausdrücken. Die Zeitreihe ist in Abbildung 7.17 dargestellt.

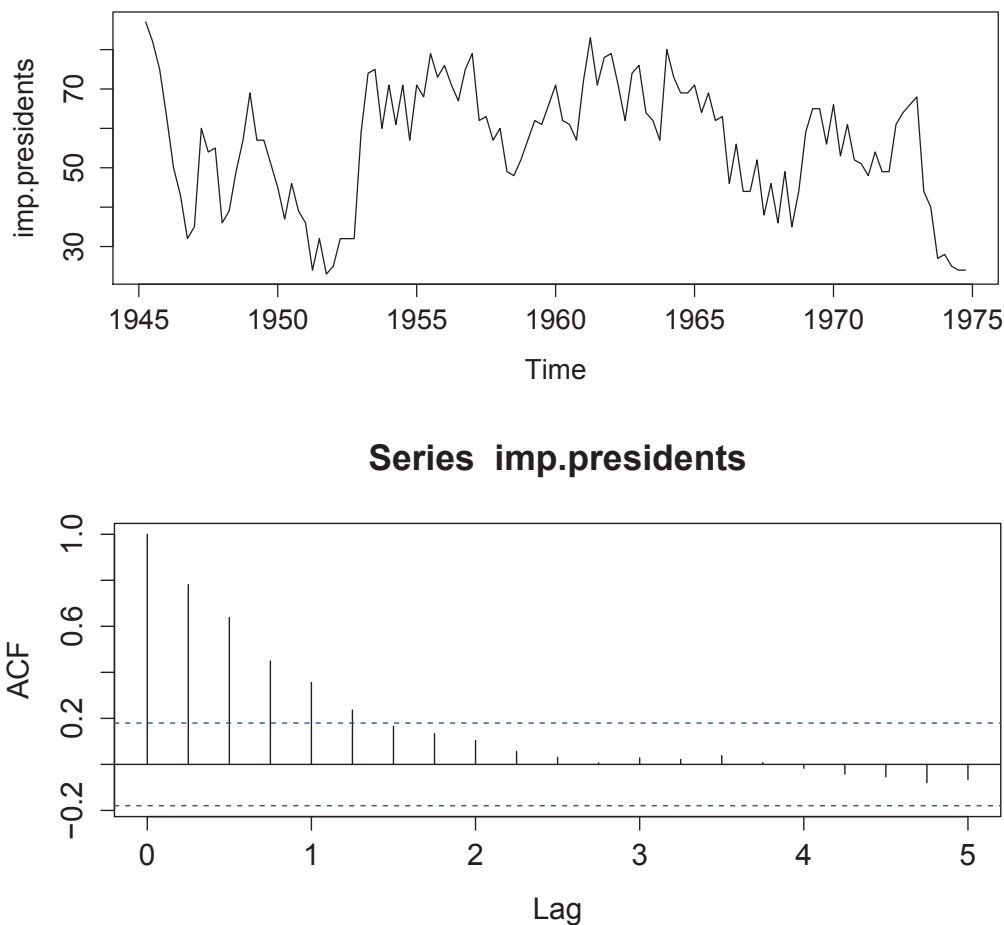
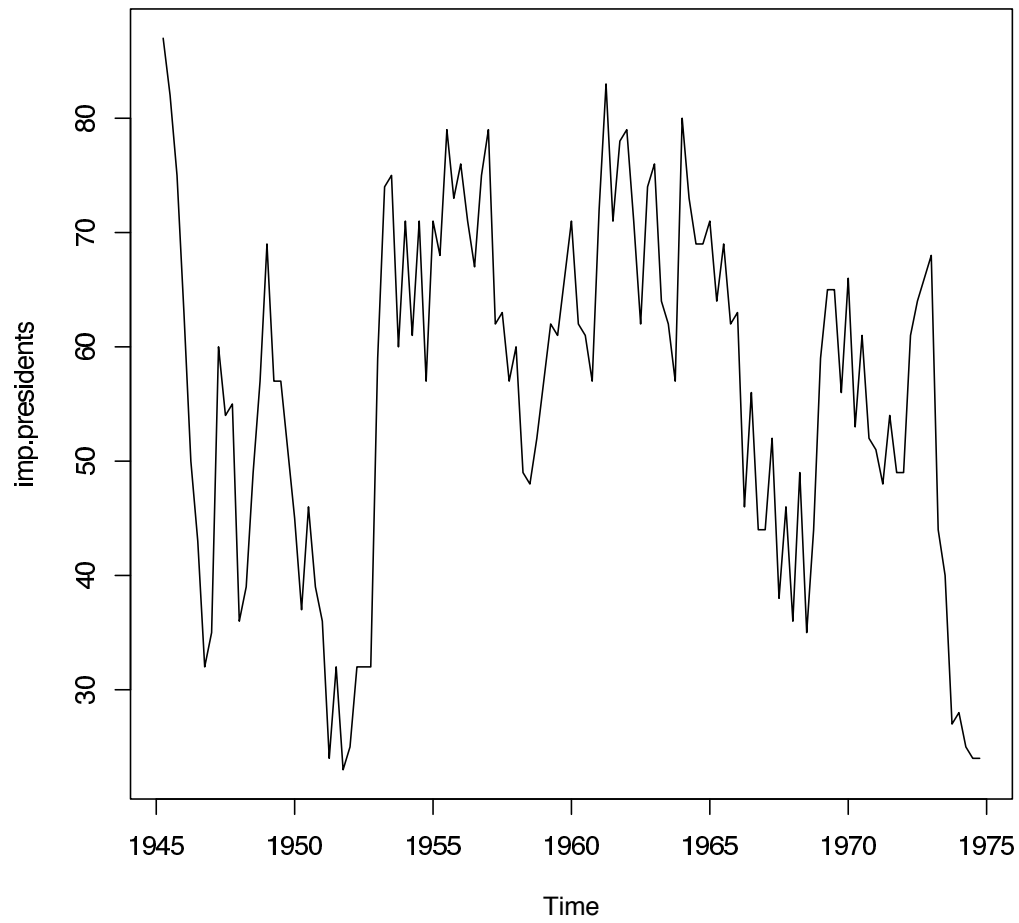


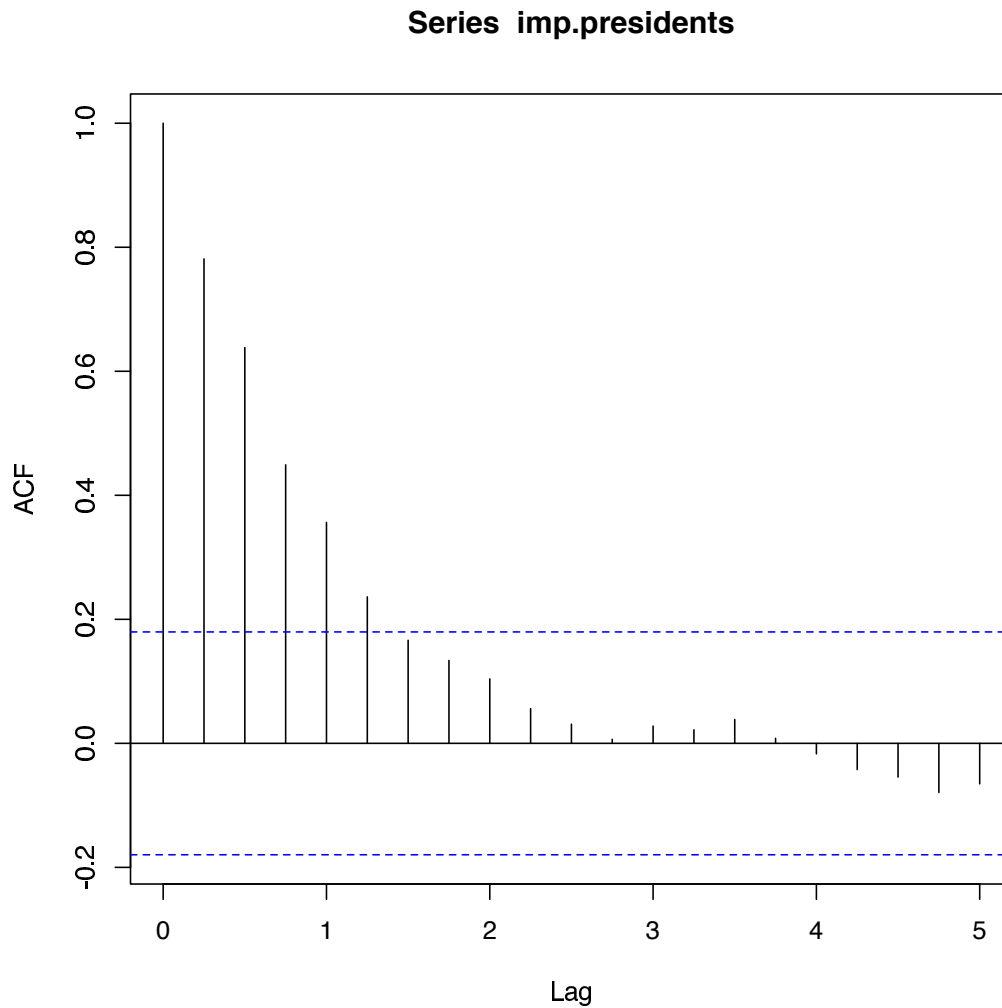
Abbildung 7.17.: (*oben:*) Die Zustimmung (in Prozent) der Präsidenten in den USA ist als Zeitreihe dargestellt und (*unten:*) das Korrelogramm der zugehörigen empirischen Autokorrelationsfunktion $\varphi(\tau)$, wobei τ mit Lag bezeichnet ist. Die Einheit von τ entspricht in diesem Fall einer Zeitperiode von einem Vierteljahr.

Aus dem Korrelogramm in Abbildung 7.17 entnehmen wir, dass es tatsächlich eine Korrelation innerhalb eines Zeitraumes von etwa einem Jahr zwischen den Zustimmungen für den Präsidenten gibt. Das obige Korrelogramm kann mit R folgendermassen erstellt werden (mehr zu Korrelogrammen in Kapitel 7.6):

```
pos.na <- which(is.na(presidents))
x.time <- time(presidents)
x <- x.time[-pos.na]
y <- presidents[-pos.na]
xy.smooth <- smooth.spline(x, y)
pred.times <- x.time[pos.na]
pred.pres <- predict(xy.smooth, pred.times)
imp.presidents <- presidents
imp.presidents[pos.na] <- round(pred.pres$y)
imp.presidents <- window(imp.presidents, start = c(1945,
  2))
plot(imp.presidents)
```



```
acf(imp.presidents)
```

□

7.4.2. Die Kreuzkorrelationsfunktion

Nachdem in Abschnitt 7.4.1 die Korrelationsfunktion eines Signals $S(t)$ mit sich selbst gebildet wurde, kann diese auch für zwei verschiedene Signale $S_1(t)$ und $S_2(t)$ berechnet werden. In diesem Fall spricht man von der **Kreuzkorrelationsfunktion** $\varphi_{12}(\tau)$ der beiden Signale.

Kreuzkorrelationsfunktion

Für stationäre und ergodische Zufallsprozesse ist die **Kreuzkorrelationsfunktio-**

on wie folgt definiert:

$$\varphi_{12} := \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s_1(t) s_2(t + \tau) dt. \quad (7.10)$$

Die Kreuzkorrelationsfunktion stellt ein Mass für die Ähnlichkeit oder Korrelation zweier verschiedener Signale dar. Es ist natürlich möglich, anstelle von (7.10) die Funktion φ_{21} zu bilden:

$$\varphi_{21}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s_2(t) s_1(t + \tau) dt.$$

Man kann leicht zeigen, dass gilt

$$\varphi_{12}(\tau) = \varphi_{21}(-\tau).$$

Man beachte jedoch, dass die Kreuzkorrelationsfunktion im Gegensatz zur Autokorrelationsfunktion im allgemeinen weder eine gerade noch eine ungerade Funktion von τ ist und das Maximum auch nicht notwendigerweise bei $\tau = 0$ liegt. Analog zur Autokorrelationsfunktion existieren auch für die Kreuzkorrelationsfunktion die modifizierten Definitionen für periodische Signale und Signale mit endlicher Energie sowie die Kurzzeit-Kreuzkorrelationsfunktionen.

7.5. Detektion eines deterministischen Signals in einem verrauschten Signal

Im Folgenden wollen wir ein Beispiel betrachten, wie mit Hilfe von Auto- bzw. Kreuzkorrelationsfunktionen ein stark verrauschtes Signal verarbeitet werden kann. Es soll entschieden werden, ob in einem empfangenen Signal $S(t)$ ein Rechtecksignal $r(t)$ mit der Periode T_0 enthalten ist oder nicht. Die Abbildung 7.18 links zeigt das (deterministische) Rechtecksignal $r(t)$. Unser Ziel ist es, eine Methode zu entwickeln, damit wir am Eingang eines Detektors das Signal

$$S(t) = r(t) + N(t)$$

von einem Rauschsignal

$$N(t),$$

unterscheiden können. Wir nehmen im folgenden an, dass das Rauschsignal $N(t)$ **weisses Rauschen** darstellt. Weisses Rauschen hat die Eigenschaften, dass der Scharmittelwert μ null ist, die Scharstandardabweichung σ zeitlich konstant und die Autokorrelationsfunktion $\varphi_n(\tau)$ null für alle $\tau \neq 0$. Oft gilt $N(t) \sim \mathcal{N}(0, \sigma^2)$. Ein Korrelogramm von weissem Rauschen wird in Abbildung 7.22 gezeigt.

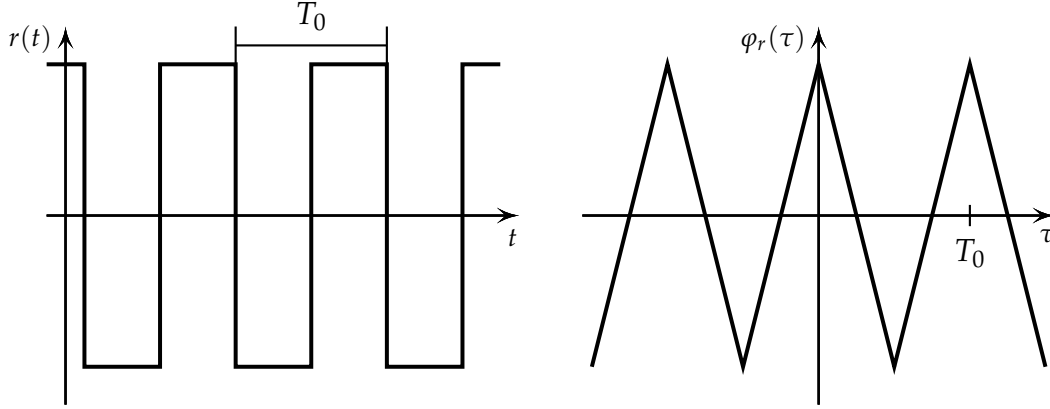


Abbildung 7.18.: (links:) Rechtecksignal $r(t)$ und (rechts:) die zugehörige Autokorrelationsfunktion $\varphi_r(\tau)$.

Abbildung 7.19 zeigt eine Realisierung $s(t)$ des Empfangssignals $S(t)$, wenn $r(t)$ darin enthalten ist. Die Kurzzeit-Autokorrelationsfunktion ergibt sich gemäss (7.8) zu

$$\begin{aligned} \varphi_T(\tau) &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} [r(t) + n(t)][r(t + \tau) + n(t + \tau)] dt \\ &= \varphi_{T;r}(\tau) + \varphi_{T;n}(\tau) + \varphi_{T;(r,n)}(\tau) + \varphi_{T;(n,r)}(\tau) . \end{aligned}$$

wobei $\varphi_{T;n}(\tau)$ die Kurzzeit-Autokorrelationsfunktion des Rauschsignals $N(t)$, $\varphi_{T;r}(\tau)$ die Kurzzeit-Autokorrelationsfunktion des Rechtecksignals und $\varphi_{T;(n,r)}(\tau)$, respektive $\varphi_{T;(r,n)}(\tau)$ die Kurzzeit-Kreuzkorrelationsfunktionen von den Signalen $r(t)$ und $N(t)$ bezeichnen. Für die Kurzzeit-Kreuzkorrelationsfunktion $\varphi_{T;(r,n)}(\tau)$ eines deterministischen Signals $r(t)$ mit einem Rauschsignal $N(t)$ gilt bei genügend grossem Intervall T unter der Voraussetzung, dass die beiden Signale statistisch unabhängig sind und mindestens eines davon auch mittelwertfrei ist,

$$\varphi_{T;(r,n)} \approx \varphi_{r,n}(\tau) = \varphi_{n,r}(-\tau) = 0 .$$

Denn in diesem Fall heben sich die positiven und negativen Anteile bei der Integration infolge der Unabhängigkeit von $r(t)$ und $N(t + \tau)$ auf.

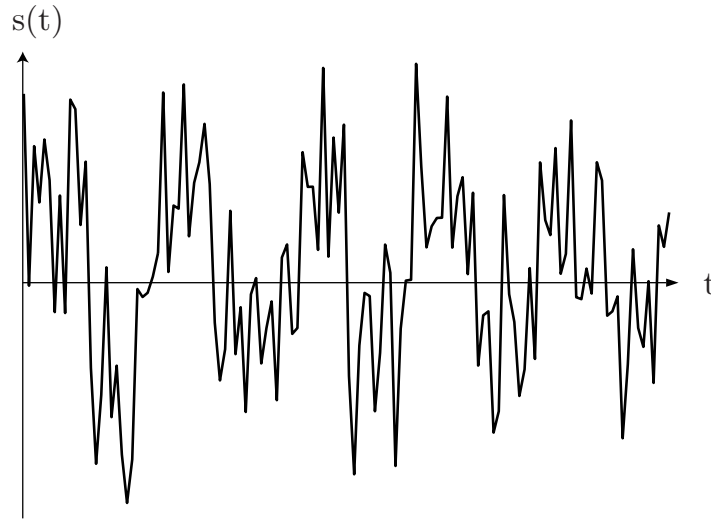


Abbildung 7.19.: Realisierung eines verrauschten Rechtecksignals.

Damit liefert der Detektor die Summe der Autokorrelationsfunktionen von $r(t)$ und $N(t)$. Die Autokorrelationsfunktion von $r(t)$ berechnet sich gemäss (7.9) mit

$$\varphi_r(\tau) = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} r(t)r(t+\tau) dt .$$

Die Autokorrelationsfunktion $\varphi_r(\tau)$ für das Rechtecksignal $r(t)$ wird im Beispiel C.5.3 des Kapitels C.5 des Anhangs berechnet und ist in Abbildung 7.18 dargestellt.

Die Autokorrelationsfunktion des Rauschsignals $N(t)$ kann mit dem Wiener-Khintchine-Theorem (siehe dazu Anhang C.5) ermittelt werden zu

$$\varphi_n(\tau) = \Phi_0 \frac{\omega_g}{\pi} \frac{\sin(\omega_g \tau)}{\omega_g \tau} ,$$

wobei ω_g die Kreisfrequenz des bandbegrenzten Rauschens bezeichnet – bei $\omega_g \rightarrow \infty$ wäre die Autokorrelationsfunktion die Dirac-Funktion $\delta(\tau)$. $\varphi_n(\tau)$ strebt also gegen null für grosses τ .

In Abbildung 7.20 ist die Autokorrelationsfunktion $\varphi_s(\tau)$ des Signals $S(t) = r(t) + N(t)$ dargestellt als Summe der Autokorrelationsfunktionen $\varphi_r(\tau)$ und $\varphi_n(\tau)$. Für den Fall, dass $r(t)$ nicht in $S(t)$ enthalten ist, ergibt sich $\varphi_s(\tau) = \varphi_n(\tau)$. Für genügend grosses τ zeigt sich deutlich ein periodischer Anteil in $\varphi_s(\tau)$ aufgrund der Autokorrelationsfunktion $\varphi_r(\tau)$ des periodischen Signals $r(t)$.

Die Entscheidung, ob die periodische Komponente $r(t)$ im empfangenen Signal enthalten ist oder nicht, kann mit einem Schwellwertdetektor vorgenommen werden,

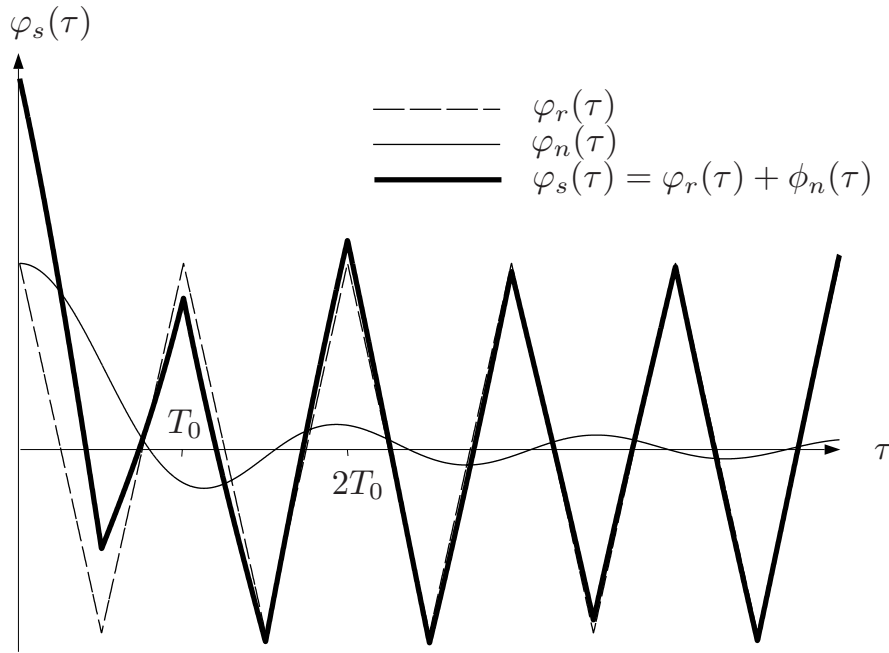


Abbildung 7.20.: Autokorrelationsfunktion $\varphi_s(\tau)$ des Signals $S(t)$ für den Fall, dass $r(t)$ in $S(t)$ enthalten ist.

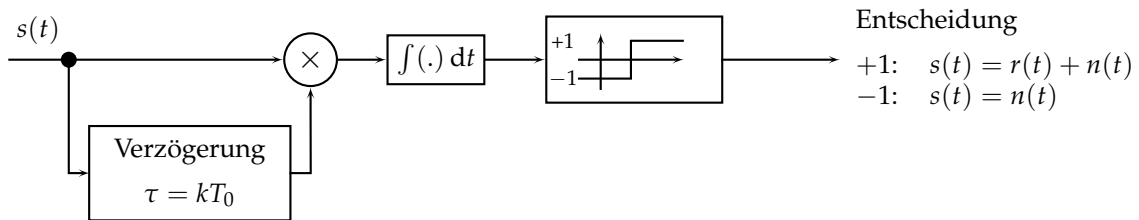


Abbildung 7.21.: Blockschaltbild des Detektors für $r(t)$.

der mit Vorteil an einer Stelle $\tau = kT_0, k \in \mathbb{N}$ arbeitet, wobei T_0 die Periodendauer des Rechtecksignals bezeichnet. Das Signal $s(t)$ wird also mit dem verzögerten Signal $s(t + \tau)$ multipliziert und über eine bestimmte Zeitspanne integriert, was der Kurzzeit-Autokorrelationsfunktion entspricht. Handelt es sich bloss um ein Rauschsignal, dann wird ein Wert nahe null resultieren. Ist der Wert deutlich grösser als ein Schwellwert, - als Schwellwert wird in der Regel der Wert $\varphi_{r,\max}/2$ verwendet - dann schliesst man, dass im Signal das Rechtecksignal enthalten ist. Ein Blockschaltbild der gesamten Anordnung ist in [Abbildung 7.21](#) dargestellt.

7.6. Zeitreihen und Autokorrelationsfunktionen mit R

Im Prinzip können die Beobachtungen eines zeitabhängigen stochastischen Prozesses – allgemein einer **Zeitreihe**, wobei die Beobachtungen einer Zeitreihe zu äquidistanten und aufeinanderfolgenden Zeitpunkten erhoben wurden – in R durch einen gewöhnlichen numerischen Vektor dargestellt werden. Funktionen, die speziell zur Behandlung von Zeitreihen gedacht sind, können mit solchen Vektoren arbeiten. Andererseits können Zeitreihen aber auch über eine eigene Datenstruktur abgebildet werden.

Wir erzeugen im folgenden $n = 120$ **unabhängige** Beobachtungen aus einer Normalverteilung mit Parametern $\mu = 0$ und $\sigma = 1$. Wir fassen diese Beobachtungen als aufeinanderfolgende monatliche Beobachtungen von 10 aufeinanderfolgenden Jahren 1964 bis 1973 auf.

In der Zeitreihentheorie wird eine derartige (hier künstlich erzeugte) Reihe auch als **weisses Rauschen** bezeichnet. Zunächst erzeugen wir den Vektor der Beobachtungen.

```
set.seed(1)
x <- rnorm(120)
```

Die grundlegende R-Funktion, mit der explizit eine Zeitreihe gebildet werden kann, ist `ts()`.

```
white.noise <- ts(x)
```

Es wird gezeigt, dass `white.noise` eine Zeitreihe ist, die zum Zeitpunkt $t = 1$ startet und zum gleichen Zeitpunkt $t = 120$ endet, wobei die Anzahl der Beobachtungen pro Zeitpunkt gleich 1 ist.

Wollen wir jedoch als grundlegende Zeitpunkte die Jahre 1964 bis 1973 ansehen, für die jeweils 12 Beobachtungen vorliegen, so können wir dies mittels

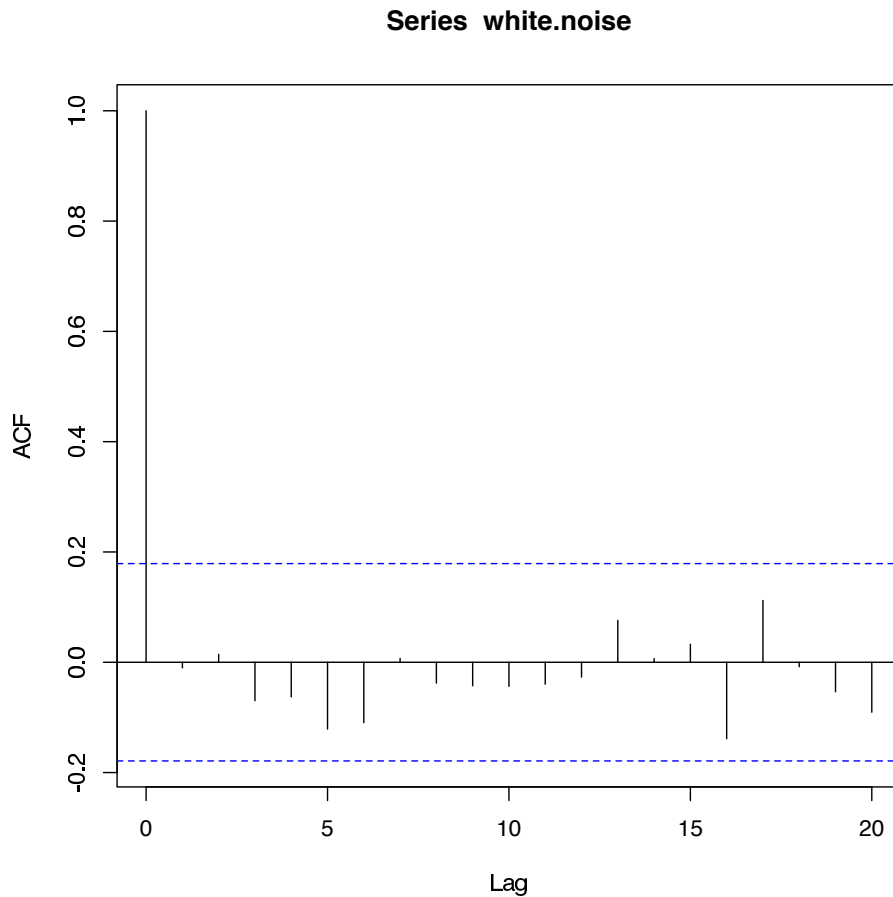
```
white.noise
```

erreichen. Bei der Anzeige der Zeitreihe werden durch das Setzen des Arguments `frequency = 12` automatisch Monatsnamen verwendet.

Zwischen Beobachtungen x_1, \dots, x_n einer Zeitreihe bestehen in der Regel stochastische Abhängigkeiten. Ein Ziel der Zeitreihenanalyse besteht darin, diese näher zu untersuchen, bzw., durch geeignete Modelle zu beschreiben, um den zugrundeliegenden Prozess zu verstehen und prognostizieren zu können. So kann es sein, dass für

ein festes τ zwischen Beobachtungen x_t und $x_{t+\tau}$ zu verschiedenen Zeitpunkten t ein (stochastischer) linearer Zusammenhang besteht. Man spricht dann von Autokorrelation zum *Lag* k . Die graphische Darstellung der Autokorrelationsfunktion $\varphi(\tau)$ für Lags $\tau = 0, 1, \dots, n - 1$ als Stabdiagramm heisst auch **Korrelogramm**. Häufig wird der maximale gezeichnete Lag kleiner als $n - 1$ gewählt. Mit

```
acf.white.noise <- acf(white.noise)
```



erhält man die Graphik aus Abbildung 7.22 und ein Objekt, aus dem weitere Informationen extrahiert werden können.

Die ACF wird hier von Lag 0 bis Lag 20 gezeichnet, wobei der Wert zum Lag 0 stets gleich 1 ist. Die Benennung der x -Achse bezieht sich auf die grundlegende Zeiteinheit, also hier Jahre. Ausserdem werden horizontale Linien $\pm 1.96 / \sqrt{n}$ als ein approximatives 95 % Konfidenzintervall eingezeichnet. Liegen Werte deutlich ausserhalb dieses Bereiches, so kann dies ein Kennzeichen für Autokorrelation sein. Wie bei einem weissen Rauschen zu erwarten, sind keine auffälligen Autokorrelationen im Korrelogramm von Abbildung 7.22 erkennbar.

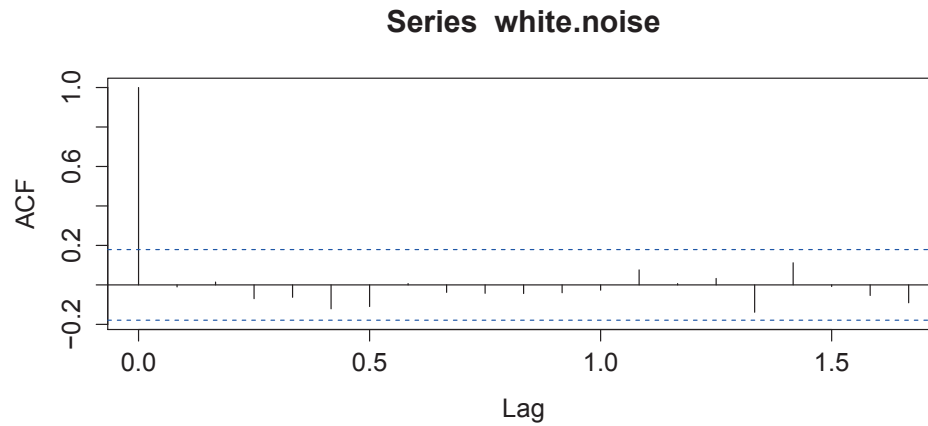


Abbildung 7.22.: Korrelogramm für weisses Rauschen

Lernziele

- ☐ Sie verstehen das Konzept eines stochastischen Prozesses.
- ☐ Sie können erkennen, wann ein stochastischer Prozess stationär und wann ergodisch ist.
- ☐ Sie können die Autokorrelationsfunktion und die Kreuzkorrelationsfunktion von einfachen stochastischen Prozessen berechnen.
- ☐ Sie können entscheiden, ob in einem verrauschten Signal ein vermutetes (deterministisches) Signal enthalten ist.

Anhang A.

Ergänzungen zur Binomialverteilung und Poissonverteilung

A.1. Herleitung des Binomialkoeffizienten

Wir betrachten n unabhängige Experimente mit Ausgang Erfolg oder Misserfolg. Die Erfolgswahrscheinlichkeit im i -ten Experiment sei $\pi_i \in \{0, 1\}$.

Frage: Was ist die Wahrscheinlichkeit, dass wir im gesamten x Erfolge beobachten? Zum Beispiel $x = 3$? Die Wahrscheinlichkeit für genau eine Auswahl von allen möglichen Ausgängen mit x Erfolgen bei n Experimenten ist gegeben durch

$$\pi^x (1 - \pi)^{n-x},$$

wobei die Experimente als unabhängig angenommen wurden. In Tabelle A.1 haben wir ein Feld eines „Experiments“ mit dem Symbol \bullet markiert, wenn Erfolg eintritt und mit dem Symbol \circ bei Misserfolg.

| | | | | | | | | | |
|-----------|---------|---------|-----------|---------|-----------|---------|---------|---------|---------|
| 1 | 2 | 3 | 4 | 5 | 6 | ... | ... | n-1 | n |
| \bullet | \circ | \circ | \bullet | \circ | \bullet | \circ | \circ | \circ | \circ |

Tabelle A.1.: Symbole für Erfolg und Misserfolg bei einem Experiment

Um die Wahrscheinlichkeit zu berechnen, dass im gesamten x Erfolge eintreten, müssen wir alle möglichen Ausgänge von den n Experimenten betrachten, die zu diesem Ergebnis führen. Die Reihenfolge innerhalb einer Auswahl spielt keine Rolle, d. h. es interessiert uns nicht, ob zuerst Experiment 4 und erst dann Experiment 1 Erfolg hat oder umgekehrt. In der Tabelle A.1 interessieren uns daher nur die verschiedenen „Muster“ und nicht, in welcher spezifischen Reihenfolge wir ein einzelnes Muster „angemalt“ haben.

Um den ersten Erfolg zu platzieren, haben wir n Möglichkeiten, für den zweiten verbleiben noch $n - 1$ und so weiter; bis für den letzten dann noch $n - x + 1$ Möglichkeiten übrig sind. Das gibt im gesamten $n(n - 1) \cdots (n - x + 1)$ Möglichkeiten.

Hier haben wir aber jeweils stillschweigend unterschieden, in welcher Reihenfolge die Erfolge eintreten. In obenstehender Tabelle hätten wir jeweils die Auswahl mit den Reihenfolgen der Erfolge $1 \leftrightarrow 4 \leftrightarrow 6$, $1 \leftrightarrow 6 \leftrightarrow 4$, $4 \leftrightarrow 1 \leftrightarrow 6$, $4 \leftrightarrow 6 \leftrightarrow 1$, $6 \leftrightarrow 1 \leftrightarrow 4$ und $6 \leftrightarrow 4 \leftrightarrow 1$ einzeln gezählt, obwohl wir dies ja eigentlich nicht unterscheiden wollen, da alle zum selben Muster führen. Für eine gegebene Auswahl gibt es $x!$ verschiedene mögliche Reihenfolgen, die Erfolge zu platzieren. Also haben wir genau so viel Mal zu viel gezählt. Wenn wir dies korrigieren wollen, erhalten wir

$$\frac{n(n-1) \cdots (n-x+1)}{x!}$$

verschiedene Möglichkeiten, x Erfolge bei n Experimenten zu platzieren, ohne die Reihenfolge der Erfolge zu berücksichtigen. Dies können wir auch schreiben als

$$\frac{n!}{x!(n-x)!},$$

was wir mit dem Binomialkoeffizienten $\binom{n}{x}$ abkürzen können („ n tief x “). Wir haben $\binom{n}{x}$ verschiedene Möglichkeiten, die alle zum Resultat „im gesamten x Erfolge“ führen. Jede dieser Möglichkeiten hat die gleiche Wahrscheinlichkeit $\pi^x(1-\pi)^{n-x}$. Die Wahrscheinlichkeit, im gesamten x Erfolge zu beobachten, ist also gegeben durch

$$\binom{n}{x} \pi^x (1-\pi)^{n-x}.$$

A.2. Erwartungswert und Varianz der Binomialverteilung

Den Erwartungswert μ errechnet man direkt aus der Definition

$$\mu = \sum_{i=1}^n x_i \cdot P(X_i = x_i)$$

und dem binomischen Lehrsatz zu:

$$\begin{aligned}
 \mu &= \sum_{k=0}^n k \binom{n}{k} \pi^k (1-\pi)^{n-k} \\
 &= n\pi \sum_{k=0}^n k \frac{(n-1)!}{(n-k)!k!} \pi^{k-1} (1-\pi)^{(n-1)-(k-1)} \\
 &= n\pi \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} \pi^{k-1} (1-\pi)^{(n-1)-(k-1)} \\
 &= n\pi \sum_{k=1}^n \binom{n-1}{k-1} \pi^{k-1} (1-\pi)^{(n-1)-(k-1)} \\
 &= n\pi \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} \pi^{\ell} (1-\pi)^{(n-1)-\ell} \quad \text{mit } \ell := k-1 \\
 &= n\pi \sum_{\ell=0}^m \binom{m}{\ell} \pi^{\ell} (1-\pi)^{m-\ell} \quad \text{mit } m := n-1 \\
 &= n\pi (\pi + (1-\pi))^m = n\pi 1^m = n\pi.
 \end{aligned}$$

Alternativ lässt sich der Erwartungswert μ mit der Summenregel für Erwartungswerte berechnen, wenn man berücksichtigt, dass die identischen Einzelprozesse X_i der Bernoulli-Verteilung mit $E(X_i) = \pi$ genügen. Dann ergibt sich, dass

$$X = X_1 + \cdots + X_n$$

einer $\text{Bin}(n, \pi)$ -Verteilung folgt, und:

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = n E(X_1) = n\pi.$$

Die Varianz bestimmt sich direkt aus dem Verschiebungssatz

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

zu

$$\text{Var}(X) = \sum_{k=0}^n k^2 \binom{n}{k} \pi^k (1-\pi)^{n-k} - n^2 \pi^2 = n\pi(1-\pi) = n\pi q$$

oder alternativ aus der Summenregel für die Varianz unabhängiger Zufallsvariablen, wenn man berücksichtigt, dass die identischen Einzelprozesse X_i der Bernoulli-Verteilung mit

$$\text{Var}(X_i) = \pi(1-\pi) = \pi q$$

genügen, zu

$$\begin{aligned}\operatorname{Var}(X) &= \operatorname{Var}(X_1 + \cdots + X_n) \\ &= \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n) \\ &= n \operatorname{Var}(X_1) \\ &= n\pi(1 - \pi) = n\pi q.\end{aligned}$$

Die zweite Gleichheit gilt, da die Einzelexperimente unabhängig sind, so dass die Einzelvariablen unkorreliert sind.

A.3. Herleitung der Poisson-Verteilung

Die Poisson-Verteilung lässt sich aus grundlegenden Prozesseigenschaften, den sogenannten Poissonschen Annahmen, herleiten. Wenn diese Eigenschaften einem Geschehen in guter Näherung zugeordnet werden können, wird die Ereignishäufigkeit Poisson-verteilt sein.

Man betrachtet ein Zeitkontinuum t , auf dem zählbare Ereignisse mit konstanter mittlerer Anzahl g pro Einheitsintervall stattfinden, also pro Sekunde ist die mittlere Anzahl Ereignisse konstant.

Nun richtet man den Blick auf ein genügend kleines Zeitintervall Δt . Was sich dort ereignet, bestimmt die globale Verteilung auf der Zeitachse.

Die drei Poissonschen Annahmen lauten:

1. Innerhalb des Zeitintervalls $[t, t + \Delta t]$ gibt es höchstens ein Ereignis und beliebig viele Momente, in denen nichts geschieht (Seltenheit).
2. Die Wahrscheinlichkeit, ein Ereignis im Zeitintervall Δt zu finden, ist proportional zur Länge des Zeitintervalls Δt . Da g konstant ist, ist diese Wahrscheinlichkeit damit auch unabhängig von t .
3. Das Eintreten eines Ereignisses im Intervall Δt wird nicht beeinflusst von Ereignissen, die in der Vorgeschichte stattgefunden haben (Geschichtslosigkeit).

Mit Annahme 1 und 2 ist die Wahrscheinlichkeit, ein Ereignis im Intervall Δt zu finden, gegeben als

$$p_1(\Delta t) = g \cdot \Delta t$$

sowie die Wahrscheinlichkeit eines leeren Intervalls durch

$$p_0(\Delta t) = 1 - p_1(\Delta t) = 1 - g \cdot \Delta t.$$

Nach Annahme 3 ist die Wahrscheinlichkeit eines leeren Intervalls Δt unabhängig vom Auftreten irgendwelcher Ereignisse im Bereich t davor. So berechnet man die Wahrscheinlichkeit für kein Ereignis bis zum Zeitpunkt $t + \Delta t$ zu

$$p_0(t + \Delta t) = p_0(t) \cdot p_0(\Delta t) = p_0(t) - g \cdot p_0(t) \cdot \Delta t.$$

Das ergibt näherungsweise die Differentialgleichung $dp_0(t)/dt = -g \cdot p_0(t)$ mit der Lösung

$$p_0(t) = e^{-g \cdot t}$$

mit der Randbedingung $p_0(0) = 1$. Ebenso findet man die Wahrscheinlichkeit für m Ereignisse bis zum Zeitpunkt $t + \Delta t$

$$\begin{aligned} p_m(t + \Delta t) &= p_m(t) \cdot p_0(\Delta t) + p_{m-1}(t) \cdot p_1(\Delta t) \\ &= p_m(t) - g \cdot p_m(t) \cdot \Delta t + g \cdot p_{m-1}(t) \cdot \Delta t. \end{aligned}$$

Jedes angehängte Zeitintervall Δt darf nach Annahme 1 nur entweder kein oder ein Ereignis enthalten. Die entsprechende Differentialgleichung

$$dp_m(t)/dt = -g \cdot p_m(t) + g \cdot p_{m-1}(t)$$

hat die Lösung

$$p_m(t) = \frac{(g \cdot t)^m}{m!} e^{-g \cdot t}.$$

Identifiziert man nun in diesem Ausdruck, der die Wahrscheinlichkeit des Eintretens von m Ereignissen im Kontinuumsbereich t beschreibt, die Parameter $(g \cdot t)$ mit λ und m mit k , stimmt er mit der Formel der Poisson-Verteilung überein:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Die Zahl λ ergibt sich in vielen Aufgabenstellungen als Produkt einer „Rate“ (Anzahl von Ereignissen pro Einheitsintervall) und einem Vielfachen des Einheitsintervalls.

A.4. Erwartungswert und Varianz der Poissonverteilung

Ist die Zufallsvariable X Poisson-verteilt, also $X \sim \text{Pois}(\lambda)$, so ist λ zugleich Erwartungswert und Varianz, denn:

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

$$\begin{aligned}
 E(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
 &= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \left(\sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right) \\
 &= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
 &= \lambda^2 \cdot e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda^2 + \lambda.
 \end{aligned}$$

Nach dem Verschiebungssatz gilt nun: $\text{Var}(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Alternativ lassen sich Erwartungswert und Varianz auch folgendermassen berechnen: Seien X_1, \dots, X_n n unabhängige Bernoulli-verteilte Zufallsvariablen mit $\pi = \lambda/n$ und sei $X := X_1 + \dots + X_n$. Für $n \rightarrow \infty$ gilt $X \sim \text{Pois}(\lambda)$ und

$$\begin{aligned}
 E(X) &= E(X_1) + \dots + E(X_n) = \underbrace{\frac{\lambda}{n} + \dots + \frac{\lambda}{n}}_{n \text{ mal}} = \lambda \rightarrow \lambda \\
 \text{Var}(X) &= \text{Var}(X_1) + \dots + \text{Var}(X_n) \\
 &= \underbrace{\frac{\lambda}{n} \left(1 - \frac{\lambda}{n}\right) + \dots + \frac{\lambda}{n} \left(1 - \frac{\lambda}{n}\right)}_{n \text{ mal}} = \lambda \left(1 - \frac{\lambda}{n}\right) \rightarrow \lambda.
 \end{aligned}$$

A.5. Poisson-Verteilung als Grenzfall der Binomialverteilung

Die Wahrscheinlichkeitsverteilung der Binomialverteilung $\text{Bin}(n, \pi)$ lautet

$$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k} = \frac{n!}{(n-k)!k!} \pi^k (1 - \pi)^{n-k}.$$

Die Poisson-Verteilung lässt sich aus der Binomialverteilung herleiten. Sie ist die Grenzverteilung der Binomialverteilung bei sehr kleinen Anteilen der interessierten Merkmale und sehr grossem Stichprobenumfang: $n \rightarrow \infty$ und $\pi \rightarrow 0$ unter der Nebenbedingung, dass das Produkt $n\pi = \lambda$ einen Wert annimmt, der weder null noch unendlich ist. λ ist dann für alle in der Grenzwertbildung betrachteten Binomialverteilungen wie auch für die resultierende Poisson-Verteilung der Erwartungswert.

Der Wert einer Poisson-verteilten Zufallsvariable an der Stelle k ist der Grenzwert $n \rightarrow \infty$ einer Binomialverteilung mit $\pi = \frac{\lambda}{n}$ an der Stelle k :

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(X = k) &= \lim_{n \rightarrow \infty} \frac{n!}{k! (n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \lim_{n \rightarrow \infty} \left(\frac{\lambda^k}{k!}\right) \left(\frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &= \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} \underbrace{\left(\frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-k+1}{n}\right)}_{\rightarrow 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\rightarrow e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\rightarrow 1} \\
 &= \frac{\lambda^k e^{-\lambda}}{k!}.
 \end{aligned}$$

Bei grossen Stichproben lässt sich folglich die Binomialverteilung gut durch die Poisson-Verteilung approximieren.

A.6. Berechnung von Erwartungswert und Varianz mit Hilfe von charakteristischen Funktionen

Die charakteristische Funktion $\Phi_x(k)$ einer Wahrscheinlichkeitsverteilung $f(x)$ ist definiert als der Erwartungswert von e^{ikx} , d. h., die Fourier-Transformierte der Wahrscheinlichkeitsverteilung $f(x)$:

$$\Phi_x(k) = E[e^{ikx}] = \int_{-\infty}^{\infty} e^{ikx} f(x) dx. \quad (\text{A.1})$$

Für diskrete Zufallsvariablen X_j definiert man

$$\Phi_x(k) = E[e^{ikx}] = \sum_j e^{ikx} f(x_j).$$

Falls das Integral (A.1) existiert, existiert auch die Umkehrung

$$f(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikz} \Phi_z(k) dk.$$

Die charakteristische Funktion für die Binomialverteilung lautet:

$$\Phi_x(k) = \left(\pi(e^{ik} - 1) + 1\right)^n$$

für die Poisson-Verteilung findet sich

$$\Phi_x(k) = \exp\left(\lambda(e^{ik} - 1)\right).$$

Falls die charakteristische Funktion einer Wahrscheinlichkeitsverteilung gegeben ist, kann man die **Momente der Ordnung** m über folgendes Differential erhalten:

$$\begin{aligned} \frac{d^m}{dk^m} \Phi_x(k) \Big|_{k=0} &= \frac{d^m}{dk^m} \int_{-\infty}^{\infty} e^{ikx} f(x) dx \Big|_{k=0} \\ &= i^m \int_{-\infty}^{\infty} x^m f(x) dx \\ &= i^m E[X^m]. \end{aligned}$$

Mit Hilfe der charakteristischen Funktion lassen sich nun die Erwartungswerte und Varianzen der verschiedenen Wahrscheinlichkeitsverteilungen berechnen. Für die Binomialverteilung ergibt sich zum Beispiel für den Erwartungswert:

$$i^1 E[X] = \frac{d}{dk} \left(\pi(e^{ik} - 1) + 1 \right)^n \Big|_{k=0} = i \cdot n\pi.$$

Für die Varianz erhalten wir

$$\begin{aligned} \text{Var}(X) &= E[X^2] - E^2[X] \\ &= i^{-2} \frac{d^2}{dk^2} \left(\pi(e^{ik} - 1) + 1 \right)^n \Big|_{k=0} - (n\pi)^2 = n\pi(1 - \pi). \end{aligned}$$

A.7. Erwartungstreue der Varianz

Sie haben sich bestimmt schon gefragt, warum in der Definition der empirischen Varianz

$$\frac{1}{n-1} \sum_{i=1}^n n(x_i - \bar{x}_n)^2$$

der Faktor $\frac{1}{n-1}$ und nicht etwa $\frac{1}{n}$ steht. Nun dies hat damit zu tun, dass wir eine Eigenschaft von der empirischen Varianz einfordern, die **Erwartungstreue** genannt wird. Mit der empirischen Varianz schätzen wir ja den wahren, uns unbekannten Wert der Varianz. Ein Schätzer heisst **erwartungstreu**, wenn sein Erwartungswert gleich dem wahren Wert des zu schätzenden Parameters ist.

Betrachten wir zuerst ein einfaches Beispiel: den Schätzer des Erwartungswertes. Zur Schätzung des Erwartungswertes μ aus einem Datensatz x_1, \dots, x_n wird üblicherweise das arithmetische Mittel

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

verwendet. Wiederholen wir die Messung und damit die Schätzung des Erwartungswertes, so können wir die Messpunkte x_1, \dots, x_n als Realisierungen von den Zufallsvariablen X_1, \dots, X_n und das arithmetische Mittel als Realisierung der Zufallsvariablen

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

auffassen. Werden alle Stichprobenvariablen X_i zufällig aus der Grundgesamtheit gezogen, so haben alle den Erwartungswert $E(X_i) = \mu_X$. Damit berechnet sich der Erwartungswert des arithmetischen Mittels zu

$$\begin{aligned} E[\bar{X}_n] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n} \sum_{i=1}^n E(X_i) \\ &= \frac{1}{n} \cdot n \cdot \mu_X \\ &= \mu_X \end{aligned}$$

Das arithmetische Mittel ist also ein erwartungstreuer Schätzer des unbekannten Verteilungsparameters μ_X .

Fassen wir nun die empirische Varianz als Realisierung der Zufallsvariablen

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

auf, so lässt sich der Erwartungswert vom diesem Schätzer folgendermassen berechnen

$$\begin{aligned}
 E \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n ((X_i - \mu_X) - (\bar{X}_n - \mu_X))^2 \right] \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n \{ (X_i - \mu_X)^2 - 2(X_i - \mu_X)(\bar{X}_n - \mu_X) + (\bar{X}_n - \mu_X)^2 \} \right] \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu_X)^2 - 2n(\bar{X}_n - \mu_X)(\bar{X}_n - \mu_X) + n(\bar{X}_n - \mu_X)^2 \right] \\
 &= \frac{1}{n-1} E \left[\sum_{i=1}^n (X_i - \mu_X)^2 - n(\bar{X}_n - \mu_X)^2 \right] \\
 &= \frac{1}{n-1} (n\text{Var}(X) - n\text{Var}(\bar{X}_n)) \\
 &= \frac{1}{n-1} (n\text{Var}(X) - \text{Var}(X)) \\
 &= \text{Var}(X)
 \end{aligned}$$

Wir haben dabei folgende Definition der Varianz einer Zufallsvariablen X benutzt:

$$\text{Var}(X) = E \left[(X - \mu_X)^2 \right]$$

und die Eigenschaft der Varianz

$$\begin{aligned}
 n\text{Var}(\bar{X}_n) &= n\text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \\
 &= \frac{n}{n^2} \text{Var} \left(\sum_{i=1}^n X_i \right) \\
 &= \text{Var}(X)
 \end{aligned}$$

Somit haben wir also gezeigt, dass die empirische Varianz erwartungstreu ist.

A.8. Beweis des Zentralen Grenzwertsatzes

Zentraler Grenzwertsatz

Die Folge X_1, X_2, \dots bezeichne unabhängige Zufallsvariablen, die alle den Erwartungswert 0 und Varianz σ^2 haben, eine gemeinsame kumulative Verteilungsfunktion F und dieselbe momentengenerierende Funktion

$$M(t) = E[e^{tx}] = \int_{-\infty}^{\infty} f(x)e^{tx} dx$$

besitzen, die in der Nähe von 0 definiert sei. Es sei

$$S_n = \sum_{i=1}^n X_i$$

Dann gilt

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)$$

für alle $x \in \mathbb{R}$.

Beweis Es sei $Z_n = S_n/(\sigma\sqrt{n})$. Wir werden zeigen, dass die momentengenerierende Funktion von Z_n gegen die momentengenerierende Funktion der Standardnormalverteilung strebt, nämlich $\exp(t^2/2)$. Da S_n eine Summe von unabhängigen Zufallsvariablen ist, gilt

$$M_{S_n}(t) = [M(t)]^n$$

und

$$M_{Z_n}(t) = \left[M\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Die Entwicklung von $M(s)$ in eine Taylor-Reihen um $s = 0$ lautet

$$M(s) = M(0) + sM'(0) + \frac{1}{2}s^2M''(0) + \epsilon_s$$

wobei $\epsilon_s/s^2 \rightarrow 0$ für $s \rightarrow 0$. Es gilt, dass $E(X) = 0$, $M'(0) = 0$ und $M''(0) = \sigma^2$ und $t/(\sigma\sqrt{n}) \rightarrow 0$ für $n \rightarrow \infty$. Somit ist

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \epsilon_n$$

wobei $\epsilon_n/(t^2/(n\sigma^2)) \rightarrow 0$ für $n \rightarrow \infty$. Wir haben folglich

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \epsilon_n\right)^n$$

Da für $a_n \rightarrow a$ gilt, dass

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = \exp(a)$$

folgt

$$M_{Z_n(t)} \rightarrow \exp(t^2/2) \quad \text{für } n \rightarrow \infty$$

wobei $\exp(t^2/2)$ die momentengenerierende Funktion der Standardnormalverteilung ist, was wir zeigen wollten. \square

Diese Formulierung des Zentralen Grenzwertsatzes ist eine der simpelsten von vielen Formulierungen unterschiedlicher Allgemeinheit und von variierendem Abstraktionsgrad. Wir haben den Zentralen Grenzwertsatz unter der Annahme bewiesen, dass eine momentengenerierende Funktion existiert, was eine eher starke Annahme darstellt. Für praktische Zwecke, insbesondere für die Statistik, ist die Grenzwertbetrachtung selber nicht von herausragendem Interesse. Statistiker interessieren sich eher für die Näherung mit beschränkten Werten für n . Es ist allerdings unmöglich, eine abschliessende und präzise Aussage zu machen, wie gut eine Approximation ist. Es bestehen bloss gewisse Faustregeln: je symmetrischer die Verteilung ist und je schneller die Ausläufer der Verteilung abfallen, desto besser ist die Approximation für bereits kleine Werte von n .

Anhang B.

Aus der Normalverteilung hergeleitete Verteilungen

Die **Chi-Quadrat-Verteilung**¹ (χ^2 -Verteilung) ist eine stetige Wahrscheinlichkeitsverteilung über der Menge der positiven reellen Zahlen. Üblicherweise ist mit „Chi-Quadrat-Verteilung“ die zentrale Chi-Quadrat-Verteilung gemeint. Ihr einziger Parameter n muss eine natürliche Zahl sein und wird **Freiheitsgrad** genannt.

Sie ist eine der Verteilungen, die aus der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ hergeleitet werden kann. Hat man n Zufallsvariablen X_i , die unabhängig und standardnormalverteilt sind, so ist die Chi-Quadrat-Verteilung mit n Freiheitsgraden definiert als die Verteilung der Summe der quadrierten Zufallsvariablen $X_1^2 + \dots + X_n^2$.

B.1. Dichtefunktion der Chi-Quadrat-Verteilung

Chi-Quadrat-Verteilung

Bezeichnen wir mit Z eine standardnormalverteilte Zufallsvariable, dann wird die Verteilung von $X = Z^2$ **Chi-Quadrat-Verteilung mit einem Freiheitsgrad** genannt.

Wir haben also $X = Z^2$, wobei $Z \sim \mathcal{N}(0, 1)$. Somit haben wir

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(Z^2 \leq x) \\ &= P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\ &= \Phi(\sqrt{x}) - \Phi(-\sqrt{x}). \end{aligned}$$

¹Zu Risiken und Nebenwirkungen lesen Sie diesen Anhang mit Gelassenheit und fragen Sie Ihren Dozenten oder Tutoren. Im Falle eines Falles: Zu den riesigen Nebenwirkungen fressen Sie den Anhang und erschlagen Sie den irren Dozenten Ihres Tutoren.

Anhang B. Aus der Normalverteilung hergeleitete Verteilungen

Wir finden die Dichtefunktion von X , indem wir die Kumulative Verteilungsfunktion $F_X(x)$ nach x ableiten. Da $\Phi'(x) = \varphi(x)$, ergibt sich mit der Kettenregel

$$\begin{aligned} f_X(x) &= \frac{1}{2} x^{-1/2} \varphi(\sqrt{x}) + \frac{1}{2} \varphi(-\sqrt{x}) \\ &= x^{-1/2} \varphi(\sqrt{x}), \end{aligned}$$

wobei wir im letzten Schritt verwendet haben, dass φ symmetrisch bezüglich der y -Achse ist. Dies ergibt dann

$$f_X(x) = \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-x/2}.$$

Wir bezeichnen diese Dichtefunktion auch als $f_{\chi_1^2}$, also der Chi-Quadrat-Verteilung mit einem Freiheitsgrad.

Wir stellen fest, dass wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, dann ist $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$ und somit

$$[(X - \mu)/\sigma]^2 \sim \chi_1^2.$$

Wir stellen fest, dass die Chi-Quadrat-Verteilung mit einem Freiheitsgrad ein Spezialfall der Gammaverteilung ist.

Gammaverteilung

Die **Gammaverteilung** ist eine kontinuierliche Wahrscheinlichkeitsverteilung über der Menge der positiven reellen Zahlen. Sie ist eine direkte Verallgemeinerung der Exponentialverteilung und hat folgende Dichtefunktion

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0. \end{cases}$$

wobei der Ausdruck $\Gamma(\alpha)$ für den Funktionswert der Gammafunktion steht, nach der die Verteilung auch benannt ist. Die Gammafunktion ist für positive reelle Zahlen α über das Integral

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

definiert. Die Werte von $\Gamma(\alpha)$ kann man auch berechnen mit $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$, $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$ mit $\alpha \in \mathbb{R}^+$.

Die Chi-Quadrat-Verteilung mit einem Freiheitsgrad ist also ein Spezialfall der Gammaverteilung, wobei $\alpha = 1/2$ und $\lambda = 1/2$.

Die **Momentengenerierende Funktion** $M(t)$ der Gammaverteilung ist gegeben durch:

$$\begin{aligned} M(t) &\equiv E[e^{tx}] \\ &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \left(\frac{\lambda}{\lambda - t} \right)^\alpha. \end{aligned}$$

Mit der Momentengenerierenden Funktion lassen sich Erwartungswert und Varianz einer Wahrscheinlichkeitsverteilung sehr elegant berechnen:

$$M'(t=0) = E[X] = \frac{\alpha}{\lambda}$$

und

$$M''(t=0) = E[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2},$$

woraus sich für die Varianz folgendes Resultat ergibt

$$\begin{aligned} \text{Var}[X] &= E[X^2] - (E[X])^2 \\ &= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} \\ &= \frac{\alpha}{\lambda^2}. \end{aligned}$$

Über den kleinen Umweg der Gammaverteilung und deren Momentengenerierenden Funktion lässt sich nun die Dichte der Chi-Quadrat-Verteilung mit n Freiheitsgraden herleiten.

χ^2 -Quadrat-Verteilung mit n Freiheitsgraden

Die χ^2 -**Quadrat-Verteilung mit n Freiheitsgraden** ist definiert als die Summe aus n stochastisch unabhängiger quadrierter standardnormalverteilter Zufallsvariablen

$$X_1^2 + \dots + X_n^2,$$

mit $X_i \sim \mathcal{N}(0,1)$ für $i = 1, \dots, n$ und wir mit χ_n^2 bezeichnet. Die Summe „quadrierter“ Größen kann keine negativen Werte annehmen.

Anhang B. Aus der Normalverteilung hergeleitete Verteilungen

Die Dichte der Zufallsvariable

$$V = X_1^2 + \dots + X_n^2,$$

mit X_1, \dots, X_n unabhängig und standardnormalverteilt, ergibt sich aus der Momentengenerierenden Funktion

$$\begin{aligned} M(t) &= E[e^{tV}] \\ &= E[e^{tX_1^2} \cdot e^{tX_2^2} \dots e^{tX_n^2}] \\ &= E[e^{tX_1^2}] \cdot E[e^{tX_2^2}] \dots E[e^{tX_n^2}], \end{aligned}$$

wobei wir im letzten Schritt verwendet haben, dass alle X_i stochastisch unabhängig sind. Nun haben wir aber bereits $E[e^{tX_i^2}]$ berechnet, und zwar ist dies nichts anderes als die Momentengenerierende Funktion der Chi-Quadrat-Verteilung mit einem Freiheitsgrad:

$$\left(\frac{\lambda}{\lambda - t} \right)^\alpha.$$

Somit finden wir für die Momentengenerierende Funktion der Chi-Quadrat-Verteilung

$$\begin{aligned} M(t) &= \left(\frac{\lambda}{\lambda - t} \right)^{n \cdot \alpha} \Big|_{\alpha=\frac{1}{2}, \lambda=\frac{1}{2}} \\ &= (1 - 2t)^{-n/2} \end{aligned}$$

Dies ist aber nun nichts anderes als die Momentengenerierende Funktion der Gammaverteilung mit den Parameterwerten $\alpha = n/2$ und $\lambda = 1/2$.

χ_n^2 -Verteilung mit n Freiheitsgraden

Die Dichte der χ_n^2 -Verteilung mit n Freiheitsgraden ist also gegeben durch:

$$f(v) = \begin{cases} \frac{v^{\frac{n}{2}-1} e^{-\frac{v}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} & v > 0 \\ 0 & v \leq 0 \end{cases}$$

Der Erwartungswert der Chi-Quadrat-Verteilung mit n Freiheitsgraden ergibt sich nun direkt aus der Momentengenerierenden Funktion

$$\begin{aligned} M'(t=0) &= E[V] \\ &= -n/2 \cdot (1 - 2t)^{-n/2-1} \cdot (-2) \Big|_{t=0} \\ &= n. \end{aligned}$$

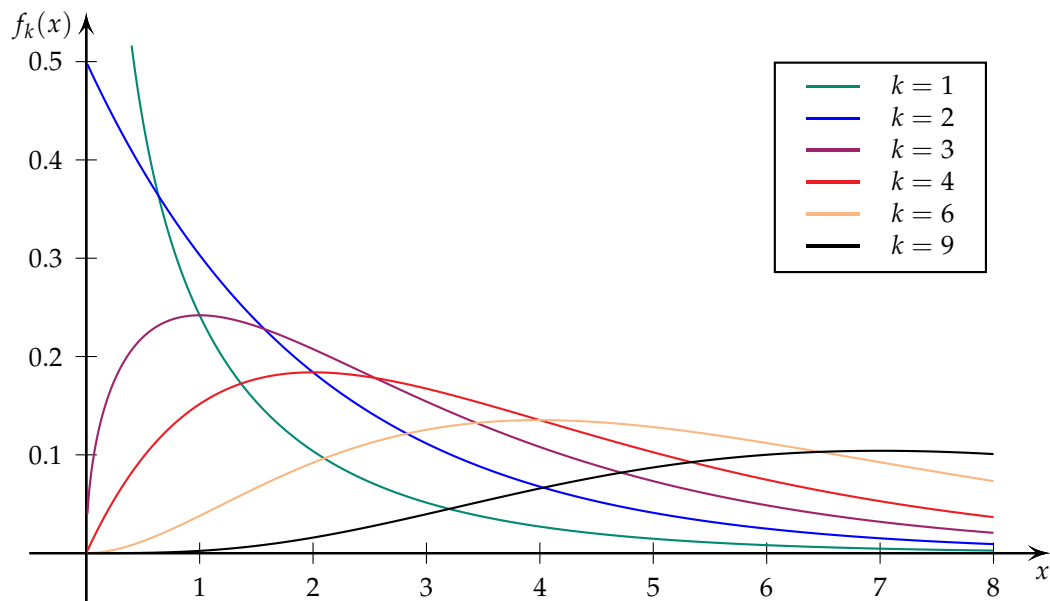


Abbildung B.1.: Dichten der Chi-Quadrat-Verteilung mit verschiedenen Freiheitsgraden k

Eine ähnliche Rechnung ergibt für die Varianz

$$\text{Var}[V] = 2n.$$

B.2. t -Verteilung

t -Verteilung

Es seien die Zufallsvariablen Z standardnormalverteilt, also $Z \sim \mathcal{N}(0, 1)$, und U χ_n^2 -verteilt, also $U \sim \chi_n^2$. Falls Z und U stochastisch unabhängig sind, dann folgt die Zufallsvariable

$$\frac{Z}{\sqrt{U/n}}$$

der t -Verteilung mit n Freiheitsgraden.

Die Wahrscheinlichkeitsdichte der t -Verteilung lässt sich herleiten aus der gemeinsamen Dichte der beiden unabhängigen Zufallsvariablen Z und χ_n^2 , die standardnor-

Anhang B. Aus der Normalverteilung hergeleitete Verteilungen

mal, beziehungsweise χ_n^2 -verteilt sind:

$$f_{Z, \chi_n^2}(z, y) = \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \cdot \frac{y^{\frac{n}{2}-1} e^{-\frac{1}{2}y}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}.$$

Mit der Transformation

$$t = z / \sqrt{y/n}, \quad v = y,$$

bekommt man die gemeinsame Dichte von $T = Z / \sqrt{\chi_n^2/n}$ und χ_n^2 , wobei $-\infty < t < \infty$ und $0 \leq v < \infty$.

Die Jacobideterminante dieser Transformation ist:

$$\det \frac{\partial(z, y)}{\partial(t, v)} = \begin{vmatrix} \sqrt{\frac{v}{n}} & 0 \\ \diamond & 1 \end{vmatrix} = \sqrt{\frac{v}{n}}.$$

Der Wert \diamond ist unwichtig, weil er bei der Berechnung der Determinante mit 0 multipliziert wird. Die neue Dichtefunktion schreibt sich also

$$f_{T, \chi_n^2}(t, v) = \frac{e^{-\frac{1}{2}v \frac{t^2}{n}}}{\sqrt{2\pi}} \cdot \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} v^{\frac{n}{2}-1} e^{-\frac{1}{2}v} \cdot \sqrt{\frac{v}{n}}.$$

Gesucht ist nun die Randverteilung $f(t)$ als Integral über die nicht interessierende Variable v :

$$\begin{aligned} f(t) &= \int_0^\infty f_{T, \chi_n^2}(t, v) dv \\ &= \frac{1}{\sqrt{n\pi} 2^{(n+1)/2} \Gamma(n/2)} \int_0^\infty v^{(n-1)/2} e^{-v(1+t^2/n)/2} dv \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \end{aligned}$$

***t*-Verteilung mit n Freiheitsgraden**

Die Dichtefunktion der ***t*-Verteilung mit n Freiheitsgraden** ist gegeben durch

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Anhang B. Aus der Normalverteilung hergeleitete Verteilungen

Wenn die unabhängigen Zufallsvariablen X_1, X_2, \dots, X_n identisch normalverteilt sind mit Erwartungswert μ und Standardabweichung σ_X , dann sind der Stichprobenmittelwert gegeben durch

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

und die Stichprobenvarianz durch

$$\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Es kann gezeigt werden, dass \bar{X}_n und $\hat{\sigma}_X$ stochastisch unabhängig sind.

Die Zufallsgrösse

$$\frac{\bar{X}_n - \mu}{\sigma_X / \sqrt{n}}$$

folgt einer Standardnormalverteilung. Nun möchten wir aber wissen, welche Verteilung die Grösse

$$\frac{\bar{X}_n - \mu}{\hat{\sigma}_X / \sqrt{n}}$$

hat. Dazu schreiben wir als erstes die obige Grösse etwas um:

$$\begin{aligned} \frac{\bar{X}_n - \mu}{\hat{\sigma}_X / \sqrt{n}} &= \frac{\bar{X}_n - \mu}{\hat{\sigma}_X / \sqrt{n}} \cdot \frac{\sigma_X}{\sigma_X} \\ &= \frac{\bar{X}_n - \mu}{\sigma_X / \sqrt{n}} \cdot \frac{\sigma_X}{\hat{\sigma}_X} \\ &= \frac{\bar{X}_n - \mu}{\sigma_X / \sqrt{n}} \bigg/ \left(\frac{\hat{\sigma}_X}{\sigma_X} \right). \end{aligned}$$

Nun stellt sich die Frage, wie $\frac{\hat{\sigma}_X}{\sigma_X}$ verteilt ist. Als erstes stellen wir fest, dass

$$\begin{aligned} (n-1) \cdot \left(\frac{\hat{\sigma}_X}{\sigma_X} \right)^2 &= (n-1) \cdot \frac{1}{\sigma_X^2} \cdot \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \frac{1}{\sigma_X^2} \cdot \sum_{i=1}^n (X_i - \bar{X}_n)^2. \end{aligned}$$

Nun wissen wir, dass

$$\frac{1}{\sigma_X^2} \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2.$$

Dieser Ausdruck lässt sich aber umformen in

$$\frac{1}{\sigma_X^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{1}{\sigma_X^2} \sum_{i=1}^n [(X_i - \bar{X}_n) + (\bar{X}_n - \mu)]^2.$$

Multiplizieren wir den quadratischen Term aus und benutzen, dass $\sum_{i=1}^n (X_i - \bar{X}_n)^2 = 0$, so erhalten wir

$$\underbrace{\frac{1}{\sigma_X^2} \sum_{i=1}^n (X_i - \mu)^2}_W = \underbrace{\frac{1}{\sigma_X^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2}_U + \underbrace{\left(\frac{\bar{X}_n - \mu}{\sigma_X / \sqrt{n}} \right)^2}_V$$

Dies ist eine Beziehung von der Form $W = U + V$, wobei U und V stochastisch unabhängig (was etwas schwieriger zu zeigen ist). Somit gilt für die Momentengenerierenden Funktionen

$$M_W(t) = \frac{M_U(t)}{M_V(t)}.$$

Da sowohl W wie auch V χ_n^2 -verteilt sind, gilt

$$\begin{aligned} M_U(t) &= \frac{M_W(t)}{M_V(t)} \\ &= \frac{(1 - 2t)^{-n/2}}{(1 - 2t)^{-1/2}} \\ &= (1 - 2t)^{-(n-1)/2}. \end{aligned}$$

Also folgt U , respektive der Ausdruck $\frac{1}{\sigma_X^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ einer χ_{n-1}^2 -Verteilung. $\left(\frac{\hat{\sigma}_X}{\sigma_X} \right)^2$ ist also eine χ_{n-1}^2 -verteilte Zufallsvariable dividiert durch die Anzahl Freiheitsgrade. Daraus schliessen wir, dass

$$\frac{\bar{X}_n - \mu}{\hat{\sigma}_X / \sqrt{n}}$$

wie eine Zufallsvariable $Z / \sqrt{U/(n-1)}$ verteilt ist, wobei $Z \sim \mathcal{N}(0, 1)$ und $U \sim \chi_{n-1}^2$, somit folgt $\frac{\bar{X}_n - \mu}{\hat{\sigma}_X / \sqrt{n}}$ einer t_{n-1} -Verteilung.

B.3. Chi-Quadrat Test

Beim Chi-Quadrat Test wird eine Hypothese über die unbekannte Verteilung der Zufallsvariablen X geprüft, woraus sich der Name Anpassungstest, Verteilungstest bzw. Goodness-of-fit-Test ergibt.

Wir fassen die n unabhängigen Beobachtungen x_1, \dots, x_n als Realisationen der Zufallsvariablen X auf. Es muss nun eine Intervallbildung der beobachteten Werte in disjunkte, aneinander angrenzende Klassen erfolgen. Mit k als Anzahl der Klassen ($k \geq 2$) können die Klassen allgemein wie folgt geschrieben werden:

$$(x_0^*, x_1^*), (x_1^*, x_2^*), \dots, (x_{k-1}^*, x_k^*)$$

bzw. (x_{j-1}^*, x_j^*) für $j = 1, \dots, k$. Es bezeichne $h_j(x_{j-1}^* < X \leq x_j^*) = h_j$ die beobachtete absolute Häufigkeit in der j -ten Klasse der Stichprobe für $j = 1, \dots, k$. Die Intervallbildung sollte so erfolgen, dass $h_j \geq 5$ für alle $j = 1, \dots, k$.

Wir bezeichnen mit p_j für $j = 1, \dots, k$ die Wahrscheinlichkeit, dass die Zufallsvariable X in die j -te Klasse (x_{j-1}^*, x_j^*) fällt, wenn ihr die hypothetische Verteilung $F_0(x)$ zugrunde liegt, d.h. wenn die Nullhypothese H_0 gilt:

$$p_j = P(x_{j-1}^* < X \leq x_j^* | H_0) .$$

Die Tatsache, dass die beobachteten absoluten Häufigkeiten Zufallsvariablen H_j sind, lässt sich wie folgt zeigen. Aus der Grundgesamtheit wird ein Element zufällig gezogen und festgestellt, ob das Ereignis $\{X = x_j | x_j \in (x_{j-1}^*, x_j^*)\}$ eingetreten ist oder nicht. Es gibt somit nur zwei mögliche Ergebnisse des Zufallsexperimentes. Die Wahrscheinlichkeit für das Eintreten des Ereignisses $\{X = x_j | x_j \in (x_{j-1}^*, x_j^*)\}$ beträgt bei Gültigkeit der Nullhypothese p_j und die Wahrscheinlichkeit für das Nichteintreten $1 - p_j$. Es liegt somit ein Bernoulli-Experiment vor. Das Zufallsexperiment wird n mal wiederholt, wobei die einzelnen Versuche unabhängig voneinander sind. Bei n -maliger Durchführung der Versuche interessiert die Gesamtzahl des Eintretens von $\{X = x_j | x_j \in (x_{j-1}^*, x_j^*)\}$, d.h. die absolute Häufigkeit H_j in der Stichprobe. Diese Häufigkeit kann von Stichprobe zu Stichprobe unterschiedlich sein, so dass H_j eine diskrete Zufallsvariable ist, die die Werte $0, \dots, n$ annehmen kann. Die Zufallsvariable H_j ist binomialverteilt und zwar bei Gültigkeit der Nullhypothese H_0 mit den Parametern n und p_j : $H_j \sim \text{Bin}(n, p_j)$. Der Erwartungswert von H_j ist $E(H_j) = n \cdot p_j$ und damit die bei Gültigkeit der Nullhypothese erwartete Häufigkeit des Ereignisses $\{X = x_j | x_j \in (x_{j-1}^*, x_j^*)\}$ in der Stichprobe. Die Variation der absoluten Häufigkeiten für $\{X = x_j | x_j \in (x_{j-1}^*, x_j^*)\}$ wird durch die Varianz $\text{Var}(H_j) = np_j(1 - p_j)$ erfasst.

Für die Konstruktion der Teststatistik wird die Abweichung der Zufallsvariable von ihrem Erwartungswert gebildet:

$$H_j - n \cdot p_j .$$

Zur Vermeidung, dass sich positive und negative Abweichungen aufheben, erfolgt eine Quadrierung: $(H_j - n \cdot p_j)^2$. Mit der Division durch die erwartete Häufigkeit

$n \cdot p_j$ wird der Einfluss des Stichprobenumfanges und der Wahrscheinlichkeit p_j berücksichtigt und der unterschiedlichen Bedeutung der Abweichungen Rechnung getragen. Eine Differenz $h_j - n \cdot p_j = 5$ fällt bei $n \cdot p_j = 10$ stärker ins Gewicht als bei $n \cdot p_j = 100$. Weiter beobachten wir, dass für kleine p_j und n genügend gross gilt, dass $\text{Var}(X) \approx n \cdot p_j$ und somit

$$\frac{(H_j - n \cdot p_j)}{\sqrt{n \cdot p_j}} \sim \mathcal{N}(0, 1).$$

Betrachten wir folgende Teststatistik

$$V = \sum_{j=1}^k \frac{(H_j - n \cdot p_j)^2}{n \cdot p_j}$$

so ist auch V eine Zufallsvariable, da die H_j Zufallsvariablen sind. Bei Gültigkeit der Nullhypothese, hinreichend grossem Stichprobenumfang n und Einhaltung der Approximationsbedingungen ($n \cdot p_j \geq 5$) ist die Teststatistik V approximativ χ^2 -verteilt mit $f = k - m - 1$ Freiheitsgraden. Dies gilt unabhängig davon, welche Verteilung unter H_0 angenommen wurde. Sind die Approximationsbedingungen nicht erfüllt, müssen vor der Anwendung des Tests benachbarte Werte bzw. Klassen zusammengefasst werden.

Bei der Ermittlung der Freiheitsgrade ist zu berücksichtigen, dass ein Freiheitsgrad grundsätzlich verloren geht, weil die beobachteten absoluten Häufigkeiten nicht unabhängig voneinander sind. Für vorgegebenen Stichprobenumfang n und aufgrund der Bedingung $\sum_j h_j = n$ folgt, dass jede Häufigkeit h_j durch die anderen $k - 1$ Häufigkeiten bestimmt ist. Weitere Freiheitsgrade gehen verloren, wenn die hypothetische Verteilung $F_0(x)$ nicht mit allen ihren Parametern bekannt ist, sondern diese Parameter aus der Stichprobe geschätzt werden müssen. Mit m als Anzahl der zu schätzenden Parameter ergibt sich die Anzahl der Freiheitsgrade zu: $f = k - m - 1$. Wurden keine Parameter geschätzt, so ist die Anzahl Freiheitsgrade gegeben durch die Anzahl Klassen minus 1.

Bei einem Signifikanzniveau α wird H_0 abgelehnt, wenn $V > \chi^2_{(1-\alpha; k-m-1)}$ gilt, wenn also der aus der Stichprobe erhaltene Wert der Prüfgrösse grösser als das $(1 - \alpha)$ -Quantil der χ^2 -Verteilung mit $k - m - 1$ Freiheitsgraden ist.

Beispiel B.3.1

In einem berühmten Artikel, den Rutherford, Chadwick und Ellis 1920 publiziert hatten, befindet sich die Tabelle B.1 mit Messwerten.

Es wurde die Anzahl der von einer radioaktiven Substanz emittierten Teilchen in einem Zeitintervall von 7.5 Sekunden gemessen. Im gesamten wurden die Anzahlen Zerfälle in 2608 Zeitintervallen beobachtet. Nun wurde postuliert, dass die Anzahl

Anhang B. Aus der Normalverteilung hergeleitete Verteilungen

| Anzahl Zerfälle | Beobachtet | Erwartet | $(O - E)^2 / E$ | |
|-----------------|------------|----------|-----------------|--|
| 0 | 57 | 54.40 | 0.12 | |
| 1 | 203 | 210.50 | 0.27 | |
| 2 | 383 | 407.40 | 1.46 | |
| 3 | 525 | 525.50 | 0 | |
| 4 | 532 | 508.40 | 1.10 | |
| 5 | 408 | 393.50 | 0.053 | |
| 6 | 273 | 253.80 | 1.45 | |
| 7 | 139 | 140.30 | 0.01 | |
| 8 | 45 | 67.80 | 7.67 | |
| 9 | 27 | 29.20 | 0.17 | |
| 10 | 10 | 11.30 | 0.15 | |
| > 11 | 6 | 5.80 | 0.01 | |

Tabelle B.1.: Anzahl Zerfälle in 7.5 Sekunden und Anzahl Experimente (von insgesamt 2608), in denen diese Anzahl Zerfälle beobachtet wurde.

emittierter Teilchen pro Zeitintervall eine Zufallsvariable mit Poisson-Verteilung ist. Um diese Vermutung zu überprüfen, führen wir einen Chi-Quadrat-Test durch.

In diesem Fall, ist der Erwartungswert der Poisson-Verteilung unbekannt. Wir schätzen den Erwartungswert, indem wir den empirischen Mittelwert zu 3.870 ermitteln. Für D^2 mit $12 - 1 - 1 = 10$ Freiheitsgraden ergibt sich dann aus dem Datensatz der Wert 12.94. Der Verwerfungsbereich auf dem 1 % Signifikanzniveau beginnt bei 23.2. Der von uns ermittelte Wert überschreitet diese Schwelle nicht, so dass wir daraus schliessen können, dass die Anzahl emittierter Teilchen pro Zeitintervall in guter Übereinstimmung mit der Poisson-Verteilung ist.

Mit R ermittelt sich der das 99 %-Quantil mit

```
qchisq(0.99, df = 10)
```

```
## [1] 23.20925
```

der P-Wert wie folgt

Anhang B. Aus der Normalverteilung hergeleitete Verteilungen

```
1 - pchisq(12.93714, df = 10)

## [1] 0.2272077
```

Der Chi-Quadrat Test kann auch folgendermassen in R ausgeführt werden

```
beobachtet <- c(57, 203, 383, 525, 532, 408, 273, 139, 45,
  27, 10, 6)
erwartet <- c(54.4, 210.5, 407.4, 525.5, 508.4, 393.5, 253.8,
  140.3, 67.8, 29.2, 11.3, 5.8)
chisq.test(beobachtet, erwartet)

##
## Pearson's Chi-squared test
##
## data: beobachtet and erwartet
## X-squared = 132, df = 121, p-value = 0.2329
```

□

Anhang C.

Signaltheorie und das Wiener-Khintchine Theorem

C.1. Herleitung der Lösung der Diffusionsgleichung aus Random Walk

Wir können den Zusammenhang des makroskopisch beschriebenen Phänomens der **Diffusion** mit dem mikroskopischen Phänomen der **Brownschen Bewegung** eines Partikels verstehen, indem wir zu einer Kontinuumsbeschreibung des Random Walks übergehen und die Schrittlänge Δx sowie den zeitlichen Abstand Δt immer kleiner machen. Insbesondere können wir die zeitabhängige Teilchendichtefunktion

$$n(x, t) = \frac{n}{\sqrt{4\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}} \quad (\text{C.1})$$

aus der Punktwahrscheinlichkeit für die Position des betrunkenen Bargängers

$$P(M_N = m) = \frac{2}{\sqrt{8\pi Npq}} e^{-\frac{(m-N(p-q))^2}{8Npq}} \quad (\text{C.2})$$

herleiten. Wir definieren

$$x = m\Delta x \quad \text{und} \quad t = N\Delta t$$

und führen die Parameter

$$D \equiv 2pq \frac{(\Delta x)^2}{\Delta t} \quad \text{und} \quad v \equiv (p - q) \frac{\Delta x}{\Delta t}$$

ein. D wird **Diffusionskonstante** und v wird **mittlere Driftgeschwindigkeit** genannt. Dann können wir die Normalverteilung in Gleichung (C.2) in die folgende Form bringen

$$P(X = m\Delta x; N\Delta t) = \frac{2\Delta x}{\sqrt{4\pi Dt}} e^{-\frac{(x-vt)^2}{4Dt}}.$$

Wir führen nun den Grenzübergang $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$ bei konstantem D und konstantem v durch. Dies bedeutet, dass wir die beiden „makroskopischen“ Charakteristika des Zufallspfad, die mittlere Driftgeschwindigkeit v und die Diffusionskonstante D vorgeben, aber Raum und Zeit kontinuierlich machen. Bei diesem Prozess schrumpft Δx mit der Wurzel von Δt , denn bei konstantem D gilt $(\Delta x)^2 \propto \Delta t$ oder $\Delta x \propto \sqrt{\Delta t}$. Andererseits muss bei konstantem v gelten, dass die Parameter p und q sich so ändern, dass $p - q \propto \Delta x$ ist. Für $\Delta x \rightarrow 0$ nähern sich also beide Parameter p und q dem Wert $1/2$ an, wobei v konstant bleibt. Für $\Delta x \rightarrow 0$ wird aber auch die Punktwahrscheinlichkeit $P(X = m\Delta x; N\Delta t)$ null, was wir bei einer kontinuierlichen Wahrscheinlichkeitsverteilung natürlich erwarten. Nun entspricht der Ausdruck $P(X = m\Delta x; N\Delta t)$ für die diskrete Zufallsvariable X der Wahrscheinlichkeit

$$\begin{aligned} P(x < X \leq x + \Delta x; t) &= F(x + \Delta x, t) - F(x, t) \\ &= \Delta F(x, t) \\ &= \frac{2\Delta x}{\sqrt{4\pi Dt}} e^{-\frac{(x-vt)^2}{4Dt}}, \end{aligned}$$

also der Differenz der kumulativen Wahrscheinlichkeitsfunktion. Wir erhalten dann folgende Wahrscheinlichkeitsdichte für die Position x des Brownschen Partikels zur Zeit t

$$\frac{\Delta F(x, t)}{2\Delta x} \xrightarrow{\Delta x \rightarrow 0} f(x; t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{(x-vt)^2}{4Dt}}, \quad (\text{C.3})$$

Die Position des Brownschen Partikels zur Zeit $t = 0$ bezeichnen wir mit x_0 . In unserer Herleitung aus dem Random Walk haben wir stets $x_0 = 0$ angenommen. Falls dies nicht der Fall ist, ersetzen wir x einfach durch $x - x_0$. Der Parameter D wird **Diffusionskonstante** und v wird **mittlere Driftgeschwindigkeit** genannt.

Nehmen wir einfachheitshalber an, dass die Driftgeschwindigkeit verschwindet (falls $p = q$, dann ist $v = 0$) und die Ausgangsposition x_0 nicht null sein muss, dann ergibt sich die Teilchendichtefunktion (C.1) nun, indem wir die n Brownschen Teilchen mit $f(x; t)$ multiplizieren:

$$n(x, t) = n \cdot f(x; t) = \frac{n}{\sqrt{4\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}} \quad (\text{C.4})$$

Diffusionsgleichung

Die Teilchendichte

$$n(x, t) = \frac{n}{\sqrt{4\pi Dt}} e^{-\frac{(x-x_0)^2}{4Dt}}$$

löst die **Diffusionsgleichung**

$$\frac{\partial n(x, t)}{\partial t} = D \frac{\partial^2 n(x, t)}{\partial x^2}, \quad (\text{C.5})$$

mit der Randbedingung, dass sich alle Teilchen zum Zeitpunkt $t = 0$ bei $x = 0$ befinden. Der Random Walk in der Kontinuumsbeschreibung wird **Brownsche Bewegung** genannt und kann also als **Diffusionsprozess** aufgefasst werden.

Die beiden Terme auf der rechten Seite von Gleichung (C.5) haben eine einfache anschauliche Bedeutung: Der erste Term (mit v) verschiebt die Wahrscheinlichkeitsdichte $f(x, t)$ mit Geschwindigkeit v nach rechts (wenn v positiv ist; sonst wird die Verteilung nach links geschoben). Man nennt diesen Term den **Driftterm**. Für $p = q$ ist $v = 0$, und der Driftterm fällt weg. Der Term auf der rechten Seite von Gleichung (C.5) macht die Verteilung breiter; er wird deshalb **Diffusionsterm** genannt. Physikalisch stellen wir uns einen Tintentropfen vor, der in heisses Wasser gegeben wird: durch Diffusion verteilt sich die Tinte im Wasser aufgrund der Stösse zwischen den sich zufällig bewegenden Wassermolekülen und Tintenmolekülen.

C.2. Korrelation von Signalen und Wiener-Khintchine-Theorem

Eine der interessantesten Fragen in der Signalverarbeitung lautet: Wie ähnlich sind zwei Signale? Diese Frage stellt sich immer dann, wenn man ein bekanntes Muster in einem Signal wiederfinden will, sei es um einen Funkempfänger mit dem Trägersignal einer Übertragung zu synchronisieren, oder um in einer Sprachaufnahme Worte herauszufiltern. Ein Konzept, das hierfür zum Einsatz kommt, heisst Korrelation („Zusammenhang“).

C.3. Energie, Leistung, Korrelation

Die Energie eines Spannungssignal in einem Stromkreis mit einem bekannten Widerstand R ist

$$E = \int P(t) dt = \int U(t)I(t) dt = \frac{1}{R} \int U^2(t) dt.$$

Sie hängt also vom Quadrat der Spannung ab.

Signalenergie

In diesem Sinne definiert man die **Signalenergie** als Integral über das Quadrat eines Signals:

$$E_s = \int_{t_1}^{t_2} |s(t)|^2 dt ,$$

wobei der Betrag gebildet wird, damit die Formel auch für komplexwertige Signale gültig bleibt. Falls $s(t)$ ein periodisches Signal ist, bietet es sich an, über eine Periode zu integrieren. Um einen vergleichbaren Wert zu erhalten, wenn man über unterschiedliche Zeitintervalle integriert, normiert man die Energie mit der Zeit, über die man integriert.

Signalleistung

Wir definieren deshalb die **Signalleistung** als

$$P_s = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} |s(t)|^2 dt .$$

C.3.1. Ähnlichkeitsmass für Signale

Die mittlere quadratische Abweichung zweier Signale $s(t)$ und $g(t)$ ist

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} [s(t) - g(t)]^2 dt &= \lim_{T \rightarrow \infty} \frac{1}{T} \left(\int_{-T/2}^{T/2} s(t)^2 dt + \int_{-T/2}^{T/2} g(t)^2 dt - 2 \int_{-T/2}^{T/2} s(t)g(t) dt \right) \\ &= P_s + P_g - 2 \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t)g(t) dt \end{aligned}$$

Beim Vergleich zweier Signale interessiert man sich nun nicht für absolute Leistungen, so dass bloss der letzte Summand im obigen Ausdruck von Interesse ist.

Kreuzkorrelationskoeffizient

Ein Mass für die Ähnlichkeit zweier reellwertiger Signale $s(t)$ und $g(t)$, das von absoluten Leistungswerten unabhängig ist, ist der **Kreuzkorrelationskoeffizient**, den wir definieren als

$$\Phi_{sg} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s^*(t)g(t) dt ,$$

wobei $s^*(t)$ das zu $s(t)$ komplex konjugierte Signal bezeichnet.

Der Kreuzkorrelationskoeffizient wird genau dann null, wenn zwischen zwei Signalen kein linearer Zusammenhang besteht. Diese Definition der Korrelation ist nur dann nützlich, wenn man eine feste zeitliche Lage der Funktionen zueinander annimmt. Können stochastische Signale auch verschoben zueinander auftreten, wie es in der Technik in der Regel der Fall ist, macht es Sinn, den Kreuzkorrelationskoeffizienten zu einer Funktion der Zeitverschiebung zu erweitern.

Kreuzkorrelationsfunktion

Wir definieren die **Kreuzkorrelationsfunktion** für **stationäre** stochastische Signale $s(t)$ und $g(t)$

$$\varphi_{sg}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s^*(t)g(t + \tau) dt .$$

Den Kreuzkorrelationskoeffizienten erhält man aus der Kreuzkorrelationsfunktion ganz trivial mit $\Phi_{sg} = \varphi_{sg}(0)$. Eine der häufigeren Anwendungen dieser Definition ist der Vergleich eines Signals mit sich selbst, genannt Autokorrelation.

Autokorrelationsfunktion

Die **Autokorrelationsfunktion** (AKF) ist definiert als

$$\varphi_{ss}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s^*(t)s(t + \tau) dt .$$

Insbesondere gilt

$$\varphi_{ss}(0) = \Phi_{ss} = P_s .$$

Zu den Anwendungen der AKF gehört einerseits festzustellen, wie repetitiv ein Signal ist, d.h. ob Periodizitäten vorliegen. In diesem Fall hat die AKF auch an anderen Stellen als $\varphi(0)$ Werte, die von null verschieden sind. Andererseits ist nur ein Hintergrundrauschen dermassen zufällig, dass Periodizitäten nicht auftreten und sich das Rauschsignal nie wiederholt. Man kann ein störendes Rauschen also eben daran erkennen, dass die AKF an der Nullstelle einen sehr scharfen Peak hat, sonst überall verschwindet.

C.4. Korrelation und Faltung

Im folgenden nehmen wir an, dass die Signale $s(t)$ und $g(t)$ reellwertig sind, also $s^* = s(t)$. Weiter nehmen wir an, dass wir nur über ein endliches Zeitintervall T integrieren. Wir schreiben also im folgenden

$$\varphi_{sg}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} = \int_{-T/2}^{T/2} s(t)g(t+\tau) dt \approx \frac{1}{T} \int_{-T/2}^{T/2} s(t)g(t+\tau) dt = \varphi_{sg;T}(\tau)$$

Man stellt bei der Kreuzkorrelationsfunktion eine gewisse Ähnlichkeit zur Faltung fest:

$$\begin{aligned} \varphi_{sg;T}(\tau) &= \frac{1}{T} \int_{-T/2}^{T/2} s(t)g(t+\tau) dt \\ &= \frac{1}{T} \int_{-T/2}^{T/2} \underbrace{s(-t')g(\tau - t')}_{s(-\tau)*g(\tau)} dt' \end{aligned}$$

wobei wir folgende Substitution verwendet haben: $t' = -t$.

C.4.1. Kreuzkorrelationstheorem

Es bietet sich nun geradezu an, die Signale mit Hilfe der Fourier-Transformation in den Bildraum zu überführen. Im Bildraum berechnet sich die Faltung als Produkt der

Fourier-Transformierten der Signale. Es ergibt sich:

$$\begin{aligned}
 \Phi_{sg;T}(\omega) &\equiv \int_{-\infty}^{\infty} \varphi_{sg;T}(\tau) e^{-j\omega\tau} d\tau \\
 &= \mathcal{F}\{\varphi_{sg;T}(\tau)\} \\
 &= \frac{1}{T} \mathcal{F}\{s(-\tau) \star g(\tau)\} \\
 &= \frac{1}{T} \mathcal{F}\{s(-\tau)\} \cdot \mathcal{F}\{g(\tau)\} \\
 &= \frac{1}{T} S^*(\omega) \cdot G(\omega),
 \end{aligned}$$

wobei $S^*(\omega)$ die komplex konjugierte Funktion von $S(\omega)$ bezeichnet und T die Integrationszeit für die Faltung ist. Wir haben dabei benützt, dass für ein reellwertiges $s(t)$ gilt

$$\begin{aligned}
 \mathcal{F}\{s(-t)\} &= \int_{-\infty}^{\infty} s(-t) e^{-j\omega t} dt \\
 &= \int_{-\infty}^{\infty} s(t') e^{+j\omega t'} dt' \\
 &= S^*(\omega),
 \end{aligned}$$

Kreuzkorrelationstheorem

Das **Kreuzkorrelationsfunktionstheorem** lautet also

$$\Phi_{sg}(\omega) = \frac{1}{T} \cdot S^*(\omega) \cdot G(\omega).$$

C.5. Wiener-Khintchine Theorem

Wiener-Khintchine Theorem

Das Kreuzkorrelationstheorem ergibt im Falle der Autokorrelationsfunktion, also im Fall $s(\tau) = g(\tau)$, das Wiener-Khintchine-Theorem

$$\mathcal{F}\{\varphi_{ss;T}(\tau)\} = \Phi_{ss;T}(\omega) = \frac{1}{T} \cdot S^*(\omega) \cdot S(\omega) = \frac{1}{T} |S(\omega)|^2$$

$\frac{|S(\omega)|^2}{T}$ wird das **Leistungsdichtespektrum** des Signals $s(t)$ genannt. Die Fourier-Transformierte der Autokorrelationsfunktion $\varphi_{ss;T}(\tau)$ ergibt also das Leistungsdichtespektrum des Signals.

Der Grund, warum $\frac{|S(\omega)|^2}{T}$ Leistungsdichtespektrum genannt wird, wird aus folgender Rechnung ersichtlich

$$\begin{aligned} \varphi_{ss;T}(\tau = 0) &= \mathcal{F}^{-1}\{\Phi_{ss;T}(\omega)\} \Big|_{\tau=0} \\ &= \mathcal{F}^{-1}\left\{\frac{1}{T} |S(\omega)|^2\right\} \Big|_{\tau=0} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{T} e^{j\omega\tau} d\omega \Big|_{\tau=0} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{T} d\omega \end{aligned}$$

Erinnern wir uns an die Definition der Signalleistung, so ergibt sich¹

$$\varphi_{ss}(0) = P_s = \frac{1}{T} \int_{-T/2}^{T/2} |s(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{T} d\omega .$$

$\frac{1}{T} |S(\omega)|^2 d\omega$ gibt also die Teilleistung des Signals an, die auf das Frequenzband $[\omega, \omega + d\omega]$ entfällt. Wenn man ein Spannungssignal in einem realen Stromkreis betrachtet, dann hat die Autokorrelationsfunktion im Frequenzbereich die Einheit $1/\Omega \cdot V^2 s^3 = W/Hz$.

¹Dies ist im übrigen nichts anderes als das Parseval'sche Theorem.

Signalleistung

Die **Signalleistung** P_s eines stochastischen Signals $S(t)$ berechnet sich mit Hilfe des Wiener-Khintchine Theorems gemäss

$$P_s = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{ss}(\omega) d\omega ,$$

wobei $\Phi_{ss}(\omega)$ die Fourier-Transformierte der Autokorrelationsfunktion $\varphi_{ss}(\tau)$ ist.

Beispiel C.5.1

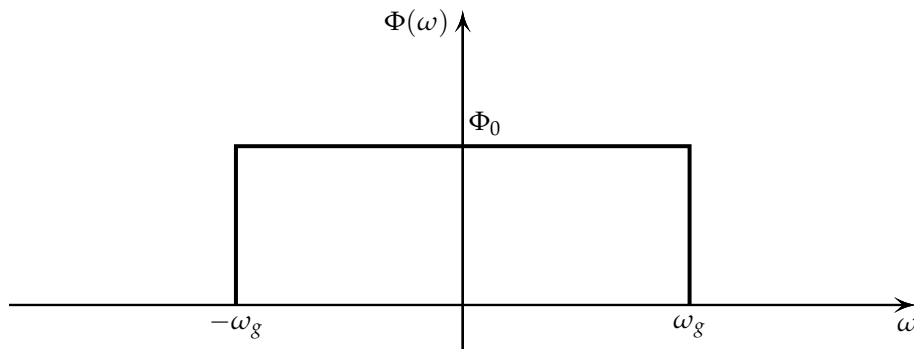


Abbildung C.1.: Leistungsdichtespektrum von weissem bandbegrenztem Rauschen.

Wir können nun das Wiener-Khintchine-Theorem auf das Beispiel des Gauss'schen Rauschsignals anwenden. Wir nehmen für das Rauschsignal $N(t)$ weisses, bei der Kreisfrequenz ω_g bandbegrenztes Rauschen an – „weiss“ bedeutet ein Leistungsdichtespektrum $\Phi(\omega)$, das konstant ist über alle vorkommenden Frequenzen, also $\Phi(\omega) = \Phi_0$. Dies trifft auf das Leistungsspektrum $\Phi(\omega)$ gemäss der Abbildung C.1 zu. Aus $\Phi(\omega)$ lässt sich nun die Autokorrelationsfunktion $\varphi(\tau)$ folgendermassen be-

rechnen:

$$\begin{aligned}
 \varphi_n(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(\omega) e^{j\omega\tau} d\omega = \frac{1}{2\pi} \int_{-\omega_g}^{\omega_g} \Phi_0 e^{j\omega\tau} d\omega \\
 &= \frac{\Phi_0}{2\pi} \left[\frac{e^{j\omega\tau}}{j\tau} \right]_{-\omega_g}^{\omega_g} = \frac{\Phi_0}{2j\pi\tau} \left[e^{j\omega_g\tau} - e^{-j\omega_g\tau} \right] \\
 &= \Phi_0 \frac{\omega_g}{\pi} \frac{\sin(\omega_g\tau)}{\omega_g\tau}.
 \end{aligned} \tag{C.6}$$

□

Beispiel C.5.2

Wir betrachten im folgenden das Rechtecksignal

$$s(t) = -A\sigma(t + T_0/2) + 2A\sigma(t) - A\sigma(t - T_0/2).$$

Die Fourier-Transformierte des Signals $s(t)$ ist gegeben durch

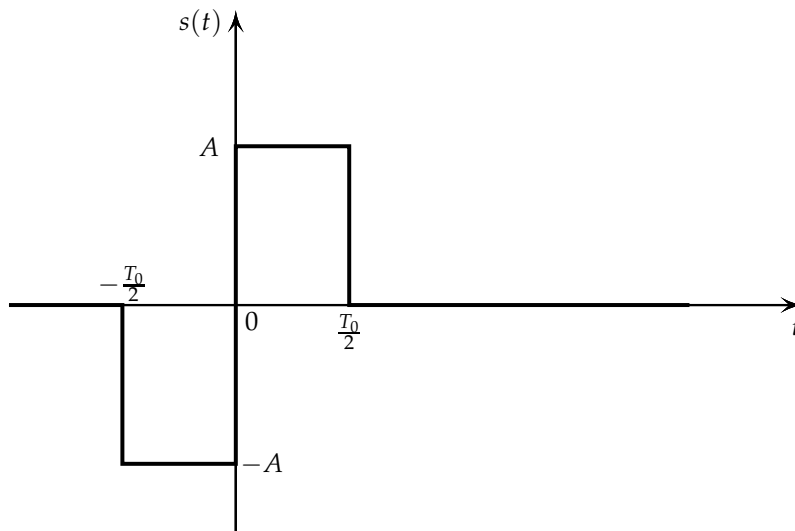


Abbildung C.2.: Rechtecksignal $s(t) = -A\sigma(t + T_0/2) + 2A\sigma(t) - A\sigma(t - T_0/2)$.

$$\begin{aligned}
 S(\omega) &= \int_{-\infty}^{\infty} \left(-A\sigma(t + T_0/2) + 2A\sigma(t) - A\sigma(t - T_0/2) \right) e^{-j\omega t} dt \\
 &= \int_{-T_0/2}^0 -A \cdot e^{-j\omega t} dt + \int_0^{T_0/2} A \cdot e^{-j\omega t} dt \\
 &= \frac{A}{j\omega} \left(1 - e^{+j\omega T_0/2} \right) + \frac{A}{j\omega} \left(1 - e^{-j\omega T_0/2} \right) \\
 &= \frac{A}{j\omega} \left(2 - e^{+j\omega T_0/2} - e^{-j\omega T_0/2} \right) \\
 &= \frac{2A}{j\omega} \left(1 - \cos(\omega T_0/2) \right) \\
 &= \frac{4A}{j\omega} \sin^2(\omega T_0/4)
 \end{aligned}$$

Die Fourier-Transformierte eines reellwertigen Signals $s(-t)$ ist gegeben durch

$$\begin{aligned}
 \mathcal{F}\{s(-t)\} &= \int_{-\infty}^{\infty} s(-t) e^{-j\omega t} dt \\
 &= \int_{-\infty}^{\infty} s(t') e^{+j\omega t'} dt' \\
 &= S^*(\omega),
 \end{aligned}$$

wobei wir mit $S^*(\omega)$ die zu $S(\omega)$ komplex konjugierte Funktion bezeichnen. Somit ist

$$\begin{aligned}
 \mathcal{F}\{s(-t)\} &= \frac{-A}{j\omega} \left(2 - e^{-j\omega T_0/2} - e^{+j\omega T_0/2} \right) \\
 &= \frac{-2A}{j\omega} \left(1 - \cos(\omega T_0/2) \right) \\
 &= \frac{-4A}{j\omega} \sin^2(\omega T_0/4)
 \end{aligned}$$

Im Bildbereich hat die Autokorrelationsfunktion $\varphi(\tau)$ folgende Form

$$\begin{aligned}\mathcal{F}\{\varphi(\tau)\} &= \frac{1}{T_0} S^*(\omega) \cdot S(\omega) \\ &= \frac{A^2}{\omega^2 T_0} \left(2 - e^{+j\omega T_0/2} - e^{-j\omega T_0/2} \right)^2 \\ &= \frac{A^2}{\omega^2 T_0} \left(6 - 4e^{-j\omega T_0/2} - 4e^{+j\omega T_0/2} + e^{-j\omega T_0} + e^{j\omega T_0} \right).\end{aligned}$$

Transformieren wir zurück in den Zeitraum, finden wir

$$\begin{aligned}\varphi(\tau) &= \frac{A^2}{T_0} \left(-6\tau\sigma(\tau) + 4(\tau - T_0/2)\sigma(\tau - T_0/2) + 4(\tau + T_0/2)\sigma(\tau + T_0/2) \right. \\ &\quad \left. - (\tau - T_0)\sigma(\tau - T_0) - (\tau + T_0)\sigma(\tau + T_0) \right) \\ &= \begin{cases} -A^2(1 + \tau/T_0), & -T_0 \leq \tau \leq -T_0/2 \\ +A^2(1 + 3\tau/T_0), & -T_0/2 \leq \tau \leq 0 \\ +A^2(1 - 3\tau/T_0), & 0 \leq \tau \leq T_0/2 \\ +A^2(-1 + \tau/T_0), & T_0/2 \leq \tau \leq T_0. \end{cases}\end{aligned}$$

Den Graphen sehen Sie in Abbildung C.3.

□

Beispiel C.5.3

Wir betrachten im folgenden das T_0 -periodische Rechtecksignal. Das T_0 -periodische Signal kann als Fourier-Reihe geschrieben werden:

$$s(t) = \sum_{k=-\infty}^{+\infty} c_k e^{jk\omega_0 t},$$

wobei $\omega_0 = 2\pi/T_0$ und

$$c_k = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} s(t) e^{-jk\omega_0 t} dt.$$

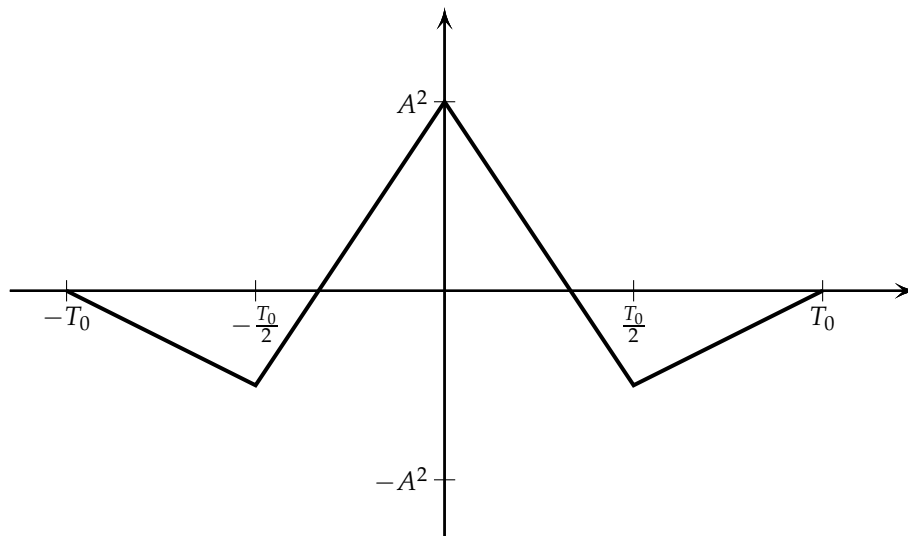


Abbildung C.3.: Rücktransformation in den Zeitraum

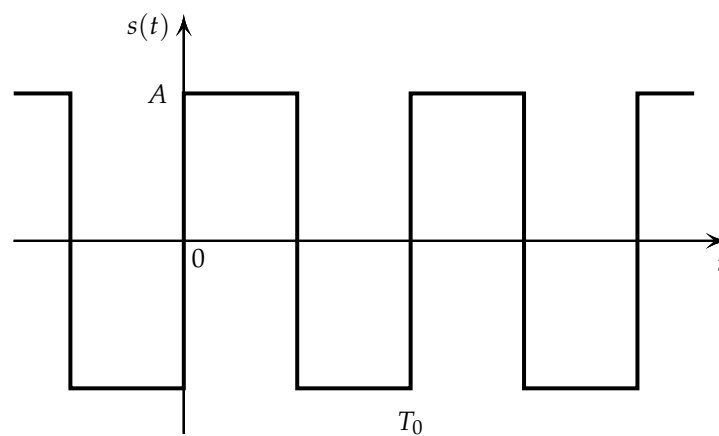


Abbildung C.4.: Das T_0 -periodische Rechtecksignal

Aus Beispiel B folgt

$$\begin{aligned}
 c_k &= \frac{4A}{jk\omega_0 T_0} \sin^2(k\omega_0 T_0/4) \\
 &= \frac{4A}{jk\omega_0 T_0} \sin^2(k\pi/2) \\
 &= \begin{cases} \frac{2A}{jk\pi} & k \text{ ungerade} \\ 0 & k \text{ gerade} \end{cases}
 \end{aligned}$$

Somit gilt für jedes $k \in \mathbb{N}_0$

$$c_{2k} = 0, \quad c_{2k-1} = \frac{2A}{j\pi} \cdot \frac{1}{2k-1}.$$

Die Fourier-Reihe für das T_0 -periodische Rechtecksignal lautet also:

$$s(t) = \frac{2A}{j\pi} \sum_{k=-\infty}^{+\infty} \frac{1}{2k-1} e^{j(2k-1)\omega_0 t},$$

Die Autokorrelationsfunktion für eine T_0 -periodische Funktion ist definiert als

$$\begin{aligned} \varphi_{ss}(\tau) &= \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} s(t)s(t+\tau) dt \\ &= \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \left(\frac{2A}{j\pi} \sum_{k=-\infty}^{+\infty} \frac{1}{2k-1} e^{j(2k-1)\omega_0 t} \right) \cdot \left(\frac{2A}{j\pi} \sum_{m=-\infty}^{+\infty} \frac{1}{2m-1} e^{j(2m-1)\omega_0(t+\tau)} \right) dt \\ &= \frac{-4A^2}{\pi^2} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} \left(\sum_{k=-\infty}^{+\infty} \frac{1}{2k-1} \right) \cdot \left(\sum_{m=-\infty}^{+\infty} \frac{1}{2m-1} \right) e^{j(2k-1)\omega_0 t} e^{j(2m-1)\omega_0(t+\tau)} dt \\ &= \frac{-4A^2}{\pi^2} \left(\sum_{k=-\infty}^{+\infty} \frac{1}{2k-1} \right) \cdot \left(\sum_{m=-\infty}^{+\infty} \frac{1}{2m-1} \right) e^{j(2m-1)\omega_0 \tau} \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} e^{j(2k+2m-2)\omega_0 t} dt \end{aligned}$$

Nun gilt für $m, n \in \mathbb{N}_0$

$$\frac{1}{T_0} \int_{-T_0/2}^{T_0/2} e^{j(2k+2m-2)\omega_0 t} dt = \begin{cases} 0, & \text{für } 2m+2k-2 \neq 0 \\ 1, & \text{für } 2m+2k-2 = 0 \end{cases}$$

Beim Produkt der unendlichen Reihe bleiben nur die Term übrig, die $k+m-1=0$ erfüllen. Also finden wir

$$\begin{aligned} \varphi_{ss}(\tau) &= \frac{-4A^2}{\pi^2} \sum_{m,k: m+k=1} \left(\frac{1}{2m-1} \cdot \frac{1}{2k-1} \right) e^{j(2m-1)\omega_0 \tau} \\ &= \frac{4A^2}{\pi^2} \sum_{m=-\infty}^{\infty} \left(\frac{1}{2m-1} \right)^2 e^{j(2m-1)\omega_0 \tau}. \end{aligned}$$

Die Autokorrelationsfunktion ergibt dann die T_0 -periodische Fortsetzung von

$$\varphi(\tau) = \begin{cases} A^2(1 - 4\frac{\tau}{T_0}), & 0 \leq \tau \leq \frac{T_0}{2} \\ A^2(1 - 4\frac{\tau}{T_0}), & -\frac{T_0}{2} \leq \tau \leq 0. \end{cases}$$

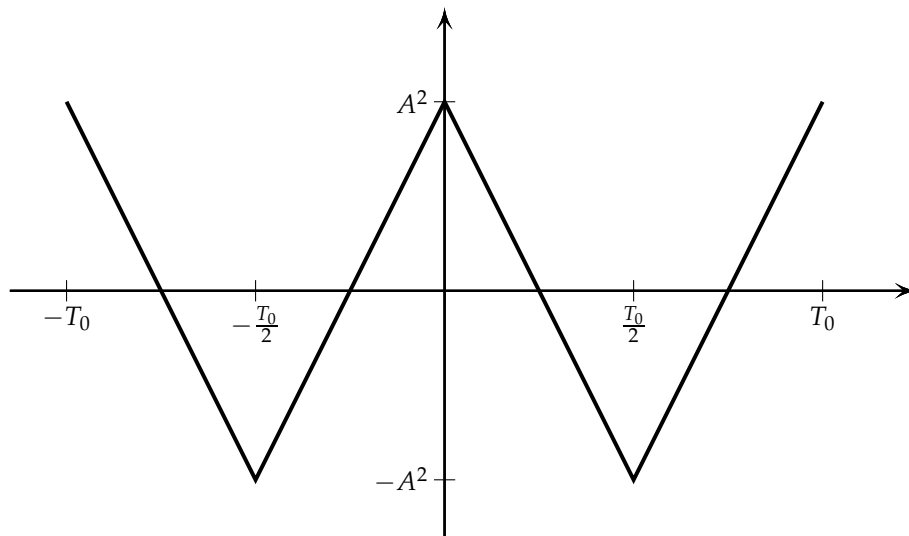


Abbildung C.5.: T_0 -periodische Fortsetzung

Den Graphen sehen Sie in [Abbildung C.5](#).

□

Index

- P*-Wert, 114, 131
- \sqrt{n} -Gesetz, 173
- t*-Test, 199, 217
- t*-Verteilung, 200, 277, 278
- z*-Test, 195
- Gesetz der grossen Zahlen**, 173
- Zentraler Grenzwertsatz (ZGWS)**, 174

- Abstand
 - quadratisch gemittelte, 229
- Abweichung, 13
 - mittlere absolute, 13
- Alphateilchen, 46
- Alternativhypothese, 119
- Aphastrahler, 46
- Autokorrelationsfunktion (AKF), 233, 244, 288
 - Kurzzeit-, 246
- Axiom
 - Kolmogorov, 53
- Bernoulli-Verteilung, 80
- Bewegung
 - Brownsche, 226, 286
- Binomialkoeffizient, 80, 83, 260
- Binomialverteilung, 78, 81, 95, 100, 103, 265
- Boltzmann-Verteilung, 168
- Boxplot, 23

- Chi-Quadrat Test, 280
- Chi-Quadrat-Verteilung, 273, 275

- Daten, 131
- De Morgan'sche Regeln, 51
- Diffusion, 227, 284
- Diffusionsgleichung, 285
- Diffusionskoeffizient, 229
- Diffusionskonstante, 228, 284, 285
- Diffusionsprozess, 237, 286
- Diffusionsterm, 286
- disjunkt, 51
- Driftgeschwindigkeit, 228, 284, 285

- Einstein-Smoluchowski-Gleichung, 229
- Elementarereignis, 48
 - günstig, 57
 - möglich, 57
- Energie, 286
- Ensemble, 234
- Ereignis, 48, 50
- Ergodenhypothese, 242
- Ergodizität, 238
- Erwartungswert, 88, 149, 150, 261, 264
- Exponentialverteilung, 153, 154

- Fehler
 - 1. Art, 125, 199
 - 2. Art, 125, 127, 199
 - absoluter, 178
 - relativer, 178
- Fehlerrechnung, 177
- Freiheitsgrad, 273, 275
- Funktion, 74
 - charakteristische, 266, 267
 - momentengenerierende, 275

- Gammaverteilung, 274
- Gesetz
 - der totalen Wahrscheinlichkeit, 72
- Grundraum, 48

- Häufigkeit
 - absolute, 47

- relative, 48
- Häufigkeitsverteilung, 47
- Halbwertszeit, 156
- Histogramm, 20
- Interpretation, 54
 - Bayes'sche, 54
 - frequentistische, 54
- Kennzahlen, 149, 171, 172
 - einer Verteilung, 88
- Klasse, 20
- Konfidenzintervall, 99, 134
- Korrelation, 286
 - empirische, 41
- Korrelogramm, 247, 258
- Kreuzkorrelationsfunktion, 250, 251, 288
- Kreuzkorrelationskoeffizient, 288
- Kreuzkorrelationstheorem, 289, 290
- Laplace-Wahrscheinlichkeit, 57
- Leistung, 286
 - normierte, 246
- Leistungsdichtespektrum, 291
- Likelihood-Funktion, 107, 110, 111, 192, 271
- log-Likelihood-Funktion, 111, 271
- Macht, 125, 128, 199
- Mann-Whitney-Test, 219
- Maximum-Likelihood Schätzer, 112
- Maximum-Likelihood-Schätzermethode, 100, 103
- Median, 14, 150
- Methode
 - kleinste Quadrate, 33
- Mittel
 - arithmetisches, 12, 170
- Modell, 119, 143, 183
 - für Zählraten, 45
 - von Laplace, 57
- Momentenmethode, 100, 101, 108, 109
- Momentenschätzer, 109
- Musterfunktion, 234
- Normal-Plot, 187
- Normalapproximation, 175
- Normalverteilung, 157, 273
- Nullhypothese, 119, 131
- P-Wert, 131, 132
- Parameterschätzung, 36
- Partitionierung, 71
- Poisson-Verteilung, 93, 100, 156, 263, 265
- Prozess
 - stochastischer, 221, 234
 - ergodischer, 240
 - schwach stationär, 236
 - stark stationär, 236
- Punkt-Schätzung, 99, 100
- Q-Q-Plot, 184
- Quantil, 150
 - empirische α , 19
- Quartil, 16
- Quartilsdifferenz, 14, 18
- Random Walk, 221
- Rauschen
 - thermisches, 230, 236
 - weisses, 252, 255
- Realisierung, 75, 99, 234
- Rechenregel
 - für Erwartungswert, 90
 - für Varianz, 90
- Rechtecksignal, 252, 295
- Regression
 - lineare, 31
- Residuum, 34
- Robustheit, 16
- Scatterplot, 30
- Scharmittelwert, 238
- Scharvarianz, 238
- Signal
 - stochastisches, 233
- Signalenergie, 287
- Signalleistung, 287, 292
- Signifikanzniveau, 120
- Stabdiagramm, 86
- Standardabweichung, 12, 88, 178

- Standardfehler, 172, 173
- Standardnormalverteilung, 160
- Stationarität, 235
- Statistik, 1
 - für Messdaten, 143, 183
 - deskriptive, 8
 - zweidimensionale Daten, 29
 - eindimensionaler Daten, 8
 - für Zählraten, 98
- Stichprobe
 - gepaarte, 214
 - ungepaarte, 216
- Stochastik, 2
- stochastisch, 231
- Streudiagramm, 30
- Summe, 170
- Test
 - einseitiger, 125
 - statistischer, 99, 112, 124, 132
 - zweiseitiger, 121, 124
- Testentscheid, 120
- Teststatistik, 120
- Theorem
 - Bayes, 71
- Transformation
 - lineare, 162
 - nichtlineare, 167
- Unabhängigkeit, 58, 96, 170
- Urne, 2
- Varianz, 12, 149, 150, 261, 264
 - empirische, 12
- Venn-Diagramm, 52
- Vermutung, 1
- Verschiebungssquadrat, 226
- Verteilung, 174
 - stetige, 149, 151
 - uniforme, 151
- Verteilungsfunktion
 - empirische kumulative, 26, 27
 - kumulative, 91
 - kumulative, 148
- Vertrauensintervall, 99, 134, 137, 206
 - einsitiges, 138
 - zweiseitiges, 206
- Verwerfungsbereich, 120
- Vorzeichenstest, 208
- Wahrscheinlichkeit, 1, 48
 - bedingte, 65
 - geschätzte, 48
- Wahrscheinlichkeitsbaum, 61
- Wahrscheinlichkeitsdichte, 147, 148
- Wahrscheinlichkeitsdichtefunktion, 235
- Wahrscheinlichkeitsmodell, 48
 - diskretes, 55
- Wahrscheinlichkeitsverteilung, 77, 78, 143, 232
- Widerstandsrauschen, 230
- Wiener-Khintchine-Theorem, 253, 291
- Wilcoxon-Test, 211, 213
- Zählraten, 93
- Zeitmittelwert, 238
- Zeitreihe, 255
- Zerfall
 - Alpha-, 46, 95
- Zufallsexperiment, 48
- Zufallsprozess
 - diskreter, 231
 - zeitdiskreter, 231
 - kontinuierlicher, 234
 - stationärer, 236
- Zufallsvariable, 74, 143, 161, 169
 - Poisson-verteilte, 96
 - standardisieren, 165
 - stetige, 143