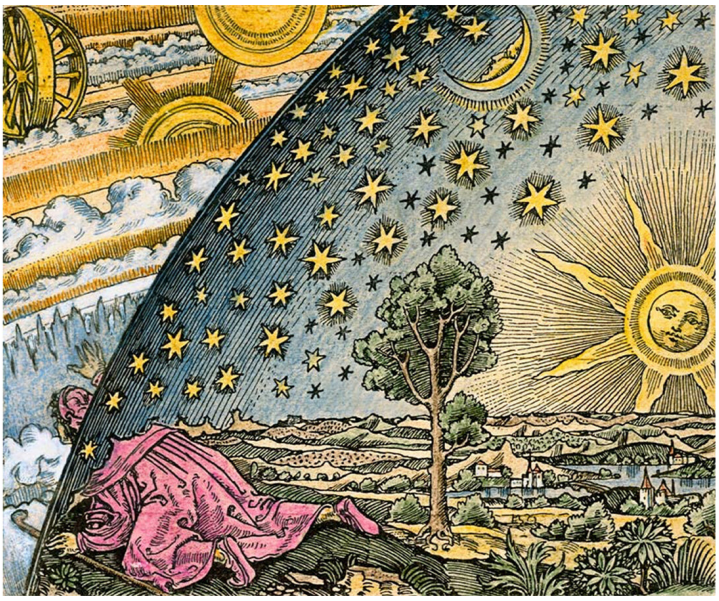
Kurs 5.1: Unendliche Weiten

Sampling in hochdimensionalen Räumen und die Statistik des Big Bang



„Wanderer am Weltenrand“ von Camille Flammarion, 1888

Stell dir vor, du begibst dich auf eine Reise zu den Grenzen der Raumzeit, als unser Universum vor knapp 14 Milliarden Jahren in Erscheinung trat: Der Raum selbst ist gerade erst entstanden und expandiert doch schon um dich herum in alle Richtungen. Das Thermometer zeigt eine Temperatur an, die milliardenfach heißer ist als jede Sonne. Nach knapp einer Minute hat sich die Umgebung so sehr verdünnt und dadurch abgekühlt, dass aus dem dichten Elementarteilchen-Plasma die ersten Elemente entstehen. Nicht mehr lange, dann werden sich die ersten Gaswolken um Ansammlungen dunkler Materie zusammenfinden. Es folgt die Entstehung von unzähligen Sternen und Galaxien und nicht zuletzt von einer gewissen Lebensform, die sich heute den Kopf über den Anfang der Welt zerbricht.

Um zu dieser Erzählung vom Ursprung unserer Welt zu gelangen, brauchte es viele hundert Jahre Forschungsanstrengungen, große Mengen an Daten aus dem Labor und von Teleskopen sowie ausgefuchste Modelle, die versuchen, alle diese Beobachtungen in einem konsistenten Bild zu kombinieren. In diesem Kurs lernen die Teilnehmenden die notwendigen statistischen Werkzeuge, um aus heutigen Beobachtungen auf das frühe Universum zurückzuschließen und analysieren beispielhaft einen Datensatz, über dessen Interpretation die Forschungswelt noch heute uneins ist: Wurden Gravitationswellen aus dem Big Bang gemessen oder handelt es sich um das Hintergrundrauschen von der Verschmelzung supermassiver schwarzer Löcher?

Die Teilnehmenden erhalten Einblicke in das kosmologische Standardmodell und lernen, wie Modellannahmen mit realen Messungen in Verbindung gebracht werden. Für die Auswertung der im Kurs verwendeten Datensätze sind erste Programmierkenntnisse von großem Vorteil. Außerdem sind Grundkenntnisse der Differential- und Integralrechnung für die Kursinhalte hilfreich.

DEUTSCHE SCHÜLERAKADEMIE
Bildung & Begabung

Deutsche SchülerAkademie, 09.07.–25.07.2026, Roßleben-Wiehe

Inhaltsverzeichnis

1 Überblick über den Kurs **2**
- 1.1 Das Universum auf einen Blick . . . 2
- 1.2 Woraus bestehen wir? . . . 3
- 1.3 Zwei Theorien, eine Spannung . . . 4
- 1.4 Das Standardmodell der Kosmologie: ΛCDM . . . 5
- 1.5 Eine Zeitleiste des Universums . . . 6
- 1.6 Größenordnungen: Länge und Zeit im Universum . . . 8
- 1.7 Lücken im Standardmodell . . . 9

2 Kosmologie: Rotverschiebung, Expansion, Inflation und das Hubble-Gesetz **11**
- 2.1 Das kosmologische Prinzip . . . 11
- 2.2 Die Friedmann-Gleichungen . . . 12
- 2.3 Rotverschiebung . . . 12
- 2.4 Das Hubble-Gesetz . . . 13
- 2.5 Inflation . . . 15

3 Differentialrechnung **17**
- 3.1 Ableitungen: eine kurze Wiederholung . . . 17
- 3.2 Differentialgleichungen: Gleichungen, deren Lösungen Funktionen sind . . . 17
- 3.3 Die Tangente als lineare Approximation . . . 18

4 Integralrechnung **20**
- 4.1 Das bestimmte Integral . . . 20
- 4.2 Der Hauptsatz der Differential- und Integralrechnung . . . 21
- 4.3 Differentialgleichungen lösen: Separation der Variablen . . . 22
- 4.4 Numerische Integration . . . 23
- 4.5 Existenz von Lösungen: der Satz von Picard–Lindelöf . . . 23

5 Natürliche Einheiten **25**
- 5.1 Länge und Zeit . . . 25
- 5.2 Energie und Frequenz . . . 26
- 5.3 Temperatur und Energie . . . 28
- 5.4 Energie als universelle Einheit . . . 28

6 Quantengravitation und die Planck-Skala **33**
- 6.1 Die Planck-Skala . . . 33
- 6.2 Aufgabe: Wann wird ein Photon zum Schwarzen Loch? . . . 33
- 6.3 Warum wir GR und QFT trotzdem vertrauen können . . . 34
- 6.4 Inflation und empirische Evidenz . . . 36
  - 6.4.1 Die Energieskala der Inflation . . . 36
  - 6.4.2 Bonus: Vorhersagen und CMB-Beobachtungen . . . 36

7 Taylor-Reihen **38**
- 7.1 Von der Tangente zum Polynom . . . 38
- 7.2 Ein Beispiel: $f(x) = x[\sin(x) + \cos(x)]$ . . . 39
- 7.3 Taylor-Reihen visualisieren . . . 40

8 Friedmann-Gleichungen und ihre Lösung **43**
8.1 Wie verdünnen sich Energiedichten? . . . 43
8.2 Materie-dominiertes Universum: $a(t) \propto t^{2/3}$ . . . 45
8.3 Strahlungsdominiertes Universum: $a(t) \propto t^{1/2}$ . . . 46
8.4 Vakuum-dominiertes Universum: exponentielle Expansion . . . 47
8.5 Die allgemeine Mischung . . . 47
8.6 Berechnung des Alters des Universums . . . 47

9 Dunkle Materie **51**
9.1 Rotationskurven von Galaxien . . . 51
9.2 Was für ein Teilchen ist Dunkle Materie? . . . 52
9.3 Freeze-out: wie viel Dunkle Materie bleibt übrig? . . . 53

10 Fourier-Analysis **57**
10.1 Wellen: $c = \lambda f$ . . . 57
10.2 Motivation: Wie hoch ist die Dusche? . . . 57
10.3 Funktionen als Vektoren . . . 57
10.4 Ein Beispiel: die Rechteckwelle . . . 58
10.5 Spektren . . . 59

11 Gravitationswellen **62**
11.1 Was ist eine Gravitationswelle? . . . 62
11.2 Wie entstehen Gravitationswellen? . . . 63
11.3 Messung: das Michelson-Interferometer . . . 63
11.4 Frequenzbänder und Detektoren . . . 63
11.5 Das Gravitationswellenspektrum $\Omega_{\mathrm{gw}}(f)$ . . . 65

12 Schwarze Löcher **68**
12.1 Was ist ein schwarzes Loch? . . . 68
12.2 Typen schwarzer Löcher . . . 68
12.3 Zeitdilatation im Gravitationsfeld . . . 70
12.4 Ein Fall ins schwarze Loch . . . 71

13 Einführung in die Wahrscheinlichkeitstheorie **73**
13.1 Modelle, Daten und die Rolle der Stochastik . . . 73
13.2 Ein unfairer Würfel . . . 73
13.3 Wahrscheinlichkeitsräume . . . 74
13.4 Bedingte Wahrscheinlichkeit und der Satz von Bayes . . . 75
13.5 Aufgaben zur bedingten Wahrscheinlichkeit und dem Satz von Bayes . . . 75
13.6 Häufige diskrete Wahrscheinlichkeitsverteilungen . . . 79
13.6.1 Binomialverteilung . . . 79
13.6.2 Poissonverteilung . . . 79
13.7 Statistische Modelle . . . 80
13.8 Aufgaben zu Verteilungen und statistischen Modellen . . . 80
13.9 Übergang zu kontinuierlichen Wahrscheinlichkeitsverteilungen . . . 86
13.9.1 Gleichverteilung . . . 86
13.9.2 Normalverteilung (Gauß-Verteilung) . . . 86
13.9.3 Zentraler Grenzwertsatz . . . 87

14 Statistik **97**
14.1 Stichproben und Verteilungen . . . 97
14.1.1 Deskriptive Statistik . . . 97
14.1.2 Wahrscheinlichkeitsverteilungen . . . 97
14.1.3 Erwartungswerte . . . 97
14.1.4 Kumulative Verteilungsfunktion (CDF) . . . 97
14.1.5 Gemeinsame Verteilungen . . . 98
14.1.6 Fehlerfortpflanzung . . . 98
14.2 Schätztheorie . . . 98
14.2.1 Schätzer . . . 98
14.2.2 Maximum-Likelihood-Methode . . . 99
14.2.3 Methode der kleinsten Quadrate . . . 100
14.2.4 Aufgabe: Messungen kombinieren . . . 101
14.2.5 Die χ^2-Verteilung . . . 101
14.3 Konfidenzintervalle . . . 102
14.3.1 Deskriptive Intervalle . . . 102
14.3.2 Konfidenzintervalle aus ML und LS . . . 102
14.3.3 Mehrdimensionale Konfidenzregionen . . . 102
14.4 Hypothesentests . . . 102
14.4.1 Grundbegriffe . . . 102
14.4.2 Neyman-Pearson-Test . . . 103
14.4.3 p-Werte und Look-elsewhere-Effekt . . . 103
14.4.4 χ^2-Anpassungstest . . . 103
14.4.5 Likelihood-Quotienten-Test und Wilks' Theorem . . . 103
14.5 Bayessche Statistik . . . 103
14.5.1 Satz von Bayes . . . 103
14.5.2 Wahl des Priors . . . 104
14.5.3 Marginalisierung . . . 104
14.5.4 Modellvergleich . . . 104

15 Monte-Carlo-Methoden **111**
15.1 Zufallszahlen aus beliebigen Verteilungen . . . 111
15.1.1 Transformationsmethode . . . 111
15.1.2 Verwerfungsmethode (Acceptance-Rejection) . . . 111
15.2 Von χ^2 zur Likelihood . . . 112
15.3 Von der Likelihood zur Posterior: Satz von Bayes . . . 113
15.4 Warum nicht einfach ein Gitter ausprobieren? . . . 114
15.5 Ein Zufallsspaziergang durch den Parameterraum . . . 115
15.6 Warum funktioniert das Metropolis-Kriterium? (*Detailed Balance*) . . . 116
15.7 Der Metropolis-Hastings-Sampler . . . 118

16 Schuhpendel: Bayessche Inferenz mit eigenem MCMC **122**
16.1 Modell und Physik . . . 122
16.2 Likelihood, Prior und Posterior . . . 125
16.3 Metropolis-Hastings-Inferenz . . . 125

17 Pulsar-Timing-Arrays **128**
17.1 Pulsare als kosmische Uhren . . . 128
17.2 Warum ein *Array*? . . . 129
17.3 Das „Free Spectrum": Messung des Gravitationswellenhintergrunds . . . 129
17.4 Alles zusammen: ein erstes Spektrum fitten . . . 131
17.5 Extra-Aufgabe: Mehr Parameter, derselbe Sampler . . . 134

18 Phasenübergänge im frühen Universum **138**
18.1 Was ist ein Feld? . . . 138
18.2 Das mechanische Bild: eine Kugel im Potential . . . 138
18.3 Bubble Nucleation und Kollisionen . . . 139
18.4 Das Spektrum eines Phasenübergangs . . . 140
18.5 Der Phasenübergang als physikalisches Modell . . . 144

19 SMBHB, Phasenübergang und reale Daten **149**

20 Globale Fits und moderne Sampling-Methoden **156**
20.1 Was ist ein globaler Fit? . . . 156
20.2 Phasenübergänge im frühen Universum: ein Forschungsbeispiel . . . 156
20.3 Ausblick: Jenseits von Metropolis-Hastings . . . 156
20.3.1 Nested Sampling . . . 157
20.3.2 Parallel Tempering MCMC (PTMCMC) . . . 159
20.3.3 Wahl des Samplers in der Praxis . . . 159

Ausblick **161**

Glossar **162**

A Satz von Picard–Lindelöf **168**

Vorwort

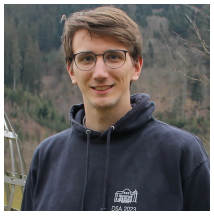
Dieses Skript hat zwei Aufgaben gleichzeitig. Zum einen ist es ein Nachschlagewerk für uns, Simon und Carlo, in dem die mathematischen und physikalischen Grundlagen unseres Kurses an der Deutschen SchülerAkademie so aufgeschrieben sind, dass wir uns bei der Vorbereitung jeder Kursstunde schnell wieder daran erinnern können, wie eine Herleitung ging, welche Aufgabe an welcher Stelle gut passt und welche Musterlösung dazugehört. Zum anderen ist es ein Skript, das wir am Ende des Kurses an alle Teilnehmenden weitergeben: als Erinnerung, zum Nachlesen und vielleicht auch zum Weiterstöbern.

Das Skript ist deshalb so aufgebaut, dass es *linear* gelesen werden kann: Jedes Kapitel baut, wo immer es geht, auf den vorherigen auf. Die meisten Kapitel enthalten neben dem eigentlichen Text auch Aufgaben mit Musterlösungen, von denen viele sich über die letzten Jahre in unseren Kursen bewährt. An dieser Stelle sei deswegen nocheinmal allen ehemaligen Teilnehmenden gedankt, die mitgeholfen haben, unsere Kurse so spannend, witzig und für alle Beteiligten lehrreich zu machen. Dieses, vielleicht sogar unterhaltsame, Skript ist das Ergebnis dieser gemeinsamen Anstrengungen und hoffentlich eine nützliche Ressource für alle, die sich für Kosmologie und Physik interessieren.

Am Ende der Reise, nach Kosmologie, Analysis, Friedmann-Gleichungen, Fourier-Analysis, Phasenübergängen, Gravitationswellen, Pulsar-Timing-Arrays und Dunkler Materie, stehen die Monte-Carlo-Methoden, mit denen wir die gesamte vorherige Mathematik und Physik schließlich zu einem eigenen kleinen Datenanalyse-Projekt zusammenführen.

Das Skript enthält mehr Stoff, als in zwei Wochen behandelt werden kann – das ist Absicht. Für den roten Faden genügen die Kapitel 1–5 (Überblick, Kosmologie und das mathematische Handwerkszeug), 8–12 (Friedmann, Dunkle Materie, Fourier, Phasenübergänge, Gravitationswellen), 14–17 (PTA, Statistik, Monte Carlo, Globale Fits). Die Kapitel 6 (Quantengravitation), 7 (Taylor-Reihen) und 13 (Schwarze Löcher) sowie der Anhang (Satz von Picard–Lindelöf) sind interessante Vertiefungen, aber keine Voraussetzung für das, was danach kommt.

Viel Freude beim Lesen, Rechnen und (Wieder-)Entdecken!
Simon und Carlo



Simon Schwarz (geb. 1999) hat 2025 in Göttingen über stochastische Prozesse und diskrete Differentialgeometrie promoviert. Studiert hat er Mathematik und Theologie an mehreren deutschen Universitäten. Aktuell arbeitet er in der Entwicklung bei Airbus und engagiert sich nebenbei als Rettungssanitäter.

Carlo Tasillo (geb. 1997) hat Physik mit Schwerpunkt Kosmologie und theoretische Teilchenphysik in Dortmund, Aachen und Perugia studiert. Promoviert hat er am DESY Hamburg über Gravitationswellenhintergründe aus dem frühen Universum, gefolgt von Forschungsaufenthalten in den USA und Schweden. Aktuell arbeitet er in Valencia, Spanien.

1 Überblick über den Kurs

1.1 Das Universum auf einen Blick

Abbildung 1.1 zeigt das beobachtbare Universum als logarithmische Karte, mit unserem **Sonnensystem im Zentrum**. Die Darstellung ist *logarithmisch*: Jeder Schritt nach außen entspricht der vervielfachung der Entfernung von unserem Sonnensystem. So passen Erde, Sonnensystem, Milchstraße, Galaxienhaufen und der Rand des beobachtbaren Universums, über 60 Größenordnungen, auf ein einziges Bild.
Was sieht man von innen nach außen?

- **Sonnensystem, Milchstraße, Nachbargalaxien:** Das uns vertraute Nahfeld. Diese Objekte sind mit bloßem Auge oder einfachen Teleskopen sichtbar.
- **Kosmisches Netz** (*cosmic web*): Galaxien sind nicht zufällig verteilt, sondern bilden ein großräumiges Geflecht aus Filamenten (Fäden), Voids (leere Bereiche) und Galaxienhaufen an den Knotenpunkten. Die Beobachtung dieser Strukturen benötigt moderne Teleskope.
- **Rotverschiebung** (rötlicher werdende Außenringe): Licht aus immer größeren Entfernungen hat einen längeren Weg durch das expandierende Universum zurückgelegt; seine Wellenlänge wurde dabei gestreckt und erscheint rotverschoben (Kapitel 2). Seit weniger als 100 Jahren, seit 1929, wissen wir dank Hubble, dass die Rotverschiebung mit dem Abstand zunimmt. Auf diese Beobachtung stützt sich die Urknalltheorie.
- **Dark Ages** (dunkler Bereich nach der Rekombination): Je weiter wir in die Entfernung blicken, desto frühere Epochen des Universums sehen wir. Letztendlich erreichen wir Zeiten, in denen es noch keine Sterne gab. Erst nach dem Erkalten des sog. primordialen Plasmas, nach einigen hundert Millionen Jahren zündeten die ersten Sterne und beendeten damit die *Dark Ages*.
- **Roter Ring = kosmischer Mikrowellenhintergrund (CMB):** Das älteste Licht des Universums, ausgesandt ca. $\sim 380\,000$ Jahre nach dem Urknall bei einer Rotverschiebung von $z \approx 1100$. Der CMB ist die wichtigste kosmologische Beobachtungsgröße überhaupt; er wird uns in Kapitel 8 und 6 (Inflation) immer wieder begegnen. Der CMB ist seit 1965 bekannt und wurde in den letzten 20 Jahren präzise vermessen.
- **Grauer äußerster Bereich = Gravitationswellenhintergrund:** Jenseits des CMB liegt das Universum vor der Rekombination, für Photonen undurchdringlich, aber für Gravitationswellen transparent. Genau hier suchen wir nach Spuren des frühen Universums und sind auf der Suche nach Signalen, die uns etwas über die frühe Geschichte des Universums verraten (Kapitel 11, 17). Ziel dieses Kurses soll sein, nachvollziehbar zu machen, wie wir heute nach solchen Signalen suchen und was sie uns über das frühe Universum mitteilen können.

Bemerkenswert: Von *jedem* Punkt im Universum würde eine solche Karte ähnlich aussehen – überall dasselbe kosmische Netz, dieselbe Hintergrundstrahlung am Rand. Das ist das *kosmologische Prinzip* (Kapitel 2).

Bevor wir uns in den nächsten Kapiteln Schritt für Schritt durch Differential- und Integralrechnung, Natürliche Einheiten, Friedmann-Gleichungen, Fourier-Analysis, Phasenübergänge, Gravitationswellen, Pulsar-Timing-Arrays und Monte-Carlo-Methoden arbeiten, wollen wir uns einmal die ganze Landkarte ansehen, auf der wir uns bewegen. Dieses Kapitel beantwortet noch nichts im Detail; es soll aber schon einmal zeigen, *wohin* die Reise geht und *warum* die einzelnen Bausteine zusammengehören. Ein Amuse-Gueule für alle, die diesem Text weiter folgen wollen.

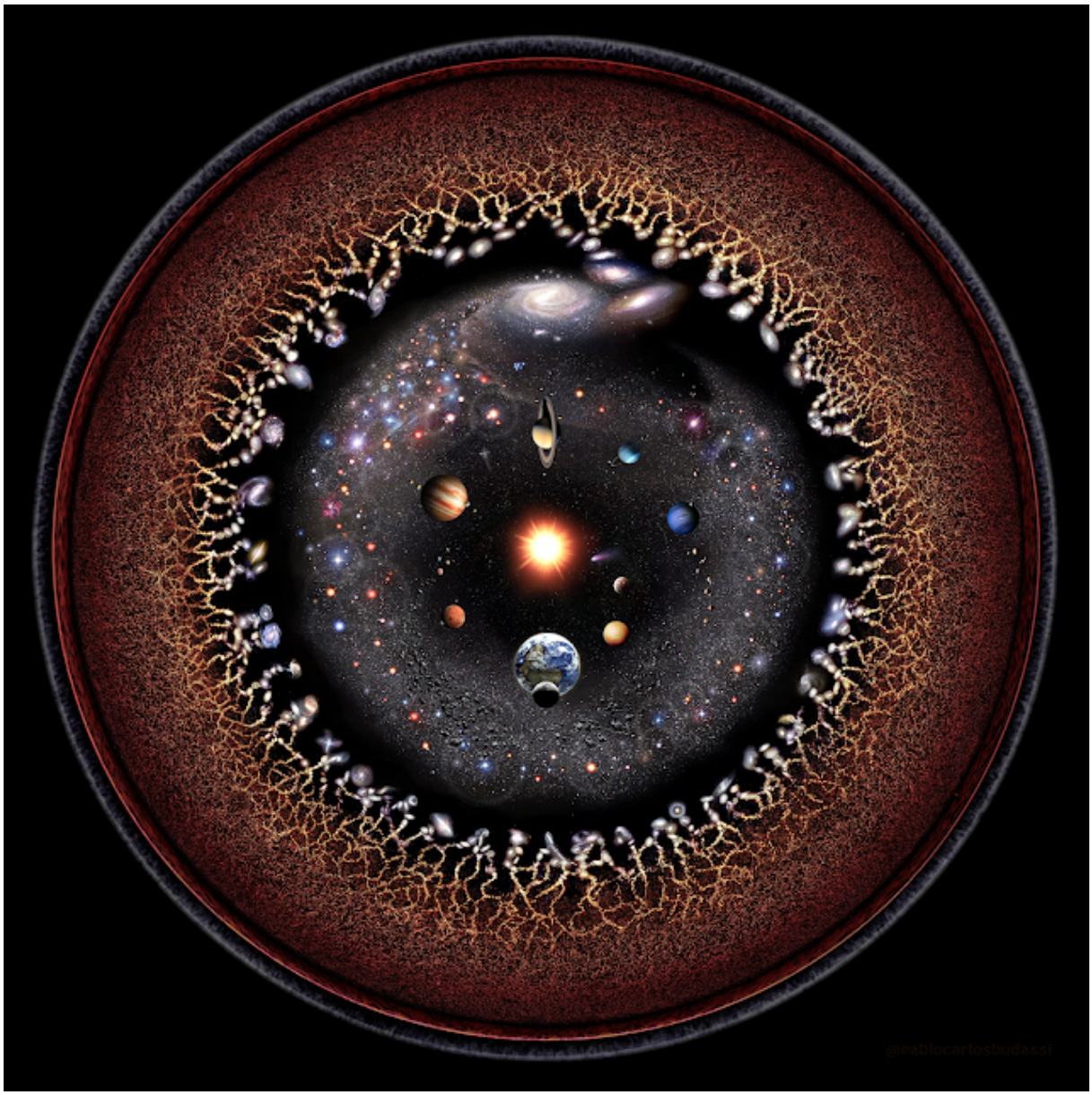


Abbildung 1.1: Logarithmische Karte des beobachtbaren Universums, mit unserem Sonnensystem im Zentrum. Illustration: Pablo Carlos Budassi [1].

1.2 Woraus bestehen wir?

Ein guter Anfang ist immer: bei dem anfangen, was man schon kennt. Alles, was wir um uns herum sehen, ob Stühle, Bäume, Sterne, oder uns selbst, besteht aus Atomen. Atome bestehen aus einem Kern (Protonen und Neutronen) und Elektronen, die ihn umkreisen. Protonen und Neutronen bestehen wiederum aus Quarks, die durch die *starke Kernkraft* zusammengehalten werden, vermittelt durch Teilchen, die *Gluonen* genannt werden. Die elektromagnetische Kraft, die Elektronen an den Kern bindet, wird durch *Photonen* vermittelt. Und dann gibt es noch die *schwache Kernkraft*, vermittelt durch die $W^{\pm}$- und Z-Bosonen, die für radioaktive Zerfälle verantwortlich ist. Schließlich gibt es noch das *Higgs-Boson*, dessen zugehöriges Feld den anderen Teilchen über den sogenannten Higgs-Mechanismus ihre Masse verleiht (dazu mehr in Kapitel 18). Zusätzlich gibt es noch Anti-Teilchen zu allen Teilchen, welche die gleichen Eigenschaften wie die entsprechenden Teilchen haben, aber mit umgekehrten Ladungen.

All das zusammen, nämlich Quarks, Elektronen (und ihre schwereren Geschwister Myon und

Tau sowie die zugehörigen Neutrinos), Photon, Gluonen, $W^{\pm}$-, Z- und Higgs-Boson, bildet das *Standardmodell der Teilchenphysik* (SM). Es ist eines der experimentell am genauesten überprüften Modelle, vielleicht sogar das am besten getestete Modell überhaupt. In Abbildung 1.2 befindet sich eine Übersicht über die Elementarteilchen des Standardmodells.

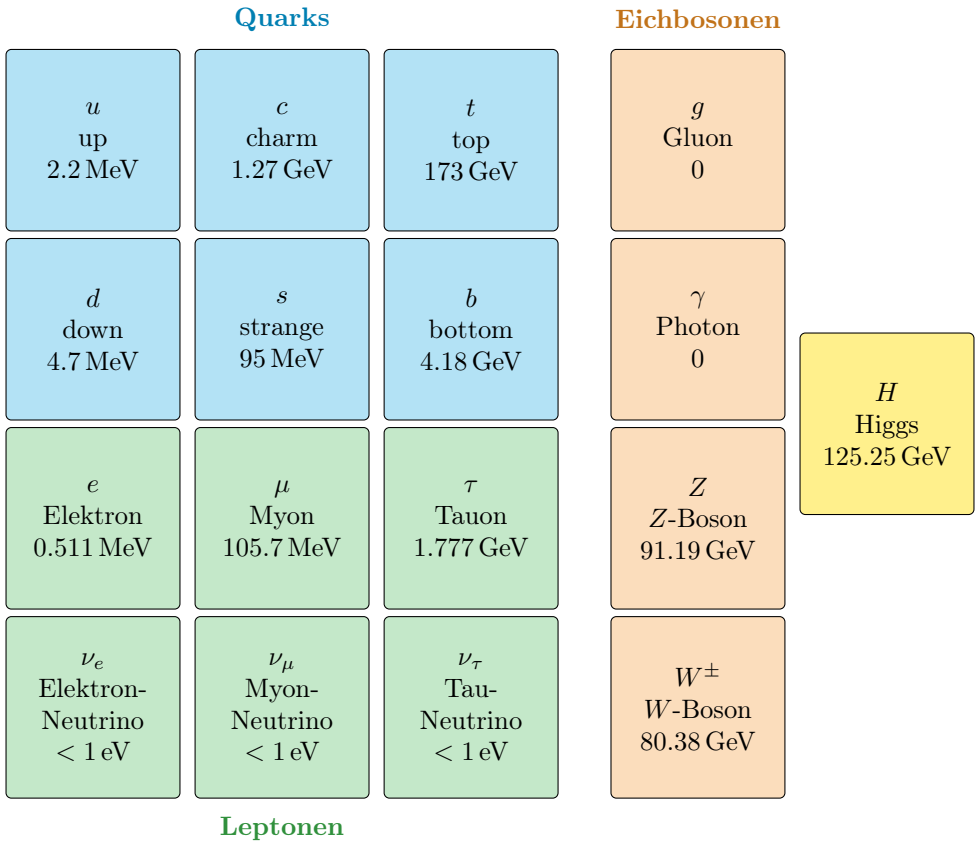


Abbildung 1.2: Die Elementarteilchen des Standardmodells mit ihren (Ruhe-)Massen. Quarks und Leptonen (links, je drei „Generationen“) sind Fermionen und bilden die Materie; Gluon, Photon, Z- und W-Bosonen (rechts) sind die Austauschteilchen der starken, elektromagnetischen und schwachen Wechselwirkung. Das Higgs-Boson ist für die Massen der anderen Teilchen verantwortlich.

1.3 Zwei Theorien, eine Spannung

Das Standardmodell ist eine *Quantenfeldtheorie* (QFT): Teilchen werden als Anregungen von Feldern (siehe Kapitel 18) beschrieben, die den ganzen Raum ausfüllen, und ihre Wechselwirkungen folgen den Regeln der Quantenmechanik.

Die Gravitation hingegen wird nicht durch das Standardmodell beschrieben, sondern durch Einsteins *Allgemeine Relativitätstheorie* (ART, oder GR für *general relativity*). In der ART ist die Gravitation keine Kraft im klassischen Sinne, sondern eine Folge der *Krümmung der Raumzeit* durch Energie und Materie. Kurz zusammengefasst kann man sagen, dass das aus der Schule wohlbekannte Newtonsche Gravitationsgesetz und die Coulomb-Kraft, welche die Anziehung und Abstoßungvon Ladungen beschreibt, beide auf sehr unterschiedliche Weise formuliert werden:

- $F_G = G_N \frac{mM}{r^2}$: beschreibt die Gravitation, wird im Allgemeinen durch die Raumzeitkrümmung abgelöst.
- $F_C = \frac{1}{4\pi\epsilon_0} \frac{qQ}{r^2}$: beschreibt elektrische Felder, wird im Allgemeinen zusammen mit Magent-

feldern, der schwachen und der starken Kraft als Quantenfeld beschrieben.

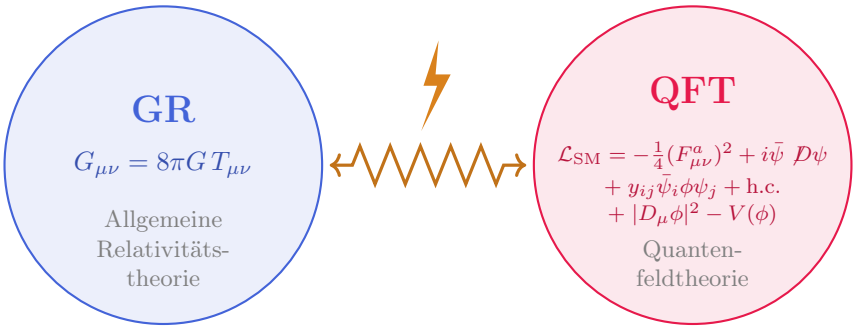


Abbildung 1.3: Die zwei Säulen der modernen Physik: Die Allgemeine Relativitätstheorie (GR) beschreibt Gravitation als Raumzeitkrümmung (Einstein-Gleichung $G_{\mu\nu} = 8\pi G\, T_{\mu\nu}$). Die Quantenfeldtheorie (QFT) beschreibt die drei anderen Grundkräfte über den Standardmodell-Lagrangian $\mathcal{L}_{\text{SM}}$. Beide Theorien sind für sich extrem präzise, aber bisher gibt es keine konsistente *Quantengravitation*, die beide vereint.

Bislang gibt es keine anerkannte Theorie der *Quantengravitation*, die beide Theorien vereint (Abbildung 1.3). Ein großer Teil dieses Kurses besteht darin, nachzuvollziehen, *warum wir trotzdem unseren Rechnungen vertrauen können*, nämlich solange wir uns auf Energie-, Längen- und Zeitskalen bewegen, auf denen beide Theorien getrennt voneinander hervorragend funktionieren. Wo genau diese Grenze liegt, werden wir in Kapitel 5 (Natürliche Einheiten) ausrechnen: Es ist die sogenannte *Planck-Skala*.

1.4 Das Standardmodell der Kosmologie: ΛCDM

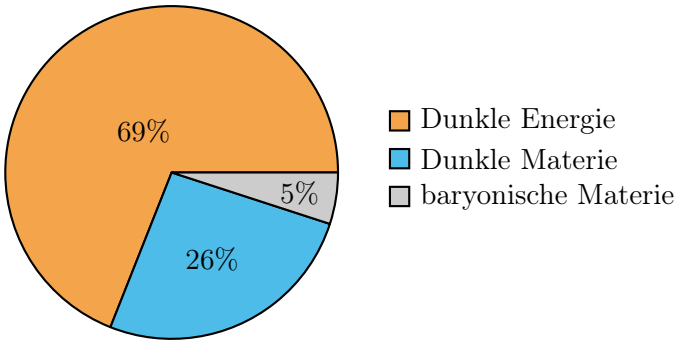
Während das Standardmodell der Teilchenphysik beschreibt, *woraus* das Universum besteht, beschreibt das Standardmodell der Kosmologie, ΛCDM, *wie* sich das Universum als Ganzes entwickelt. Die Buchstaben stehen für:

- Λ: die kosmologische Konstante, oft auch *Dunkle Energie* genannt – verantwortlich für die beschleunigte Expansion des heutigen Universums. Einstein selbst hat sie einmal (in einem anderen Kontext) als seine größte Eselei bezeichnet. Auch heute noch gibt sie der Forschung Rätsel auf.
- CDM: *cold dark matter*, Dunkle Materie, eine bislang unbekannte Form von Materie, die Galaxien zusammenhält, aber nicht mit Licht wechselwirkt. Womöglich handelt es sich hierbei um eine neue Form von Elementarteilchen, welche dem Schema in Abb. 1.2 hinzugefügt werden müsste.

Die heutige Energiebilanz des Universums ist in Tabelle 1.1 zusammengefasst.

Bestandteil	Anteil an der Gesamtenergiedichte heute
Dunkle Energie (Λ)	$\sim 69\,\%$
Dunkle Materie	$\sim 26\,\%$
gewöhnliche (baryonische) Materie	$\sim 5\,\%$
Strahlung (Photonen, Neutrinos)	$< 0.1\,\%$

Tabelle 1.1: Heutige Energiebilanz des Universums nach Planck 2018 [2]. Knapp 95 % bestehen aus Dunkler Energie und Dunkler Materie, Komponenten, die wir bislang nur über ihre gravitative Wirkung kennen.



Nicht eingezeichnet: **Strahlung** (Photonen, Neutrinos) trägt mit $< 0.1\,\%$ einen so winzigen Anteil zur heutigen Energiebilanz bei, dass er im Tortendiagramm nicht sichtbar wäre.

Abbildung 1.4: Die heutige Energiebilanz des Universums (Werte aus Tabelle 1.1). Knapp 95 % des Universums bestehen aus Komponenten, nämlich Dunkle Energie und Dunkle Materie, die wir bislang nur über ihre gravitative Wirkung kennen; nur etwa 5 % sind die uns vertraute, baryonische Materie, aus der Sterne, Planeten und wir selbst bestehen.

Abbildung 1.4 veranschaulicht diese Aufteilung als Tortendiagramm: Der weit überwiegende Teil des Universums besteht aus Dunkler Energie und Dunkler Materie, denen wir in Kapitel 9 (Dunkle Materie) noch ausführlich begegnen werden. Mit nur diesen wenigen Zutaten (plus den *Friedmann-Gleichungen*, die wir in Kapitel 8 herleiten werden) lässt sich die gesamte Entwicklung des Universums von Sekundenbruchteilen nach dem Urknall bis heute beschreiben.

1.5 Eine Zeitleiste des Universums

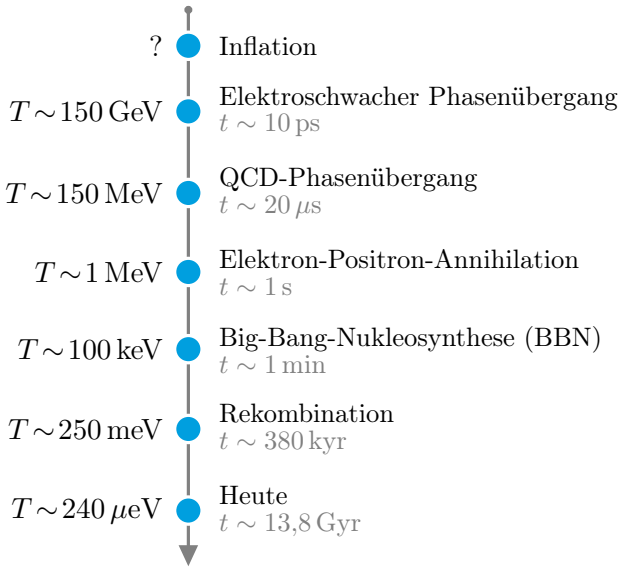


Abbildung 1.5: Zeitleiste der thermischen Geschichte des Universums, vom Ende der Inflation (obten) bis heute (unten), mit Temperaturen T links und Zeitpunkten t nach dem Ende der Inflation rechts. Die eingetragenen Ereignisse (elektroschwacher und QCD-Phasenübergang, BBN, Rekombination) sind die wichtigsten Wegmarken, die uns als roter Faden durch den Kurs begleiten.

Abbildung 1.5 zeigt die wichtigsten Stationen der thermischen Geschichte grafisch als Zeitleiste, von der Inflation (unten, höchste Temperaturen) bis heute (oben). Die Temperaturen sind in Elektronenvolt (eV) angegeben (Kapitel 5). Diese Zeitleiste wird uns als roter Faden durch den ganzen Kurs begleiten.

Was passiert jeweils bei den einzelnen Punkten?

Inflation Direkt nach dem Urknall vermuten wir, dass das Universum eine Phase extrem schneller, exponentieller Ausdehnung durchlaufen hat, vermutlich angetrieben durch ein hypothetisches Skalarfeld (das sog. Inflaton-Feld). Dieser Vorgang glättet das Universum auf großen Skalen, und bläst die ersten quantenmechanischen Fluktuationen auf makroskopische Skalen auf, sodass diese die die Keime für spätere Galaxien werden. Ferner erklärt die Inflation, warum das Universum heute so homogen und flach ist. Zum Ende der Inflation zerfällt das Inflatonfeld in das primordiale Plasma, womöglich also auch in die Teilchen des Standardmodells. Dadurch erhitzt sich das Universum extrem stark und wir sprechen vom "Reheating". Wenn Menschen in der Populärwissenschaft vom "Big Bang" oder "Urknall" sprechen, meinen sie oft die Inflation. Manchmal meinen sie aber auch den Zeitpunkt *vor* der Inflation, über den wir noch weniger empirische Evidenz haben (Kapitel 6). In der Forschung meint "Urknall" für gewöhnlich den Zeitpunkt des Reheatings, da danach das sog. "Hot Big Bang"-Szenario anschließt, also die Ausdehnung und zunehmende Abkühlung des primordialen Plasmas.

Elektroschwacher Phasenübergang ($T \sim 150\,\mathbf{GeV}$, $t \sim 10\,\mathbf{ps}$) Das Higgs-Feld „friert" in sein heutiges Minimum ein: Die elektrische und die schwache Wechselwirkung, die bei sehr hohen Temperaturen vereint waren, trennen sich. Die W- und Z-Bosonen, sowie die Fermionen (also die Quarks und die Leptonen) erhalten durch den Higgs-Mechanismus ihre Masse. Dies ist eines der wichtigsten Beispiele für einen kosmologischen Phasenübergang, den wir in Kapitel 18 ausführlich behandeln. Viele Menschen vermuten, dass zu diesem Zeitpunkt auch die winzige Asymmetrie zwischen Materie und Antimaterie entstanden ist, die dazu geführt hat dass heute auf ca. 10 Milliarden Licht-Teilchen ein Materieteilchen kommt, welches aus der Annihilation von Materie und Antimaterie in der folgenden Entwicklung des Universums übrig geblieben ist.

QCD-Phasenübergang ($T \sim 150\,\mathbf{MeV}$, $t \sim 20\,\mu\mathbf{s}$) Quarks und Gluonen, die vorher frei im Quark-Gluon-Plasma herumschwirren, werden durch die starke Kernkraft zu Protonen und Neutronen „eingesperrt" (Confinement). Dieser Übergang ist dafür verantwortlich, dass die heute bekannte Kernmaterie überhaupt existiert. QCD steht für Quantenchromodynamik, also die Theorie der starken Wechselwirkung zwischen Quarks und Gluonen.

Elektron-Positron-Annihilation ($T \sim 1\,\mathbf{MeV}$, $t \sim 1\,\mathbf{s}$) Das Universum kühlt so weit ab, dass Elektronen und Positronen nicht mehr thermisch produziert werden können. Sie annihilieren massenweise zu Photonen; der übrige Überschuss an Elektronen (eines für jedes Proton) bildet die Elektronen in heutigen Atomen.

Big-Bang-Nukleosynthese (BBN, $T \sim 100\,\mathbf{keV}$, $t \sim 1\,\mathbf{min}$) Die Temperatur fällt weit genug ab, damit Protonen und Neutronen zu leichten Atomkernen fusionieren können: vor allem Helium-4, ein wenig Deuterium, Helium-3 und Lithium. Etwa 25 % der baryonischen Masse wird so in Helium umgewandelt, die restlichen 75 % bleiben als noch ionisierter Wasserstoff, also Protonen zurück. Die BBN-Vorhersagen stimmen hervorragend mit astronomischen Beobachtungen überein und sind ein starker Beleg für das Urknall-Modell. Die Übereinstimmung ist so gut, dass die BBN auch als Test für Ergänzungen des Standardmodells dient (Kapitel 20).

Rekombination ($T \sim 250\,\mathbf{meV}$, $t \sim 380\,000\,\mathbf{yr}$) Elektronen und Protonen kombinieren sich zu neutralem Wasserstoff. Das Universum wird für Photonen transparent: Licht kann jetzt erstmals frei reisen. Die dabei freigesetzten Photonen bilden den *kosmischen Mikrowellenhintergrund* (CMB), den wir heute als gleichmäßige Wärmestrahlung bei $\sim 2.7\,\mathrm{K}$ messen. Der CMB ist die bei weitem wichtigste Informationsquelle über das frühe Universum. Die Vermessung des CMB hat das ΛCDM-Modell erst möglich gemacht.

Heute ($T \sim 240\,\mu$eV, $t \sim 13.8\,$Gyr) Das Universum ist transparent, von Galaxien durchzogen und befindet sich in einer Phase beschleunigter Expansion durch die Dunkle Energie (Λ). Kursleitende und -teilnehmende zerbrechen sich nach wie vor den Kopf über seinen Ursprung.

Diese Zeitleiste begleitet uns durch den gesamten Kurs: Kapitel 18 behandelt Phasenübergänge wie auch den elektroschwachen und QCD-Phasenübergang im Detail, Kapitel 8.6 beschreibt, wie wie jedem Ereignis ein Alter des Universum zuordnen können und Kapitel 11 und 17 zeigen, wie wir in neuen Forschungsdaten nach Spuren dieser frühen Ereignisse suchen können.

Abbildung 1.6 ergänzt diese Zeitleiste um eine quantitative Perspektive: Sie zeigt, wie sich die Energiedichten der drei Komponenten ($\rho_{\rm rad}$, $\rho_{\rm mat}$, ρ_Λ) über die gesamte kosmische Geschichte entwickeln, von der BBN über die Rekombination bis heute. Ziel von Kapitel 8 wird es sein, diese Entwicklung nachzuvollziehen.

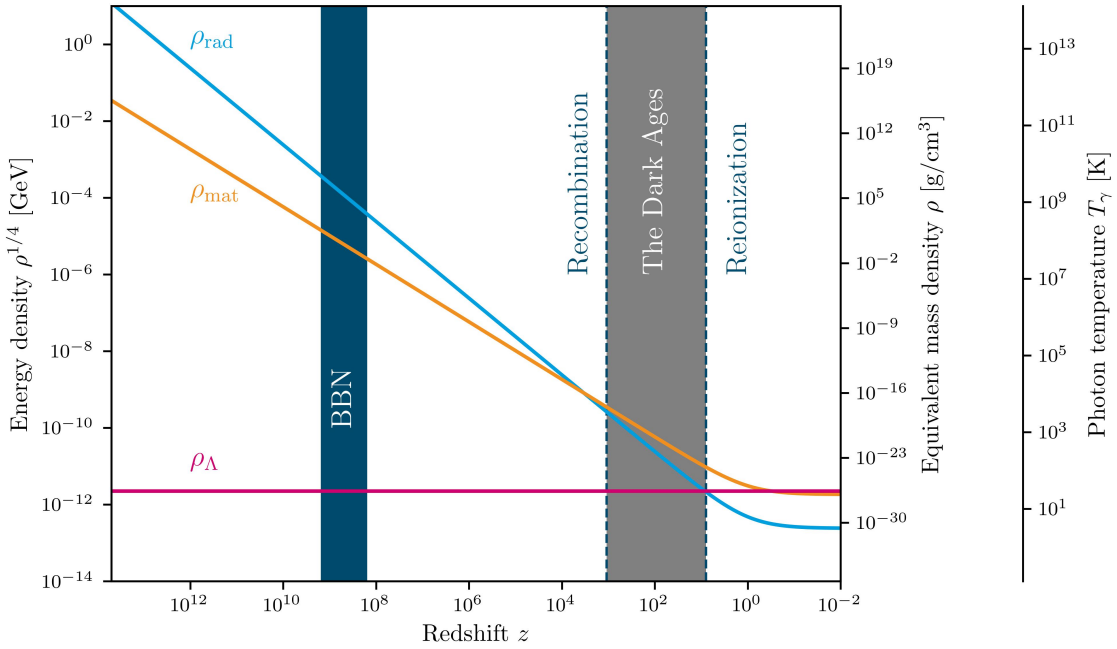


Abbildung 1.6: Entwicklung der Energiedichten $\rho^{1/4}$ (in GeV) als Funktion der Rotverschiebung z, von der Urknall-Ära ($z \sim 10^{12}$, links) bis heute ($z \sim 0$, rechts). Strahlung (blau, $\rho_{\rm rad} \propto (1+z)^4$) dominiert bei hohen z; Materie (orange, $\rho_{\rm mat} \propto (1+z)^3$) übernimmt nach der Materie-Strahlungs-Gleichheit; die Vakuumenergie ρ_Λ (pink) dominiert seit $z \lesssim 0.3$. Die markierten Epochen (BBN, Rekombination, Dark Ages, Reionisation) entsprechen den Wegmarken in der Zeitleiste. Die beiden zusätzlichen vertikalen Achsen vergleichen die Energiedichte (dank $E = mc^2$) mit einer Massendichte und einer Temperatur des primordialen Plasmas zum angegebenen Zeitpunkt.

1.6 Größenordnungen: Länge und Zeit im Universum

Ein zentrales Werkzeug, um sich in den enormen Größenordnungen der Kosmologie zurechtzufinden, sind die *natürlichen Einheiten* (Kapitel 5). Sie erlauben es, Längen, Zeiten, Energien und Temperaturen ineinander umzurechnen. Tabelle 1.2 gibt eine Übersicht der relevanten Längen- und Zeit-Skalen, von der kleinsten bekannten (Planck-Skala) bis zur größten (dem beobachtbaren Universum), welche wir letztendlich auch in natürlichen Einheiten auszudrücken lernen werden.

Zwischen der kleinsten und der größten Zahl in Tabelle 1.2 liegen mehr als 60 Größenordnungen, und trotzdem werden wir sehen, dass dieselben physikalischen Gesetze (Friedmann-Gleichungen, Quantenfeldtheorie) auf allen diesen Skalen eine Rolle spielen.

Skala	Länge	Zeit
Planck-Skala	$1.6 \times 10^{-35}\,\mathrm{m}$	$5.4 \times 10^{-44}\,\mathrm{s}$
Proton	$\sim 10^{-15}\,\mathrm{m}$	–
Atom	$\sim 10^{-10}\,\mathrm{m}$	–
Mensch	$\sim 1\,\mathrm{m}$	–
Erde–Sonne	$1.5 \times 10^{11}\,\mathrm{m}$	$\sim 8\,\mathrm{min}$
Milchstraße	$\sim 10^{21}\,\mathrm{m}$	$\sim 10^{5}\,\mathrm{yr}$
beobachtbares Universum	$\sim 10^{26}\,\mathrm{m}$	$\sim 13.8 \times 10^{9}\,\mathrm{yr}$

Tabelle 1.2: Charakteristische Längen- und Zeitskalen von der Planck-Skala bis zum beobachtbaren Universum. Zwischen dem kleinsten und dem größten Eintrag liegen über 60 Größenordnungen.

In der Astronomie verwendet man der Übersicht halber zwei besonders praktische (da: *astronomisch große*) Längeneinheiten. Das *Lichtjahr* (ly) ist die Strecke, die Licht in einem Jahr zurücklegt:

$$1\,\mathrm{ly} = c \cdot 1\,\mathrm{yr} \approx 9.46 \times 10^{15}\,\mathrm{m}\,. \tag{1.1}$$

Das *Parsec* (pc) ist die Einheit, die in der professionellen Astronomie und Kosmologie noch häufiger verwendet wird. Es ist definiert über den Winkel, unter dem die Erde-Sonne-Distanz ($1\,\mathrm{AU} = 1.496 \times 10^{11}\,\mathrm{m}$, eine sog. astronomische Einheit) von einem Stern aus erscheint: Ein Stern hat einen Abstand von einem Parsec, wenn seine jährliche Parallaxe genau eine Bogensekunde ($1''$) beträgt, wenn er sich im Laufe eines Jahres also um den angegebenen Winkel bewegt. Es gilt

$$1\,\mathrm{pc} \approx 3.086 \times 10^{16}\,\mathrm{m} \approx 3.26\,\mathrm{ly}\,. \tag{1.2}$$

Für kosmologische Skalen verwendet man Kiloparsec ($\mathrm{kpc} = 10^3\,\mathrm{pc}$) und Megaparsec ($\mathrm{Mpc} = 10^6\,\mathrm{pc}$): Die Milchstraße hat einen Durchmesser von $\sim 30\,\mathrm{kpc}$, Galaxienhaufen sind ~ 1–$10\,\mathrm{Mpc}$ groß, und das beobachtbare Universum hat einen Radius von $\sim 14\,\mathrm{Gpc}$. Die Hubble-Konstante H_0 wird typischerweise in km/s/Mpc gemessen, Mpc ist also die natürliche Längenskala der Kosmologie.

1.7 Lücken im Standardmodell

Das Standardmodell der Teilchenphysik ist eine Triumph der modernen Physik, beantwortet aber längst nicht alle Fragen. Es gibt mehrere gut bekannte Phänomene, für die das SM keine Erklärung liefert. Diese „Lücken“ sind zugleich die spannendsten offenen Fragen der Physik:

Dunkle Materie. Wir sehen, dass Galaxien sich schneller drehen, als sie es dürften, wenn ihre sichtbare Materie allein gravitativ wirken würde. Ebenso biegen Galaxienhaufen das Licht stärker ab, als ihr sichtbarer Inhalt es erklären könnte. Das Standardmodell hat kein Teilchen, das diese *Dunkle Materie* erklären könnte, da sie kaum mit normalem Material wechselwirkt, leuchtet nicht, und macht dennoch $\sim 26\,\%$ der gesamten Energiedichte des Universums aus. Kapitel 9 widmet sich dieser Frage.

Dunkle Energie. Das Universum expandiert nicht nur, sondern *beschleunigt*. Diese Beschleunigung wird einer unbekannten Form von Energie zugeschrieben, der *Dunklen Energie* (oft modelliert als kosmologische Konstante Λ), die $\sim 69\,\%$ der gesamten Energiedichte ausmacht. Charakteristische Eigenschaft der dunklen Energie ist, dass sie, anders als gewöhnliche Energiedichten, einen negativen Druck hat. Das kann intuitiv damit gleichgesetzt werden, dass “sie sich gerne ausdehnen würde”. Woher sie kommt, ist völlig unklar: Die theoretische Vorhersage aus der

Quantenfeldtheorie liegt 10^{120}-mal über dem beobachteten Wert, das größte Versagen der theoretischen Physik und eine der besten Motivationen um sich mit Theorien der quantengravitation (Kapitel 6) auseinanderzusetzen.

Das Hierarchieproblem. Die Higgs-Masse beträgt 125 GeV, obwohl die Quantenfluktuationen (in der Sprache der Quantenfeldtheorie: die *Schleifenkorrekturen*, engl. loop corrections) schieben sie eigentlich auf die Planck-Skala $M_{\mathrm{Pl}} \sim 10^{19}\,\mathrm{GeV}$. Damit die Higgs-Masse so klein bleibt, müssten sich die Korrekturen auf 30 Dezimalstellen genau wegheben. Das erscheint unnatürlich. Theorien wie Supersymmetrie oder Composit-Higgs-Modelle (also solche, in denen das Higgs-Boson kein Elementarteilchen, sondern ein zusammengesetztes Teilchen wie ein Atom ist) versuchen dieses Problem zu lösen, bisher ohne experimentellen Beleg.

Neutrinomassen. Im Standardmodell sind Neutrinos masselos. Aber Experimente zeigen, dass Neutrinos *oszillieren*, d.h. sie wechseln zwischen Elektron-, Myon- und Tau-Neutrino-Typ hin und her, was nur möglich ist, wenn die verschiedenen Neutrino-Typen unterschiedliche Massen haben. Das SM muss erweitert werden, um das zu erklären; welcher Mechanismus hinter den Neutrinomassen steckt, ist noch unbekannt. Häufig wird aber angenommen, dass Neutrinos auch ihre Massen durch das Higgs-Boson erhalten.

All diese Lücken zeigen: Hinter dem Standardmodell steckt noch mehr Physik. Genau nach dieser „neuen Physik“ sucht ein Großteil der modernen experimentellen und theoretischen Physik – auch mit den Werkzeugen, die wir in diesem Kurs erarbeiten werden.

Abbildung 1.7: Das hat mit dem Big Bang *nichts* zu tun.

2 Kosmologie: Rotverschiebung, Expansion, Inflation und das Hubble-Gesetz

2.1 Das kosmologische Prinzip

Die gesamte moderne Kosmologie beruht auf einer überraschend einfachen Annahme, dem *kosmologischen Prinzip*:

- **Isotropie**: Das Universum sieht (auf großen Skalen) in jede Richtung gleich aus – es ist egal, in welche Richtung man schaut.
- **Homogenität**: Das Universum sieht (auf großen Skalen) an jedem Ort gleich aus, es ist egal, an welchem Ort man sich befindet.

Natürlich ist das Universum *lokal* alles andere als homogen: es gibt Galaxien, Sterne, Planeten und leere Räume dazwischen. Aber wenn man groß genug zoomt (Skalen von einigen hundert Megaparsec), verschwinden diese Unterschiede im statistischen Mittel. Abbildung 1.1 aus Kapitel 1 veranschaulicht das eindrücklich: Von jedem Punkt im Universum aus würde eine solche Karte des beobachtbaren Universums ähnlich aussehen: überall dasselbe kosmische Netz, dieselbe Hintergrundstrahlung am Rand.

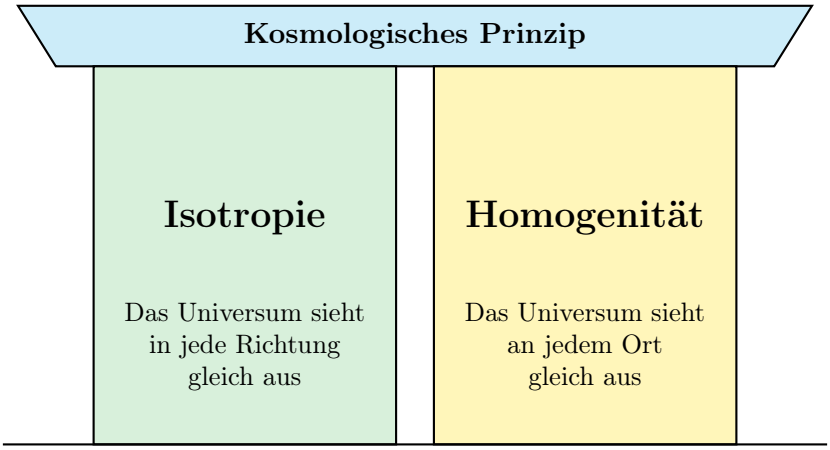


Abbildung 2.1: Das kosmologische Prinzip „ruht“ auf den beiden gleichberechtigten Säulen Isotropie und Homogenität. Erst beide gemeinsam erlauben es, die gesamte Geometrie des Universums durch eine einzige Funktion, den Skalenfaktor $a(t)$, zu beschreiben.

Aus dem kosmologischen Prinzip (Abbildung 2.1) folgt, dass sich die Geometrie des Universums vollständig durch eine einzige Funktion der Zeit beschreiben lässt: den *Skalenfaktor* $a(t)$. Er beschreibt, wie sich Abstände zwischen zwei beliebigen (mitbewegten) Punkten im Universum mit der Zeit verändern – für $a(t_1) < a(t_2)$ war der Abstand zur Zeit t_1 kleiner als zur Zeit t_2, das Universum hat sich also ausgedehnt. Per Konvention wird der heutige Wert auf $a(t_0) = 1$ gesetzt. Der Abstand L zwischen zwei Punkten (die sich *frei fallend* bewegen, also nur durch die Gravitation beeinflusst sind) lässt sich in Abhängigkeit von der Zeit über den Skalenfaktor beschreiben:

$$L(t) = \frac{a(t)}{a_*} \cdot L_*, \tag{2.1}$$

wobei a_* und L_* der Wert des Skalenfaktors bzw. des Abstands zu einem anderen, bekannten Zeitpunkt sind.

Mit dem Skalenfaktor definieren wir die *Hubble-Rate*

$$H(t) = \frac{1}{a(t)} \frac{\mathrm{d}a}{\mathrm{d}t}(t)\,, \tag{2.2}$$

die angibt, wie schnell sich das Universum zum Zeitpunkt t ausdehnt. Ihr heutiger Wert $H_0 = H(t_0)$ heißt *Hubble-Konstante*. Die Einheit der Hubble-Rate ist $[H] = T^{-1}$, also „pro Zeit“. Wie wir unten sehen werden, ist die geläufigere Einheit der Hubble-Konstanten aber $[H] = \mathrm{km\,(s\,Mpc)^{-1}}$.

2.2 Die Friedmann-Gleichungen

Setzt man das kosmologische Prinzip in die Einsteingleichungen der ART ein, erhält man die beiden *Friedmann-Gleichungen*. Eine schöne, von John Wheeler geprägte Merkhilfe (ursprünglich für die Einsteingleichungen im Allgemeinen, aber genauso passend hier) hilft, sich die Bedeutung der Gleichungen klarzumachen:

Friedmann-Gleichung I

$$H^2(t) = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\,\rho(t) = \frac{1}{3\,m_{\mathrm{Pl}}^2}\,\rho(t) \tag{2.3}$$

„*Matter tells space how to curve*“: Die Materie-Energiedichte ρ bestimmt, wie schnell sich das Universum ausdehnt (H).

Friedmann-Gleichung II

$$\dot{\rho}(t) = -3H(t)[\rho(t) + p(t)] \tag{2.4}$$

„*Space tells matter how to move*“: Die Expansionsrate H bestimmt, wie schnell die Materie-Energiedichte ρ mit der Zeit abnimmt.

Dabei sind $\rho(t)$ die mittlere Energiedichte und $p(t)$ der Druck des kosmischen Plasmas, sowie $m_{\mathrm{Pl}} = 1/\sqrt{8\pi G}$ die (reduzierte) Planck-Masse (siehe Kapitel 5). Diese beiden Gleichungen, zusammen mit einer Zustandsgleichung $p = p(\rho)$, die festlegt, um welche Art von Materie (Strahlung, Staub, Vakuumenergie) es sich handelt, werden wir in Kapitel 8 explizit lösen und in Abschnitt 8.6 numerisch auf unser eigenes Universum anwenden, um sein Alter zu berechnen.

2.3 Rotverschiebung

Stellen wir uns vor, ein fernes Atom sendet Licht mit Wellenlänge λ_{e} aus (Index e für *emittiert*), zu einem Zeitpunkt, an dem der Skalenfaktor den Wert $a_{\mathrm{e}} = a(t_{\mathrm{e}})$ hatte. Dieses Licht erreicht uns heute, bei $a_0 = a(t_0) = 1$, mit einer Wellenlänge λ_0. Da sich der Raum selbst, und mit ihm die Wellenlänge des Lichts, um den Faktor a_0/a_{e} ausgedehnt hat, gilt

$$\frac{\lambda_{\mathrm{e}}}{a_{\mathrm{e}}} = \frac{\lambda_0}{a_0} \quad\Longrightarrow\quad \lambda_0 = \lambda_{\mathrm{e}} \cdot \frac{a_0}{a_{\mathrm{e}}} = \lambda_{\mathrm{e}} \cdot \frac{1}{a_{\mathrm{e}}}\,. \tag{2.5}$$

Da $a_{\mathrm{e}} < 1$ (der Skalenfaktor war in der Vergangenheit kleiner, alles war dichter), gilt $\lambda_0 > \lambda_{\mathrm{e}}$: Das Licht wird beim Reisen durch das expandierende Universum zu größeren Wellenlängen hin verschoben – „rotverschoben“, da Rot die langwelligste Farbe des sichtbaren Spektrums ist. Man definiert die *Rotverschiebung* z über

$$z := \frac{\lambda_0 - \lambda_{\mathrm{e}}}{\lambda_{\mathrm{e}}} \quad\Longrightarrow\quad \lambda_0 = \lambda_{\mathrm{e}}\,(1+z) = \frac{\lambda_{\mathrm{e}}}{a_{\mathrm{e}}} \quad\Longrightarrow\quad a_{\mathrm{e}} = \frac{1}{1+z}\,. \tag{2.6}$$

Die Rotverschiebung z ist damit eine direkt messbare Größe (man vergleicht beobachtete Spektrallinien mit bekannten Laborwerten von λ_{e}), die uns sagt, wie dicht das Universum war, als

das beobachtete Licht ausgesandt wurde. Abbildung 2.2 veranschaulicht diesen Effekt: Das ausgesandte Licht (blau, kurze Wellenlänge) wird durch die kosmische Expansion zu längeren Wellenlängen hin gestreckt und kommt rot verschoben bei uns an.

Abbildung 2.2: Rotverschiebung durch die kosmische Expansion: Das von der Quelle emittierte Licht mit Wellenlänge λ_e (blau, links) wird mit dem expandierenden Raum mitgedehnt und erreicht uns mit längerer Wellenlänge $\lambda_0 > \lambda_e$ (rot, rechts); das Licht „verschiebt sich ins Rote“.

2.4 Das Hubble-Gesetz

Für kleine Rotverschiebungen (also für nicht allzu weit entfernte Objekte) lässt sich zeigen, dass die Fluchtgeschwindigkeit v einer Galaxie näherungsweise proportional zu ihrer Entfernung d_L von uns ist:

$$v \approx H_0 \cdot d_L \,. \tag{2.7}$$

Dies ist das berühmte *Hubble-Gesetz*. Es war Edwin Hubbles Beobachtung (1929), dass weiter entfernte Galaxien sich schneller von uns entfernen, die als erster überzeugender Hinweis auf ein expandierendes Universum gilt und damit indirekt auf einen „Anfang“ (den Urknall).

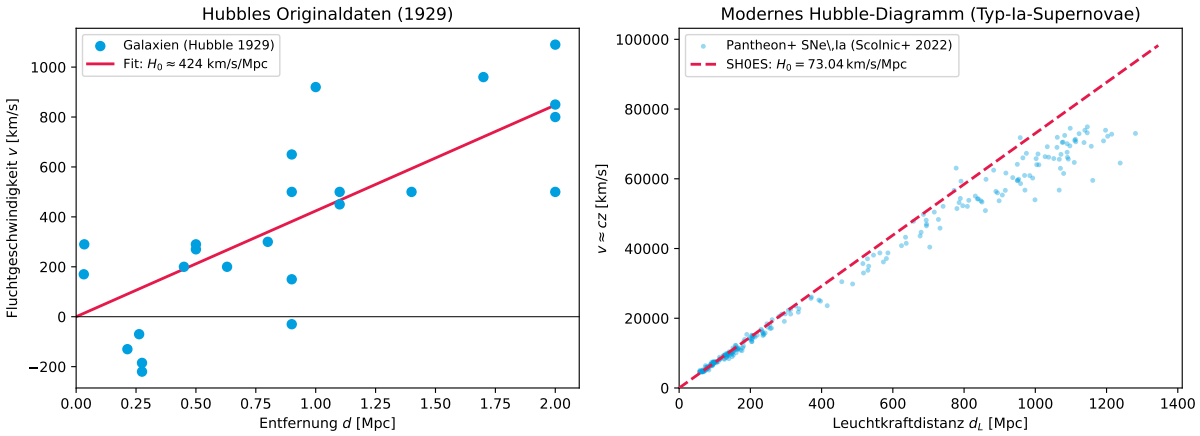


Abbildung 2.3: *Links:* Hubbles Originaldaten von 1929: Fluchtgeschwindigkeit v gegen Entfernung d_L. Die Steigung ist direkt H_0; Hubbles Wert $H_0 \approx 424\,\mathrm{km/s/Mpc}$ war wegen fehlerhafter Entfernungskalibrierung zu groß. Daten: Hubble [3]. *Rechts:* Modernes Hubble-Diagramm mit repräsentativen Typ-Ia-Supernovae aus dem Pantheon+ / SH0ES-Datensatz [4]: die gestrichelte Linie zeigt die SH0ES-Referenz $H_0 = 73.04\,\mathrm{km/s/Mpc}$. Bei $d_L \gtrsim 300\,\mathrm{Mpc}$ weichen die Datenpunkte systematisch nach unten ab, ein Effekt der Näherung $v \approx cz$ (s. Text).

Bei hohen Luminositätsdistanzen ($d_L \gtrsim 300\,\mathrm{Mpc}$, entsprechend $z \gtrsim 0.1$) liegen die Datenpunkte im modernen Hubble-Diagramm (Abbildung 2.3) *unterhalb* der SH0ES-Referenzlinie. Das liegt an der Näherung $v \approx cz$: Für $z \gtrsim 0.1$ unterschätzt $v = cz$ die tatsächliche Fluchtgeschwindigkeit, weil die Luminositätsdistanz im expandierenden Universum schneller anwächst als cz/H_0. Genauer gilt

$$d_L = \frac{c}{H_0}(1+z)\int_0^z \frac{dz'}{E(z')}, \quad E(z) = \sqrt{\Omega_{\mathrm{mat}}(1+z)^3 + \Omega_\Lambda}\,, \tag{2.8}$$

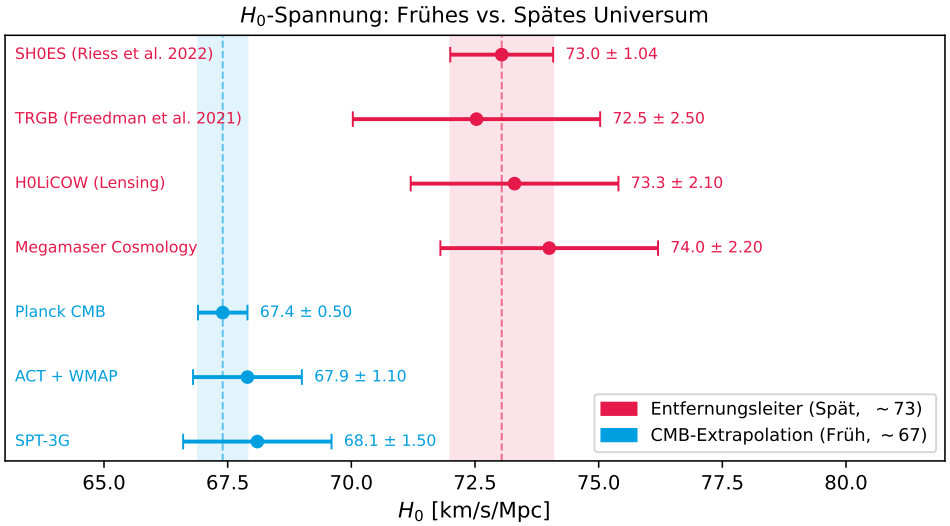


Abbildung 2.4: Die moderne H_0-Spannung: Messungen über die kosmische Entfernungsleiter (rot, „spätes Universum“, z.B. SH0ES: $H_0 = 73.04 \pm 1.04\,\mathrm{km/s/Mpc}$) ergeben systematisch höhere Werte als CMB-Extrapolationen (blau, „frühes Universum“, z.B. Planck: $H_0 = 67.4 \pm 0.5\,\mathrm{km/s/Mpc}$). Diese Diskrepanz von $\gtrsim 5\sigma$ ist bis heute ungeklärt und könnte auf neue Physik hindeuten.

wobei $\Omega_{\mathrm{mat}} \approx 0.31$ und $\Omega_\Lambda \approx 0.69$. Bei $z = 0.1$ beträgt die Korrektur bereits $\sim 5\,\%$, bei $z = 0.2$ schon $\sim 10\,\%$. Das Hubble-Gesetz $v = H_0 d_L$ mit der Näherung $v \approx cz$ gilt daher nur für kleine Rotverschiebungen $z \ll 1$.

Aufgabe 2.1

Das Hubble-Gesetz

$$v = H_0 \cdot d_{\mathrm{L}}$$

beschreibt, mit welcher Geschwindigkeit v sich Galaxien in einer (Luminositäts-)Distanz d_{L} von uns wegbewegen. Durch spektroskopische Analysen des Lichts von Himmelskörpern und die Analyse des CMB konnte $H_0 = 67.4\,\frac{\mathrm{km}}{\mathrm{s\,Mpc}}$ bestimmt werden.

a) Ein Parsec ist die Distanz, aus welcher der mittlere Erdbahnradius (1 au), also der mittlere Abstand zwischen Sonne und Erde, unter einem Winkel von einer Bogensekunde erscheint. Es gilt

$$1\,\mathrm{pc} = 206{,}265\,\mathrm{au} = 3.26\,\mathrm{ly} = 30.9 \times 10^{12}\,\mathrm{km}\,.$$

Nutze obige Umrechnungshilfe um den Hubble-Parameter H_0 in der Einheit s^{-1} anzugeben.

b) Der inverse Hubble-Parameter H_0^{-1} definiert also eine Zeitskala. Gib diese in einer der Größenordnung angemessenen Einheit an. Was fällt dir auf?

c) Welche Längenskala gibt cH_0^{-1} an? Wähle auch hier eine Einheit, die der Kosmologie angemessen ist.

Hinweis: Ein Jahr ist ziemlich genau $\pi \times 10^7\,\mathrm{s}$ lang.

Musterlösung 2.1

a) Wir rechnen Mpc in km um, $1\,\mathrm{Mpc} = 10^6\,\mathrm{pc} = 3.09 \times 10^{19}\,\mathrm{km}$, und s in km/h ... bzw. direkt:

$$H_0 = 67.4\,\frac{\mathrm{km}}{\mathrm{s\,Mpc}} \cdot \frac{1\,\mathrm{Mpc}}{3.09 \times 10^{19}\,\mathrm{km}} = 2.18 \times 10^{-18}\,\mathrm{s}^{-1}\,. \tag{2.9}$$

b) Der Kehrwert ist

$$H_0^{-1} = \frac{1}{2.18 \times 10^{-18}\,\mathrm{s}} = 4.59 \times 10^{17}\,\mathrm{s}\,. \tag{2.10}$$

Mit $1\,\mathrm{yr} \approx \pi \times 10^7\,\mathrm{s}$ folgt

$$H_0^{-1} \approx \frac{4.59 \times 10^{17}}{\pi \times 10^7}\,\mathrm{yr} \approx 14.6 \times 10^9\,\mathrm{yr} = 14.6\,\mathrm{Gyr}\,. \tag{2.11}$$

Das fällt auf: Diese Zahl liegt sehr nahe am Alter des Universums ($\approx 13.8\,\mathrm{Gyr}$, vgl. Kapitel 1.5 und 8.6)! H_0^{-1} heißt auch *Hubble-Zeit*, eine grobe Abschätzung für das Alter des Universums, die wir in Kapitel 8.6 präzisieren werden.

c)

$$cH_0^{-1} \approx c \times 14.6\,\mathrm{Gyr} = 14.6\,\mathrm{Gly} \approx 4.5\,\mathrm{Gpc}\,, \tag{2.12}$$

wobei wir $1\,\mathrm{pc} \approx 3.26\,$Lichtjahre benutzt haben. cH_0^{-1} heißt in dieser Form auch *Hubble-Radius*, eine grobe Abschätzung für die Größe des beobachtbaren Universums.

Aus der Beobachtung der Ausdehnung des Universums folgt also, unter der ggf. naiven Annahme, dass die Ausdehnungsgeschwindigkeit zu jedem Zeitpunkt gleich ist, dass das Universum einen Anfang gehabt haben muss, der kann 15 Milliarden Jahre zurückliegt. Dass das knapp daneben ist, liegt daran, dass die Ausdehnung nicht konstant ist, sondern sich im Laufe der Zeit verändert. Wie genau sich die Ausdehnungsgeschwindigkeit mit der Zeit ändert und wie sich das auf das Alter des Universums auswirkt, werden wir in den folgenden Kapiteln genauer betrachten.

2.5 Inflation

Das Hubble-Gesetz und die Friedmann-Gleichungen beschreiben die Expansion des Universums sehr erfolgreich, werfen aber auch Fragen auf. Zwei der bekanntesten:

- **Horizont-Problem**: Warum sehen wir in entgegengesetzten Richtungen des Himmels (CMB) nahezu exakt dieselbe Temperatur, obwohl diese Regionen, die gemäß der „normalen“ Expansionsgeschichte, niemals in kausalem Kontakt gestanden haben sollten?
- **Flachheits-Problem**: Warum ist die Geometrie des Universums (soweit wir messen können) extrem nahe an „flach“ ($\sum_i \Omega_i \approx 1$, vgl. Kapitel 8), obwohl diese Eigenschaft mit der Zeit eigentlich instabil sein sollte?

Eine elegante Lösung für beide Probleme ist die *Inflation*: eine extrem kurze Phase (Bruchteile einer Sekunde) ganz am Anfang des Universums, in der der Skalenfaktor $a(t)$ *exponentiell* anwächst, $a(t) \propto \mathrm{e}^{H_{\mathrm{inf}}\,t}$ mit nahezu konstantem H_{inf}. Eine solche exponentielle Expansion entspricht, wie wir in Kapitel 8 sehen werden, einem Universum, dessen Energiedichte von einer (nahezu) konstanten Vakuumenergie dominiert wird.

Am Ende der Inflation muss diese im Inflationsfeld gespeicherte Energie in die Teilchen des Standardmodells umgewandelt werden; dieser Prozess heißt *Reheating* und ist selbst eine Art

Phasenübergang (Kapitel 18). Erst nach dem Reheating beginnt die „klassische“ heiße Expansionsgeschichte, die wir in der Zeitleiste (Kapitel 1.5) zusammengefasst haben.

3 Differentialrechnung

Bevor wir die Friedmann-Gleichungen lösen können, brauchen wir handwerkliches Rüstzeug aus der Analysis. Wir beginnen mit der Differentialrechnung; in Kapitel 4 folgt die Integralrechnung, und in Kapitel 7 bauen wir beides zu Taylor-Reihen zusammen.

3.1 Ableitungen: eine kurze Wiederholung

Die Ableitung $f'(x) = \frac{\mathrm{d}f}{\mathrm{d}x}(x)$ einer Funktion f an einer Stelle x gibt die Steigung der Tangente an den Graphen von f in diesem Punkt an. Formal ist sie definiert als

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}\,. \tag{3.1}$$

Die wichtigsten Rechenregeln, die wir im Kurs immer wieder brauchen werden:

Ableitungsregeln

- Potenzregel: $\frac{\mathrm{d}}{\mathrm{d}x} x^n = n\,x^{n-1}$
- Kettenregel: $\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x)) \cdot g'(x)$
- Produktregel: $\frac{\mathrm{d}}{\mathrm{d}x}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x)$
- Exponentialfunktion: $\frac{\mathrm{d}}{\mathrm{d}x} \mathrm{e}^{\beta x} = \beta\,\mathrm{e}^{\beta x}$

3.2 Differentialgleichungen: Gleichungen, deren Lösungen Funktionen sind

Eine *gewöhnliche Differentialgleichung* (englisch *ordinary differential equation*, kurz ODE) ist, im Gegensatz zu „gewöhnlichen“ Gleichungen wie $ax^2 + bx + c = 0$, deren Lösungen Zahlen sind, eine Gleichung, deren Lösungen *Funktionen* sind, und die Ableitungen dieser Funktionen enthält. Ein Beispiel:

$$f'(x) = \beta\, f(x)\,. \tag{3.2}$$

Gesucht ist eine Funktion f, die, eingesetzt in diese Gleichung, eine wahre Aussage liefert. Genau solche Gleichungen werden uns in der Kosmologie ständig begegnen: Die Friedmann-Gleichungen (Kapitel 8) sind nichts anderes als ODEs für den Skalenfaktor $a(t)$. Mehr noch: Unsere Naturgesetze sind als Differentialgleichungen formuliert! Wenn wir verstehen wollen, wie die Natur funktioniert, kommen wir um Differentialgleichungen also nicht herum.

Aufgabe 3.1

1. Zeige, dass für alle $\alpha, \beta \in \mathbb{R}$ die Funktion $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \alpha \exp(\beta \cdot x)$ die Differentialgleichung

   $$f' = \beta f$$

   löst.

2. Finde eine Lösung der Differentialgleichung

   $$\ddot{f} = 2f\,,$$

   sodass die Anfangswertbedingung $f(0) = 1$ erfüllt ist. Ist die Lösung eindeutig?

Musterlösung 3.1

1. Mit der Kettenregel (und $\frac{\mathrm{d}}{\mathrm{d}x}(\beta x) = \beta$) gilt

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\alpha \mathrm{e}^{\beta x}\right] = \alpha\beta\,\mathrm{e}^{\beta x} = \beta \underbrace{\alpha \mathrm{e}^{\beta x}}_{=f(x)} = \beta f(x)\,, \tag{3.3}$$

also löst $f(x) = \alpha \mathrm{e}^{\beta x}$ tatsächlich $f' = \beta f$, für *beliebige* α, β.
2. Für $\ddot{f} = 2f$ machen wir denselben Ansatz $f(x) = \alpha \mathrm{e}^{\beta x}$. Zweimaliges Ableiten (Kettenregel) gibt

$$\ddot{f}(x) = \alpha\beta^2 \mathrm{e}^{\beta x} = \beta^2 f(x)\,. \tag{3.4}$$

Damit $\ddot{f} = 2f$ gilt, brauchen wir $\beta^2 = 2$, also $\beta = \pm\sqrt{2}$. Die Anfangsbedingung $f(0) = \alpha \mathrm{e}^0 = \alpha \overset{!}{=} 1$ liefert $\alpha = 1$. Damit sind sowohl

$$f_+(x) = \mathrm{e}^{\sqrt{2}x} \qquad \text{als auch} \qquad f_-(x) = \mathrm{e}^{-\sqrt{2}x} \tag{3.5}$$

gültige Lösungen, die beide $f(0) = 1$ erfüllen! Die Lösung ist also *nicht eindeutig* – eine ODE zweiter Ordnung (höchste auftretende Ableitung: zweite Ableitung) braucht im Allgemeinen *zwei* Anfangsbedingungen (z.B. $f(0)$ *und* $\dot{f}(0)$), um eine eindeutige Lösung festzulegen.

3.3 Die Tangente als lineare Approximation

Eine geometrische Sichtweise auf die Ableitung, die für Kapitel 7 (Taylor-Reihen) zentral wird: Die Tangente an den Graphen von f im Punkt a ist die *beste lineare Approximation* von f in der Nähe von a (Abbildung 3.1).

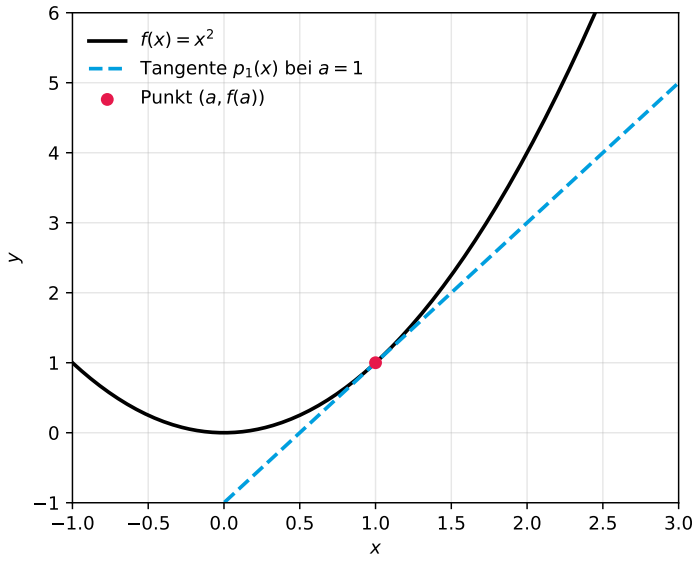


Abbildung 3.1: Die Funktion $f(x) = x^2$ (durchgezogen) und ihre Tangente $p_1(x)$ im Punkt $a = 1$ (gestrichelt). In unmittelbarer Nähe von a ist die Tangente praktisch nicht von f zu unterscheiden; sie stimmt mit f im Punkt a sowohl im Funktionswert als auch in der Steigung überein. Weiter entfernt von a weichen Funktion und Tangente jedoch zunehmend voneinander ab.

Aufgabe 3.2

Gegeben sei eine differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$. Bestimme die Funktionsvorschrift der Tangente $p_1(x) = m \cdot x + b$ von f am Punkt $a \in \mathbb{R}$.
Hinweis: Mache dir klar, dass die Tangente folgende zwei Gleichungen erfüllt:

$$p_1(a) = f(a) \qquad \text{und} \qquad p_1'(a) = f'(a)\,.$$

Was bedeutet das geometrisch?

Musterlösung 3.2

Die zweite Bedingung, $p_1'(a) = f'(a)$, legt die Steigung fest: Da $p_1'(x) = m$ für alle x (Ableitung einer linearen Funktion ist konstant), muss $m = f'(a)$ gelten. Die erste Bedingung, $p_1(a) = f(a)$, legt dann den y-Achsenabschnitt b fest:

$$p_1(a) = f'(a) \cdot a + b \overset{!}{=} f(a) \quad \Longrightarrow \quad b = f(a) - f'(a) \cdot a\,. \tag{3.6}$$

Insgesamt:

$$p_1(x) = f(a) + f'(a) \cdot (x - a)\,. \tag{3.7}$$

Geometrisch bedeutet das: Die Tangente p_1 „berührt“ den Graphen von f im Punkt $(a, f(a))$ und hat dort dieselbe Steigung wie f; sie ist also die Gerade, die f in unmittelbarer Nähe von a am besten approximiert. Diese Idee, f durch ein Polynom zu approximieren, das an einer Stelle a in möglichst vielen Ableitungen mit f übereinstimmt – werden wir in Kapitel 7 auf Polynome höheren Grades verallgemeinern.

4 Integralrechnung

> „*Differentiation ist Arbeit, Integration ist Kunst.*“

Das Ableiten einer Funktion folgt einem Regelwerk: Potenzregel, Produktregel, Kettenregel; jede Ableitung lässt sich mechanisch ausführen. Das Integrieren ist das Gegenteil: Es gibt kein allgemeines Rezept. Ob man ein Integral in geschlossener Form lösen kann, hängt von Erfahrung, der richtigen Substitution und häufig auch einfach von Glück ab. Genau deshalb spielen *numerische* Integrationsmethoden eine so große Rolle, und am Ende des Kurses werden wir mit *Monte-Carlo-Integration* (Kapitel 15) sogar Integrale in hochdimensionalen Räumen lösen, die analytisch schlicht nicht handhabbar wären. In der Schule werden diese Integrale nicht behandelt; in der „realen Welt“ der Technik und Forschung hingegen ginge nichts ohne sie.

4.1 Das bestimmte Integral

Während die Ableitung die *Steigung* einer Funktion misst, misst das *Integral* die *Fläche* unter ihrem Graphen (Abbildung 4.1). Die folgende Aufgabe rekonstruiert die formale Definition des bestimmten Integrals, in Form eines kleinen Puzzles.

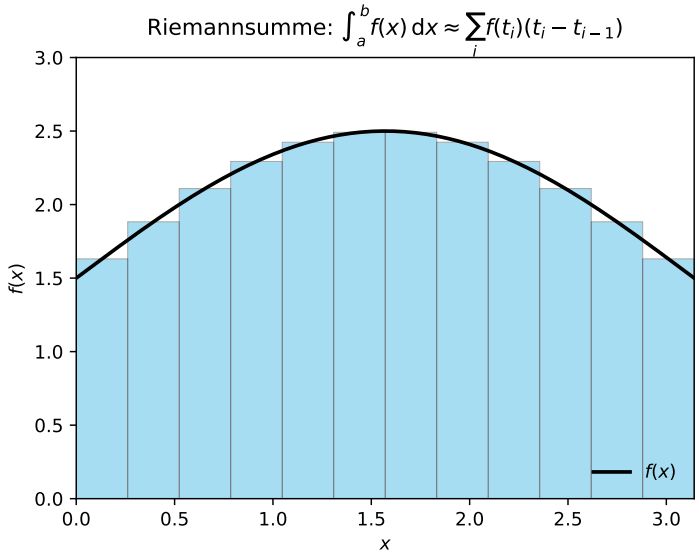


Abbildung 4.1: Eine Riemannsumme approximiert die Fläche unter dem Graphen von f (schwarz) durch schmale Rechtecke der Breite $(t_i - t_{i-1})$ und Höhe $f(t_i)$. Je feiner die Zerlegung des Intervalls $[a, b]$ (je kleiner $h = \max_i |t_i - t_{i-1}|$), desto besser approximiert die Summe der Rechteckflächen die tatsächliche Fläche; im Grenzwert $h \to 0$ erhält man das Integral $\int_a^b f(x)\,dx$.

Aufgabe 4.1

Oh nein! :(Das Unwetter von gestern hat die Definition des bestimmten Integrals von a bis b einer stetigen Funktion $f : \mathbb{R} \to \mathbb{R}$ völlig durcheinandergebracht:

$$f(t_i)\,, \quad f(t)\,dt\,, \quad \sum_{i=1}^{n}\,, \quad \lim_{h\to 0}\,, \quad \int_a^b\,, \quad :=\,, \quad (t_i - t_{i-1})\,,$$

wobei $a = t_0, \ldots, b = t_n$ eine Zerlegung des Intervalls $[a, b]$ ist und $h := \max_i |t_i - t_{i-1}|$. Setzt die Definition wieder richtig zusammen. Zeichnet (mindestens) ein Bild dazu und tauscht

euch darüber aus, was ihr über Integrale wisst.

Musterlösung 4.1

$$\int_a^b f(t)\,\mathrm{d}t := \lim_{h\to 0} \sum_{i=1}^{n} f(t_i)\cdot(t_i - t_{i-1})\,. \tag{4.1}$$

Jeder Summand $f(t_i)\cdot(t_i - t_{i-1})$ ist der Flächeninhalt eines schmalen Rechtecks mit Höhe $f(t_i)$ und Breite $(t_i - t_{i-1})$. Die Summe all dieser Rechtecke (eine sogenannte *Riemannsumme*) approximiert die Fläche unter dem Graphen von f als „Treppenfunktion“. Je feiner die Zerlegung wird (also je kleiner h wird), desto besser wird diese Approximation; im Grenzwert $h \to 0$ erhalten wir die exakte Fläche, das Integral $\int_a^b f(t)\,\mathrm{d}t$.

4.2 Der Hauptsatz der Differential- und Integralrechnung

Der folgende Satz verbindet Differential- und Integralrechnung und macht das Integral für uns praktisch berechenbar:

> **Hauptsatz der Differential- und Integralrechnung (Teil 1).** Sei $F : I = [a,b] \to \mathbb{R}$ eine differenzierbare Funktion und $f : I \to \mathbb{R}$ die Ableitung von F, d.h. es gelte $F' = f$. Dann kann für beliebige $a, b \in I$ das bestimmte Integral wie folgt berechnet werden:
>
> $$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)\,. \tag{4.2}$$

Aufgabe 4.2

Berechnet (mit Hilfe des Hauptsatzes der Differential- und Integralrechnung) die folgenden zwei Integrale:

$$\int_{-1}^{1} x^2\,\mathrm{d}x\,, \qquad \int_1^2 \frac{2}{x}\,\mathrm{d}x\,.$$

Musterlösung 4.2

Für das erste Integral suchen wir eine Stammfunktion F_1 mit $F_1'(x) = x^2$. Mit der Potenzregel ist $F_1(x) = \frac{1}{3}x^3$ eine solche Stammfunktion (denn $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{3}x^3 = \frac{3}{3}x^2 = x^2$). Damit

$$\int_{-1}^{1} x^2\,\mathrm{d}x = F_1(1) - F_1(-1) = \frac{1}{3} - \left(-\frac{1}{3}\right) = \frac{2}{3}\,. \tag{4.3}$$

Für das zweite Integral ist $F_2(x) = 2\ln x$ eine Stammfunktion von $f(x) = \frac{2}{x}$ (denn $\frac{\mathrm{d}}{\mathrm{d}x}\ln x = \frac{1}{x}$). Damit

$$\int_1^2 \frac{2}{x}\,\mathrm{d}x = F_2(2) - F_2(1) = 2\ln 2 - 2\ln 1 = 2\ln 2\,, \tag{4.4}$$

da $\ln 1 = 0$.

4.3 Differentialgleichungen lösen: Separation der Variablen

Eine der wichtigsten Techniken, um ODEs zu lösen (und die Technik, mit der wir in Kapitel 8 die Friedmann-Gleichungen lösen werden) ist die *Separation der Variablen*. Die Idee: Wir benutzen den (von Physikern gerne und unbekümmert verwendeten) „Trick“, $\frac{\mathrm{d}f}{\mathrm{d}t}$ wie einen *Bruch* zu behandeln und mit $\mathrm{d}t$ „durchzumultiplizieren“.

Aufgabe 4.3

1. Gegeben sei die folgende ODE:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \sqrt{f}\,.$$

   Finde mit Hilfe des Hauptsatzes der Differential- und Integralrechnung eine Lösung.

2. Funktioniert dieser Lösungsansatz auch bei der ODE

$$\left(\frac{\dot{f}}{f}\right)^2 = f^{-4} + f^{-3} + 1 \quad ?$$

Hinweis: Vielleicht hilft es, den bereits im Kurs erwähnten ominösen Trick der Physiker „Multiplizieren mit $\mathrm{d}t$“ zu benutzen.

Musterlösung 4.3

1. Wir „multiplizieren mit $\mathrm{d}t$“ und teilen durch $\sqrt{f}$:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \sqrt{f} \quad\Longrightarrow\quad \mathrm{d}f = \sqrt{f}\,\mathrm{d}t \quad\Longrightarrow\quad \frac{\mathrm{d}f}{\sqrt{f}} = \mathrm{d}t\,. \tag{4.5}$$

Nun integrieren wir beide Seiten: links über f, rechts über t:

$$\int f^{-1/2}\,\mathrm{d}f = \int \mathrm{d}t \quad\Longrightarrow\quad 2\sqrt{f} = t + C\,, \tag{4.6}$$

wobei wir die Stammfunktion $\int f^{-1/2}\,\mathrm{d}f = 2f^{1/2}$ benutzt haben (Probe: $\frac{\mathrm{d}}{\mathrm{d}f}\left[2f^{1/2}\right] = 2\cdot\frac{1}{2}f^{-1/2} = f^{-1/2}$ ✓). Auflösen nach f liefert

$$f(t) = \left(\frac{t+C}{2}\right)^2. \tag{4.7}$$

Probe: $\frac{\mathrm{d}f}{\mathrm{d}t} = 2\cdot\frac{t+C}{2}\cdot\frac{1}{2} = \frac{t+C}{2} = \sqrt{f(t)}$ ✓.

2. Wir versuchen denselben Trick. Zunächst formen wir um:

$$\dot{f}^2 = f^2\left(\frac{1}{f^4} + \frac{1}{f^3} + 1\right) \quad\Longrightarrow\quad \dot{f} = f\sqrt{f^{-4} + f^{-3} + 1}\,. \tag{4.8}$$

„Multiplizieren mit $\mathrm{d}t$“ und Trennen der Variablen gibt

$$\frac{\mathrm{d}f}{f\sqrt{f^{-4} + f^{-3} + 1}} = \mathrm{d}t \quad\Longrightarrow\quad \int \frac{\mathrm{d}f}{f\sqrt{f^{-4} + f^{-3} + 1}} = \int \mathrm{d}t = t + C\,. \tag{4.9}$$

Hier kommen wir nicht weiter: Das Integral auf der linken Seite hat *keine* elementare (also durch endlich viele „bekannte“ Funktionen ausdrückbare) Stammfunktion! Der Trick funktioniert also *formal* genauso wie in Teil 1, jedoch lässt sich das entstehende Integral nicht analytisch lösen.

Wichtig für später: Diese zweite ODE ist kein Zufall: sie ist (bis auf Umbenennung von f in den Skalenfaktor a) genau die Friedmann-Gleichung für ein Universum mit Strahlung, Materie *und* Vakuumenergie zugleich (Kapitel 8). Dass sie nicht analytisch lösbar ist, ist der Grund, warum wir in Kapitel 8.6 das Alter des Universums *numerisch* berechnen müssen.

4.4 Numerische Integration

Nicht jedes Integral lässt sich mit dem Hauptsatz lösen, entweder weil wir keine Stammfunktion finden (wie eben gesehen), oder weil die Funktion gar nicht in geschlossener Form vorliegt (z.B. Messdaten). In solchen Fällen müssen wir das Integral *numerisch* approximieren, also direkt über die Riemannsummen-Idee von Aufgabe 4.1, nur cleverer.

Die *Simpson-Regel* approximiert die Funktion stückweise durch Parabeln statt durch Rechtecke (oder Trapeze) und ist dadurch für glatte Funktionen sehr genau:

Aufgabe 4.4

Informiert euch (z.B. auf Wikipedia) über die Trapez- und Simpson-Regel. Implementiert die Simpson-Regel in `python`. Testet die Methode an den Integralen aus Aufgabe 4.2 und 4.3.

Musterlösung 4.4

Eine mögliche Implementierung der Simpson-Regel für N (gerade) Teilintervalle:

```
def simpson(f, a, b, N):
    '''Approximiere das Integral von f(x) von a bis b
    mit der Simpson-Regel'''
    if N % 2 != 0:
        print("N ist nicht gerade!")
        return np.nan
    else:
        h = (b - a) / N
        I = f(a) + f(b)
        x_k = a + h
        for k in range(1, round(N/2) + 1):
            I += 4 * f(x_k)
            x_k += 2 * h
        x_k = a + 2 * h
        for k in range(1, round(N/2)):
            I += 2 * f(x_k)
            x_k += 2 * h
        return (h / 3) * I
```

Test an $\int_{-1}^{1} x^2 \, \mathrm{d}x = \frac{2}{3}$: `simpson(lambda x: x**2, -1, 1, 100)` liefert (bis auf Rundungsfehler) ≈ 0.6667. Diese Funktion werden wir in Kapitel 8.6 wiederverwenden, um das nicht-analytisch-lösbare Integral aus Aufgabe 4.3 (Teil 2), dort als Integral für das Alter des Universums t_0, numerisch auszuwerten.

4.5 Existenz von Lösungen: der Satz von Picard–Lindelöf

Warum haben ODEs überhaupt Lösungen? Diese Frage führt auf einen fundamentalen Satz der Analysis, den *Satz von Picard–Lindelöf*, der Existenz und Eindeutigkeit von Lösungen für Lipschitz-stetige rechte Seiten garantiert. Den vollständigen Beweis (inklusive des Banachschen

Fixpunktsatzes als zentrales Hilfsmittel und einer Programmierübung zur Picard-Iteration) findet ihr in Anhang A.

5 Natürliche Einheiten

Im Folgenden wollen wir uns anhand einiger alltagsnaher Überlegungen dem (zugegebenermaßen anfangs eher unbequemen) Thema der natürlichen Einheiten widmen. Früher oder später lernen jedoch alle im Rahmen eines Physikstudiums, dass sich das anfängliche Leiden, gepaart mit einiger Verwirrung, sich dann doch auszahlt. Das wollen wir nun auch nachempfinden. Also los!

5.1 Länge und Zeit

Dortmund und Aachen sind 2h voneinander entfernt. Dieser Satz ergibt keinen Sinn. Dortmund und Aachen sind zwei Städte in Deutschland. Wie können zwei Städte, zwei Punkte im geometrischen Raum, ein Zeitintervall voneinander entfernt sein? Städte sind doch eigentlich nur räumlich voneinander entfernt? Hat das schon etwas mit Raumzeit zu tun?

Natürlich *nicht*. Wenn wir sagen, dass Dortmund und Aachen 2h voneinander entfernt sind, meinen wir *eigentlich*, dass es mit dem Auto 2h dauert um von der einen in die andere Stadt zu fahren. Was hier nach einem No-brainer klingt, ist eigentlich eine ganz schön verblüffende Erkenntnis: Haben wir eine Referenzgeschwindigkeit und sind wir uns alle über diese Referenzgeschwindigkeit einig (im Beispiel: dass ein Auto im Schnitt 100 km/h zurücklegt), werden auf einmal Längendifferenzen und Zeitintervalle die gleiche Größe; oder zumindest können wir beide Angaben äquivalent zueinander verstehen:

$$2\,\mathrm{h} = 200\,\mathrm{km}\,. \tag{5.1}$$

Was hier wie eine üble Missachtung der Mathematik aussieht, ist eigentlich gar kein Problem, solange wir die obige Gleichung nicht etwa als $2 = 200$ oder $1\,\mathrm{h} = 1\,\mathrm{km}$ falsch verstehen: Nur weil ein Auto ca. 100 km/h zurücklegt, heißt das noch nicht, dass Stunden und Kilometer das Gleiche sind oder wir Probleme beim Zählen bekommen. Es bedeutet doch lediglich, dass wir von nun an auch ganz andere Entfernungen mit Zeitintervallen angeben können.

Widmen wir uns nun zum Beispiel der Frage, wie weit die Sonne von der Erde entfernt ist. Schlagen wir dies bei Wikipedia nach, bekommen wir prompt die folgende Antwort: $d_{\mathrm{SE}} = 149,597,870\,\mathrm{km}$. Nun ist es nicht unbedingt die *natürlichste* Art der Fortbewegung mit dem Auto zur Sonne zu fahren. (Würde man es dennoch tun, würde man eine mäßig interessante zeitliche Entfernung von $\Delta t_{\mathrm{SE}} = 149,597,870\,\mathrm{km} = 149,599\,\mathrm{h}$ erhalten.)

Nun ist es so, dass die Sonne vor allem[1] Licht produziert, durch welches wir sie dann beobachten können. Und da wir uns nicht mit Autos durch den Weltraum bewegen, ist es hier eine deutlich *natürlichere* Einheit, die Lichtgeschwindigkeit

$$c = 299,792.458\,\frac{\mathrm{km}}{\mathrm{s}} \tag{5.2}$$

als allgemein akzeptierte Geschwindigkeit der Fortbewegung zu wählen. Tatsächlich ist es eine fundamentale Annahme, dass die Lichtgeschwindigkeit im Vakuum an jedem Ort und zu jeder Zeit immer den gleichen Wert hat; anders also als jedes Auto. Ohne diese Annahme, würde die Allgemeine Relativitätstheorie nicht funktionieren. Die folgende Rechnung ist also, wenn man so will, *natürlich*: Nutzen wir diese Naturkonstante als Umrechnungsfaktor, erhalten wir

$$d_{\mathrm{SE}} = 149,597,870\,\mathrm{km} = 149,597,870\,\mathrm{km} = 499\,\mathrm{s} = 8\,\mathrm{min}\,19\mathrm{s}\,. \tag{5.3}$$

Diese Zahl ist doch sogar ganz handlich! Das Licht braucht also knapp acht Minuten von der Sonne bis zu uns. Anders ausgedrückt: Die Sonne ist acht Lichtminuten von uns entfernt.

Nochmal ganz langsam! Was ist denn hier passiert? Wie kann denn mathematisch $149\,597\,870\,\mathrm{km} = 499\,\mathrm{s}$ hinhauen? Die Antwort ist eigentlich ganz einfach: Behandeln wir

[1] Sie strahlt auch eine gigantische Menge Neutrinos ab – mehr sogar eigentlich noch als Photonen, also Lichtteilchen. Doch dazu ein andermal.

die Einheiten hinter den Zahlen als Variablen, können wir diese nach Belieben algebraisch manipulieren. Versuchen wir es einmal an dem gegebenen Beispiel:

$$149,597,870\,\mathrm{km} = 499\,\mathrm{s} \tag{5.4}$$

$$\Leftrightarrow \quad \frac{149,597,870}{499}\frac{\mathrm{km}}{\mathrm{s}} = 1 \tag{5.5}$$

$$\Leftrightarrow \quad 299,792.458\,\mathrm{km/s} = 1 \tag{5.6}$$

$$\Leftrightarrow \quad c = 1 \tag{5.7}$$

Wow! Da steht nun ganz eindeutig dass $c = 1$ gilt.

Wem sich bei diesem Anblick zunächst einmal der Magen umdreht: Seid beruhigt. Ihr seid in guter Gesellschaft. Dieses Ergebnis zu akzeptieren (und vielmehr: es zu benutzen) braucht ein wenig Zeit. Im Physikstudium wird man mit diesem Ergebnis erst im fünften Semester konfrontiert, wenn man sich in der theoretischen Physik spezialisiert. Spätestens ab dem sechsten Semester dann hat man allerdings auch schon vergessen, welchen Wert die Lichtgeschwindigkeit c denn "eigentlich" im SI-Einheitensystem in den Einheiten km/s hatte.

Warum das? Bisher haben wir doch im Prinzip nur gesagt, dass das Licht acht Minuten von der Sonne bis zu uns benötigt? Stimmt. Bisher steht hier noch nichts besonders magisches. Lassen wir uns aber noch einen Moment länger auf die Konsequenzen von $c = 1$ ein. Wir können nun nicht nur den Abstand der Erde zur Sonne, sondern jeden beliebigen Abstand in Zeit-Einheiten angeben. Was wäre denn beispielsweise ein Meter in Sekunden?

$$c = \frac{299,792.458\,\mathrm{km}}{1\,\mathrm{s}} = \frac{299,792,458\,\mathrm{m}}{1\,\mathrm{s}} \tag{5.8}$$

$$\Leftrightarrow \quad 299,792,458\,\mathrm{m} = 1\,\mathrm{s} \tag{5.9}$$

$$\Leftrightarrow \quad 1\,\mathrm{m} = \frac{1\,\mathrm{s}}{299,792,458} = 3.3\,\mathrm{ns} \tag{5.10}$$

Ein Meter ist also in dem von $c = 1$ definierten Einheitensystem das Gleiche wie drei Nanosekunden. Wir können dies wieder physikalisch interpretieren: Das Licht benötigt drei Nanosekunden um eine Strecke von einem Meter zu durchqueren. Technisch gesehen ist es nun absolut korrekt zu sagen: „*Wow, Simon! Was haben dir deine Eltern nur zu Essen gegeben. Du bist fast sieben Nanosekunden groß!*".

5.2 Energie und Frequenz

Der tiefere Sinn dieses Einheitensystems kann sich einem nur ganz erschließen, wenn nicht nur eine (hier gewissermaßen willkürlich ausgewählte) Naturkonstante mathematisch auf 1 gesetzt wird. Gehen wir mal einen Schritt weiter (sind wir abenteuerlustig!) und setzen wir auch das allseits beliebte reduzierte Planck'sche Wirkungsquantum[2] $\hbar = 1$. Schauen wir auch hier wieder einmal auf Wikipedia nach: Dort finden wir[3]

$$\hbar = \frac{h}{2\pi} = 1.054,571,82 \times 10^{-34}\,\frac{\mathrm{m^2\,kg}}{\mathrm{s}} = 1.054,571,82 \times 10^{-34}\,\mathrm{J\,s}\,. \tag{5.11}$$

[2]Solltest du noch nie etwas von dieser Naturkonstante gehört haben: Gar kein Problem. An dieser Stelle sei nur gesagt, dass sie die fundamentale Größe der Quantenmechanik ist. Weil sie so klein ist, spielt Quantenmechanik in unserer Erfahrungswelt des Alltags keine Rolle. Wäre $\hbar$ größer, würden wir womöglich auch durch Wände tunneln können... Also angenommen, die Atome und Moleküle aus denen wir bestehen, würden immer noch stabil sein und sich nicht einfach in nullkommanichts in „Luft" auflösen.

[3]Warum ist ein Joule das gleiche wie $1\,\mathrm{kgm^2/s^2}$? Schaue dir dafür z.B. einmal an, welche Energie $E = F \cdot d$ benötigt wird um eine Masse mit einer Kraft von $F = 1\,\mathrm{N} = 1\,\mathrm{kg\,m/s^2}$ um $d = 1\,\mathrm{m}$ fortzubewegen. Warum ist ein Newton das Gleiche wie $1\,\mathrm{kg\,m/s^2}$? Nimm dir dafür $F = m \cdot a$ zu Herzen und wähle eine Testmasse von einem Kilogramm und eine Beschleunigung von $1\,\mathrm{m/s^2}$.

Setzen wir nun das Planck'sche Wirkungsquantum auf 1, folgen daraus

$$1\,\mathrm{J} = \frac{1}{1.054,571,82 \times 10^{-34}\,\mathrm{s}} = 9.5 \times 10^{-35}\,\mathrm{s}^{-1} = 9.5 \times 10^{33}\,\mathrm{Hz} \tag{5.12}$$

und

$$1\,\mathrm{s} = \frac{1}{1.054,571,82 \times 10^{-34}\,\mathrm{J}} = 9.5 \times 10^{33}\,\mathrm{J}^{-1}\,. \tag{5.13}$$

Handelsübliche Energien entsprechen also gigantisch hohen Frequenzen. Um bei gewöhnlichen Frequenzen zu landen (sagen wir: Wie wir sie vom Radiohören können, irgendwas mit Gigahertz) brauchen wir also sehr sehr kleine Energien.

Vielleicht kommt dir die folgende Energieeinheit bekannt vor:

$$1\,\mathrm{eV} = 1.602,176,634 \times 10^{-19}\,\mathrm{J}\,. \tag{5.14}$$

Das ist die Energiemenge, die nötig ist um ein Elektron eine Potentialdifferenz von einem Volt heraufzubewegen (daher der Name: *ein Elektronenvolt*). Die Skala die diese Energie angibt ist die Skala, auf der die Quantenmechanik relevant ist: Als Vergleich dient hier zum Beispiel auch die Energie $E_{\mathrm{Ryd}} = 13.6\,\mathrm{eV}$, die ein Elektron in der „untersten Schale" im Wasserstoffatom besitzt, wenn man dem Bohrschen Atommodel folgt. Diese Energiemenge wird auch also Rydberg-Energie bezeichnet.

Nutzen wir nun einmal die in der Quantenmechanik etwas handlichere Einheit Elektronenvolt um die Konsequenzen von $\hbar = 1$ zu verdeutlichen:

$$\hbar = 1.054,571,82 \times 10^{-34}\,\mathrm{Js} = 6.582,119,569 \times 10^{-16}\,\mathrm{eV\,s}\,. \tag{5.15}$$

Daraus können wir nun direkt

$$1\,\mathrm{eV} = \frac{1}{6.582,119,569 \times 10^{-16}\,\mathrm{s}} = 1.52 \times 10^{15}\,\mathrm{Hz} = 1.52\,\mathrm{THz}\,, \tag{5.16}$$

und

$$1\,\mathrm{eV}^{-1} = 6.582,119,569 \times 10^{-16}\,\mathrm{s} = 0.658\,\mathrm{fs} \tag{5.17}$$

ableiten. Eine physikalische Interpretation dessen wird zum Beispiel durch die Heisenberg'sche Energie-Zeit-Unschärferelation $\Delta E \cdot \Delta t \geq \hbar/2$ möglich: Im Vakuum treten ständig Quantenfluktuationen auf. Eine Energie-Fluktuation in Höhe von 1eV wird dabei maximal auf der Femtosekunden-Skala auftreten. Größere Energiefluktuationen sind dementsprechend noch deutlich kurzlebiger. Gleichzeitig können wir auch die Wellenlänge der emittierten Photonen bei Übergängen im Wasserstoffatom berechnen: Die dadurch freigesetzten Energien sind auf der 10 eV-Skala und die freigesetzte Strahlung ist im 10 THz-Bereich. Wenn wir diese Strahlung messen wollen, müssen wir also einen Detektor bauen, der elektromagnetische Strahlung bei solch hohen Frequenzen messen kann.

Cool! Nur durch das Setzen von $\hbar = 1$ haben wir jetzt schon tiefe Einblicke in die Skala bekommen, auf der sich die Quantenmechanik abspielt und mit welcher Art von Experiment wir ihre Vorhersagen überprüfen können. Aber lasst uns doch noch einen Schritt weiter gehen. Was passiert, wenn wir nun $c = \hbar = 1$ setzen? Das sieht noch einmal verrückter aus als das, was wir bisher schon gemacht haben, aber gehen wir systematisch an die Sache heran: $c = 1$ hatte eigentlich nur zur Folge, dass Zeiten und Längen mathematisch die gleiche „Dimension" angeben. Die Gleichung $\hbar = 1$ hatte zur Folge, dass Zeiten und (inverse) Energien die gleiche Dimension angeben. Kombinieren wir beides miteinander, können wir nun Längen mit (inversen) Energie-Einheiten beschreiben!

Im Beispiel von oben bedeutet das ganz plastisch

$$E_{\mathrm{Ryd}} = 13.6\,\mathrm{eV} = 20.6\,\mathrm{THz} = \frac{1}{14.55\,\mathrm{nm}}\,. \tag{5.18}$$

Wir bekommen also das Ergebnis, dass die elektromagnetische Strahlung bei Übergängen im Wasserstoffatom irgendwie mit der 10 nm-Skala zusammenhängt. Eine weitere schnelle Wikipedia-Suche später stellen wir fest, dass das die Wellenlänge von Licht im ultravioletten Bereich, kurz unterhalb des sichtbaren Spektrums ist. Wow! Wir haben bisher noch kein Naturgesetz benutzt und doch wissen wir schon, dass wir irgendeine Art von Effekt mit UV-Licht erwarten können. *Es ist und bleibt magisch.*

5.3 Temperatur und Energie

In der theoretischen Hochenergie-Physik und Kosmologie wird standardmäßig ein Einheitensystem benutzt, das sich *natürliche Einheiten* nennt und durch das Gleichsetzen von

$$c = \hbar = k_{\mathrm{B}} = 1 \tag{5.19}$$

definitert ist. Die letzte Naturkonstante darin ist die Boltzmann-Konstante

$$k_{\mathrm{B}} = 1.381 \times 10^{-23}\,\frac{\mathrm{J}}{\mathrm{K}} = 86.17\,\frac{\mu\mathrm{eV}}{\mathrm{K}}\,, \tag{5.20}$$

welche es erlaubt die mittlere thermische Energie eines Teilchens in einem Gas (oder eben dem primordialen Plasma aus Elementarteilchen) zu berechnen. Über den Daumen gilt $E_{\mathrm{therm}} \simeq k_{\mathrm{B}}\,T$, wobei T die Temperatur des Gases beschreibt[4]. Bei Raumtemperatur finden wir also, dass ein Teilchen eine thermische Energie von ca. $293\,\mathrm{K} \approx 25\,\mathrm{meV}$ besitzt. Wild. Und ganz schön wenig kinetische Energie, siehe zum Beispiel die obige Umrechnung von Joule in eV.

Aber es wird noch wilder! Dadurch, dass wir aus der Gleichung

$$1\,\mathrm{J} = 1\,\mathrm{Nm} = 1\,\mathrm{kg}\frac{\mathrm{m}}{\mathrm{s}^2}\mathrm{m} = 1\,\mathrm{kg}\frac{\mathrm{m}^2}{\mathrm{s}^2} \tag{5.21}$$

gewissermaßen den Meter-pro-Sekunde-Bruch rauskürzen können, weil ja $c \approx 3 \cdot 10^8\,\mathrm{m/s} = 1$, finden wir

$$1\,\mathrm{J} = 1\,\mathrm{kg}\frac{\mathrm{m}^2}{\mathrm{s}^2} \cdot \frac{1}{c^2} = 1\,\mathrm{kg}\frac{\mathrm{m}^2}{\mathrm{s}^2}\frac{\mathrm{s}^2}{2.997^2 \cdot 10^{16}\,\mathrm{m}^2} = 1.113 \times 10^{-17}\,\mathrm{kg}\,. \tag{5.22}$$

Energie und Masse *sind* das gleiche! Das erinnert natürlich schon sehr an Einstein's berühmte Gleichung $E = mc^2$.

5.4 Energie als universelle Einheit

Durch die Kombination dieser drei Naturkonstanten können wir nach ein bisschen Algebra die Umrechnungsfaktoren in Tabelle 5.1 herleiten.

[4] Wenn man es ganz genau nimmt, gilt $E_{\mathrm{therm}} = \frac{3}{2}k_{\mathrm{B}}\,T$ für ein monoatomares Gas im dreidimensionalen Raum. Bei mehr Freiheitsgraden wird der Faktor 3/2 entsprechend größer.

Größe	**Einheit (geschrieben)**	**Wert in SI-Einheiten**
Energie	1 eV	$1{,}602 \cdot 10^{-19}$ J
Länge	$\frac{1}{1\ \mathrm{eV}}$	$1{,}973 \cdot 10^{-7}$ m
Zeit	$\frac{1}{1\ \mathrm{eV}}$	$6{,}582 \cdot 10^{-16}$ s
Masse	1 eV	$1{,}783 \cdot 10^{-36}$ kg
Dichte	$1\ \mathrm{eV}^4$	$2{,}320 \cdot 10^{-16}$ kg/m^3
Impuls	1 eV	$5{,}344 \cdot 10^{-28}$ N · s
Temperatur	1 eV	$1{,}160 \cdot 10^{4}$ K

Tabelle 5.1: Umrechnungsfaktoren für die wichtigsten physikalischen Größen zwischen natürlichen Einheiten ($c = \hbar = k_B = 1$, Basiseinheit eV) und SI-Einheiten.

Aufgabe 5.1

Das Newtonsche Gravitationsgesetz lautet

$$F_G = G_\mathrm{N} \frac{mM}{r^2}\,,$$

wobei

$$G_\mathrm{N} = 6.67 \cdot 10^{-11}\,\frac{\mathrm{m}^3}{\mathrm{kg\,s}^2}$$

die Gravitationskonstante ist. In unseren natürlichen Einheiten mit $c = k_\mathrm{B} = \hbar = 1$ gilt ferner (vgl. Tabelle 5.1)

$$1\,\mathrm{eV} = 1.16 \cdot 10^4\,\mathrm{K} = 1.8 \cdot 10^{-36}\,\mathrm{kg} = 1.5 \cdot 10^{15}\,\mathrm{s}^{-1} = 5.1 \cdot 10^6\,\mathrm{m}^{-1}\,.$$

a) Gib G_N in natürlichen Einheiten an. „In natürlichen Einheiten“ heißt „in Einheiten von eV^α“, wobei der Exponent α zunächst zu bestimmen ist. Mache dir dazu klar, dass $c = 1$ bewirkt, dass Längen und Zeiten in denselben Einheiten angegeben werden können, sodass $\mathrm{m}^3/(\mathrm{kg\,s}^2)$ gewissermaßen um $\mathrm{m}^2/\mathrm{s}^2$ gekürzt werden kann (unter Berücksichtigung des nötigen Faktors). Setze dann für die verbliebenen SI-Einheiten die entsprechenden Werte in natürlichen Einheiten ein.

b) Welchen Wert hat $1/\sqrt{G_\mathrm{N}}$ in natürlichen Einheiten? Dieser Wert wird als Planck-Energie, oder auch als Planck-Masse M_Pl bezeichnet.

c) Oh nein! Ein Tornado ist durch die Planck-Skala gewirbelt. :(Dadurch ist in der folgenden Tabelle die Zuordnung zwischen Zahlenwerten und dazugehörigen SI-Einheiten durcheinandergekommen. Ordne die folgenden Zahlenwerte den Planck-Einheiten der linken Spalte zu:

$$5.4 \cdot 10^{-44}, \quad 2.2 \cdot 10^{-8}, \quad 1.6 \cdot 10^{-35}, \quad 1.5 \cdot 10^{32}.$$

Planck-Einheit	Zahlenwert	Einheit
M_Pl		kg
T_Pl		K
L_Pl		m
t_Pl		s

Die Planck-Skala beschreibt die minimale Längen- und Zeitskala sowie die maximale Energie-, Massen- und Temperaturskala, bis zu der wir Allgemeine Relativitätstheorie und Quantenfeldtheorie als separate Theorien verstehen können. Unsere Naturgesetze lassen sich demnach erst auf Zeitskalen, die deutlich größer als eine Planck-Zeit t_{Pl} (z.B. nach der Anfangssingularität), sicher anwenden.

Musterlösung 5.1

a) Mit $c = 1$ gilt $3 \cdot 10^8\,\mathrm{m/s} = 1$, also können wir $\mathrm{m}^3/(\mathrm{kg\,s}^2)$ zu $(3 \cdot 10^8)^{-2}\,\mathrm{m/kg}$ „kürzen“. Einsetzen der natürlichen Einheiten:

$$\begin{aligned} G_{\mathrm{N}} &= 6.67 \times 10^{-11}\,\frac{\mathrm{m}^3}{\mathrm{kg\,s}^2} = \frac{6.67 \times 10^{-11}}{(2.998 \cdot 10^8)^2} \cdot \frac{5.1 \times 10^6}{1\,\mathrm{eV}} \cdot \frac{1.8 \times 10^{-36}}{1\,\mathrm{eV}} \\ &\approx 6.8 \times 10^{-57}\,\mathrm{eV}^{-2} = 6.8 \times 10^{-39}\,\mathrm{GeV}^{-2}\,. \end{aligned}$$

Also $\alpha = -2$.
b) $1/\sqrt{G_{\mathrm{N}}} \approx 1/\sqrt{6.8 \times 10^{-39}\,\mathrm{GeV}^{-2}} \approx 1.2 \times 10^{19}\,\mathrm{GeV} \approx M_{\mathrm{Pl}}$. Das ist die Planck-Masse!
c) Zuordnung:

Planck-Einheit	Zahlenwert	Einheit
M_{Pl}	2.2×10^{-8}	kg
T_{Pl}	1.5×10^{32}	K
L_{Pl}	1.6×10^{-35}	m
t_{Pl}	5.4×10^{-44}	s

Aufgabe 5.2

Hinweis: In der folgenden Aufgabe vergleichen wir Teilchen*massen* m (normalerweise in kg oder MeV/c^2) direkt mit *Temperaturen* T (normalerweise in Kelvin) und schreiben beide einfach in GeV. Das ist kein Zufall: Wie wir gerade in diesem Kapitel gesehen haben, gilt in den *natürlichen Einheiten* $c = \hbar = k_B = 1$, sodass Massen, Energien und Temperaturen alle dieselbe Einheit (z.B. eV oder GeV) tragen und direkt vergleichbar werden.
In einem heißen Plasma, wie es im frühen Universum vorlag, können nur Teilchenspezies mit Masse $m \lesssim T$ effektiv produziert werden und tragen damit signifikant zur Energiedichte bei; Teilchen mit $m \gg T$ sind exponentiell unterdrückt (vgl. Kapitel 8). Schau dir die Massen der Teilchen des Standardmodells an (z.B. in Abbildung 1.2) und überlege:

1. Welche Teilchenspezies sind bei einer Temperatur von $T = 10^{16}\,\mathrm{GeV}$ (ungefähr die Temperatur am Ende der Inflation) Teil des Plasmas?
2. Welche bei $T = 10\,\mathrm{GeV}$?
3. Welche bei $T = 0.1\,\mathrm{MeV}$ (kurz vor der Elektron-Positron-Annihilation)?

Was fällt dir auf, wenn du die drei Fälle vergleichst?

Musterlösung 5.2

1. Bei $T = 10^{16}\,\mathrm{GeV}$ sind *alle* Teilchen des Standardmodells leichter als T (selbst das Top-Quark mit $m_t \approx 173\,\mathrm{GeV}$), das Plasma besteht also aus allen SM-Teilchen.

2. Bei $T = 10\,\mathrm{GeV}$ ist das Top-Quark ($173\,\mathrm{GeV}$), das Higgs-Boson ($125\,\mathrm{GeV}$) sowie die $W^{\pm}$- und Z-Bosonen (~ 80–$91\,\mathrm{GeV}$) bereits zu schwer, um effektiv produziert zu werden; sie sind „Boltzmann-unterdrückt“. Das heißt, dass ihre Anzahldichte mit dem Faktor $\mathrm{e}^{-m/T}$ unterdrückt ist. Übrig bleiben die leichteren Quarks, Leptonen, Photonen und Gluonen.

3. Bei $T = 0.1\,\mathrm{MeV}$ sind nur noch Photonen, Elektronen/Positronen (knapp, $m_e \approx 0.511\,\mathrm{MeV}$) und Neutrinos relativistisch, während alle Quarks (in Protonen/Neutronen gebunden) und alle anderen Leptonen/Bosonen sind längst aus dem Plasma verschwunden bzw. in leichtere Teilchen zerfallen.

Die Beobachtung: *Je kälter das Plasma wird, desto weniger Teilchenspezies sind Teil davon.* Das sich abkühlende Universum „Boltzmann-unterdrückt“ sukzessive die schwersten Teilchen zuerst, was in Kapitel 8 über die effektiven Freiheitsgrade g_* noch wichtig werden.

Aufgabe 5.3

Die erste Friedmanngleichung verknüpft die Ausdehnungsgeschwindigkeit des Universums mit dessen mittlerer Energiedichte ρ:

$$H^2 = \frac{8\pi}{3M_{\mathrm{Pl}}^2}\rho\,.$$

Benutze die Hubble-Konstante H_0, also die Hubble-Rate $H(t)$ zum heutigen Zeitpunkt t_0, um $\rho(t_0)$ zu finden.

a) Gib die Energiedichte in $\mathrm{GeV/m^3}$ und $\mathrm{kg/m^3}$ an. Interpretiere die Größenordnung. *Tipp: Vergleiche mit der Masse eines Protons, $m_p \approx 0.938\,GeV$.*

b) Gib die Energiedichte nun auch in eV^4 an. Das Standardmodell der Teilchenphysik impliziert, dass das Vakuum eine Energiedichte in der Größenordnung von $\rho_{\mathrm{SM}} \simeq M_{\mathrm{Pl}}^4$ hat. Wie gut oder eher schlecht sagt das Standardmodell die gemessene Vakuumenergie vorher?

Musterlösung 5.3

Auflösen nach $\rho(t_0)$:

$$\rho(t_0) = \frac{3M_{\mathrm{Pl}}^2}{8\pi} H_0^2 \,. \tag{5.23}$$

Mit $H_0 \approx 1.45 \times 10^{-33}\,\mathrm{eV}$ (aus Aufgabe 2.1, umgerechnet über Kapitel 5) und $M_{\mathrm{Pl}} = 1.22 \times 10^{19}\,\mathrm{GeV}$ erhält man

$$\rho(t_0) \approx 3.5 \times 10^{-11}\,\mathrm{eV}^4 \approx (2\,\mathrm{meV})^4 \,. \tag{5.24}$$

a) Mit der Umrechnungstabelle aus Kapitel 5 (Dichte: $1\,\mathrm{eV}^4 = 2.32 \times 10^{-16}\,\mathrm{kg/m^3}$):

$$\rho(t_0) \approx 8.5 \times 10^{-27}\,\mathrm{kg/m^3} \approx 4.6\,\frac{\mathrm{GeV}}{\mathrm{m}^3} \,. \tag{5.25}$$

Eine Protonenmasse entspricht $\approx 1\,\mathrm{GeV}$; die heutige mittlere Energiedichte des Universums entspricht also nur ungefähr *4-5 Protonen pro Kubikmeter*! Das Universum ist also fast vollkommen leer.
b) In eV^4: $\rho(t_0) \approx 3.5 \times 10^{-11}\,\mathrm{eV}^4$. Das Standardmodell sagt $\rho_{\mathrm{SM}} \sim M_{\mathrm{Pl}}^4 \approx (1.22 \times 10^{28}\,\mathrm{eV})^4 \approx 10^{112}\,\mathrm{eV}^4$ vorher. Das Verhältnis

$$\frac{\rho(t_0)}{\rho_{\mathrm{SM}}} \sim \frac{10^{-11}}{10^{112}} \sim 10^{-123} \tag{5.26}$$

ist eine der spektakulärsten Diskrepanzen zwischen Theorie und Beobachtung in der gesamten Physik: das berühmte *Hierarchie-* bzw. *kosmologische Konstanten-Problem*. Es ist bis heute ungelöst und einer der wichtigsten Hinweise darauf, dass unser Verständnis von Quantenfeldtheorie und Gravitation noch unvollständig ist (vgl. Kapitel 1).

6 Quantengravitation und die Planck-Skala

In Kapitel 1 haben wir gesehen, dass GR und QFT die zwei großen Säulen der modernen Physik sind – und dass man sich bislang nicht einig ist, wie diese Theorien in der Natur vereint sind. In Kapitel 5 haben wir gelernt, wie man Energien, Längen, Zeiten und Temperaturen in denselben Einheiten schreibt. Jetzt können wir ausrechnen, *wo genau* die Grenze liegt, jenseits derer beide Theorien gleichzeitig wichtig werden und damit eine Theorie der Quantengravitation unvermeidlich ist.

6.1 Die Planck-Skala

Um uns dieser Frage zu nähern, starten wir ein Gedankenexperiment: Aus der Allgemeinen Relativitätstheorie wissen wir, dass große Energien die Raumzeit krümmen; soweit bis dann sogar ein schwarzes Loch entstehen kann, wenn die Energiedichte nur groß genug ist. Aus der Quantenmechanik wissen wir, dass Photonen, also Lichtteilchen, eine Energie haben, die proportional zu ihrer Frequenz ist. Um herauszubekommen, wann sich GR und QFT also in die Quere kommen, betrachten wir den Fall, wo die Energie eines Objekts so groß ist, dass seine Eigenschaften durch beide Theorien beschrieben werden müssen.

Die Rechnung ist einfach: Ein (angenommen: kugelförmiges, nicht-geladenes und auch nicht rotierendes) Objekt verhält sich wie ein Schwarzes Loch, wenn sein Energie E so groß ist, dass sein *Schwarzschild-Radius* r_S kleiner als sein „tatsächlicher Radius“ wird. Dieser wird im Rahmen der Quantenmechanik durch die *Compton-Wellenlänge* λ_C beschrieben; der kürzesten Länge, auf der man ein einzelnes Teilchen lokalisieren kann:

- **Schwarzschild-Radius:** Aus der GR: Ein Objekt der Masse $m = E/c^2$ hat den Schwarzschild-Radius

$$r_\mathrm{S} = \frac{2Gm}{c^2} = \frac{2GE}{c^4} \tag{6.1}$$

- **Compton-Wellenlänge:** Aus der QM: Ein Photon der Energie E hat die Wellenlänge[5]

$$\lambda_\mathrm{C} = \frac{\hbar c}{E} \tag{6.2}$$

Wenn $r_\mathrm{S} \sim \lambda_\mathrm{C}$ gilt, ist weder die Allgemeine Relativitätstheorie noch die Quantenmechanik allein ausreichend um das Objekt zu beschreiben: Gravitation und Quanteneffekte sind gleich wichtig.

6.2 Aufgabe: Wann wird ein Photon zum Schwarzen Loch?

Aufgabe 6.1

a) Setze $\lambda_\mathrm{C} = r_\mathrm{S}$ und löse nach der Energie E auf (in SI-Einheiten). Das Ergebnis ist die *Planck-Energie* E_Pl.

b) Berechne E_Pl numerisch. Nutze $\hbar = 1.055 \times 10^{-34}\,\mathrm{Js}$, $c = 3 \times 10^8\,\mathrm{m/s}$, $G = 6.674 \times 10^{-11}\,\mathrm{m^3 kg^{-1} s^{-2}}$. Gib das Ergebnis in GeV an (mit $1\,\mathrm{J} = 6.242 \times 10^9\,\mathrm{GeV}$).

c) Der Large-Hadron-Collider LHC am Europäischen Kernforschungszentrum CERN beschleunigt Protonen auf bis zu $7\,\mathrm{TeV} = 7 \times 10^3\,\mathrm{GeV}$. Wie viele Größenordnungen fehlen

[5] In der Schule wird häufig statt dem $\hbar$ ein $h = 2\pi\hbar$ verwendet und die hier definierte Größe als „reduzierte“ Compton-Wellenlänge bezeichnet. Das ist uns an dieser Stelle egal. In der Teilchenphysik rechnet wirklich niemand mit h statt $\hbar$. Am Ende ist es aber eine Frage der Konvention. Wir hätten ja auch $h = 1$ sezen können.

bis zur Planck-Energie? Was sagt dir das über die Chancen, Quantengravitationseffekte im Teilchenbeschleuniger zu messen?

d) (*Bonus*) Bei welcher Temperatur T_{Pl} würde ein thermisches Photon ($E \sim k_{\mathrm{B}}T$) diese Energie haben? Vergleiche mit der Zeitleiste aus Kapitel 1.5.

Musterlösung 6.1

a) In SI-Einheiten: Compton-Wellenlänge $\lambda_{\mathrm{C}} = \hbar c/E$, Schwarzschild-Radius $r_{\mathrm{S}} = 2GE/c^4$. Gleichsetzen:

$$\frac{\hbar c}{E} = \frac{2GE}{c^4} \quad \Longrightarrow \quad E^2 = \frac{\hbar c^5}{2G} \quad \Longrightarrow \quad E \simeq \sqrt{\frac{\hbar c^5}{2G}} \approx 10^9\,\mathrm{J}\,. \tag{6.3}$$

Wir finden also dass das *einzelne* Photon eine gigantische Energie von 10^9 J haben müsste.
b)

$$E \approx 10^9\,\mathrm{J} \approx 10^{19}\,\mathrm{GeV}\,. \tag{6.4}$$

Bis auf einen Faktor von $\sqrt{2}$ ist das genau die Planck-Energie! (Der Faktor $\sqrt{2}$ ist Schall und Rauch. Hier geht es nur um eine Abschätzung einer Größenordnung. Die Planck-Energie ist normalerweise ohne den Faktor 2 definiert.)
c) Der LHC erreicht 7×10^3 GeV; die Planck-Energie beträgt $\sim 10^{19}$ GeV. Das ergibt:

$$\frac{E_{\mathrm{Pl}}}{E_{\mathrm{LHC}}} \sim \frac{10^{19}}{10^4} = 10^{15}\,. \tag{6.5}$$

Es fehlen **15 Größenordnungen**. Kein denkbarer, irdischer Teilchenbeschleuniger könnte jemals Quantengravitationseffekte direkt messen; die Planck-Skala ist unvorstellbar weit jenseits heutiger Technologie. Es besteht aber die Hoffnung, dass wir vielleicht mit einer anderen Art von Experimenten die Quantengravitation untersuchen können. Tatsächlich könnte uns sogar auch das sehr frühe Universum Aufschlüsse darüber liefern, sofern wir denn Signale aus diesen Zeiten messen können.
d) Die Planck-Temperatur: $T_{\mathrm{Pl}} = E_{\mathrm{Pl}}/k_{\mathrm{B}} \approx 10^{32}\,\mathrm{K}$. Wären wir in der Lage solch hohen Temperaturen zu erreichen, könnten wir direkt die Effekte der Quantengravitation beobachten. Es ist aber fraglich, ob solch hohe Temperaturen jemals im Universum erreicht wurden. Selbst nach dem Ende der Inflation geht man davon aus, dass die Temperatur um einige Größenordnungen unter der Planck-Temperatur lag.

6.3 Warum wir GR und QFT trotzdem vertrauen können

Das Ergebnis der Aufgabe, nämlich 15 Größenordnungen Abstand zwischen LHC und Planck-Skala, klingt frustrierend. Es ist aber zugleich eine gute Nachricht: Es bedeutet, dass GR und QFT in allen Experimenten, die wir je durchgeführt haben, *getrennt* angewendet werden dürfen, ohne dass man sich um die jeweils andere Theorie kümmern muss.

Das ist im Übrigen auch der Grund dafür, dass die meisten Physik-Studierenden, insbesondere die, die sich nicht in der theoretischen Physik spezialisieren wollen, nie etwas mit der Allgemeinen Relativitätstheorie zu tun haben.

Dieses Prinzip nennt man *Skalentrennnung* (engl. *scale separation*). GR beschreibt Gravitation auf kosmischen Skalen, wo Quanteneffekte vernachlässigbar klein sind. QFT beschreibt die Teilchenphysik auf subatomaren Skalen, wo Gravitation so schwach ist, dass man sie komplett ignorieren kann; das Graviton (falls es existiert) koppelt mit einer Stärke $\sim E/E_{\mathrm{Pl}}$ an andere

Teilchen, was für $E \sim$ TeV einem Faktor $\sim 10^{-15}$ entspricht. Zum Vergleich: Die Kopplungskonstante in der Elektrodynamik ist ca. $\alpha \sim 1/137$. Beide Theorien sind in ihrer jeweiligen Domäne präzise getestet und bestätigt.

Aus diesem Grund spricht man von *effektiven Feldtheorien*: Jede Theorie ist nur bis zu einer bestimmten Energieskala gültig, und das ist in Ordnung. Newtons Mechanik ist eine effektive Theorie der Relativitätstheorie für $v \ll c$; die Thermodynamik ist eine effektive Theorie für viele Teilchen. Die GR und das Standardmodell der Teilchenphysik sind effektive Beschreibungen für Energien weit unterhalb der Planck-Skala.

In gewisser Hinsicht ist das, was wir gerade gemacht haben, zutiefst verblüffend: Jede andere Art von Theorie außerhalb der Physik, nimm die Biologie, die Psychologie oder die Politologie, hat eine Grenze, in der sie nicht mehr anwendbar ist oder jenseits von der sie nicht mehr frei von Widersprüchen ist. Diese Grenzen auszuloten, ist ein ständiges Thema in anderen Wissenschaften. Die Physik hat hier eine Art Sonderstellung: Die Theorie bringt praktischerweise schon gleich ihren Gültigkeitsbereich mit und wir können sie im Rahmen dessen einfach anwenden. Außerhalb davon, haben wir keine Ahnung, wie sich die Natur verhält.

Eng damit verbunden ist die Frage „Was war vor dem Urknall“: Abhängig davon, was man nun als *Urknall* definiert, kommt man zu ganz unterschiedlichen Antworten. Wenn man an dieser Stelle aber die Anfangssingularität als *Urknall* bezeichnet, also noch vor dem *hot Big Bang*-Szenario, vor dem Reheating und vor der Inflation, dann ist die Antwort darauf: *Wir wissen es nicht.* Aus den oben genannten Gründen, ergibt die Frage aber auch gewissermaßen keinen Sinn: Wir wissen ja nicht mal, ob Zeiten kleiner als die Planck-Zeit überhaupt existieren. Aussagen jenseits der Planck-Skala benötigen zu ihrer Beantwortung eine vereinheitlichte Theorie der Quantengravitation. Von denen gibt es aktuelle eine handvoll (siehe unten). Und diese kommen zu fundamental unterschiedlichen Antworten auf die gestellte Frage. Es bleibt fraglich, ob wir die Frage jemals mit empirischen Methoden beantworten können werden. Eng damit verbunden ist im Übrigen auch die Frage: *Wie sieht das Zentrum von einem Schwarzen Loch aus?* Auch diese Frage lässt sich ohne eine Theorie der Quantengravitation nicht beantworten.

Theorien der Quantengravitation Was passiert *jenseits* der Planck-Skala, also auf Zeitskalen unterhalb von $\sim 10^{-44}\,\mathrm{s}$? Hier versagen beide Theorien, und eine vollständige Theorie der *Quantengravitation* wäre nötig. Bisher gibt es keine experimentell bestätigte solche Theorie. Wichtige Kandidaten sind:

- **Stringtheorie:** Ersetzt punktförmige Teilchen durch ausgedehnte Objekte (Strings). Enthält Quantengravitation automatisch und ist mathematisch konsistent, macht aber (bisher) kaum testbare Vorhersagen auf erreichbaren Energieskalen.
- **Schleifenquantengravitation (LQG):** Quantisiert den Raum direkt, ohne ihn in einen flachen Hintergrund einzubetten. Raum und Zeit werden diskret; die kleinstmögliche Länge ist die Planck-Länge $\ell_{\mathrm{Pl}} \approx 1.6 \times 10^{-35}\,\mathrm{m}$.
- **Weitere Ansätze:** Kausal dynamische Triangulierungen, asymptotische Sicherheit, Nicht-kommutative Geometrie: Das Feld ist offen und aktiv.

Keiner dieser Ansätze ist heute zweifelsfrei bestätigt. Die Planck-Skala ist möglicherweise für immer experimentell unzugänglich, was die Quantengravitation für viele Menschen zu einem der faszinierendsten offenen Probleme der theoretischen Physik macht. Für Phänomenologen wie Carlo ist sich auch faszinierend, aber auch frustrierend, da die Rechnungen für Vorhersagen über die reale Welt und ihre Phänomene einfach zu schwierig sind, als dass er sie im Detail verstehen oder nachzuvollziehen könnte.

6.4 Inflation und empirische Evidenz

In Abschnitt 2 haben wir die Inflation als kurze Phase exponentieller Expansion eingeführt, die das Horizont- und das Flachheitsproblem löst. Hier betrachten wir, *warum* Inflation so nahe an der Planck-Skala liegt und welche empirischen Spuren sie hinterlassen hat.

6.4.1 Die Energieskala der Inflation

Damit die Inflation das Flachheitsproblem löst, muss der Skalenfaktor während der Inflation um mindestens einen Faktor e^{60} anwachsen (*60 e-Folds*). Das inflationstreibende Feld das *Inflaton* ϕ, rollt dabei langsam ein flaches Potential $V(\phi)$ herunter (*slow-roll*). Die Energiedichte während der Inflation ist näherungsweise

$$\rho_{\text{inf}} \approx V(\phi) \approx \text{const}, \tag{6.6}$$

was einer kosmologischen Konstante entspricht und zu exponentiellem Wachstum führt. Die Energieskala ist durch die Friedmann-Gleichung bestimmt:

$$H_{\text{inf}}^2 = \frac{8\pi G_N}{3} V(\phi) \quad \Rightarrow \quad E_{\text{inf}} = V(\phi)^{1/4} \lesssim 10^{16}\,\text{GeV} \approx 10^{-3}\,M_{\text{Pl}}. \tag{6.7}$$

Die Inflation findet also bei Energien statt, die zwar deutlich unter der Planck-Masse liegen, aber weit jenseits jedes heutigen Beschleunigers (LHC: $\sim 10^4\,\text{GeV}$). Dennoch ist Inflation beobachtbar, nämlich durch die Quantenfluktuationen des Inflatonfelds, die sich zu kosmischen Dichtefluktuationen aufgeblasen haben.

6.4.2 Bonus: Vorhersagen und CMB-Beobachtungen

Die folgenden Vorhersagen sind wenig pädagogisch aufbereitet, aber dennoch interessant: Die wichtigste empirische Spur der Inflation ist das *Leistungsspektrum* der CMB-Temperaturanisotropien. Quantenfluktuationen des Inflatons während der Inflation erzeugen ein nahezu skaleninvariantes Spektrum von Dichteschwankungen, die späteren Keime für Galaxien und *Large-Scale-Structure*. Konkret sagt das Standard-Inflations-Szenario vorher:

- **Skalarer Spektralindex n_s:** Das Leistungsspektrum folgt einem Potenzgesetz $P(k) \propto k^{n_s-1}$. Perfekte Skaleninvarianz wäre $n_s = 1$ (Harrison-Zel'dovich-Spektrum); die Slow-Roll-Näherung ergibt $n_s < 1$ (roter Tilt). Planck 2018 misst:

  $$n_s = 0.9649 \pm 0.0042 \quad (68\%\,\text{CL})\,. \tag{6.8}$$

- **Tensor-zu-Skalar-Verhältnis r:** Inflation erzeugt neben skalaren auch *tensorielle* Fluktuationen, nämlich primordiale Gravitationswellen. Diese würden im CMB eine B-Moden-Polarisation hinterlassen. Die aktuelle Obergrenze (BICEP/Keck 2021):

  $$r < 0.036 \quad (95\%\,\text{CL})\,. \tag{6.9}$$

  Einfache Modelle wie die Starobinsky-Inflation (R^2-Modell) sagen $r \approx 0.004$ vorher und liegen damit klar unter der Schranke. Bisher wurden keine Spuren von Tensorfluktuationen im CMB entdeckt. Sollten diese gefunden werden, wären das großartige Neuigkeiten, die sicherlich mit einem Nobelpreis belohnt würden.

- **Gaussianität:** Einfache Einfeldmodelle sagen nahezu gausssche Fluktuationen vorher. Planck findet keine signifikante Nicht-Gaussianität ($f_{\text{NL}} \approx 0$), konsistent mit dieser Vorhersage.

Die Gesamtheit der CMB-Daten (Planck, BICEP/Keck, ACT, SPT) ist konsistent mit dem Bild einer Inflation mit einem einzelnen Inflatonfeld, und legt damit nahe, dass das frühe Universum durch eine extrem flache Potentiallandschaft bei $\sim 10^{16}\,\mathrm{GeV}$ geprägt wurde. Was genau das Inflaton ist, ob es ein Teilchen des Standardmodells, eine neue Skalarfeld-Erweiterung oder ein Artefakt einer fundamentaleren Theorie wie der Stringtheorie ist –, bleibt eine der zentralen offenen Fragen der Kosmologie.

Die Messung von $r > 0$ wäre eine der größten Entdeckungen der Physik: Sie würde beweisen, dass Gravitationswellen während der Inflation erzeugt wurden, also ein direktes Fenster in Quanteneffekte bei Planck-nahen Energien.

7 Taylor-Reihen

7.1 Von der Tangente zum Polynom

In Aufgabe 3.2 haben wir gesehen, dass die Tangente

$$p_1(x) = f(a) + f'(a) \cdot (x - a) \tag{7.1}$$

die beste *lineare* Approximation einer Funktion f in der Nähe von a ist: p_1 stimmt mit f im Punkt a im Funktionswert *und* in der ersten Ableitung überein. Die naheliegende Idee: Können wir eine noch bessere Approximation finden, die auch in der zweiten, dritten, ... Ableitung übereinstimmt? Ziel ist es, eine (ggf. sehr komplizierte) Funktion f durch ein *Polynom* zu approximieren, denn Polynome sind einfach auszuwerten, abzuleiten und zu integrieren.

Für $a = 0$ definieren wir das *Taylor-Polynom n-ten Grades* von f:

$$p_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!} \frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \cdots + \frac{x^n}{n!} \frac{\mathrm{d}^n f}{\mathrm{d}x^n}(0)\,. \tag{7.2}$$

Aufgabe 7.1

Für alle $n \in \mathbb{N}$ und eine differenzierbare Funktion f definieren wir das Polynom n-ten Grades $p_n : \mathbb{R} \to \mathbb{R}$,

$$p_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!} \frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0)\,.$$

Zeige, dass für alle $j = 0, \ldots, n$

$$\frac{\mathrm{d}^j f}{\mathrm{d}x^j}(0) = \frac{\mathrm{d}^j p_n}{\mathrm{d}x^j}(0)$$

gilt, d.h. dass f und p_n im Punkt $x_0 = 0$ die gleichen Ableitungen bis zur n-ten Ordnung haben.

Musterlösung 7.1

Wir betrachten ein einzelnes Glied der Summe, $g_k(x) = \frac{x^k}{k!} \frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0)$ (beachte: $\frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0)$ ist nur eine *Zahl*, keine Funktion von x!). Ableiten nach x gibt (Potenzregel)

$$g_k'(x) = \frac{k\, x^{k-1}}{k!} \frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0) = \frac{x^{k-1}}{(k-1)!} \frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0)\,, \qquad k \geq 1\,, \tag{7.3}$$

da $\frac{k}{k!} = \frac{1}{(k-1)!}$. Jede Ableitung „verschiebt“ also den Exponenten und die Fakultät um eins nach unten, ohne den eigentlichen Koeffizienten $\frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0)$ zu verändern. Wendet man das j-mal an, erhält man

$$\frac{\mathrm{d}^j p_n}{\mathrm{d}x^j}(x) = \sum_{k=j}^{n} \frac{x^{k-j}}{(k-j)!} \frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0) \tag{7.4}$$

(die Terme mit $k < j$ verschwinden, da sie öfter abgeleitet werden als ihr Grad hoch ist). Jetzt setzen wir $x = 0$ ein. Wegen $0^0 = 1$ und $0^m = 0$ für $m \geq 1$ überlebt nur der Term mit

$k = j$:

$$\frac{\mathrm{d}^j p_n}{\mathrm{d}x^j}(0) = \underbrace{\frac{0^0}{0!}\frac{\mathrm{d}^j f}{\mathrm{d}x^j}(0)}_{k=j} + \underbrace{\frac{0^1}{1!}\frac{\mathrm{d}^{j+1} f}{\mathrm{d}x^{j+1}}(0)}_{k=j+1} + \cdots = \frac{\mathrm{d}^j f}{\mathrm{d}x^j}(0)\,. \qquad \blacksquare \tag{7.5}$$

Damit stimmen f und p_n am Punkt 0 in allen Ableitungen bis zur Ordnung n überein; p_n ist also tatsächlich die „beste“ Approximation von f durch ein Polynom n-ten Grades in der Nähe von $x = 0$.

7.2 Ein Beispiel: $f(x) = x[\sin(x) + \cos(x)]$

Wir testen das Konzept an einem konkreten Beispiel, das uns die nächsten beiden Aufgaben begleiten wird.

Aufgabe 7.2

Wir betrachten die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = x \cdot [\sin(x) + \cos(x)]$. Bestimme die Koeffizienten des quadratischen Polynoms, d.h. die Zahlen $a, b, c \in \mathbb{R}$ in $p_2(x) = ax^2 + bx + c$, sodass $f(0) = p_2(0)$ und $f'(0) = p_2'(0)$ und $f''(0) = p_2''(0)$ gilt.

Musterlösung 7.2

Nach Definition ist

$$p_2(x) = \sum_{k=0}^{2} \frac{x^k}{k!}\frac{\mathrm{d}^k f}{\mathrm{d}x^k}(0) = f(0) + xf'(0) + \frac{x^2}{2}f''(0)\,. \tag{7.6}$$

Wir berechnen die Ableitungen von $f(x) = x[\sin x + \cos x]$ bei $x = 0$:

$$f(0) = 0 \cdot [\sin 0 + \cos 0] = 0\,, \tag{7.7}$$
$$f'(x) = [\sin x + \cos x] + x[\cos x - \sin x] \quad \Rightarrow \quad f'(0) = \sin 0 + \cos 0 = 1\,, \tag{7.8}$$
$$f''(x) = 2[\cos x - \sin x] - x[\sin x + \cos x] \quad \Rightarrow \quad f''(0) = 2[\cos 0 - \sin 0] = 2\,. \tag{7.9}$$

Also

$$p_2(x) = 0 + x \cdot 1 + \frac{x^2}{2} \cdot 2 = x^2 + x\,. \tag{7.10}$$

Aufgabe 7.3

Wie in der letzten Aufgabe sei $f : \mathbb{R} \to \mathbb{R}$ gegeben durch $f(x) = x \cdot [\sin(x) + \cos(x)]$. Bestimme die Koeffizienten des kubischen Polynoms, d.h. die Zahlen $a, b, c, d \in \mathbb{R}$ in $p_3(x) = ax^3 + bx^2 + cx + d$, sodass $f(0) = p_3(0)$ und $f'(0) = p_3'(0)$ und $f''(0) = p_3''(0)$ und $f'''(0) = p_3'''(0)$ gilt.

Musterlösung 7.3

Es gilt $p_3(x) = p_2(x) + \frac{x^3}{3!}\frac{\mathrm{d}^3 f}{\mathrm{d}x^3}(0)$. Die dritte Ableitung ist

$$f'''(x) = -3[\cos x - \sin x] - x[\cos x - \sin x] + x[\sin x + \cos x] \cdot (-1) \tag{7.11}$$

$$\Rightarrow \quad f'''(0) = -3[\cos 0 - \sin 0] = -3\,. \tag{7.12}$$

Also

$$p_3(x) = x^2 + x + \frac{x^3}{6} \cdot (-3) = -\frac{x^3}{2} + x^2 + x\,. \tag{7.13}$$

Konvergenzfrage: Konvergiert $p_n(x) \to f(x)$ für $n \to \infty$, für jedes $x \in \mathbb{R}$? Das hängt von der Funktion f ab – für $f(x) = x[\sin x + \cos x]$ (eine Kombination aus Sinus, Kosinus und einem Polynom) ist die Antwort *ja*: Die Taylorreihe konvergiert für *alle* $x \in \mathbb{R}$ gegen f. Das ist aber nicht für jede Funktion so Allerdings gibt es Funktionen, deren Taylorreihe nur in einem begrenzten Bereich um 0 konvergiert (*Konvergenzradius*), oder die noch nicht einmal gegen f selbst konvergiert.
Kleinwinkelnäherung: Ein Spezialfall, der in der Physik dauernd vorkommt: Für $\sin(x)$, $\tan(x)$ und $x \mapsto x$ selbst stimmen die linearen Taylorpolynome (p_1) bei $x = 0$ überein – „$\sin(x) \approx x \approx \tan(x)$“ für kleine x. Das steckt z.B. hinter der Näherung, mit der man die Schwingungsdauer eines Pendels herleitet.

7.3 Taylor-Reihen visualisieren

Aufgabe 7.4

a) Installiere `matplotlib` in deiner Python-Umgebung (z.B. mit `conda install matplotlib`) und importiere es mit `import matplotlib.pyplot as plt`.

b) Plotte die Funktion $f(x) = x[\sin(x)+\cos(x)]$ mit `matplotlib` für $x \in [-6, 6]$. Verwende `plt.plot(x,y)`, wobei `x = np.linspace(-6,6,N)`.

c) Füge dem Plot nun die Graphen der zuvor berechneten Polynome $p_1(x)$, $p_2(x)$ und $p_3(x)$ hinzu.

d) Schreibe eine Funktion, die die Graphen von $f(x)$ und $p_n(x)$ für beliebiges $n \in \mathbb{N}$ plottet.

e) Füge dem Plot ein weiteres Panel hinzu, in dem du die Differenz von $f(x)$ und $p_n(x)$ plottest. Welche Beobachtung machst du für große n?

Musterlösung 7.4

Wir definieren f, die Polynome und den Plot in Python:

```
import numpy as np
import matplotlib.pyplot as plt
import math

def f(x):
    return x * (np.cos(x) + np.sin(x))
```

```
x = np.linspace(-6, 6, 1000)

def p1(x): return x
def p2(x): return x + x**2
def p3(x): return x + x**2 - x**3/2

# allgemeine n-te Ableitung von f bei x=0 (Muster: +-1,+-2,...)
def dnf_dxn(n):
    return (-1)**np.floor((n-1)/2) * n

def taylor(x, n):
    s = 0
    for k in range(n):
        s += dnf_dxn(k) * x**k / math.factorial(k)
    return s

def error(x, n):
    return f(x) - taylor(x, n)

fig, axes = plt.subplots(2)
axes[0].plot(x, f(x), color="black", lw=2, label="f(x)")
axes[1].plot(x, np.zeros(len(x)), color="black", lw=2)
for i in [1, 3, 5, 10, 15]:
    axes[0].plot(x, taylor(x, i), ls="--", label=f"$p_{{{i}}}(x)$")
    axes[1].plot(x, error(x, i), ls="--", label=f"$e_{{{i}}}(x)$")
for ax in axes:
    ax.set_ylim(-10, 10)
    ax.set_xlim(-6, 6)
    ax.grid()
    ax.legend()
axes[1].set_xlabel("x")
fig.savefig("taylor_musterloesung.pdf")
```

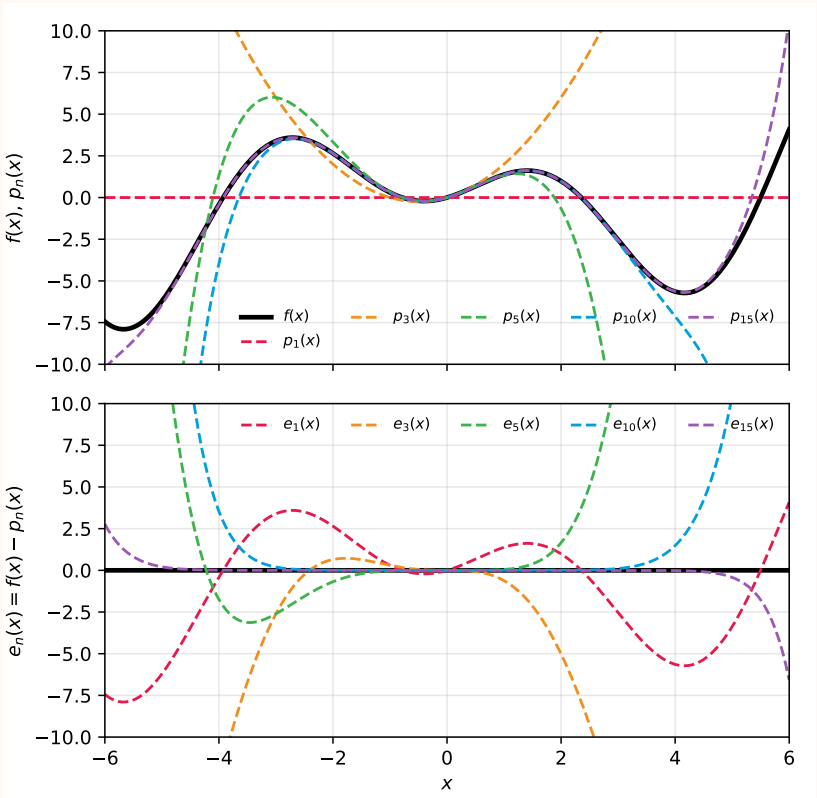


Abbildung 7.1: Oben: die Funktion $f(x) = x[\sin(x) + \cos(x)]$ (schwarz) und ihre Taylor-Polynome $p_n(x)$ für $n = 1, 3, 5, 10, 15$ (gestrichelt). Unten: der jeweilige Approximationsfehler $e_n(x) = f(x) - p_n(x)$. Man erkennt deutlich, wie die Taylor-Polynome mit wachsendem Grad n auf einem immer größeren Bereich um $x = 0$ mit f übereinstimmen.

Beobachtung für große n (Abbildung 7.1): Im Bereich $|x| \lesssim 5$ wird die Differenz $e_n(x) = f(x) - p_n(x)$ mit wachsendem n immer kleiner; die Taylorpolynome nähern sich f immer weiter an. Am Rand des betrachteten Intervalls braucht man höhere n, damit die Approximation gut wird: Die Taylorreihe konvergiert „von innen nach außen“.

8 Friedmann-Gleichungen und ihre Lösung

Jetzt haben wir alle Werkzeuge zusammen, um die Friedmann-Gleichung (2.3) aus Kapitel 2 tatsächlich zu *lösen*, für die drei wichtigsten Materiearten des Universums: „Staub“, Strahlung und Vakuumenergie.

8.1 Wie verdünnen sich Energiedichten?

Stellen wir uns eine Box im primordialen Universum vor, einen mitexpandierenden Würfel mit Kantenlänge $\propto a(t)$ und Volumen $V(t) \propto a^3(t)$.

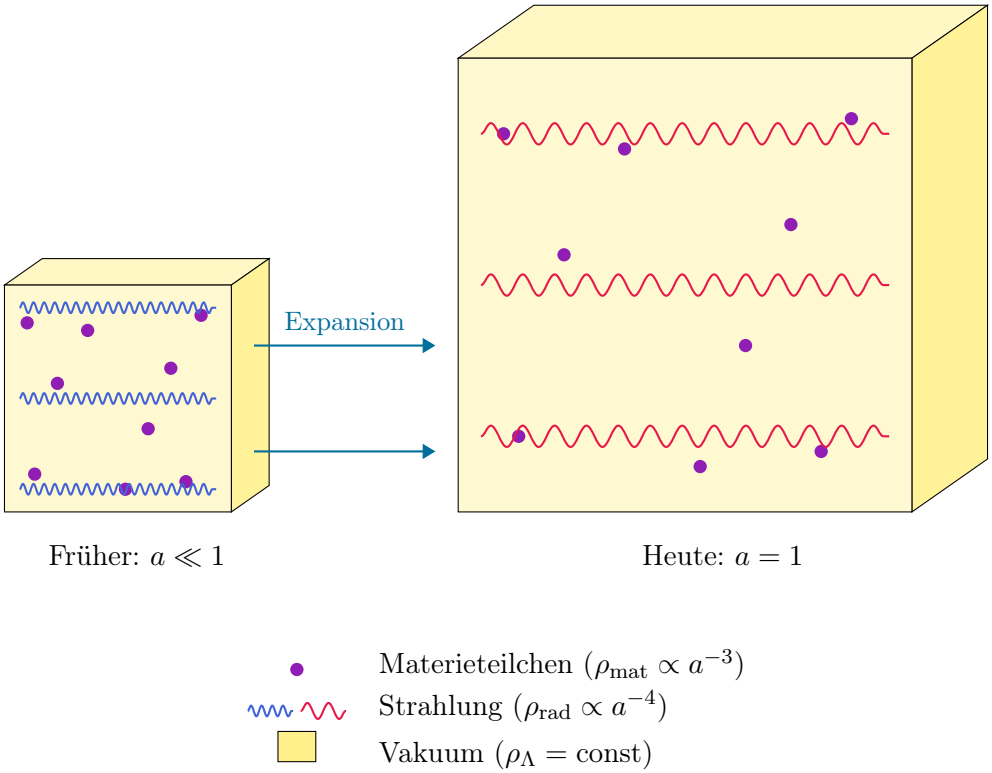


Abbildung 8.1: Veranschaulichung der unterschiedlichen Verdünnung der drei Energiedichten beim Übergang von einem frühen, kleinen mitexpandierenden Volumen ($a \ll 1$, links) zu einem späteren, größeren Volumen ($a = 1$, rechts), jeweils als Würfel dargestellt. Die *Anzahl* der Materieteilchen (violette Punkte) bleibt gleich, ihre Dichte sinkt also mit dem Volumen $\propto a^{-3}$. Die Strahlung (Wellenlinien) wird zusätzlich zur Verdünnung auch noch rotverschoben (größere Wellenlänge, kleinere Amplitude, blau→rot), insgesamt $\rho_{\text{rad}} \propto a^{-4}$. Die Vakuumenergie (gelbe Füllung) füllt den Raum dagegen mit *konstanter* Dichte, unabhängig von der Größe des Volumens.

„Staub“: Die Anzahl N der Materieteilchen im Würfel ändert sich nicht, nur das Volumen wächst. Die Energiedichte $\rho_{\text{mat}} = E/V$ mit $E = N \cdot m = \text{const}$ verdünnt sich daher rein durch das wachsende Volumen:

$$\rho_{\text{mat}}(t) = \rho^0_{\text{mat}} \left(\frac{a_0}{a(t)} \right)^3 = \rho^0_{\text{mat}}\, a^{-3}(t)\,, \tag{8.1}$$

wobei wir $a_0 = a(t_0) = 1$ verwenden (Konvention).

Strahlung: Auch hier wächst das Volumen wie a^3, aber zusätzlich verliert jedes einzelne

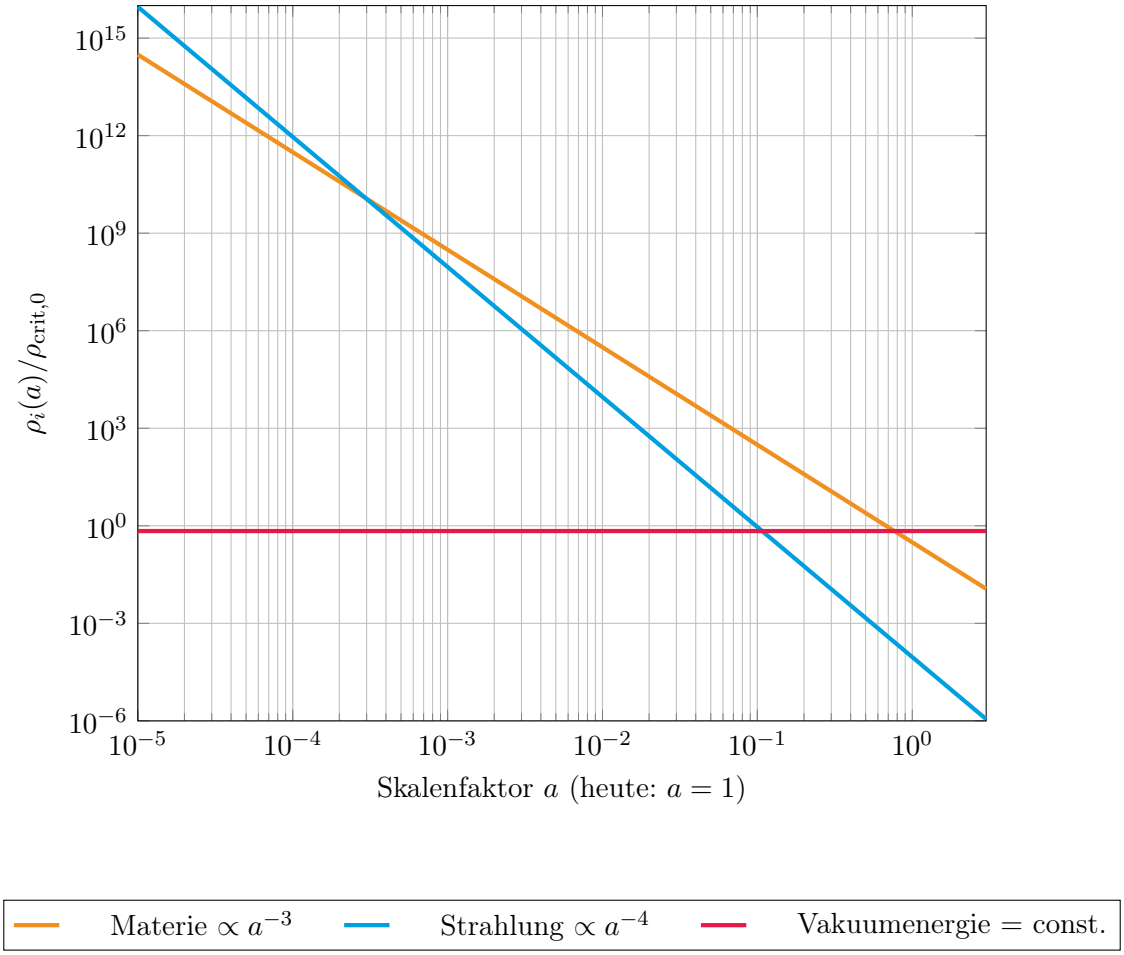


Abbildung 8.2: Verdünnung der drei Energiedichten mit dem Skalenfaktor a (doppeltlogarithmisch). Da $a = 1/(1+z)$, entspricht die x-Achse zugleich (umgekehrt) der Rotverschiebung z: Ganz links ($a \ll 1$, also $z \gg 1$) dominiert Strahlung, weil $\rho_{\mathrm{rad}} \propto a^{-4}$ am schnellsten ansteigt; dazwischen dominiert Materie ($\rho_{\mathrm{mat}} \propto a^{-3}$); ganz rechts (heute und in Zukunft, $a \gtrsim 1$) dominiert die konstante Vakuumenergiedichte ρ_Λ. Die Kreuzungspunkte der Geraden markieren genau die Übergänge „strahlungsdominiert → materiedominiert“ (bei a_{eq}, vgl. Aufgabe 8.2) und „materiedominiert → vakuumdominiert“.

Photon Energie durch die Rotverschiebung, $E_\gamma = \hbar f \propto \lambda^{-1} \propto a^{-1}$ (vgl. Kapitel 2). Insgesamt:

$$\rho_{\mathrm{rad}}(t) = \rho^0_{\mathrm{rad}}\, a^{-4}(t)\,. \tag{8.2}$$

Vakuumenergie: Per Definition hängt die Vakuumenergiedichte nicht von der „Größe der Box“ ab:

$$\rho_\Lambda(t) = \rho^0_\Lambda = \mathrm{const}\,. \tag{8.3}$$

Abbildung 8.1 veranschaulicht diese drei Verdünnungsgesetze anschaulich am Beispiel eines mitexpandierenden Würfels. Abbildung 8.2 zeigt die drei eben hergeleiteten Potenzgesetze $\rho_{\mathrm{mat}}(a) \propto a^{-3}$, $\rho_{\mathrm{rad}}(a) \propto a^{-4}$ und $\rho_\Lambda(a) = \mathrm{const}$ gemeinsam über viele Größenordnungen in a (bzw. äquivalent in der Rotverschiebung z). Welche Komponente zu welcher Zeit *dominiert*, also die größte Energiedichte hat und damit die Friedmann-Gleichung (2.3) und die Expansion $a(t)$ bestimmt, hängt also einfach davon ab, welche Gerade an der jeweiligen Stelle a am höchsten liegt. Das ist die Grundlage für die drei folgenden Abschnitte.

Exkurs: die effektiven Freiheitsgrade g_*. Genauer betrachtet steckt in ρ_{rad} nicht nur Strahlung im engeren Sinne (Photonen), sondern *alle* relativistischen Teilchenspezies, die gerade Teil des heißen Plasmas sind – also $\rho_{\text{rad}}(T) = \frac{\pi^2}{30} g_*(T)\, T^4$ (in natürlichen Einheiten, Kapitel 5). Die Zahl $g_*(T)$ zählt dabei (gewichtet nach Spin und Bosonen/Fermionen-Statistik) genau die Teilchensorten, die bei Temperatur T noch $m \lesssim T$ erfüllen und damit relativistisch im Plasma mitschwimmen: exakt die Größe, die in Aufgabe 5.2 schon eine Rolle gespielt hat! Da g_* mit sinkender Temperatur immer wieder sprunghaft kleiner wird (jedes Mal, wenn eine weitere Teilchenspezies Boltzmann-unterdrückt wird), ist $\rho_{\text{rad}} \propto a^{-4}$ aus Gleichung (2.3) streng genommen nur stückweise gültig, nämlich für unsere Zwecke (Größenordnungsabschätzungen) reicht das aber völlig aus. Wir werden g_* in Kapitel 9 bei der Berechnung des thermischen „Freeze-out" der Dunklen Materie wiedersehen, welche nicht einfach Boltzmann-unterdrückt wird und heute immer noch im Universum vorhanden ist.

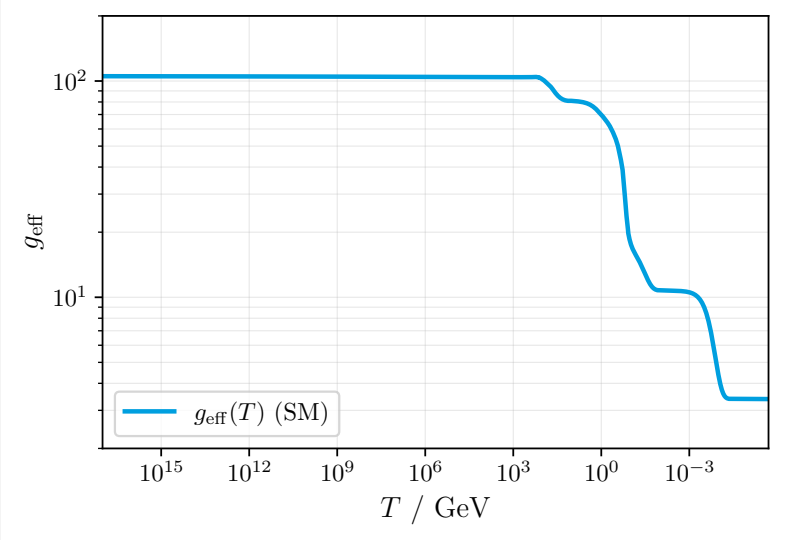


Abbildung 8.3: Effektive Freiheitsgrade $g_{\text{eff}}(T)$ des Standardmodells als Funktion der Temperatur T. Die stufenweisen Sprünge zeigen das sukzessive Ausfrieren der Teilchenspezies: Bei hohen Temperaturen ($T \gtrsim 200\,\text{GeV}$) tragen alle SM-Teilchen bei ($g_{\text{eff}} \approx 106.75$), beim QCD-Phasenübergang ($T \sim 150\,\text{MeV}$) sinkt g_{eff} stark, und nach der Elektron-Positron-Annihilation ($T \sim 0.5\,\text{MeV}$) verbleiben nur Photonen und Neutrinos ($g_{\text{eff}} \approx 3.4$). Daten aus Referenz [5].

Abbildung 8.3 zeigt den Verlauf von $g_{\text{eff}}(T)$ für das Standardmodell: Beim QCD-Phasenübergang ($T \sim 150\,\text{MeV}$) fällt g_{eff} von ~ 60 auf ~ 10, und nach der Elektron-Positron-Annihilation verbleiben nur noch Photonen und Neutrinos mit $g_{\text{eff}} \approx 3.4$.

8.2 Materie-dominiertes Universum: $a(t) \propto t^{2/3}$

Wir setzen $\rho_{\text{mat}}(t) = \rho^0_{\text{mat}} a^{-3}(t)$ in die erste Friedmann-Gleichung (2.3) ein und nehmen an, dass ρ^0_{mat} die *gesamte* heutige Energiedichte ist, sodass $\frac{8\pi}{3m_{\text{Pl}}^2}\rho^0_{\text{mat}} = H_0^2$:

$$\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2\, a^{-3}(t) \quad \Longrightarrow \quad \left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2\, a^{-1}(t) \quad \Longrightarrow \quad \frac{\mathrm{d}a}{\mathrm{d}t} = H_0\, a^{-1/2}\,. \tag{8.4}$$

Das ist genau vom Typ der ODE aus Aufgabe 4.3! Separation der Variablen liefert

$$a^{1/2}\,\mathrm{d}a = H_0\,\mathrm{d}t \quad\Longrightarrow\quad \int a^{1/2}\,\mathrm{d}a = H_0\int \mathrm{d}t \quad\Longrightarrow\quad \frac{2}{3}a^{3/2}(t) = H_0 t + C\,. \tag{8.5}$$

Randbedingungen: Zur Zeit $t = 0$ (Urknall) ist die Energiedichte unendlich, also $a(t = 0) = 0$; damit folgt sofort $C = 0$. Bei $t = t_0$ (heute) ist $a(t_0) = 1$, also

$$\frac{2}{3}\cdot 1 = H_0 t_0 \quad\Longrightarrow\quad \boxed{t_0 = \frac{2}{3}H_0^{-1}}\,. \tag{8.6}$$

Mit $H_0^{-1} \approx 14.6\,\mathrm{Gyr}$ (Aufgabe 2.1) ergibt das $t_0 \approx 9.7\,\mathrm{Gyr}$, was für ein *rein materiedominiertes* Universum. Allgemein gilt für $t < t_0$ (mit $C = 0$):

$$a(t) = \left(\frac{3}{2}H_0 t\right)^{2/3} \quad\Longrightarrow\quad a(t) \propto t^{2/3}\,. \tag{8.7}$$

8.3 Strahlungsdominiertes Universum: $a(t) \propto t^{1/2}$

Aufgabe 8.1

Wir haben gemeinsam hergeleitet, wie man das Alter t_0 des Universums abschätzen kann, wenn man annimmt, dass es mit Staub gefüllt ist (Abschnitt 8.2). Dafür haben wir die Friedmann-Gleichung mit dem allseits beliebten „Integraltrick“ (Separation der Variablen) gelöst. Für Staub gilt $\rho_{\mathrm{mat}}(t) \propto a^{-3}(t)$. Als nächstes wollen wir nun das Alter des Universums unter der Annahme berechnen, dass es stattdessen von Strahlung dominiert wird. Für Strahlung gilt $\rho_{\mathrm{rad}}(t) \propto a^{-4}(t)$. Das ist intuitiv, da die einzelnen Teilchen nicht nur durch das Anwachsen des Volumens $V \propto a^3$, in dem sie sich befinden, verdünnt werden, sondern sich auch die Energie der einzelnen Strahlungsteilchen (z.B. Photonen) $E = h\cdot f \propto a^{-1}$ durch die kosmische Rotverschiebung $f \propto a^{-1}$ verringert. Wiederhole die im Kurs besprochene Berechnung von t_0, wobei du nun anstelle von $\rho = \rho_{\mathrm{mat}}$ ein strahlungsdominiertes Universum $\rho = \rho_{\mathrm{rad}}$ annimmst.

Musterlösung 8.1

Wir setzen $\rho_{\mathrm{rad}}(t) = \rho^0_{\mathrm{rad}} a^{-4}(t)$ mit $\frac{8\pi}{3m_{\mathrm{Pl}}^2}\rho^0_{\mathrm{rad}} = H_0^2$ in Gleichung (2.3) ein:

$$\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2\,a^{-4}(t) \quad\Longrightarrow\quad \left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2\,a^{-2}(t) \quad\Longrightarrow\quad \frac{\mathrm{d}a}{\mathrm{d}t} = H_0\,a^{-1}(t)\,. \tag{8.8}$$

Separation der Variablen:

$$a\,\mathrm{d}a = H_0\,\mathrm{d}t \quad\Longrightarrow\quad \int a\,\mathrm{d}a = H_0\int \mathrm{d}t \quad\Longrightarrow\quad \frac{1}{2}a^2(t) = H_0 t + C\,. \tag{8.9}$$

Wieder gilt $a(t = 0) = 0 \Rightarrow C = 0$. Bei $t = t_0$ mit $a(t_0) = 1$:

$$\frac{1}{2} = H_0 t_0 \quad\Longrightarrow\quad \boxed{t_0 = \frac{1}{2}H_0^{-1}} \approx \frac{1}{2}\cdot 14.6\,\mathrm{Gyr} \approx 7.3\,\mathrm{Gyr}\,. \tag{8.10}$$

Allgemein folgt $a(t) = (2H_0 t)^{1/2} \propto t^{1/2}$; der Skalenfaktor wächst in einem strahlungsdominierten Universum *langsamer* als in einem materiedominierten ($t^{1/2}$ vs. $t^{2/3}$), und das geschätzte Alter t_0 ist entsprechend kleiner.

8.4 Vakuum-dominiertes Universum: exponentielle Expansion

Für $\rho_\Lambda(t) = \rho_\Lambda^0 = \text{const}$ wird die Friedmann-Gleichung (2.3) besonders einfach:

$$\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2 \quad \Longrightarrow \quad \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} = H_0 \quad \Longrightarrow \quad \frac{\mathrm{d}a}{\mathrm{d}t} = H_0\, a(t)\,. \tag{8.11}$$

Das ist genau die ODE aus Aufgabe 3.1! Die Lösung ist eine Exponentialfunktion:

$$a(t) = a(0) \cdot \mathrm{e}^{H_0 t}\,, \tag{8.12}$$

also *exponentielles* Wachstum mit konstanter Rate H_0, ganz anders als die „langsamen" Potenzgesetze $t^{2/3}$ bzw. $t^{1/2}$ für Materie bzw. Strahlung. Genau dieses exponentielle Verhalten ist die Grundlage der Inflation (Kapitel 2) und beschreibt auch die (durch Dunkle Energie angetriebene) zunehmend beschleunigte Expansion unseres heutigen Universums.

8.5 Die allgemeine Mischung

In Wirklichkeit enthält unser Universum alle drei Komponenten gleichzeitig:

$$\rho(a) = \rho_{\mathrm{rad}}^0\, a^{-4} + \rho_{\mathrm{mat}}^0\, a^{-3} + \rho_\Lambda^0\,. \tag{8.13}$$

Wir definieren die heutigen *relativen* Energiedichten (Dichteparameter)

$$\Omega_i = \frac{\rho_i^0}{\rho_0}\,, \qquad \rho_0 = \rho_{\mathrm{rad}}^0 + \rho_{\mathrm{mat}}^0 + \rho_\Lambda^0 = \frac{3m_{\mathrm{Pl}}^2}{8\pi} H_0^2\,, \qquad \Omega_{\mathrm{rad}} + \Omega_{\mathrm{mat}} + \Omega_\Lambda = 1\,. \tag{8.14}$$

Damit wird die erste Friedmann-Gleichung (2.3) zu

$$\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = H_0^2\left[\Omega_{\mathrm{rad}}\, a^{-4} + \Omega_{\mathrm{mat}}\, a^{-3} + \Omega_\Lambda\right], \tag{8.15}$$

bzw. nach Auflösen nach $\frac{\mathrm{d}a}{\mathrm{d}t}$:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = H_0\left[\Omega_{\mathrm{rad}}\, a^{-2} + \Omega_{\mathrm{mat}}\, a^{-1} + \Omega_\Lambda\, a^2\right]^{1/2}. \tag{8.16}$$

Diese Gleichung ist vom selben Typ wie der zweite, „unlösbare" Teil von Aufgabe 4.3 – Separation der Variablen führt formal zu einem Integral, das wir analytisch nicht mehr lösen können. Genau dieses Integral werden wir in Kapitel 8.6 numerisch (mit der Simpson-Regel aus Aufgabe 4.4) auswerten, um das tatsächliche Alter unseres Universums zu berechnen.

8.6 Berechnung des Alters des Universums

In Kapitel 8 haben wir die allgemeine Friedmann-Gleichung (8.16) für eine Mischung aus Strahlung, Materie und Vakuumenergie aufgestellt:

$$\frac{\mathrm{d}a}{\mathrm{d}t} = H_0\left[\Omega_{\mathrm{rad}}\, a^{-2} + \Omega_{\mathrm{mat}}\, a^{-1} + \Omega_\Lambda\, a^2\right]^{1/2}. \tag{8.17}$$

Separation der Variablen liefert

$$\mathrm{d}t = \frac{1}{H_0}\,\frac{\mathrm{d}a}{\left[\Omega_{\mathrm{rad}}\, a^{-2} + \Omega_{\mathrm{mat}}\, a^{-1} + \Omega_\Lambda\, a^2\right]^{1/2}}\,. \tag{8.18}$$

Mit den Randbedingungen $a(t=0) = 0$ und $a(t_0) = 1$ (heute) erhalten wir formal

$$\boxed{t_0 = \frac{1}{H_0}\int_0^1 \frac{\mathrm{d}a}{\left[\Omega_{\mathrm{rad}}\, a^{-2} + \Omega_{\mathrm{mat}}\, a^{-1} + \Omega_\Lambda\, a^2\right]^{1/2}}}\,. \tag{8.19}$$

Anders als in den drei Spezialfällen aus Kapitel 8 (reine Materie, reine Strahlung, reines Vakuum) lässt sich dieses Integral für eine *Mischung* aller drei Komponenten nicht mehr analytisch (also „mit Stift und Papier“) lösen. Hier kommt die numerische Integration aus Aufgabe 4.4 ins Spiel: Wir werten Gleichung (8.19) mit der Simpson-Regel aus und können so endlich das Alter *unseres* Universums berechnen!

Aufgabe 8.2

Jetzt wollen wir das Alter t_0 *unseres* Universums berechnen. Yay :)

a) Das beste Modell für die Entwicklung und Zusammensetzung unseres Universums ist das kosmologische Standardmodell ΛCDM. Die heutige Energiezusammensetzung kann durch die relativen Energiedichten

$$\Omega_i = \frac{\rho_i}{\rho(t_0)} \quad \text{mit} \quad \rho(t_0) = \frac{3M_{\mathrm{Pl}}^2}{8\pi} H_0^2 \quad \text{und} \quad \Omega_{\mathrm{mat}} + \Omega_{\mathrm{rad}} + \Omega_\Lambda = 1$$

beschrieben werden kann. Du kannst die kosmologischen Parameter des ΛCDM-Modells dem Artikel Planck 2018 [2] entnehmen. Rufe darin Tabelle 2 auf und finde die am besten bestimmten Werte von H_0, Ω_{mat} und Ω_Λ.

b) Der Anteil Ω_{rad} der Strahlungsenergie am heutigen Energie-Inhalt des Universums lässt sich nicht direkt dem Artikel entnehmen. Du kannst ihn aber über die Gleichung

$$\Omega_{\mathrm{rad}} = \frac{\Omega_{\mathrm{mat}}}{1 + z_{\mathrm{eq}}}$$

berechnen, wobei z_{eq} wieder der Tabelle 2 zu entnehmen ist. Berechne Ω_{rad}.

c) Mit H_0, Ω_{mat}, Ω_{rad} und Ω_Λ sind nun alle nötigen Größen bekannt um t_0 für das ΛCDM-Modell zu berechnen. Im Kurs haben wir ein Integral für t_0 angegeben (Gleichung (8.19)), welches wir nicht analytisch lösen konnten. Implementiere den Ausdruck als `python`-Funktion und berechne t_0.

c) Herzlichen Glückwunsch! Du hast gerade das Alter des Universums berechnet. Sei stolz, genieße den Moment und spiele mit den Parametern Ω_{mat}, Ω_{rad} und Ω_Λ herum um deren Effekt auf t_0 besser zu verstehen. Beachte, dass $\Omega_{\mathrm{mat}} + \Omega_{\mathrm{rad}} + \Omega_\Lambda = 1$ in jedem Fall garantiert sein muss.

d) Setze nun $\Omega_{\mathrm{rad}} = 0$ und nutze, dass dann $\Omega_{\mathrm{mat}} + \Omega_\Lambda = 1$ gilt um t_0 in Abhängigkeit von $\Omega_{\mathrm{mat}} \in [0, 1]$ zu plotten.

e) Zuvor haben wir Ω_{rad} aus dem Zeitpunkt der matter-radiation equality berechnet, welcher durch $\rho_{\mathrm{rad}}(t_{\mathrm{eq}}) = \rho_{\mathrm{mat}}(t_{\mathrm{eq}})$ definiert ist. Der Redshift, der zu diesem Zeitpunkt korrespondiert, wird mit z_{eq} bezeichnet. Leite die oben benutzte Gleichung $\Omega_{\mathrm{rad}} = \Omega_{\mathrm{mat}}/(1 + z_{\mathrm{eq}})$ her.

Musterlösung 8.2

a) Aus Tabelle 2 des Planck-2018-Papers ([2], Spalte „TT,TE,EE+lowE+lensing“):

$$H_0 = 67.36 \, \frac{\mathrm{km}}{\mathrm{s\,Mpc}}\,, \qquad \Omega_{\mathrm{mat}} = 0.3153\,, \qquad \Omega_\Lambda = 0.6847\,. \tag{8.20}$$

(Verschiedene Datensätze liefern leicht unterschiedliche Werte, z.B. $\Omega_{\text{mat}} = 0.3104$, $\Omega_\Lambda = 0.6894$ – für unsere Zwecke spielt das keine große Rolle.)
b) Ebenfalls aus Tabelle 2: $z_{\text{eq}} \approx 3387$. Damit

$$\Omega_{\text{rad}} = \frac{\Omega_{\text{mat}}}{1 + z_{\text{eq}}} = \frac{0.3104}{3388} \approx 9.16 \times 10^{-5} \,. \tag{8.21}$$

Ω_{rad} ist also winzig im Vergleich zu Ω_{mat} und Ω_Λ; Strahlung trägt heute praktisch nicht mehr zur Energiebilanz des Universums bei (anders als kurz nach dem Urknall, vgl. Kapitel 1 und 18).
c) Wir implementieren Gleichung (8.19) mit der Simpson-Regel aus Aufgabe 4.4:
Zunächst die Konstanten aus den Tabellen sowie eine Implementierung der Simpson-Regel:

```
import numpy as np

# Define constants
H0_Gyr = 14.46                          # inverse Hubble constant in Gyr
Om_vac = 0.6894                         # Relative amount of Dark energy
Om_mat = 0.3104                         # Relative amount of matter energy
↪  (dark + baryonic)
z_eq   = 3387                           # Redshift of matter radiation equality
Om_rad = Om_mat / (1 + z_eq)            # Relative amount of radiation energy
N      = 1000                           # Number of subintervals to evaluate
↪  integral

def simpson(f, a, b, N):
    '''Approximate the integral of f(x) from a to b using Simpson's rule

    Parameters
    ----------
    f : function
        Vectorized function of a single variable
    a, b : float
        Interval of integration [a,b]
    N : (even) integer
        Number of subintervals of [a,b]
    '''
    if N % 2 != 0:
        print("N ist nicht gerade!")
        return np.nan
    else:
        h = (b - a) / N
        I = f(a) + f(b)
        x_k = a + h
        for k in range(1, round(N/2) + 1):
            I += 4 * f(x_k)
            x_k += 2 * h
        x_k = a + 2 * h
        for k in range(1, round(N/2)):
            I += 2 * f(x_k)
            x_k += 2 * h
        return (h / 3) * I
```

Nun der Integrand aus Gleichung (8.19) und die Auswertung des Integrals:

```
# integrand
def integrand(x):
```

```
    vac, mat, rad = Om_vac, Om_mat, Om_rad
    if x != 0:
        return 1/( x * np.sqrt(vac + mat * x**(-3) + rad * x**(-4)) )
    else:
        return 0

# Value of the integral over the scale factor from 0 to 1 (heute),
# depending on the number of subintervals N in Simpson's rule
def integral(N):
    return simpson(integrand, 0, 1, N)

age_Gyr = H0_Gyr * integral(N) # age of the Universe in billion years
print("Das Universum ist", round(age_Gyr,2), "Mrd. Jahre alt.")
```

Hier wurde im Integranden $a^{-2} = a \cdot a^{-3}$ und $a^2 = a \cdot a^4 \cdot a^{-2}$ etwas umsortiert; man kann leicht nachrechnen, dass

$$\left[\Omega_{\mathrm{rad}} a^{-2} + \Omega_{\mathrm{mat}} a^{-1} + \Omega_\Lambda a^2\right]^{1/2} = a \cdot \left[\Omega_{\mathrm{rad}} a^{-4} + \Omega_{\mathrm{mat}} a^{-3} + \Omega_\Lambda\right]^{1/2} \tag{8.22}$$

gilt, sodass `integrand(x)` genau dem Integranden in Gleichung (8.19) entspricht. Das Skript liefert

$$\boxed{t_0 \approx 13.8\,\mathrm{Gyr}}\,, \tag{8.23}$$

in exzellenter Übereinstimmung mit dem aus dem CMB bestimmten Alter des Universums.
d) Setzt man $\Omega_{\mathrm{rad}} = 0$ und $\Omega_\Lambda = 1 - \Omega_{\mathrm{mat}}$, kann man $t_0(\Omega_{\mathrm{mat}})$ für $\Omega_{\mathrm{mat}} \in [0, 1]$ plotten (z.B. mit `matplotlib`, indem man `Om_mat` in einer Schleife durchläuft und jeweils `integral(N)` aufruft). Man beobachtet: Für $\Omega_{\mathrm{mat}} \to 1$ (rein materiedominiert) nähert sich $t_0 \to \frac{2}{3} H_0^{-1} \approx 9.7\,\mathrm{Gyr}$ (vgl. Abschnitt 8.2) an, während für $\Omega_{\mathrm{mat}} \to 0$ (rein vakuumdominiert) $t_0 \to \infty$ divergiert: ein beschleunigt expandierendes, von Λ dominiertes Universum ist „unendlich alt“ in dem Sinne, dass $a(t) \to 0$ für $t \to -\infty$ (vgl. die Exponentialfunktion aus Kapitel 8). Das tatsächliche $t_0 \approx 13.8\,\mathrm{Gyr}$ liegt zwischen diesen beiden Extremen, näher am rein materiedominierten Wert, passend dazu, dass Λ erst in der jüngeren Vergangenheit des Universums dominant wurde.
e) Zur Zeit der „matter-radiation equality“ gilt $\rho_{\mathrm{rad}}(a_{\mathrm{eq}}) = \rho_{\mathrm{mat}}(a_{\mathrm{eq}})$. Mit $\rho_{\mathrm{rad}}(a) = \rho^0_{\mathrm{rad}} a^{-4}$ und $\rho_{\mathrm{mat}}(a) = \rho^0_{\mathrm{mat}} a^{-3}$ (Kapitel 8) folgt

$$\rho^0_{\mathrm{rad}} a_{\mathrm{eq}}^{-4} = \rho^0_{\mathrm{mat}} a_{\mathrm{eq}}^{-3} \quad \Longrightarrow \quad a_{\mathrm{eq}} = \frac{\rho^0_{\mathrm{rad}}}{\rho^0_{\mathrm{mat}}} = \frac{\Omega_{\mathrm{rad}}}{\Omega_{\mathrm{mat}}}\,, \tag{8.24}$$

wobei wir $\rho_0 = \rho(t_0)$ gekürzt haben, also $\Omega_i = \rho_i^0 / \rho_0$. Mit der Rotverschiebungs-Skalenfaktor-Beziehung $1 + z = 1/a$ (Kapitel 2, mit $a_0 = 1$) folgt

$$1 + z_{\mathrm{eq}} = \frac{1}{a_{\mathrm{eq}}} = \frac{\Omega_{\mathrm{mat}}}{\Omega_{\mathrm{rad}}} \quad \Longrightarrow \quad \boxed{1 + z_{\mathrm{eq}} = \frac{\Omega_{\mathrm{mat}}}{\Omega_{\mathrm{rad}}}} \quad \Longrightarrow \quad \boxed{\Omega_{\mathrm{rad}} = \frac{\Omega_{\mathrm{mat}}}{1 + z_{\mathrm{eq}}}}\,, \tag{8.25}$$

genau die Formel aus Teil b).

9 Dunkle Materie

In Kapitel 1 haben wir gesehen, dass rund 26 % der heutigen Energiedichte des Universums aus *Dunkler Materie* (DM) besteht, mehr als fünfmal so viel wie aus der uns vertrauten, baryonischen Materie. In diesem Kapitel fragen wir: Woher wissen wir das überhaupt? Und welche Art von Teilchen könnte Dunkle Materie sein?

9.1 Rotationskurven von Galaxien

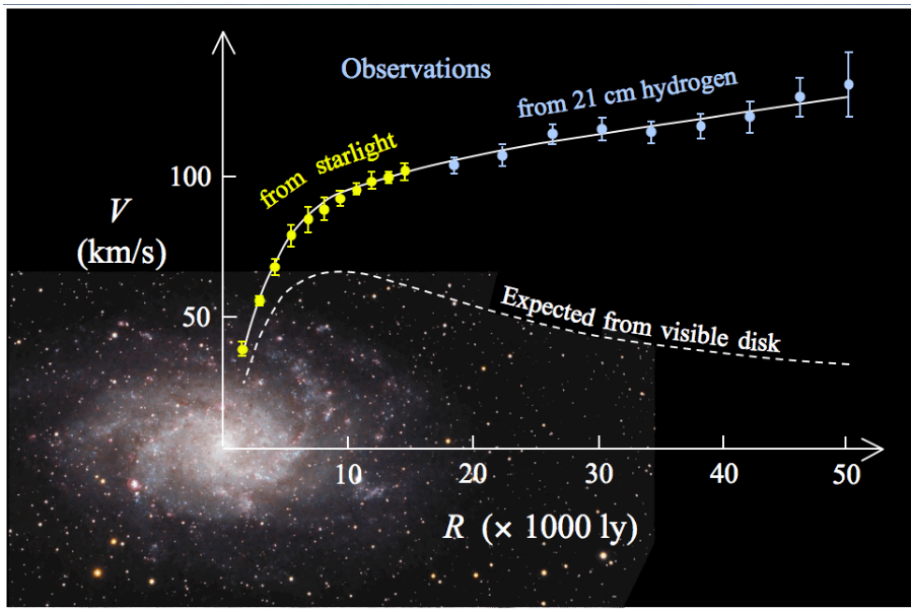


Abbildung 9.1: Beobachtete Rotationsgeschwindigkeit $v_c(r)$ einer typischen Spiralgalaxie als Funktion des Abstands r vom Galaxienzentrum. Statt wie erwartet für $r > R$ (Rand der sichtbaren Scheibe, gestrichelt) gemäß $v_c \propto r^{-1/2}$ abzufallen, bleibt v_c nahezu konstant, ein Hinweis auf eine zusätzliche, unsichtbare Masseverteilung (den „Dark-Matter-Halo“).

Eine typische Spiralgalaxie hat einen sichtbaren Radius $R \sim 10\,\mathrm{kpc} \approx 3 \times 10^4$ Lichtjahre. Die Rotationsgeschwindigkeit $v_c(r)$ von Sternen und Gas auf einer Kreisbahn mit Radius r folgt aus dem Gleichgewicht von Gravitationskraft und Zentripetalkraft,

$$\frac{GM(r)}{r^2} = \frac{v_c^2(r)}{r} \quad \Rightarrow \quad v_c(r) = \sqrt{\frac{GM(r)}{r}}\,, \tag{9.1}$$

wobei $M(r) = 4\pi \int_0^r \mathrm{d}r'\, r'^2 \rho(r')$ die Masse innerhalb des Radius r ist. Für $r > R$ erwarten wir $M(r) \approx M(R) = \text{const}$, also $v_c(r) \propto r^{-1/2}$; die Rotationsgeschwindigkeit sollte mit wachsendem Abstand *abfallen*, genau wie die Umlaufgeschwindigkeit der Planeten um die Sonne abnimmt.

Die Beobachtung (Abbildung 9.1, gemessen z.B. über die Dopplerverschiebung der 21 cm-Linie von neutralem Wasserstoffgas weit außerhalb der sichtbaren Scheibe) zeigt aber: $v_c(r) \approx$ const für $r > R$. Aus Gleichung (9.1) folgt damit $M(r) \propto r$, also

$$\rho(r) \propto r^{-2}\,. \tag{9.2}$$

Es muss also, weit über den sichtbaren Rand R der Galaxie hinaus, zusätzliche Masse vorhanden sein, deren Dichte mit r^{-2} abfällt.

Aufgabe 9.1

Welche Art von System hat ein Energiedichteprofil $\rho(r) \propto r^{-2}$? Betrachte eine kugelsymmetrische Wolke aus Teilchen der Masse m, die wie ein ideales Gas der Zustandsgleichung $p = \frac{k_B T}{m}\rho$ gehorcht und *isothermal* ist, d.h. $T(r) = T = \text{const}$. Im hydrostatischen Gleichgewicht hält der Druckgradient der Eigengravitation die Waage:

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{\rho(r)\, M(r)\, G}{r^2}\,. \tag{9.3}$$

Zeige mit dem Potenzansatz $\rho(r) = C \cdot r^{-b}$, dass $b = 2$ und $C = \frac{k_B T}{2\pi G m}$ die einzige Lösung ist, die $\rho \propto r^{-2}$ erfüllt.

Musterlösung 9.1

Mit der idealen Gasgleichung gilt $\mathrm{d}p/\,\mathrm{d}r = \frac{k_B T}{m}\,\mathrm{d}\rho/\,\mathrm{d}r$. Einsetzen in die hydrostatische Gleichung und Multiplizieren mit r^2/ρ gibt

$$\frac{k_B T}{G m}\frac{r^2}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r} = -M(r)\,. \tag{9.4}$$

Ableiten nach r und Verwenden von $\frac{\mathrm{d}M}{\mathrm{d}r} = 4\pi r^2 \rho(r)$ liefert

$$\frac{k_B T}{G m}\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{r^2}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r}\right] = -4\pi r^2 \rho(r)\,. \tag{9.5}$$

Einsetzen des Ansatzes $\rho(r) = Cr^{-b}$ (also $\frac{\mathrm{d}\rho}{\mathrm{d}r} = -bCr^{-b-1}$) ergibt auf der linken Seite $-\frac{bk_BT}{Gm}$ (also eine *Konstante*, unabhängig von r), auf der rechten Seite $-4\pi C r^{2-b}$. Damit beide Seiten für alle r übereinstimmen, muss $2 - b = 0$, also $b = 2$, und

$$C = \frac{k_B T}{2\pi G m}\,. \tag{9.6}$$

Ein isothermales, sich selbst gravitierendes Gas hat also automatisch $\rho(r) = \frac{k_B T}{2\pi G m}\, r^{-2}$; genau das Profil, das flache Rotationskurven erzeugt. Man spricht vom *Dark-Matter-Halo*: einer kugelförmigen Wolke aus unsichtbaren, untereinander nur gravitativ wechselwirkenden Teilchen, in deren Zentrum die sichtbare Galaxie eingebettet ist.

9.2 Was für ein Teilchen ist Dunkle Materie?

Typische Zahlen: Für Galaxien gilt $R \sim 100\,\mathrm{kpc}$, $M(R) \sim 10^{12}\, M_\odot$; für Zwerggalaxien $R \sim 1\,\mathrm{kpc}$, $M(R) \sim 10^7\, M_\odot$. Nehmen wir an, Dunkle Materie bestehe aus einem einzelnen Elementarteilchen der Masse m_{DM}; was lässt sich daraus über m_{DM} lernen?

Aufgabe 9.2

a) *De-Broglie-Argument:* Damit der DM-Halo überhaupt auf einer Längenskala R lokalisiert sein kann, muss die de-Broglie-Wellenlänge $\lambda_{\mathrm{dB}} = \hbar/(m_{\mathrm{DM}} v)$ der DM-Teilchen kleiner als R sein. Mit $v \approx v_c = \sqrt{GM(R)/R}$ aus Gleichung (9.1) folgt

$$\lambda_{\mathrm{dB}} \simeq \frac{\hbar}{m_{\mathrm{DM}}}\sqrt{\frac{R}{GM(R)}} < R\,. \tag{9.7}$$

Löse diese Ungleichung nach m_{DM} auf und schätze m_{DM} für eine Galaxie und eine Zwerggalaxie ab.
b) *Tremaine-Gunn-Argument (Pauli-Prinzip):* Wäre DM ein *Fermion*, dürfte pro Phasenraumzelle der Größe h^3 höchstens ein Teilchen sitzen, $N \lesssim \mathcal{V}/h^3$ mit $\mathcal{V} \simeq R^3(m_{\mathrm{DM}}v)^3$. Zeige, dass daraus

$$m_{\mathrm{DM}} \gtrsim \left(G^3 M(R) R^3 h^{-6}\right)^{-1/8} \tag{9.8}$$

folgt, und werte dies für Galaxien und Zwerggalaxien aus.

Musterlösung 9.2

a) Auflösen nach m_{DM}:

$$m_{\mathrm{DM}} > \frac{\hbar}{\sqrt{R\,M(R)\,G}} \approx \begin{cases} 5\times 10^{-25}\,\mathrm{eV} & \text{Galaxien} \\ 10^{-21}\,\mathrm{eV} & \text{Zwerggalaxien} \end{cases}. \tag{9.9}$$

b) Mit $M(R) = m_{\mathrm{DM}} N(R) \leq m_{\mathrm{DM}} \cdot R^3 (m_{\mathrm{DM}} v)^3 h^{-3}$ und $v = \sqrt{GM(R)/R}$ folgt $M(R) \leq m_{\mathrm{DM}}^4 R^3 [GM(R)/R]^{3/2} h^{-3}$, also

$$m_{\mathrm{DM}} \gtrsim \left(G^3 M(R) R^3 h^{-6}\right)^{-1/8} \approx \begin{cases} 20\,\mathrm{eV} & \text{Galaxie} \\ 500\,\mathrm{eV} & \text{Zwerggalaxie} \end{cases}. \tag{9.10}$$

Da Neutrinomassen auf $m_\nu < 2\,\mathrm{eV}$ beschränkt sind, scheiden Neutrinos als (alleinige) DM aus. Das ist die zentrale Erkenntnis dieses Abschnitts: **Es gibt kein Teilchen im Standardmodell, das Dunkle Materie sein kann**: entweder interagieren die Teilchen zu stark (die Materie wäre nicht „dunkel“), oder sie sind (wie Neutrinos) mit obigen Schranken inkompatibel. Dunkle Materie erfordert also *neue Physik* jenseits des Standardmodells, oder alternative Erklärungen wie primordiale Schwarze Löcher.

Neben Rotationskurven gibt es viele weitere, unabhängige Hinweise auf DM: der kosmische Mikrowellenhintergrund, N-Body-Simulationen der Strukturbildung, Gravitationslinseneffekte, der „Bullet Cluster“ und die Stabilität von Galaxienhaufen.

9.3 Freeze-out: wie viel Dunkle Materie bleibt übrig?

Ein populärer Mechanismus, um die heutige DM-Menge zu erklären, ist der *Freeze-out*: Direkt nach der Inflation ist DM im thermischen Gleichgewicht mit dem Standardmodell-Plasma, DM + DM $\leftrightarrow$ SM + SM. Die (über den Impulsraum integrierte) Boltzmann-Gleichung für die DM-Teilchenzahldichte n_{DM} lautet

$$\dot{n}_{\mathrm{DM}} + 3Hn_{\mathrm{DM}} = -\langle\sigma_{\mathrm{ann}} v\rangle \left[n_{\mathrm{DM}}^2 - \left(n_{\mathrm{DM}}^{\mathrm{eq}}\right)^2\right], \tag{9.11}$$

wobei der erste Term die Verdünnung durch die Hubble-Expansion (Kapitel 8) und der zweite Term Annihilation bzw. Produktion von DM-Paaren beschreibt. Mit der Entropiedichte $s \propto a^{-3}$ (Kapitel 8) definiert man den *Ertrag* $Y_{\mathrm{DM}} = n_{\mathrm{DM}}/s$ und die dimensionslose Variable $x = m_{\mathrm{DM}}/T$; daraus wird

$$\frac{\mathrm{d}Y_{\mathrm{DM}}}{\mathrm{d}x} = -C(x)\, x^{-n-2}\left(Y_{\mathrm{DM}}^2 - \left(Y_{\mathrm{DM}}^{\mathrm{eq}}\right)^2\right), \tag{9.12}$$

mit $C(x) = \sqrt{\pi/45}\sqrt{g_*}\, m_{\mathrm{Pl}} m_{\mathrm{DM}} \sigma_0$ und $\langle\sigma_{\mathrm{ann}} v\rangle = \sigma_0 x^{-n}$.

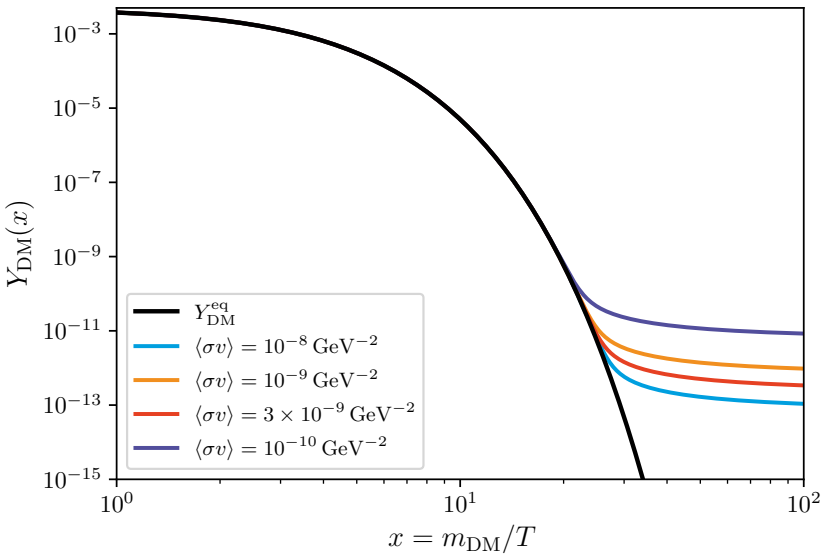


Abbildung 9.2: Numerische Lösungen der Freeze-out-Gleichung für $m_{\mathrm{DM}} = 1\,\mathrm{TeV}$ und verschiedene Annihilations-Wirkungsquerschnitte $\langle\sigma_{\mathrm{ann}}v\rangle$ (farbig) im Vergleich zum Gleichgewichtswert $Y_{\mathrm{DM}}^{\mathrm{eq}}(x)$ (schwarz). Solange $Y_{\mathrm{DM}} \approx Y_{\mathrm{DM}}^{\mathrm{eq}}$, folgt die DM-Dichte exponentiell dem Gleichgewicht; sobald die Annihilationsrate unter die Hubble-Rate fällt („Freeze-out“), bleibt Y_{DM} auf einem konstanten Plateauwert Y_{DM}^{∞} „eingefroren“.

Solange die Annihilationsrate $\Gamma_{\mathrm{ann}} = \langle\sigma_{\mathrm{ann}}v\rangle n_{\mathrm{DM}}$ groß gegen H ist, folgt Y_{DM} dicht dem Gleichgewichtswert $Y_{\mathrm{DM}}^{\mathrm{eq}}(x) \propto x^{3/2}\mathrm{e}^{-x}$, der wegen der Boltzmann-Unterdrückung (vgl. Kapitel 8) exponentiell abfällt. Sobald $\Gamma_{\mathrm{ann}} \lesssim H$, d.h. die Expansion „überholt“ die Annihilation, kann das Gleichgewicht nicht mehr aufrechterhalten werden, und Y_{DM} „friert“ bei einem konstanten Wert Y_{DM}^{∞} aus (Abbildung 9.2). *Größere* Wirkungsquerschnitte $\langle\sigma_{\mathrm{ann}}v\rangle$ führen zu *späterem* Freeze-out und damit zu *kleinerem* Y_{DM}^{∞}, also weniger übrig gebliebene DM.

Aufgabe 9.3

In dieser Aufgabe löst ihr die Boltzmann-Gleichung numerisch und reproduziert Abbildung 9.2. Für velocity-unabhängige s-Wellen-Annihilation ($n = 0$, $\langle\sigma_{\mathrm{ann}}v\rangle = \mathrm{const}$) lautet die Gleichung für den Ertrag $Y_{\mathrm{DM}}(x)$ mit $x = m_{\mathrm{DM}}/T$:

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\frac{\lambda}{x^2}\left[Y^2 - Y_{\mathrm{eq}}^2(x)\right], \tag{9.13}$$

mit $\lambda = \sqrt{\pi/45}\, m_{\mathrm{Pl}}\, m_{\mathrm{DM}}\, \sqrt{g_*}\, \langle\sigma_{\mathrm{ann}}v\rangle$ und dem Gleichgewichtswert

$$Y_{\mathrm{eq}}(x) = \frac{45}{4\pi^4}\,\frac{g_{\mathrm{DM}}}{g_*}\, x^2\, K_2(x), \tag{9.14}$$

wobei K_2 die modifizierte Bessel-Funktion zweiter Art ist (`scipy.special.kn(2,x)`), $g_{\mathrm{DM}} = 2$ die internen Freiheitsgrade (Majorana-Fermion), $g_* = 100$ die effektiven Freiheitsgrade beim Freeze-out und $m_{\mathrm{Pl}} = 1.22 \times 10^{19}\,\mathrm{GeV}$ die Planck-Masse.

1. Implementiert eine Funktion `Y_eq(x)`, die den Gleichgewichtswert berechnet. *Hinweis:* Verwendet `from scipy.special import kn`.

2. Löst Gleichung (9.13) mit `scipy.integrate.solve_ivp` (Methode `'Radau'`, `rtol=1e-8`, `atol=1e-18`) für $x \in [1, 100]$ mit Anfangsbedingung $Y(1) = Y_{\mathrm{eq}}(1)$ und $m_{\mathrm{DM}} = 1\,\mathrm{TeV}$.

3. Wiederholt die Rechnung für fünf verschiedene Wirkungsquerschnitte: $\langle\sigma v\rangle \in \{10^{-8}, 3\times10^{-8}, 10^{-9}, 3\times10^{-9}, 10^{-10}\}\,\mathrm{GeV}^{-2}$.

4. Plottet $Y_{\mathrm{DM}}(x)$ und $Y_{\mathrm{eq}}(x)$ auf einer doppelt-logarithmischen Skala und vergleicht eure Abbildung mit Abbildung 9.2.

Musterlösung 9.3

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.special import kn
from scipy.integrate import solve_ivp

MP    = 1.22e19   # Planck-Masse in GeV
m_DM  = 1e3       # DM-Masse in GeV (1 TeV)
g_DM  = 2         # interne Freiheitsgrade
g_eff = 100.0     # effektive Freiheitsgrade beim Freeze-out

def Y_eq(x):
    return (45 / (4 * np.pi**4)) * (g_DM / g_eff) * x**2 * kn(2, x)

def rhs(x, Y, sigmav):
    lam = np.sqrt(np.pi / 45) * MP * m_DM * np.sqrt(g_eff) * sigmav
    return [-(lam / x**2) * (Y[0]**2 - Y_eq(x)**2)]

x_eval = np.logspace(0, 2, 600)   # x von 1 bis 100

sigmav_vals = [1e-8, 3e-8, 1e-9, 3e-9, 1e-10]
labels = [r'$10^{-8}$', r'$3\times10^{-8}$', r'$10^{-9}$',
↪  r'$3\times10^{-9}$', r'$10^{-10}$']

fig, ax = plt.subplots(figsize=(5.5, 3.8))
ax.plot(x_eval, Y_eq(x_eval), 'k-', lw=2,
↪  label=r'$Y_\mathrm{DM}^\mathrm{eq}$')

for sigmav, lab in zip(sigmav_vals, labels):
    sol = solve_ivp(rhs, (1.0, 100.0), [Y_eq(1.0)],
                    args=(sigmav,), method='Radau', t_eval=x_eval,
                    rtol=1e-8, atol=1e-18)
    ax.plot(sol.t, sol.y[0], lw=1.8, label=lab)

ax.set_xscale('log'); ax.set_yscale('log')
ax.set_xlim(1, 100); ax.set_ylim(1e-15, 5e-3)
ax.set_xlabel(r'$x = m_\mathrm{DM}/T$')
ax.set_ylabel(r'$Y_\mathrm{DM}(x)$')
ax.legend(loc='lower left', fontsize=9.5, title=r'$\langle\sigma
↪  v\rangle\;[\mathrm{GeV}^{-2}]$')
plt.tight_layout()
plt.savefig('img/dm_freezeout.pdf', bbox_inches='tight')
plt.show()
```

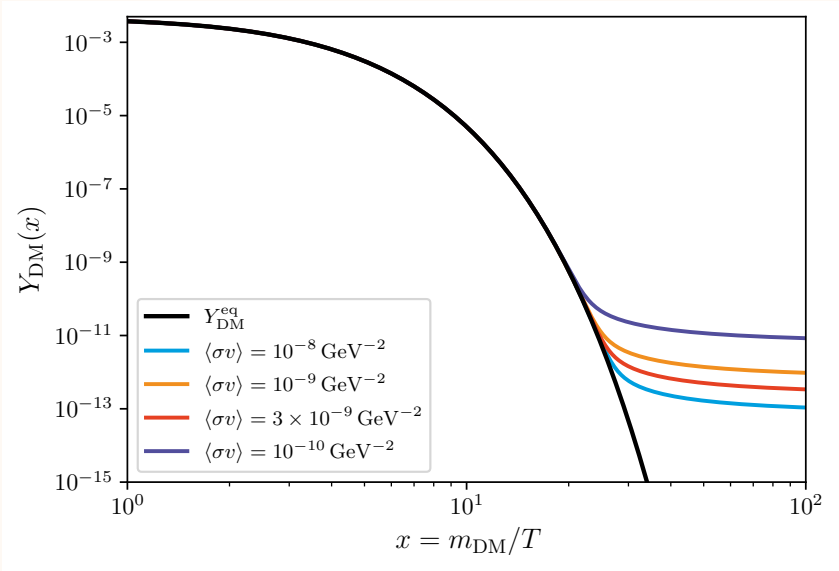


Abbildung 9.3: Lösung der Boltzmann-Gleichung für fünf Wirkungsquerschnitte $\langle\sigma v\rangle$. Je kleiner $\langle\sigma v\rangle$, desto früher friert die Dunkle Materie aus und desto größer ist der heutige Yield Y^{∞}_{DM}.

Die heutige DM-Dichte ergibt sich aus $\Omega_{\mathrm{DM}}h^2 = m_{\mathrm{DM}}Y^{\infty}_{\mathrm{DM}}s_0/(\rho^0_{\mathrm{tot}}/h^2)$. Der gemessene Wert $\Omega_{\mathrm{DM}}h^2 = 0.12$ (Kapitel 1) legt damit für eine gegebene Masse m_{DM} den benötigten Wirkungsquerschnitt fest; man findet näherungsweise

$$\Omega_{\mathrm{DM}}h^2 \approx \frac{3\times10^{-27}\,\mathrm{cm}^3/\mathrm{s}}{\langle\sigma_{\mathrm{ann}}v\rangle}\,. \tag{9.15}$$

Todo: WIMP exclusion plot. Abbildung 9.3 zeigt: Für DM-Massen $m_{\mathrm{DM}} \gtrsim \mathcal{O}(1\,\mathrm{GeV})$ sind die für $\Omega_{\mathrm{DM}}h^2 = 0.12$ benötigten Wirkungsquerschnitte inzwischen experimentell ausgeschlossen. Das klassische „WIMP“-Szenario (*weakly interacting massive particle*) steht damit unter Druck, und alternative Mechanismen, z.B. Freeze-in, dunkle Photonen, Axionen oder primordiale Schwarze Löcher, werden zunehmend wichtig. Welcher Mechanismus tatsächlich für die Entstehung der Dunklen Materie verantwortlich war, ist bis heute eine der großen offenen Fragen der Kosmologie und Teilchenphysik.

10 Fourier-Analysis

Wir verlassen jetzt für einen Moment die Kosmologie und wenden uns einem Werkzeug zu, das wir in den letzten Kapiteln dringend brauchen werden: der *Fourier-Analysis*. Sie erlaubt es, (fast) jedes periodische Signal als Überlagerung von einfachen Sinus- und Kosinusschwingungen darzustellen und wird uns helfen, Gravitationswellen (Kapitel 11) und das „Rauschen“ von Pulsaren (Kapitel 17) zu verstehen.

10.1 Wellen: $c = \lambda f$

Wir beginnen mit einer Beziehung, die du vermutlich schon aus der Schule kennst und die uns in diesem und den nächsten Kapiteln noch oft begegnen wird: Für jede Welle, die sich mit Geschwindigkeit c ausbreitet, hängen Wellenlänge λ und Frequenz f über

$$c = \lambda f \tag{10.1}$$

zusammen. Für elektromagnetische Wellen *und* für Gravitationswellen ist c die Lichtgeschwindigkeit, hohe Frequenz entspricht also einer kurzen Wellenlänge und umgekehrt.

10.2 Motivation: Wie hoch ist die Dusche?

Bevor wir formal werden, ein kleines Experiment, das du mit deinem Handy und einer Bluetooth-Box nachmachen kannst: Stelle die Box in eine Duschkabine, spiele ein „weißes Rauschen“ ab (alle Frequenzen gleichzeitig, gleich laut) und nimm das Ergebnis mit dem Mikrofon des Handys auf. Im aufgenommenen Spektrum wirst du feststellen, dass manche Frequenzen lauter sind als andere: Die Kabine „verstärkt“ ganz bestimmte Frequenzen und „dämpft“ andere.

Der Grund: Zwischen den parallelen Wänden der Duschkabine (Abstand L) bilden sich *stehende Schallwellen* aus – Resonanzen, bei denen genau eine halbe Wellenlänge (oder ein ganzzahliges Vielfaches davon) zwischen die Wände passt,

$$L = n \cdot \frac{\lambda_n}{2}\,, \qquad n = 1, 2, 3, \ldots \tag{10.2}$$

Mit der Schallgeschwindigkeit $c \approx 343\,\mathrm{m/s}$ und der Wellengleichung $c = \lambda f$ aus Gleichung (10.1) entsprechen diese Resonanzwellenlängen den Frequenzen

$$f_n = \frac{c}{\lambda_n} = n \cdot \frac{c}{2L}\,. \tag{10.3}$$

Misst man also im aufgenommenen Spektrum (mehr dazu gleich) die Frequenz f_1 der niedrigsten Resonanz, lässt sich daraus die Breite der Dusche bestimmen: $L = c/(2f_1)$. Für $L \approx 1\,\mathrm{m}$ erwarten wir $f_1 \approx 170\,\mathrm{Hz}$, ein Wert im hörbaren Bereich liegt und sich mit einem Smartphone-Mikrofon problemlos messen lässt!

Um aus einer Tonaufnahme (einem Signal $f(t)$, das angibt, wie laut es zu jedem Zeitpunkt t ist) herauszufinden, *welche* Frequenzen besonders stark vertreten sind, brauchen wir genau das Werkzeug, das wir in diesem Kapitel entwickeln: die *Fourier-Analysis*. Sie zerlegt ein beliebiges Signal $f(t)$ in seine Frequenzanteile und macht aus einer unübersichtlichen Wellenform ein „Spektrum“, in dem die Resonanzfrequenzen der Dusche als klar erkennbare Spitzen auftauchen.

10.3 Funktionen als Vektoren

Aus der linearen Algebra kennst du Vektoren im $\mathbb{R}^3$, die man in einer beliebigen Orthonormalbasis $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ zerlegen kann:

$$\vec{v} = \sum_{i=1}^{3} v_i\, \vec{e}_i\,, \qquad v_i = \vec{v} \cdot \vec{e}_i\,. \tag{10.4}$$

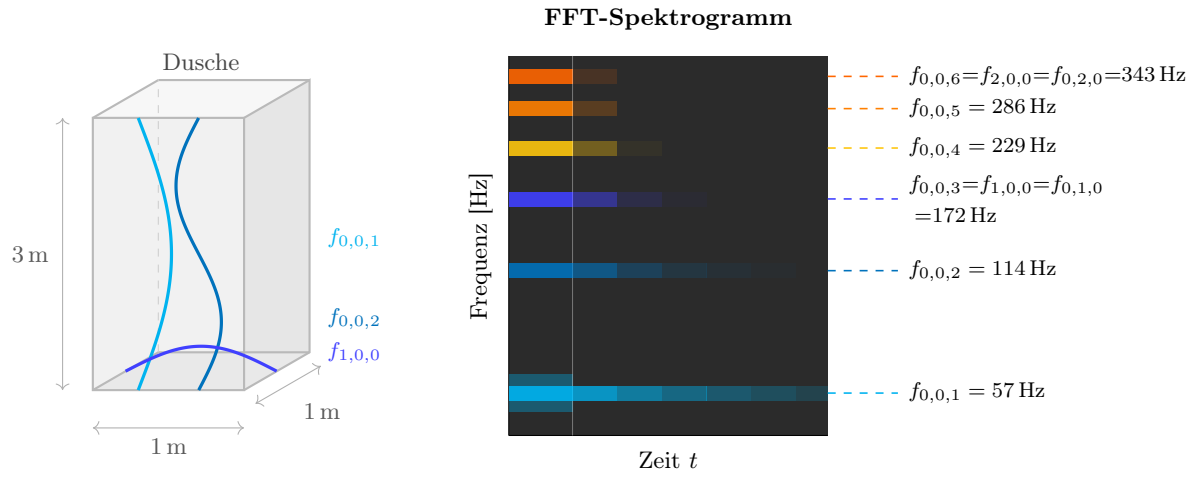


Abbildung 10.1: Schematische Darstellung der Dusche ($1\,\mathrm{m} \times 1\,\mathrm{m} \times 3\,\mathrm{m}$, $c = 343\,\mathrm{m/s}$) als akustischer Resonator. *Links:* Aufrechte Duschkabine mit stehenden Wellen: Vertikale Moden $f_{0,0,1}$ und $f_{0,0,2}$ entlang der 3 m-Höhe (hellblau/blau) und horizontale Mode $f_{1,0,0}$ auf dem Boden entlang der 1 m-Breite (blau). *Rechts:* FFT-Spektrogramm (logarithmische Frequenzachse): Ein Ton nahe 57 Hz regt alle Resonanzen an (helle Bänder links); danach klingen die Moden frei aus, wobei tiefere Frequenzen länger nachhallen. Die Mode $f_{0,0,3}$=$f_{1,0,0}$=$f_{0,1,0}$ = 172 Hz ist dreifach entartet (alle drei Raumachsen liefern bei gleicher Frequenz eine stehende Welle).

Die Idee der Fourier-Analysis ist es, dasselbe mit *Funktionen* zu tun: Eine (hinreichend gutartige, periodische) Funktion $f(t)$ mit Periode T kann als „Vektor“ in einem unendlich-dimensionalen Funktionenraum aufgefasst werden, und die Funktionen

$$\left\{1,\ \cos\left(\frac{2\pi n t}{T}\right),\ \sin\left(\frac{2\pi n t}{T}\right)\right\}_{n=1,2,3,\ldots} \tag{10.5}$$

bilden darin eine Orthogonalbasis, die sogenannte *Fourier-Basis*. Jede periodische Funktion lässt sich also als

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n \cos\left(\frac{2\pi n t}{T}\right) + b_n \sin\left(\frac{2\pi n t}{T}\right)\right] \tag{10.6}$$

schreiben. Die Koeffizienten a_n, b_n geben an, „wie viel“ von der Schwingung mit Frequenz $f_n = n/T$ (der n-ten *Harmonischen*) in $f(t)$ enthalten ist, ganz analog zu den Komponenten v_i eines Vektors.

10.4 Ein Beispiel: die Rechteckwelle

Ein klassisches Beispiel ist die Rechteckwelle (Periode $T = 2\pi$, Amplitude ± 1). Ihre Fourier-Reihe enthält nur ungerade Harmonische:

$$f_{\text{Rechteck}}(t) = \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{\sin((2k-1)t)}{2k-1} = \frac{4}{\pi}\left(\sin t + \frac{\sin 3t}{3} + \frac{\sin 5t}{5} + \ldots\right). \tag{10.7}$$

Je mehr Terme man mitnimmt, desto besser approximiert die Summe die „eckige“ Rechteckwelle – an den Sprungstellen bleibt allerdings immer ein kleines Überschwingen (das sogenannte *Gibbs-Phänomen*). Abbildung 10.2 zeigt diese Konvergenz für $N = 1, 3, 5, 11$ Terme.

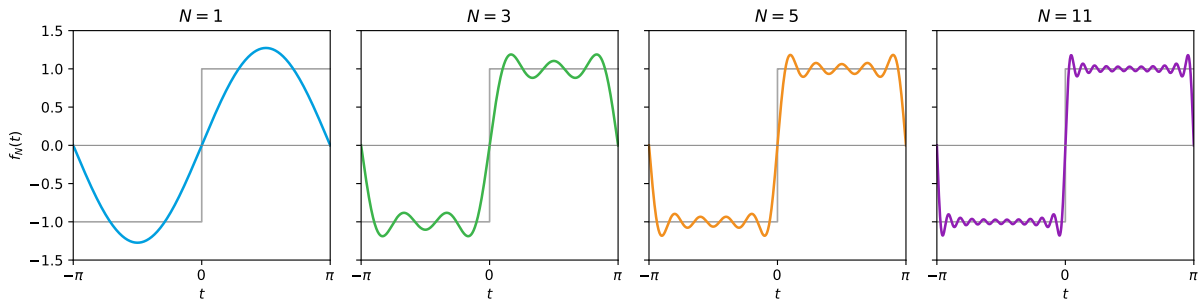


Abbildung 10.2: Fourier-Zerlegung der Rechteckwelle: die Partialsumme $f_N(t) = \frac{4}{\pi}\sum_{k=1}^{N}\frac{\sin[(2k-1)t]}{2k-1}$ für $N = 1, 3, 5, 11$ Terme (farbig) im Vergleich zur exakten Rechteckwelle (grau). Mit wachsendem N approximiert die Summe die Rechteckwelle immer besser; an den Sprungstellen bleibt jedoch stets ein kleines Überschwingen sichtbar (*Gibbs-Phänomen*).

10.5 Spektren

Statt ein Signal $f(t)$ im *Zeitbereich* zu betrachten (also als Funktion der Zeit t), kann man es im *Frequenzbereich* betrachten: Man trägt einfach die Koeffizienten a_n, b_n (bzw. ihre Beträge) gegen die Frequenz f_n auf. Man spricht dann vom *Spektrum* des Signals. Betrachten wir als Beispiel

$$f(t) = 2\sin(t) + 4\sin(5t) + \sin(8t)\,. \tag{10.8}$$

Im Zeitbereich sieht $f(t)$ wie eine komplizierte, „verbeulte“ Schwingung aus, während er im Frequenzbereich besteht sein Spektrum jedoch nur aus drei „Linien“ bei den Frequenzen 1, 5 und 8 (in Einheiten von $1/(2\pi)$), mit Höhen 2, 4 und 1. Die gesamte Information über $f(t)$ steckt also vollständig (und oft viel übersichtlicher!) in seinem Spektrum.

Diese Idee ist überall in der Physik zentral: Ein Prisma zerlegt Licht in sein Spektrum (die Frequenzen der enthaltenen elektromagnetischen Wellen), ein Mikrofon plus Computer kann einen Akkord in seine einzelnen Tonfrequenzen zerlegen – und ein Gravitationswellendetektor oder ein Pulsar-Timing-Array zerlegt ein verrauschtes Zeitsignal in ein Frequenzspektrum, um darin nach den charakteristischen Signaturen von Gravitationswellen zu suchen (Kapitel 11 und 17).

Aufgabe 10.1

Betrachte das Signal

$$f(t) = 2\sin(t) + 4\sin(5t) + \sin(8t)\,.$$

a) Plotte $f(t)$ für $t \in [0, 2\pi]$ (z.B. mit `numpy` und `matplotlib`). Ist aus dem Zeitsignal allein leicht erkennbar, aus welchen Schwingungen $f(t)$ zusammengesetzt ist?

b) Erstelle nun einen zweiten Plot, der das *Spektrum* von f zeigt: auf der x-Achse die (Kreis-)Frequenzen $1, 5, 8$, auf der y-Achse die zugehörigen Amplituden $2, 4, 1$ (als Balken oder einzelne Punkte). *Tipp: In Python kann man dieses „Spektrum von Hand“ einfach als zwei Arrays* `frequenzen = [1,5,8]` *und* `amplituden = [2,4,1]` *eintragen und mit* `plt.stem` *oder* `plt.bar` *darstellen.*

c) Eine Gravitationswelle mit Frequenz $f = 100\,\mathrm{Hz}$ erreicht die Erde mit Lichtgeschwindigkeit $c \approx 3 \times 10^8\,\mathrm{m/s}$. Welche Wellenlänge λ hat diese Welle (Gleichung (10.1))? Vergleiche λ mit dem Durchmesser der Erde ($\approx 12{,}700\,\mathrm{km}$).

d) Pulsar-Timing-Arrays messen Gravitationswellen bei Frequenzen von $f \sim 1\,\mathrm{nHz}$ ($1\,\mathrm{nHz} = 10^{-9}\,\mathrm{Hz}$). Berechne die Wellenlänge dieser Wellen in Lichtjahren ($1\,\mathrm{Lj} \approx$

9.46×10^{15} m). Was sagt dir dieses Ergebnis über die Größe eines geeigneten Detektors?

Musterlösung 10.1

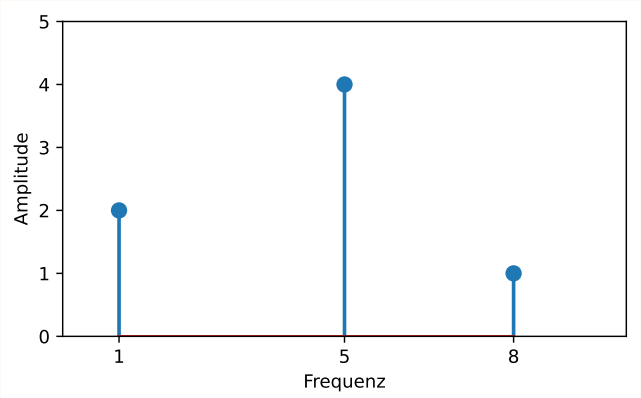
a) Pythoncode:

```
import numpy as np
import matplotlib.pyplot as plt

t = np.linspace(0, 2*np.pi, 1000)
f = 2*np.sin(t) + 4*np.sin(5*t) + np.sin(8*t)

plt.plot(t, f)
plt.xlabel("t")
plt.ylabel("f(t)")
plt.show()
```

Im Zeitbereich sieht $f(t)$ aus wie eine unregelmäßige, „verbeulte“ Schwingung; man erkennt zwar, dass mehrere Frequenzen überlagert sind, aber nicht ohne Weiteres, *welche* das genau sind oder mit welchem relativen Gewicht sie auftreten.

b) Pythoncode:

```
frequenzen = [1, 5, 8]
amplituden = [2, 4, 1]

plt.stem(frequenzen, amplituden)
plt.xlabel("Frequenz")
plt.ylabel("Amplitude")
plt.show()
```



Im Spektrum sieht man auf einen Blick drei „Linien“ bei $f = 1, 5, 8$ mit Höhen $2, 4, 1$, exakt die drei Bausteine, aus denen $f(t)$ besteht. Diese Darstellung ist viel informativer als die Zeitreihe aus a): Man liest die Zusammensetzung des Signals direkt ab.

c) Aus Gleichung (10.1):

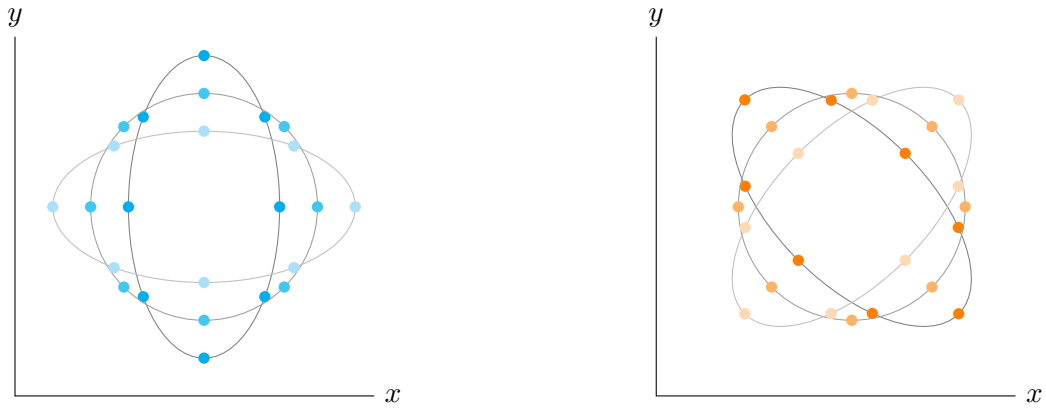
$$\lambda = \frac{c}{f} = \frac{3 \times 10^8\,\mathrm{m/s}}{100\,\mathrm{Hz}} = 3 \times 10^6\,\mathrm{m} = 3000\,\mathrm{km}\,. \tag{10.9}$$

Diese Wellenlänge ist von derselben Größenordnung wie der Durchmesser der Erde ($\approx$ 12,700 km, also nur etwa 4-mal größer)! Allgemein gilt, dass die Aperatur mit der man Gravitationswellen messen will ungefähr so groß sein muss wie die Wellenlänge der Welle. So lassen

sich Resonanzeffekte ausnutzen, die aus einem kleinen Signal eine messbare Größe machen. Natürlich sind die Aperaturen, mit denen wir Gravitationswellen auf der Erde messen, keine tausenden von Kilometern groß. Durch ausgeklügelte Spiegelsysteme lassen sich diese Aperate auch um einen Faktor 1000 kleiner bauen - mehr dazu in Kapitel 11. Das erklärt, warum Gravitationswellendetektoren wie LIGO Ausmaße von der Größenordnung von Kilometern haben müssen und warum man für Gravitationswellen mit noch viel größeren Wellenlängen (wie sie von supermassiven Schwarzen-Loch-Paaren erzeugt werden) ein Pulsar-Timing-Array von der Größe der Milchstraße braucht (Kapitel 17).
d) Aus $\lambda = c/f$:

$$\lambda = \frac{3 \times 10^8\,\mathrm{m/s}}{10^{-9}\,\mathrm{Hz}} = 3 \times 10^{17}\,\mathrm{m} \approx \frac{3 \times 10^{17}\,\mathrm{m}}{9.46 \times 10^{15}\,\mathrm{m/Lj}} \approx 32\,\mathrm{Lj}\,. \tag{10.10}$$

Die Wellenlänge einer nHz-Gravitationswelle beträgt also etwa 30 Lichtjahre – ein Detektor müsste (grob gesagt) von ähnlicher Größe sein. Kein auf der Erde oder im Sonnensystem baubareres Interferometer kommt auch nur annähernd daran heran. Deshalb nutzen Pulsar-Timing-Arrays die Galaxis selbst als Detektor, mit Pulsaren in Abständen von Tausenden von Lichtjahren als „Spiegel“ – genau das werden wir in Kapitel 17 sehen.



(a) +-Polarisierung (b) ×-Polarisierung

Abbildung 11.1: Schematische Darstellung des Effekts einer +-polarisierten (links) und ×-polarisierten Gravitationswelle auf einen Ring von Testmassen in der x-y-Ebene. Die Farbsättigung der Punkte zeigt die Zeitabhängigkeit.

11 Gravitationswellen

In Kapitel 2 haben wir Einsteins Bild der Allgemeinen Relativitätstheorie (ART) kennengelernt: „Matter tells space how to curve, and space tells matter how to move.“ Massen krümmen die Raumzeit, und diese Krümmung bestimmt, wie sich andere Massen bewegen. Doch die Raumzeit ist nicht starr, sondern kann auch *schwingen*. Genau wie eine gespannte Membran, die man anstößt, Wellen aussendet, können beschleunigte Massen Wellen in der Raumzeit selbst erzeugen: *Gravitationswellen*.

Einstein sagte Gravitationswellen bereits 1916 voraus, aber es dauerte fast 100 Jahre, bis sie 2015 erstmals direkt gemessen werden konnten: Die Kollaborationen LIGO und Virgo beobachteten das Verschmelzen zweier schwarzer Löcher mit Massen von 36 ± 6 und 29 ± 4 Sonnenmassen, wobei eine Energie von etwa 3 Sonnenmassen als Gravitationswellen abgestrahlt wurde. 2017 gelang zusätzlich die Beobachtung eines verschmelzenden Neutronensternpaares, begleitet von einem Gammablitz (GRB), der am Himmel praktisch gleichzeitig eintraf. Daraus folgt, dass sich Gravitationswellen mit einer Geschwindigkeit c_{gw} ausbreiten, die sich von der Lichtgeschwindigkeit c um höchstens $\sim 10^{-15}$ unterscheidet, eine der präzisesten Tests der Allgemeinen Relativitätstheorie überhaupt. Inzwischen wurden mehr als einhundert solcher „Binärsysteme“ beobachtet. Dieses Kapitel soll eine Intuition dafür liefern, was eine Gravitationswelle ist, wovon sie verursacht werden können und wie man sie mathematisch beschreiben und beobachten kann. Wir halten uns dabei nah am Übersichtsartikel von Hindmarsh, Lüben, Lumma & Pauly [6], insbesondere Kapitel 8 davon.

11.1 Was ist eine Gravitationswelle?

Eine Gravitationswelle ist eine sich ausbreitende „Kräuselung“ der Raumzeit, ein periodisches Dehnen und Stauchen von Längen, das sich mit Lichtgeschwindigkeit c ausbreitet. Stell dir einen Ring aus frei schwebenden Testmassen vor, durch dessen Mitte eine Gravitationswelle läuft (senkrecht zur Ringebene): Während die Welle durchläuft, wird der Ring abwechselnd in die eine Richtung gestreckt und in die dazu senkrechte Richtung gestaucht und danach umgekehrt. (Man spricht auch von der „quadrupolaren“ Natur von Gravitationswellen.) Diese Verzerrung oszilliert mit der Frequenz f der Gravitationswelle. Abb. 11.1 zeigt eine Skizze dieser möglichen Verformungen.

Man quantifiziert diese Verzerrung durch die dimensionslose *Strain* (Dehnung)

$$h = \frac{\Delta L}{L}\,, \tag{11.1}$$

wobei L die ursprüngliche Länge und ΔL die durch die Gravitationswelle hervorgerufene Längenänderung ist. Gravitationswellen von astrophysikalischen Quellen, die auf der Erde ankommen, haben typischerweise winzige Amplituden von $h \sim 10^{-21}$ oder kleiner; das ist der Grund, warum es so unglaublich schwer ist, sie zu messen (und warum der erste direkte Nachweis erst 2015 gelang, ziemlich genau 100 Jahre nach Einsteins Vorhersage)!

11.2 Wie entstehen Gravitationswellen?

Damit ein System Gravitationswellen abstrahlt, muss es eine zeitlich veränderliche, asymmetrische Massenverteilung haben; eine perfekt kugelsymmetrische, pulsierende Masse strahlt z.B. *keine* Gravitationswellen ab. Die „lautesten“ Quellen sind daher Systeme aus zwei umeinander kreisenden, sehr kompakten und massereichen Objekten:

- **Verschmelzende Doppelsysteme** aus Neutronensternen oder schwarzen Löchern: Während die beiden Objekte umeinander kreisen, strahlen sie Energie in Form von Gravitationswellen ab, rücken dadurch näher zusammen und kreisen immer schneller, bis sie schließlich verschmelzen. Das resultierende Signal ist ein charakteristischer „Chirp“: eine Welle, deren Frequenz und Amplitude mit der Zeit zunehmen.

- **Stochastische Hintergründe**, z.B. aus Phasenübergängen im frühen Universum (Kapitel 18) oder aus der Überlagerung unzähliger supermassiver Schwarzer-Loch-Doppelsysteme (Kapitel 17). Hier ist kein einzelnes „Chirp“-Ereignis sichtbar, sondern ein konstantes, verrauschtes Signal aus vielen überlagerten Quellen.

11.3 Messung: das Michelson-Interferometer

Wie misst man eine relative Längenänderung von $\Delta L/L \sim 10^{-21}$? Die Antwort: mit einem *Laser-Interferometer*, wie es z.B. bei den Detektoren LIGO, Virgo oder (zukünftig) LISA verwendet wird. Das Grundprinzip ist ein *Michelson-Interferometer*:

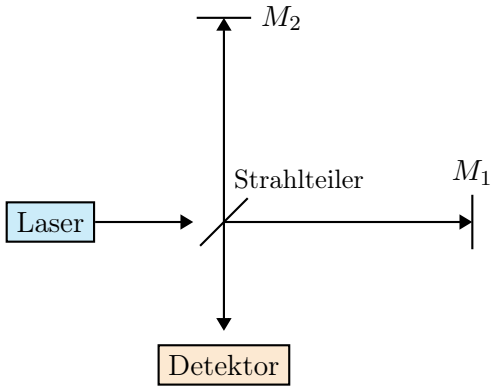
Ein Laserstrahl wird an einem Strahlteiler in zwei Teilstrahlen aufgespalten, die senkrecht zueinander zwei lange Arme (bei LIGO: $L = 4\,\mathrm{km}$) durchlaufen, an Spiegeln reflektiert werden und sich am Strahlteiler wieder überlagern. Die beiden Teilstrahlen interferieren, je nachdem ob sich die Wellenberge des einen Strahls mit den Wellenbergen oder den Wellentälern des anderen überlagern, hängt extrem empfindlich von der *Längendifferenz* der beiden Arme ab (auf Bruchteile der Laserwellenlänge λ genau, vgl. Kapitel 10, Gleichung (10.1)).

Läuft nun eine Gravitationswelle durch den Detektor, dehnt sie einen Arm minimal und staucht gleichzeitig den anderen (oder umgekehrt); die Längendifferenz ΔL zwischen den Armen ändert sich, und damit ändert sich das Interferenzmuster am Ausgang. Aus dieser winzigen Änderung des Interferenzmusters lässt sich $h = \Delta L/L$ und damit das Gravitationswellensignal selbst rekonstruieren (Abbildung 11.2).

11.4 Frequenzbänder und Detektoren

Genau wie Licht (Radiowellen, Mikrowellen, sichtbares Licht, Röntgenstrahlung, ...) treten Gravitationswellen über ein riesiges Spektrum von Frequenzen auf, und je nach Frequenz braucht man völlig unterschiedliche „Antennen“, um sie zu messen:

- $\sim$ 10–1000 Hz: Verschmelzende Neutronensterne und stellare schwarze Löcher. Gemessen von bodengebundenen Laserinterferometern wie LIGO und Virgo (Armlänge $\mathcal{O}$(km)).



(a) Schematischer Aufbau: Laser, Strahlteiler, Spiegel M_1 und M_2, Detektor.

(b) Kreisförmiges Interferenzmuster (konzentrische helle und dunkle Ringe) am Detektor.

Abbildung 11.2: Michelson-Interferometer (a) und Interferenzmuster (b), wie es bei LIGO, Virgo und KAGRA eingesetzt wird. Eine durchlaufende Gravitationswelle ändert die relative Länge der beiden Arme; je nach Phasenlage wechseln helle und dunkle Ringe ihren Radius.

- $\sim 10^{-4}$–10^{-1} Hz: Verschmelzende mittelschwere/massereiche schwarze Löcher. Geplant: weltraumgestützte Interferometer wie LISA (Armlänge $\mathcal{O}(10^6\,\mathrm{km})$).
- $\sim 10^{-9}$–10^{-7} Hz (nHz): Supermassive Schwarze-Loch-Doppelsysteme und ggf. Phasenübergänge im frühen Universum (Kapitel 18). Gemessen mit *Pulsar-Timing-Arrays*, „Antennen“ von der Größe der Milchstraße! Dazu mehr in Kapitel 17.

Je niedriger die Frequenz der Gravitationswelle, desto größer muss (gemäß $c = \lambda f$, Kapitel 10) die entsprechende Wellenlänge sein, und desto größer muss auch der „Detektor“ sein, der diese Wellenlänge auflösen kann.

In Tabelle 11.1 sind aktuelle und geplante bodengebundene Laserinterferometer zusammengestellt; ihre Armlängen liegen alle in der Größenordnung weniger Kilometer, passend zum $\sim$ 10–1000 Hz-Band.

Name	Standort	Armlänge
GEO	Deutschland	0.6 km
(a)LIGO	USA (2×)	4 km
(a)Virgo	Italien	3 km
KAGRA	Japan	4 km
LIGO-India*	Indien	4 km

Tabelle 11.1: Aktuelle und geplante (*) bodengebundene Gravitationswellendetektoren mit Laserinterferometrie, nach [6].

Für *kosmologische* Quellen ist neben dem heutigen Frequenzband auch interessant, bei welcher Frequenz ein Signal *ursprünglich* erzeugt wurde. Als grobe Abschätzung gilt für ein Ereignis zur Zeit t die Frequenz $f \sim t^{-1} \sim H$; durch die kosmische Expansion wird diese Frequenz bis heute auf $f_0 = \frac{a(t)}{a(t_0)} f$ rotverschoben. Tabelle 11.2 listet die minimalen (heute beobachtbaren) Frequenzen für einige potentielle Quellen im frühen Universum.

Für das mHz-Band, in dem EW-Phasenübergänge ihr Signal hinterlassen, ist die geplante Weltraummission *LISA* (Laser Interferometer Space Antenna) konzipiert (Abbildung 11.3).

Ereignis	Zeit [s]	Temperatur [GeV]	Frequenz [Hz]
QCD-Phasenübergang	10^{-3}	0.1	10^{-8}
EW-Phasenübergang	10^{-11}	100	10^{-5}
Ende der Inflation	$\geq 10^{-36}$	$\leq 10^{16}$	$\geq 10^{8}$

Tabelle 11.2: Heute beobachtbare Minimalfrequenzen von Gravitationswellen aus verschiedenen potentiellen Quellen im frühen Universum, nach [6]. Der elektroschwache (EW) Phasenübergang – das Thema von Kapitel 18, liegt bei $\sim 10^{-5}\,\mathrm{Hz}$, also genau im LISA-Band.

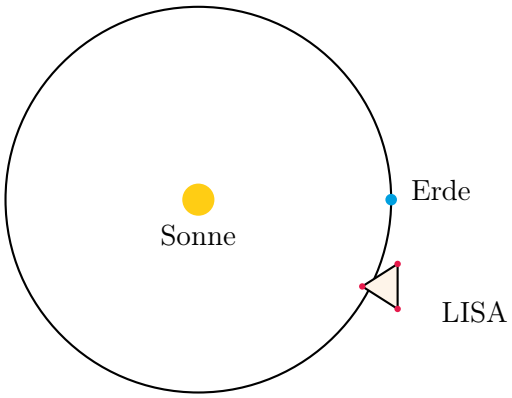


Abbildung 11.3: Schematische Darstellung der LISA-Mission: Drei Satelliten bilden ein gleichseitiges Dreieck mit Armlängen von $2.5 \times 10^6\,\mathrm{km}$ und umkreisen, der Erde leicht voraus- bzw. nachlaufend, gemeinsam die Sonne. Über Laserlinks zwischen den Satelliten misst LISA Gravitationswellen im $\sim 10^{-4}$–$10^{-1}\,\mathrm{Hz}$-Band – geplanter Start: 2034.

11.5 Das Gravitationswellenspektrum $\Omega_{\mathrm{gw}}(f)$

Eine Gravitationswelle trägt selbst Energie. Genauso wie wir in Kapitel 2 von einer *Energiedichte* ρ für Materie, Strahlung und Vakuum gesprochen haben, können wir daher auch einer Gravitationswelle eine Energiedichte ρ_{gw} zuordnen. Da Gravitationswellen über ein ganzes Spektrum von Frequenzen f verteilt sein können (z.B. ein stochastischer Hintergrund), betrachtet man üblicherweise den Beitrag $\mathrm{d}\rho_{\mathrm{gw}}$ eines logarithmischen Frequenzintervalls $\mathrm{d}\ln f$ und vergleicht ihn mit der *kritischen* Energiedichte ρ_{crit}, die nach Gleichung (2.3) über $H^2 = \frac{8\pi G}{3}\rho_{\mathrm{crit}}$ mit der heutigen Hubble-Rate zusammenhängt. Daraus definiert man das dimensionslose *Gravitationswellenspektrum*

$$\Omega_{\mathrm{gw}}(f) := \frac{1}{\rho_{\mathrm{crit}}}\frac{\mathrm{d}\rho_{\mathrm{gw}}}{\mathrm{d}\ln f}\,. \tag{11.2}$$

$\Omega_{\mathrm{gw}}(f)$ gibt also an, welcher *Anteil* der heutigen kritischen Energiedichte des Universums in Form von Gravitationswellen bei der Frequenz f vorliegt, typischerweise winzige Werte von $\Omega_{\mathrm{gw}} \sim 10^{-10}$ oder kleiner. Diese Größe wird uns in den nächsten Kapiteln immer wieder begegnen: Sowohl der Gravitationswellenhintergrund aus Phasenübergängen (Kapitel 18) als auch der von Pulsar-Timing-Arrays gesuchte nHz-Hintergrund (Kapitel 17) werden über ihr jeweiliges Spektrum $\Omega_{\mathrm{gw}}(f)$ charakterisiert.

Die Kombination $h^2\Omega_{\mathrm{gw}}$. In der Praxis verwendet man statt Ω_{gw} oft die Kombination

$$h^2\Omega_{\mathrm{gw}}(f)\,, \quad h := \frac{H_0}{100\,\frac{\mathrm{km}}{\mathrm{s\,Mpc}}} \approx 0.674\,, \tag{11.3}$$

wobei h der dimensionslose Hubble-Parameter ist. Der Vorteil: Da $\rho_{\mathrm{crit}} = 3H_0^2/(8\pi G) \propto H_0^2$ und der Beitrag $d\rho_{\mathrm{gw}}$ selbst nicht von H_0 abhängt, ist $\Omega_{\mathrm{gw}} \propto H_0^{-2}$, also $h^2\Omega_{\mathrm{gw}}$ *unabhängig von*

H_0. Angesichts der ungelösten H_0-Spannung ($\sim 5\sigma$ Diskrepanz zwischen frühem und spätem Universum, Abbildung 2.4, vgl. Kapitel 2) ist das ein wichtiger praktischer Vorteil: Verschiedene Experimente können $h^2\Omega_{\mathrm{gw}}$ direkt miteinander vergleichen, ohne sich auf einen gemeinsamen H_0-Wert einigen zu müssen. Abbildung 11.4 zeigt typische Spektren für verschiedene kosmologische Quellen zusammen mit den Sensitivitätskurven geplanter Detektoren.

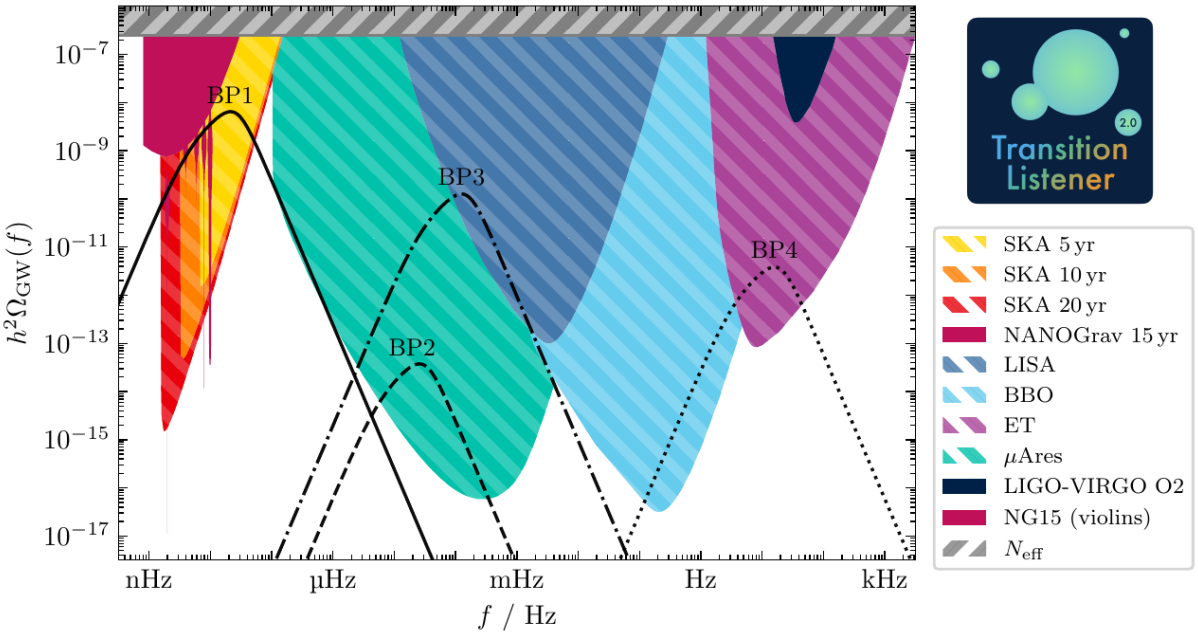


Abbildung 11.4: Gravitationswellenspektren für vier Benchmark-Punkte (BP1–BP4) des TransitionListener-Codes, zusammen mit Sensitivitätskurven aktueller und geplanter Detektoren: SKA (5/10/20 Jahre), NANOGrav 15yr (Violinen = Messung), LISA, BBO, ET, μAres und LIGO-VIRGO O2. Das schraffierte Band am oberen Rand zeigt die N_{eff}-Schranke aus BBN und CMB. BP1 (QCD-Phasenübergang, $T_{\mathrm{reh}} \sim 14\,\mathrm{MeV}$) liegt im nHz-Band von PTAs, BP2 im μHz-Band, BP3 im mHz-Band (LISA), BP4 im Hz-Band (ET/BBO). Abbildung aus Referenz [7].

Aufgabe 11.1

LIGO hat Armlängen von $L = 4\,\mathrm{km}$ und kann Strain-Amplituden von bis zu $h \sim 10^{-21}$ messen.

a) Berechne die zugehörige Längenänderung $\Delta L = h \cdot L$ in Metern. Vergleiche ΔL mit dem Durchmesser eines Protons ($\approx 1.7 \times 10^{-15}\,\mathrm{m}$).

b) Der bei LIGO verwendete Laser hat eine Wellenlänge von $\lambda = 1064\,\mathrm{nm}$. Wie oft passt ΔL aus a) in λ? Was sagt dir das über die Anforderungen an die Messtechnik?

c) Ein Gravitationswellensignal mit Frequenz $f = 100\,\mathrm{Hz}$ hat nach Aufgabe 10.1 eine Wellenlänge $\lambda_{\mathrm{GW}} = 3000\,\mathrm{km}$. Vergleiche diese Wellenlänge mit der LIGO-Armlänge $L = 4\,\mathrm{km}$. Wieso kann LIGO trotzdem ein so „langwelliges“ Signal messen, obwohl $L \ll \lambda_{\mathrm{GW}}$?

d) Der Laser in LIGO wird nicht einfach einmal durch den Arm geschickt, sondern mithilfe von Fabry-Pérot-Resonatoren N-mal pro Richtung hin- und hergespiegelt. Dadurch wird die effektive Armlänge auf $L_{\mathrm{eff}} = N \cdot L$ vergrößert. Dies nennt man den *Quality-Faktor* $Q \approx N$ des Resonators.

- Aus Teil b) wisst ihr, dass die heutige Technologie eine Phasenverschiebung von etwa $\Delta\phi/\lambda \sim 10^{-12}$ messen kann. Um das Signal tatsächlich nachzuweisen, muss die *effektive* Längenänderung $\Delta L_{\text{eff}} = N \cdot \Delta L$ mindestens $\sim 10^{-12} \cdot \lambda$ betragen. Berechne, wie groß N mindestens sein muss.
- Die tatsächliche Finesse (Resonatorgüte) von LIGO beträgt $\mathcal{F} \approx 280$, also $N \approx 280$ effektive Umläufe. Bestätige, dass dieser Wert deine Abschätzung erfüllt.

Musterlösung 11.1

a)

$$\Delta L = h \cdot L = 10^{-21} \cdot 4 \times 10^{3}\,\text{m} = 4 \times 10^{-18}\,\text{m}\,. \tag{11.4}$$

Das ist etwa 1000-mal *kleiner* als der Durchmesser eines Protons (1.7×10^{-15} m)! LIGO misst also Längenänderungen, die winzig im Vergleich zu Atomkernen sind.

b)

$$\frac{\Delta L}{\lambda} = \frac{4 \times 10^{-18}\,\text{m}}{1064 \times 10^{-9}\,\text{m}} \approx 4 \times 10^{-12}\,. \tag{11.5}$$

Die zu messende Längenänderung ist also etwa 10^{-12}-mal kleiner als die Laserwellenlänge selbst. Das verdeutlicht, warum LIGO als eines der präzisesten Messinstrumente der Menschheitsgeschichte gilt: Es muss Phasenverschiebungen des Laserlichts auf einem Niveau von einem Billionstel einer Wellenlänge auflösen.

c) Die LIGO-Armlänge $L = 4\,\text{km}$ ist tatsächlich viel kleiner als die Wellenlänge $\lambda_{\text{GW}} = 3000\,\text{km}$ des Signals. Das ist aber kein Problem: Damit ein Detektor eine Welle messen kann, muss er nicht so groß wie ihre Wellenlänge sein; er muss nur empfindlich genug sein, um die durch die Welle hervorgerufene *Dehnung* $h = \Delta L/L$ zu registrieren, egal wie klein L im Vergleich zu λ_{GW} ist. (Bei Pulsar-Timing-Arrays in Kapitel 17 ist das anders: Dort nutzt man die enormen Distanzen zu Pulsaren als „Armlänge“ L, um Gravitationswellen mit noch viel größeren Wellenlängen $\lambda_{\text{GW}} \sim$ Lichtjahre zu messen.)

d) Für die effektive Längenänderung muss gelten:

$$N \cdot \Delta L \geq 10^{-12} \cdot \lambda = 10^{-12} \cdot 1064 \times 10^{-9}\,\text{m} = 1.064 \times 10^{-18}\,\text{m}\,. \tag{11.6}$$

Mit $\Delta L = 4 \times 10^{-18}\,\text{m}$ aus Teil a) folgt:

$$N \geq \frac{1.064 \times 10^{-18}\,\text{m}}{4 \times 10^{-18}\,\text{m}} \approx 0.27\,. \tag{11.7}$$

Bereits $N = 1$ würde also laut dieser groben Abschätzung reichen – aber das ist täuschend: Die Messkette in LIGO ist deutlich komplexer (Photonenrauschen, Schrotrauschen, seismisches Rauschen etc.), und mit $N \approx 280$ Umläufen durch die Fabry-Pérot-Resonatoren wird die effektive Längenänderung um den Faktor 280 verstärkt:

$$\Delta L_{\text{eff}} = 280 \cdot 4 \times 10^{-18}\,\text{m} \approx 10^{-15}\,\text{m} \approx \text{Protonenradius}\,, \tag{11.8}$$

was das Signal technisch messbar macht. Der Quality-Faktor des Resonators ist also entscheidend für LIGOs Empfindlichkeit – und tatsächlich einer der größten technischen Aufwände beim Bau!

12 Schwarze Löcher

12.1 Was ist ein schwarzes Loch?

Stell dir vor, du wirfst einen Ball senkrecht nach oben. Je mehr Schwung du gibst, desto höher fliegt er, aber irgendwann kommt er immer zurück. Außer wenn er *schnell genug* ist: Die *Fluchtgeschwindigkeit* v_{esc} ist die Mindestgeschwindigkeit, die ein Objekt braucht, um dem Gravitationsfeld eines Körpers der Masse M und Radius R zu entkommen. Aus Energieerhaltung ($\frac{1}{2}mv^2 = \frac{GMm}{R}$) folgt

$$v_{\text{esc}} = \sqrt{\frac{2GM}{R}}\,. \tag{12.1}$$

Was passiert, wenn $v_{\text{esc}} = c$, wenn also selbst Licht nicht mehr entkommen kann? Dann nennt man den entsprechenden Radius den **Schwarzschild-Radius**:

$$r_s = \frac{2GM}{c^2}\,. \tag{12.2}$$

Ein Objekt, das auf einen Radius kleiner als r_s komprimiert wurde, ist ein **schwarzes Loch**: Keine Materie, kein Licht, keine Information kann den Bereich $r < r_s$ verlassen.

Achtung: Newtonsche Herleitung

Die Formel (12.2) stammt formal aus einer newtonschen Rechnung mit $v_{\text{esc}} = c$, aber das Ergebnis ist exakt! Es ist ein glücklicher Zufall, dass die newtonsche Herleitung den richtigen Schwarzschild-Radius liefert. Die volle Begründung stammt aus der Allgemeinen Relativitätstheorie (Schwarzschild-Metrik, 1916).

Zum Vergleich einige Schwarzschild-Radien:

Objekt	Masse	r_s
Erde	$5{,}97 \times 10^{24}\,\text{kg}$	8,87 mm
Sonne	$1{,}99 \times 10^{30}\,\text{kg}$	2,95 km
Stellar-BH	$10\,M_\odot$	29,5 km
Sgr A*	$4 \times 10^6\,M_\odot$	0,08 AU
M87*	$6{,}5 \times 10^9\,M_\odot$	127 AU

Die Erde würde zu einem schwarzen Loch, wenn man sie auf die Größe einer Murmel ($r_s \approx 9\,\text{mm}$) komprimieren könnte.

12.2 Typen schwarzer Löcher

Schwarze Löcher existieren in sehr unterschiedlichen Massenklassen:

Stellar (3–$100\,M_\odot$) Entstehen beim Kollaps massereicher Sterne am Ende ihres Lebens in einer Supernova. Sie sind die Quellen der LIGO/Virgo-Gravitationswellensignale aus Kapitel 11.

Intermediär (10^2–$10^5\,M_\odot$) Eine noch wenig verstandene Klasse; mögliche Entstehungswege umfassen den Kollaps dichter Sternhaufen oder die Verschmelzung vieler stellarer schwarzer Löcher.

Supermassiv (10^6–$10^{10}\,M_\odot$) Sitzen im Zentrum nahezu jeder großen Galaxie. Im Zentrum unserer Milchstraße befindet sich *Sagittarius A** mit $\sim 4 \times 10^6\,M_\odot$; sein Schwarzschild-Radius $r_s \approx 0{,}08\,\text{AU}$ ist kleiner als die Merkurbahn! Das schwarze Loch in der Galaxie M87, *M87**, hat $\sim 6{,}5 \times 10^9\,M_\odot$ mit $r_s \approx 127\,\text{AU}$, also etwa so groß wie das gesamte Sonnensystem!

Pluto umkreist die Sonne bei $\approx 39{,}5\,\mathrm{AU}$; der Durchmesser seiner Bahn beträgt $\approx 79\,\mathrm{AU}$. Voyager 1, die am weitesten von der Erde entfernte Raumsonde, überschritt die *Heliopause* (die äußere Grenze des Sonnensystems) bei $\approx 121\,\mathrm{AU}$; der Schwarzschild-Radius von M87* ist noch größer.

Das *Event Horizon Telescope* (EHT), ein weltweites Netz von Radioteleskopen, das gemeinsam wie ein erdgroßes Teleskop arbeitet, hat 2019 das erste direkte Bild eines schwarzen Lochs aufgenommen [8]: M87*, 55 Millionen Lichtjahre entfernt (Abbildung 12.1). 2022 folgte das erste Bild von Sagittarius A* im Zentrum unserer eigenen Milchstraße. Abbildung 12.2 gibt eine beschriftete Übersicht der Bildmerkmale, die in realen EHT-Aufnahmen sichtbar sind.

Abbildung 12.1: Erstes direktes Bild eines schwarzen Lochs: M87* in der Galaxie M87, aufgenommen vom *Event Horizon Telescope* im April 2019 [8]. Der leuchtende orange-rote Ring besteht aus heißem Plasma der Akkretionsscheibe, dessen Licht durch die extreme Raumzeitkrümmung zu einem Ring gebogen wird. Die dunklere Region in der Mitte ist der *Schatten* des schwarzen Lochs, der Bereich, aus dem kein Licht entkommen kann. Bildnachweis: EHT Collaboration, CC BY 4.0.

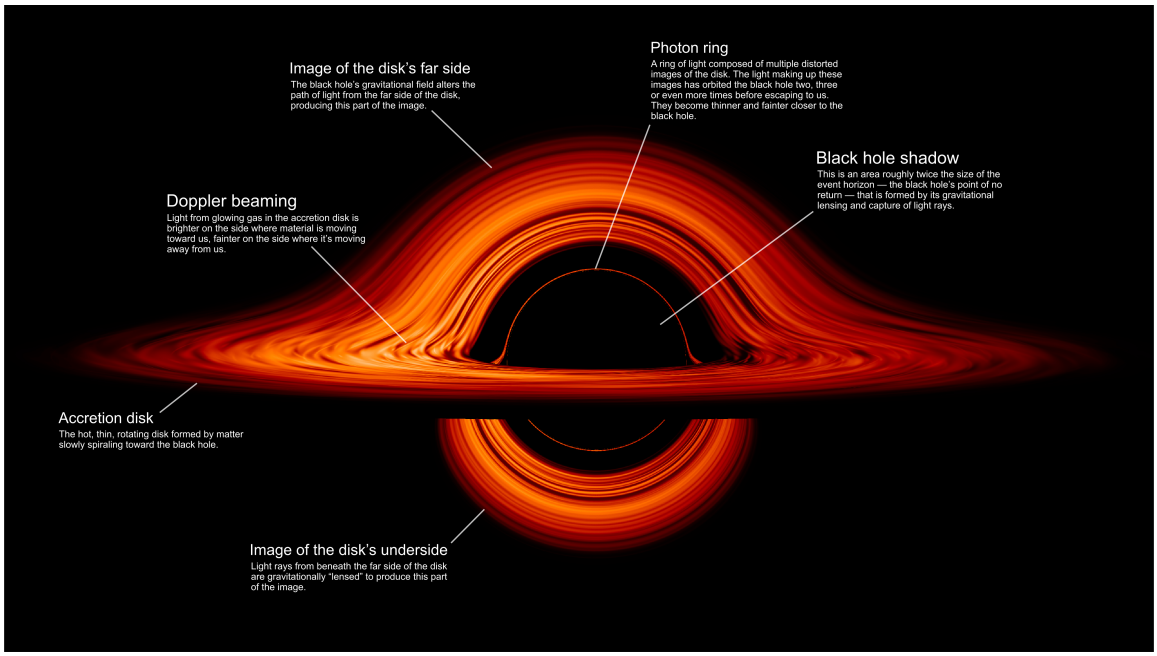


Abbildung 12.2: Beschriftetes Schema der Bildmerkmale eines schwarzen Lochs [9]. Die Simulation zeigt eine Akkretionsscheibe aus heißem Plasma um ein schwarzes Loch. *Doppler-Beaming:* Die Seite der Scheibe, die sich auf uns zubewegt, erscheint heller. *Bild der Rückseite der Scheibe:* Licht vom hinteren Teil wird über den Schatten hinweggebogen und oben sichtbar. *Bild der Unterseite:* Licht von unterhalb der Scheibe wird nach unten abgelenkt. *Photonring:* Licht, das das schwarze Loch zwei- oder dreimal umkreist hat, bevor es entkommt, extrem schmal und hell. *Schwarzer Schatten:* Der Bereich ohne Lichtaustritt, etwa doppelt so groß wie der Ereignishorizont. Bildnachweis: NASA/ESA, L. Calçada.

12.3 Zeitdilatation im Gravitationsfeld

Die Allgemeine Relativitätstheorie sagt voraus, dass Zeit im Gravitationsfeld *langsamer* läuft. Für eine kugelsymmetrische, nicht rotierende Masse M (Schwarzschild-Metrik) gilt für eine ruhende Uhr im Abstand r:

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \sqrt{1 - \frac{r_s}{r}}\,, \tag{12.3}$$

wobei $\mathrm{d}\tau$ die Eigenzeit der Uhr und $\mathrm{d}t$ die Zeit eines weit entfernten Beobachters ist. Je näher an der Masse, desto langsamer läuft die Uhr; am Ereignishorizont ($r = r_s$) gilt $\mathrm{d}\tau/\,\mathrm{d}t = 0$: Die Zeit friert ein.

Aufgabe 12.1

Verwende Gleichung (12.3) und den Schwarzschild-Radius (12.2).
a) Berechne den Schwarzschild-Radius der Erde: $M_\oplus = 5{,}97 \times 10^{24}\,\mathrm{kg}$, $G = 6{,}67 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$.
b) Wie viel langsamer läuft eine Uhr auf der Erdoberfläche ($R_\oplus = 6{,}37 \times 10^6\,\mathrm{m}$) relativ zu einer Uhr weit weg im All? Berechne die Zeitdifferenz pro Tag.
Tipp: Für $x \ll 1$ gilt $\sqrt{1-x} \approx 1 - x/2$.
c) Was passiert nach Gleichung (12.3) am Ereignishorizont ($r \to r_s$)? Interpretiere das Ergebnis für einen außenstehenden Beobachter.
d) (*Bonus*) GPS-Satelliten fliegen in $h = 20.200\,\mathrm{km}$ Höhe über der Erde. Berechne, um wie viel schneller eine GPS-Borduhr pro Tag läuft als eine Uhr auf der Erdoberfläche (nur

Gravitationseffekt). Warum ist dieser Effekt für GPS wichtig?

Musterlösung 12.1

a)

$$r_s = \frac{2GM_\oplus}{c^2} = \frac{2 \times 6{,}67 \times 10^{-11} \times 5{,}97 \times 10^{24}}{(3 \times 10^8)^2} \approx 8{,}87 \times 10^{-3}\,\mathrm{m} \approx 8{,}9\,\mathrm{mm}\,. \tag{12.4}$$

Die gesamte Erde müsste auf die Größe einer Murmel zusammengepresst werden.
b) Mit $r_s/R_\oplus = 8{,}87 \times 10^{-3}/6{,}37 \times 10^6 \approx 1{,}39 \times 10^{-9} \ll 1$ und $\sqrt{1-x} \approx 1 - x/2$:

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} \approx 1 - \frac{r_s}{2R_\oplus} \approx 1 - 6{,}97 \times 10^{-10}\,. \tag{12.5}$$

Pro Tag ($\Delta t = 86.400\,\mathrm{s}$) läuft die Uhr auf der Erdoberfläche um

$$\Delta\tau \approx 6{,}97 \times 10^{-10} \times 86.400\,\mathrm{s} \approx 60\,\mu\mathrm{s} \tag{12.6}$$

langsamer als eine Uhr weit weg im Weltraum. Pro Jahr akkumulieren sich so $\approx 22\,\mathrm{ms}$.
c) Am Ereignishorizont gilt $r_s/r = 1$, also $\mathrm{d}\tau/\,\mathrm{d}t = 0$: Aus Sicht eines weit entfernten Beobachters scheint alles, was sich dem Horizont nähert, immer langsamer zu fallen. Das einfallende Objekt wird zunehmend rotverschoben und kommt scheinbar *niemals* am Horizont an; es friert für den Außenstehenden ein.
d) Für GPS-Satelliten bei $r = R_\oplus + h = 6{,}37 \times 10^6 + 2{,}02 \times 10^7 = 2{,}66 \times 10^7\,\mathrm{m}$:

$$\begin{aligned}\Delta\left(\frac{\mathrm{d}\tau}{\mathrm{d}t}\right) &= \frac{r_s}{2}\left(\frac{1}{R_\oplus} - \frac{1}{r}\right) \\ &= \frac{8{,}87 \times 10^{-3}}{2}\left(\frac{1}{6{,}37 \times 10^6} - \frac{1}{2{,}66 \times 10^7}\right) \approx 5{,}3 \times 10^{-10}\,.\end{aligned} \tag{12.7}$$

Pro Tag: $5{,}3 \times 10^{-10} \times 86.400\,\mathrm{s} \approx 45{,}8\,\mu\mathrm{s}$. GPS-Borduhren laufen gravitativ $\approx 45{,}8\,\mu\mathrm{s}$ pro Tag *schneller* als Uhren auf der Erde (sie befinden sich weiter vom Erdzentrum). Da GPS auf Nanosekunden-Präzision angewiesen ist, muss dieser Effekt in der Satellitensoftware korrigiert werden: Die Allgemeine Relativitätstheorie ist hier kein akademisches Spielzeug, sondern eine Ingenieurnotwendigkeit.

12.4 Ein Fall ins schwarze Loch

Was passiert einem Astronauten, der in ein schwarzes Loch fällt? Die Antwort hängt vom Standpunkt ab.

Sicht des außenstehenden Beobachters. Von außen scheint der Astronaut immer langsamer zu fallen (Gleichung (12.3)): Er friert ein, wird zugleich gravitativ rotverschoben und immer schwächer, bis er praktisch unsichtbar wird. Aus dieser Perspektive überquert er den Horizont scheinbar *nie*.

Sicht des fallenden Astronauten. Für den Astronauten selbst vergeht seine Zeit normal. Er überquert den Ereignishorizont in *endlicher Eigenzeit*, ohne dabei etwas Besonderes zu spüren; der Horizont ist kein physikalisches Hindernis, keine Wand, kein Vorhang. *Am Ereignishorizont selbst passiert lokal nichts Besonderes.* (Das gilt insbesondere für supermassive schwarze Löcher, bei denen die Gezeitenkräfte am Horizont extrem gering sind.)

Spaghettifizierung. Bei kompakten, stellaren schwarzen Löchern wirken schon weit vor dem Horizont starke *Gezeitenkräfte*: Da die Gravitationskraft mit $1/r^2$ abfällt, ist die Anziehung an

den Füßen stärker als am Kopf. Die Gezeitenkraft ΔF ist die Differenz der Gravitationskraft zwischen zwei Punkten des Körpers, die im Abstand ℓ (z. B. der Körpergröße des Astronauten) radial voneinander entfernt sind:

$$\Delta F = \frac{GMm}{r^2} - \frac{GMm}{(r+\ell)^2} \approx \frac{2GMm\ell}{r^3} \quad (\ell \ll r)\,. \tag{12.8}$$

Am Ereignishorizont ($r = r_s = 2GM/c^2$) wird daraus

$$\Delta F\big|_{r=r_s} = \frac{2GMm\ell}{r_s^3} = \frac{2GMm\ell}{(2GM/c^2)^3} = \frac{m\ell\, c^6}{4\,G^2M^2} \propto \frac{1}{M^2}\,. \tag{12.9}$$

Die Gezeitenkraft am Horizont *sinkt* also mit wachsender Masse, ein kontraintuitives Ergebnis. Für konkrete Zahlen nehmen wir $m = 80\,\mathrm{kg}$ und $\ell = 1{,}8\,\mathrm{m}$:

$$\begin{aligned}\Delta F\big|_{r=r_s} &= \frac{80\,\mathrm{kg} \times 1{,}8\,\mathrm{m} \times (3\times10^8\,\mathrm{m/s})^6}{4 \times (6{,}67\times10^{-11})^2\,\mathrm{m^6kg^{-2}s^{-4}} \times M^2} \\ &\approx \frac{1{,}05\times10^{53}\,\mathrm{N\cdot kg^2}}{M^2}\,.\end{aligned} \tag{12.10}$$

Einsetzen für verschiedene schwarze Löcher:

Objekt	Masse M	ΔF am Horizont
Stellar-BH	$10\,M_\odot = 2\times10^{31}\,\mathrm{kg}$	$\approx 1{,}5\times10^{10}\,\mathrm{N}$
Sgr A*	$4\times10^6\,M_\odot = 8\times10^{36}\,\mathrm{kg}$	$\approx 0{,}09\,\mathrm{N}$
M87*	$6{,}5\times10^9\,M_\odot = 1{,}3\times10^{40}\,\mathrm{kg}$	$\approx 35\,\mathrm{nN}$

Ein stellares schwarzes Loch zerreißt den Astronauten mit $\sim 10^{10}\,\mathrm{N}$, also etwa das 10^7-fache des eigenen Körpergewichts ($\approx 800\,\mathrm{N}$ auf der Erde), und das passiert schon *lange* vor dem Erreichen des Horizonts. Bei Sgr A* entspräche die Gezeitenkraft etwa dem Gewicht eines Apfels, spürbar, aber harmlos. Bei M87* ist sie mit 35 nN vollkommen vernachlässigbar: Man könnte den Ereignishorizont problemlos überqueren, ohne etwas zu bemerken.

Dieser Unterschied dehnt den Körper radial und quetscht ihn senkrecht dazu; der Astronaut wird zu einem langen, dünnen Faden (**Spaghettifizierung**).

Optische Effekte. Die starke Raumzeitkrümmung lenkt Licht dramatisch um das schwarze Loch herum (*Gravitationslinseneffekt*). Ein fallender Astronaut würde sehen, wie das gesamte Universum hinter ihm zu einem engen *Einstein-Ring* zusammengepresst wird, während der Blick nach vorne in die Dunkelheit des Horizonts fällt. Nach der Horizontpassage kann er nichts mehr nach außen schicken; alle Zukunftsrichtungen im Raumzeit-Diagramm zeigen auf die Singularität zu, die er nicht vermeiden kann.

13 Einführung in die Wahrscheinlichkeitstheorie

13.1 Modelle, Daten und die Rolle der Stochastik

Physik lebt vom Wechselspiel zweier Richtungen: Auf der einen Seite steht ein **Modell** unserer Welt (ein physikalisches Gesetz mit einer Handvoll Parametern), auf der anderen Seite stehen **Daten**: Messwerte aus dem Labor oder von Teleskopen. Wie hängen die beiden zusammen? Genau diese Frage beantwortet die **Stochastik**, die Lehre vom Rechnen mit zufällig erscheinenden Phänomenen. Sie zerfällt in zwei komplementäre Disziplinen, siehe Abbildung 13.1:

- Die **Wahrscheinlichkeitstheorie** denkt *vorwärts*: Ist das Modell gegeben, sagt sie vorher, welche Daten wir mit welcher Wahrscheinlichkeit beobachten sollten.
- Die **Statistik** denkt *rückwärts*: Sind die Daten gegeben, versucht sie herauszufinden, welches Modell – bzw. welche Parameter eines Modells – am besten zu ihnen passt.

Abbildung 13.1: Modell und Daten stehen über die Stochastik in Verbindung: Die Wahrscheinlichkeitstheorie schließt vorwärts vom Modell auf mögliche Daten, die Statistik rückwärts von den Daten auf das zugrundeliegende Modell. Dass die „Daten“-Kopie rechts deutlich unordentlicher ausfällt als das „Modell“ links, ist Absicht: Ein Modell ist ein sauberes, idealisiertes Abbild der Welt, während echte Messdaten immer mit Rauschen, Unsicherheiten und Unvollkommenheiten behaftet sind.

Um diesen Rückschluss (Statistik) überhaupt führen zu können, müssen wir zunächst verstehen, was eine Wahrscheinlichkeit überhaupt *ist* und welche Rechenregeln für sie gelten. Das ist das Ziel dieses Kapitels; in Kapitel 14 bauen wir darauf auf.

13.2 Ein unfairer Würfel

Als Einstieg betrachten wir einen sechsseitigen Würfel (Abbildung 13.2), von dem wir aus vielen vorherigen Würfen folgende Beobachtungen kennen:

- Eine 1, 2 oder 3 wird in sieben von zehn Würfen geworfen.
- Die Augenzahlen 4, 5 und 6 werden alle gleich wahrscheinlich geworfen.
- Eine 1 oder 2 wird in sechs von zehn Würfen geworfen.
- Eine 1 wird doppelt so oft geworfen wie eine 2.

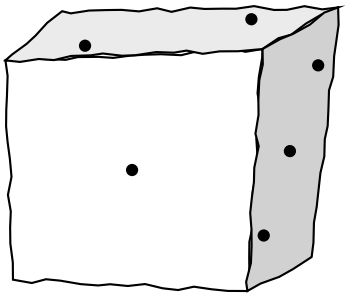


Abbildung 13.2: Ein *unfairer* Würfel: Schon die schiefen, ungleichmäßigen Seitenflächen verraten, dass hier nicht alles mit rechten Dingen zugeht. Sichtbar sind die Augenzahlen 1 (vorne), 2 (oben) und 3 (rechts) – genau die Gruppe, die laut Beobachtung in 7 von 10 Würfen fällt. Die Rückseiten 4, 5 und 6 sind dagegen gleichverteilt.

Aufgabe 13.1

Wie modellieren wir einen solchen Würfel mathematisch?

Es hilft, nicht über einzelne Würfe, sondern über *Ereignisse* nachzudenken – also über Teilmengen der möglichen Ausgänge wie $\{1, 2, 3\}$ oder $\{4\}$.

13.3 Wahrscheinlichkeitsräume

Definition 13.1

Ein **Wahrscheinlichkeitsraum** ist ein Tripel (Ω, Σ, P) mit:

- der **Grundmenge** Ω, der Menge aller möglichen Ausgänge;
- dem **Ereignissystem** Σ, einer Menge von Teilmengen von Ω (den *Ereignissen*); im diskreten Fall wählt man meist die volle **Potenzmenge** $\Sigma = \mathcal{P}(\Omega) := \{U \mid U \subset \Omega\}$;
- dem **Wahrscheinlichkeitsmaß** $P : \Sigma \to [0, 1]$, das jedem Ereignis eine Wahrscheinlichkeit zuordnet und die folgenden (Kolmogorov-)Axiome erfüllt:

$$P(E) \geq 0 \quad \forall E \in \Sigma\,, \tag{13.1}$$

$$P(\Omega) = 1\,, \tag{13.2}$$

$$P(A \cup B) = P(A) + P(B) \quad \text{für } A, B \in \Sigma \text{ mit } A \cap B = \emptyset\,. \tag{13.3}$$

Musterlösung 13.1

Für unseren Würfel ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $\Sigma = \mathcal{P}(\Omega)$. Aus den Beobachtungen oben wissen wir:

$$P(\{1,2,3\}) = \frac{7}{10}\,, \qquad P(\{4\}) = P(\{5\}) = P(\{6\})\,, \tag{13.4}$$

$$P(\{1,2\}) = \frac{6}{10}\,, \qquad P(\{1\}) = 2\,P(\{2\})\,. \tag{13.5}$$

Mit Axiom (13.3) (angewandt auf die disjunkten Ereignisse $\{4\}, \{5\}, \{6\}$) und (13.2) folgt $P(\{4\}) + P(\{5\}) + P(\{6\}) = 1 - \frac{7}{10} = \frac{3}{10}$, also $P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{10}$.

Da $\{1,2,3\} = \{1,2\} \cup \{3\}$ eine disjunkte Vereinigung ist, liefert Axiom (13.3) außerdem

$$P(\{3\}) = P(\{1,2,3\}) - P(\{1,2\}) = \frac{7}{10} - \frac{6}{10} = \frac{1}{10}\,. \tag{13.6}$$

Ebenso ist $\{1,2\} = \{1\} \cup \{2\}$ disjunkt, also $P(\{1\}) + P(\{2\}) = P(\{1,2\}) = \frac{6}{10}$. Setzt man $P(\{1\}) = 2P(\{2\})$ ein, folgt $3P(\{2\}) = \frac{6}{10}$, also

$$P(\{2\}) = \frac{2}{10} = \frac{1}{5}\,, \qquad P(\{1\}) = 2P(\{2\}) = \frac{4}{10} = \frac{2}{5}\,. \tag{13.7}$$

Damit ist unser unfairer Würfel bereits vollständig modelliert – ganz ohne den Begriff der bedingten Wahrscheinlichkeit, den wir erst in Abschnitt 13.4 einführen und direkt an diesem Würfel illustrieren:

Augenzahl k	1	2	3	4	5	6
$P(\{k\})$	0.4	0.2	0.1	0.1	0.1	0.1

(Probe: Die sechs Werte summieren sich, wie es Axiom (13.2) verlangt, zu 1.)

13.4 Bedingte Wahrscheinlichkeit und der Satz von Bayes

Definition 13.2

Für zwei Ereignisse $A, B \in \Sigma$ mit $P(B) > 0$ ist die **bedingte Wahrscheinlichkeit** von A gegeben B definiert als

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}\,. \tag{13.8}$$

Als Beispiel nutzen wir unseren inzwischen (Abschnitt 13.3) vollständig bekannten Würfel: Wie wahrscheinlich ist eine 1, wenn wir schon wissen, dass eine 1, 2 oder 3 gewürfelt wurde? Mit Gleichung (13.8) und $\{1\} \cap \{1,2,3\} = \{1\}$ folgt

$$P(\{1\} \mid \{1,2,3\}) = \frac{P(\{1\})}{P(\{1,2,3\})} = \frac{0.4}{0.7} = \frac{4}{7}\,. \tag{13.9}$$

Diese bedingte Wahrscheinlichkeit ist *größer* als die unbedingte $P(\{1\}) = 0.4$: Das Wissen, dass eine 1, 2 oder 3 gefallen ist, schließt die Möglichkeiten $4, 5, 6$ (zusammen 0.3 Wahrscheinlichkeit) aus und macht eine 1 dadurch relativ wahrscheinlicher. Genauso findet man $P(\{2\} \mid \{1,2,3\}) = 0.2/0.7 = 2/7$ und $P(\{3\} \mid \{1,2,3\}) = 0.1/0.7 = 1/7$ – die drei bedingten Wahrscheinlichkeiten summieren sich korrekt zu 1, wie es sein muss, wenn wir uns auf das Ereignis $\{1,2,3\}$ beschränken.

13.5 Aufgaben zur bedingten Wahrscheinlichkeit und dem Satz von Bayes

Aufgabe 13.2

a) Zeige, dass für zwei Ereignisse $A, B \in \Sigma$

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{13.10}$$

gilt. Diese Beziehung heißt Satz von Bayes und ist wichtig in unserem Kurs.

b) Findet eine intuitive Erklärung für den Satz von Bayes, die eure Eltern verstehen

würden.

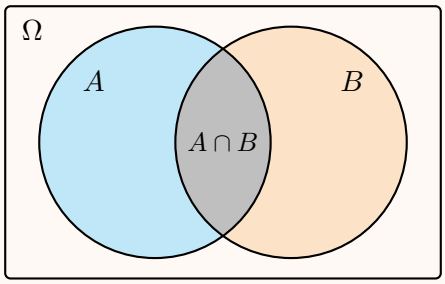
c) Sei $A \in \Sigma$ und A^c das *Gegenereignis* zu A, d.h. die Menge $A^c \in \Sigma$ mit $A \cap A^c = \emptyset$ und $A \cup A^c = \Omega$. Zeige, dass für ein beliebiges Ereignis $B \in \Sigma$

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid A^c) \cdot P(A^c) \tag{13.11}$$

gilt (man spricht auch vom *Gesetz der totalen Wahrscheinlichkeit*). Versucht, auch diesen Zusammenhang intuitiv zu verstehen.

Musterlösung 13.2

Zur Veranschaulichung der beteiligten Mengen:



a) Für $P(A), P(B) > 0$ liefert Definition 13.2 zwei Ausdrücke für dieselbe Schnittmenge (die grau schattierte Fläche $A \cap B$ oben):

$$P(A \cap B) = P(A \mid B)\, P(B)\,, \qquad P(A \cap B) = P(B \mid A)\, P(A)\,, \tag{13.12}$$

wobei wir im zweiten Schritt $A \cap B = B \cap A$ benutzt haben. Gleichsetzen und Division durch $P(B)$ liefert die Behauptung.
b) Eine mögliche Formulierung: Der Satz von Bayes sagt, wie du eine Vermutung über eine *Ursache* anpassen solltest, nachdem du eine *Beobachtung* gemacht hast. Dabei kombinierst du zwei Dinge: wie plausibel die Ursache *vorher* schon war (Prior $P(A)$) und wie gut die Beobachtung zu dieser Ursache passt (Likelihood $P(B \mid A)$). Beispiel: Du kommst morgens runter und der Rasen ist nass (Beobachtung B). Hat es geregnet (Ursache A) oder lief der Rasensprenger? Wenn es in deiner Gegend selten regnet ($P(A)$ klein), aber Regen fast immer nassen Rasen bedeutet ($P(B \mid A)$ groß), reicht die nasse Wiese allein noch nicht, um sicher auf Regen zu schließen – du musst beides gegeneinander abwägen. Genau das tut der Satz von Bayes rechnerisch.
c) Da $A \cup A^c = \Omega$, gilt für jedes $B \in \Sigma$: $B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$. Diese Vereinigung ist *disjunkt*, denn $(B \cap A) \cap (B \cap A^c) \subset A \cap A^c = \emptyset$. Mit dem dritten Kolmogorov-Axiom (13.3) folgt

$$P(B) = P(B \cap A) + P(B \cap A^c)\,. \tag{13.13}$$

Wendet man auf beide Summanden die Definition der bedingten Wahrscheinlichkeit an ($P(B \cap A) = P(B \mid A)P(A)$ und analog für A^c), folgt die Behauptung. Intuitiv: Jedes Ereignis B tritt entweder zusammen mit A oder zusammen mit A^c ein – ein Drittes gibt es nicht. Die Gesamtwahrscheinlichkeit von B ist deshalb der gewichtete Mittelwert seiner Wahrscheinlichkeit in diesen beiden Fällen, gewichtet mit der Wahrscheinlichkeit des jeweiligen Falls.

Der Satz von Bayes ist einer der folgenreichsten Sätze der Wahrscheinlichkeitstheorie:

Satz von Bayes

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}\,. \tag{13.14}$$

Eine bedingte Wahrscheinlichkeit lässt sich damit „umdrehen“: Aus $P(B \mid A)$ wird $P(A \mid B)$. In der Sprache der Statistik nennt man $P(A)$ den **Prior** (das Vorwissen über A, bevor wir B beobachtet haben), $P(B \mid A)$ die **Likelihood** (wie wahrscheinlich die Beobachtung B unter der Annahme A ist), $P(A \mid B)$ den **Posterior** (unser aktualisiertes Wissen über A, nachdem wir B beobachtet haben) und $P(B)$ die **Evidenz** (eine Normierungskonstante).

Diesen Begriffen begegnen wir in Kapitel 16 wieder, wenn wir Bayessche Inferenz an einem konkreten physikalischen System durchführen.

Aufgabe 13.3

Ein medizinischer Test wird eingesetzt, um eine seltene Krankheit in der Bevölkerung zu erkennen. Die Krankheit tritt bei 0.2 % der Bevölkerung auf. Der Test erkennt erkrankte Personen mit einer Wahrscheinlichkeit von 99 %. Gesunde Personen werden mit einer Wahrscheinlichkeit von 2 % fälschlicherweise als krank markiert.

a) Definiere einen geeigneten Wahrscheinlichkeitsraum für das „Zufallsexperiment“, dass eine Person auf die Krankheit getestet wird. Lege dazu Grundmenge sowie geeignete Ereignisse für den Gesundheitszustand und das Testergebnis fest und gib die bekannten Wahrscheinlichkeiten an.

b) Die Person wird positiv getestet. Berechne die Wahrscheinlichkeit, dass diese Person tatsächlich an der Krankheit leidet.

Hinweis: Beide Gleichungen aus Aufgabe 13.2 sind für die Berechnung hilfreich.

Musterlösung 13.3

a) Grundmenge $\Omega = \{(\text{krank}, +), (\text{krank}, -), (\text{gesund}, +), (\text{gesund}, -)\}$, mit $\Sigma = \mathcal{P}(\Omega)$. Wir betrachten die Ereignisse

$$K = \text{„Person ist krank“}\,, \qquad T = \text{„Test ist positiv“}\,, \tag{13.15}$$

mit Gegenereignissen K^c (gesund) und T^c (negativ). Gegeben sind:

$$P(K) = 0.002\,, \qquad P(T \mid K) = 0.99\,, \qquad P(T \mid K^c) = 0.02\,. \tag{13.16}$$

b) Gesucht ist $P(K \mid T)$. Mit dem Satz von Bayes (Aufgabe 13.2a):

$$P(K \mid T) = \frac{P(T \mid K)\, P(K)}{P(T)}\,. \tag{13.17}$$

Den Nenner $P(T)$ erhalten wir mit dem Gesetz der totalen Wahrscheinlichkeit (Aufgabe 13.2c), angewandt auf die Partition $\{K, K^c\}$:

$$P(T) = P(T \mid K)\, P(K) + P(T \mid K^c)\, P(K^c) = 0.99 \cdot 0.002 + 0.02 \cdot 0.998 \tag{13.18}$$

$$= 0.00198 + 0.01996 = 0.02194\,. \tag{13.19}$$

Damit

$$P(K \mid T) = \frac{0.00198}{0.02194} \approx 0.090\,, \tag{13.20}$$

also nur etwa 9 %! Obwohl der Test zu 99 % zuverlässig ist, liegt bei einem positiven Ergebnis in neun von zehn Fällen *keine* Erkrankung vor. Der Grund: Die Krankheit ist so selten ($P(K) = 0.2\,\%$), dass die (wenigen) falsch-positiven Ergebnisse unter den vielen Gesunden die (wenigen) richtig-positiven Ergebnisse unter den wenigen Kranken zahlenmäßig überwiegen. Diese „Basisraten-Falle“ ist einer der wichtigsten Gründe, warum man Testergebnisse nie isoliert, sondern immer zusammen mit der Prävalenz $P(K)$ interpretieren sollte.

Aufgabe 13.4

In einer Spielshow stehen drei verschlossene Türen zur Auswahl. Hinter einer Tür befindet sich ein Mittel für Weltfrieden, hinter den beiden anderen jeweils ein feucht-schwitziger Händedruck. Eine Kandidatin wählt zunächst zufällig eine Tür. Anschließend öffnet der Moderator eine der beiden übrigen Türen, hinter der sich sicher der feucht-schwitzige Händedruck befindet. Danach bietet er der Kandidatin an, bei ihrer ursprünglichen Wahl zu bleiben oder zur anderen, noch geschlossenen Tür zu wechseln.

a) Definiere einen geeigneten Wahrscheinlichkeitsraum. Beschreibe die Grundmenge und geeignete Ereignisse, die den Ort des Wundermittels und die erste Wahl der Kandidatin berücksichtigen.

b) Sollte die Kandidatin die Tür wechseln oder bei ihrer ursprünglichen Wahl bleiben? Berechne die jeweiligen Wahrscheinlichkeiten.

Musterlösung 13.4

a) Sei $D \in \{1,2,3\}$ die Tür mit dem Wundermittel und $C \in \{1,2,3\}$ die erste Wahl der Kandidatin. Beide sind unabhängig und gleichverteilt, also

$$\Omega = \{1,2,3\} \times \{1,2,3\}\,, \qquad P((d,c)) = \frac{1}{9} \quad \forall (d,c) \in \Omega\,. \tag{13.21}$$

Das Ereignis „Treffer“ $T = \{(d,c) \in \Omega \mid d = c\}$ (die erste Wahl ist bereits richtig) hat 3 der 9 gleichwahrscheinlichen Elementarereignisse, also $P(T) = \frac{1}{3}$ und $P(T^c) = \frac{2}{3}$.
b) Der Moderator öffnet *immer* eine Tür mit Händedruck, die weder die Preistür D noch die gewählte Tür C ist – das ist ihm garantiert möglich, egal was D und C sind. Entscheidend ist: Ob Wechseln zum Gewinn führt, hängt nur davon ab, ob die *erste* Wahl schon richtig war.

- Tritt T ein (Wahrscheinlichkeit $\frac{1}{3}$): Die erste Wahl war schon die Preistür. Die einzige noch geschlossene Tür nach dem Öffnen des Moderators enthält den Händedruck – Wechseln verliert, Bleiben gewinnt.
- Tritt T^c ein (Wahrscheinlichkeit $\frac{2}{3}$): Die erste Wahl war falsch. Der Moderator kann dann gar nicht anders, als die *einzige* verbleibende Tür mit Händedruck zu öffnen – die Preistür bleibt zwangsläufig geschlossen. Wechseln gewinnt, Bleiben verliert.

Also gilt $P(\text{Gewinn beim Wechseln}) = P(T^c) = \frac{2}{3}$ und $P(\text{Gewinn beim Bleiben}) = P(T) = \frac{1}{3}$. Die Kandidatin sollte **wechseln**: Das verdoppelt ihre Gewinnchance von 33 % auf 67 %.

Intuitiv: Die erste Wahl ist zu $\frac{2}{3}$ falsch; durch das gezielte Öffnen einer Verliererтür „bündelt“ der Moderator diese $\frac{2}{3}$ Wahrscheinlichkeit auf die verbleibende Tür.

13.6 Häufige diskrete Wahrscheinlichkeitsverteilungen

Neben unserem Würfel mit endlich vielen Ausgängen gibt es auch für diskrete Zufallsvariablen mit *unbeschränkt* vielen möglichen Werten (etwa eine Anzahl $k \in \{0, 1, 2, \ldots\}$) etablierte Standardverteilungen. Zwei der wichtigsten sind die Poisson- und die Binomialverteilung.

13.6.1 Binomialverteilung

Bei n unabhängigen Versuchen mit Erfolgswahrscheinlichkeit p folgt die Anzahl der Erfolge k der **Binomialverteilung**:

$$B(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \langle k \rangle = np, \quad \sigma_k^2 = np(1-p)\,. \tag{13.22}$$

Sie beschreibt z.B. die Anzahl der Erfolge bei wiederholten Münzwürfen oder allgemein bei Experimenten mit nur zwei möglichen Ausgängen („Erfolg“ oder „Misserfolg“); wir begegnen ihr in Aufgabe 13.6 bei einer unfairen Münze wieder. Dabei ist das Symbol $\binom{n}{k} = n!/(k!(n-k)!)$ der sog. Binomialkoeffizient. Die Schreibweise $n! = n \times (n-1) \times \cdots \times 1$ steht für die sog. Fakultät von n.

13.6.2 Poissonverteilung

Im Grenzfall $n \to \infty$, $p \to 0$ bei festem $\lambda = np$ geht die Binomialverteilung in die **Poissonverteilung** über:

$$P(k; \lambda) = \frac{\lambda^k}{k!}\, \mathrm{e}^{-\lambda}, \quad \langle k \rangle = \lambda, \quad \sigma_k^2 = \lambda\,. \tag{13.23}$$

Sie beschreibt seltene Ereignisse mit konstanter Rate (z.B. radioaktiver Zerfall, Photonendetektion) und hat die bemerkenswerte Eigenschaft $\sigma_k = \sqrt{\lambda}$: die relative Unsicherheit $\sigma_k/\langle k \rangle = 1/\sqrt{\lambda}$ geht für viele Ereignisse gegen null. Abbildung 13.3 zeigt beide Verteilungen im Vergleich.

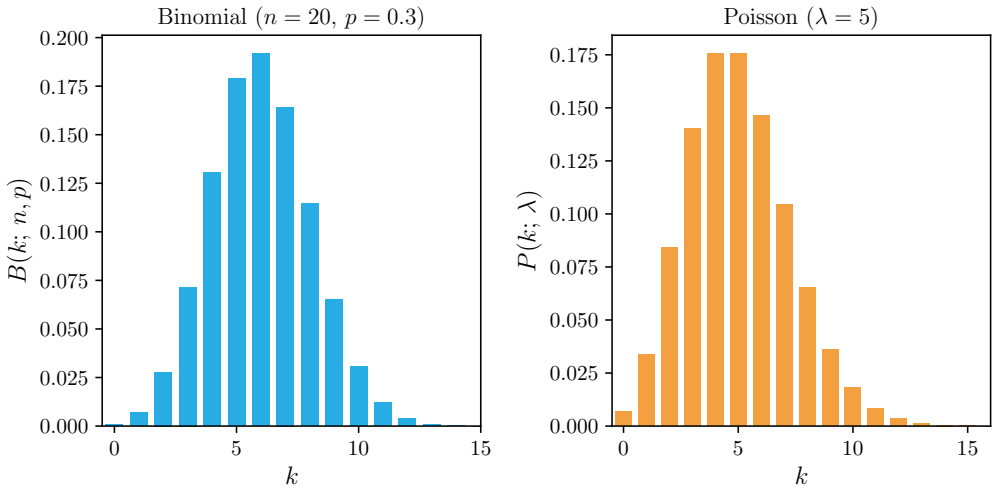


Abbildung 13.3: *Links:* Binomialverteilung $B(k; n = 20, p = 0.3)$. *Rechts:* Poissonverteilung $P(k; \lambda = 5)$ – ihr Grenzfall für seltene Ereignisse mit konstanter Rate.

Bevor wir die Poissonverteilung an einem simulierten Datensatz überprüfen, brauchen wir zwei Kennzahlen, mit denen man einen Datensatz $\{x_1, \ldots, x_N\}$ zusammenfasst: den **Stichprobenmittelwert**

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{13.24}$$

und die **Stichprobenstandardabweichung**

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}\,. \tag{13.25}$$

Beide schätzen aus endlich vielen Beobachtungen den (unbekannten) Erwartungswert $\langle x \rangle$ bzw. die Streuung σ_x der zugrundeliegenden Verteilung – mehr dazu in Kapitel 14.

13.7 Statistische Modelle

Bisher hatten wir ein Wahrscheinlichkeitsmaß P, das vollständig feststand. In der Praxis kennen wir P aber selten exakt – wir kennen höchstens seine *Form*, mit ein paar offenen Parametern, die wir aus Daten bestimmen wollen. Das führt zum zentralen Objekt der Statistik:

Definition 13.3

Ein Tripel $(\Omega, \Sigma, (P_\theta)_{\theta \in \Theta})$ heißt **statistisches Modell**, wobei Θ der Raum aller möglichen Parameter ist und P_θ für jedes $\theta \in \Theta$ ein Wahrscheinlichkeitsmaß auf (Ω, Σ) ist.

Beispiele:

i) Eine Münze mit $P_p(\text{„Kopf“}) = p$: Grundmenge $\Omega = \{\text{Kopf}, \text{Zahl}\}$, Parameterraum $\Theta = [0, 1]$. Für $p = \frac{1}{2}$ ist die Münze fair, sonst nicht.

ii) Die Anzahl k der Photonen, die ein Detektor in einem festen Zeitintervall registriert, folgt für viele unabhängige, seltene Ereignisse einer Poissonverteilung $P_\lambda(k)$ mit unbekannter Rate $\lambda > 0$ (Parameterraum $\Theta = (0, \infty)$), vgl. Abschnitt 13.6.

iii) Die Körpergröße x der Teilnehmenden in diesem Raum folgt näherungsweise einer Normalverteilung $P_{(\mu,\sigma)}(x) = \mathcal{N}(x;\, \mu, \sigma^2)$ (vgl. Abschnitt 13.9) mit unbekanntem Mittelwert μ und unbekannter Streuung σ. Hier ist $\theta = (\mu, \sigma)$ selbst schon ein Parameter*vektor* und $\Theta = \mathbb{R} \times (0, \infty)$ entsprechend zweidimensional – Θ muss also kein einfaches Intervall sein, sondern kann jede beliebige Dimension haben.

Das Ziel der Statistik ist es nun, ausgehend von beobachteten Daten $x_1, \ldots, x_N$ auf den „richtigen“ (oder zumindest plausibelsten) Parameter θ zurückzuschließen – also den Pfeil in Abbildung 13.1 rückwärts zu gehen. Mit den Werkzeugen dieses Kurses im Gepäck wenden wir uns dieser Aufgabe in Kapitel 14 zu.

13.8 Aufgaben zu Verteilungen und statistischen Modellen

Aufgabe 13.5

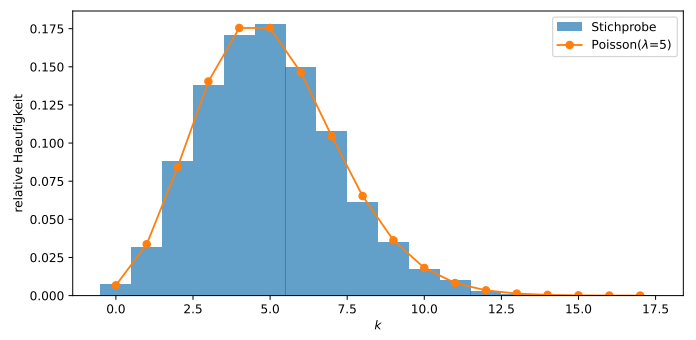
In dieser Aufgabe erkunden wir die Poissonverteilung mit Python. Du darfst in dieser Aufgabe (und auch allen folgenden Aufgaben) natürlich gerne auch das Internet zu Rate ziehen, wenn du noch Fragen hast, wie man in `Python` programmiert. (Simon und Carlo machen das auch so.)

a) Erzeuge mit `numpy.random.poisson(λ=5, size=10000)` einen Datensatz von 10 000 Poisson-Zufallszahlen. Plotte das Histogramm (normiert, `density=True`) und vergleiche es mit der theoretischen Poisson-Verteilung $P(k;\lambda)$ aus dem Text.

b) Berechne für diesen Datensatz den Stichprobenmittelwert $\bar{k}$ und die Stichprobenstandardabweichung s. Vergleiche mit den theoretischen Werten $\langle k\rangle = \lambda$ und $\sigma_k = \sqrt{\lambda}$.

Musterlösung 13.5

```
import numpy as np
import matplotlib.pyplot as plt
from math import factorial, exp

def poisson_wahrscheinlichkeit(k, lam):
    return lam**k / factorial(k) * exp(-lam)

# Teil a): Poisson-Histogramm
lam = 5
data = np.random.poisson(lam, size=10000)
k = np.arange(0, data.max() + 1)
plt.figure(figsize=(8, 4))
plt.hist(data, bins=np.arange(-0.5, data.max() + 1.5, 1), density=True,
         label='Stichprobe', alpha=0.7)
plt.plot(k, [poisson_wahrscheinlichkeit(ki, lam) for ki in k], 'o-',
         label=rf'Poisson($\lambda$={lam})')
plt.xlabel('$k$'); plt.ylabel('relative Haeufigkeit')
plt.legend(); plt.tight_layout(); plt.show()

# Teil b): Kenngroessen
def stichprobenmittelwert(x):
    return sum(x) / len(x)

def stichprobenstandardabweichung(x):
    xbar = stichprobenmittelwert(x)
    return (sum((xi - xbar)**2 for xi in x) / len(x))**0.5

# (np.mean(data) und np.std(data) wuerden hier genauso funktionieren.)
print(f'Mittelwert: {stichprobenmittelwert(data):.3f}  (theoretisch: {lam})')
print(f'Std:        {stichprobenstandardabweichung(data):.3f}  (theoretisch:
↪  {np.sqrt(lam):.3f})')
```



Aufgabe 13.6

Wir haben eine exotische Münze, die mit Wahrscheinlichkeit p Kopf zeigt. Die Münze wird 14-mal geworfen; dabei werden 10-mal Kopf und 4-mal Zahl beobachtet (Daten D). Wir interessieren uns für die Wahrscheinlichkeit, dass die nächsten beiden Würfe Kopf zeigen.

a) Beschreibe das statistische Modell für dieses Experiment.

b) Der sogenannte **Maximum-Likelihood-Schätzer** aus der frequentistischen Statistik schätzt die unbekannte Wahrscheinlichkeit p durch

$$\hat{p} = \frac{\text{Anzahl Kopf}}{\text{Anzahl Würfe}}\,. \tag{13.26}$$

Berechne den Schätzwert $\hat{p}$ und bestimme damit die Wahrscheinlichkeit, dass die nächsten beiden Würfe Kopf zeigen.

c) Schreibe in `Python` eine Funktion, die für ein gegebenes $p = 5/7$ aufeinander folgende 14 Würfe simuliert. Werte diese Funktion 10 000-mal aus und mache eine Liste der Male, in denen genau 10-mal Kopf und 4-mal Zahl beobachtet wurde. Welche frequentistische Wahrscheinlichkeit hat also dieses Ergebnis (10-mal Kopf und 4-mal Zahl in 14 Würfen)?

Hinweis: Die Funktion `np.random.binomial(n, p)` zieht dir direkt eine solche Stichprobe. Google gerne mal, wie man diese importiert und wofür n dabei steht.

d) Bei n unabhängigen Versuchen mit Erfolgswahrscheinlichkeit p folgt die Anzahl der Erfolge k der Verteilung

$$B(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{13.27}$$

Schreibe in `Python` eine Funktion, die die Wahrscheinlichkeit für k-mal Kopf bei n Würfen mit vorgegebenem p berechnet. Berechne damit für

$$p = 0.2,\ 0.4,\ 0.6,\ 0.8,\ 0.9$$

jeweils die Wahrscheinlichkeit, bei 14 Würfen genau 10-mal Kopf und 4-mal Zahl zu beobachten.

Hinweis: Hier ist `np.random.binomial` *nicht* das richtige Werkzeug – das liefert dir nur eine zufällige Stichprobe, aber keine Wahrscheinlichkeit. Berechne stattdessen $n!$ selbst (oder mit `math.factorial`) und setze damit den Binomialkoeffizienten und $B(k; n, p)$ direkt aus der obigen Formel zusammen.

e) Plotte die Ergebnisse der Aufgaben c) und d) gemeinsam und vergleiche sie miteinander. Da c) bisher nur eine einzige frequentistische Wahrscheinlichkeit bei $p = \hat{p}$ liefert, musst du die Simulation aus c) dafür zunächst erweitern: Wiederhole sie für jeden der fünf Werte $p = 0.2, 0.4, 0.6, 0.8, 0.9$ aus Teil d). Trage anschließend sowohl die theoretischen (Teil d) als auch die simulierten (Teil c) Wahrscheinlichkeiten für $P(\text{10-mal Kopf})$ über p auf.

f) (*Bonus*) In der Bayesschen Statistik wird p als zufällige Größe modelliert. Wir nehmen zunächst an, dass alle Werte zwischen 0 und 1 gleich wahrscheinlich sind (*a-priori*-Verteilung). Versuche, die *a-posteriori*-Verteilung (numerisch oder auf dem Papier) zu

berechnen. Wie groß ist die Wahrscheinlichkeit, dass die nächsten beiden Würfe Kopf zeigen, wenn du diese Verteilung zur Berechnung verwendest? Unterscheidet sich das von dem frequentistischen Ergebnis? Basierend auf deiner Analyse: Solltest du Geld darauf verwetten, dass die nächsten beiden Würfe Kopf zeigen?
Hinweis: Erwartungswerte werden wie folgt berechnet:

$$\langle f(p) \rangle = \int_0^1 f(p)\, P(p \mid D)\, \mathrm{d}p\,. \tag{13.28}$$

Wenn du noch nicht gelernt hast, wie man das relevante Integral löst, recherchiere gerne im Internet, wie das numerisch in `Python` geht.

Musterlösung 13.6

a) Die interessierende Zufallsgröße ist die Anzahl k der Köpfe bei $n = 14$ unabhängigen Würfen mit Kopfwahrscheinlichkeit p; sie kann jeden Wert zwischen 0 und 14 annehmen. Das statistische Modell (Definition 13.3, Abschnitt 13.7) ist also $(\Omega, \Sigma, (P_p)_{p\in\Theta})$ mit Grundmenge $\Omega = \{0, 1, \ldots, 14\}$, Ereignissystem $\Sigma = \mathcal{P}(\Omega)$ und Parameterraum $\Theta = [0, 1]$; welche konkrete Wahrscheinlichkeit P_p jedem k zuordnet, leiten wir in Teil d) her.
b) $\hat{p} = 10/14 = 5/7 \approx 0.714$. Unter der Annahme $p = \hat{p}$ (und Unabhängigkeit der Würfe) ist die Wahrscheinlichkeit zweier weiterer Köpfe $\hat{p}^2 = (5/7)^2 = 25/49 \approx 0.510$.
c)

```
import numpy as np

def muenzwuerfe_simulieren(p, n=14):
    """Simuliert n aufeinanderfolgende Muenzwuerfe mit Kopfwahrscheinlichkeit
    ↪  p; gibt Anzahl Kopf zurueck."""
    return np.random.binomial(n, p)

p_hat = 5 / 7  # Ergebnis aus Teil b)
ergebnisse = [muenzwuerfe_simulieren(p_hat) for _ in range(10000)]
treffer = sum(1 for k in ergebnisse if k == 10)
print(f'Frequentistische Wahrscheinlichkeit: {treffer / 10000:.3f}
↪  ({treffer}/10000 Treffer)')
```

Wir simulieren also nicht mit einem beliebigen p, sondern gezielt mit unserem Schätzwert $\hat{p} = 5/7$ aus Teil b). In einem Durchlauf ergaben sich z.B. 2342 von 10 000 Wiederholungen mit genau 10-mal Kopf, also eine frequentistische Wahrscheinlichkeit von ≈ 0.234 (der genaue Wert schwankt von Durchlauf zu Durchlauf leicht). Das Ergebnis wird uns in Teil d) als Vergleichswert dienen.
d) Binomialkoeffizient und Fakultät lassen sich direkt aus ihrer in der Aufgabenstellung gegebenen Definition heraus programmieren:

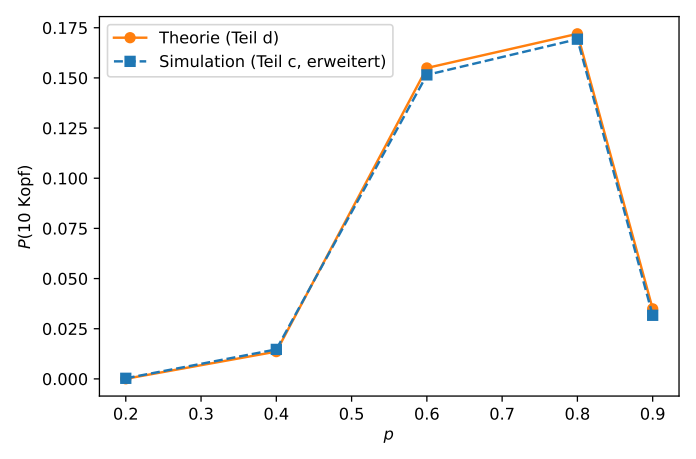
```
from math import factorial

def binomialkoeffizient(n, k):
    return factorial(n) // (factorial(k) * factorial(n - k))

def binomial_wahrscheinlichkeit(k, n, p):
    return binomialkoeffizient(n, k) * p**k * (1 - p)**(n - k)

for p in [0.2, 0.4, 0.6, 0.8, 0.9]:
    print(f'p={p}: P(10 Kopf) = {binomial_wahrscheinlichkeit(10, 14,
    ↪  p):.3g}')
```

p	0.2	0.4	0.6	0.8	0.9
$P(D \mid p)$	4.2×10^{-5}	1.4×10^{-2}	0.155	0.172	0.035

Die Wahrscheinlichkeit steigt also zunächst mit p, erreicht aber *nicht* ihr Maximum bei $p = 0.9$, sondern fällt dorthin schon wieder deutlich ab: Bei $p = 0.8$ ist sie mit 0.172 noch größer als bei $p = 0.9$ mit nur 0.035. Das passt genau zum beobachteten Anteil $10/14 \approx 0.71$, der zwischen 0.6 und 0.8 liegt – die Wahrscheinlichkeit hat dort ihr Maximum und nimmt in *beide* Richtungen wieder ab. Zur Probe werten wir dieselbe Funktion bei $p = \hat{p} = 5/7$ aus Teil b) aus: `binomial_wahrscheinlichkeit(10, 14, 5/7)` ≈ 0.231 – in guter Übereinstimmung mit der frequentistischen Schätzung ≈ 0.234 aus Teil c)!

e) Um die Ergebnisse aus c) und d) auf einem gemeinsamen Plot vergleichen zu können, müssen sie dieselbe Größe als Funktion derselben Variable darstellen. Wir ergänzen dazu die Simulation aus c), die bisher nur für den einen Wert $p = \hat{p}$ lief, um dieselben fünf Werte $p = 0.2, 0.4, 0.6, 0.8, 0.9$ wie in Teil d):

```
import matplotlib.pyplot as plt

p_werte = [0.2, 0.4, 0.6, 0.8, 0.9]
frequentistisch = []
for p in p_werte:
    ergebnisse = [muenzwuerfe_simulieren(p) for _ in range(10000)]
    treffer = sum(1 for k in ergebnisse if k == 10)
    frequentistisch.append(treffer / 10000)

theoretisch = [binomial_wahrscheinlichkeit(10, 14, p) for p in p_werte]

plt.plot(p_werte, theoretisch, 'o-', label='Theorie (Teil d)')
plt.plot(p_werte, frequentistisch, 's--', label='Simulation (Teil c,
↪  erweitert)')
plt.xlabel('$p$'); plt.ylabel(r'$P(10\ \mathrm{Kopf})$')
plt.legend(); plt.tight_layout(); plt.show()
```



Simulation und Theorie stimmen für alle fünf Werte von p sehr gut überein – ein weiterer Beleg dafür, dass die in Teil d) hergeleitete Formel tatsächlich dieselbe Verteilung beschreibt, die wir in c) simuliert haben. Der Wert $p = 0.9$ macht dabei besonders deutlich, dass die Wahrscheinlichkeit tatsächlich ein *Maximum* bei $p \approx 10/14$ hat und keineswegs mit p immer weiter ansteigt.

f) *(Bonus)* Prior $\pi(p) = 1$ für $p \in [0, 1]$ (vgl. die Gleichverteilung $\mathcal{U}(0, 1)$ in Abschnitt 13.9, der Standard-uninformative Prior). Die Likelihood ist die Binomialverteilung aus Teil d),

$P(D \mid p) = \binom{14}{10} p^{10}(1-p)^4$. Mit dem Satz von Bayes:

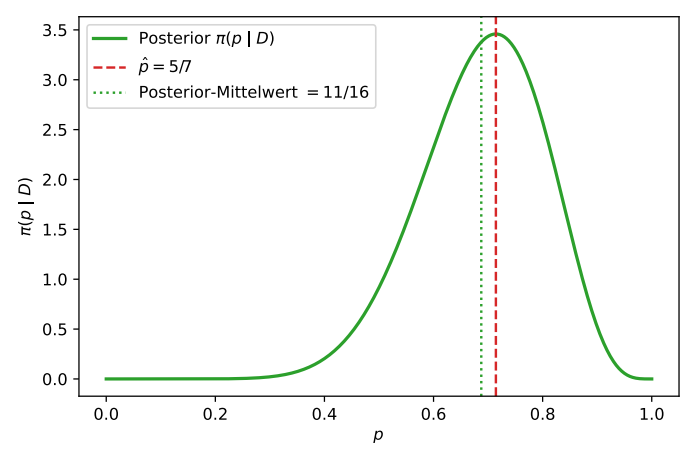
$$\pi(p \mid D) = \frac{P(D \mid p)\,\pi(p)}{P(D)} = \frac{\binom{14}{10} p^{10}(1-p)^4 \cdot 1}{P(D)} \propto p^{10}(1-p)^4\,, \tag{13.29}$$

da weder $\binom{14}{10}$ noch $P(D)$ von p abhängen und deshalb in der Proportionalitätskonstante verschwinden. Die fehlende Normierungskonstante ist das Integral $Z = \int_0^1 p^{10}(1-p)^4\,\mathrm{d}p$; damit ist $\pi(p \mid D) = p^{10}(1-p)^4/Z$. Statt diese Größe von Hand auszurechnen, werten wir $p^{10}(1-p)^4$ auf einem feinen Gitter aus und integrieren numerisch (z.B. mit der Trapezregel). Auf demselben Weg erhalten wir sofort auch die eigentlich gesuchte Größe: die Wahrscheinlichkeit zweier weiterer Köpfe ist ja gerade der Erwartungswert $\langle p^2 \rangle$ von $P(HH \mid p) = p^2$ unter der Posterior-Verteilung,

$$P(HH \mid D) = \langle p^2 \rangle_{\pi(p|D)} = \int_0^1 p^2\,\pi(p \mid D)\,\mathrm{d}p\,. \tag{13.30}$$

```
import numpy as np

p_grid = np.linspace(0, 1, 1000)
unnormiert = p_grid**10 * (1 - p_grid)**4

Z = np.trapezoid(unnormiert, p_grid)
posterior = unnormiert / Z

p_mittelwert = np.trapezoid(p_grid * posterior, p_grid)
p2_erwartung = np.trapezoid(p_grid**2 * posterior, p_grid)

print(f'Z = {Z:.4g}, <p> = {p_mittelwert:.4f}, <p^2> = {p2_erwartung:.4f}')
```

Das liefert $Z \approx 6.66 \times 10^{-5}$, $\langle p \rangle \approx 0.6875$ und $\langle p^2 \rangle \approx 0.4853$.



Der Posterior-Mittelwert $\langle p \rangle \approx 0.6875$ liegt etwas unter $\hat{p} = 5/7 \approx 0.714$: Der uninformative Prior zieht das Ergebnis leicht in Richtung seines eigenen Mittelwerts $\langle p \rangle = \frac{1}{2}$ („Shrinkage“), ein Effekt, der mit mehr Daten immer kleiner wird. Genau denselben Effekt sehen wir bei der eigentlich gesuchten Antwort: Die Bayessche Wahrscheinlichkeit für zwei weitere Köpfe, $P(HH \mid D) = \langle p^2 \rangle \approx 0.485$, liegt ebenfalls minimal unter der naiven frequentistischen Schätzung $\hat{p}^2 \approx 0.510$ aus Teil b) – auch hier zieht der prior die Wahrscheinlichkeit etwas herunter. Im Grenzwert vieler “Messungen” verschwindet die Abhängigkeit vom Prior und die frequentistische und die Bayessche Aussagen nähern sich einander an: Die Daten siegen am Ende über unsere subjektiven Vorannahmen.

13.9 Übergang zu kontinuierlichen Wahrscheinlichkeitsverteilungen

Im bisherigen Verlauf war unsere Grundmenge Ω stets *diskret* (abzählbar), wie die sechs Augenzahlen eines Würfels: Jedem einzelnen Ausgang $\omega \in \Omega$ konnte eine Wahrscheinlichkeit $P(\{\omega\}) > 0$ zugeordnet werden, und Σ war einfach die volle Potenzmenge. Die meisten physikalischen Messgrößen – eine Zeit, eine Energie, eine Position – sind aber *kontinuierlich*: Ω ist dann etwa $\mathbb{R}$ oder ein Intervall darin, mit überabzählbar vielen möglichen Ausgängen. Für ein einzelnes Element $x \in \mathbb{R}$ ergibt $P(\{x\}) = 0$ dann keinen Sinn mehr – Wahrscheinlichkeiten lassen sich hier nur noch *Intervallen* zuordnen, beschrieben durch eine **Wahrscheinlichkeitsdichtefunktion** (die formale Definition folgt in Kapitel 14). Wir lernen hier die beiden wichtigsten kontinuierlichen Verteilungen kennen – die Gleichverteilung und die Normalverteilung –, sowie den zentralen Grenzwertsatz, der erklärt, warum die Normalverteilung praktisch überall in der Natur begegnet.

13.9.1 Gleichverteilung

Die **Gleichverteilung** $\mathcal{U}(a, b)$ beschreibt eine Zufallsvariable, die jeden Wert im Intervall $[a, b]$ mit gleicher Wahrscheinlichkeit annimmt:

$$\mathcal{U}(x;\, a, b) = \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & \text{sonst} \end{cases}, \qquad \langle x \rangle = \frac{a+b}{2}\,, \quad \sigma_x^2 = \frac{(b-a)^2}{12}\,. \tag{13.31}$$

Sie ist der einfachste *uninformative Prior* in der Bayesschen Statistik: Hat man über den möglichen Wertebereich eines Parameters keinerlei Vorinformation, wählt man $p(\theta) \propto \mathcal{U}(\theta;\, a, b)$.

13.9.2 Normalverteilung (Gauß-Verteilung)

$$\mathcal{N}(x;\, \mu, \sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \langle x \rangle = \mu, \quad \sigma_x^2 = \sigma^2\,. \tag{13.32}$$

Wir schreiben $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn eine Zufallsvariable X dieser Verteilung folgt. Die Normalverteilung besitzt bekannte Eigenschaften: 68.3% der Wahrscheinlichkeit liegen in $[\mu - \sigma, \mu + \sigma]$, 95.4% in $[\mu - 2\sigma, \mu + 2\sigma]$, 99.7% in $[\mu - 3\sigma, \mu + 3\sigma]$. Abbildung 13.4 zeigt ihre Form für verschiedene Parameter.

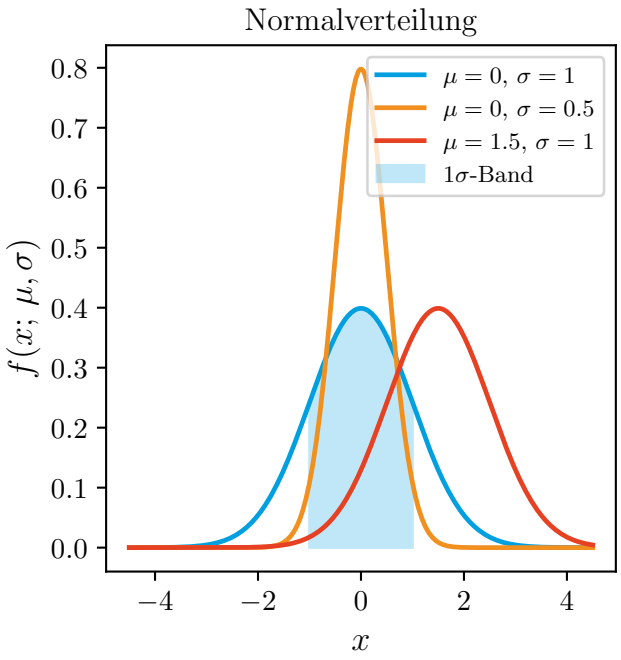


Abbildung 13.4: Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ für verschiedene Parameter; die blaue Fläche markiert das 1σ-Intervall (68.3% der Wahrscheinlichkeit).

Mehrdimensionale Normalverteilung. Für einen Vektor $\vec{x} = (x_1, \ldots, x_n)^T$ mit Mittelwert $\vec{\mu}$ und symmetrischer Kovarianzmatrix V ($V_{ij} = \mathrm{cov}(x_i, x_j)$) ist die Dichte der n-dimensionalen Normalverteilung durch

$$f(\vec{x}; \vec{\mu}, V) = \frac{1}{(2\pi)^{n/2}\sqrt{\det V}} \exp\left[-\tfrac{1}{2}(\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu})\right] \ , \tag{13.33}$$

gegeben, wobei $\det V$ die *Determinante* der Kovarianzmatrix V bezeichnet.

13.9.3 Zentraler Grenzwertsatz

Bisher haben wir vor allem diskrete Verteilungen mit wenigen, klar unterscheidbaren Werten betrachtet (den Würfel, die Binomial- und die Poissonverteilung). Was passiert, wenn eine solche diskrete Zufallsgröße immer mehr mögliche Werte annehmen kann? Dieser Frage nähern wir uns über Computersimulationen – und stoßen dabei auf den vielleicht wichtigsten Satz der gesamten Wahrscheinlichkeitstheorie: den **zentralen Grenzwertsatz (ZGS)**.

Aufgabe 13.7

Wir untersuchen, was mit der Binomialverteilung $B(k; n, p)$ passiert, wenn n immer größer und p *nicht* kleiner wird (im Gegensatz zum Grenzübergang zur Poissonverteilung aus Abschnitt 13.6).

a) Wähle $p = 0.3$ fest. Erzeuge für $n = 10, 50, 200, 1000$ jeweils mit `np.random.binomial(n, p, size=10000)` einen Datensatz von 10 000 Stichproben. Plotte für jedes n das normierte Histogramm (`density=True`), am besten in vier Teilplots (`plt.subplots`) nebeneinander. Was fällt dir an der *Form* der Histogramme auf, je größer n wird?

b) Die vier Histogramme aus a) sind auf sehr unterschiedlichen k-Achsen nicht direkt vergleichbar, da sowohl Lage als auch Breite der Verteilung von n abhängen (erinnere

dich: $\langle k \rangle = np$ und $\sigma_k^2 = np(1-p)$). Um die *Form* unabhängig von n vergleichen zu können, *standardisiere* jede Stichprobe k gemäß

$$z = \frac{k - \langle k \rangle}{\sigma_k} = \frac{k - np}{\sqrt{np(1-p)}}\,. \tag{13.34}$$

Plotte nun die normierten Histogramme der standardisierten Werte z für alle vier n *gemeinsam in einem einzigen Plot* (z. B. mit `alpha=0.5` und unterschiedlichen Farben, damit man sie unterscheiden kann). Was beobachtest du für wachsendes n?

c) Für $n = 10$ kann k nur 11 verschiedene Werte annehmen, für $n = 1000$ dagegen 1001 verschiedene Werte – auf demselben (standardisierten) Bereich, den die Verteilung im Wesentlichen abdeckt. Die „Lücken" zwischen den möglichen Werten von z werden also immer kleiner. Im Grenzfall $n \to \infty$ liegt schließlich *jeder* reelle Wert im Bereich der Verteilung – z wird zu einer **kontinuierlichen** Zufallsgröße, beschrieben durch eine **Wahrscheinlichkeitsdichte** statt einzelner Wahrscheinlichkeiten $P(k)$ (die formale Definition folgt in Kapitel 14). Die Grenzkurve, der sich die standardisierten Binomialverteilungen annähern, ist genau die Normalverteilung $\mathcal{N}(x; \mu, \sigma^2)$ aus dem vorigen Abschnitt, hier mit $\mu = 0$ und $\sigma = 1$ – die sogenannte **Standardnormalverteilung** $\mathcal{N}(x; 0, 1)$. Ergänze deinen Plot aus b) um diese Kurve.

d) Der ZGS gilt nicht nur für Summen von Bernoulli-Variablen (also für die Binomialverteilung), sondern für Summen *beliebiger* unabhängiger, identisch verteilter Zufallsgrößen mit endlicher Varianz – ganz gleich, wie diese einzeln verteilt sind! Überprüfe das:

- Ziehe mit `np.random.uniform(0, 1, size=(10000, n))` jeweils 10 000 Stichproben von n gleichverteilten Zufallszahlen und bilde deren Summe S_n (Summation über die richtige Achse).
- Standardisiere die Stichproben von S_n (du benötigst dafür Mittelwert und Varianz einer einzelnen $\mathcal{U}(0, 1)$-Zufallsgröße, siehe Abschnitt 13.9).
- Plotte das normierte Histogramm für $n = 1, 2, 5, 30$ zusammen mit $\mathcal{N}(x; 0, 1)$.

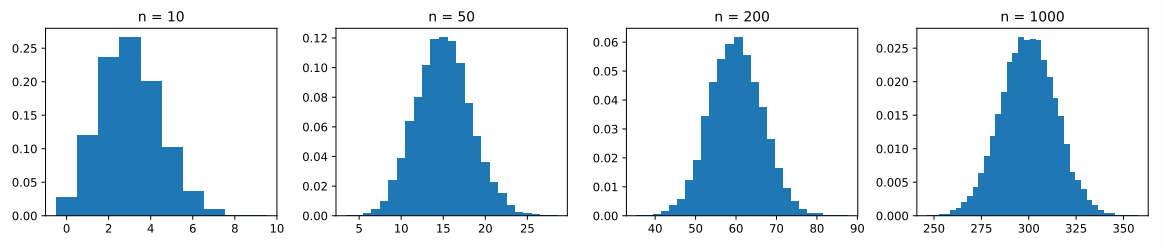
e) Der ZGS ist einer der Hauptgründe, warum die Normalverteilung in der Physik – und insbesondere bei der Beschreibung von **Messunsicherheiten** – eine so herausragende Rolle spielt. Die Grundidee: Eine einzelne Messung (z. B. die Länge eines Tisches mit einem Lineal) wird praktisch nie durch *eine* einzige Fehlerquelle verfälscht, sondern durch eine *Überlagerung vieler kleiner, voneinander unabhängiger Störungen* – Ablesefehler, kleine Temperaturschwankungen, Erschütterungen, Ungenauigkeiten im Material des Lineals, und so weiter. Begründe in eigenen Worten, weshalb der ZGS ein gutes Argument dafür liefert, warum zufällige Messunsicherheiten in der Physik häufig durch eine Normalverteilung modelliert werden – auch wenn man die einzelnen, elementaren Fehlerquellen einer Messung oft gar nicht genau kennt. In welchen Situationen könnte diese Rechtfertigung *nicht* gut funktionieren?

Musterlösung 13.7

a)

```
import numpy as np
import matplotlib.pyplot as plt
```

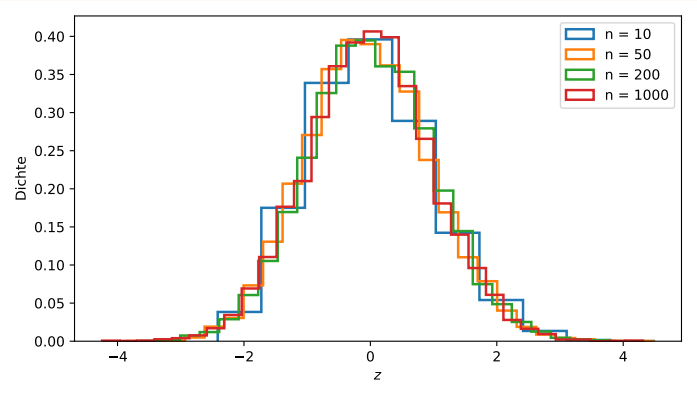
```
p = 0.3
n_werte = [10, 50, 200, 1000]

def diskrete_bins(daten, max_bins=40):
    k_min, k_max = daten.min(), daten.max()
    m = int(np.ceil((k_max - k_min + 1) / max_bins))
    return np.arange(k_min - 0.5, k_max + m + 0.5, m)

fig, axes = plt.subplots(1, 4, figsize=(14, 3))
for ax, n in zip(axes, n_werte):
    stichprobe = np.random.binomial(n, p, size=10000)
    ax.hist(stichprobe, bins=diskrete_bins(stichprobe), density=True)
    ax.set_title(f'n = {n}')
plt.tight_layout(); plt.show()
```



Wichtig ist hier die Wahl der Bins: Da k nur ganzzahlige Werte annehmen kann, dürfen Bin-Kanten nur zwischen zwei ganzzahligen Werten liegen, nie mittendrin – sonst landen mal ein, mal zwei benachbarte k-Werte im selben Balken, je nachdem wie die Kanten zufällig mit den k-Werten zusammenfallen, und das Histogramm sieht „zackig“ aus (vgl. das Poisson-Histogramm in Aufgabe 13.5). Bei kleinem n bekommt so jeder Balken genau einen k-Wert; bei großem n gäbe das bei fester Stichprobengröße 10 000 aber sehr viele, nur noch dünn besetzte Bins und damit ein unnötig verrauschtes Bild – `diskrete_bins` fasst deshalb bei Bedarf mehrere benachbarte k-Werte zu einem (weiterhin exakt ausgerichteten) Bin zusammen, sodass alle vier Panels ähnlich viele Bins haben.
Für kleine n ist die Verteilung noch deutlich schief (asymmetrisch) und „kantig“, da k nur wenige ganzzahlige Werte annehmen kann. Mit wachsendem n wird die Form zunehmend symmetrisch und glockenförmig.

b)

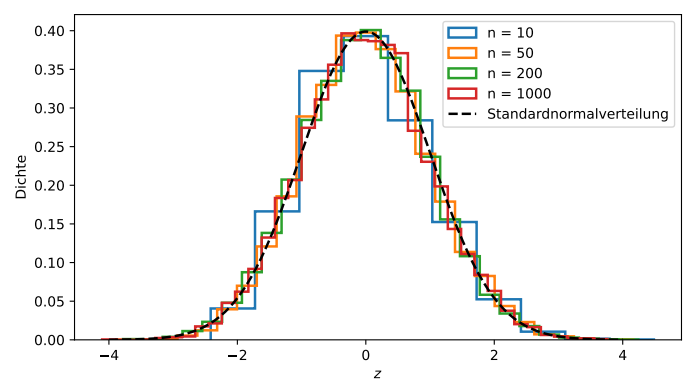
```
plt.figure(figsize=(7, 4))
for n in n_werte:
    stichprobe = np.random.binomial(n, p, size=10000)
    erwartung = n * p
    streuung = np.sqrt(n * p * (1 - p))
    z = (stichprobe - erwartung) / streuung
    z_bins = (diskrete_bins(stichprobe) - erwartung) / streuung
    plt.hist(z, bins=z_bins, density=True, histtype='step', linewidth=1.8,
    ↪  label=f'n = {n}')
plt.xlabel('$z$'); plt.ylabel('Dichte')
plt.legend(); plt.tight_layout(); plt.show()
```



Wir zeichnen die vier Histogramme hier als reine Umrisslinien (`histtype='step'`) statt gefüllter, transparenter Flächen: Bei vier übereinanderliegenden Verteilungen verschmilzt die Alpha-Überlagerung sonst schnell zu einem unleserlichen Farbbrei, während die Umrisse auch im dichten Zentrum klar unterscheidbar bleiben. Für wachsendes n nähern sich alle vier standardisierten Histogramme derselben glockenförmigen Kurve an – unabhängig davon, dass sich Lage und Breite der *unstandardisierten* Verteilungen stark unterscheiden. Nur bei $n = 10$ (blau) erkennt man noch deutlich die einzelnen, diskreten Stufen.

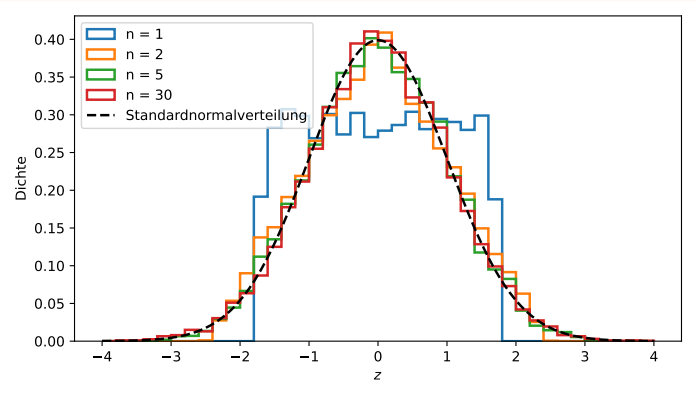
c)

```
z_werte = np.linspace(-4, 4, 200)
dichte_standard = 1 / np.sqrt(2 * np.pi) * np.exp(-z_werte**2 / 2)

plt.figure(figsize=(7, 4))
for n in n_werte:
    stichprobe = np.random.binomial(n, p, size=10000)
    erwartung = n * p
    streuung = np.sqrt(n * p * (1 - p))
    z = (stichprobe - erwartung) / streuung
    z_bins = (diskrete_bins(stichprobe) - erwartung) / streuung
    plt.hist(z, bins=z_bins, density=True, histtype='step', linewidth=1.8,
    ↪  label=f'n = {n}')
plt.plot(z_werte, dichte_standard, 'k--', linewidth=1.8,
↪  label='Standardnormalverteilung')
plt.xlabel('$z$'); plt.ylabel('Dichte')
plt.legend(); plt.tight_layout(); plt.show()
```



d) Für $\mathcal{U}(0,1)$ gilt nach der Formel aus Abschnitt 13.9 $\langle x \rangle = 1/2$ und $\sigma_x^2 = 1/12$:

```
n_werte_2 = [1, 2, 5, 30]
mittelwert_uniform = 0.5
varianz_uniform = 1 / 12

plt.figure(figsize=(7, 4))
bins = np.linspace(-4, 4, 41)  # gemeinsame Bins fuer alle n (S_n ist
↪  kontinuierlich)
for n in n_werte_2:
    stichprobe = np.random.uniform(0, 1, size=(10000, n))
    S_n = stichprobe.sum(axis=1)
    erwartung = n * mittelwert_uniform
    streuung = np.sqrt(n * varianz_uniform)
    z = (S_n - erwartung) / streuung
    plt.hist(z, bins=bins, density=True, histtype='step', linewidth=1.8,
    ↪  label=f'n = {n}')
plt.plot(z_werte, dichte_standard, 'k--', linewidth=1.8,
↪  label='Standardnormalverteilung')
plt.xlabel('$z$'); plt.ylabel('Dichte')
plt.legend(); plt.tight_layout(); plt.show()
```



Schon ab $n = 5$ ist die standardisierte Summe kaum noch von der Standardnormalverteilung zu unterscheiden – obwohl die einzelnen x_i gleichverteilt und damit alles andere als glockenförmig sind! Nur $n = 1$ (die unveränderte Gleichverteilung selbst) weicht noch deutlich ab. Genau das ist die Kernaussage des ZGS: Für die Form der Summe spielt die Form der Einzelverteilungen im Grenzfall vieler Summanden keine Rolle mehr.

e) Solange der Gesamtfehler aus *vielen kleinen, voneinander unabhängigen* Beiträgen zusammengesetzt ist, sagt der ZGS voraus, dass die Summe näherungsweise normalverteilt ist – unabhängig davon, wie genau die einzelnen Beiträge verteilt sind. Das rechtfertigt, Messunsicherheiten pragmatisch als normalverteilt anzunehmen, ohne jede einzelne Fehlerquelle im Detail zu kennen. Diese Rechtfertigung funktioniert *nicht* gut, wenn

- ein einzelner Fehlerbeitrag den Gesamtfehler dominiert (z. B. ein grober Ablesefehler oder ein defektes Messgerät) statt vieler gleichberechtigter kleiner Beiträge,
- die einzelnen Fehlerquellen nicht unabhängig, sondern korreliert sind (z. B. ein systematischer Fehler, der alle Messungen gleichermaßen verschiebt),
- nur sehr wenige Fehlerquellen beitragen, sodass der (asymptotische) Grenzfall des ZGS noch nicht erreicht ist – insbesondere in den Flanken der Verteilung (seltene, große Abweichungen) kann die tatsächliche Verteilung dann noch spürbar von einer Normalverteilung abweichen.

Bemerkung: Dies lässt sich auch explizit simulieren. Modelliert man den wahren Messfehler als Summe $\varepsilon = \sum_{i=1}^{12} \varepsilon_i$ von $n = 12$ unabhängig auf $[-0.5, 0.5]$ gleichverteilten Einzelbeiträgen (in geeigneten Einheiten, z. B. Millimeter) und erzeugt so 10 000 simulierte „Messungen“, so ist das resultierende Histogramm bereits sehr gut durch eine Normalverteilung $\mathcal{N}(x; \mu, \sigma^2)$ mit aus der Stichprobe bestimmten μ und σ beschrieben.

Nun formalisieren wir unsere Erkenntnisse aus Aufgabe 13.7 und stellen eine weitverbreitete Version des *Zentralen Grenzwertsatzes* vor:

Satz 13.1

Zentraler Grenzwertsatz (ZGS), Lindeberg–Lévy-Form. Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ unabhängige, identisch verteilte Zufallsvariablen mit $\langle X_i \rangle = \mu$ und $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Für jedes $n \in \mathbb{N}$ seien

$$S_n := \sum_{i=1}^{n} X_i \,, \qquad Z_n := \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \,. \tag{13.35}$$

Dann konvergiert Z_n für $n \to \infty$ *in Verteilung* gegen die Standardnormalverteilung, d. h. für jedes $a \in \mathbb{R}$ gilt

$$\lim_{n\to\infty} P(Z_n \le a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} \, \mathrm{d}x \,. \tag{13.36}$$

Der Beweis findet sich in Aufgabe 13.8. Der ZGS gilt auch für unabhängige, nicht notwendig identisch verteilte Zufallsvariablen $X_1, \ldots, X_n$ mit verschiedenen Mittelwerten μ_i und endlichen Varianzen σ_i – allerdings nur, sofern *keine einzelne* Varianz die summierte Größe $\sigma^2 = \sum_i \sigma_i^2$ dominiert (sog. Lindeberg-Bedingung; andernfalls gibt allein diese eine Variable die Form der Summe vor, siehe die Diskussion in Teil e) von Aufgabe 13.7). Unter dieser Voraussetzung konvergiert die Zufallsvariable

$$Z_n := \frac{\sum_{i=1}^{n} X_i - \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \tag{13.37}$$

für $n \to \infty$ in Verteilung gegen die Standardnormalverteilung unabhängig von der Form der Einzelverteilungen der X_i. Diese allgemeinere Form steckt hinter der Anwendung auf Messunsicherheiten in Aufgabe 13.7.

Aufgabe 13.8

Beweis des Zentralen Grenzwertsatzes: In dieser Aufgabe beweisen wir den ZGS in der Lindeberg–Lévy-Form aus Satz 13.1, also für unabhängige, *identisch verteilte* (i.i.d.) Zufallsvariablen $X_1, \ldots, X_n$ mit Mittelwert μ und endlicher Streuung σ^2. Der Beweis nutzt ein Werkzeug der Wahrscheinlichkeitstheorie, die **momenterzeugende Funktion** (MEF) einer Zufallsvariable X:

$$M_X(t) := \langle e^{tX} \rangle \,, \tag{13.38}$$

sofern dieser Erwartungswert für t in einer Umgebung von 0 existiert (eine technische Zusatzvoraussetzung, die für die meisten in der Praxis auftretenden Verteilungen erfüllt ist – der vollständige ZGS unter der schwächeren Voraussetzung endlicher Varianz benötigt statt der MEF die sogenannte *charakteristische Funktion* und ist deutlich technischer zu beweisen). Ohne Beschränkung der Allgemeinheit nehmen wir an, dass $\mu = 0$ gilt (sonst betrachten wir

$X_i - \mu$ statt X_i).

a) Zeige durch Reihenentwicklung von e^{tX} (Taylor-Reihe, vgl. Abschnitt 7 oder Wikipedia), dass für $t \to 0$

$$M_X(t) = 1 + \langle X \rangle\, t + \frac{\langle X^2 \rangle}{2} t^2 + \mathcal{O}(t^3) \tag{13.39}$$

gilt (wir nehmen an, dass Reihenentwicklung und Erwartungswert vertauscht werden dürfen – eine Rechtfertigung dafür würde den Rahmen dieser Aufgabe sprengen). Nutze (13.39), um

$$M_X(t) = 1 + \frac{\sigma^2}{2} t^2 + \mathcal{O}(t^3) \tag{13.40}$$

für $t \to 0$ zu zeigen.

b) Zeige: Für zwei *unabhängige* Zufallsvariablen X, Y gilt $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ (Tipp: $e^{t(X+Y)} = e^{tX} e^{tY}$, und für unabhängige X, Y faktorisiert der Erwartungswert eines Produkts, $\langle f(X) g(Y) \rangle = \langle f(X) \rangle \langle g(Y) \rangle$). Folgere daraus, dass für n unabhängige, identisch verteilte $X_1, \ldots, X_n$ mit $Y = \sum_i X_i$

$$M_Y(t) = \big[M_X(t) \big]^n \tag{13.41}$$

gilt.

c) Wir betrachten die *standardisierte* Summe $Z_n = Y/(\sigma\sqrt{n})$ (beachte $\mu = 0$). Zeige mit Gleichung (13.38), dass

$$M_{Z_n}(t) = M_Y\left(\frac{t}{\sigma\sqrt{n}} \right) \tag{13.42}$$

gilt. Nutze die bisher hergeleiteten Gleichungen, um

$$M_{Z_n}(t) = \left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2}) \right]^n \tag{13.43}$$

im Grenzwert $n \to \infty$ zu zeigen.

d) Logarithmiere Gleichung (13.43):

$$\ln M_{Z_n}(t) = n \ln\left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2}) \right] . \tag{13.44}$$

Nutze die Taylor-Entwicklung $\ln(1+u) = u + \mathcal{O}(u^2)$ um $u = 0$ mit $u = t^2/(2n) + \mathcal{O}(n^{-3/2})$, um für festes t und $n \to \infty$ den Grenzwert $M_{Z_n}(t) \to e^{t^2/2}$ zu zeigen.

e) Zeige durch quadratische Ergänzung im Exponenten, dass die momenterzeugende Funktion der Standardnormalverteilung $\mathcal{N}(x; 0, 1)$ genau $M(t) = e^{t^2/2}$ ist. (Tipp: Schreibe $tx - x^2/2 = -\frac{1}{2}(x-t)^2 + \frac{t^2}{2}$ und nutze, dass $\mathcal{N}(x; t, 1)$ als Wahrscheinlichkeitsdichte über ganz $\mathbb{R}$ zu 1 integriert.)

f) Es stellt sich die Frage, ob aus der Konvergenz der MEF $M_{Z_n}(t) \to M(t)$ die Konvergenz der Verteilung folgt. Ein zentraler Baustein für die Antwort auf diese Frage sind die Taylor-Koeffizienten der MEF, die gerade die Momente der Verteilungen sind.

Verallgemeinere dein Ergebnis aus a) auf beliebige Ordnung $k \in \mathbb{N}_0$ und zeige (durch k-faches Differenzieren von $M_X(t) = \langle e^{tX} \rangle$ nach t, wobei Ableitung und Erwartungswert vertauscht werden dürfen), dass

$$M_X(t) = \sum_{k=0}^{\infty} \frac{\langle X^k \rangle}{k!} t^k \,, \qquad \text{also} \qquad M_X^{(k)}(0) = \langle X^k \rangle \tag{13.45}$$

gilt. Hier bezeichnet $M_X^{(k)}$ die k-te Ableitung nach t. Nutze das, um zu begründen (unter der – hier nicht bewiesenen – Tatsache, dass man bei Potenzreihen mit gemeinsamem Konvergenzradius Ableitung und Grenzwert $n \to \infty$ vertauschen darf), dass aus $M_{Z_n}(t) \to M(t)$ für jedes feste k

$$\langle Z_n^k \rangle \xrightarrow{n\to\infty} \langle N^k \rangle \,, \qquad N \sim \mathcal{N}(0,1) \tag{13.46}$$

folgt – *alle* Momente von Z_n konvergieren also gegen die entsprechenden Momente der Standardnormalverteilung. Überlege dir abschließend (ohne es beweisen zu müssen): Warum sollte eine Wahrscheinlichkeitsverteilung durch die Gesamtheit *aller* ihrer Momente eindeutig festgelegt sein?

Informiere dich im Internet über das Momentenproblem.

g) Folgere aus den bisherigen Ergebnissen, dass der Zentrale Grenzwertsatz in der Lindeberg–Lévy-Form gilt, sofern die MEF der Zufallsvariablen existiert.

Musterlösung 13.8

a) Aus der Taylor-Reihe der Exponentialfunktion, $e^{tX} = \sum_{k=0}^{\infty} (tX)^k/k! = 1 + tX + \frac{1}{2}t^2X^2 + \mathcal{O}(t^3)$, folgt durch (angenommenermaßen erlaubtes) gliedweises Bilden des Erwartungswerts

$$M_X(t) = \langle e^{tX} \rangle = 1 + t\langle X \rangle + \frac{t^2}{2}\langle X^2 \rangle + \mathcal{O}(t^3) \,.$$

Für die Varianz gilt allgemein $\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2$; mit $\langle X \rangle = \mu = 0$ bleibt also $\langle X^2 \rangle = \sigma^2$. Einsetzen liefert genau Gleichung (13.40).
b) Wegen $e^{t(X+Y)} = e^{tX}e^{tY}$ und Unabhängigkeit von X, Y gilt

$$M_{X+Y}(t) = \langle e^{tX}e^{tY} \rangle = \langle e^{tX} \rangle \langle e^{tY} \rangle = M_X(t) M_Y(t) \,.$$

Induktionsanfang $n = 1$: $M_{X_1}(t) = M_X(t) = [M_X(t)]^1$, klar. Induktionsschritt: Gelte $M_{X_1+\cdots+X_{n-1}}(t) = [M_X(t)]^{n-1}$. Da X_n unabhängig von $X_1 + \cdots + X_{n-1}$ ist, liefert die soeben gezeigte Produktregel

$$\begin{aligned} M_Y(t) &= M_{(X_1+\cdots+X_{n-1})+X_n}(t) = M_{X_1+\cdots+X_{n-1}}(t) \cdot M_{X_n}(t) \\ &= [M_X(t)]^{n-1} \cdot M_X(t) = [M_X(t)]^n \,, \end{aligned}$$

wobei $M_{X_n}(t) = M_X(t)$, da alle X_i identisch verteilt sind.
c) Mit $Z_n = \frac{1}{\sigma\sqrt{n}} Y$ gilt

$$M_{Z_n}(t) = \langle e^{tZ_n} \rangle = \left\langle e^{\frac{t}{\sigma\sqrt{n}} Y} \right\rangle = M_Y\left(\frac{t}{\sigma\sqrt{n}}\right) \,.$$

Mit Gleichung (13.41) und Gleichung (13.40) bei $s = t/(\sigma\sqrt{n})$ folgt $\sigma^2 s^2/2 = t^2/(2n)$ und $\mathcal{O}(s^3) = \mathcal{O}(n^{-3/2})$, also

$$M_{Z_n}(t) = \big[M_X(s)\big]^n = \left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2})\right]^n .$$

d) Mit $u = t^2/(2n) + \mathcal{O}(n^{-3/2}) \to 0$ für $n \to \infty$ (bei festem t) liefert $\ln(1+u) = u + \mathcal{O}(u^2)$, wobei $\mathcal{O}(u^2) = \mathcal{O}(n^{-2})$:

$$\begin{aligned}\ln M_{Z_n}(t) &= n \ln\left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{-3/2})\right] \\ &= n\left[\frac{t^2}{2n} + \mathcal{O}(n^{-3/2}) + \mathcal{O}(n^{-2})\right] = \frac{t^2}{2} + \mathcal{O}(n^{-1/2}) .\end{aligned}$$

Für $n \to \infty$ verschwindet der Restterm, also $\ln M_{Z_n}(t) \to t^2/2$ und somit $M_{Z_n}(t) \to e^{t^2/2}$.
e) Mit $tx - \frac{1}{2}x^2 = -\frac{1}{2}(x-t)^2 + \frac{t^2}{2}$ gilt

$$M(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx - x^2/2}\, \mathrm{d}x = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}\, \mathrm{d}x = e^{t^2/2} \cdot 1 = e^{t^2/2} ,$$

denn das verbliebene Integral ist genau die auf 1 normierte Dichte $\mathcal{N}(x; t, 1)$.
f) Differenziert man $M_X(t) = \langle e^{tX} \rangle$ formal k-mal nach t und vertauscht dabei Ableitung und Erwartungswert, erhält man $M_X^{(k)}(t) = \langle X^k e^{tX} \rangle$, also $M_X^{(k)}(0) = \langle X^k \rangle$. Das sind gerade die Taylor-Koeffizienten von M_X um $t = 0$ (mal $k!$), also $M_X(t) = \sum_{k=0}^{\infty} \langle X^k \rangle\, t^k/k!$ – eine direkte Verallgemeinerung von Gleichung (13.39) auf beliebige Ordnung k, Gleichung (13.45).
Da M_{Z_n} und M als Potenzreihen um $t = 0$ mit gemeinsamem, positivem Konvergenzradius geschrieben werden können und $M_{Z_n}(t) \to M(t)$ auf einer ganzen Umgebung von 0 gilt, folgt auch die Konvergenz sämtlicher Ableitungen an der Stelle 0: $M_{Z_n}^{(k)}(0) \to M^{(k)}(0)$ für jedes feste k. Mit Gleichung (13.45) bedeutet das gerade $\langle Z_n^k \rangle \to \langle N^k \rangle$ für alle $k \in \mathbb{N}_0$, also Gleichung (13.46).
Intuitiv legen die Momente einer Verteilung immer feiner ihre Form fest: Der nullte Moment normiert sie, der erste ist ihr Schwerpunkt (Mittelwert), der zweite beschreibt die Streuung um den Schwerpunkt, höhere Momente erfassen Schiefe, Wölbung und beliebig feine weitere Details der Form. Stimmen *alle* Momente zweier Verteilungen überein, bleibt intuitiv kein Freiheitsgrad mehr übrig, in dem sie sich noch unterscheiden könnten. Im Allgemeinen kann das trotzdem schiefgehen: Es gibt tatsächlich verschiedene Verteilungen mit exakt denselben Momenten (das sog. *Momentenproblem* ist nicht immer eindeutig lösbar) – typischerweise dann, wenn die Verteilung so schwere Flanken hat, dass ihre MEF nirgends außer bei $t = 0$ existiert. Existiert die MEF dagegen (wie hier für $\mathcal{N}(0,1)$) in einer ganzen Umgebung von 0, ist die Verteilung durch ihre Momente eindeutig bestimmt, und der Satz von Fréchet–Shohat garantiert dann tatsächlich die Konvergenz in Verteilung. Das erklärt, warum die MEF-Konvergenz $M_{Z_n}(t) \to M(t)$ tatsächlich die Konvergenz von Z_n in Verteilung nach sich zieht – das nutzen wir nun in g), um den Beweis abzuschließen.
g) In d) und e) wurde gezeigt, dass $M_{Z_n}(t) \to e^{t^2/2} = M(t)$ für jedes feste t in einer Umgebung von 0 – die MEF von Z_n strebt also gegen die MEF der Standardnormalverteilung. Nach dem in f) diskutierten Zusammenhang (Momentenkonvergenz plus Eindeutigkeit der Verteilung durch ihre Momente, Satz von Fréchet–Shohat – äquivalent: der Eindeutigkeitssatz für momenterzeugende Funktionen) folgt daraus tatsächlich die Konvergenz der Verteilungsfunktionen,

$$\lim_{n\to\infty} P(Z_n \le a) = \int_{-\infty}^{a} \mathcal{N}(x; 0, 1)\, \mathrm{d}x$$

für jedes $a \in \mathbb{R}$ – und das ist genau die in Satz 13.1 behauptete Konvergenz von Z_n in Verteilung gegen $\mathcal{N}(0,1)$ (mit $Z_n = (Y - n\mu)/(\sigma\sqrt{n})$, was nach Rücknahme der Vereinfachung $\mu = 0$ gerade der Definition von Z_n im Satz entspricht). Damit ist der ZGS für identisch verteilte Zufallsvariablen mit existierender MEF bewiesen.

Bemerkung zu charakteristische Funktionen. Die Voraussetzung, dass die MEF $M_X(t) = \langle e^{tX} \rangle$ in einer ganzen Umgebung von $t = 0$ existiert, ist tatsächlich einschränkend: Bei Verteilungen mit *heavy tails* (aber mit endlicher Varianz!) kann $\langle e^{tX} \rangle$ für jedes $t \neq 0$ divergieren, weil e^{tX} schneller wächst, als die Flanken der Verteilung abfallen. Der Ausweg ist, e^{tX} durch $e^{\mathrm{i}tX}$ zu ersetzen (mit der imaginären Einheit i) und die **charakteristische Funktion** $\varphi_X(t) := \langle e^{\mathrm{i}tX} \rangle$ zu betrachten – im Wesentlichen die Fourier-Transformierte der Wahrscheinlichkeitsdichte von X. Da $|e^{\mathrm{i}tX}| = 1$ für jedes reelle t und jedes reelle X gilt, ist $\varphi_X(t)$ als Erwartungswert einer beschränkten Größe für *jede* Zufallsvariable X und jedes $t \in \mathbb{R}$ endlich – die technische Zusatzvoraussetzung entfällt also vollständig. Die Schritte a)–d) dieser Aufgabe lassen sich fast wörtlich auf φ_X übertragen; nur der letzte Schritt (aus punktweiser Konvergenz der charakteristischen Funktionen auf Konvergenz in Verteilung zu schließen) erfordert echte Fourier-Analysis – den sog. *Lévyschen Stetigkeitssatz* –, was den vollständigen Beweis des ZGS unter der schwächeren Voraussetzung endlicher Varianz deutlich technischer macht als die hier gezeigte Version mit der MEF.

14 Statistik

In diesem Kapitel erarbeiten wir die statistischen Grundlagen, die für die Datenanalyse am Ende des Kurses und in der modernen Physik allgemein unverzichtbar sind. Die Inhalte folgen dem Aufbau einer klassischen Statistikvorlesung, angepasst an unsere Bedürfnisse.

14.1 Stichproben und Verteilungen

14.1.1 Deskriptive Statistik

Ein Datensatz $\{x_1, \ldots, x_N\}$ lässt sich mit wenigen Kenngrößen charakterisieren:

- **Arithmetisches Mittel:** $\bar{x} = \dfrac{1}{N}\sum_{i=1}^{N} x_i$
- **Varianz:** $V(x) = \dfrac{1}{N}\sum_i (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$
- **Standardabweichung:** $\sigma = \sqrt{V(x)}$
- **Median:** Der Wert, unterhalb dessen genau die Hälfte der Messwerte liegt.

Für *Paare* (x_i, y_i) misst die **Kovarianz** den linearen Zusammenhang:

$$\operatorname{cov}(x,y) = \frac{1}{N}\sum_i (x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\bar{y}\,. \tag{14.1}$$

Der dimensionslose **Korrelationskoeffizient** $\rho = \operatorname{cov}(x,y)/(\sigma_x\sigma_y)$ liegt stets in $[-1,1]$: $\rho = +1$ bedeutet perfekte positive, $\rho = -1$ perfekte negative Korrelation, $\rho = 0$ zeigt *keine lineare* Abhängigkeit (aber nicht notwendigerweise Unabhängigkeit!).

14.1.2 Wahrscheinlichkeitsverteilungen

Woher kommt ein Datensatz? Er stammt aus einem zugrundeliegenden *Zufallsprozess*, beschrieben durch eine **Wahrscheinlichkeitsverteilung** $P(x)$:

- **Diskret:** $P(x)$ ist die Wahrscheinlichkeit, den Wert x zu beobachten; es gilt $\sum_x P(x) = 1$.
- **Stetig:** $P(x)\,\mathrm{d}x$ ist die Wahrscheinlichkeit, einen Wert in $[x, x+\mathrm{d}x]$ zu beobachten; es gilt $\int_{-\infty}^{\infty} P(x)\,\mathrm{d}x = 1$. Man nennt $P(x)$ die *Wahrscheinlichkeitsdichtefunktion* (PDF).

14.1.3 Erwartungswerte

Der **Erwartungswert** einer Funktion $f(x)$ ist der gewichtete Mittelwert über alle möglichen Ausgänge:

$$\langle f(x)\rangle = \int f(x)\,P(x)\,\mathrm{d}x\,. \tag{14.2}$$

Insbesondere: $\langle x\rangle$ ist der Mittelwert der Verteilung, $V(x) = \langle x^2\rangle - \langle x\rangle^2$ ihre Varianz.

14.1.4 Kumulative Verteilungsfunktion (CDF)

Die **kumulative Verteilungsfunktion** gibt die Wahrscheinlichkeit an, einen Wert $\le a$ zu beobachten:

$$F(a) = \int_{-\infty}^{a} P(x)\,\mathrm{d}x = p(x \le a)\,. \tag{14.3}$$

Es gilt $F(-\infty) = 0$, $F(+\infty) = 1$ und per Definition $F(x_{\mathrm{med}}) = \frac{1}{2}$ für den Median.

14.1.5 Gemeinsame Verteilungen

Für zwei Zufallsvariablen x und y definiert man die *gemeinsame* PDF $P(x, y)$ mit $\iint P(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1$. Die **Marginalverteilung** $P_1(x) = \int P(x, y)\,\mathrm{d}y$ ist die Wahrscheinlichkeitsdichte für x unabhängig vom Wert von y. x und y sind *unabhängig* genau dann, wenn $P(x, y) = P_1(x) \cdot P_2(y)$; dann gilt $\mathrm{cov}(x, y) = 0$. Aber: $\mathrm{cov}(x, y) = 0$ impliziert *nicht* Unabhängigkeit!

14.1.6 Fehlerfortpflanzung

Gegeben eine Funktion $y(x_1, \ldots, x_n)$, wobei x_i die Unsicherheit σ_i haben. Durch Taylor-Entwicklung um die Mittelwerte erhält man:

$$\sigma_y = \sqrt{\sum_i \left(\frac{\partial y}{\partial x_i}\right)^2 \sigma_i^2} \qquad \text{(unkorrellierte } x_i\text{)}\,. \tag{14.4}$$

Bei korrelierten Variablen mit Kovarianzmatrix $V_{ij} = \mathrm{cov}(x_i, x_j)$ gilt:

$$\sigma_y = \sqrt{\sum_{ij} \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} V_{ij}}\,. \tag{14.5}$$

Wichtige Konsequenz: *Statistische* Fehler (unkorreliert) mitteln sich mit $1/\sqrt{N}$ heraus, *systematische* Fehler (korreliert) nicht!

Aufgabe 14.1

Zeige mit Gleichung (14.4), dass der Fehler des Mittelwerts $\bar{x} = \frac{1}{N}\sum_i x_i$ von N unabhängigen Messungen mit demselben Fehler σ gleich $\sigma/\sqrt{N}$ ist.

Musterlösung 14.1

Mit $y = \frac{1}{N}\sum_i x_i$ gilt $\frac{\partial y}{\partial x_i} = \frac{1}{N}$ für alle i. Einsetzen in Gleichung (14.4):

$$\sigma_y = \sqrt{\sum_i \frac{1}{N^2}\sigma^2} = \sqrt{\frac{N\sigma^2}{N^2}} = \frac{\sigma}{\sqrt{N}}\,. \tag{14.6}$$

Jede weitere Messung verringert also die Unsicherheit, allerdings nur mit $1/\sqrt{N}$, d.h. um den Fehler zu halbieren, braucht man viermal so viele Messungen.

14.2 Schätztheorie

14.2.1 Schätzer

Ein **Schätzer** $\hat{a}$ ist eine Rechenvorschrift, die aus einer Stichprobe einen Schätzwert für einen Parameter a der zugrundeliegenden Verteilung liefert. Gute Schätzer sind:

- **Konsistent:** $\lim_{N\to\infty} \hat{a} = a$ (konvergiert zum wahren Wert)
- **Erwartungstreu (unbiased):** $\langle \hat{a} \rangle = a$ (kein systematischer Fehler)
- **Effizient:** Minimale Varianz $V(\hat{a})$ unter allen erwartungstreuen Schätzern

Die untere Schranke für die Varianz eines erwartungstreuen Schätzers ist die *Cramér-Rao-Schranke*: $V(\hat{a}) \geq 1/I(a)$, wobei $I(a) = -\langle \mathrm{d}^2 \log L/\mathrm{d}a^2 \rangle$ die **Fisher-Information** ist.

14.2.2 Maximum-Likelihood-Methode

Sei $P(x;a)$ eine PDF mit Parameter a. Für eine Stichprobe $(x_1,\ldots,x_N)$ ist die **Likelihood**:

$$L(\vec{x};a) = \prod_i P(x_i;a)\,. \tag{14.7}$$

Der **Maximum-Likelihood-Schätzer** (ML-Schätzer) $\hat{a}$ maximiert L (äquivalent: minimiert $-\log L$):

$$\left.\frac{\mathrm{d}\log L}{\mathrm{d}a}\right|_{a=\hat{a}} = 0\,. \tag{14.8}$$

Eigenschaften des ML-Schätzers: konsistent, für große N erwartungstreu und effizient. Außerdem *invariant unter Umparametrisierung*: Ist $b = f(a)$, so ist $\hat{b} = f(\hat{a})$.

Die Unsicherheit des ML-Schätzers erhält man aus der Krümmung der Log-Likelihood:

$$V(\hat{a}) \approx \left(-\left.\frac{\mathrm{d}^2\log L}{\mathrm{d}a^2}\right|_{a=\hat{a}}\right)^{-1} \quad\Rightarrow\quad \log L(\hat{a}\pm\sigma_{\hat{a}}) \approx \log L(\hat{a}) - \tfrac{1}{2}\,. \tag{14.9}$$

Aufgabe 14.2

Gegeben sei ein Datensatz von $N = 50$ Messungen, die normalverteilt mit unbekanntem Mittelwert μ und bekannter Streuung $\sigma = 1$ sind. Die Log-Likelihood lautet:

$$\log L(\mu) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 + \text{const}\,.$$

a) Erzeuge $N = 50$ Datenpunkte mit `numpy.random.normal(`$\mu_{\texttt{true}} = 2.5$`, `$\sigma = 1$`, size=50)`.

b) Plotte $\log L(\mu)$ als Funktion von μ für $\mu \in [1,4]$. Markiere das Maximum.

c) Finde den ML-Schätzer analytisch ($\hat{\mu} = \bar{x}$) und numerisch mit `scipy.optimize.minimize_scalar`. Stimmen sie überein?

d) Bestimme das 68%-Konfidenzintervall aus der Bedingung $\log L(\hat{\mu}\pm\sigma_{\hat{\mu}}) = \log L(\hat{\mu}) - \frac{1}{2}$.

Musterlösung 14.2

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import minimize_scalar

mu_true, sigma, N = 2.5, 1.0, 50
rng = np.random.default_rng(42)
x = rng.normal(mu_true, sigma, N)

def log_likelihood(mu):
    return -0.5 / sigma**2 * np.sum((x - mu)**2)

# b) Likelihood-Kurve
mu_grid = np.linspace(1, 4, 300)
ll = [log_likelihood(m) for m in mu_grid]
plt.plot(mu_grid, ll)
```

```
plt.xlabel(r'$\mu$'); plt.ylabel(r'$\log L(\mu)$')

# c) ML-Schaetzer
mu_hat_analytic = x.mean()
res = minimize_scalar(lambda m: -log_likelihood(m), bounds=(1, 4),
↪  method='bounded')
mu_hat_num = res.x
print(f'analytisch: {mu_hat_analytic:.4f}, numerisch: {mu_hat_num:.4f}')

# d) 68%-Konfidenzintervall aus log L = log L_max - 0.5
ll_max = log_likelihood(mu_hat_analytic)
sigma_hat = sigma / np.sqrt(N)      # analytisch: sigma/sqrt(N)
print(f'68%-KI: [{mu_hat_analytic - sigma_hat:.3f}, {mu_hat_analytic +
↪  sigma_hat:.3f}]')
plt.axhline(ll_max - 0.5, color='gray', linestyle='--', label=r'$\log L_\max -
↪  \frac{1}{2}$')
plt.legend(); plt.show()
```

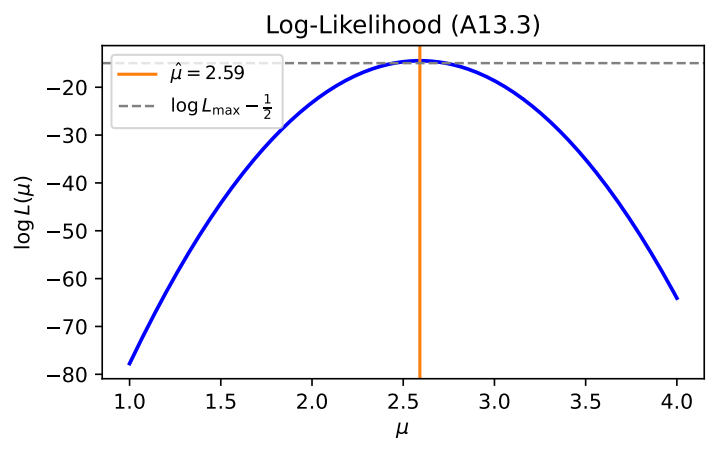


Abbildung 14.1: Log-Likelihood $\log L(\mu)$ als Funktion von μ (blau), ML-Schätzer $\hat{\mu} = \bar{x}$ (orange) und die $-\frac{1}{2}$-Linie (grau gestrichelt) zur Bestimmung des 68 %-KI.

Der analytische und numerische ML-Schätzer stimmen überein ($\hat{\mu} = \bar{x}$). Das 68%-KI hat Breite $\sigma/\sqrt{N} = 1/\sqrt{50} \approx 0.14$.

14.2.3 Methode der kleinsten Quadrate

Ziel: Modell $y = f(x; a)$ an Messdaten $(x_i, y_i \pm \sigma_i)$ anpassen. Da die y_i normalverteilt sind, vereinfacht sich die Log-Likelihood zu:

$$-\log L = \frac{1}{2}\chi^2 + \text{const}, \quad \text{mit} \quad \chi^2 = \sum_i \left(\frac{y_i - f(x_i; a)}{\sigma_i} \right)^2 . \tag{14.10}$$

ML-Schätzer und **kleinste Quadrate** (LS) sind also äquivalent: Minimierung von χ^2 maximiert L. Die Unsicherheit erhält man analog: $\chi^2(\hat{a} \pm \sigma_{\hat{a}}) = \chi^2(\hat{a}) + 1$.

14.2.4 Aufgabe: Messungen kombinieren

Aufgabe 14.3

Eine Naturkonstante y wurde von n unabhängigen Experimenten mit Ergebnissen $y_i \pm \sigma_i$ gemessen.

a) Zeige mit der LS-Methode ($\chi^2 = \sum_i (y_i - y_0)^2/\sigma_i^2$, Minimierung nach y_0), dass der optimale Schätzer und seine Unsicherheit gegeben sind durch:

$$\hat{y} = \frac{\sum_i w_i y_i}{\sum_i w_i}, \quad \frac{1}{\sigma_{\hat{y}}^2} = \sum_i \frac{1}{\sigma_i^2}, \quad \text{mit } w_i = \frac{1}{\sigma_i^2}\,. \tag{14.11}$$

b) ATLAS misst die Higgs-Boson-Masse als $124.88 \pm 0.37\,\text{GeV}$, CMS als $125.26 \pm 0.20\,\text{GeV}$. Was ist die kombinierte Messung?

c) Planck misst $H_0 = 67.4 \pm 0.5\,\text{km/s/Mpc}$, SH0ES misst $H_0 = 73.04 \pm 1.04\,\text{km/s/Mpc}$. Berechne den kombinierten Schätzer. Was fällt auf?

Musterlösung 14.3

a) Die Bedingung $\mathrm{d}\chi^2/\mathrm{d}y_0 = 0$ ergibt:

$$-2\sum_i \frac{y_i - \hat{y}}{\sigma_i^2} = 0 \;\Rightarrow\; \hat{y} = \frac{\sum_i y_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}\,. \tag{14.12}$$

Die Unsicherheit folgt aus $V(\hat{y}) = \left(\frac{1}{2}\,\mathrm{d}^2\chi^2/\mathrm{d}y_0^2\right)^{-1} = \left(\sum_i 1/\sigma_i^2\right)^{-1}$.
b) $w_{\text{ATLAS}} = 1/0.37^2 = 7.31$, $w_{\text{CMS}} = 1/0.20^2 = 25.0$. Damit $\hat{m}_h = (7.31 \cdot 124.88 + 25.0 \cdot 125.26)/32.31 = \mathbf{125.17 \pm 0.18}\,\text{GeV}$. Die kombinierte Messung ist präziser als beide Einzelmessungen.
c) $w_{\text{Planck}} = 1/0.5^2 = 4.0$, $w_{\text{SH0ES}} = 1/1.04^2 = 0.925$. Damit $\hat{H}_0 = (4.0 \cdot 67.4 + 0.925 \cdot 73.04)/4.925 \approx \mathbf{68.5 \pm 0.45}\,\text{km/s/Mpc}$. Der kombinierte Wert liegt zwischen den Messungen, aber beide weichen um mehrere σ voneinander ab: Das ist die H_0-Spannung (vgl. Kapitel 2)!

14.2.5 Die χ^2-Verteilung

Für ein Modell mit k freien Parametern und l Datenpunkten gilt: χ^2_{min} folgt einer **χ^2-Verteilung** mit $n = l - k$ Freiheitsgraden und Erwartungswert $\langle\chi^2\rangle = n$. Als Faustregel:

$$\frac{\chi^2_{\text{min}}}{n} \approx 1 \quad \text{(guter Fit)}\,. \tag{14.13}$$

Ein deutlich größerer Wert deutet auf ein falsches Modell oder unterschätzte Fehler hin; ein deutlich kleinerer auf überschätzte Fehler. Abbildung 14.2 zeigt die χ^2-Verteilung für verschiedene Freiheitsgrade.

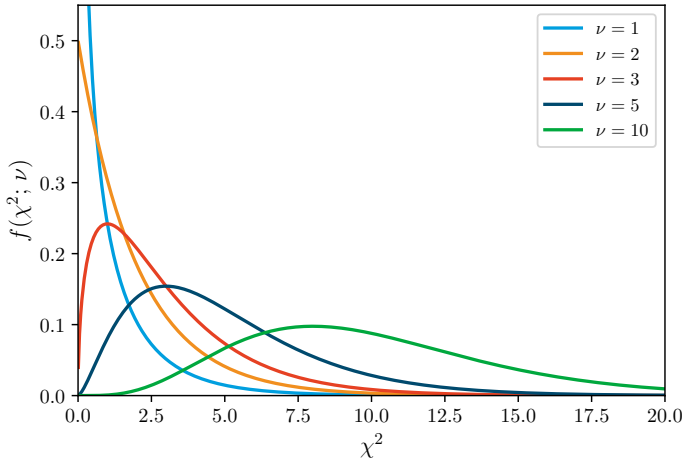


Abbildung 14.2: χ^2-Verteilung $f(\chi^2;\nu)$ für verschiedene Freiheitsgrade ν. Mit wachsendem ν verschiebt sich der Peak zu höheren Werten und die Verteilung wird symmetrischer (CLT: für $\nu \to \infty$ konvergiert sie zur Normalverteilung um ν). Der Erwartungswert ist stets $\langle \chi^2 \rangle = \nu$.

14.3 Konfidenzintervalle

14.3.1 Deskriptive Intervalle

Für eine normalverteilte Messgröße $x \sim \mathcal{N}(\mu, \sigma)$ enthält das Intervall $[\mu - k\sigma, \mu + k\sigma]$ die Wahrscheinlichkeit $C = 2F(k) - 1$ (mit F der CDF der Normalverteilung). Umgekehrt: Für ein vorgegebenes *Konfidenzniveau* C bestimmt man k so, dass $\int_{x_-}^{x_+} P(x)\,\mathrm{d}x = C$.

14.3.2 Konfidenzintervalle aus ML und LS

Für normalverteilte $\hat{a}$ (gültig für große N): Das 68%-Konfidenzintervall ergibt sich direkt aus $\sigma_{\hat{a}}$. Allgemein für Konfidenzniveau C:

$$\log L(a_{\min}) = \log L(a_{\max}) = \log L(\hat{a}) - \frac{N^2}{2}, \quad N = \sqrt{2}\,\mathrm{erf}^{-1}(C)\,. \tag{14.14}$$

Für LS analog: $\chi^2(a_{\min}) = \chi^2(a_{\max}) = \chi^2(\hat{a}) + N^2$.

14.3.3 Mehrdimensionale Konfidenzregionen

Für mehrere Parameter $\hat{\vec{a}}$ ist die Konfidenzregion der Bereich, für den

$$-2\,\Delta \log L(\vec{a}) = -2[\log L(\vec{a}) - \log L(\hat{\vec{a}})] \leq Q_C(n)\,, \tag{14.15}$$

wobei $Q_C(n)$ das C-Quantil der χ^2-Verteilung mit n (= Anzahl der Parameter) Freiheitsgraden ist. Für zwei Parameter und $C = 68\%$ ist $Q = 2.30$, für $C = 95\%$ ist $Q = 5.99$.

14.4 Hypothesentests

14.4.1 Grundbegriffe

Eine **Hypothese** ist eine Aussage über eine PDF, einfach (legt sie vollständig fest) oder zusammengesetzt (erlaubt mehrere PDFs).

- **Signifikanz** α: Wahrscheinlichkeit, H zu verwerfen, obwohl H wahr ist (Fehler 1. Art).
- **Power** $1-\beta$: Wahrscheinlichkeit, H zu verwerfen, wenn H falsch ist; β ist die Wahrscheinlichkeit für Fehler 2. Art.
- **Teststatistik:** Eine Funktion $t(\vec{x})$ der Daten, mit der die Entscheidung getroffen wird.

14.4.2 Neyman-Pearson-Test

Für zwei einfache Hypothesen H und A ist der optimale Test (minimiert β bei festem α) der **Likelihood-Quotienten-Test**:

$$t(\vec{x}) = \frac{L_A(\vec{x})}{L_H(\vec{x})} = \frac{\prod_i P_A(x_i)}{\prod_i P_H(x_i)}\,. \tag{14.16}$$

Verwirf H, falls $t > t_{\max}$ (wobei $t_{\max}$ durch α festgelegt wird).

14.4.3 p-Werte und Look-elsewhere-Effekt

Der **p-Wert** ist die Wahrscheinlichkeit, ein Ergebnis mindestens so extrem wie das beobachtete zu erhalten, unter der Annahme dass H wahr ist:

$$p = \int_{t_{\text{obs}}}^{\infty} P_H(t)\,\mathrm{d}t\,. \tag{14.17}$$

In der Teilchenphysik spricht man von *Entdeckung* erst ab $p < 2.87 \times 10^{-7}$ (5σ).

Look-elsewhere-Effekt: Testet man n unabhängige Hypothesen, ist die Wahrscheinlichkeit für *mindestens ein* zufällig signifikantes Ergebnis $\approx n\alpha$ statt α. Korrektur: lokaler p-Wert muss durch den *Trialfaktor* n dividiert werden.

14.4.4 χ^2-Anpassungstest

Hypothese H_0: Daten folgen Modell $f(x;\vec{a})$. Teststatistik:

$$\chi^2_{\min} = \sum_i \left(\frac{y_i - f(x_i;\hat{\vec{a}})}{\sigma_i}\right)^2\,, \tag{14.18}$$

folgt unter H_0 einer χ^2-Verteilung mit $n = l-k$ Freiheitsgraden (l Messpunkte, k freie Parameter). Man berechnet den p-Wert und verwirft H_0 bei $p < \alpha$.

14.4.5 Likelihood-Quotienten-Test und Wilks' Theorem

Für eine Nullhypothese $a = a_0$ gegen $a \neq a_0$ in einem Modell $f(x;a)$:

$$q = -2\log\frac{L(a_0)}{L(\hat{a})}\,. \tag{14.19}$$

Wilks' Theorem: Falls H_0 korrekt ist, folgt q einer χ^2-Verteilung mit k Freiheitsgraden (= Anzahl der unter H_0 fixierten Parameter). Damit kann q direkt als Teststatistik verwendet werden.

Nuisance-Parameter: Oft hängt das Modell von uninteressanten Parametern θ (Untergrundnormierung, Kalibrationsunsicherheiten etc.) ab. Man maximiert die Likelihood über θ für jeden Wert von a – das ergibt die **Profil-Likelihood** $L(a) = L(a,\hat{\theta}(a))$ und das Profil-Likelihood-Verhältnis $\Lambda = L(a_0)/L(\hat{a})$. Wilks' Theorem gilt auch hier, mit k = Anzahl der *freien* (nicht als Nuisance-Parameter behandelten) Parameter.

14.5 Bayessche Statistik

14.5.1 Satz von Bayes

Die bayessche Statistik weist Wahrscheinlichkeiten direkt Modellparametern zu; sie ist eine *subjektive* Wahrscheinlichkeitsinterpretation („Grad der Überzeugung"). Grundlage ist der **Satz von**

Bayes:

$$\underbrace{p(\theta \mid D)}_{\text{Posterior}} = \frac{\underbrace{p(D \mid \theta)}_{\text{Likelihood}} \cdot \underbrace{p(\theta)}_{\text{Prior}}}{\underbrace{p(D)}_{\text{Evidenz}}} \propto p(D \mid \theta) \cdot p(\theta)\,. \tag{14.20}$$

- **Prior** $p(\theta)$: Vorwissen über θ *vor* der Messung (z.B. $\theta > 0$ aus physikalischen Gründen).
- **Likelihood** $p(D \mid \theta)$: Wahrscheinlichkeit der Daten D für gegebene Parameter θ (entspricht $L(\vec{x}; \theta)$ aus dem ML-Formalismus).
- **Posterior** $p(\theta \mid D)$: Aktualisiertes Wissen nach Beobachtung von D.

Das Ergebnis sind **Kredibilitätsintervalle** $[H_{\min}, H_{\max}]$ mit $\int_{H_{\min}}^{H_{\max}} p(\theta \mid D)\,\mathrm{d}\theta = C$. Diese haben die intuitivere Interpretation: „θ liegt mit Wahrscheinlichkeit C in diesem Bereich“.

14.5.2 Wahl des Priors

- **Gleichverteilungs-Prior** (*flat prior*): $p(\theta) = \text{const}$ auf dem erlaubten Bereich. Geeignet, wenn keine Information vorliegt und eine Verschiebung $\theta \to \theta + c$ das Problem nicht ändert.
- **Log-Prior**: $p(\theta) \propto 1/\theta$. Geeignet für Größenordnungsprobleme (Skaleninvarianz: $p(1 \leq \theta \leq 2) = p(100 \leq \theta \leq 200)$).
- Wichtig: Der Prior ist immer subjektiv. Für große Datensätze überwiegt die Likelihood, und die Abhängigkeit vom Prior verschwindet.

14.5.3 Marginalisierung

Hängt das Modell von mehreren Parametern (θ, ϕ) ab, interessiert man sich aber nur für θ (*Nuisance-Parameter* ϕ), so erhält man die **marginalisierte Posterior**:

$$p(\theta \mid D) = \int p(\theta, \phi \mid D)\,\mathrm{d}\phi\,. \tag{14.21}$$

Dies ist das bayessche Analogon zur Profil-Likelihood im frequentistischen Formalismus.

14.5.4 Modellvergleich

Gegeben zwei Modelle A und B, berechnet man den **Bayes-Faktor**:

$$K = \frac{p(A \mid D)}{p(B \mid D)} = \frac{p(D \mid A)}{p(D \mid B)} \quad (\text{gleiche Modell-Priors } p(A) = p(B))\,, \tag{14.22}$$

wobei $p(D \mid A) = \int p(D \mid \theta, A)\, p(\theta \mid A)\,\mathrm{d}\theta$ die *Bayessche Evidenz* ist. Interpretation nach der *Jeffreys-Skala*: $K > 10$ ist starke Evidenz für A, $K > 100$ entscheidend.

Dies verkörpert **Ockhams Rasiermesser**: Erklären zwei Modelle die Daten gleich gut, wird das einfachere bevorzugt, weil ein komplexeres Modell die Evidenz über einen größeren Parameterraum „verteilt“.

Frequentistisch vs. Bayesianisch. Beide Ansätze haben ihre Stärken (Tabelle 14.1):

	Frequentistisch	**Bayesianisch**
Parameterbestimmung	ML + Konfidenzintervalle	Posterior + Kredibilitätsintervalle
Nuisance-Parameter	Profil-Likelihood	Marginalisierung
Modellentscheidung	Hypothesentests	Bayes-Faktor

Tabelle 14.1: Vergleich frequentistischer und bayesianischer Statistik.

Aufgabe 14.4

Wir führen eine einfache Bayessche Inferenz für den Mittelwert μ einer Normalverteilung durch, diesmal *auf einem Gitter* statt analytisch.
Gegeben: $N = 10$ Messungen x_i mit bekannter Streuung $\sigma = 1$, Prior $p(\mu) = \text{const}$ auf $\mu \in [-5, 5]$.

a) Erzeuge $N = 10$ Datenpunkte mit `numpy.random.normal(`$\mu_{\texttt{true}} = 1.5$`, `$\sigma = 1$`, size=10)`.

b) Berechne auf einem feinen Gitter $\mu \in [-5, 5]$ die Likelihood $L(\mu) = \prod_i \mathcal{N}(x_i; \mu, 1)$ (arbeite mit der Log-Likelihood und nehme am Ende `np.exp` um numerische Probleme zu vermeiden). Normiere zur Posterior.

c) Plotte Prior, Likelihood und Posterior übereinander. Was passiert mit der Posterior, wenn du $N = 1$ vs. $N = 100$ Datenpunkte verwendest?

d) Berechne das 95%-Kredibilitätsintervall als das kürzeste Intervall, das 95% der Posterior-Wahrscheinlichkeit enthält.

Musterlösung 14.4

```
import numpy as np
import matplotlib.pyplot as plt

mu_true, sigma = 1.5, 1.0
rng = np.random.default_rng(0)
x = rng.normal(mu_true, sigma, size=10)

mu_grid = np.linspace(-5, 5, 1000)
dm = mu_grid[1] - mu_grid[0]

# Log-Likelihood (Summe ueber Datenpunkte)
log_L = np.array([-0.5 * np.sum((x - m)**2) / sigma**2 for m in mu_grid])
log_L -= log_L.max()                # numerische Stabilitaet
posterior = np.exp(log_L)
posterior /= posterior.sum() * dm   # Normierung

# Flat Prior
prior = np.ones_like(mu_grid) / 10.0

plt.plot(mu_grid, prior / prior.max(), '--', label='Prior (flach)')
plt.plot(mu_grid, np.exp(log_L) / np.exp(log_L).max(), label='Likelihood')
plt.plot(mu_grid, posterior / posterior.max(), label='Posterior')
plt.axvline(mu_true, color='k', linestyle=':', label=r'wahrer Wert $\mu$')
plt.xlabel(r'$\mu$'); plt.legend(); plt.show()
```

```
# 95%-Kredibilitaetsintervall (HDI via kumulierter Posterior)
cdf = np.cumsum(posterior) * dm
lo = mu_grid[np.searchsorted(cdf, 0.025)]
hi = mu_grid[np.searchsorted(cdf, 0.975)]
print(f'95%-Kredibilitaetsintervall: [{lo:.2f}, {hi:.2f}]')
```

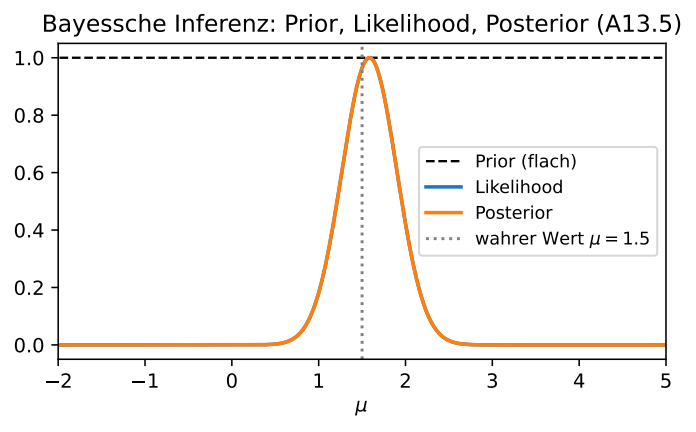


Abbildung 14.3: Prior (gestrichelt), Likelihood und Posterior für $N = 10$ Messungen. Die Posterior ist schärfer als die Likelihood allein und zentriert auf den wahren Wert.

Mit mehr Daten wird die Posterior immer schärfer und zentriert sich auf den wahren Wert: Das ist das bayessche Lernen. Bei $N \to \infty$ stimmt das Kredibilitätsintervall mit dem frequentistischen Konfidenzintervall überein; für kleine N kann es abweichen, je nach Prior.

Aufgabe 14.5

Bayes-Modellvergleich: Welches Polynom passt am besten?
Wir haben $m = 12$ verrauschte Messpunkte (x_i, y_i), die aus einem Polynom dritten Grades mit Gaußschem Rauschen ($\sigma = 0.4$) erzeugt wurden. Wir wollen herausfinden, welcher Polynomgrad $n = 0, 1, 2, 3, 4, 5$ am besten zu den Daten passt, ohne dabei zu überfitten.

a) Erzeuge den Datensatz:

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(42)
m = 12
x = np.linspace(-2, 2, m)
sigma = 0.4
y = 0.35 * x**3 - 0.8 * x**2 + 0.4 * x + 1.2 + sigma * np.random.randn(m)
```

Fitte für jedes $n \in \{0, 1, 2, 3, 4, 5\}$ ein Polynom mit `np.polyfit(x, y, n)` und plotte alle Fits zusammen mit den Daten.

b) Unser Datenmodell nimmt an, dass die Messwerte y_i um den wahren Polynomwert herum *gaußsch verrauscht* sind: $y_i = f(x_i) + \varepsilon_i$ mit $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, unabhängig und mit bekanntem σ. Der Fit vom Grad n aus a) liefert an jeder Stelle x_i eine Vorhersage $\hat{y}_i$ – das ist einfach der Wert, den das gefittete Polynom n-ten Grades an der Stelle x_i annimmt (in `numpy` erhältst du die $\hat{y}_i$ mit `np.polyval(coeffs, x)`, wobei `coeffs` die

von `np.polyfit` zurückgegebenen Koeffizienten sind). Die Log-Likelihood der Daten unter diesem Modell ist

$$\ln L = \sum_i \ln \mathcal{N}(y_i; \hat{y}_i, \sigma^2) = -\frac{1}{2\sigma^2} \sum_i (y_i - \hat{y}_i)^2 - \frac{m}{2} \ln(2\pi\sigma^2)\,. \tag{14.23}$$

Da σ als bekannt vorausgesetzt wird, hängt nur der erste Term von den Fit-Parametern ab – die Likelihood zu maximieren ist also äquivalent dazu, $\sum_i (y_i - \hat{y}_i)^2$ zu minimieren. Genau das tut `np.polyfit` bereits (kleinste Quadrate): Der Fit aus a) ist also schon der Maximum-Likelihood-Fit für den jeweiligen Grad n. **Implementiere eine Funktion** `log_likelihood(y, y_hat, sigma)`, die obige Formel berechnet, **und werte sie für jedes n am zugehörigen Fit aus a) aus** (also mit den $\hat{y}_i =$ `np.polyval(coeffs, x)` des jeweiligen Polynomfits vom Grad n).

Das Bayessche Informationskriterium (BIC) approximiert die Log-Evidenz direkt über diese maximale Log-Likelihood $\ln L_{\max}(n)$:

$$\ln Z_n \approx \ln L_{\max}(n) - \frac{n+1}{2} \ln m\,, \tag{14.24}$$

wobei $n+1$ die Anzahl freier Parameter des Polynoms vom Grad n ist (die $n+1$ Koeffizienten). **Berechne den Log-Bayes-Faktor** relativ zum Nullmodell ($n = 0$): $\ln B_{n0} = \ln Z_n - \ln Z_0$ **und plotte ihn als Funktion von n.**

c) Welcher Polynomgrad wird durch den Bayes-Faktor bevorzugt? Die maximale Log-Likelihood $\ln L_{\max}(n)$ kann mit wachsendem n nur größer werden oder gleich bleiben, nie kleiner – wieso? Warum sorgt der Straf-Term $-\frac{n+1}{2} \ln m$ trotzdem dafür, dass nicht einfach das komplexeste Modell ($n = 5$) gewinnt?

d) Die BIC-Näherung aus b) ist nur eine (asymptotische) Näherung für die tatsächliche Bayessche Evidenz $Z_n = \int p(D \mid \theta, n)\, p(\theta \mid n)\, \mathrm{d}\theta$ – das Integral über den gesamten Parameterraum, gewichtet mit dem Prior $p(\theta \mid n)$. Um dieses Integral tatsächlich auszuwerten, brauchst du (i) einen expliziten Prior für die $n+1$ Polynomkoeffizienten θ und (ii) ein Verfahren, das dieses (oft hochdimensionale) Integral numerisch berechnet – *Nested Sampling* ist genau dafür gebaut (im Unterschied zu MCMC, das dir Stichproben aus der Posterior liefert, aber nicht direkt Z). **Installiere dazu das Paket** `dynesty` (`pip install dynesty` bzw. `conda install -c conda-forge dynesty`).

`dynesty` erwartet den Prior nicht als Dichte, sondern als `prior_transform(u)`: eine Funktion, die einen Punkt $u \in [0,1]^{n+1}$ aus dem Einheitswürfel auf einen Parametervektor θ abbildet, sodass θ gemäß deines gewünschten Priors verteilt ist. Für einen flachen (uniformen) Prior $\theta_k \sim \mathcal{U}(-a, a)$ auf jedem Koeffizienten ist das einfach $\theta_k = a\,(2u_k - 1)$. **Wähle einen sinnvollen Wert für a** (die wahren Koeffizienten liegen alle im Bereich $[-1, 2]$, sodass $a = 5$ großzügig, aber nicht unsinnig ist) **und implementiere** `prior_transform(u)` **sowie** `loglike(theta)` (Letzteres ruft einfach deine `log_likelihood`-Funktion aus b) auf, mit $\hat{y}_i =$ `np.polyval(theta, x)`). **Erzeuge für jedes n einen** `dynesty.NestedSampler` **mit** `ndim=n+1`**, führe** `sampler.run_nested()` **aus und lies die tatsächliche Log-Evidenz** `sampler.results.logz[-1]` **aus** (dynesty liefert zusätzlich eine geschätzte Unsicherheit `results.logzerr[-1]` – die Schätzung basiert ja selbst auf einer endlichen Anzahl Stichproben und ist daher mit einem Fehler behaftet). **Berechne damit den tatsächlichen Log-Bayes-Faktor $\ln B_{n0}$ und vergleiche ihn in einem Plot mit der BIC-Näherung aus b).** Stimmen die beiden Kurven überein? Welchen Grad n bevorzugt die tatsächliche Evidenz?

Musterlösung 14.5

a)

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(42)
m = 12
x = np.linspace(-2, 2, m)
sigma = 0.4
y = 0.35 * x**3 - 0.8 * x**2 + 0.4 * x + 1.2 + sigma * np.random.randn(m)

ns = [0, 1, 2, 3, 4, 5]
x_plot = np.linspace(-2, 2, 200)

fig, ax = plt.subplots(figsize=(6, 4))
ax.plot(x, y, 'o', color='k', label='Daten')
for n in ns:
    coeffs = np.polyfit(x, y, n)
    ax.plot(x_plot, np.polyval(coeffs, x_plot), label=f'n={n}')
ax.set_xlabel('x'); ax.set_ylabel('y'); ax.legend(fontsize=8)
plt.tight_layout(); plt.show()
```

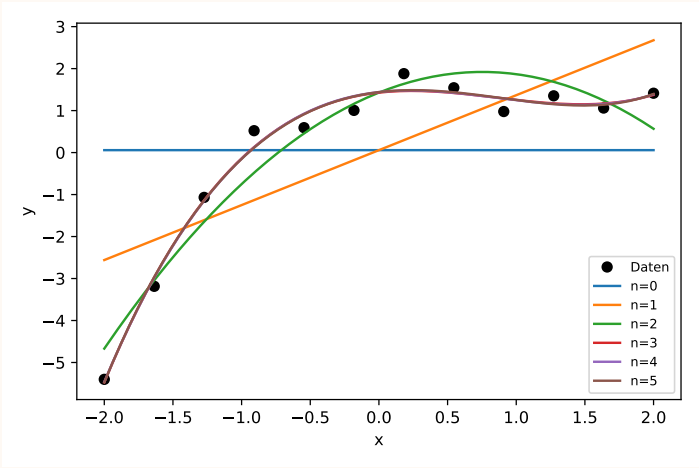


Abbildung 14.4: Datenpunkte und Polynomfits vom Grad $n = 0, \ldots, 5$. Die Fits ab $n = 3$ liegen praktisch übereinander – höhere Grade bringen kaum noch eine sichtbare Verbesserung.

b)

```
def log_likelihood(y, y_hat, sigma):
    m = len(y)
    return -0.5 * np.sum((y - y_hat)**2) / sigma**2 - 0.5 * m * np.log(2 *
    ↪  np.pi * sigma**2)

lnZ = np.zeros(len(ns))
for i, n in enumerate(ns):
    coeffs = np.polyfit(x, y, n)
    y_hat = np.polyval(coeffs, x)
    lnL_max = log_likelihood(y, y_hat, sigma)
    lnZ[i] = lnL_max - 0.5 * (n + 1) * np.log(m)

lnB = lnZ - lnZ[0]
```

```
fig, ax = plt.subplots(figsize=(6, 4))
ax.plot(ns, lnB, 'o-')
ax.axhline(0, color='gray', ls='--')
ax.set_xlabel('Polynomgrad n'); ax.set_ylabel(r'$\ln B_{n0}$')
plt.tight_layout(); plt.show()
```

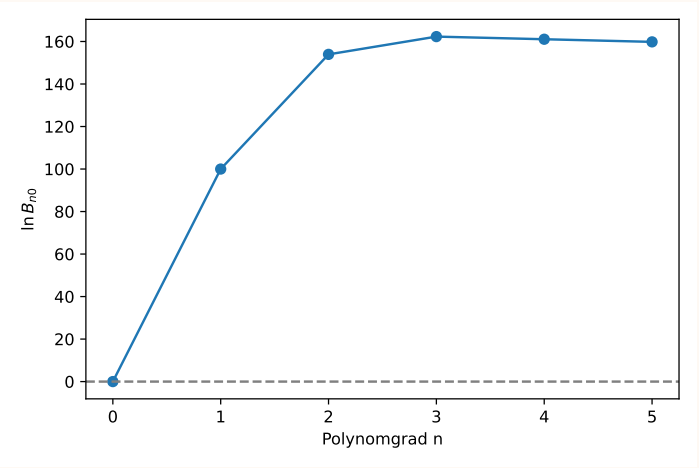


Abbildung 14.5: Log-Bayes-Faktor $\ln B_{n0}$ (BIC-Näherung) relativ zu $n = 0$. Das Maximum liegt bei $n = 3$, dem wahren Grad des erzeugenden Polynoms.

c) Für diesen (mit `seed=42` reproduzierbaren) Datensatz ergibt sich $\ln B_{n0} \approx 0, 100, 154, 162, 161, 160$ für $n = 0, \ldots, 5$ – der Bayes-Faktor bevorzugt also $n = 3$, den *wahren* Grad des erzeugenden Polynoms.
$\ln L_{\max}(n)$ kann mit wachsendem n nie kleiner werden, weil ein Polynom vom Grad n jedes Polynom niedrigeren Grades als Spezialfall enthält (z. B. mit Koeffizient 0 vor x^n) – die kleinste-Quadrate-Lösung kann also immer mindestens so gut sein wie die für $n - 1$. Tatsächlich sieht man das auch numerisch: $\ln L_{\max}$ steigt von $n = 0$ bis $n = 3$ stark an, wird ab $n = 3$ aber praktisch flach ($-2.38 \to -2.38 \to -2.37$ für $n = 3, 4, 5$) – die zusätzlichen Parameter bei $n = 4, 5$ verbessern den Fit kaum noch, weil das Rauschen bereits bei $n = 3$ ausgeschöpft ist. Der Straf-Term $-\frac{n+1}{2} \ln m$ wächst hingegen *linear* in n und wiegt diesen marginalen Gewinn an Likelihood ab $n = 3$ nicht mehr auf: Ockhams Rasiermesser bevorzugt daher das einfachste Modell, das die Daten noch gut erklärt, statt eines unnötig komplexeren Modells mit vernachlässigbarem Likelihood-Zugewinn.
d)

```
from dynesty import NestedSampler

a = 5.0  # Prior: theta_k ~ U(-a, a) fuer jeden Koeffizienten

def make_loglike(n):
    def loglike(theta):
        y_hat = np.polyval(theta, x)
        return log_likelihood(y, y_hat, sigma)
    return loglike

def prior_transform(u):
    return a * (2 * u - 1)

lnZ_ns = np.zeros(len(ns))
```

```
for i, n in enumerate(ns):
    sampler = NestedSampler(make_loglike(n), prior_transform, ndim=n + 1,
                            nlive=200, bound='multi', sample='rwalk')
    sampler.run_nested(print_progress=False)
    lnZ_ns[i] = sampler.results.logz[-1]

lnB_ns = lnZ_ns - lnZ_ns[0]

fig, ax = plt.subplots(figsize=(6, 4))
ax.plot(ns, lnB, 'o-', label='BIC-Näherung')
ax.plot(ns, lnB_ns, 's--', label='dynesty (exakt)')
ax.axhline(0, color='gray', ls='--')
ax.set_xlabel('Polynomgrad n'); ax.set_ylabel(r'$\ln B_{n0}$'); ax.legend()
plt.tight_layout(); plt.show()
```

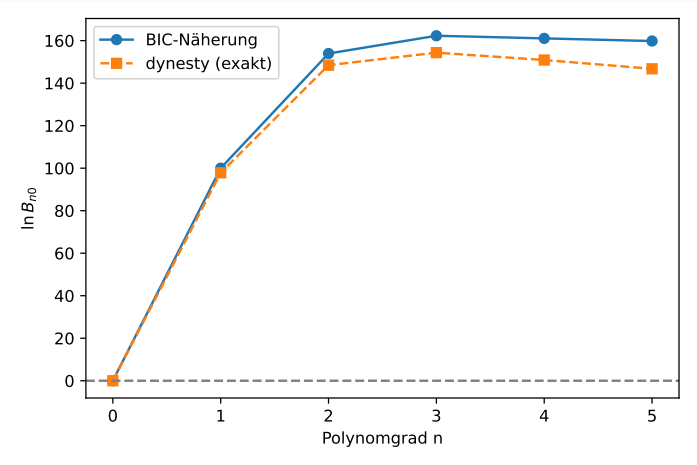


Abbildung 14.6: Vergleich des Log-Bayes-Faktors $\ln B_{n0}$ aus der BIC-Näherung (b) mit der tatsächlichen, über Nested Sampling berechneten Evidenz (d). Beide bevorzugen $n = 3$; die exakten Werte liegen wegen des Ockham-Faktors durchweg etwas unter der BIC-Näherung.

Für den Datensatz aus a) ergibt sich $\ln B_{n0}^{\text{dynesty}} \approx 0,\ 97,\ 149,\ 154,\ 150,\ 146$ für $n = 0, \ldots, 5$ (mit einer geschätzten Unsicherheit von jeweils $\lesssim 0.4$ in $\ln Z_n$, also viel kleiner als die Abstände zwischen den n) – die tatsächliche Evidenz bevorzugt also ebenfalls $n = 3$, in guter Übereinstimmung mit der BIC-Näherung aus b) ($\ln B_{n0} \approx 0,\ 100,\ 154,\ 162,\ 161,\ 160$). Die beiden Kurven haben dieselbe Form und dasselbe Maximum bei $n = 3$; die tatsächlichen Werte liegen aber durchweg etwas *unter* der BIC-Näherung, besonders für $n \geq 2$. Das liegt am sogenannten *Ockham-Faktor*: Das echte Integral bestraft zusätzliche Parameter nicht nur über deren Anzahl (wie der Term $-\frac{n+1}{2} \ln m$ es pauschal tut), sondern auch dafür, dass ihr Prior-Volumen $(2a)^{n+1}$ mit jedem weiteren Parameter wächst, während nur ein kleiner Teil davon tatsächlich zu den Daten passt – ein Effekt, den die BIC-Näherung komplett ignoriert (sie hängt gar nicht von a ab). Für einen deutlich größeren Prior (größeres a) würde die tatsächliche Evidenz komplexerer Modelle daher noch stärker absinken, während sich an der BIC-Näherung nichts ändern würde.

Im letzten Kapitel werden wir die bayessche Methode direkt anwenden: Der Markov-Chain-Monte-Carlo-Algorithmus (Kapitel 15) liefert Samples aus der Posterior $p(\theta \mid D)$, ohne dass wir das hochdimensionale Integral der Evidenz berechnen müssen.

15 Monte-Carlo-Methoden

Wir stehen jetzt vor folgendem Problem: Wir haben 14 Datenpunkte (Aufgabe 17.1) und ein physikalisches Modell mit ein paar freien Parametern θ, z.B. $\theta = (\alpha, \beta/H, T_*)$ für das Phasenübergangsspektrum aus Kapitel 18, oder $\theta = (\Omega_{\mathrm{mat}}, \Omega_\Lambda, \dots)$ für das ΛCDM-Modell aus Kapitel 8.6. Wir wollen wissen: *Welche Werte von θ passen am besten zu den Daten, und mit welcher Unsicherheit?* Die Antwort liefert die *Bayessche Statistik*, kombiniert mit einer cleveren Computersimulation: dem *Markov-Chain-Monte-Carlo-* bzw. *Metropolis-Hastings*-Algorithmus.

> **Hinweis zu den Abbildungen:** Die Plots in den Musterlösungen dieses Kapitels werden von euch selbst im Rahmen der Aufgaben erzeugt. Die Musterlösungen zeigen jeweils ein Referenzbeispiel – eure Plots werden ähnlich aussehen, aber aufgrund von Zufallszahlen leicht abweichen.

15.1 Zufallszahlen aus beliebigen Verteilungen

Bevor wir uns dem MCMC-Algorithmus widmen, schauen wir uns an, wie man Computern beibringt, Zufallszahlen nach einer *beliebigen* Verteilung zu erzeugen. Alle Programmiersprachen (und auch `numpy`) liefern standardmäßig Zufallszahlen mit *Gleichverteilung* auf $[0, 1]$. Wie kommt man davon zu anderen Verteilungen?

15.1.1 Transformationsmethode

Grundidee: Ist x gleichverteilt auf $[0, 1]$ und $y(x)$ eine monotone Funktion, dann folgt y einer Verteilung $g(y)$. Das Verhältnis der Wahrscheinlichkeitsdichten bleibt erhalten:

$$g(y)\,\mathrm{d}y = \mathrm{d}x \quad \Rightarrow \quad y(x) = F^{-1}(x)\,, \tag{15.1}$$

wobei F die kumulative Verteilungsfunktion der Zielverteilung ist. Man invertiert also die CDF analytisch.

Beispiel: Erzeuge Zufallszahlen mit PDF $f(x) = \frac{1}{2}\sin x$ für $x \in [0, \pi]$. Die CDF ist $F(x) = \int_0^x \frac{1}{2}\sin x'\,\mathrm{d}x' = \frac{1}{2}(1 - \cos x)$. Auflösen nach x: $y(u) = F^{-1}(u) = 2\arcsin(\sqrt{u})$. Zieht man u gleichverteilt in $[0, 1]$ und berechnet $y(u)$, so folgt y der gewünschten Verteilung.

Aufgabe 15.1

Zeige mit der Transformationsmethode, wie man aus gleichverteilten Zufallszahlen $u \in [0, 1]$ Zufallszahlen mit Exponentialverteilung $f(t) = \lambda\,\mathrm{e}^{-\lambda t}$ $(t \geq 0)$ erzeugt. *Hinweis: berechne zuerst $F(t)$ und löse dann $F(t) = u$ nach t auf.*

Musterlösung 15.1

Die CDF der Exponentialverteilung: $F(t) = \int_0^t \lambda\,\mathrm{e}^{-\lambda t'}\,\mathrm{d}t' = 1 - \mathrm{e}^{-\lambda t}$. Setze $F(t) = u$:

$$1 - \mathrm{e}^{-\lambda t} = u \;\Rightarrow\; t = -\frac{\ln(1-u)}{\lambda}\,. \tag{15.2}$$

Da $1 - u$ für gleichverteiltes u ebenfalls gleichverteilt ist, kann man auch kürzer schreiben: $t = -\ln(u)/\lambda$.

15.1.2 Verwerfungsmethode (Acceptance-Rejection)

Ist die CDF nicht analytisch invertierbar, nutzt man einen geometrischen Ansatz:

1. Wähle eine obere Schranke $f_{\max}$ mit $f(x) \le f_{\max}$ auf $[x_{\min}, x_{\max}]$.
2. Ziehe einen gleichverteilten Vorschlag x aus $[x_{\min}, x_{\max}]$.
3. Ziehe eine zweite gleichverteilte Zahl y aus $[0, f_{\max}]$.
4. Akzeptiere x, falls $y < f(x)$; andernfalls verwerfe und wiederhole.

Die Wahrscheinlichkeit, x zu akzeptieren, ist proportional zu $f(x)$; der Datensatz der akzeptierten Werte folgt also genau der Zielverteilung $f(x)$. Nachteil: Je kleiner der Anteil akzeptierter Vorschläge ($f_{\max} \cdot (x_{\max} - x_{\min}) \gg \int f \,\mathrm{d}x$), desto ineffizienter die Methode.

15.2 Von χ^2 zur Likelihood

Aufgabe 15.2

Viele Experimente messen N unabhängige Größen: Messwerte μ_k mit Unsicherheiten σ_k, $k = 1, \ldots, N$. Jede Messung ist Gauß-verteilt. Ein Modell liefert Vorhersagen x_k; wie gut passt das Modell? Aus der Schule kennt ihr die χ^2-Methode:

$$\chi^2 = \sum_{k=1}^{N} \left(\frac{x_k - \mu_k}{\sigma_k} \right)^2 .$$

Je kleiner χ^2, desto besser das Modell.

a) Die Wahrscheinlichkeitsdichte, einen Modellwert x_k bei Messung (μ_k, σ_k) zu beobachten, ist

$$\mathcal{L}_k(x_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{1}{2} \left(\frac{x_k - \mu_k}{\sigma_k} \right)^2 \right).$$

Zeige, dass $\log \mathcal{L}_k(x_k) = -\frac{1}{2} \left(\frac{x_k - \mu_k}{\sigma_k} \right)^2 - \frac{1}{2} \log(2\pi\sigma_k^2)$, also bis auf eine Konstante gerade $-\frac{1}{2}$ mal der k-te χ^2-Summand.

b) Die N Messungen sind unabhängig: die Gesamt-Likelihood ist das *Produkt* der Einzel-Likelihoods. Was wird aus dem Produkt, wenn man den Logarithmus nimmt? Warum ist das numerisch vorteilhaft?

c) Implementiere eine allgemeine Funktion

```
def log_likelihood(model_vals, mu_vals, sigma_vals):
    ...
    return ...
```

die $\sum_{k=1}^{N} \log \mathcal{L}_k(x_k)$ zurückgibt. *Hinweis: zwei Zeilen, keine Schleife – **numpy** rechnet elementweise.* Teste mit

```
import numpy as np
mu_vals    = np.array([-10.0, -9.5, -9.0, -8.8, -8.5])
sigma_vals = np.array([  0.3,  0.4,  0.3,  0.5,  0.4])
print(log_likelihood(mu_vals, mu_vals, sigma_vals))       # Modell trifft
↪  genau
print(log_likelihood(mu_vals + 1.0, mu_vals, sigma_vals)) # Modell daneben
```

Erkläre, welcher Aufruf die größere log-Likelihood liefert und warum.

d) Welche Beziehung gilt zwischen dem ML-Schätzer (Maximierung der Likelihood) und der Methode der kleinsten Quadrate (Minimierung von χ^2)?

Hinweis: In Kapitel 17 werdet ihr genau diese Funktion auf 14 NANOGrav-Frequenzbins anwenden – dann heißen die Argumente `mu_log10` *und* `sigma_log10`*.*

Musterlösung 15.2

a) Logarithmieren von $\mathcal{L}_k(x_k)$: $\log(e^y) = y$ und $\log(1/\sqrt{y}) = -\frac{1}{2}\log y$:

$$\log \mathcal{L}_k(x_k) = -\frac{1}{2}\log(2\pi\sigma_k^2) - \frac{1}{2}\left(\frac{x_k - \mu_k}{\sigma_k}\right)^2. \tag{15.3}$$

Der zweite Term ist genau $-\frac{1}{2}$ mal der k-te χ^2-Summand; der erste ist eine Konstante (unabhängig vom Modell).
b) $\log \prod_k \mathcal{L}_k = \sum_k \log \mathcal{L}_k$. Eine Summe ist numerisch stabiler als ein Produkt aus N teils winzigen Zahlen (die leicht zu `0` underflowen).
c)

```
def log_likelihood(model_vals, mu_vals, sigma_vals):
    chi2 = ((model_vals - mu_vals) / sigma_vals)**2
    return np.sum(-0.5 * chi2 - 0.5 * np.log(2 * np.pi * sigma_vals**2))
```

Der erste Aufruf (`model_vals = mu_vals`) liefert den maximal möglichen Wert (alle χ^2-Terme verschwinden). Der zweite (`mu_vals + 1.0`) hat $\chi^2 > 0$, also kleinere log-Likelihood – ein Modell, das systematisch daneben liegt, ist weniger wahrscheinlich.
d) Maximierung von $\log \mathcal{L}$ ist äquivalent zur Minimierung von χ^2, da $\log \mathcal{L} = -\frac{1}{2}\chi^2 + \text{const}$. ML-Schätzer und Methode der kleinsten Quadrate sind für Gauß-Likelihood identisch.

15.3 Von der Likelihood zur Posterior: Satz von Bayes

Die Likelihood $\mathcal{L}(\text{model}(\theta))$ aus Aufgabe 15.2 sagt uns, wie wahrscheinlich die *Daten* sind, *gegeben* ein Parametersatz θ. Eigentlich interessiert uns aber die umgekehrte Frage: Wie wahrscheinlich ist ein Parametersatz θ, *gegeben* die Daten D, die wir beobachtet haben? Die Antwort liefert der *Satz von Bayes*:

$$\underbrace{p(\theta \mid D)}_{\text{Posterior}} \propto \underbrace{p(D \mid \theta)}_{\text{Likelihood}} \cdot \underbrace{p(\theta)}_{\text{Prior}}. \tag{15.4}$$

Der *Prior* $p(\theta)$ kodiert unser Vorwissen über θ, *bevor* wir die Daten betrachtet haben (z.B. „α kann nur positiv sein“ oder „$\beta/H \geq 1$“ aus Kapitel 18). Die *Posterior* $p(\theta \mid D)$ ist das, was wir eigentlich wissen wollen: die Wahrscheinlichkeitsverteilung der Parameter, *nachdem* wir die Daten berücksichtigt haben.

Aufgabe 15.3

Vor ein paar Tagen habt ihr Bayes' Theorem kennengelernt: für Parameter θ eines Modells und Daten D gilt

$$\underbrace{p(\theta \mid D)}_{\text{Posterior}} \propto \underbrace{p(D \mid \theta)}_{\text{Likelihood}} \cdot \underbrace{p(\theta)}_{\text{Prior}}.$$

a) Die Likelihood ist genau das, was du in Aufgabe 15.2 implementiert hast: $p(D \mid \theta) =$

$\mathcal{L}(\text{model}(\theta))$, wobei $\text{model}(\theta)$ die 14 Modellwerte für einen Parametersatz θ sind. Was beschreibt der Prior $p(\theta)$, und welche einfache Form könnte er haben, wenn wir z.B. nur wissen, dass ein Parameter $\log_{10} A$ (der Logarithmus einer Amplitude A) zwischen -13 und -4 liegen *kann*, aber sonst keine Vorinformation haben?

b) Logarithmiere Bayes' Theorem: $\log p(\theta \mid D) = \log p(D \mid \theta) + \log p(\theta) + \text{const}$. Wir nennen die (unnormierte) linke Seite ab jetzt `log_posterior(theta)`. Wie sollte `log_posterior` für ein θ aussehen, das *außerhalb* des erlaubten Bereichs (Prior = 0) liegt? *Hinweis: was ist* $\log 0$*, und wie geht* `numpy` *damit um (*`-np.inf`*)?*

Musterlösung 15.3

a) Der Prior $p(\theta)$ beschreibt unser Wissen (bzw. unsere Annahmen) über θ, *bevor* wir die konkreten Daten D gesehen haben, z.B. aus physikalischen Überlegungen (wie $\beta/H \geq 1$ in Kapitel 18) oder aus früheren Messungen. Wissen wir nur, dass $\log_{10} A \in [-13, -4]$ liegen kann, aber sonst nichts, ist die einfachste Wahl ein *Gleichverteilungs-Prior* (uniform prior):

$$p(\log_{10} A) = \begin{cases} \text{const.} & \text{falls } -13 \leq \log_{10} A \leq -4 \\ 0 & \text{sonst} \end{cases} . \tag{15.5}$$

b) Liegt θ außerhalb des erlaubten Bereichs, ist $p(\theta) = 0$, also $\log p(\theta) = \log 0 = -\infty$. Damit wird auch $\log p(\theta \mid D) = -\infty$, unabhängig vom Wert der Likelihood. In `numpy` entspricht das `-np.inf`. `log_posterior(theta)` sollte also für θ außerhalb des Prior-Bereichs sofort `-np.inf` zurückgeben, ohne überhaupt erst die (unter Umständen aufwendige) Likelihood auszuwerten.

15.4 Warum nicht einfach ein Gitter ausprobieren?

Ab jetzt schreiben wir `log_posterior` immer als Summe `log_prior + log_likelihood`, wobei `log_prior` entweder 0 (innerhalb des erlaubten Bereichs) oder `-np.inf` (außerhalb) ist.

Eine naive Idee: Werte `log_posterior` einfach für viele Werte von θ auf einem feinen Gitter aus und schaue, wo das Maximum liegt. Für *einen* Parameter $\theta \in [0, 1]$ mit 1000 Gitterpunkten ist das kein Problem. Aber: Unser Phasenübergangsmodell aus Kapitel 18 hat *drei* Parameter $(\alpha, \beta/H, T_*)$; mit 1000 Werten pro Parameter wären das $1000^3 = 10^9$ Kombinationen! Bei vier Parametern (Kapitel 17, verallgemeinertes Modell) wären es bereits 10^{12}. Dieses explosionsartige Wachstum der Rechenzeit mit der Anzahl der Parameter nennt man den *Fluch der Dimensionen* (*curse of dimensionality*).

Selbst wenn die Rechenzeit kein Problem wäre, würde uns ein Gitter nur den *einen besten* Punkt liefern. Wir wollen aber mehr wissen: Wie *breit* ist die Posterior um diesen besten Punkt herum – d.h. wie groß ist die Unsicherheit auf θ? Sind verschiedene Parameter *korreliert*? Eine Aussage wie „$\alpha = 0.3 \pm 0.05$“ braucht mehr als nur den besten Punkt.

Die Lösung: ein Algorithmus, der (i) auch in hohen Dimensionen funktioniert und (ii) nicht nur den besten Punkt liefert, sondern *viele* Punkte θ, und zwar so, dass Bereiche mit hoher Posterior-Wahrscheinlichkeit *häufiger* besucht werden als Bereiche mit niedriger Wahrscheinlichkeit. Eine solche Sammlung von Punkten heißt *Posterior-Sample*; aus ihr lassen sich Mittelwerte, Streuungen und Korrelationen direkt ablesen, indem man einfach die Häufigkeiten der Samples in den verschiedenen Bereichen des Parameterraums auszählt.

15.5 Ein Zufallsspaziergang durch den Parameterraum

Aufgabe 15.4

Die Idee: wir starten an einem Punkt θ_{curr} im Parameterraum und "laufen" Schritt für Schritt zu neuen Punkten, ähnlich wie eine Irrfahrt (*random walk*).

a) Ein neuer Vorschlag θ_{prop} soll "in der Nähe" von θ_{curr} liegen. Wie könntest du mit `rng.normal(0.0, step_sizes, size=ndim)` (einem `numpy`-Zufallszahlengenerator `rng`, der normalverteilte Zufallszahlen liefert) aus θ_{curr} einen solchen Vorschlag θ_{prop} konstruieren? Was bewirkt der Parameter `step_sizes` (ein Array mit einer Schrittweite pro Parameter)?

b) *Naive Regel:* "Gehe zu θ_{prop}, falls $\texttt{log\_posterior}(\theta_{\text{prop}}) > \texttt{log\_posterior}(\theta_{\text{curr}})$, sonst bleibe bei θ_{curr}." Was würde mit diesem "immer bergauf"-Spaziergang nach vielen Schritten passieren? Würdest du damit ein Posterior-*Sample* bekommen, oder eher einen einzelnen Punkt?

c) Damit der Spaziergang sich *proportional zur Posterior-Wahrscheinlichkeit* in der Umgebung aufhält (also auch mal in etwas schlechtere Bereiche geht, dort aber seltener als in gute Bereiche), brauchen wir eine Regel, die "schlechtere" Vorschläge nicht immer, aber *manchmal* akzeptiert. Vervollständige die folgende Idee:

> Berechne das Verhältnis $\alpha = \dfrac{p(\theta_{\text{prop}} \mid D)}{p(\theta_{\text{curr}} \mid D)}$. Falls $\alpha \geq 1$ (der Vorschlag ist mindestens so gut wie der aktuelle Punkt): .... Falls $\alpha < 1$ (der Vorschlag ist schlechter): akzeptiere ihn trotzdem, aber nur mit Wahrscheinlichkeit ....

Hinweis: Diese Regel heißt das Metropolis-Kriterium.

d) In Aufgabe 15.3 habt ihr gesehen, dass `log_posterior` statt $p(\theta \mid D)$ direkt $\log p(\theta \mid D)$ zurückgibt, u.a. weil $p(\theta \mid D)$ selbst oft astronomisch klein ist und `numpy` dann nur noch 0 ausgeben würde. Schreibe α aus c) so um, dass nur noch `log_posterior`-Werte vorkommen (also $\log \alpha = \ldots$, ohne dass du α selbst ausrechnen musst).

e) *Ein kleiner Numerik-Trick:* Mit `u = rng.random()` bekommt man eine Zufallszahl gleichverteilt in $[0, 1)$. Überlege dir, warum die Bedingung

$$\log(u) < \log \alpha$$

beide Fälle aus c) auf einmal korrekt abdeckt, also sowohl "$\alpha \geq 1$: immer akzeptieren" als auch "$\alpha < 1$: mit Wahrscheinlichkeit α akzeptieren". *Hinweis: für welche u ist $\log u$ positiv, negativ, oder 0? Und welche Werte kann $\log \alpha$ jeweils annehmen?*

Musterlösung 15.4

a) Mit $\theta_{\text{prop}} = \theta_{\text{curr}} + \texttt{rng.normal(0.0, step\_sizes, size=ndim)}$ erhält man einen Vorschlag, der um einen zufälligen, normalverteilten Versatz von θ_{curr} entfernt liegt. `step_sizes` legt die typische Schrittweite (Standardabweichung) für jeden Parameter einzeln fest: Große `step_sizes` bedeuten große, „mutige" Schritte (der Spaziergang erkundet den Parameterraum schnell, aber viele Vorschläge landen in schlechten Bereichen und werden abgelehnt), kleine `step_sizes` bedeuten kleine, „vorsichtige" Schritte (fast immer akzeptiert, aber der Spaziergang bewegt sich nur sehr langsam durch den Parameterraum).

b) Mit der „immer bergauf“-Regel würde der Spaziergang sich immer weiter zum (lokalen) Maximum von `log_posterior` bewegen und dort „festkleben“; sobald alle benachbarten Vorschläge schlechter sind als der aktuelle Punkt, bewegt er sich nicht mehr. Das liefert nur den *einen besten Punkt* (ein Optimierungsverfahren), aber kein Posterior-Sample im Sinne der vorherigen Diskussion.
c) Vervollständigt:

> Falls $\alpha \geq 1$: akzeptiere θ_{prop} *immer*. Falls $\alpha < 1$: akzeptiere θ_{prop} trotzdem, aber nur mit Wahrscheinlichkeit α (sonst bleibe bei θ_{curr}).

Das ist das *Metropolis-Kriterium*.
d) Mit Gleichung (15.4) und $p(\theta \mid D) = \exp(\log p(\theta \mid D))$ gilt

$$\alpha = \frac{p(\theta_{\text{prop}} \mid D)}{p(\theta_{\text{curr}} \mid D)} = \exp[\log p(\theta_{\text{prop}} \mid D) - \log p(\theta_{\text{curr}} \mid D)]\,, \tag{15.6}$$

also

$$\log \alpha = \texttt{log\_posterior}(\theta_{\text{prop}}) - \texttt{log\_posterior}(\theta_{\text{curr}})\,. \tag{15.7}$$

e) Für $u \in [0, 1)$ ist $\log u \leq 0$, wobei $\log u \to -\infty$ für $u \to 0$ und $\log u \to 0$ für $u \to 1$. Ist $\log \alpha \geq 0$ (also $\alpha \geq 1$), ist die Bedingung $\log u < \log \alpha$ immer erfüllt (da $\log u < 0 \leq \log \alpha$, außer im Grenzfall $u \to 1, \log \alpha = 0$); das deckt den Fall „immer akzeptieren“ ab. Ist $\log \alpha < 0$ (also $0 < \alpha < 1$), ist $\log u < \log \alpha$ genau dann erfüllt, wenn $u < \alpha$; da u gleichverteilt in $[0, 1)$ ist, passiert das mit Wahrscheinlichkeit α. Genau das ist das Metropolis-Kriterium aus c).

15.6 Warum funktioniert das Metropolis-Kriterium? (*Detailed Balance*)

Aufgabe 15.5

Der Zufallsspaziergang aus Aufgabe 15.4 definiert eine sogenannte *Markov-Kette*: ausgehend von θ_{curr} wird mit einer bestimmten Wahrscheinlichkeit $T(\theta_{\text{curr}} \to \theta')$ der nächste Punkt θ' erreicht, egal wie die Kette vorher dort hingekommen ist. Wir wollen verstehen, warum diese Kette nach vielen Schritten Punkte θ mit einer Häufigkeit besucht, die proportional zu $p(\theta \mid D)$ ist.

a) $T(\theta \to \theta')$ setzt sich aus zwei Schritten zusammen: zuerst wird θ' *vorgeschlagen* (mit einer Wahrscheinlichkeitsdichte $q(\theta \to \theta')$, vgl. Aufgabe 15.4a) der Gauß-Schritt), dann wird der Vorschlag mit der Metropolis-Wahrscheinlichkeit

$$a(\theta \to \theta') = \min\left(1, \frac{p(\theta' \mid D)}{p(\theta \mid D)}\right)$$

akzeptiert, also $T(\theta \to \theta') = q(\theta \to \theta') \cdot a(\theta \to \theta')$. Begründe: warum gilt für den Gauß-Schritt aus Aufgabe 15.4a) $q(\theta \to \theta') = q(\theta' \to \theta)$ (man sagt, der Vorschlag ist *symmetrisch*)?

b) Eine Verteilung $\pi(\theta)$ heißt *stationär* für die Markov-Kette, wenn ein Schritt eine Stichprobe aus π wieder in eine Stichprobe aus π überführt:

$$\sum_{\theta} \pi(\theta)\, T(\theta \to \theta') = \pi(\theta') \qquad \text{für alle } \theta'.$$

Wenn $p(\theta \mid D)$ stationär ist, bleibt die Kette also, einmal bei $p(\theta \mid D)$ "angekommen" – für immer so verteilt; genau das brauchen wir, damit unsere Kette ein Posterior-Sample liefert.

c) *Detailed balance* ist eine (stärkere, aber einfacher zu prüfende) hinreichende Bedingung für Stationarität: für *je zwei* Punkte θ, θ' soll gelten

$$\pi(\theta)\, T(\theta \to \theta') = \pi(\theta')\, T(\theta' \to \theta).$$

Zeige, dass aus dieser Bedingung b) folgt: summiere beide Seiten über θ und benutze, dass $\sum_\theta T(\theta' \to \theta) = 1$ (die Kette geht von θ' ja *irgendwohin*).

d) Jetzt die Hauptaufgabe: zeige, dass $\pi(\theta) = p(\theta \mid D)$ mit $T = q \cdot a$ und dem Metropolis-a aus a) *detailed balance* erfüllt. Unterscheide die zwei Fälle $p(\theta' \mid D) \geq p(\theta \mid D)$ und $p(\theta' \mid D) < p(\theta \mid D)$ und zeige in beiden Fällen

$$p(\theta \mid D)\, q(\theta \to \theta')\, a(\theta \to \theta') \;=\; p(\theta' \mid D)\, q(\theta' \to \theta)\, a(\theta' \to \theta).$$

Hinweis: Im Fall $p(\theta' \mid D) \geq p(\theta \mid D)$ ist $a(\theta \to \theta') = 1$ und $a(\theta' \to \theta) = p(\theta \mid D)/p(\theta' \mid D)$. Setze beides ein und benutze a).

Musterlösung 15.5

a) Der Gauß-Schritt aus Aufgabe 15.4a) schlägt $\theta' = \theta + \xi$ vor, wobei ξ aus einer Normalverteilung mit Mittelwert 0 gezogen wird. Die Wahrscheinlichkeitsdichte für diesen Schritt hängt also nur von der *Differenz* $\theta' - \theta$ ab, $q(\theta \to \theta') = g(\theta' - \theta)$ für eine symmetrische Gauß-Dichte g (d.h. $g(\xi) = g(-\xi)$, da die Normalverteilung um 0 symmetrisch ist). Damit folgt

$$q(\theta \to \theta') = g(\theta' - \theta) = g(-(\theta - \theta')) = g(\theta - \theta') = q(\theta' \to \theta)\,, \tag{15.8}$$

also ist der Vorschlag tatsächlich symmetrisch.
b) (Reine Definition, keine Rechnung nötig; siehe Aufgabenstellung.)
c) Wir summieren beide Seiten von $\pi(\theta)T(\theta \to \theta') = \pi(\theta')T(\theta' \to \theta)$ über θ:

$$\sum_\theta \pi(\theta)T(\theta \to \theta') = \sum_\theta \pi(\theta')T(\theta' \to \theta) = \pi(\theta') \sum_\theta T(\theta' \to \theta) = \pi(\theta') \cdot 1 = \pi(\theta')\,, \tag{15.9}$$

wobei wir im zweiten Schritt $\pi(\theta')$ aus der Summe über θ herausgezogen haben (es hängt nicht von θ ab) und im dritten Schritt $\sum_\theta T(\theta' \to \theta) = 1$ benutzt haben. Die linke Seite ist aber genau die linke Seite der Stationaritätsbedingung aus b); also folgt b) aus detailed balance.
d) Fall 1: $p(\theta' \mid D) \geq p(\theta \mid D)$**.** Dann ist $\frac{p(\theta'|D)}{p(\theta|D)} \geq 1$, also $a(\theta \to \theta') = \min(1, \dots) = 1$ und $a(\theta' \to \theta) = \min\left(1, \frac{p(\theta|D)}{p(\theta'|D)}\right) = \frac{p(\theta|D)}{p(\theta'|D)}$ (da $\frac{p(\theta|D)}{p(\theta'|D)} \leq 1$). Damit:

$$\text{LHS} = p(\theta \mid D)\, q(\theta \to \theta') \cdot 1 = p(\theta \mid D)\, q(\theta \to \theta')\,, \tag{15.10}$$

$$\text{RHS} = p(\theta' \mid D)\, q(\theta' \to \theta) \cdot \frac{p(\theta \mid D)}{p(\theta' \mid D)} = p(\theta \mid D)\, q(\theta' \to \theta)\,. \tag{15.11}$$

Mit $q(\theta \to \theta') = q(\theta' \to \theta)$ aus a) ist LHS = RHS.

Fall 2: $p(\theta' \mid D) < p(\theta \mid D)$**.** Dann tauschen die Rollen: $a(\theta' \to \theta) = 1$ und $a(\theta \to \theta') = \frac{p(\theta'|D)}{p(\theta|D)}$. Damit:

$$\text{LHS} = p(\theta \mid D)\, q(\theta \to \theta') \cdot \frac{p(\theta' \mid D)}{p(\theta \mid D)} = p(\theta' \mid D)\, q(\theta \to \theta')\,, \tag{15.12}$$

$$\text{RHS} = p(\theta' \mid D)\, q(\theta' \to \theta) \cdot 1 = p(\theta' \mid D)\, q(\theta' \to \theta)\,. \tag{15.13}$$

Wieder mit $q(\theta \to \theta') = q(\theta' \to \theta)$ aus a) folgt LHS = RHS.
In beiden Fällen gilt also detailed balance für $\pi(\theta) = p(\theta \mid D)$. Zusammen mit c) und b) folgt: $p(\theta \mid D)$ ist eine stationäre Verteilung der Markov-Kette mit dem Metropolis-Kriterium – nach einer „Einschwingphase“ (*burn-in*) besucht die Kette also Punkte θ mit relativen Häufigkeiten proportional zu $p(\theta \mid D)$. Genau das macht den Metropolis-Algorithmus zu einem korrekten Verfahren, um aus der Posterior zu samplen!

15.7 Der Metropolis-Hastings-Sampler

Aufgabe 15.6

Jetzt habt ihr alle Bausteine selbst hergeleitet: Zeit, sie zu einem Algorithmus zusammenzusetzen! Lege eine Datei `mh_sampler.py` an.

a) Schreibe eine Funktion

```
def metropolis_hastings(log_post, x0, step_sizes, n_steps, seed=0):
    rng = np.random.default_rng(seed)

    ndim = len(x0)
    x_curr = np.array(x0, dtype=float)
    logp_curr = log_post(x_curr)

    chain = np.zeros((n_steps, ndim))
    log_post_chain = np.zeros(n_steps)
    n_accept = 0

    for i in range(n_steps):
        # 1) Vorschlag: ein zufaelliger kleiner Schritt (Aufgabe 6a)
        x_prop = ...

        # 2) Posterior am Vorschlag auswerten
        logp_prop = ...

        # 3) Metropolis-Kriterium in log-Form (Aufgabe 6d)
        log_alpha = ...

        # 4) Annehmen oder verwerfen (Aufgabe 6e)
        #    Achtung: log_post kann -np.inf sein (ausserhalb des Priors) --
        #    das darf niemals akzeptiert werden!
        if ...:
            x_curr, logp_curr = x_prop, logp_prop
            n_accept += 1

        chain[i] = x_curr
        log_post_chain[i] = logp_curr
```

```
    return chain, log_post_chain, n_accept / n_steps
```

und ergänze die vier Lücken mit dem, was du in Aufgabe 15.4 hergeleitet hast. *Hinweis zu Lücke 4:* `np.isfinite(logp_prop)` *ist* `False`*, wenn* `logp_prop` *gleich* `-np.inf` *ist.*

b) **Teste deinen Sampler an einem Beispiel, dessen Lösung du schon kennst!** Bevor wir den Sampler auf unsere (unbekannte) PTA-Likelihood anwenden, probieren wir ihn an einer 2D-Gauß-Verteilung mit *vorgegebenem* Mittelwert μ und *vorgegebener* Kovarianz Σ, um zu prüfen, ob unser Sampler überhaupt funktioniert:

```
mu = np.array([1.0, -0.5])
Sigma = np.array([[1.0, 0.6],
                  [0.6, 0.5]])
Sigma_inv = np.linalg.inv(Sigma)

def log_post_gauss(theta):
    d = theta - mu
    return -0.5 * d @ Sigma_inv @ d
```

Rufe deinen Sampler auf mit

```
chain, logp, acc = metropolis_hastings(
    log_post_gauss, x0=[0.0, 0.0], step_sizes=[0.7, 0.7],
    n_steps=20000, seed=1)
```

Verwerfe die ersten 2000 Schritte als "*burn-in*" (Einschwingphase, in der der Spaziergang noch vom Startpunkt zur eigentlichen Verteilung findet): `samples = chain[2000:]`.

c) Vergleiche `samples.mean(axis=0)` mit `mu` und `np.cov(samples.T)` mit `Sigma`. Passt es? Gib außerdem die `acc` (Akzeptanzrate) aus; sie sollte zwischen 0.2 und 0.5 liegen.

d) Plotte `chain[:,0]` gegen den Schritt-Index (sog. *trace plot*) sowie `samples[:,0]` gegen `samples[:,1]` als Streudiagramm.

e) *Experimentiere:* Was passiert mit der Akzeptanzrate `acc`, wenn du `step_sizes` stark vergrößerst (z.B. `[5.0, 5.0]`) oder stark verkleinerst (z.B. `[0.01, 0.01]`)? Wie sieht der trace plot in beiden Fällen aus? Was bedeutet das jeweils für die Erkundung des Parameterraums?

Musterlösung 15.6

a) Mit den Ergebnissen aus Aufgabe 15.4:

```
def metropolis_hastings(log_post, x0, step_sizes, n_steps, seed=0):
    rng = np.random.default_rng(seed)

    ndim = len(x0)
    x_curr = np.array(x0, dtype=float)
    logp_curr = log_post(x_curr)

    chain = np.zeros((n_steps, ndim))
    log_post_chain = np.zeros(n_steps)
    n_accept = 0
```

```
    for i in range(n_steps):
        # 1) Vorschlag: ein zufaelliger kleiner Schritt (Aufgabe 6a)
        x_prop = x_curr + rng.normal(0.0, step_sizes, size=ndim)

        # 2) Posterior am Vorschlag auswerten
        logp_prop = log_post(x_prop)

        # 3) Metropolis-Kriterium in log-Form (Aufgabe 6d)
        log_alpha = logp_prop - logp_curr

        # 4) Annehmen oder verwerfen (Aufgabe 6e)
        if np.isfinite(logp_prop) and np.log(rng.random()) < log_alpha:
            x_curr, logp_curr = x_prop, logp_prop
            n_accept += 1

        chain[i] = x_curr
        log_post_chain[i] = logp_curr

    return chain, log_post_chain, n_accept / n_steps
```

b)/c) Nach dem Aufruf mit den angegebenen Einstellungen und `samples = chain[2000:]` ergibt sich `samples.mean(axis=0)`$\approx (1.0, -0.5)$, in guter Übereinstimmung mit `mu` – und `np.cov(samples.T)`$\approx \Sigma$. Die Akzeptanzrate liegt bei diesen `step_sizes` bei acc ≈ 0.3–0.4, also im gewünschten Bereich von 0.2–0.5. Der Sampler funktioniert also (Abbildung 15.1): Er rekonstruiert eine Verteilung, deren Parameter wir vorher selbst festgelegt haben, korrekt aus den Samples allein.

d) Der trace plot `chain[:,0]` gegen den Schritt-Index zeigt zunächst (in den ersten $\sim$ wenigen hundert Schritten) eine kurze „Einschwingphase“, in der die Kette vom Startpunkt $(0, 0)$ zum Bereich um $\mu = (1, -0.5)$ wandert; das ist der *burn-in*. Danach oszilliert `chain[:,0]` stationär um $\mu_1 = 1.0$. Das Streudiagramm `samples[:,0]` gegen `samples[:,1]` zeigt eine elliptische „Punktwolke“; die Form und Ausrichtung der Ellipse entspricht genau der Kovarianzmatrix Σ (positive Korrelation, da $\Sigma_{12} = 0.6 > 0$).

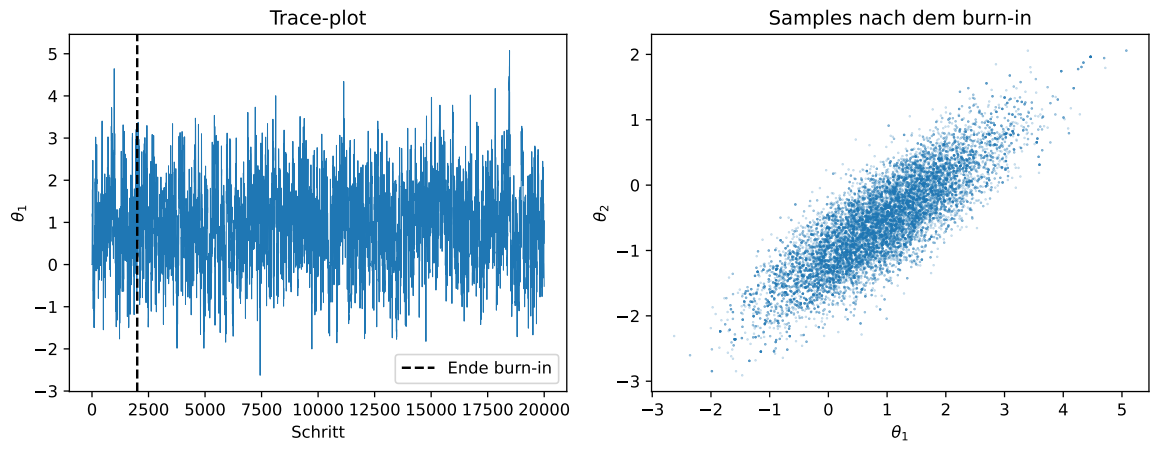


Abbildung 15.1: Trace plot von `chain[:,0]` (links, mit sichtbarem burn-in am Anfang) und Streudiagramm der `samples` nach dem burn-in (rechts) für den 2D-Gauß-Test. Die elliptische Punktwolke entspricht der vorgegebenen Kovarianzmatrix Σ.

e) Für sehr große `step_sizes` (z.B. $[5, 5]$) landen die meisten Vorschläge weit außerhalb der Gauß-Verteilung, wo log post sehr klein ist; die Akzeptanzrate sinkt stark (oft $\ll 0.1$). Der

trace plot zeigt lange „flache“ Abschnitte, in denen die Kette stecken bleibt, unterbrochen von wenigen großen Sprüngen. Für sehr kleine `step_sizes` (z.B. $[0.01, 0.01]$) wird fast jeder Vorschlag akzeptiert (Akzeptanzrate $\to 1$), aber die Kette bewegt sich nur in winzigen Schritten – der trace plot sieht aus wie eine sehr „glatte“, langsam mäandernde Kurve. In beiden Extremfällen braucht die Kette *viel länger*, um den Parameterraum gut zu erkunden, als bei einer „vernünftigen“ Akzeptanzrate von 0.2–0.5.

16 Schuhpendel: Bayessche Inferenz mit eigenem MCMC

Szenario: Ein Unwetter hat alle Zollstöcke, Lineale und Maßbänder zerstört. Du hast nur dein Handy (mit einer g-Force-messenden Sensor-App), etwas Schnur und einen Schuh – und musst trotzdem eine Länge bestimmen. Du baust dir ein Pendel, verpackst dein Handy sicher im Schuh und lässt es während des Schwingens seine eigene Beschleunigung aufzeichnen. Ziel: mithilfe des Metropolis-Hastings-Samplers aus Kapitel 15 gleichzeitig die Pendellänge L, die Erdbeschleunigung g und die Dämpfung deines Pendels aus den Messdaten zu bestimmen. *(Das vollständige Übungsblatt mit allen Zwischenschritten – inklusive der Aufnahme deiner eigenen Daten und der Auflösung eines kleinen Rätsels um eine verdoppelte Frequenz – findest du separat als* `uebungsblatt_inselpendel.tex`*.)*

Die Messdaten stehen in einer CSV-Datei mit den Spalten Zeit t (in Sekunden) und Beschleunigung a_x, a_y, a_z (in Einheiten von g), wie sie z. B. von den Apps *phyphox* oder *Physics Toolbox* exportiert wird.

16.1 Modell und Physik

Aufgabe 16.1

Lade die Daten aus deiner CSV-Datei ein und plotte alle drei Beschleunigungskomponenten gegen die Zeit.

- Identifiziere die Achse, die eine saubere, regelmäßige Schwingung zeigt, sowie ein Zeitfenster, das nur das freie Schwingen (ohne Anfassen des Handys) enthält.
- Zentriere die Daten in diesem Fenster (ziehe den Mittelwert ab).

Musterlösung 16.1

```
import numpy as np
import matplotlib.pyplot as plt

t, ax, ay, az, atot = np.loadtxt('meine_messung.csv', delimiter=',',
                                 skiprows=1, unpack=True)
fig, axes = plt.subplots(3, 1, figsize=(7, 5.5), sharex=True)
axes[0].plot(t, ax); axes[0].set_ylabel('$a_x/g$')
axes[1].plot(t, ay); axes[1].set_ylabel('$a_y/g$')
axes[2].plot(t, az); axes[2].set_ylabel('$a_z/g$')
axes[2].set_xlabel('$t$ [s]')
plt.tight_layout(); plt.show()
```

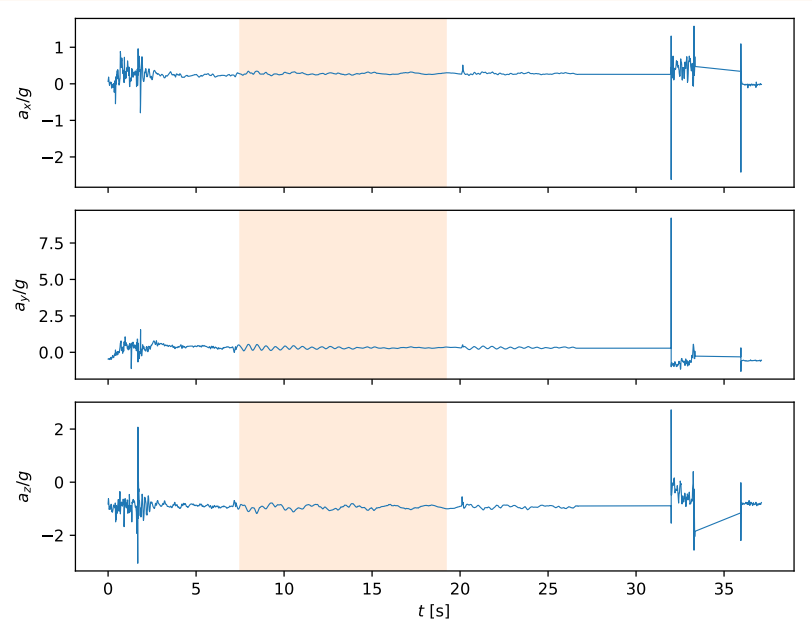


Abbildung 16.1: Rohe Beschleunigungsdaten aller drei Achsen. Die z-Achse (entlang der Schnur ausgerichtet) zeigt die sauberste, langsam abklingende Schwingung.

```
start_time, end_time = 7.5, 19.2
s = np.searchsorted(t, start_time)
e = np.searchsorted(t, end_time)

t_win = t[s:e] - start_time
z_win = az[s:e] - np.mean(az[s:e])
```

Ein Pendel der Länge L schwingt in der Kleinwinkelnäherung ($\sin\varphi \approx \varphi$) wie ein harmonischer Oszillator. Die Bewegungsgleichung lautet:

$$\varphi''(t) + \frac{g}{L}\,\varphi(t) = 0\,. \tag{16.1}$$

Gestartet mit Auslenkung φ_0 und Ruhe ($\varphi'(0) = 0$) ergibt sich die Lösung

$$\varphi(t) = \varphi_0\,\cos(\omega t), \qquad \omega = \sqrt{\frac{g}{L}}\,. \tag{16.2}$$

Das Handy misst allerdings nicht direkt den Winkel φ, sondern die Beschleunigung entlang seiner eigenen (z-)Achse – und die hängt über die Projektion der Schwerkraft bzw. die Zentripetalbeschleunigung von $\varphi(t)$ *in zweiter Ordnung* ab (z. B. über $\cos\varphi \approx 1 - \varphi^2/2$). Mit der Doppelwinkelformel $\cos^2(\omega t) = \tfrac{1}{2}(1 + \cos(2\omega t))$ folgt daraus, dass das gemessene Signal mit der *doppelten* Kreisfrequenz 2ω oszilliert, nicht mit ω selbst (die vollständige Herleitung dieses Effekts über eine Projektionsskizze findet sich im Übungsblatt). Zusätzlich klingt die Amplitude eines realen Pendels durch Luftwiderstand mit der Zeit ab, ungefähr wie $\mathrm{e}^{-\gamma t}$. Das vollständige Modell für die gemessene Beschleunigung lautet damit

$$a(t) = A\,\mathrm{e}^{-\gamma(t-t_0)}\,\cos\!\big(2\omega(t-t_0) + \psi_0\big)\,, \qquad \omega = \sqrt{\frac{g}{L}}\,. \tag{16.3}$$

Die Modellparameter, die wir aus den Daten bestimmen wollen, sind also die Amplitude A, die Phase ψ_0, die Pendellänge L, die Erdbeschleunigung g und die Dämpfung γ.

Aufgabe 16.2

- Implementiere `acc_z_model(t, theta)` mit `theta = (A, psi0, L, g, gamma)` gemäß Gleichung (16.3).
- Plotte Modell und Daten im selben Graph (mit einem Testpunkt, z. B. $L = 4.9\,\mathrm{m}$, $g = 9.81\,\mathrm{m/s^2}$, $\gamma = 0.05\,\mathrm{s^{-1}}$, $A = 0.16$, $\psi_0 = 1.3$), um zu prüfen, ob die Kurve qualitativ passt.

Musterlösung 16.2

```
def acc_z_model(t, theta):
    A, psi0, L, g, gamma = theta
    omega = np.sqrt(g / L)
    return A * np.exp(-gamma * t) * np.cos(2 * omega * t + psi0)

# Kurzcheck mit Testwerten
test = (0.16, 1.3, 4.9, 9.81, 0.05)
plt.plot(t_win, z_win, '.', ms=3, label='Daten')
plt.plot(t_win, acc_z_model(t_win, test), label='Modell (Test)')
plt.xlabel('$t-t_0$ [s]'); plt.ylabel(r'$a_z/g$ (zentriert)'); plt.legend();
↪  plt.show()
```

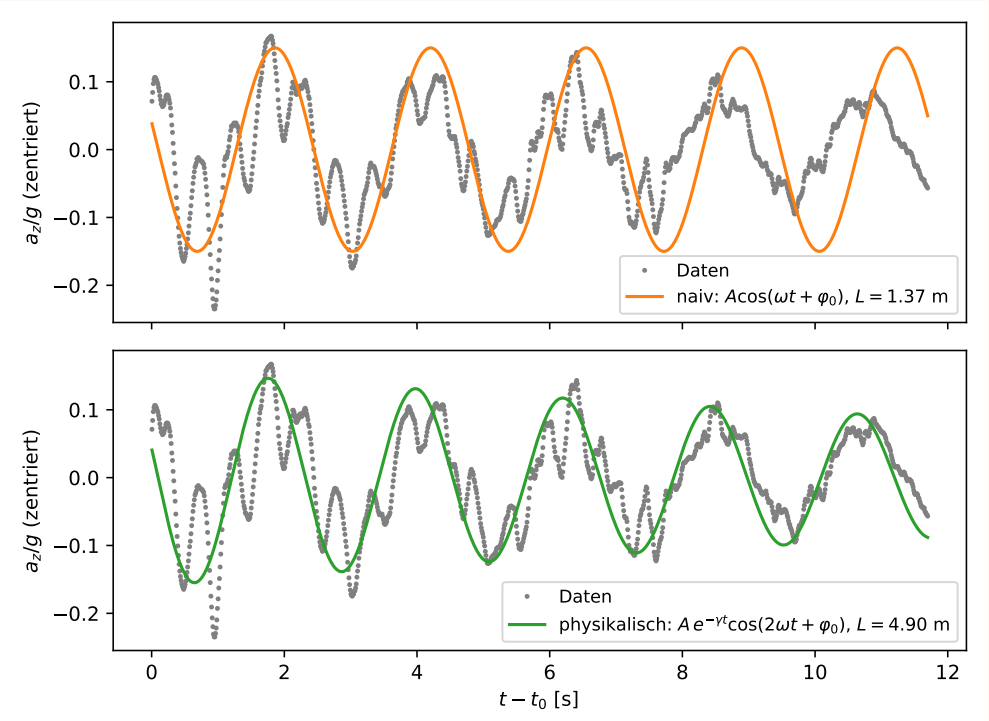


Abbildung 16.2: Oben: ein naives, ungedämpftes Ein-Frequenz-Modell trifft nur die Periode, nicht die abklingende Hüllkurve. Unten: das vollständige Modell aus Gleichung (16.3) mit $L = 4.9\,\mathrm{m}$, $g = 9.81\,\mathrm{m/s^2}$, $\gamma = 0.05\,\mathrm{s^{-1}}$ trifft die Daten über das ganze Fenster hinweg deutlich besser. Stimmt die Periode und die Abklingrate qualitativ, muss der MCMC-Sampler nur noch die genauen Werte finden.

16.2 Likelihood, Prior und Posterior

Aufgabe 16.3

- Die Messung hat eine geschätzte Genauigkeit von $\sigma \approx 0.1$ (in Einheiten von g). Implementiere eine Gauß-Likelihood für die Daten:

$$\log p(D \mid \theta) = -\frac{1}{2\sigma^2} \sum_i \big[a_{z,i} - a(t_i;\theta)\big]^2 + \text{const.} \tag{16.4}$$

- Implementiere die Priors: $A > 0$, $0 \le \psi_0 \le 2\pi$, $L > 0$, $\gamma \ge 0$ (jeweils flach), sowie einen informativen Gauß-Prior auf g, da wir aus anderen Experimenten schon relativ genau wissen, was g sein sollte:

$$g \sim \mathcal{N}(9.81\,\mathrm{m/s^2},\, (0.05\,\mathrm{m/s^2})^2)\,. \tag{16.5}$$

Musterlösung 16.3

```
sigma_meas = 0.1   # Messgenauigkeit in Einheiten von g

def log_likelihood(theta):
    model = acc_z_model(t_win, theta)
    return -0.5 / sigma_meas**2 * np.sum((z_win - model)**2)

def log_prior(theta):
    A, psi0, L, g, gamma = theta
    if not (0 < A < 10):
        return -np.inf
    if not (0 <= psi0 <= 2 * np.pi):
        return -np.inf
    if not (0 < L < 10):
        return -np.inf
    if not (0 <= gamma < 2):
        return -np.inf
    return -0.5 * ((g - 9.81) / 0.05)**2
```

16.3 Metropolis-Hastings-Inferenz

Aufgabe 16.4

Verwende den Metropolis-Hastings-Sampler aus Aufgabe 15.6 (Datei `mh_sampler.py`) um aus der Posterior $p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta)$ mit $\theta = (A, \psi_0, L, g, \gamma)$ zu sampeln.

a) Definiere `log_posterior(theta)`.

b) Rufe den Sampler auf (Startpunkt nahe deinen Testwerten aus Aufgabe 16.2; $n \gtrsim 10^5$ Schritte; verwerfe die ersten 10 % als Burn-in). Stelle sicher, dass die Akzeptanzrate zwischen 0.2 und 0.5 liegt.

c) Erzeuge einen Trace-Plot für alle fünf Parameter und prüfe, ob die Kette konvergiert ist.

d) Erstelle einen Corner-Plot. Welche Werte für L und g bevorzugen die Daten? Stimmt g mit dem Erwartungswert $9.81\,\mathrm{m/s^2}$ überein?

Musterlösung 16.4

```python
from mh_sampler import metropolis_hastings

def log_posterior(theta):
    lp = log_prior(theta)
    if not np.isfinite(lp):
        return -np.inf
    return lp + log_likelihood(theta)

x0 = [0.15, 1.3, 1.4, 9.81, 0.05]
step_sizes = [0.02, 0.02, 0.05, 0.02, 0.01]

chain, logp, acc = metropolis_hastings(
    log_posterior, x0, step_sizes, n_steps=200000, seed=0)
samples = chain[20000:]   # 10% burn-in
print(f'Akzeptanzrate: {acc:.2f}')
```

Die Abbildungen 16.3–16.4 zeigen die Ergebnisse der Musterlösung.

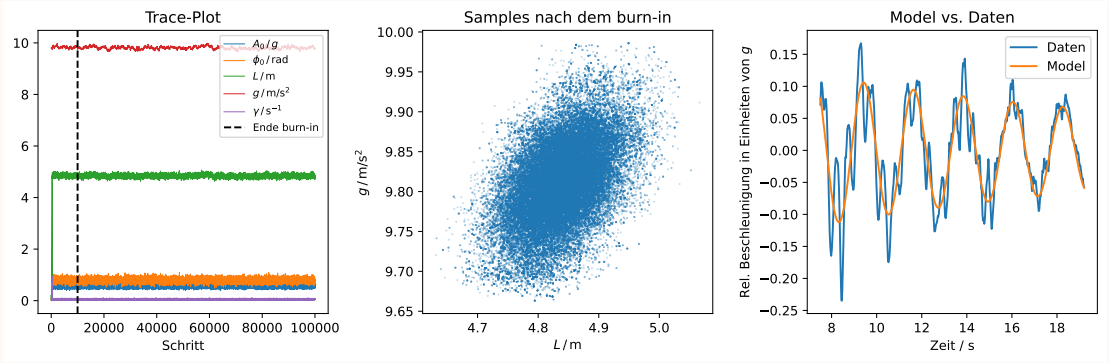


Abbildung 16.3: Trace-Plot (links), Samples im (L, g)-Unterraum nach dem Burn-in (Mitte) und Modell vs. Daten für den MAP-Fit (rechts).

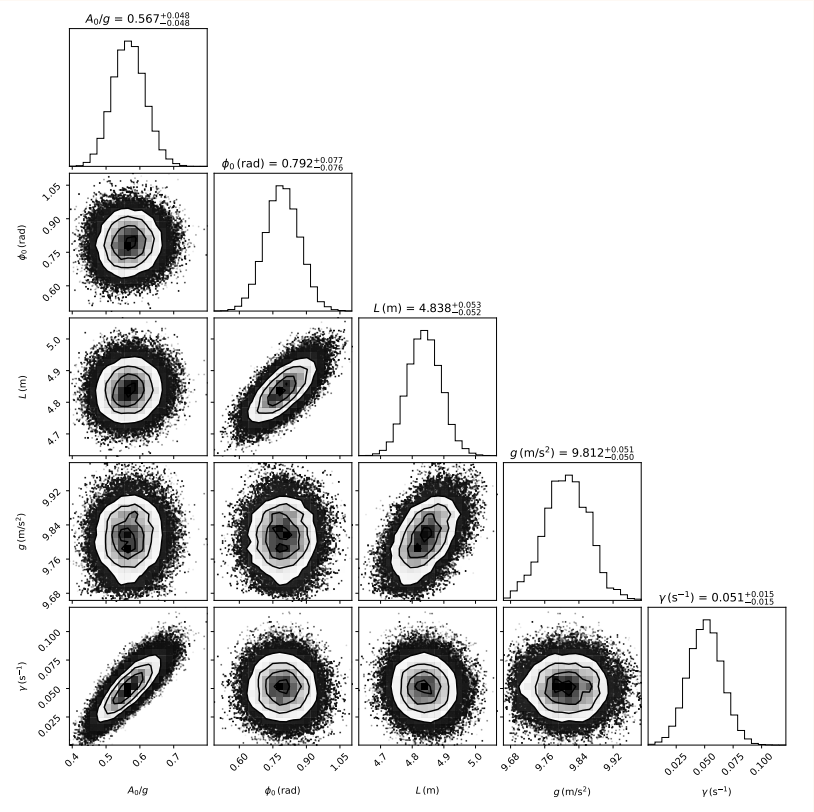


Abbildung 16.4: Corner-Plot mit marginalisierten Posteriors (Diagonale) und 2D-Konfidenzregionen für alle fünf Parameter.

Die Referenzmessung liefert $L = 4.84 \pm 0.05\,\mathrm{m}$, $g = 9.81 \pm 0.05\,\mathrm{m/s^2}$ und $\gamma = 0.051 \pm 0.015\,\mathrm{s^{-1}}$ – g stimmt gut mit dem Erwartungswert $9.81\,\mathrm{m/s^2}$ überein, und L ist durch die Daten (nicht durch den Prior) bestimmt.

Aufgabe 16.5

[Für die Schnellen: Herleitung der Bewegungsgleichung] Der Drehmomentansatz für ein Pendel der Länge L mit Masse m liefert

$$\varphi''(t) = -\frac{g}{L}\sin\varphi(t)\,.$$

a) Wende die Kleinwinkelnäherung $\sin\varphi \approx \varphi$ an und zeige, dass daraus $\varphi''(t) + (g/L)\,\varphi(t) = 0$ folgt.

b) Prüfe durch Einsetzen, dass $\varphi(t) = \varphi_0\cos(\omega t)$ mit $\omega = \sqrt{g/L}$ eine Lösung ist, wenn $\varphi'(0) = 0$ gilt.

Mit dem Schuhpendel haben wir den Metropolis-Hastings-Algorithmus an einem vollständig kontrollierbaren Beispiel geübt: Modell bekannt, Antwort bekannt, Ergebnis überprüfbar. Ab jetzt wenden wir dieselbe Maschinerie auf echte Physik an — zunächst auf Phasenübergänge im frühen Universum und dann auf Pulsar-Timing-Arrays.

17 Pulsar-Timing-Arrays

Gravitationswellen, welche von Paaren von supermassiven Schwarzen Löchern oder von Phasenübergängen bei MeV-Temperaturen im primordialen Plasma erzeugt werden, haben extrem lange Wellenlängen von der Größenordnung von Lichtjahren, Schwingungsfrequenzen auf der nHz-Skala und Periodendauern von Jahren bis Jahrzehnten. Um Gravitationswellen von solchen *astronomischen* Ausmaßen zu detektieren braucht es auch Messgeräte entsprechender Größe. In diesem Kapitel wollen wir uns mit Pulsar-Timing-Arrays beschäftigen, welche genau diese Aufgabe erfüllen. Dazu nutzt man die gesamte *Galaxie* als Detektor, und misst die kleinen Längenänderungen, die durch die Gravitationswellen verursacht werden, mithilfe von *Pulsaren*.

17.1 Pulsare als kosmische Uhren

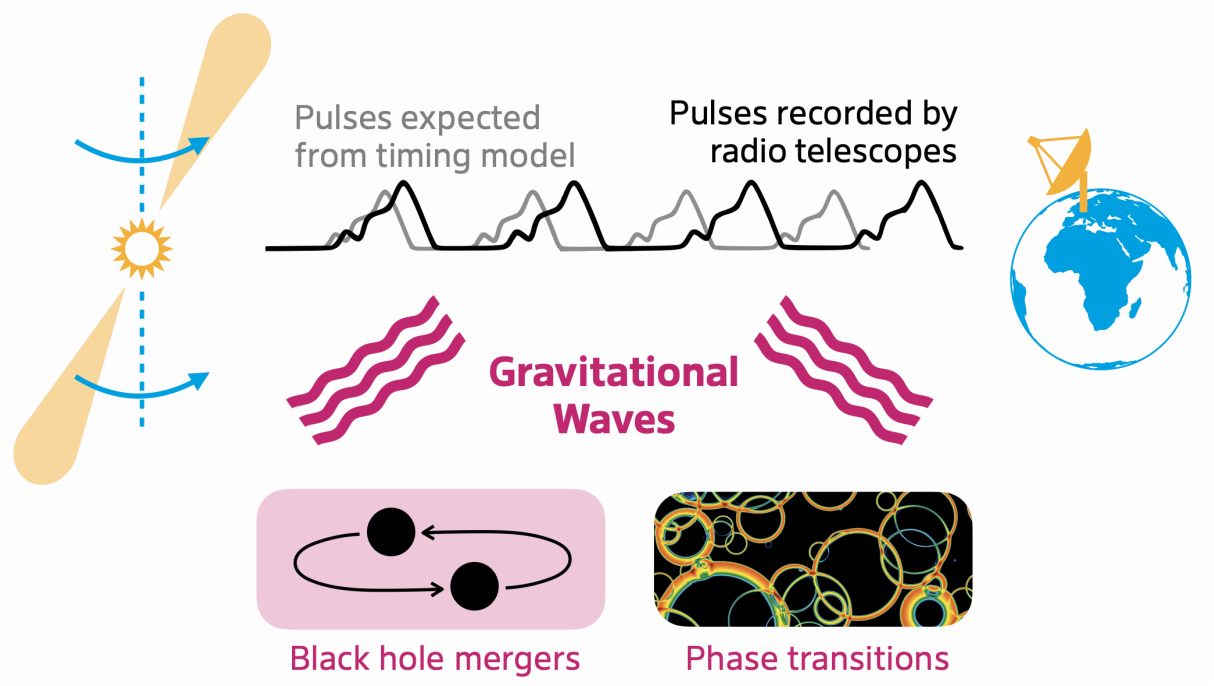


Abbildung 17.1: Schaubild aus einem Forschungshighlight der Universität Hamburg (Oktober 2024): „Der kosmische Brummen – woher kommt das Hintergrundrauschen aus Gravitationswellen?“. Das Signal, das NANOGrav und andere PTA-Kollaborationen messen, könnte von supermassiven Schwarzen-Loch-Binärsystemen oder von einem kosmologischen Phasenübergang im frühen Universum stammen. In Referenz [10] wurde untersucht, ob ein *Phasenübergang in einem dunklen Sektor* bei MeV-Temperaturen das Signal erklären kann und welche kosmologischen Nebenbedingungen dabei eingehalten werden müssen. Abbildung: C. Tasillo [10].

Ein Pulsar ist ein extrem schnell rotierender Neutronenstern, der wie ein kosmischer Leuchtturm einen Strahl elektromagnetischer Strahlung aussendet. Zeigt dieser Strahl bei jeder Rotation zur Erde, registrieren wir ein regelmäßiges „Ticken“, bei manchen sogenannten *Millisekundenpulsaren* mit einer Regelmäßigkeit, die mit Atomuhren konkurriert. Diese Pulsare sind also extrem präzise „kosmische Uhren“.

Läuft nun eine Gravitationswelle zwischen Erde und Pulsar durch die Raumzeit, verändert sie die effektive Distanz zwischen Erde und Pulsar geringfügig. Dadurch kommt das Signal des Pulsars minimal früher oder später an als erwartet. Diese kleinen Abweichungen zwischen beobachteter und erwarteter Ankunftszeit nennt man *Timing-Residuen*.

17.2 Warum ein *Array*?

Das Signal eines Gravitationswellenhintergrunds in den Timing-Residuen eines einzelnen Pulsars ist winzig. Folglich ist es sehr schwierig, es von anderen Störeffekten (intrinsisches „Rauschen“ des Pulsars, Messfehler, Störungen durch das interstellare Medium, ...) zu unterscheiden. Der Trick: Ein Gravitationswellenhintergrund beeinflusst *alle* vermessenen Pulsare am Himmel *gleichzeitig* und mit einer ganz bestimmten, vorhersagbaren (sog. *Hellings-Downs-*) *Korrelation* zwischen je zwei Pulsaren, die nur vom Winkelabstand dieser beiden Pulsare am Himmel abhängt.

Indem man die Timing-Residuen von vielen ($\mathcal{O}(10\text{–}100)$) über die ganze Galaxie verteilten Millisekundenpulsaren gemeinsam analysiert, ein sog. *Pulsar-Timing-Array* (PTA), kann man nach genau diesem charakteristischen, korrelierten Muster suchen, das von zufälligem Rauschen klar unterscheidbar ist. Effektiv wird die Milchstraße selbst zu einem riesigen Gravitationswellendetektor, dessen „Armlängen“ die Distanzen zwischen Erde und den einzelnen Pulsaren sind (typischerweise $\mathcal{O}(10^3)$ Lichtjahre), passend zur enormen Wellenlänge der gesuchten nHz-Gravitationswellen.

17.3 Das „Free Spectrum“: Messung des Gravitationswellenhintergrunds

Kollaborationen wie NANOGrav (*North American Nanohertz Observatory for Gravitational Waves*) beobachten seit Jahrzehnten dutzende Pulsare und werten die Korrelationen ihrer Timing-Residuen aus. Das Ergebnis wird oft als sogenanntes *free spectrum* dargestellt: Für jede von 14 Frequenzen $f_k = k/T_{\text{obs}}$ ($k = 1, \ldots, 14$, mit $T_{\text{obs}} = 16$ Jahren Beobachtungsdauer) wird ein Messwert für $\log_{10}\big(h^2\Omega_{\text{gw}}(f_k)\big)$ mit einer Unsicherheit angegeben, wobei $\Omega_{\text{gw}}(f)$ die Energiedichte der Gravitationswellen pro logarithmischem Frequenzintervall ist, normiert auf die kritische Energiedichte des Universums. Die Multiplikation mit $h^2 = 0.674^2 = 0.454$ (dem quadrierten Hubble-Parameter in Einheiten von $100\,\text{km/s/Mpc}$) macht die Messgröße unabhängig vom exakten Wert der Hubble-Konstante und folglich auch von der sog. *Hubble-Tension*. Die Beschreibung als free spectrum erlaubt also eine direkte Beschreibung der Stärke $h^2\Omega_{\text{gw}}(f_k)$ des kosmischen Gravitationswellenhintergrunds bei einer bestimmer Frequenz f_k.

Aufgabe 17.1

NANOGrav berichtet sein Ergebnis als “free spectrum”: für jede von 14 Frequenzen $f_k = k/T_{\text{obs}}$ ($k = 1, \ldots, 14$, $T_{\text{obs}} = 16$ Jahre die Beobachtungsdauer) gibt es einen Messwert für $\log_{10}\left(h^2\Omega_{\text{gw}}(f_k)\right)$ mit einer Unsicherheit. Damit ihr nicht auf einen Download von echten (sehr großen) Datensätzen warten müsst, geben wir euch eine vereinfachte Version dieser Messung als Zahlenwerte vor; sie ist aus den echten NANOGrav-15yr-Daten abgeleitet.
Das “free spectrum” modelliert jeden Frequenz-Bin k als unabhängige Gauß-Messung:

$$\log_{10}\big(h^2\Omega_{\text{gw}}(f_k)\big) \;\sim\; \mathcal{N}(\mu_k,\, \sigma_k^2)\,. \tag{17.1}$$

Dabei sind μ_k und σ_k der Erwartungswert und die Standardabweichung von $\log_{10}(h^2\Omega_{\text{gw}})$ – also *nicht* $\log_{10}\mu_k$ oder $\log_{10}\sigma_k$! Im Code unten heißen diese Größen entsprechend `mu_log10` (“Mittelwert im log10-Raum”) und `sigma_log10` (“Streuung im log10-Raum”).

a) Lege eine Datei `mock_likelihood.py` an und kopiere folgenden Code hinein:

```
import numpy as np

T_obs = 16.0 * 365.25 * 24 * 3600.0    # 16 Jahre, in Sekunden
f0 = 1.0 / T_obs                       # niedrigste Fourier-Frequenz (Hz)
n_bins = 14
freqs = f0 * np.arange(1, n_bins + 1)  # 14 Frequenz-Bins (Hz)
```

```
nHz = 1e-9                                  # 1 nHz in Hz, zur Umrechnung

# Mittelwert und Unsicherheit von log10(h^2 Omega_gw) in den 14
↪ Frequenz-Bins
mu_log10 = np.array([
    -10.1074, -8.9820, -8.9061, -8.9676, -10.0959, -10.3628, -10.6969,
     -8.1848, -10.1441, -9.1966, -9.2726, -9.6692, -9.6163, -9.3616,
])

sigma_log10 = np.array([
    0.3978, 0.1909, 0.2624, 0.5017, 1.7334, 1.9396, 1.3870, 1.2318,
    1.4141, 1.6662, 1.6218, 1.3132, 1.2425, 1.2703,
])
```

b) Plotte diese „Mock-Daten“: `freqs` auf der x-Achse (logarithmisch, `plt.xscale("log")`), `mu_log10` auf der y-Achse mit Fehlerbalken `yerr=sigma_log10` (benutze `plt.errorbar`).

Hinweis: `freqs` ist in Hz gegeben, das sind sehr kleine, unhandliche Zahlen. Überlegt euch, in welcher Einheit man Frequenzen im PTA-Band sinnvoll angibt (Größenordnung der Frequenzen? Es gibt schon eine passende Konstante in eurem Code…) und rechnet `freqs` entsprechend um, bevor ihr sie plottet.

c) *Diskussion:* Bei welchen Frequenzen ist das Frequenzspektrum am genauesten bestimmt?

Musterlösung 17.1

a) Die Datei `mock_likelihood.py` mit den Daten oben anlegen; diese Datei wird uns ab jetzt durch den Rest des Kurses begleiten.
b) Da `freqs` Werte um $10^{-9}\,\mathrm{Hz} = 1\,\mathrm{nHz}$ hat, bietet es sich an, in Einheiten von `nHz` zu plotten:

```
import matplotlib.pyplot as plt

plt.errorbar(freqs / nHz, mu_log10, yerr=sigma_log10, fmt="o")
plt.xscale("log")
plt.xlabel(r"$f$ [nHz]")
plt.ylabel(r"$\log_{10}(h^2 \Omega_{\rm gw})$")
plt.show()
```

c) Die kleinsten Unsicherheiten ($\sigma_{\log 10} \approx 0.19$–$0.26$) findet man bei den Frequenz-Bins 2 und 3 (f_2, f_3); hier ist die Messung also am genauesten. Bin 8 sticht heraus: hier ist $\mu \approx -8.18$ deutlich größer (also der Hintergrund stärker) als in den Nachbar-Bins, allerdings mit relativ großer Unsicherheit ($\sigma \approx 1.23$). Insgesamt erkennt man bereits in den Rohdaten (Abbildung 17.2) eine „Beule“ im Spektrum bei niedrigen Frequenzen – genau diese Beule werden wir in Kapitel 18 mit einem physikalischen Modell (Phasenübergang) erklären.

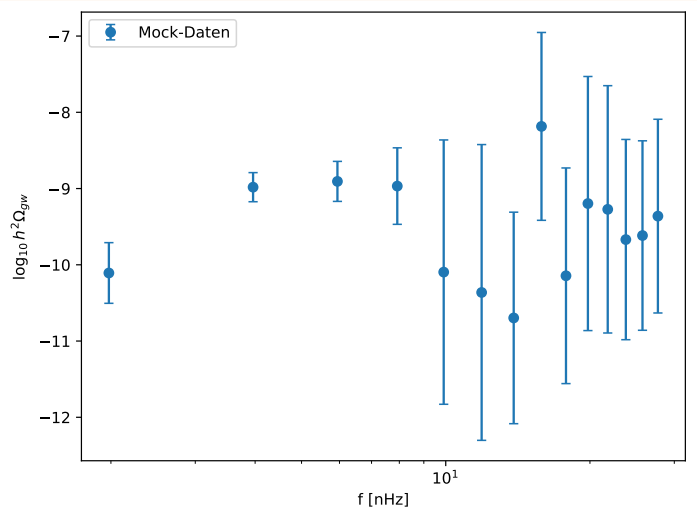


Abbildung 17.2: Die Mock-Daten: $\log_{10}(h^2\Omega_{\mathrm{gw}})$ als Funktion der Frequenz f (in nHz) mit Fehlerbalken $\sigma_{\log 10}$. Bei niedrigen Frequenzen erkennt man eine „Beule“ im Spektrum, genau die Struktur, die wir später mit einem Phasenübergangsmodell erklären werden.

Wir haben nun die Daten einer echten Gravitationswellenmessung der NANOGrav-Kollaboration (2023) in der Hand. Die spannende Frage ist: *Welches physikalische Modell passt am besten zu diesen 14 Datenpunkten, und mit welcher Unsicherheit lassen sich seine Parameter bestimmen?* Um diese Frage zu beantworten, benutzen wir die gleichen Methoden, die wir bereits im Schuhpendel-Beispiel kennengelernt haben: Wir definieren ein Modell, eine Likelihood, einen Prior und sampeln dann die Posterior mit dem Metropolis-Hastings-Algorithmus. Dann können wir uns Gedanken darüber machen, ob diese Modellparameter mit anderen kosmologischen Beobachtungen konsistent sind, und ob das Modell überhaupt physikalisch plausibel ist.

17.4 Alles zusammen: ein erstes Spektrum fitten

Jetzt bringen wir alles zusammen: den Metropolis-Hastings-Sampler aus Aufgabe 15.6 und die Likelihood basierend auf den Daten aus Aufgabe 17.1. Als Modell benutzen wir zunächst ein einfaches „Toy model“ mit der *Form* eines Peaks:

$$h^2\Omega_{\mathrm{gw}}(f) = A \cdot \frac{x}{1+x^2}\,, \qquad x = \frac{f}{f_{\mathrm{peak}}}\,. \tag{17.2}$$

Man kann leicht zeigen, dass $x/(1+x^2)$ sein Maximum bei $x=1$, also bei $f = f_{\mathrm{peak}}$, hat; f_{peak} ist also die Peak-Frequenz, und A legt die Höhe des Spektrums am Peak fest. Logarithmiert man Gleichung (17.2), wird aus dem Produkt eine Summe:

$$\log_{10}\!\big(h^2\Omega_{\mathrm{gw}}(f)\big) = \log_{10} A + \log_{10}\!\left(\frac{x}{1+x^2}\right). \tag{17.3}$$

Genau wie f_{peak} über viele Größenordnungen variieren kann, sampeln wir also $\log_{10} A$ und $\log_{10} f_{\mathrm{peak}}$ statt A und f_{peak} selbst.

Aufgabe 17.2

a) Schreibe eine Funktion

```
def broken_powerlaw_log10(f_hz, log10A, f_pivot_hz):
```

```
    x = f_hz / f_pivot_hz
    return 10**log10A + np.log10(x / (1.0 + x**2))
```

b) Definiere einen Gleichverteilungs-Prior für $\theta = (\log_{10} A, \log_{10} f_{\text{peak}})$:

```
LOG_A_MIN, LOG_A_MAX = -13.0, -4.0
LOG_F_PEAK_MIN, LOG_F_PEAK_MAX = -10.0, -7.0   # f_peak in [0.1, 100] nHz

def log_prior(theta):
    log10A, log_f_peak = theta
    if (LOG_A_MIN <= log10A <= LOG_A_MAX) and (LOG_F_PEAK_MIN <=
    ↪  log_f_peak <= LOG_F_PEAK_MAX):
        return 0.0
    return -np.inf
```

und schreibe, analog zu Aufgabe 15.3, eine Funktion `log_posterior(theta)`, die (i) `log_prior(theta)` auswertet und sofort `-np.inf` zurückgibt, falls dieser nicht endlich ist, (ii) sonst mit `broken_powerlaw_log10` das Modell an den 14 Frequenzen `freqs` berechnet, und (iii) `log_prior + log_likelihood(model)` zurückgibt.

c) Rufe den Sampler auf:

```
chain, logp, acc = metropolis_hastings(
    log_posterior, x0=[-8.0, -9.0], step_sizes=[0.45, 0.22],
    n_steps=50000, seed=42)
samples = chain[5000:]   # burn-in
```

Gib die Akzeptanzrate aus, erstelle einen corner plot von `samples` und bestimme den best-fit-Punkt sowie Median und 16%/84%-Perzentile für A und $\log_{10} f_{\text{pivot}}$. Diese befinden sich im Corner-Plot oberhalb der Panels auf der Diagonalen.

d) Plotte die Mock-Daten (wie in Aufgabe 17.1) zusammen mit der Modellkurve für den besten Parametersatz:

```
f_plot = np.logspace(np.log10(freqs[0]) - 1, np.log10(freqs[-1]) + 1, 500)
```

Welche Peak-Frequenz f_{peak} bevorzugen die Daten (in nHz)? Wie gut ist sie eingegrenzt?

Musterlösung 17.2

a)/b) Mit den Ergebnissen aus Aufgabe 15.3:

```
def broken_powerlaw_log10(f_hz, log10A, f_pivot_hz):
    x = f_hz / f_pivot_hz
    return 10**log10A + np.log10(x / (1.0 + x**2))

LOG_A_MIN, LOG_A_MAX = -13.0, -4.0
LOG_F_PEAK_MIN, LOG_F_PEAK_MAX = -10.0, -7.0

def log_prior(theta):
    log10A, log_f_peak = theta
    if (LOG_A_MIN <= log10A <= LOG_A_MAX) and (LOG_F_PEAK_MIN <= log_f_peak <=
    ↪  LOG_F_PEAK_MAX):
        return 0.0
    return -np.inf
```

```
def log_posterior(theta):
    lp = log_prior(theta)
    if not np.isfinite(lp):
        return -np.inf
    log10A, log_f_peak = theta
    model = broken_powerlaw_log10(freqs, log10A, 10**log_f_peak)
    return lp + log_likelihood(model)
```

c) Der Aufruf von `metropolis_hastings` mit den angegebenen Einstellungen liefert eine Akzeptanzrate von acc ≈ 0.3, wieder im gewünschten Bereich 0.2–0.5. Der corner plot (Abbildung 17.3) zeigt für $\log_{10} A$ und $\log_{10} f_{\mathrm{peak}}$ jeweils eine annähernd glockenförmige marginalisierte Verteilung; im 2D-Panel erkennt man eine leichte Korrelation zwischen $\log_{10} A$ und $\log_{10} f_{\mathrm{peak}}$ (eine höhere Amplitude A „passt“ zu einer leicht anderen Peak-Frequenz, ohne die Daten schlechter zu erklären, eine *Entartung*, ganz analog zu der zwischen β/H und T_* in Kapitel 18). Best-fit-Punkt und die 16%/84%-Perzentile lassen sich direkt aus dem Corner-Plot ablesen.

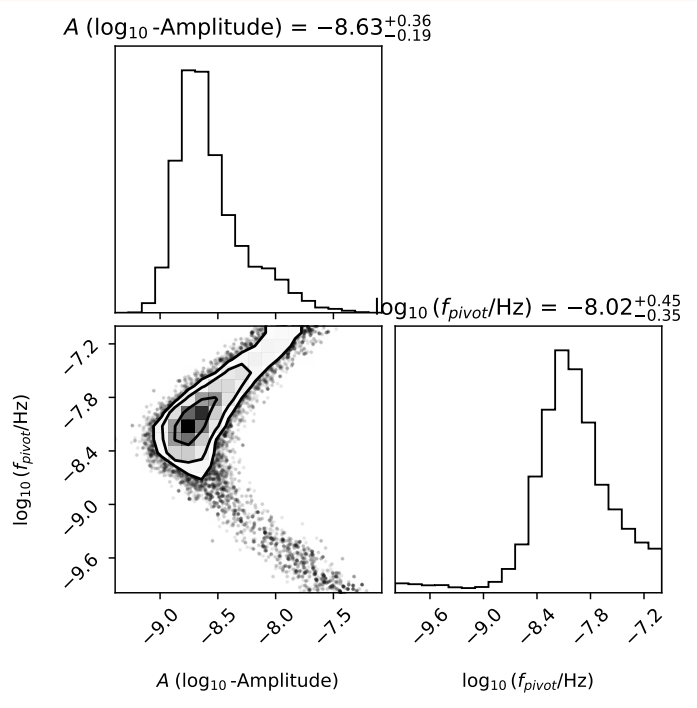


Abbildung 17.3: Corner plot der Samples für $\log_{10} A$ und $\log_{10} f_{\mathrm{peak}}$. Beide Parameter sind gut bestimmt (annähernd glockenförmige marginalisierte Verteilungen); im 2D-Panel ist eine leichte Korrelation (Entartung) zwischen den beiden Parametern zu erkennen.

d) Die beste Modellkurve folgt der „Beule“ in den Mock-Daten (vgl. Aufgabe 17.1c)) und hat ihren Peak bei einer Frequenz f_{peak} von einigen nHz – also genau im Bereich, in dem die Mock-Daten die größten Werte von $\log_{10} h^2\Omega_{\mathrm{gw}}$ zeigen. Das 16%/84%-Intervall für $\log_{10} f_{\mathrm{peak}}$ erstreckt sich dabei über etwa eine Größenordnung – die Daten legen die Peak-Frequenz also schon recht gut fest, aber nicht beliebig genau.

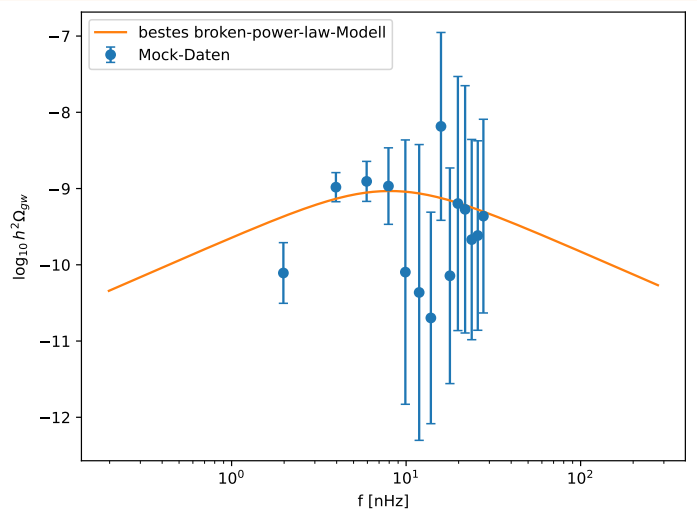


Abbildung 17.4: Die PTA-Daten (Punkte mit Fehlerbalken) zusammen mit der besten Modellkurve $h^2\Omega_{\mathrm{gw}}(f)$ für den best-fit-Parametersatz. Die Modellkurve folgt der „Beule“ bei niedrigen Frequenzen.

Damit haben wir zum ersten Mal erfolgreich ein Modell an einen (simulierten) Gravitationswellenhintergrund gefittet (Abbildung 17.4)! Im Rest dieses Kapitels ersetzen wir dieses noch unmotivierte „Toy model“ schrittweise durch *physikalische* Modelle, zunächst eine Verallgemeinerung des Toy models, dann das Spektrum eines Phasenübergangs im frühen Universum mit den Parametern $\alpha, \beta/H, T_*$ aus Kapitel 18. Der entscheidende Punkt dabei: Der Sampler `metropolis_hastings` selbst bleibt dabei *komplett unverändert*; nur das Modell in `log_posterior` wird ausgetauscht. Das ist die ganze Stärke eines allgemeinen MCMC-Samplers: Einmal geschrieben, lässt er sich auf praktisch *jedes* Modell anwenden, das sich als `log_posterior(theta)` schreiben lässt.

17.5 Extra-Aufgabe: Mehr Parameter, derselbe Sampler

Das Toy model aus Gleichung (17.2) aus Aufgabe 17.2 hat eine feste Form: links von f_{peak} steigt $h^2\Omega_{\mathrm{gw}}(f) \propto x$, rechts davon fällt es $\propto x^{-1}$ (mit $x = f/f_{\mathrm{peak}}$).

Aufgabe 17.3

[Für die Schnellen] Verallgemeinere das Toy model zu

$$\log_{10}\!\left(h^2\Omega_{\mathrm{gw}}(f)\right) = \log_{10} A + \log_{10}\!\left(\frac{x^a}{1+x^c}\right), \qquad x = \frac{f}{f_{\mathrm{peak}}}, \tag{17.4}$$

mit zwei zusätzlichen freien Parametern a und c (für $a = 1, c = 2$ ist das genau das Modell aus Gleichung (17.2) aus Aufgabe 17.2).

a) Was bedeuten a und $c - a$ für die Steigung des Spektrums links bzw. rechts von f_{peak} (jeweils auf einer doppelt-logarithmischen Achse)?

b) Erweitere `log_prior` und `log_posterior` so, dass $\theta = (\log_{10} A, \log_{10} f_{\mathrm{peak}}, a, c)$ jetzt *vier* Einträge hat. Wähle sinnvolle (gleichverteilte) Prior-Bereiche für a und c, z.B. $a \in [0.5, 5]$, $c \in [1, 8]$.

c) Muss sich an `metropolis_hastings` selbst irgendetwas ändern, damit es mit 4 statt 2 Parametern funktioniert? Probiere es aus; du brauchst nur `x0` und `step_sizes` auf vier Einträge zu erweitern, z.B. `x0=[-8.0,-9.0,1.0,2.0]`, `step_sizes=[0.2,0.2,0.2,0.35]`, `n_steps=100000`. Was sagt dir das über die Allgemeinheit des Algorithmus aus Aufgabe 15.6?

d) Erstelle wieder einen corner plot und einen Daten-Plot. Welche Werte für a und c bevorzugen die Daten? Vergleiche mit $a = 1, c = 2$ aus Aufgabe 17.2; ist das einfachere Toy model eine gute Näherung?

Musterlösung 17.3

a) Für $x \ll 1$ (also $f \ll f_{\text{peak}}$) gilt $x^c \ll 1$, sodass $\frac{x^a}{1+x^c} \approx x^a$ und damit $\log_{10}(h^2\Omega_{\text{gw}}) \approx \log_{10} A + a \log_{10} x$; a ist also die Steigung des Spektrums *links* von f_{peak} in einer doppeltlogarithmischen Darstellung. Für $x \gg 1$ gilt $\frac{x^a}{1+x^c} \approx x^{a-c}$, also $\log_{10}(h^2\Omega_{\text{gw}}) \approx \log_{10} A + (a - c) \log_{10} x$ – die Steigung *rechts* von f_{peak} ist $a - c$, also um $c - a$ *flacher* (steiler abfallend) als links. Für $a = 1, c = 2$ ergibt das genau die Steigungen $+1$ und -1 des ursprünglichen Toy models (17.2).
b) Wir erweitern Prior, Modellfunktion und Posterior um die zwei zusätzlichen Parameter:

```
A_MIN, A_MAX = -13.0, -4.0
LOG_F_PEAK_MIN, LOG_F_PEAK_MAX = -10.0, -7.0
A_EXP_MIN, A_EXP_MAX = 0.5, 5.0
C_EXP_MIN, C_EXP_MAX = 1.0, 8.0

def log_prior(theta):
    A, log_f_peak, a, c = theta
    if not (A_MIN <= A <= A_MAX):
        return -np.inf
    if not (LOG_F_PEAK_MIN <= log_f_peak <= LOG_F_PEAK_MAX):
        return -np.inf
    if not (A_EXP_MIN <= a <= A_EXP_MAX):
        return -np.inf
    if not (C_EXP_MIN <= c <= C_EXP_MAX):
        return -np.inf
    return 0.0

def broken_powerlaw_general_log10(f_hz, A, f_peak_hz, a, c):
    x = f_hz / f_peak_hz
    return A + np.log10(x**a / (1.0 + x**c))

def log_posterior(theta):
    lp = log_prior(theta)
    if not np.isfinite(lp):
        return -np.inf
    A, log_f_peak, a, c = theta
    model = broken_powerlaw_general_log10(freqs, A, 10**log_f_peak, a, c)
    return lp + log_likelihood(model)
```

c) An `metropolis_hastings` selbst muss *nichts* geändert werden; `ndim` wird automatisch aus der Länge von `x0` bestimmt, und `rng.normal(0.0, step_sizes, size=ndim)` funktioniert für beliebig viele Dimensionen. Der Algorithmus kennt `log_posterior` nur als „Black Box“, die für ein θ eine einzelne Zahl zurückgibt; wie viele Einträge θ hat, spielt für den Sampler keine Rolle.

d) Der corner plot (Abbildung 17.5) zeigt, dass a und c von den Daten tatsächlich eingegrenzt werden, allerdings nicht sonderlich präzise: Für nur 14 Datenpunkte sind 4 freie Parameter schon eine ganze Menge. Die bevorzugten Werte liegen typischerweise nahe bei $a \approx 3$ und $c \approx 4.5$, also weit entfernt vom einfachen Toy model aus Aufgabe 17.2: Die zusätzliche Freiheit in a und c wird von den Daten nicht zwingend benötigt, das einfachere Modell ist also bereits eine gute Näherung (Abbildung 17.6).

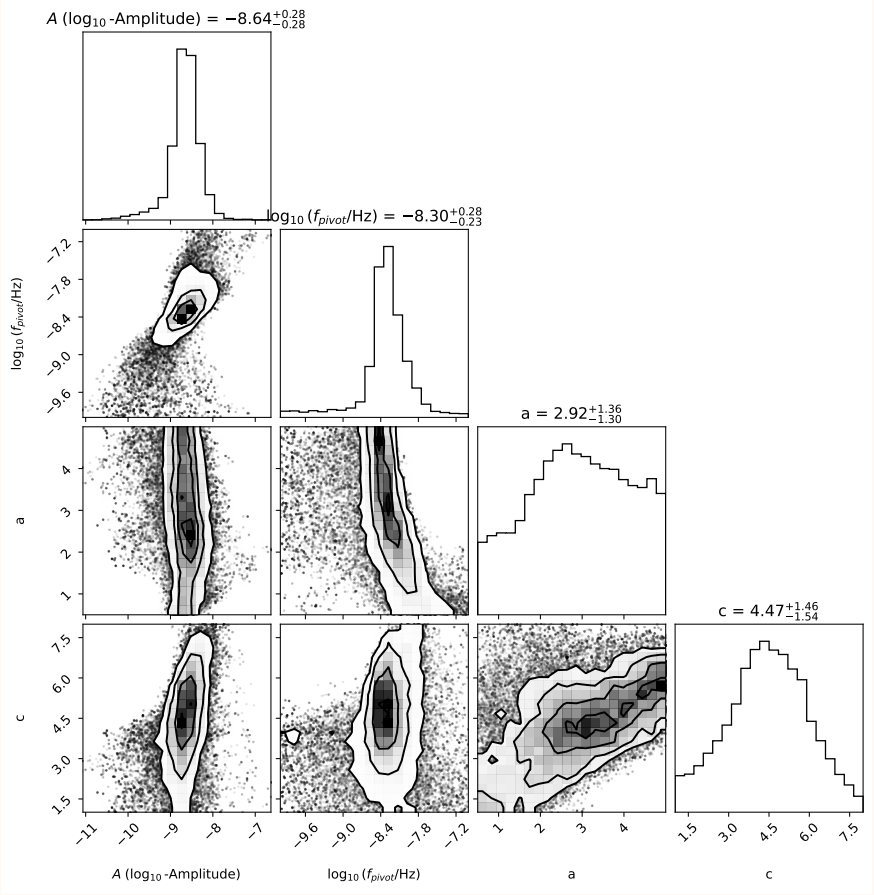


Abbildung 17.5: Corner plot für das verallgemeinerte Modell aus Gleichung (17.4) mit den vier Parametern $\log_{10} A$, $\log_{10} f_{\text{peak}}$, a und c.

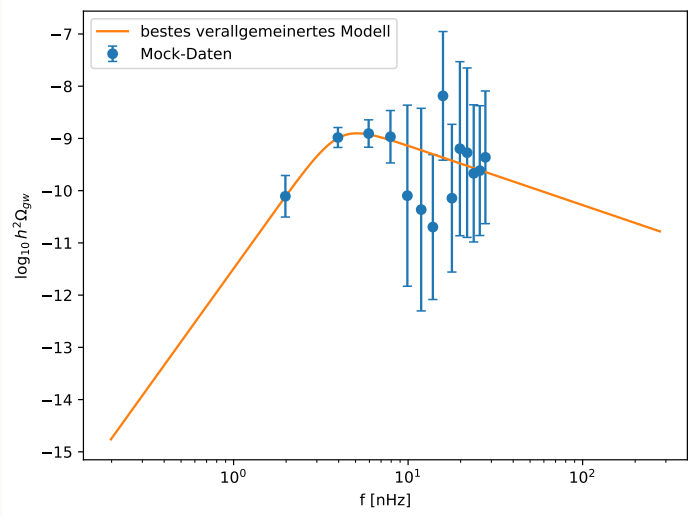


Abbildung 17.6: Mock-Daten und bestes verallgemeinertes Modell aus Gleichung (17.4).

18 Phasenübergänge im frühen Universum

In Kapitel 1 haben wir gesehen, dass das frühe Universum extrem heiß und dicht war und sich seitdem stetig abgekühlt hat. Genau wie Wasser beim Abkühlen von Dampf zu Flüssigkeit zu Eis *Phasenübergänge* durchläuft, durchlief auch das frühe Universum mehrere solcher Übergänge, z.B. den elektroschwachen Phasenübergang (bei dem das Higgs-Feld seinen heutigen Wert annahm) und den QCD-Phasenübergang (bei dem Quarks und Gluonen sich zu Protonen und Neutronen zusammenschlossen, vgl. Kapitel 1).

18.1 Was ist ein Feld?

Bevor wir uns Phasenübergängen im Detail widmen, klären wir einen zentralen Begriff der modernen Physik: das *Feld*. Intuitiv ist ein Feld eine Größe, die *jedem Punkt im Raum* (und ggf. in der Zeit) einen Wert zuordnet.

Skalarfelder ordnen jedem Raumpunkt eine einzige Zahl zu. Ein alltägliches Beispiel ist die *Temperatur*: Hält man ein Thermometer in einen Raum, misst man an jedem Ort (x, y, z) eine Temperatur $T(x, y, z, t)$, eine Zahl pro Raumpunkt. In der Physik ist das *Higgs-Feld* $\phi(x, y, z, t)$ ein Skalarfeld: Es hat an jedem Punkt der Raumzeit einen einzigen reellen (oder komplexen) Wert, der die Stärke der Wechselwirkung mit den Elementarteilchen angibt.

Vektorfelder ordnen jedem Raumpunkt einen *Vektor* zu, also eine Richtung und eine Stärke. Ein anschauliches Beispiel ist die *Windgeschwindigkeit*: An jedem Ort gibt es einen Windvektor $\vec{v}(x, y, z, t)$, der angibt, wie schnell und in welche Richtung die Luft strömt. In der Physik sind das *elektrische Feld* $\vec{E}(\vec{x}, t)$ und das *magnetische Feld* $\vec{B}(\vec{x}, t)$ Vektorfelder, jeweils drei Zahlen pro Raumpunkt.

Tensorfelder sind die nächste Stufe: Sie ordnen jedem Punkt eine *Matrix* von Zahlen zu. Das wichtigste Beispiel in der Physik ist das *Gravitationsfeld*, genauer die *Metrik* $g_{\mu\nu}(x)$ der Allgemeinen Relativitätstheorie. Sie beschreibt, wie Abstände und Zeiten in der gekrümmten Raumzeit gemessen werden. Da $\mu, \nu \in \{0, 1, 2, 3\}$ (Zeit + drei Raumkoordinaten), hat die Metrik $4 \times 4 = 16$ Komponenten pro Raumpunkt – aber wegen der Symmetrie $g_{\mu\nu} = g_{\nu\mu}$ sind es nur 10 unabhängige Zahlen. Das Gravitationsfeld ist also fundamentell komplizierter als ein Skalar- oder Vektorfeld: Es beschreibt nicht eine einzelne Größe oder eine Richtung, sondern die gesamte Geometrie des Raums.

Diese Hierarchie aus Skalaren, Vektoren und Tensoren ist grundlegend für die Physik der Felder. Phasenübergänge sind Situationen, in denen ein Skalarfeld (wie das Higgs-Feld) seinen Gleichgewichtswert ändert.

18.2 Das mechanische Bild: eine Kugel im Potential

Stell dir eine Kugel vor, die sich in einer „Landschaft“ $V(\phi)$ bewegt; ϕ ist hier ein Platzhalter für ein Feld (z.B. das Higgs-Feld), $V(\phi)$ seine potentielle Energie. Bei hohen Temperaturen hat diese Landschaft typischerweise ein einziges Minimum bei $\phi = 0$ (die Kugel liegt „oben“, das Feld ist im symmetrischen Zustand). Kühlt das Universum ab, kann sich die Form von $V(\phi)$ ändern: es kann sich ein zweites, tieferes Minimum bei $\phi = \phi_0 \neq 0$ ausbilden. Das System „will“ in dieses neue, energetisch günstigere Minimum übergehen: Das ist der Phasenübergang.

Es gibt dabei zwei sehr unterschiedliche Szenarien:

- **Crossover (glatter Übergang):** Das alte Minimum verschiebt sich kontinuierlich zum neuen: Die Kugel „rollt“ sanft von $\phi = 0$ nach $\phi = \phi_0$, ohne dass es jemals eine Barriere gibt. Das gesamte Universum macht diesen Übergang gleichzeitig und gleichmäßig durch. Es werden *keine* nennenswerten Gravitationswellen erzeugt.

- **Phasenübergang erster Ordnung (Übergang mit Barriere):** Zwischen dem alten Minimum bei $\phi = 0$ und dem neuen, tieferen Minimum bei $\phi = \phi_0$ bildet sich eine *Potentialbarriere* aus. Das Feld kann diese Barriere nicht überall gleichzeitig „überqueren“;

stattdessen entstehen an einzelnen, zufälligen Orten kleine Bereiche („Blasen“), in denen das Feld den Sprung über die Barriere via Quantentunneln oder thermische Fluktuationen geschafft hat und bereits im neuen Minimum ϕ_0 sitzt.

Abbildung 18.1 zeigt das effektive Potential für beide Fälle.

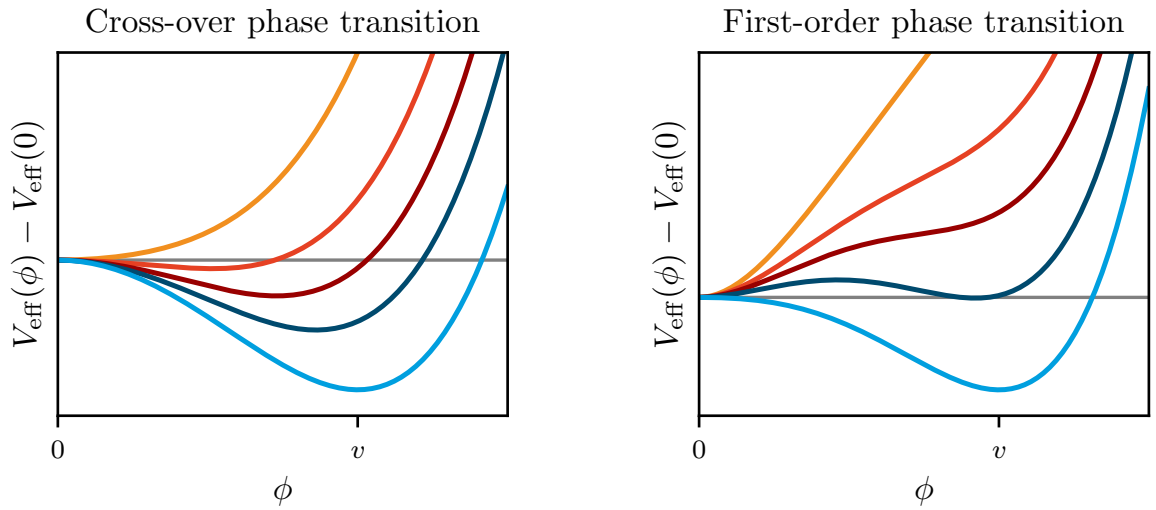


Abbildung 18.1: *Links:* Crossover-Phasenübergang: Das Minimum des effektiven Potenzials $V_{\text{eff}}(\phi, T)$ verschiebt sich beim Abkühlen kontinuierlich; die Kurven (von Hoch- zu Tieftemperatur) zeigen, wie das Minimum glatt von $\phi = 0$ nach $\phi = \phi_0$ wandert, ohne je eine Barriere zu überqueren. *Rechts:* Phasenübergang erster Ordnung: Beim Abkühlen bildet sich eine Potentialbarriere zwischen $\phi = 0$ und $\phi = \phi_0$. Das Feld muss diese Barriere via Quantentunneln oder thermische Fluktuationen überwinden.

18.3 Bubble Nucleation und Kollisionen

Bei einem Phasenübergang erster Ordnung entstehen also winzige Blasen der „neuen Phase“ ($\phi = \phi_0$) inmitten der „alten Phase“ ($\phi = 0$); man spricht von *bubble nucleation*. Da die neue Phase energetisch günstiger ist, wachsen diese Blasen: Ihre Wände breiten sich (oft nahezu mit Lichtgeschwindigkeit) ins umgebende Medium aus, bis benachbarte Blasen aufeinandertreffen und verschmelzen (*bubble-wall collisions*). Abbildung 18.2 skizziert diesen Prozess schematisch; Abbildung 18.3 zeigt Schnappschüsse einer numerischen Fluidsimulation.

Diese kollidierenden Blasenwände erzeugen, zusammen mit den Schallwellen und der Turbulenz, die sie im Plasma des frühen Universums hinterlassen, ein stochastisches Gravitationswellen-Hintergrundsignal. Man unterscheidet drei Beiträge:

1. Kollisionen der Blasenwände selbst,
2. Schallwellen, die nach den Kollisionen im Plasma weiterlaufen,
3. (magnetohydrodynamische) Turbulenz im Plasma.

Alle drei Beiträge erzeugen Gravitationswellen mit einem charakteristischen, breiten Spektrum, das ein Maximum bei einer bestimmten Peak-Frequenz f_p hat und zu höheren und niedrigeren Frequenzen hin abfällt.

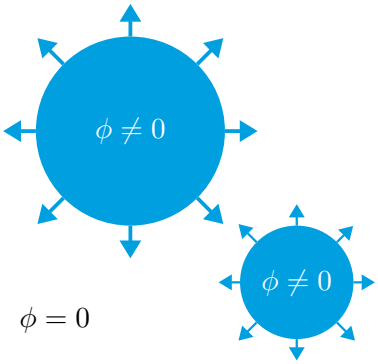


Abbildung 18.2: Skizze der Blasenbildung (*bubble nucleation*) bei einem Phasenübergang erster Ordnung. In den blauen Blasen hat das Feld bereits den Sprung über die Potentialbarriere geschafft und sitzt im neuen Minimum $\phi \neq 0$, während es im umgebenden Vakuum noch im alten Minimum $\phi = 0$ verharrt. Die Blasenwände expandieren (Pfeile) nach außen, wobei die größere Blase bereits weiter gewachsen als die kleinere, bis benachbarte Blasen kollidieren und verschmelzen.

18.4 Das Spektrum eines Phasenübergangs

Gemäß Referenz [12] lässt sich das resultierende Gravitationswellen-Spektrum durch drei physikalische Parameter beschreiben:

- α: die *Stärke* des Phasenübergangs (die beim Übergang freigesetzte Energie relativ zur Strahlungsenergiedichte des Universums zu diesem Zeitpunkt),
- β/H: die *inverse Dauer* des Phasenübergangs (große Werte = schneller Phasenübergang, H ist die Hubble-Rate zum Zeitpunkt des Übergangs),
- T_*: die *Temperatur* des Universums zum Zeitpunkt des Phasenübergangs.

Das Spektrum ist gegeben durch

$$\Omega_{\mathrm{gw}}(f) \simeq 10^{-6} \left(\frac{\alpha}{1+\alpha}\right)^2 \left(\frac{H}{\beta}\right)^2 s(f/f_p)\,, \tag{18.1}$$

mit der Peak-Frequenz

$$f_p \simeq 10\,\mathrm{nHz} \cdot \left(\frac{T_*}{100\,\mathrm{MeV}}\right) \cdot \left(\frac{\beta}{H}\right) \tag{18.2}$$

und der Spektralform

$$s(x) = \frac{3.8\,x^{2.8}}{1 + 2.8\,x^{3.8}}\,. \tag{18.3}$$

Beachte, wie $\Omega_{\mathrm{gw}}(f)$ aufgebaut ist: Der Vorfaktor $10^{-6}\left(\frac{\alpha}{1+\alpha}\right)^2\left(\frac{H}{\beta}\right)^2$ bestimmt die *Höhe* des Spektrums (stärkere und schnellere Phasenübergänge erzeugen lautere Signale), während $s(f/f_p)$ die *Form* bestimmt: ein Peak bei $f = f_p$, der zu beiden Seiten hin abfällt. Die Peak-Frequenz f_p wiederum hängt über β/H und T_* direkt mit der Physik des Phasenübergangs zusammen, vgl. Gleichung (18.2). Abbildung 18.4 zeigt den Einfluss jedes Parameters auf das Spektrum.

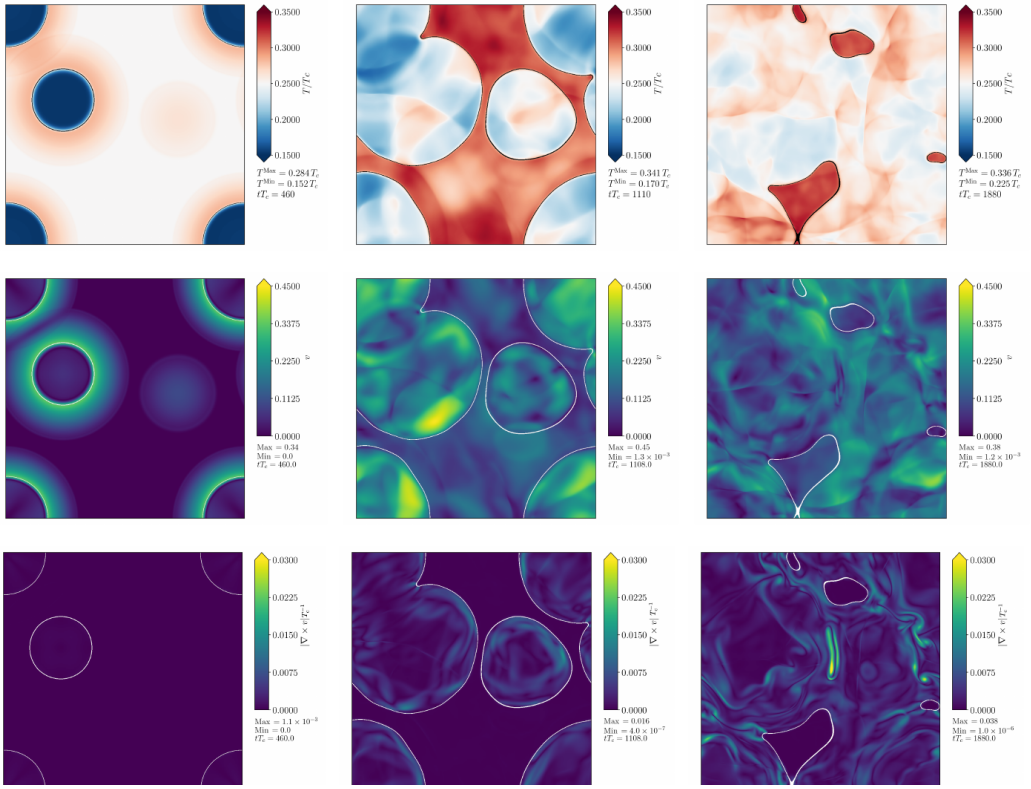


Abbildung 18.3: Numerische Fluidsimulation eines Phasenübergangs erster Ordnung (Deflagration, $v_w = 0.44$, $\alpha = 0.5$): Drei Schnappschüsse durch die (x, y)-Ebene zu frühen, mittleren und späten Zeiten (Spalten). *Oben:* Temperatur T/T_c: die Blaseninnereien (blau, heiß) expandieren ins noch heiße Außenmedium (rot). *Mitte:* Fluidgeschwindigkeit v, konzentriert um die Blasenwände. *Unten:* Vortizität $|\nabla \times v|$: ein Maß für Turbulenz, die nach den Kollisionen entsteht. Abbildung aus Referenz [11].

Warum Nanohertz? Die Antwort braucht zwei Zutaten: die Emissionsfrequenz beim Phasenübergang, und die kosmische Rotverschiebung bis heute.

Emissionsfrequenz. Die Blasen der neuen Phase wachsen, bis sie kollidieren; ihre typische Größe beim Aufeinandertreffen ist $\ell \sim \beta^{-1}$. Die emittierten Gravitationswellen haben daher Frequenz $f_{\text{em}} \sim \beta = H \cdot (\beta/H)$. In der strahlungsdominierten Ära gilt $H \sim T^2/M_{\text{Pl}}$ (Kapitel 8). Für $T_* = 100\,\text{MeV}$ und $M_{\text{Pl}} = 1{,}22 \times 10^{19}\,\text{GeV}$:

$$H(T_*) \sim \frac{(10^{-1}\,\text{GeV})^2}{1{,}22 \times 10^{19}\,\text{GeV}} \approx 8 \times 10^{-22}\,\text{GeV} \approx 1{,}2 \times 10^{3}\,\text{Hz}\,.$$

Mit $\beta/H \sim 10$–100 liegt $f_{\text{em}} \sim 10^4$–$10^5\,\text{Hz}$, also im Kilohertz- bis Megahertz-Bereich.

Woher kommt der Faktor T_0/T_*? In Kapitel 8 haben wir die Entropieerhaltung im frühen Universum gesehen: $g_{*S}\, T^3 a^3 = \text{const}$. Für näherungsweise konstante g_{*S} folgt daraus

$T \propto 1/a$ – die Temperatur fällt genau wie die Energie eines einzelnen Photons ($E_\gamma = hf \propto a^{-1}$). Daher rotverschieben Frequenzen proportional zur Temperatur:

$$\frac{f_{\text{obs}}}{f_{\text{em}}} = \frac{a_{\text{em}}}{a_0} = \frac{T_0}{T_*} \approx \frac{0{,}24\,\text{meV}}{100\,\text{MeV}} \approx 2 \times 10^{-12}\,.$$

Ergebnis.

$$f_{\text{obs}} \sim (10^4\text{–}10^5\,\text{Hz}) \times 2 \times 10^{-12} \approx (0{,}02\text{–}0{,}2)\,\text{nHz}$$

Genau das PTA-Band! Der Faktor T_0/T_* steckt explizit in Gleichung (18.2): $f_p \propto T_*$ ist nichts anderes als diese Temperatur-Rotverschiebung.

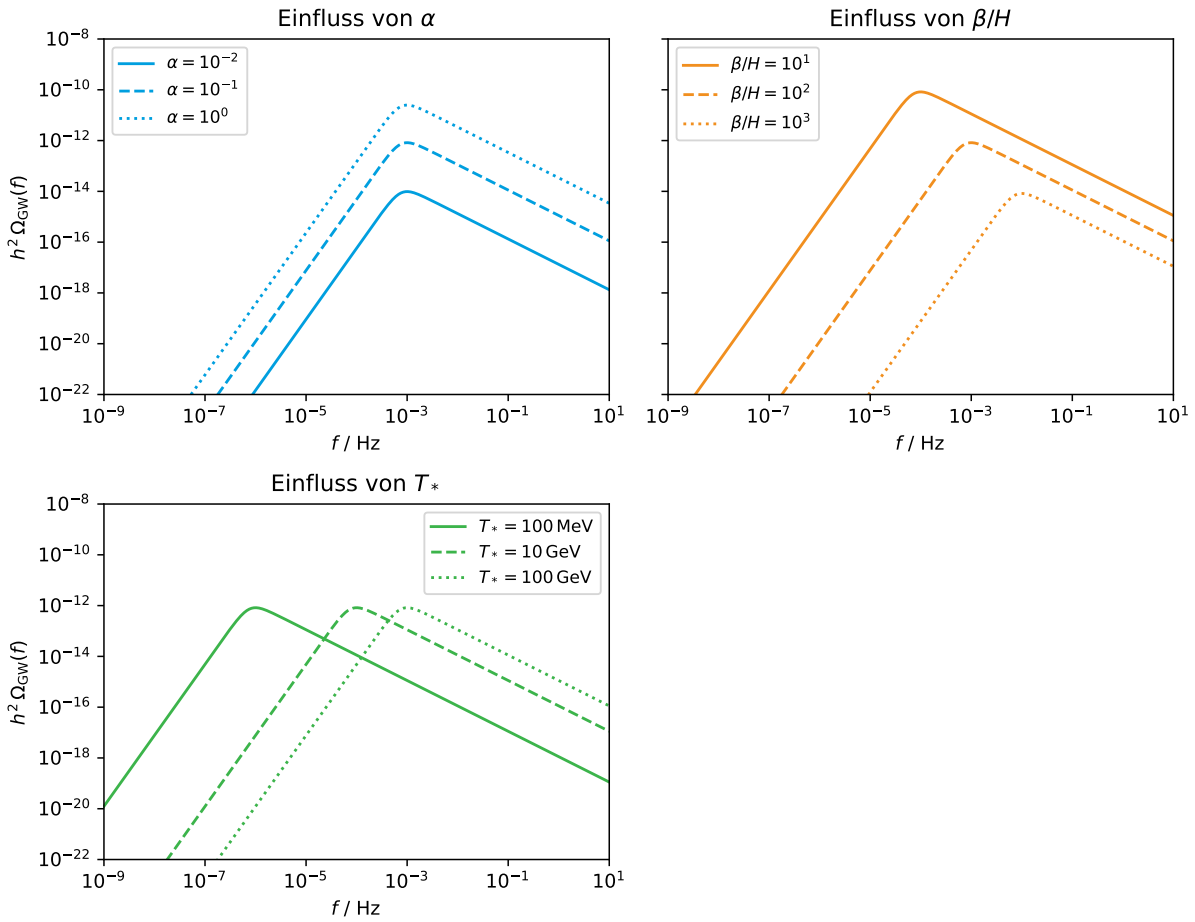


Abbildung 18.4: Einfluss der drei Parameter auf das Spektrum $h^2\Omega_{\text{gw}}(f)$, jeweils bei festgehaltenen Werten der anderen zwei. *Oben links:* Variation von α (bei $\beta/H = 100$, $T_* = 100\,\text{GeV}$) – größeres α erhöht die Amplitude deutlich. *Oben rechts:* Variation von β/H – größeres β/H senkt die Amplitude und verschiebt den Peak zu höheren Frequenzen. *Unten links:* Variation von T_* – höhere T_* verschiebt den Peak zu höheren Frequenzen (näher ans LISA-Band).

Aufgabe 18.1

Ein Phasenübergang erster Ordnung läuft ab, indem sich Blasen der neuen Phase im Universum bilden, wachsen und schließlich zusammenstoßen (*bubble-wall collisions*). Dabei wird ein Gravitationswellenhintergrund erzeugt, der sich (gemäß Referenz [13]) durch drei Parameter beschreiben lässt:

- α: Stärke des Phasenübergangs (freigesetzte Energie relativ zur Strahlungsenergiedichte des Universums),
- β/H: inverse Dauer des Phasenübergangs (große Werte = schneller Phasenübergang),
- T_* [MeV]: Temperatur des Universums zum Zeitpunkt des Phasenübergangs.

Das Spektrum ist gegeben durch

$$\Omega_{\mathrm{gw}}(f) \simeq 10^{-6} \left(\frac{\alpha}{1+\alpha}\right)^2 \left(\frac{H}{\beta}\right)^2 s(f/f_p)\,,$$

mit der Peak-Frequenz

$$f_p \simeq 10\,\mathrm{nHz} \cdot \left(\frac{T_*}{100\,\mathrm{MeV}}\right) \cdot \left(\frac{\beta}{H}\right)$$

und der Spektralform

$$s(x) = \frac{3.8\,x^{2.8}}{1+2.8\,x^{3.8}}\,.$$

a) Setze $x = 1$ in die Gleichung für $s(x)$ ein. Welche Interpretation hat die Frequenz f_p?

b) Lege eine neue Datei `phase_transition_fit.py` an. Implementiere die drei Gleichungen als Python-Funktionen:

```
def s_of_x(x):
    return 3.8 * x**2.8 / (1.0 + 2.8 * x**3.8)

def f_peak_hz(beta_over_h, t_star_mev):
    return 10.0 * nHz * (t_star_mev / 100.0) * beta_over_h

def pt_spectrum_log10(f_hz, alpha, beta_over_h, t_star_mev):
    f_p = f_peak_hz(beta_over_h, t_star_mev)
    omega = (1e-6 * (alpha / (1.0 + alpha))**2 * (1.0 / beta_over_h)**2
             * s_of_x(f_hz / f_p))
    return np.log10(omega)
```

(`nHz = 1e-9`)

c) Anders als bei einem reinen „Toy model“ haben α, β/H und T_* jetzt eine *physikalische Bedeutung*. Was würde es bedeuten, wenn $\alpha \gg 1$ wäre? Was, wenn $\beta/H \gg 1$ (verglichen mit $\beta/H \sim 1$)?

Musterlösung 18.1

a) Für $x = 1$ wird $s(1) = \frac{3.8}{1+2.8} = \frac{3.8}{3.8} = 1$. Das bedeutet, dass $s(x)$ bei $x = f/f_p = 1$, also bei $f = f_p$, den Wert 1 annimmt und damit (man kann zeigen, dass $s(x) \leq 1$ für alle

$x > 0$, mit Maximum bei $x = 1$) sein *Maximum*. f_p ist also tatsächlich die Frequenz, bei der das Spektrum $\Omega_{\mathrm{gw}}(f)$ seinen größten Wert $\sim 10^{-6}\left(\frac{\alpha}{1+\alpha}\right)^2\left(\frac{H}{\beta}\right)^2$ annimmt; der Name „Peak-Frequenz“ ist also gerechtfertigt.
b) Die drei Funktionen lassen sich direkt aus den Gleichungen (18.1)–(18.3) übersetzen:

```
import numpy as np

nHz = 1e-9

def s_of_x(x):
    return 3.8 * x**2.8 / (1.0 + 2.8 * x**3.8)

def f_peak_hz(beta_over_h, t_star_mev):
    return 10.0 * nHz * (t_star_mev / 100.0) * beta_over_h

def pt_spectrum_log10(f_hz, alpha, beta_over_h, t_star_mev):
    f_p = f_peak_hz(beta_over_h, t_star_mev)
    omega = (1e-6 * (alpha / (1.0 + alpha))**2 * (1.0 / beta_over_h)**2
             * s_of_x(f_hz / f_p))
    return np.log10(omega)
```

Man beachte, dass `pt_spectrum_log10` direkt $\log_{10}\Omega_{\mathrm{gw}}(f)$ zurückgibt; in Kapitel 17 sampeln wir nämlich wieder in Logarithmen (vgl. die Diskussion logarithmischer Größenbereiche, z.B. in Kapitel 5).
c) Für $\alpha \gg 1$ würde der Vorfaktor $\left(\frac{\alpha}{1+\alpha}\right)^2 \to 1$ gehen; die beim Phasenübergang freigesetzte Energie wäre also viel größer als die ohnehin schon im Universum vorhandene Strahlungsenergie. Ein solcher Phasenübergang wäre extrem „stark“ (engl. *strongly first-order*) und würde ein besonders lautes Gravitationswellensignal erzeugen.
Für $\beta/H \gg 1$ wäre der Vorfaktor $\left(\frac{H}{\beta}\right)^2 \ll 1$; das Signal wäre also *schwächer*. Physikalisch bedeutet $\beta/H \gg 1$, dass der Phasenübergang sehr *schnell* im Vergleich zur Expansionsrate des Universums abläuft (viele Blasen nukleieren und kollidieren, bevor sich die Skalen des Universums merklich ändern) einen „abrupten“ Übergang. Gleichzeitig verschiebt ein großes β/H über (18.2) auch die Peak-Frequenz f_p zu höheren Werten.

Die in diesem Kapitel besprochenen Gravitationswellen aus Phasenübergängen sind ein Beispiel für ein *stochastisches Gravitationswellen-Hintergrundsignal*, ein Signal, das (anders als z.B. das „Chirp“-Signal zweier verschmelzender schwarzer Löcher) nicht aus einer einzelnen, identifizierbaren Quelle stammt, sondern aus der Überlagerung unzähliger Quellen über die gesamte Geschichte des Universums. In Kapitel 11 haben wir bereits gelernt, was Gravitationswellen sind und wie man sie misst. Im nächsten Kapitel (17) sehen wir nun, wie man mit Pulsaren genau nach einem solchen stochastischen Hintergrund sucht.

18.5 Der Phasenübergang als physikalisches Modell

Das verallgemeinerte Toy model aus Aufgabe 17.3 hat zwar mehr Freiheit, aber seine vier Parameter $A, f_{\mathrm{pivot}}, a, c$ haben immer noch *keine direkte physikalische Bedeutung*. Jetzt ersetzen wir es durch das Spektrum eines Phasenübergangs erster Ordnung aus Kapitel 18, Gleichungen (18.1)–(18.3), mit den drei physikalischen Parametern $\theta = (\log_{10}\alpha, \log_{10}(\beta/H), \log_{10}T_*)$; genau wie bei A und f_{pivot} sampeln wir auch hier in den Logarithmen, da alle drei Größen über viele Größenordnungen variieren können.

Aufgabe 18.2

a) Aus physikalischen Gründen muss $\beta/H \geq 1$ sein: Wäre $\beta/H < 1$, würde sich der Raum zwischen zwei benachbarten Blasen schneller ausdehnen, als die Blasenwände wachsen können; die Blasen würden sich also nie einholen und könnten nicht kollidieren. Da das Spektrum aus Gleichung (18.1) aber gerade durch solche *bubble-wall collisions* erzeugt wird, brauchen wir $\beta/H \geq 1$. Wir wählen großzügige, aber endliche Prior-Bereiche:

```
LOG_ALPHA_MIN, LOG_ALPHA_MAX = -2.0, 1.0    # alpha in [0.01, 10]
LOG_BOH_MIN, LOG_BOH_MAX = 0.0, 2.0         # beta/H in [1, 100]
LOG_TSTAR_MIN, LOG_TSTAR_MAX = 0.0, 4.0     # T_star in [1, 10000] MeV
```

Schreibe, wie schon in Aufgabe 17.2, eine Funktion `log_prior(theta)`, die 0.0 zurückgibt, falls alle drei Werte innerhalb ihres Bereichs liegen, und sonst `-np.inf`.

b) Schreibe `log_posterior(theta)`: prüfe zuerst `log_prior`, berechne dann (für θ innerhalb des Priors) mit `pt_spectrum_log10` aus Aufgabe 18.1 das Modell an den 14 Frequenzen `freqs` (**Achtung**: `pt_spectrum_log10` erwartet $\alpha, \beta/H, T_*$ selbst, nicht ihre Logarithmen!) und gib `log_prior(theta) + log_likelihood(model)` zurück.

Musterlösung 18.2

a) Ganz analog zu Aufgabe 17.2, nur mit drei statt zwei Parametern:

```
LOG_ALPHA_MIN, LOG_ALPHA_MAX = -2.0, 1.0
LOG_BOH_MIN, LOG_BOH_MAX = 0.0, 2.0
LOG_TSTAR_MIN, LOG_TSTAR_MAX = 0.0, 4.0

def log_prior(theta):
    log_alpha, log_boh, log_tstar = theta
    if (LOG_ALPHA_MIN <= log_alpha <= LOG_ALPHA_MAX) \
       and (LOG_BOH_MIN <= log_boh <= LOG_BOH_MAX) \
       and (LOG_TSTAR_MIN <= log_tstar <= LOG_TSTAR_MAX):
        return 0.0
    return -np.inf
```

b) Der einzige Unterschied zu Aufgabe 17.2 ist, dass wir die drei gesampelten Logarithmen erst in $\alpha, \beta/H, T_*$ zurückverwandeln müssen, bevor wir `pt_spectrum_log10` aufrufen:

```
def log_posterior(theta):
    lp = log_prior(theta)
    if not np.isfinite(lp):
        return -np.inf
    log_alpha, log_boh, log_tstar = theta
    model = pt_spectrum_log10(freqs, 10**log_alpha, 10**log_boh, 10**log_tstar)
    return lp + log_likelihood(model)
```

Beachte, wie wenig sich gegenüber Aufgabe 17.2 ändert: `log_prior` bekommt drei statt zwei Bedingungen, und `broken_powerlaw_log10` wird durch `pt_spectrum_log10` ersetzt; der Rest (insbesondere `log_likelihood` und `metropolis_hastings`) bleibt unverändert.

Aufgabe 18.3

a) Rufe euren Sampler aus Aufgabe 15.6 auf (*Code-Änderungen sind nur in eurem Modell/Posterior nötig, nicht in* `metropolis_hastings` *selbst!*

```
chain, logp, acc = metropolis_hastings(
    log_posterior, x0=[0.0, 0.3, 2.0], step_sizes=[0.15, 0.15, 0.15],
    n_steps=100000, seed=7)
samples = chain[10000:]   # burn-in
```

Gib die Akzeptanzrate aus. Falls sie deutlich außerhalb von 0.2–0.5 liegt: was könntet ihr an `step_sizes` ändern, um das zu korrigieren? (Ihr müsst das nicht ausprobieren, die obigen Werte funktionieren bereits gut, aber überlegt euch, in welche Richtung ihr die Schrittweiten anpassen würdet.)

b) Erstelle einen corner plot für $(\log_{10}\alpha, \log_{10}(\beta/H), \log_{10} T_*)$ und vergleiche mit Abbildung 18.5.

c) Bestimme wieder MAP-Werte (`np.argmax(logp)`) sowie Median und 16%/84%-Perzentile für jeden der drei Parameter.

d) Plotte die Mock-Daten zusammen mit dem besten Modell (wieder mit `f_plot`, das eine Größenordnung über das PTA-Band hinausgeht). Vergleiche mit Abbildung 18.6.

e) Welche Peak-Frequenz f_p (Gleichung (18.2)) ergibt sich aus den besten Werten von β/H und T_*? Vergleiche mit dem $f_{\rm pivot}$, das ihr in Aufgabe 17.2 gefunden habt.

Musterlösung 18.3

a) Mit den angegebenen Einstellungen ergibt sich eine Akzeptanzrate von acc $\approx$ 0.2–0.3 – im gewünschten Bereich 0.2–0.5. Wäre die Akzeptanzrate deutlich kleiner als 0.2, wären die `step_sizes` zu groß (zu viele Vorschläge landen in Bereichen mit sehr kleinem `log_posterior` und werden verworfen); wäre sie deutlich größer als 0.5, wären die `step_sizes` zu klein (die Kette bewegt sich nur sehr langsam durch den Parameterraum), vgl. die Diskussion in Aufgabe 15.6e).
b) Der corner plot (Abbildung 18.5) zeigt für alle drei Parameter $\log_{10}\alpha, \log_{10}(\beta/H), \log_{10} T_*$ deutlich von den Prior-Grenzen getrennte, annähernd glockenförmige marginalisierte Verteilungen; die Daten legen alle drei Parameter also nichttrivial fest. In den 2D-Panels erkennt man „schräge“, langgezogene Konturen statt runder Kleckse: eine *Entartung* (degeneracy), insbesondere zwischen $\log_{10}(\beta/H)$ und $\log_{10} T_*$.

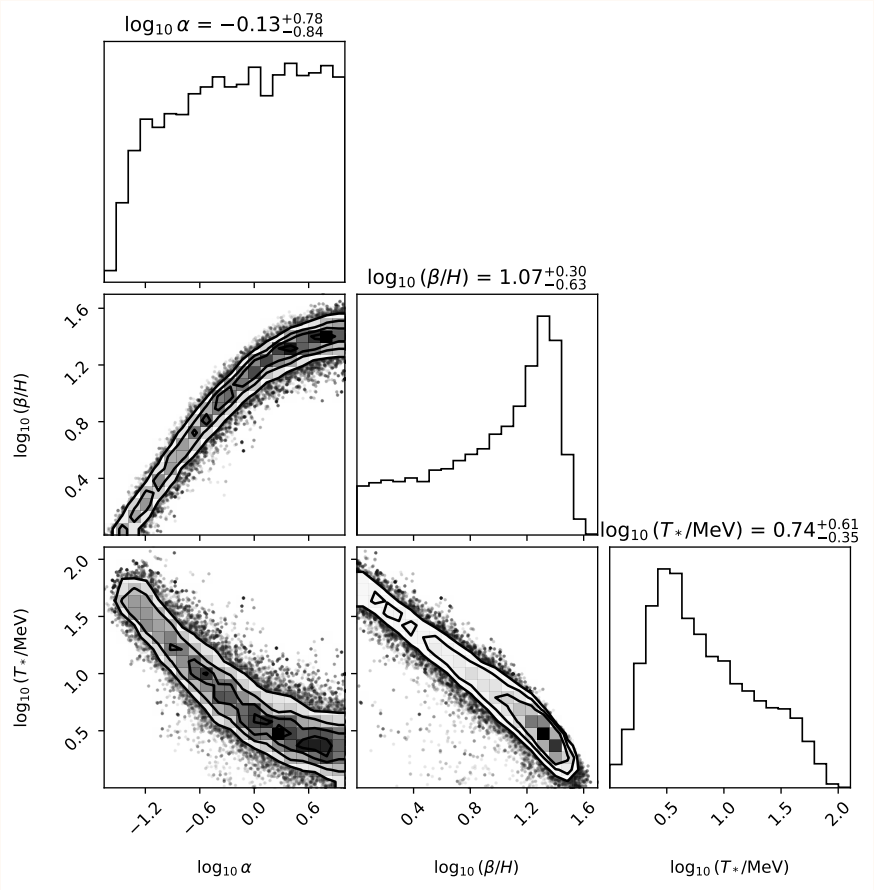


Abbildung 18.5: Corner plot für $\log_{10}\alpha$, $\log_{10}(\beta/H)$ und $\log_{10}T_*$.

c) MAP, Median und 16%/84%-Perzentile lassen sich wieder direkt mit `np.argmax(logp)` bzw. `np.percentile(samples[:,i],[16,50,84])` bestimmen; alle drei Parameter sind auf $O(1)$ Größenordnungen in $\log_{10}$ eingegrenzt.
d) Das beste Phasenübergangs-Spektrum folgt, genau wie schon das Toy model in Aufgabe 17.2, – der „Beule“ in den Mock-Daten bei niedrigen Frequenzen (Abbildung 18.6).

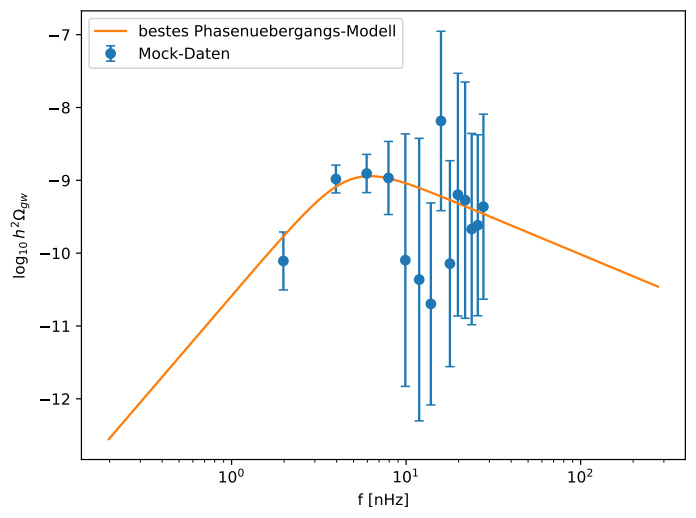


Abbildung 18.6: Mock-Daten und bestes Phasenübergangs-Spektrum.

e) Setzt man die besten Werte von β/H und T_* in Gleichung (18.2) ein, ergibt sich eine Peak-Frequenz f_p von einigen nHz, in guter Übereinstimmung mit dem f_{pivot} aus Aufgabe 17.2. Das ist kein Zufall: Beide Modelle müssen ja dieselbe „Beule“ in denselben Daten beschreiben, also muss ihre jeweilige Peak-Frequenz übereinstimmen; nur dass f_p beim Phasenübergangsmodell zusätzlich über (18.2) mit den physikalischen Größen β/H und T_* verknüpft ist.

Aufgabe 18.4

Schaut euch den corner plot (Abbildung 18.5) genau an.

a) Wie groß sind die Unsicherheiten auf $\alpha, \beta/H, T_*$, über wie viele Größenordnungen erstrecken sich die jeweiligen 16%/84%-Intervalle? Was sagt euch das darüber, wie gut die Mock-Daten den Phasenübergang festlegen?

b) *Hinweis zur Entartung aus Aufgabe 18.3b):* Könnt ihr anhand von Gleichung (18.2) verstehen, warum z.B. β/H und T_* miteinander korreliert sein könnten? Was passiert mit f_p, wenn β/H größer und T_* gleichzeitig kleiner wird?

c) War der Phasenübergang stark oder schwach (anhand von α)? Schnell oder langsam (anhand von β/H)?

d) Vergleicht die gefundene Temperatur T_* mit bekannten Energieskalen im frühen Universum – z.B. dem QCD-Phasenübergang bei ~ 100–$200\,\text{MeV}$ oder dem elektroschwachen Phasenübergang bei $\sim 100\,\text{GeV}$. Welches (hypothetische) Ereignis im frühen Universum könnte zu eurem T_* passen?

Musterlösung 18.4

a) Die 16%/84%-Intervalle erstrecken sich jeweils über $\mathcal{O}(0.5\text{–}1)$ Größenordnungen in $\log_{10}\alpha$, $\log_{10}(\beta/H)$ und $\log_{10} T_*$; die 14 Mock-Datenpunkte legen den Phasenübergang also *ungefähr*, aber nicht beliebig genau fest. Das ist typisch für PTA-Daten: Die Form des Spektrums (Höhe und Peak-Lage) ist relativ gut bestimmt, aber die Aufteilung dieser Information auf drei physikalische Parameter ist nicht eindeutig.
b) In Gleichung (18.2) gehen β/H und T_* als *Produkt* $T_* \cdot (\beta/H)$ in f_p ein. Wird β/H größer und T_* gleichzeitig im selben Verhältnis kleiner, bleibt das Produkt und damit f_p näherungsweise gleich. Solange auch die *Höhe* des Spektrums (über α und β/H in Gleichung (18.1)) ungefähr passt, erklären beide Parameterkombinationen die Daten also ähnlich gut – genau die „schräge“, langgezogene Kontur im $(\log_{10}(\beta/H), \log_{10} T_*)$-Panel von Abbildung 18.5.
c) Liegt $\alpha \lesssim \mathcal{O}(1)$, war der Phasenübergang vergleichsweise schwach (die freigesetzte Energie ist nicht viel größer als die Strahlungsenergiedichte); $\beta/H \gtrsim \mathcal{O}(10)$ entspricht einem vergleichsweise schnellen, abrupten Übergang (vgl. Aufgabe 18.1c)).
d) $T_* \sim \mathcal{O}(10\text{–}1000)\,\text{MeV}$ liegt in der Größenordnung des QCD-Phasenübergangs (~ 100–$200\,\text{MeV}$), bei dem Quarks und Gluonen zu Hadronen „einfrieren“ (vgl. Aufgabe 5.2 und Abbildung 1.2). Ein Phasenübergang in diesem Temperaturbereich wäre also kompatibel mit einem (im Standardmodell eigentlich glatten, aber durch neue Physik möglicherweise zu einem Phasenübergang erster Ordnung gemachten) QCD-artigen Übergang im frühen Universum – genau die Art von „neuer Physik“, nach der PTAs wie NANOGrav suchen.

19 SMBHB, Phasenübergang und reale Daten

Es gibt noch eine zweite, rein astrophysikalische Erklärung für einen Gravitationswellenhintergrund im PTA-Band: viele überlagerte, langsam verschmelzende supermassive Schwarze-Loch-Doppelsysteme (*SMBHBs*, Kapitel 11). Ihr Hintergrund wird üblicherweise als Potenzgesetz in der charakteristischen Strain $h_c(f)$ angegeben,

$$h_c(f) = A_{\mathrm{BHB}} \left(\frac{f}{f_{\mathrm{yr}}} \right)^{(3-\gamma)/2}, \qquad f_{\mathrm{yr}} = \frac{1}{1\,\mathrm{Jahr}}, \tag{19.1}$$

wobei $\gamma = 13/3 \approx 4.33$ der theoretisch erwartete Wert für eine Population kreisförmiger, rein durch Gravitationswellen-Abstrahlung getriebener Doppelsysteme ist. Mit der Umrechnung $h_c(f) = \frac{H_0}{h}\sqrt{\frac{3}{2}h^2\Omega_{\mathrm{gw}}(f)}/(\pi f)$ folgt daraus

$$h^2\Omega_{\mathrm{gw}}(f) = \frac{2\pi^2}{3}\left(\frac{h}{H_0}\right)^2 A_{\mathrm{BHB}}^2\, f^{5-\gamma}\, f_{\mathrm{yr}}^{\gamma-3}. \tag{19.2}$$

Aufgabe 19.1

Hinweis: Diese Aufgabe benötigt das Paket `ptarcade`. Installationsanleitung unter https://andrea-mitridate.github.io/PTArcade/getting_started/local_install/. Die Konstanten `H_0_Hz` (Hubble-Konstante in Hz) und `h` (dimensionsloser Hubble-Parameter) lassen sich dann mit `from ptarcade.models_utils import H_0_Hz, h as h_hubble` importieren.
In dieser Aufgabe vergleicht ihr, wie gut diese SMBHB-Erklärung, euer Phasenübergangs-Modell, und eine Kombination aus beiden zu den Mock-Daten passen.

a) Lege eine neue Datei `smbhb_comparison.py` an und implementiere Gleichung (19.2) mit $\theta = (\log_{10} A_{\mathrm{BHB}}, \gamma)$ als

```
F_YR = 1.0 / (365.25 * 24 * 3600.0)
H0 = H_0_Hz / h_hubble   # aus ptarcade.models_utils

def smbhb_spectrum_log10(f_hz, log10_A_BHB, gamma):
    return (np.log10(2*np.pi**2/3) - 2*np.log10(H0)
            + 2*log10_A_BHB
            + (5 - gamma) * np.log10(f_hz)
            + (gamma - 3) * np.log10(F_YR))
```

b) Anders als bei $\alpha, \beta/H, T_*$ haben wir für $(\log_{10} A_{\mathrm{BHB}}, \gamma)$ nicht nur einen Gleichverteilungs-Prior, sondern echte *Vorinformation*: Populationssynthese-Simulationen von SMBHB-Populationen [14] sagen voraus, in welchem Bereich diese beiden Parameter ungefähr liegen sollten, als 2D-Gauß-Verteilung. Implementiere diesen *astrophysikalisch informierten Prior*:

```
MU_LOG10_A, SIGMA_LOG10_A = -15.6, 0.3
MU_GAMMA, SIGMA_GAMMA = 4.7, 0.3
LOG10_A_MIN, LOG10_A_MAX = -18.0, -13.0
GAMMA_MIN, GAMMA_MAX = 0.0, 7.0

def astro_log_prior(log10_A_BHB, gamma):
    if not (LOG10_A_MIN <= log10_A_BHB <= LOG10_A_MAX):
        return -np.inf
    if not (GAMMA_MIN <= gamma <= GAMMA_MAX):
        return -np.inf
```

```
    return (-0.5 * ((log10_A_BHB - MU_LOG10_A) / SIGMA_LOG10_A)**2
            - 0.5 * ((gamma - MU_GAMMA) / SIGMA_GAMMA)**2)
```

c) **Modell 1: SMBHB allein.** Kombiniert `astro_log_prior` und `log_likelihood` zu `log_posterior_smbhb(theta)`, sampelt damit (`x0=[-15.6, 4.7]`, `step_sizes=[0.1, 0.1]`) und bestimmt den MAP-Punkt sowie `log_likelihood` (nicht `log_posterior`!) an diesem Punkt. Vergleicht außerdem Median und 16%/84%-Perzentile von $\log_{10} A_{\mathrm{BHB}}$ mit dem astro-Prior aus b): Liegt die Posterior in der Nähe des Priors, oder „zieht“ die PTA-Likelihood $\log_{10} A_{\mathrm{BHB}}$ in eine bestimmte Richtung?

d) **Modell 2: SMBHB + Phasenübergang.** Kombiniert beide Spektren *im Linearen* (nicht im Log!) und logarithmiert das Ergebnis erst danach:

```
def log_posterior_combined(theta):
    log10_A_BHB, gamma, log_alpha, log_boh, log_tstar = theta
    lp = (astro_log_prior(log10_A_BHB, gamma)
          + pt_log_prior((log_alpha, log_boh, log_tstar)))
    if not np.isfinite(lp):
        return -np.inf
    model_smbhb = smbhb_spectrum_log10(freqs, log10_A_BHB, gamma)
    model_pt = pt_spectrum_log10(freqs, 10**log_alpha, 10**log_boh,
    ↪  10**log_tstar)
    model = np.log10(10**model_smbhb + 10**model_pt)
    return lp + log_likelihood(model)
```

Sampelt mit 5 Parametern, $\theta = (\log_{10} A_{\mathrm{BHB}}, \gamma, \log_{10} \alpha, \log_{10}(\beta/H), \log_{10} T_*)$, `x0=[-15.6, 4.7, 0.0, 0.3, 2.0]`, `step_sizes=[0.1, 0.1, 0.15, 0.15, 0.15]`. Bestimmt wieder den MAP-Punkt und die zugehörige `log_likelihood` des kombinierten Modells.

e) **Modell 3: Phasenübergang allein.** Das ist genau euer `log_posterior` aus Aufgabe 18.2/18.3; bestimmt auch hier MAP-Punkt und `log_likelihood`.

f) Vergleicht die drei `log_likelihood`-Werte (Modelle 1–3) in einer kleinen Tabelle. Plottet außerdem die drei besten Spektren zusammen mit den Mock-Daten (Abbildung 19.1) und erstellt einen corner plot für Modell 2 (Abbildung 19.2).

g) *Vergleich mit der Literatur:* In [15] wird (Abschnitt 7, Abbildung 11) genau dieser Vergleich (SMBHB allein, SMBHB+Phasenübergang, Phasenübergang allein) mit echten NANOGrav-Daten durchgeführt. Vergleicht eure qualitativen Ergebnisse (welches Modell am besten/schlechtesten passt, wie groß die benötigte SMBHB-Amplitude ist) mit den dortigen Ergebnissen. Stimmt das Bild überein?

Musterlösung 19.1

a)/b) Die Implementierung folgt direkt aus Gleichung (19.2) und der 2D-Gauß-Form des astro-Priors (Code wie oben). *Achtung:* Die Werte für `MU_LOG10_A`, `SIGMA_LOG10_A`, `MU_GAMMA`, `SIGMA_GAMMA` sind eine Näherung an Gleichung (A1) aus arXiv:2306.16219.
c)–f) Vergleicht man die drei `log_likelihood`-Werte und die Spektren (Abbildung 19.1), ergibt sich ein klares Bild:

- **SMBHB allein** passt deutlich schlechter zu den Daten als die anderen beiden Modelle – das reine Potenzgesetz (19.2) kann die „Beule“ im Spektrum (Aufgabe 17.1c)) nicht reproduzieren.

- **SMBHB+Phasenübergang** und **Phasenübergang allein** liegen bei (fast) demselben, deutlich besseren `log_likelihood`-Wert; innerhalb der Genauigkeit unseres MCMC-Samplers sind sie kaum zu unterscheiden. Das Phasenübergangs-Spektrum allein erklärt die Daten also schon (nahezu) optimal; der SMBHB-Anteil im Kombi-Modell trägt kaum mehr etwas bei.
- Im Kombi-Modell (Abbildung 19.2) liegen $\log_{10} A_{\mathrm{BHB}}$ und γ sehr nahe an ihrem astro-Prior aus b); die Daten „brauchen“ im Kombi-Modell kein zusätzliches SMBHB-Signal.
- Im SMBHB-*allein*-Modell (c) dagegen liegt die Posterior von $\log_{10} A_{\mathrm{BHB}}$ deutlich *über* dem astro-Prior-Mittelwert; es wird also eine vergleichsweise *hohe* Amplitude A_{BHB} gebraucht, wenn SMBHBs den gesamten Hintergrund erklären sollen, und diese hohe Amplitude steht in Spannung zum astrophysikalischen Prior.

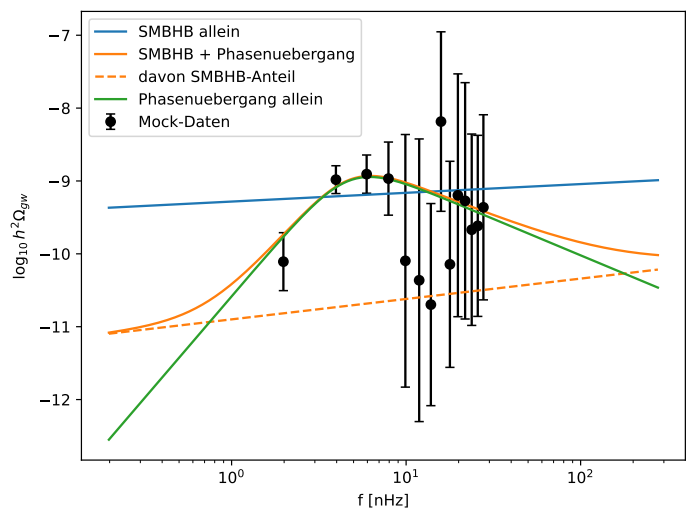


Abbildung 19.1: Mock-Daten und die besten Spektren für SMBHB allein, SMBHB+Phasenübergang und Phasenübergang allein.

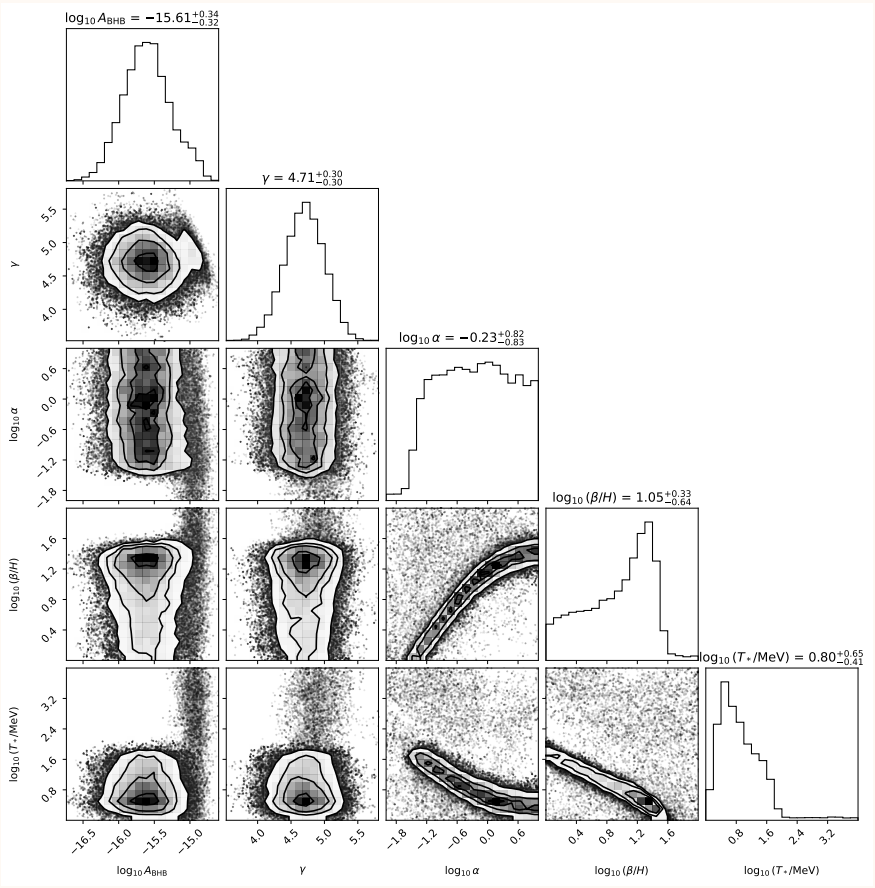


Abbildung 19.2: Corner plot für das Kombi-Modell SMBHB+Phasenübergang ($\log_{10} A_{\rm BHB}$, γ, $\log_{10}\alpha$, $\log_{10}(\beta/H)$, $\log_{10} T_*$).

g) In [15] wird mit echten NANOGrav-Daten qualitativ dasselbe Bild gefunden: Ein reines SMBHB-Potenzgesetz erklärt das beobachtete Spektrum schlechter als ein Phasenübergangsmodell, und die für SMBHB allein benötigte Amplitude liegt über der astrophysikalisch erwarteten. Unsere Mock-Daten – die ja absichtlich aus einem Phasenübergangsmodell konstruiert wurden, reproduzieren also genau das Verhalten, das man in echten Daten als „Hinweis auf neue Physik jenseits reiner SMBHBs“ interpretieren würde.

Aufgabe 19.2

[Für die Schnellen] *Ist unsere Mock-Likelihood realistisch?* Unsere Mock-Likelihood besteht aus 14 unabhängigen Gauß-Verteilungen, eine pro Frequenz-Bin (Aufgabe 15.2). Die *echte* Likelihood, die PTArcade (über das Paket `ceffyl`) für die NANOGrav-Daten benutzt, ist dagegen eine (nicht notwendigerweise Gauß-förmige, teils zweigipflige) Wahrscheinlichkeitsdichte, die direkt aus den NANOGrav-Chains bestimmt wurde. In dieser Zusatzaufgabe vergleicht ihr beide.

a) Installiere die Pakete `ceffyl` und `ptarcade` (`conda install -c conda-forge ptarcade`, `pip install ceffyl`) und lade den NANOGrav-15yr-„free spectrum“-Datensatz (Zenodo 10.5281/zenodo.10344086), aus dem auch eure Mock-Daten abgeleitet wurden:

```
from ceffyl import Ceffyl
```

```
ceffyl_pta = Ceffyl.ceffyl("ceffyl_data/30f_fs{hd}_100fDMGP_ceffyl")
```

`ceffyl_pta.density` enthält die (log-)Wahrscheinlichkeitsdichten auf einem Gitter `ceffyl_pta.rho_grid` der Variable $\log_{10}\rho$; das ist die Größe, mit der `ceffyl` intern rechnet (nicht direkt $\log_{10} h^2\Omega_{\mathrm{gw}}$).

b) Die Umrechnung zwischen unserem $\log_{10}(h^2\Omega_{\mathrm{gw}})$ und `ceffyls` $\log_{10}\rho$ folgt der Funktion `omega2cross` aus `ptarcade.models_utils`:

```
from ptarcade.models_utils import H_0_Hz, h as h_hubble

def h2omega_to_log10rho(f_hz, h2omega):
    hcf = H_0_Hz / h_hubble * np.sqrt(3.0 * h2omega / 2.0) / (np.pi * f_hz)
    S = hcf**2 / (12.0 * np.pi**2 * f_hz**3) / T_obs
    return 0.5 * np.log10(S)
```

Damit könnt ihr euer Modell `pt_spectrum_log10(freqs, ...)` in `log10rho`-Werte umrechnen, die direkt mit `ceffyl_pta.rho_grid` vergleichbar sind.

c) Schreibe eine Funktion `ceffyl_log_likelihood(ceffyl_pta, model_log10_h2omega)`, die (i) euer Modell mit `h2omega_to_log10rho` in `log10rho`-Werte umrechnet, (ii) für jeden der 14 Frequenz-Bins den nächstgelegenen Index im Gitter `ceffyl_pta.rho_grid` findet (`np.searchsorted` auf `ceffyl_pta.binedges`), (iii) dort die gespeicherte log-Dichte `ceffyl_pta.density[0, k, idx]` auswertet, und die Summe dieser 14 Werte (plus `np.log(ceffyl_pta.db)` pro Bin, wegen der Diskretisierung des Gitters) zurückgibt. *Hinweis: gib `-np.inf` zurück, falls ein `log10rho`-Wert außerhalb von `ceffyl_pta.binedges` liegt.*

d) **Test 1:** Werte sowohl `log_likelihood` (eure Gauß-Näherung aus Aufgabe 15.2) als auch `ceffyl_log_likelihood` (die echte NG15-Dichte) für ein paar zufällige $(\alpha, \beta/H, T_*)$-Kombinationen aus und vergleiche die Werte. Ist die Differenz konstant, oder hängt sie von den Parametern ab? Was sagt euch das über die Qualität eurer Gauß-Näherung?

e) **Test 2:** Führt mit `ceffyl_log_likelihood` statt `log_likelihood` eine zweite MCMC-Kette durch (gleicher Sampler, gleiches `log_prior`!) und erstellt einen corner plot. Vergleicht ihn mit eurem corner plot aus Aufgabe 18.3b) (Abbildung 19.3). Plottet außerdem beide besten Modelle zusammen mit den Mock-Daten (Abbildung 19.4).

Musterlösung 19.2

a)/b)/c) Die Implementierung folgt Schritt für Schritt aus dem Code oben; der entscheidende Punkt ist, dass `log_prior`, `metropolis_hastings` und das Modell `pt_spectrum_log10` *komplett unverändert* bleiben; nur `log_likelihood` wird durch `ceffyl_log_likelihood` ersetzt.
d) Die Differenz zwischen `log_likelihood` und `ceffyl_log_likelihood` ist *nicht* konstant, sondern hängt vom Parametersatz ab; unsere Gauß-Näherung ist also keine perfekte Beschreibung der echten NANOGrav-Likelihood, sondern approximiert sie nur in der Nähe ihres Maximums gut.
e) Der corner plot mit `ceffyl_log_likelihood` (Abbildung 19.3) sieht der Gauß-Näherung aus Abbildung 18.5 insgesamt sehr ähnlich; MAP-Werte und grobe Form der Konturen stimmen gut überein. Die größten Abweichungen treten in den Schwänzen der Verteilungen auf,

dort, wo die echte NG15-Likelihood (teils zweigipflig, nicht Gauß-förmig) am stärksten von einer einfachen Gauß-Verteilung abweicht. Beide besten Spektren liegen sehr nahe beieinander und beschreiben die „Beule“ in den Mock-Daten praktisch gleich gut (Abbildung 19.4).

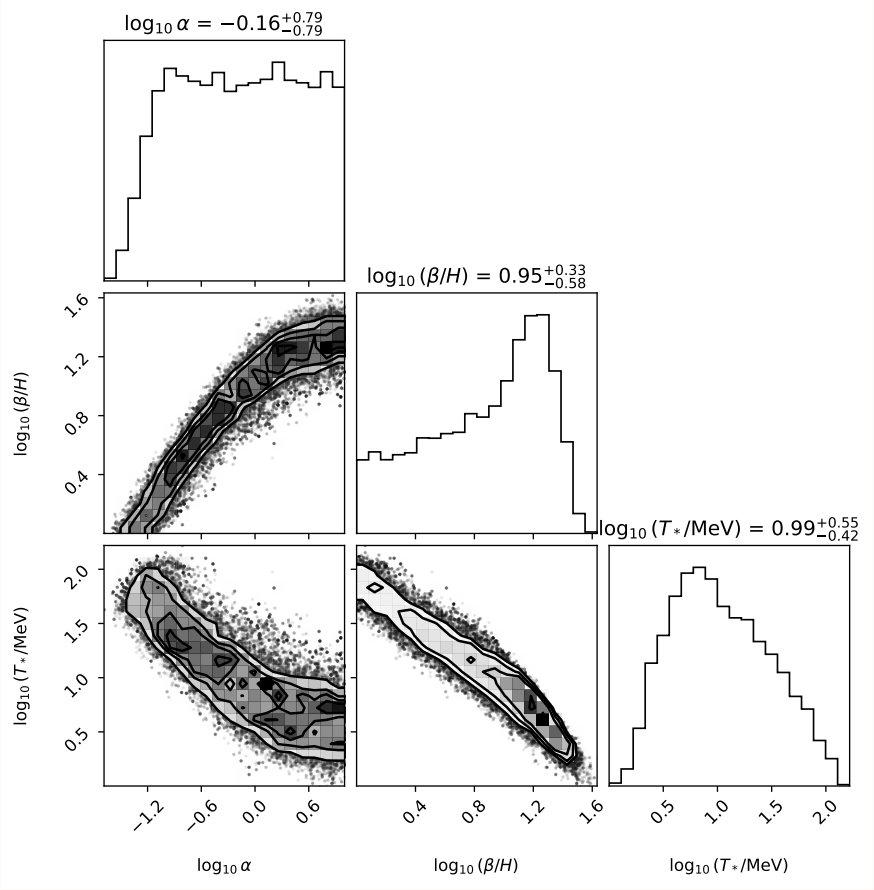


Abbildung 19.3: Corner plot mit der echten NG15-`ceffyl`-Likelihood, zum Vergleich mit Abbildung 18.5.

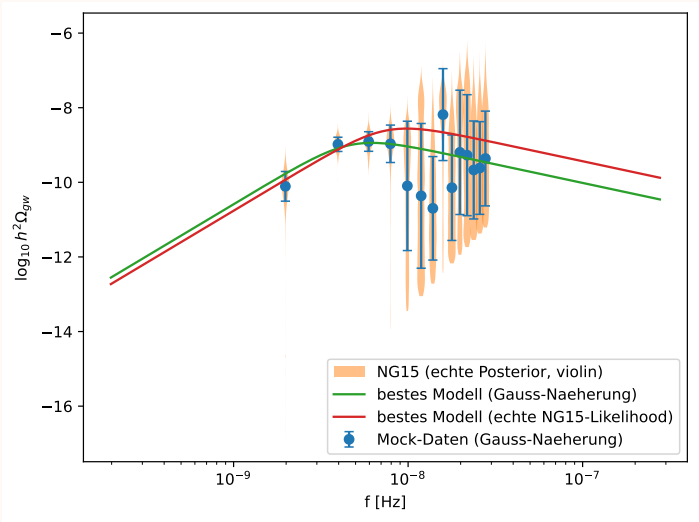


Abbildung 19.4: Beide besten Phasenübergangs-Spektren zusammen mit den Mock-Daten und den „violin plots“ der echten NG15-Posteriors.

Damit haben wir den vollen Kreis geschlossen: vom selbstgeschriebenen Sampler über ein unmotiviertes Toy model und das physikalische Phasenübergangsspektrum bis hin zum Vergleich mit der echten NANOGrav-Likelihood, mit ein und demselben `metropolis_hastings`-Algorithmus aus Aufgabe 15.6.

20 Globale Fits und moderne Sampling-Methoden

20.1 Was ist ein globaler Fit?

In Kapitel 15 haben wir den Metropolis-Hastings-Algorithmus auf ein einfaches Toy-Modell mit wenigen Parametern angewandt. In der Forschungspraxis stehen wir oft vor komplizierteren Situationen: Wir haben ein physikalisches Modell mit vielen Parametern $\theta = (\theta_1, \ldots, \theta_n)$, und wir wollen *gleichzeitig* alle Parameter an alle verfügbaren Datensätze anpassen; das nennt man einen *globalen Fit*.

Das Ziel ist es, die Posterior-Verteilung

$$p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta) \tag{20.1}$$

über den gesamten Parameterraum zu kartieren, um sowohl die besten Parameterwerte als auch ihre Unsicherheiten und Korrelationen zu bestimmen. Bei $n \gtrsim 5$ Parametern wird das Auswerten auf einem regulären Gitter unpraktikabel (N^n Punkte!), weshalb MCMC-Methoden unverzichtbar sind.

Genau diesen Fall haben wir in Kapitel 19 durchgeführt: Drei konkurrierende Modelle (SMBHB allein, Phasenübergang allein, SMBHB + Phasenübergang) wurden an dieselben Mock-PTA-Daten angepasst und ihre Likelihoods miteinander verglichen – ein typischer globaler Fit mit bis zu fünf freien Parametern.

20.2 Phasenübergänge im frühen Universum: ein Forschungsbeispiel

Ein konkretes Beispiel für einen globalen Fit aus der aktuellen Forschung ist die Frage, ob das vom NANOGrav-Pulsar-Timing-Array gemessene Gravitationswellensignal (vgl. Kapitel 17) von einem kosmologischen Phasenübergang stammen könnte.

In Referenz [10] wurde ein *Phasenübergang in einem dunklen Sektor* bei MeV-Temperaturen als Erklärung für das NANOGrav-Signal untersucht. Das Modell hat drei freie Parameter: die latente Wärme α, das Verhältnis β/H (das die Dauer des Übergangs bestimmt) und die Nukleationstemperatur T_*. Der globale Fit maximiert die kombinierte Likelihood aus dem PTA-Signal und kosmologischen Nebenbedingungen (Big-Bang-Nukleosynthese, CMB). Abbildung 18.6 zeigt das Ergebnis eines solchen Fits an simulierte PTA-Daten: Das Gravitationswellenspektrum des besten Phasenübergangs-Modells (orange) gibt die Mock-Daten (blau) gut wieder.

Das Ergebnis: Ein vollständig isolierter Phasenübergang in einem dunklen Sektor ist statistisch stark disfavorisiert gegenüber der astrophysikalischen Erklärung durch supermassive Schwarze-Loch-Binärsysteme; ein Übergang, der vor der Neutrinoentkopplung zerfällt, bleibt hingegen vereinbar.

In Referenz [15] wird gefragt, ob ein solcher Phasenübergang in einem dunklen Sektor *gleichzeitig* ein Sub-GeV-Dunkle-Materie-Kandidat sein kann. Das Modell, ein klassisch-konformer dunkler Sektor mit einem dunklen Higgs-Feld und einem dunklen Photon, verknüpft zwei Beobachtungskategorien: das PTA-Signal und zukünftige Beam-Dump-Experimente. Der globale Fit kartiert einen hochdimensionalen Parameterraum und zeigt, welche Parameterbereiche beide Phänomene erklären und experimentell zugänglich sind.

Diese Beispiele illustrieren das typische Vorgehen: Man schreibt eine (Profil-)Likelihood, die alle Datensätze kombiniert, und lässt dann einen MCMC-Sampler den Parameterraum erkunden. Die resultierenden Corner-Plots zeigen marginalisierte 1D-Posteriors (auf der Diagonalen) und 2D-Konfidenzregionen für alle Parameterpaare. Abbildung 18.5 zeigt ein Beispiel für drei Phasenübergangs-Parameter aus demselben Fit.

20.3 Ausblick: Jenseits von Metropolis-Hastings

Der Metropolis-Hastings-Algorithmus ist ein eleganter und weit verbreiteter Sampler, hat aber Schwächen, die in der Forschungspraxis wichtig werden:

- Bei **multimodalen Posteriors** (mehrere getrennte Maxima) kann die Kette in einem Modus stecken bleiben und den anderen nie finden.
- In sehr hochdimensionalen Räumen ($n \gtrsim 20$) ist die Akzeptanzrate klein und das Mixen langsam.
- Die **Bayessche Evidenz** $p(D) = \int p(D \mid \theta)\, p(\theta)\, \mathrm{d}\theta$ – die für den Bayes-Faktor beim Modellvergleich nötig ist, kann Metropolis-Hastings nicht direkt berechnen.

Für diese Situationen gibt es spezialisierte Algorithmen:

20.3.1 Nested Sampling

Nested Sampling (Skilling 2004) ist ein Algorithmus, der die Posterior-Evidenz *direkt* berechnet, indem er den Parameterraum von außen nach innen „aufrollt“. Die Grundidee:

1. Starte mit K *Live-Punkten*, die gleichmäßig aus dem Prior gezogen werden.
2. Identifiziere den Punkt mit der kleinsten Likelihood $L_{\min}$ und ersetze ihn durch einen neuen, gleichverteilten Punkt mit $L > L_{\min}$.
3. Speichere den entfernten Punkt mit seinem Gewicht $w_i \propto \Delta X_i \cdot L_i$, wobei ΔX_i das Prior-Volumen ist, das er repräsentiert.
4. Wiederhole, bis die verbleibende Evidenz vernachlässigbar ist.

Das Integral $p(D) = \int L(\theta)\, p(\theta)\, \mathrm{d}\theta$ ergibt sich aus der Summe der Gewichte: $p(D) \approx \sum_i w_i$. Als Nebenprodukt liefert Nested Sampling gewichtete Posterior-Samples. Populäre Implementierungen sind `MultiNest` [16], `dynesty` und `nautilus`.

Nested Sampling ist besonders stark bei multimodalen Posteriors und präziser Evidenzberechnung; dafür ist es in hochdimensionalen, glatten Parameterräumen oft langsamer als gut getuntes MCMC.

Aufgabe 20.1

In dieser Aufgabe verwendet ihr `dynesty`, um die Bayessche Evidenz für zwei Modelle zu berechnen und den Bayes-Faktor $\Delta \ln Z = \ln Z_{\mathrm{PT}} - \ln Z_{\mathrm{SMBHB}}$ zu bestimmen. Installiert `dynesty` via `pip install dynesty`.
Hintergrund: Metropolis-Hastings liefert Posterior-Samples, aber *keine* Evidenz $Z = p(D)$. Nested Sampling berechnet Z direkt. Der Log-Bayes-Faktor $\Delta \ln Z > 0$ bedeutet, das PT-Modell wird von den Daten bevorzugt.

1. Definiert die Log-Likelihood für das PT-Modell (aus Aufgabe 15.2) und den flachen Log-Prior-Würfel für $\theta = (\log_{10}\alpha, \log_{10}(\beta/H), \log_{10} T_*)$. Übersetzt diesen in eine `dynesty`-`ptform`-Funktion (*prior transform*), die einen Einheitswürfel $[0,1]^3$ auf den Prior-Bereich abbildet:

```
import dynesty

LOG_ALPHA_MIN, LOG_ALPHA_MAX = -1.0, 1.0
LOG_BOH_MIN,   LOG_BOH_MAX   =  0.0, 3.0
LOG_TSTAR_MIN, LOG_TSTAR_MAX =  1.0, 4.0

def ptform_pt(u):
    log_alpha = LOG_ALPHA_MIN + u[0] * (LOG_ALPHA_MAX - LOG_ALPHA_MIN)
    log_boh   = LOG_BOH_MIN   + u[1] * (LOG_BOH_MAX   - LOG_BOH_MIN)
    log_tstar = LOG_TSTAR_MIN + u[2] * (LOG_TSTAR_MAX  - LOG_TSTAR_MIN)
```

```
        return [log_alpha, log_boh, log_tstar]

    def loglike_pt(theta):
        log_alpha, log_boh, log_tstar = theta
        model = pt_spectrum_log10(freqs, 10**log_alpha, 10**log_boh,
        ↪  10**log_tstar)
        return log_likelihood(model)
```

2. Erstellt analog `ptform_smbhb` und `loglike_smbhb` für das SMBHB-Modell (Gleichung (19.2)), mit flachem Prior in $(\log_{10} A_{\mathrm{BHB}}, \gamma)$ und `bounds = [[-18, -13], [0, 7]]`.

3. Führt Nested Sampling für beide Modelle durch:

```
sampler_pt = dynesty.NestedSampler(loglike_pt, ptform_pt, ndim=3,
                                   nlive=300, sample='rwalk')
sampler_pt.run_nested(dlogz=0.1, print_progress=True)
res_pt = sampler_pt.results

sampler_smbhb = dynesty.NestedSampler(loglike_smbhb, ptform_smbhb, ndim=2,
                                      nlive=300, sample='rwalk')
sampler_smbhb.run_nested(dlogz=0.1, print_progress=True)
res_smbhb = sampler_smbhb.results
```

   Die Evidenz steht in `res.logz[-1]`, die Unsicherheit in `res.logzerr[-1]`.

4. Berechnet $\Delta \ln Z =$ `res_pt.logz[-1]` $-$ `res_smbhb.logz[-1]` und die kombinierte Unsicherheit. Wie stark wird das PT-Modell gegenüber SMBHB bevorzugt? Vergleicht mit euren qualitativen Ergebnissen aus Aufgabe 19.1.

Musterlösung 20.1

```
import numpy as np
import dynesty

# -- Prior-Grenzen (PT-Modell) --
LOG_ALPHA_MIN, LOG_ALPHA_MAX = -1.0, 1.0
LOG_BOH_MIN,   LOG_BOH_MAX   =  0.0, 3.0
LOG_TSTAR_MIN, LOG_TSTAR_MAX =  1.0, 4.0

def ptform_pt(u):
    return [LOG_ALPHA_MIN + u[0]*(LOG_ALPHA_MAX-LOG_ALPHA_MIN),
            LOG_BOH_MIN   + u[1]*(LOG_BOH_MAX-LOG_BOH_MIN),
            LOG_TSTAR_MIN + u[2]*(LOG_TSTAR_MAX-LOG_TSTAR_MIN)]

def loglike_pt(theta):
    model = pt_spectrum_log10(freqs, 10**theta[0], 10**theta[1], 10**theta[2])
    return log_likelihood(model)

# -- Prior-Grenzen (SMBHB-Modell) --
def ptform_smbhb(u):
    return [-18.0 + u[0]*5.0,    # log10_A in [-18, -13]
             0.0  + u[1]*7.0]    # gamma in [0, 7]

def loglike_smbhb(theta):
```

```
    model = smbhb_spectrum_log10(freqs, theta[0], theta[1])
    return log_likelihood(model)

# -- Nested Sampling --
sampler_pt = dynesty.NestedSampler(loglike_pt, ptform_pt, ndim=3,
                                   nlive=300, sample='rwalk')
sampler_pt.run_nested(dlogz=0.1, print_progress=False)
res_pt = sampler_pt.results

sampler_smbhb = dynesty.NestedSampler(loglike_smbhb, ptform_smbhb, ndim=2,
                                      nlive=300, sample='rwalk')
sampler_smbhb.run_nested(dlogz=0.1, print_progress=False)
res_smbhb = sampler_smbhb.results

lnZ_pt    = res_pt.logz[-1];    err_pt    = res_pt.logzerr[-1]
lnZ_smbhb = res_smbhb.logz[-1]; err_smbhb = res_smbhb.logzerr[-1]
dlnZ = lnZ_pt - lnZ_smbhb
err  = np.sqrt(err_pt**2 + err_smbhb**2)
print(f"ln Z(PT)    = {lnZ_pt:.1f} +/- {err_pt:.1f}")
print(f"ln Z(SMBHB) = {lnZ_smbhb:.1f} +/- {err_smbhb:.1f}")
print(f"Delta ln Z  = {dlnZ:.1f} +/- {err:.1f}")
```

Ein typisches Ergebnis (abhängig von den Mock-Daten) ist $\Delta \ln Z \sim 5$–15, was nach der Jeffreys-Skala als „sehr stark“ bis „entscheidend“ gilt: Das PT-Modell wird von den (absichtlich aus einem PT-Spektrum generierten) Mock-Daten gegenüber dem reinen SMBHB-Modell klar bevorzugt. Dies ist konsistent mit dem qualitativen Ergebnis aus Aufgabe 19.1, wo das SMBHB-Modell die „Beule“ im Spektrum nicht reproduzieren konnte.

20.3.2 Parallel Tempering MCMC (PTMCMC)

Parallel Tempering (auch: Replica-Exchange MCMC) löst das Multimodalitätsproblem von Metropolis-Hastings, indem es mehrere Ketten bei verschiedenen *Temperaturen* $T_1 < T_2 < \cdots < T_k$ parallel laufen lässt. Bei hoher Temperatur ist die Posterior $p(\theta \mid D)^{1/T}$ deutlich flacher; die Kette kann leicht zwischen Moden wechseln. Periodisch werden benachbarte Ketten mit der *Swap-Rate*

$$p_{\text{swap}} = \min\left(1, \frac{p(\theta_i \mid D)^{1/T_j} \cdot p(\theta_j \mid D)^{1/T_i}}{p(\theta_i \mid D)^{1/T_i} \cdot p(\theta_j \mid D)^{1/T_j}}\right) \tag{20.2}$$

ausgetauscht. Die Kette bei $T_1 = 1$ liefert die gewünschten Posterior-Samples, während die heißen Ketten als „Fährten“ durch den Parameterraum dienen.

PTMCMC ist besonders in der PTA-Gemeinschaft verbreitet: Das Paket `PTMCMCSampler` (Ellis & van Haasteren) ist speziell für hochdimensionale PTA-Analysen ausgelegt und integriert verschiedene Proposal-Methoden (Differential Evolution, Adaptive Gaussian, Prior-Draws). Es wird u.a. von NANOGrav und EPTA für die Analyse der 15-jährigen Datensätze verwendet.

20.3.3 Wahl des Samplers in der Praxis

- **Wenige Parameter, glatte Posterior:** Metropolis-Hastings mit adaptiver Schrittweite oder Hamiltonian Monte Carlo (`NumPyro`, `Stan`).
- **Multimodale Posterior oder Evidenzberechnung nötig:** Nested Sampling (`dynesty`, `nautilus`).

- **Hochdimensional, PTA-Analyse:** PTMCMC (`PTMCMCSampler`).

Die Wahl des richtigen Samplers ist oft genauso wichtig wie das physikalische Modell; ein schlecht gewählter Sampler kann zu verzerrten oder unvollständigen Posteriors führen.

Ausblick

Wir haben eine weite Reise hinter uns. Von der Rotverschiebung und dem Hubble-Gesetz über die Friedmann-Gleichungen und das Alter des Universums, vorbei an Dunkler Materie und kosmologischen Phasenübergängen, bis hin zu Gravitationswellen und Pulsar-Timing-Arrays – und schließlich zur Bayes'schen Statistik, zu MCMC-Sampling-Methoden und globalen Fits: All diese Fäden hängen zusammen.

Was wir gelernt haben. Das Universum ist ein natürliches Labor für Physik bei Energieskalen, die mit keinem Teilchenbeschleuniger erreichbar sind. Die Signale, die wir empfangen, sind kosmisch rotverschoben, verdünnt und überlagert – und brauchen deswegen präzise statistische Methoden, um interpretiert zu werden. Der Metropolis-Hastings-Algorithmus, den ihr von Grund auf selbst implementiert habt, ist dabei kein akademisches Spielzeug: Er ist in dieser oder ähnlicher Form das Werkzeug, mit dem echte Forschungsgruppen echte Messungen auswerten. Dasselbe gilt für Nested Sampling und die Bayes'sche Evidenz, die ihr in Aufgabe 20.1 berechnet habt.

Offene Fragen. Das NANOGrav-Signal, das wir in diesem Kurs modelliert haben, ist bis heute nicht eindeutig erklärt. Sind es wirklich supermassive Schwarze-Loch-Doppelsysteme? Ein kosmologischer Phasenübergang? Beides zusammen? Oder etwas ganz anderes? Ähnliche Fragen stellen sich für Dunkle Materie: Welcher Mechanismus ist für ihre Entstehung verantwortlich? Gibt es ein Signal in zukünftigen Experimenten?

Nächste Experimente. In den nächsten Jahren werden mehrere neue Beobachtungsplattformen online gehen, die direkt an die Themen dieses Kurses anknüpfen:

- **LISA** (*Laser Interferometer Space Antenna*, ESA, ∼2035): Ein weltraumgestütztes Gravitationswellenobservatorium im Millihertz-Band – sensitiv für Gravitationswellen aus Phasenübergängen bei $T_* \sim$ TeV, die im PTA-Band unsichtbar sind.
- **Einstein Telescope** (ET, Europa, ∼2035): Ein unterirdischer Gravitationswellendetektor der dritten Generation, zehnmal sensitiver als LIGO; erschließt das Decihertz-Band für Gravitationswellen aus dem frühen Universum.
- **Square Kilometre Array** (SKA, Australien/Südafrika, ab 2027): Ein Radioteleskop mit rund einer Million Antennen, das als Pulsar-Timing-Array die Sensitivität gegenüber NANOGrav um Größenordnungen verbessern wird.
- **CMB-S4** (*Cosmic Microwave Background Stage 4*, ∼2030): Ein Netzwerk von Teleskopen, das primordiale Gravitationswellen aus der Inflation und das Szenario des kosmologischen Phasenübergangs über die B-Mode-Polarisation des CMB testen wird.

Die Physik, die ihr in diesem Kurs gelernt habt, ist nicht nur Hintergrundwissen: Sie ist der Schlüssel, um die Daten dieser Experimente zu verstehen und zu interpretieren. Das Universum hat noch viele Geheimnisse – und ihr habt jetzt das Handwerkszeug, ihnen auf die Spur zu kommen.

Glossar

Dieses Glossar erläutert Fachbegriffe, die im Skript verwendet werden und über den üblichen Schulstoff hinausgehen. Die Einträge sind alphabetisch geordnet.

Allgemeine Relativitätstheorie (ART) Einsteins Theorie der Gravitation (1915), in der Massen die vierdimensionale Raumzeit krümmen und diese Krümmung die Bewegung von Körpern und Licht bestimmt. Grundlage für Kosmologie, Schwarze Löcher und Gravitationswellen.

Annihilation Kollision eines Teilchens mit seinem Antiteilchen, bei der beide verschwinden und die gesamte Masse-Energie in andere Teilchen (oft Photonen) umgewandelt wird.

Axion Hypothetisches, extrem leichtes skalares Teilchen, ursprünglich postuliert zur Lösung des starken CP-Problems der QCD. Gilt als Kandidat für Dunkle Materie.

Baryon Aus drei Quarks zusammengesetztes Teilchen; wichtigste Beispiele sind Proton und Neutron. *Baryonische Materie* bezeichnet die gewöhnliche, leuchtende Materie: sie macht nur $\approx 5\,\%$ der Energiedichte des Universums aus.

Boltzmann-Unterdrückung Exponentiell kleiner Teilchenanteil $n \propto e^{-m/T}$ für Teilchen mit Masse $m \gg T$ (Temperatur). Erklärt, warum schwere Teilchen im Plasma verschwinden, sobald die Temperatur unter ihre Masse fällt.

Boson Teilchen mit ganzzahligem Spin (0, 1, 2, ...); folgt der Bose-Einstein-Statistik. Alle Kraftträgerteilchen (Photon, Gluon, W/Z-Boson, Higgs) sind Bosonen; im Gegensatz zu Fermionen können beliebig viele Bosonen denselben Quantenzustand besetzen.

Bubble Nucleation (Blasennukleation) Entstehung und Wachstum von Blasen der neuen, energetisch günstigeren Phase in einem Erstordnungs-Phasenübergang, analog zur Bildung von Gasblasen beim Sieden von Wasser. Kollisionen dieser Blasen erzeugen Gravitationswellen.

CMB (Kosmischer Mikrowellenhintergrund) Thermische Strahlung ($T \approx 2.7\,\mathrm{K}$), die frei wurde, als sich das Universum $\sim$380 000 Jahre nach dem Urknall auf $\sim$3000 K abkühlte und Elektronen und Protonen zu neutralem Wasserstoff rekombinierten. Ältestes direkt beobachtbares „Bild“ des Universums.

Detailed Balance Gleichgewichtsbedingung für Markov-Ketten: Die Übergangsrate zwischen je zwei Zuständen $A \to B$ muss gleich der Rate $B \to A$ sein. Garantiert, dass der MCMC-Sampler die korrekte Zielverteilung erzeugt.

Dunkle Energie Räumlich gleichmäßig verteilte Energiekomponente ($\Omega_\Lambda \approx 0.68$) mit negativem Druck, die die beschleunigte Expansion des Universums antreibt. Ihre mikroskopische Natur ist ungeklärt; phänomenologisch entspricht sie Einsteins kosmologischer Konstante Λ.

Dunkle Materie (DM) Nicht leuchtende, nicht baryonische Materiekomponente ($\Omega_{\mathrm{DM}} \approx 0.27$), die nur gravitativ nachgewiesen wurde. Notwendig zur Erklärung von Rotationskurven, Galaxienhaufen und der kosmischen Strukturbildung.

e-Foldings Maß für den Expansionsfaktor während der Inflation: $N = \ln(a_{\mathrm{Ende}}/a_{\mathrm{Start}})$. Das beobachtbare Universum verlangt $N \gtrsim 60$.

Effektive Freiheitsgrade (g_{eff}) Gewichtete Anzahl relativistischer Teilchenarten im frühen Universum (Bosonen und Fermionen mit Vorfaktor 7/8 für Fermionen). Bestimmt die Energie- und Entropiedichte des heißen Plasmas.

Eichboson Spin-1-Kraftträgerteilchen in einer Eichtheorie: Photon (Elektromagnetismus), $\mathrm{W}^{\pm}/\mathrm{Z}$ (schwache Kraft), Gluon (starke Kraft).

Eichtheorie Quantenfeldtheorie, in der die Form der Wechselwirkungen durch lokale Symmetrietransformationen (Eichtransformationen) erzwungen wird. Das Standardmodell ist eine $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ Eichtheorie.

Ereignishorizont Oberfläche eines Schwarzen Lochs, von der aus kein Signal mehr nach außen entweichen kann. Ein Beobachter, der den Ereignishorizont überquert, bemerkt dabei nichts Besonderes; von außen erscheint er jedoch immer langsamer, je näher er sich dem Horizont nähert.

Fermion Teilchen mit halbzahligem Spin (1/2, 3/2, ...); folgt der Fermi-Dirac-Statistik und dem Pauli-Prinzip (kein Quantenzustand kann doppelt besetzt werden). Alle Quarks und Leptonen sind Fermionen.

Freeze-out Zeitpunkt, ab dem die Reaktionsrate einer Wechselwirkung kleiner als die Hubble-Expansionsrate wird und das chemische Gleichgewicht „einfriert“. Bestimmt z. B. die heutige Dunkle-Materie-Dichte.

Friedmann-Gleichungen Bewegungsgleichungen für den kosmologischen Skalenfaktor $a(t)$, abgeleitet aus der Allgemeinen Relativitätstheorie unter Annahme des Kosmologischen Prinzips. Verbinden die Expansionsrate H mit der Energiedichte ρ des Universumsinhalts.

Gluon Masseloses Eichboson der starken Kernkraft (QCD); trägt selbst Farbladung und vermittelt Wechselwirkungen zwischen Quarks. Im Gegensatz zum Photon können Gluonen miteinander wechselwirken.

Gravitationswelle Raumzeitkrümmungsoszillation, die sich mit Lichtgeschwindigkeit ausbreitet und durch beschleunigte, asymmetrische Massenverteilungen erzeugt wird (z. B. verschmelzende Schwarze Löcher oder Neutronensterne). 2015 erstmals direkt von LIGO detektiert.

Hierarchieproblem Ungeklärte Frage, warum die Higgs-Masse ($\approx 125\,\mathrm{GeV}$) trotz riesiger Quantenkorrekturen so viel kleiner als die Planck-Masse ($\approx 10^{19}\,\mathrm{GeV}$) ist.

Higgs-Boson Skalarteilchen (Spin 0) des Standardmodells, das durch den Higgs-Mechanismus anderen Teilchen Masse verleiht. 2012 am LHC (CERN) entdeckt.

Higgs-Mechanismus / Spontane Symmetriebrechung Mechanismus, durch den das Higgs-Feld einen nicht-verschwindenden Vakuumerwartungswert annimmt und dadurch W/Z-Bosonen sowie Fermionen Masse verleiht. Das Grundzustand des Feldes (Vakuum) besitzt weniger Symmetrie als die Feldgleichungen selbst.

Hubble-Konstante (H_0) Heutige Expansionsrate des Universums in Einheiten von $\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Verschiedene Messmethoden liefern leicht inkonsistente Werte (≈ 67 vs. ≈ 73): die sogenannte Hubble-Spannung.

Inflation Phase exponentieller Expansion des sehr frühen Universums ($a \propto e^{H_{\mathrm{inf}} t}$), angetrieben durch das Inflaton-Feld. Löst das Horizont- und Flachheitsproblem und erklärt den Ursprung der primordialen Dichteschwankungen.

Inflaton Hypothetisches skalares Quantenfeld, das die Inflation antreibt, indem es langsam ein flaches Potential hinabrollt (*Slow-roll*). Am Ende der Inflation zerfällt es in Standardmodell-Teilchen (Reheating).

Konfidenzintervall Frequentistisches Konzept: Ein Intervall, das bei vielfacher Wiederholung des Experiments in einem vorgegebenen Prozentsatz der Fälle (z. B. 95 %) den wahren Parameterwert enthält.

Kosmologische Konstante (Λ) Von Einstein 1917 eingeführter Term in den Feldgleichungen der ART, der einer konstanten Energiedichte des Vakuums entspricht. Heute identifiziert mit Dunkler Energie.

Kosmologisches Prinzip Grundannahme der Kosmologie: Das Universum ist auf großen Skalen homogen (ortsunabhängig) und isotrop (richtungsunabhängig). Erlaubt die Beschreibung der Expansion durch einen einzigen Skalenfaktor $a(t)$.

ΛCDM Standardmodell der Kosmologie; kombiniert eine kosmologische Konstante Λ (Dunkle Energie) mit kalter Dunkler Materie (*cold dark matter*, CDM). Beschreibt erfolgreich CMB, Strukturbildung, Hubble-Expansion und BBN.

Lepton Fundamentales Fermion ohne Farbladung: Elektron (e), Myon (μ), Tauon (τ) sowie die drei zugehörigen Neutrinos (ν_e, ν_μ, ν_τ).

Likelihood Wahrscheinlichkeit, die beobachteten Daten bei gegebenen Modellparametern θ zu erhalten: $\mathcal{L}(\theta) = P(\text{Daten} \mid \theta)$. Zentraler Begriff der frequentistischen und bayesschen Statistik.

LIGO (*Laser Interferometer Gravitational-Wave Observatory*) Bodengestütztes Netzwerk von Michelson-Interferometern mit Armlängen von 4 km, das 2015 erstmals Gravitationswellen aus der Verschmelzung zweier schwarzer Löcher detektierte.

LISA (*Laser Interferometer Space Antenna*) Geplantes weltraumgestütztes Gravitationswellenobservatorium der ESA (Start ∼2035); drei Raumfahrzeuge in einem Dreieck mit 2.5×10^6 km Armlänge, empfindlich im mHz-Band.

Marginalisierung Integration der Posterior-Verteilung über uninteressante ("*Nuisance-*") Parameter, um die Randverteilung in den Parametern von Interesse zu erhalten.

Markov-Kette Folge von Zuständen, bei der der nächste Zustand ausschließlich vom aktuellen abhängt (Gedächtnislosigkeit). Grundbaustein des MCMC-Samplings.

MCMC (Markov-Chain-Monte-Carlo) Numerische Methode, die aus einer komplexen mehrdimensionalen Verteilung Stichproben zieht, indem eine Markov-Kette konstruiert wird, deren stationäre Verteilung der Zielverteilung entspricht.

Metropolis-Hastings-Algorithmus Konkreter MCMC-Algorithmus: In jedem Schritt wird ein Vorschlag aus einer Sprungverteilung gezogen und anhand des Verhältnisses der Posterior-Werte akzeptiert oder abgelehnt.

Monte-Carlo-Methode Klasse numerischer Verfahren, die auf Zufallszahlen basieren; dienen z. B. zur numerischen Integration, Stichprobenziehung oder Simulation stochastischer Prozesse.

NANOGrav (*North American Nanohertz Observatory for Gravitational Waves*) US-amerikanisches Pulsar-Timing-Array; veröffentlichte 2023 starke Belege für einen stochastischen Gravitationswellenhintergrund im Nanohertz-Frequenzband.

Nested Sampling Bayessches Sampling-Verfahren, das simultan Posterior-Stichproben zieht und die Bayes-Evidenz (für Modellvergleiche) effizient berechnet. Implementiert u. a. in `MultiNest` [16], `dynesty` und `nautilus`.

Neutronenstern Extrem kompaktes Objekt ($\sim$1–2 $M_\odot$, Radius $\sim$10 km), das beim Kernkollaps einer massereichen Supernova entsteht und fast ausschließlich aus Neutronen besteht. Rotierende Neutronensterne emittieren regelmäßige Pulse und werden als Pulsare beobachtet.

Neutrino Sehr leichtes, elektrisch neutrales Lepton, das nur schwach wechselwirkt. Existiert in drei Flavors (ν_e, ν_μ, ν_τ); entsteht bei radioaktiven Betazerfällen und Kernfusionsprozessen in Sternen.

Nukleosynthese (BBN, *Big Bang Nucleosynthesis*) Entstehung leichter Atomkerne, hauptsächlich Helium-4, Deuterium und Spuren von Lithium, aus freien Protonen und Neutronen in den ersten $\sim$3 Minuten nach dem Urknall.

p-Wert Wahrscheinlichkeit, unter der Nullhypothese Daten zu beobachten, die mindestens so extrem sind wie die tatsächlich gemessenen. Ein kleiner p-Wert (typisch $p < 0.05$) gilt als Evidenz gegen die Nullhypothese.

Parsec (pc) Astronomische Längeneinheit: $1\,\mathrm{pc} \approx 3.086 \times 10^{16}\,\mathrm{m} \approx 3.26$ Lichtjahre; entspricht der Entfernung, bei der die jährliche Parallaxe eines Sterns eine Bogensekunde beträgt. Kosmologische Abstände werden in Megaparsec (Mpc) angegeben.

Phasenübergang (kosmologisch) Übergang zwischen zwei Zuständen eines Quantenfeldes im frühen Universum, z. B. die elektroschwache Symmetriebrechung ($T \sim 100\,\mathrm{GeV}$) oder der QCD-Übergang ($T \sim 150\,\mathrm{MeV}$). Kann Gravitationswellen erzeugen, wenn er von erster Ordnung ist.

Phasenübergang erster Ordnung Phasenübergang mit latenter Wärme und metastabiler Phase (vgl. Wasser vs. Dampf): Beide Phasen koexistieren vorübergehend. Erzeugt im frühen Universum Blasen und ist eine hypothetische Quelle für einen primordialen Gravitationswellenhintergrund.

Planck-Skala Energieskala der Quantengravitation: Planck-Masse $M_\mathrm{Pl} \approx 1.2 \times 10^{19}\,\mathrm{GeV}$, Planck-Länge $\ell_\mathrm{Pl} \approx 1.6 \times 10^{-35}\,\mathrm{m}$, Planck-Zeit $t_\mathrm{Pl} \approx 5.4 \times 10^{-44}\,\mathrm{s}$. Jenseits dieser Skalen versagen Allgemeine Relativitätstheorie und Quantenfeldtheorie als separate Theorien.

Posterior Bedingte Wahrscheinlichkeitsdichte der Parameter nach Beobachtung der Daten: $P(\theta \,|\, \text{Daten}) \propto \mathcal{L}(\theta) \cdot \pi(\theta)$; enthält alle Information über die Parameter nach dem Satz von Bayes.

Prior (Priori-Verteilung) Wahrscheinlichkeitsdichte der Modellparameter *vor* Beobachtung der Daten; codiert Vorwissen oder konservative Annahmen.

Pulsar Rotierender Neutronenstern, der durch sein starkes Magnetfeld gebündelte Strahlung entlang der Magnetachse aussendet; erscheint für den Beobachter als regelmäßiger Puls. Millisekunden-Pulsare (MSP) rotieren hunderte Male pro Sekunde und sind äußerst präzise kosmische Uhren.

Pulsar-Timing-Array (PTA) Netz von Millisekunden-Pulsaren, das als kohärenter Gravitationswellendetektor im Nanohertz-Band (10^{-9}–$10^{-7}\,\mathrm{Hz}$) fungiert. Bekannte Arrays: NANOGrav (USA), EPTA (Europa), PPTA (Australien), IPTA (international).

QCD (Quantenchromodynamik) Eichtheorie der starken Kernkraft; beschreibt Wechselwirkungen von Quarks (mit Farbladung rot, grün, blau) via Gluonen. QCD ist verantwortlich für den Zusammenhalt von Protonen und Neutronen und für den QCD-Phasenübergang im frühen Universum.

QFT (Quantenfeldtheorie) Theoretischer Rahmen, der Quantenmechanik mit Spezieller Relativitätstheorie vereinigt: Teilchen sind Anregungen quantisierter Felder. Das Standardmodell und alle modernen Teilchentheorien sind QFTs.

Quark Fundamentales Fermion mit Farbladung; gibt es in sechs Varianten (*Flavors*): up, down, strange, charm, bottom, top. Quarks sind stets zu farbneutralen Hadronen (Protonen, Neutronen, ...) gebunden (*Confinement*).

Reheating Aufheizung des Universums nach dem Ende der Inflation durch den Zerfall des Inflaton-Feldes in Standardmodell-Teilchen. Setzt die „übliche“ *hot Big Bang*-Entwicklung in Gang.

Rekombination Epoche ∼380 000 Jahre nach dem Ende der Inflation ($T \approx 3000\,\mathrm{K}$), in der freie Elektronen und Protonen zu neutralem Wasserstoff rekombinierten. Danach wurde das Universum für Photonen transparent – diese bilden den CMB.

Rotationskurve Diagramm der Umlaufgeschwindigkeit von Sternen und Gas als Funktion des Abstands vom Galaxienzentrum. Das beobachtete flache Verhalten ($v = \text{const}$ statt $v \propto 1/\sqrt{r}$) ist ein zentrales Argument für die Existenz Dunkler Materie.

Rotverschiebung Zunahme der Wellenlänge von Photonen; *kosmologische* Rotverschiebung entsteht durch die Expansion des Universums: $1 + z = a_0/a_{\mathrm{emit}}$. Je größer z, desto weiter in der Vergangenheit (und Entfernung) wurde das Licht emittiert.

Schwarzes Loch Objekt, aus dessen Gravitationsfeld selbst Licht nicht entweichen kann; entsteht beim Kollaps massereicher Sterne oder durch Akkretion in Galaxienkernen. Charakterisiert durch seinen Schwarzschild-Radius (Ereignishorizont).

Schwarzschild-Radius r_{S} Radius des Ereignishorizonts eines nicht-rotierenden Schwarzen Lochs: $r_{\mathrm{S}} = 2GM/c^2$. Für die Sonne wären das ≈3 km, für die Erde ≈9 mm.

SGWB (Stochastischer Gravitationswellenhintergrund) Überlagerung einer Vielzahl unaufgelöster Gravitationswellenquellen, die ein nahezu isotropes Hintergrundsignal erzeugen; astrophysikalische Quellen (z. B. SMBHB) oder kosmologische Quellen (z. B. Phasenübergänge).

Skalenfaktor $a(t)$ Dimensionslose, zeitabhängige Funktion, die die relative lineare Ausdehnung des Universums beschreibt; heute $a_0 = 1$, zur Rekombination $a \approx 1/1100$, während der Anfangssingularität $a \to 0$.

SMBH/SMBHB (*Supermassive Black Hole (Binary)*) Schwarze Löcher mit Massen mehr als $10^6\,M_\odot$; im Zentrum fast jeder großen Galaxie. Paare solcher Schwarzen Löcher (Binaries), die entstehen, wenn Galaxien fusionieren, sind die wichtigste astrophysikalische Quelle von Nanohertz-Gravitationswellen.

Spaghettifizierung Gezeitenbedingte Dehnung und seitliche Kompression eines ausgedehnten Körpers durch den Gradienten des Gravitationsfeldes nahe einem Schwarzen Loch; führt zur Zerstörung des Körpers.

Standardmodell der Teilchenphysik Quantenfeldtheorie, die alle bekannten Elementarteilchen (Quarks, Leptonen, Eichbosonen, Higgs-Boson) und drei der vier Grundkräfte (stark, schwach, elektromagnetisch) beschreibt. Extrem präzise experimentell bestätigt; enthält jedoch keine Gravitation, keine Dunkle Materie und keine Dunkle Energie.

Supersymmetrie (SUSY) Theoretische Erweiterung des Standardmodells, die jedem Teilchen einen Superpartner mit um 1/2 verändertem Spin zuordnet. Löst das Hierarchieproblem und bietet einen Dunkle-Matierie-Kandidaten; bisher nicht experimentell bestätigt.

Vakuumerwartungswert (VEV) Nicht-verschwindender Erwartungswert eines Quantenfeldes im Grundzustand: $\langle\phi\rangle \neq 0$. Das Higgs-Feld hat $\langle\phi\rangle = v \approx 246\,\mathrm{GeV}$; dies bricht die elektroschwache Symmetrie und verleiht Teilchen Masse.

WIMP (*Weakly Interacting Massive Particle*) Kandidat für Dunkle Materie mit Masse $\sim$ 1 GeV–1 TeV, der nur gravitativ und schwach wechselwirkt. Motiviert durch das *WIMP-Wunder*: der Freeze-out-Mechanismus liefert natürlich $\Omega_{\mathrm{DM}} \approx 0.27$ für schwach-wechselwirkende Teilchen mit elektroschwacher Masse.

Wirkungsquerschnitt σ Maß für die Wahrscheinlichkeit eines Streuprozesses; hat die Dimension einer Fläche (cm^2 oder GeV^{-2} in natürlichen Einheiten). Ein großer Wirkungsquerschnitt bedeutet eine starke Wechselwirkung.

A Satz von Picard–Lindelöf

Aufgabe A.1

Dieser Text behandelt die Frage, warum eine gewöhnliche Differentialgleichung der Form

$$\partial_t u(t) = f(u(t)) \tag{A.1}$$

eine Lösung hat. Hier bezeichnet $f : \mathbb{R}^n \to \mathbb{R}^n$ eine beliebige Funktion und $n \in \mathbb{N}$ eine natürliche Zahl.

Beispiel 1. Im Fall der Lotka-Volterra-Gleichungen

$$\begin{aligned}\partial_t N_1 &= N_1(\alpha - \beta N_2)\\ \partial_t N_2 &= -N_2(\gamma - \delta N_1)\end{aligned}$$

mit Konstanten $\alpha, \beta, \gamma, \delta > 0$ ist

$$u(t) = \begin{pmatrix} N_1(t) \\ N_2(t) \end{pmatrix} \quad \text{und} \quad f(u(t)) = f(N_1(t), N_2(t)) = \begin{pmatrix} N_1(t)(\alpha - \beta N_2(t)) \\ -N_2(t)(\gamma - \delta N_1(t)) \end{pmatrix}.$$

Um die Existenz möglichst einfach zu beweisen, beschränken wir uns auf Funktionen f, die Lipschitz-stetig sind.

Definition 1. *Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}^n$ ist Lipschitz-stetig, wenn eine Konstante $0 < L < \infty$ existiert, sodass*

$$\|f(x) - f(y)\| \leq L\|x - y\|$$

für alle $x, y \in \mathbb{R}^n$ gilt. Die kleinste aller möglichen Konstanten bezeichnen wir als Lipschitzkonstante.

Hinweis. Die Funktion f bildet einen Vektor aus dem $\mathbb{R}^n$ auf einen anderen Vektor ab. In Definition 1 ist

$$\|v\| = \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{\sum_{i=1}^{n} v_i^2}$$

die *Norm* des Vektors $v = (v_1, \ldots, v_n)^\top \in \mathbb{R}^n$.

Der Name *Lipschitz-stetig* ist berechtigt:

Lemma 1. *Jede Lipschitz-stetige Funktion ist stetig (im klassischen Sinne).*

Um die Existenz von Lösungen zu zeigen, benötigen wir eine Version des folgenden Theorems, das häufig in der Analysis und Numerik angewandt wird.

Theorem 1 (Banachscher Fixpunktsatz, erste Form). *Sei $M \subset \mathbb{R}^n$ eine abgeschlossene Teilmenge des $\mathbb{R}^n$ und $f : M \to M$ eine Abbildung mit Lipschitzkonstante $L < 1$. Dann hat f einen eindeutigen Fixpunkt, d.h. es existiert genau ein x^* mit $f(x^*) = x^*$.*

Fixpunkte sind zum Beispiel wichtig, weil jede Nullstelle $\alpha(x) = 0$ auch ein Fixpunkt der Funktion $x \mapsto \alpha(x) + x$ ist. Das Suchen von Nullstellen und Fixpunkten ist also äquivalent.

Beweis. Sei $x_0 \in M$. Wir definieren eine Folge $(x_k) \subset M$ durch $x_k := f(x_{k-1})$ für alle $k \in \mathbb{N}$ und verwenden die Lipschitz-Stetigkeit von f, um zu zeigen, dass die Folge gegen ein $x^* \in M$ konvergiert. Dazu schätzen wir folgendermaßen ab:

$$\|x_{k+1} - x_k\| = \|f(x_k) - f(x_{k-1})\| \leq L\|x_k - x_{k-1}\|.$$

Mehrfache Anwendung des Arguments ergibt $\|x_{k+1} - x_k\| \leq L^k\|x_1 - x_0\|$. Da $L < 1$, konvergiert $L^k \to 0$ und damit auch $\|x_{k+1} - x_k\| \to 0$ wenn $k \to \infty$. Somit ist (x_k) eine konvergente

Folge. Da M abgeschlossen ist, gibt es auch einen Grenzwert x^* (das ist gerade die Definition von Abgeschlossenheit). f ist als Lipschitz-stetige Funktion auch stetig im klassischen Sinne (Lemma 1), daher gilt

$$x^* = \lim_{k\to\infty} x_{k+1} = \lim_{k\to\infty} f(x_k) = f\left(\lim_{k\to\infty} x_k\right) = f(x^*)\,,$$

also ist x^* ein Fixpunkt. Schließlich nehmen wir an, es gäbe zwei Fixpunkte x^* und y^*. Dann folgt aus der Lipschitz-Stetigkeit $\|x^* - y^*\| = \|f(x^*) - f(y^*)\| \leq L\|x^* - y^*\|$. Diese Ungleichung kann aber nur gelten, wenn $\|x^* - y^*\| = 0$ und somit $x^* = y^*$ gilt. □

Übung 1. *Der Fixpunktsatz ist auch deshalb relevant, weil sein (konstruktiver) Beweis eine Möglichkeit aufzeigt, die Fixpunkte zu approximieren:*

a) Sei $M = [-\frac{1}{2}, \frac{1}{2}]$ und $f : M \to M$, $x \mapsto x^2$. Sind die Voraussetzungen des Banachschen Fixpunktsatzes erfüllt? Berechne die ersten fünf Folgenglieder der Folge (x_k) für ein beliebiges $x_0 \in M$. Was ist der Fixpunkt?

b) Wir wählen nun $M = [0, 1]$. Sind die Voraussetzungen immer noch erfüllt? Gibt es einen eindeutigen Fixpunkt?

Der Banachsche Fixpunktsatz gilt sogar für noch allgemeinere Funktionen, zum Beispiel Operatoren, die stetige Funktionen auf stetige Funktionen abbilden.

Übung 2. *Definiere Lipschitz-Stetigkeit analog zu Definition 1 für Operatoren $F : C^0(I, \mathbb{R}^n) \to C^0(I, \mathbb{R}^n)$, wobei $I = [a, b] \subset \mathbb{R}$ ein beschränktes und abgeschlossenes Intervall ist und $C^0(I, \mathbb{R}^n)$ die Menge aller stetigen Funktionen von I nach $\mathbb{R}^n$ bezeichnet.*
Hinweis: Verwende als Norm für stetige Funktionen $f \in C^0(I, \mathbb{R}^n)$

$$\|f\| := \|f\|_{C^0(I,\mathbb{R}^n)} = \max_{t\in I} \|f(t)\|\,.$$

Theorem 2 (Banachscher Fixpunktsatz, zweite Form). *Ein Lipschitz-stetiger Operator $A : C^0(I, \mathbb{R}^n) \to C^0(I, \mathbb{R}^n)$ mit Lipschitzkonstante $L < 1$ hat einen eindeutigen Fixpunkt.*

Übung 3. *Beweise Theorem 2, indem du dich an dem Beweis von Theorem 1 orientierst.*

Damit verfügen wir nun über das Werkzeug, um den folgenden Existenzsatz zu beweisen.

Theorem 3 (Satz von Picard–Lindelöf). *Sei $f : \mathbb{R}^n \to \mathbb{R}^n$ eine Lipschitz-stetige Funktion mit Lipschitzkonstante N. Sei außerdem T eine positive, reelle Zahl mit $T < \frac{1}{N}$. Dann existiert für alle $t_0 \in \mathbb{R}$ und $u_0 \in \mathbb{R}^n$ eine Lösung $u : J := [t_0 - T, t_0 + T] \to \mathbb{R}^n$ der gewöhnlichen Differentialgleichung (A.1) mit $u(t_0) = u_0$.*

Beweis. Sei $t_0 \in \mathbb{R}$ und $u_0 \in \mathbb{R}^n$ beliebig, aber fest. Wir definieren den Operator

$$\Phi : C^0(J, \mathbb{R}^n) \to C^0(J, \mathbb{R}^n), \qquad \Phi(u)(t) = u(t_0) + \int_{t_0}^{t} f(u(s))\ \mathrm{d}s\,.$$

Übung 4. *Verwende den Hauptsatz der Differential- und Integralrechnung, um zu zeigen, dass ein Fixpunkt von Φ differenzierbar ist und folgere, dass jeder Fixpunkt von Φ gleichzeitig eine Lösung von Gleichung (A.1) ist.*

Um den Satz von Picard–Lindelöf zu beweisen, müssen wir also zeigen, dass Φ einen Fixpunkt besitzt. Dafür verwenden wir Theorem 2 und müssen also zunächst die Lipschitz-Stetigkeit von Φ zeigen: Seien $u, v : J \to \mathbb{R}^n$ zwei Funktionen in $C^0(J, \mathbb{R}^n)$ mit $u(t_0) = v(t_0) = u_0$. Dann gilt wegen der Dreiecksungleichung

$$\begin{aligned}\|\Phi(u)(t) - \Phi(v)(t)\| &= \left\|u_0 - u_0 + \int_{t_0}^{t} f(u(s)) - f(v(s))\ \mathrm{d}s\right\| \\ &\leq \int_{t_0}^{t} \|f(u(s)) - f(v(s))\|\ \mathrm{d}s \qquad \text{(A.2)}\end{aligned}$$

für alle $t \in J$. Aus Abschätzung (A.2) folgt

$$\begin{aligned}\max_{t\in J} \|\Phi(u)(t) - \Phi(v)(t)\| &\leq \max_{t\in J} \int_{t_0}^{t} \|f(u(s)) - f(v(s))\| \, \mathrm{d}s \\ &\leq \int_{t_0}^{t_0+T} \max_{s'\in J} \|f(u(s')) - f(v(s'))\| \, \mathrm{d}s \\ &= T \cdot \max_{s\in J} \|f(u(s)) - f(v(s))\|\end{aligned}$$

und somit mit der Lipschitz-Stetigkeit von f

$$\|\Phi(u) - \Phi(v)\|_{C^0(J,\mathbb{R}^n)} \leq N \cdot T \cdot \max_{s\in J} \|u(s) - v(s)\| = N \cdot T \cdot \|u - v\|_{C^0(J,\mathbb{R}^n)} .$$

Aus der Voraussetzung $T < \frac{1}{N}$ folgt, dass Φ ein Lipschitz-stetiges Funktional mit Lipschitzkonstante $N \cdot T < 1$ ist. Theorem 2 impliziert also, dass Φ einen Fixpunkt und somit die Differentialgleichung (A.1) eine Lösung hat. □

Übung 5. *Nutze die Ideen aus Übung 1, um ein* `python`*-Skript zu schreiben, das die Lösung einer gewöhnlichen Differentialgleichung der Form (A.1) approximiert. Diese Konstruktion ist als* Picard-Iteration *bekannt, aber numerisch viel ineffektiver als die Simpson-Regel.*
Der Satz von Picard–Lindelöf ist ein erster Ausgangspunkt für weitere Fragen, zum Beispiel:

- Können wir die Existenz von Lösungen auch für lange Zeiten zeigen (und nicht nur für Intervalle der Form $[t_0 - T, t_0 + T]$)?
- Sind diese Lösungen eindeutig?
- Kann auf die Lipschitz-Stetigkeit von f verzichtet werden? Genügt zum Beispiel auch ein stetiges f?

Musterlösung A.1

Übung 1a) Auf $M = [-\frac{1}{2}, \frac{1}{2}]$ gilt $|f'(x)| = |2x| \leq 1$, also ist f Lipschitz-stetig mit $L = 1$. Das genügt nicht (wir brauchen $L < 1$), also ist Theorem 1 nicht direkt anwendbar. In der Praxis konvergiert die Folge trotzdem: $x_k = x_0^{2^k} \to 0$ für $|x_0| < 1$. Der eindeutige Fixpunkt in M ist $x^* = 0$.
Übung 1b) Auf $M = [0, 1]$ ist die Lipschitzkonstante $L = \max_{x\in[0,1]} |f'(x)| = \max_{x\in[0,1]} |2x| = 2 > 1$, also sind die Voraussetzungen des Banachschen Fixpunktsatzes nicht erfüllt. Es gibt zwei Fixpunkte: $x^* = 0$ und $x^* = 1$ (da $f(1) = 1$), also keinen eindeutigen, wie das Theorem auch nicht verspricht.
Übung 2) Ein Operator $F : C^0(I, \mathbb{R}^n) \to C^0(I, \mathbb{R}^n)$ heißt Lipschitz-stetig mit Konstante L, wenn für alle $u, v \in C^0(I, \mathbb{R}^n)$ gilt:

$$\|F(u) - F(v)\|_{C^0(I,\mathbb{R}^n)} \leq L \, \|u - v\|_{C^0(I,\mathbb{R}^n)},$$

wobei $\|u\|_{C^0(I,\mathbb{R}^n)} = \max_{t\in I} \|u(t)\|$ die Supremumsnorm ist.
Übung 3) Der Beweis verläuft analog zu Theorem 1. Sei $u_0 \in C^0(I, \mathbb{R}^n)$ beliebig. Definiere die Folge $u_k := F(u_{k-1})$. Dann gilt:

$$\|u_{k+1} - u_k\|_{C^0} = \|F(u_k) - F(u_{k-1})\|_{C^0} \leq L \, \|u_k - u_{k-1}\|_{C^0} \leq L^k \, \|u_1 - u_0\|_{C^0}.$$

Da $L < 1$, ist (u_k) eine Cauchy-Folge im vollständigen Raum $C^0(I, \mathbb{R}^n)$ (mit Supremumsnorm ist C^0 ein Banachraum). Es existiert ein Grenzwert $u^* \in C^0$, und Lipschitz-Stetigkeit von F liefert $u^* = \lim_k F(u_k) = F(u^*)$. Eindeutigkeit folgt wie in Theorem 1. □

Übung 4) Sei u^* ein Fixpunkt von Φ, d.h. $u^*(t) = u_0 + \int_{t_0}^{t} f(u^*(s))\,\mathrm{d}s$. Nach dem Hauptsatz ist u^* differenzierbar und $\partial_t u^*(t) = f(u^*(t))$. Außerdem gilt $u^*(t_0) = u_0$. Also löst u^* die Gleichung (A.1). □

Übung 5) Die Picard-Iteration approximiert die Lösung durch sukzessives Einsetzen: Man startet mit $u_0(t) \equiv u_{\text{init}}$ und berechnet $u_{k+1}(t) = u_0 + \int_{t_0}^{t} f(u_k(s))\,\mathrm{d}s$ numerisch.

```
import numpy as np
import matplotlib.pyplot as plt

def picard_iteration(f, t0, u_init, T, n_steps=200, n_iter=10):
    t = np.linspace(t0, t0 + T, n_steps)
    dt = t[1] - t[0]
    u = np.full_like(t, float(u_init))      # u_0 = konstante Funktion
    for _ in range(n_iter):
        integrand = np.array([f(u[i]) for i in range(n_steps)])
        u = u_init + np.cumsum(integrand) * dt  # Euler-Quadratur des Integrals
    return t, u

# Beispiel: u' = u, u(0) = 1  ->  exakte Loesung e^t
t, u_pic = picard_iteration(f=lambda u: u, t0=0, u_init=1.0, T=2.0)
plt.plot(t, u_pic, label='Picard-Iteration ($n=10$)')
plt.plot(t, np.exp(t), '--', label=r'exakt $\mathrm{e}^t$')
plt.legend(); plt.xlabel('$t$'); plt.ylabel('$u(t)$')
plt.title('Picard-Iteration vs. exakte Loesung')
plt.show()
```

Die Iteration konvergiert für kleine $T < 1/N$ gegen die exakte Lösung, aber wesentlich langsamer als moderne ODE-Solver (`scipy.integrate.solve_ivp`), die auf Runge-Kutta-Verfahren basieren und pro Schritt deutlich mehr Information nutzen.

Literatur

[1] Pablo Carlos Budassi. *Logarithmic Map of the Observable Universe*. Wikimedia Commons. 2012. URL: https://commons.wikimedia.org/wiki/File:Observable_Universe_Logarithmic_Map_(2020_edition)_rendered_by_Pablo_Carlos_Budassi.png.

[2] N. Aghanim u. a. „Planck 2018 results. VI. Cosmological parameters“. In: *Astron. Astrophys.* 641 (2020). [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[3] Edwin Hubble. „A relation between distance and radial velocity among extra-galactic nebulae“. In: *Proc. Nat. Acad. Sci.* 15 (1929), S. 168–173. DOI: 10.1073/pnas.15.3.168.

[4] Dan Scolnic u. a. „The Pantheon+ Analysis: The Full Data Set and Light-curve Release“. In: *Astrophys. J.* 938.2 (2022), S. 113. DOI: 10.3847/1538-4357/ac8b7a. arXiv: 2112.03863 [astro-ph.CO].

[5] Ken'ichi Saikawa und Satoshi Shirai. „Primordial gravitational waves, precisely: The role of thermodynamics in the Standard Model“. In: *JCAP* 05 (2018), S. 035. DOI: 10.1088/1475-7516/2018/05/035. arXiv: 1803.01038 [hep-ph].

[6] Mark B. Hindmarsh, Marvin Lüben, Johannes Lumma und Martin Pauly. „Phase transitions in the early universe“. In: *SciPost Phys. Lect. Notes* 24 (2021), S. 1. DOI: 10.21468/SciPostPhysLectNotes.24. arXiv: 2008.09136 [astro-ph.CO].

[7] Jonas Matuszak und Carlo Tasillo. *TransitionListener v2.0 – Robust gravitational wave predictions for cosmological phase transitions*. arXiv:2605.15259 [hep-ph]. Mai 2026. arXiv: 2605.15259 [hep-ph].

[8] Event Horizon Telescope Collaboration. „First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole“. In: *Astrophys. J. Lett.* 875 (2019), S. L1. DOI: 10.3847/2041-8213/ab0ec7. arXiv: 1906.11238 [astro-ph.GA].

[9] NASA/ESA, L. Calçada. *Anatomy of a Black Hole*. NASA/ESA press image. 2019. URL: https://www.nasa.gov/wp-content/uploads/2019/09/bh_labeled.jpg.

[10] Torsten Bringmann, Paul Frederik Depta, Thomas Konstandin, Kai Schmidt-Hoberg und Carlo Tasillo. „Does NANOGrav observe a dark sector phase transition?“ In: *JCAP* 11 (2023), S. 053. DOI: 10.1088/1475-7516/2023/11/053. arXiv: 2306.09411 [astro-ph.CO].

[11] Mark Hindmarsh, Stephan J. Huber, Kari Rummukainen und David J. Weir. „Numerical simulations of acoustically generated gravitational waves at a first order phase transition“. In: *Phys. Rev. D* 92.12 (2015), S. 123009. DOI: 10.1103/PhysRevD.92.123009. arXiv: 1504.03291 [astro-ph.CO].

[12] Martin Wolfgang Winkler und Katherine Freese. „Origin of the stochastic gravitational wave background: First-order phase transition versus black hole mergers“. In: *Phys. Rev. D* 111.8 (2025), S. 083509. DOI: 10.1103/PhysRevD.111.083509. arXiv: 2401.13729 [astro-ph.CO].

[13] Wolfram Ratzinger und Pedro Schwaller. „Whispers from the dark side: Confronting light new physics with NANOGrav data“. In: *SciPost Phys.* 10.2 (2021), S. 047. DOI: 10.21468/SciPostPhys.10.2.047. arXiv: 2009.11875 [astro-ph.CO].

[14] Adeela Afzal u. a. „The NANOGrav 15 yr Data Set: Search for Signals from New Physics“. In: *Astrophys. J. Lett.* 951.1 (2023), S. L11. DOI: 10.3847/2041-8213/acdc91. arXiv: 2306.16219 [astro-ph.HE].

[15] Sowmiya Balan, Torsten Bringmann, Felix Kahlhoefer, Jonas Matuszak und Carlo Tasillo. „Sub-GeV dark matter and nano-Hertz gravitational waves from a classically conformal dark sector“. In: *JCAP* 08 (2025), S. 062. DOI: 10.1088/1475-7516/2025/08/062. arXiv: 2502.19478 [hep-ph].

[16] F. Feroz, M. P. Hobson und M. Bridges. „MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics“. In: *Mon. Not. Roy. Astron. Soc.* 398 (2009), S. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437 [astro-ph].