DSA26 Dokumentation
Kurs 5.1 – Unendliche Weiten

Akademie Roßleben 2026

Teilnehmende des Kurses 5.1

Juli 2026

INHALTSVERZEICHNIS

GRUNDLAGEN DER KOSMOLOGIE

Allgemeine Relativitätstheorie, Rotverschiebung

von Julian Lennartz und Alex Omlor

Die Frage nach dem Ursprung, dem Aufbau und der Entwicklung des Universums gehört zu den zentralen Themen der modernen Physik. Mithilfe theoretischer Modelle und astronomischer Beobachtungen versucht die Kosmologie, die großräumigen Eigenschaften des Universums zu beschreiben und seine Entwicklung nachzuvollziehen. Eine besondere Bedeutung kommt dabei der Allgemeinen Relativitätstheorie zu, die die Grundlage des heutigen kosmologischen Standardmodells bildet. Diese Dokumentation soll dabei die wichtigsten Konzepte der modernen Kosmologie, insbesondere das kosmologische Prinzip, grundlegende kosmologische Größen sowie die Friedmann-Gleichungen als mathematische Beschreibung der Expansion des Universums erläutern.

1.1 Die Allgemeine Relativitätstheorie

Die Allgemeine Relativitätstheorie (ART) wurde 1915 von Albert Einstein entwickelt und spielt bis heute eine eine zentrale Rolle in der modernen Kosmologie. Es handelt sich dabei um ein Sytem von 10 gekoppelten, nichtlinearen, partiellen Differentialgleichungen zweiter Ordnung. Die ihr zugrundenliegenden Einsteingleichungen lauten:

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \,. \tag{1.1}$$

Aus diesen Gleichungen geht das Verständnis einer veränderlichen, dynamischen Raumzeit hervor. Im Gegensatz zur klassischen Newtonschen Physik beschreibt die ART Gravitation nicht als eine Kraft zwischen zwei Massen, sondern als Folge der Krümmung der Raumzeit. Materie und Energie beeinflussen die Geometrie der Raumzeit, wodurch sich die Bewegungen von Körpern und Lichtstrahlen ergeben. Die Einsteingleichungen verknüpfen die Verteilung von Materie und Energie mit der Krümmung der Raumzeit. Anschaulich lässt sich dieser Zusammenhang mit folgendem Zitat von John Wheeler zusammenfassen: **Matter tells space how to curve and space tells matter how to move.** Erst durch ihre Anwendung können Modelle entwickelt werden, welche die Expansion des Universums mathematisch erfassen. Auf Basis zusätzlicher Annahmen über die großräumige Struktur des Universums lassen sich schließlich auch die Friedmann-Gleichungen herleiten.

1.2 Das Kosmologische Prinzip

Das Kosmologische Prinzip beruht auf der Annahme, dass das Universum auf ausreichend großen Skalen sowohl homogen als auch isotrop ist. Homogenität bedeutet, dass das Universum im statistischen Mittel (beziehungsweise auf sehr großen Skalen) an jedem Ort dieselben Eigenschaften besitzt, während Isotropie beschreibt, dass es in jede Richtung gleich aussieht. Obwohl lokal deutliche Strukturen wie Sterne, Galaxien und Galaxienhaufen existieren, verschwinden diese Unterschiede bei Betrachtung sehr großer Skalen ab Größenodrnungen von Galxienclustern. Das Kosmologische Prinzip vereinfacht die Feldgleichungen der ART und ermöglichen somit die Erschließung der Friedmanngleichungen.

1.3 Der Skalenfaktor

Aus dem kosmologischen Prinzip folgt, dass sich die Geometrie des Universums vollständig durch eine einzige Funktion der Zeit beschreiben lässt, den sogenannten Skalenfaktor $a(t)$. Damit lässt sich beschreiben, wie sich die Abstände von zwei Punkten im Universum (die sich frei fallend bewegen, also nur durch die Gravitation beeinflusst sind) mit der Zeit verändern. Für $a(t_1) < a(t_2)$ mit $t_1 < t_2$, war der Abstand zur früheren Zeit t_1 kleiner als zur späteren Zeit t_2, das Universum hat sich also ausgedehnt. Dabei wurde festgelegt, dass $a(t_0) = 1$

für heute gilt. Dieser Abstand L lässt sich in Abhängigkeit von der Zeit über den Skalenfaktor beschreiben:

$$L(t) = \frac{a(t)}{a_*} \cdot L_* \,, \tag{1.2}$$

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

$$H(t) = \frac{1}{a(t)} \frac{\mathrm{d}a}{\mathrm{d}t}(t) \,. \tag{1.3}$$

Diese gibt an, 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 $[H] = T^{-1}$. Wie wir unten sehen werden, ist die geläufigere Einheit der Hubble-Konstanten aber $[H] = \mathrm{km\,(s\,Mpc)^{-1}}$.

1.4 Die Friedmann-Gleichungen

Grundlage für die Friedmann-Gleichungen sind das kosmologische Prinzip und die allgemeine Relativitätstheorie. Zur Herleitung wird das Kosmologische Prinzip in die Einsteingleichungen eingesetzt. Ergebnis davon sind die zwei Friedmann-Gleichungen:

$$H^2(t) = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\rho(t) = \frac{1}{3m_{\mathrm{Pl}}^2} \tag{1.4}$$

$$\dot{p}(t) = -3H(t)[\rho(t) + p(t)] \tag{1.5}$$

Aus den zwei Friedmanngleichungen folgen dann zwei grundlegende Prinzipien, die im Folgenden kurz behandelt werden. **Matter tells space how to curve**: Die Materie-Energiedichte ρ bestimmt, wie schnell sich das Universum ausdehnt (H). Und **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.

1.5 Die Rotverschiebung

Nun wollen wir die Gleichung (1.2) auf die Wellenlänge von Licht anwenden. Da der Skalenfaktor zum Zeitpunkt der Emission $a_{\mathrm{e}} < 1$ war (da das Universum in der Vergangenheit dichter war), gilt für die heute beobachtete Wellenlänge $\lambda_0 > \lambda_{\mathrm{e}}$. Das Licht wird während seiner Ausbreitung durch die kosmische Expansion zu größeren Wellenlängen hin gestreckt. Dieser Effekt wird als Rotverschiebung bezeichnet, da Rot die langwelligste Farbe des sichtbaren Spektrums ist. Man definiert die Rotverschiebung als die relative Änderung der Wellenlänge:

$$z = \frac{\lambda_0 - \lambda_{\mathrm{e}}}{\lambda_{\mathrm{e}}}. \tag{1.6}$$

Daraus lässt sich die heute beobachtete Wellenlänge mit $\lambda_0 = \lambda_{\mathrm{e}}(1 + z)$ berechnen. Unter Ausnutzung des Zusammenhangs zwischen Wellenlänge und Skalenfaktor ergibt sich für den Skalenfaktor zum Zeitpunkt der Emission:

$$a_{\mathrm{e}} = \frac{\lambda_{\mathrm{e}}}{\lambda_0} = \frac{1}{1+z}. \tag{1.7}$$

Die Rotverschiebung ist damit eine direkt messbare Größe, die uns sagt, wie dicht das Universum war, als das beobachtete Licht emittiert wurde. In Abbildung 1.1 wird dieser Effekt veranschaulicht: Wird Licht emittiert (die blaue, kurze Wellenlänge), so wird das Licht durch die kosmische Expansion zu längeren Wellenlängen hin gestreckt und kommt rot verschoben bei uns an.

Abbildung 1.1: Licht mit Wellenlänge die mit dem expandierten Raum mitgedehnt und somit rot verschoben wird, entnommen aus dem Skript von Carlo Tasillo.

1.6 Das Hubblegesetz

Das Hubble Gesetz besagt, dass für kleine Rotverschiebungen, also für nicht allzu weit entfernte Objekte die Fluchtgeschwindigkeit v einer Galaxie näherungsweise proportional zu ihrer Entfernung d_L von uns ist:

$$v \approx H_0 \cdot d_L \,. \tag{1.8}$$

Edwin Hubble beobachtete 1929, dass sich Galaxien je weiter sie von uns enfernt sind, schneller von uns entfernen. Es war somit ein überzeugender Beleg auf ein expandierendes Universum wie auch indirekt ein Beleg für die Urknall-Theorie. In Abbildung 1.2 sieht man in dem linken Diagramm, dass Hubbles Wert $H_0 \approx 424 \ \mathrm{km\,(s\,Mpc)^{-1}}$ aus 1929 aufgrund von fehlerhafter Entfernungskalibrierung zu groß war. In dem mordernen Hubble Diagramm - mit repräsentativen Typ-la-Supernovae zeigt die gestrichelte Linie die SH0ES-Referenz $H_0 = 73.03 \ \mathrm{km\,(s\,Mpc)^{-1}}$, den aktuellen Messwert der SH0ES-Kollaboration. Die Hubble-Rate ist somit keine zeitliche Konstante, sondern verändert sich dynamisch mit der Entwicklung des Universums. Ihr aktueller Wert wird als Hubble-Konstante bezeichnet. Während in der Kosmologie theoretisch oft die natürliche Einheit eV verwendet wird, ist phänomenologisch die Angabe in eben jener zuvor genannten Einheit aus 1.3 gebräuchlicher. Die Hubble-Konstante wird üblicherweise in der Astronomie als

$$H_0 = 73.03\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$$

angegeben. Da $1\,\mathrm{Mpc} = 3.086 \times 10^{19}\,\mathrm{km}$ gilt, lässt sich diese Einheit auch als inverse Zeit schreiben:

$$H_0 \approx 2.37 \times 10^{-18}\,\mathrm{s^{-1}} \approx 0.074\,\mathrm{Gyr^{-1}} \approx 1.56 \times 10^{-33}\,\mathrm{eV},$$

wobei für die letzte Umrechnung $\hbar = 6.582 \times 10^{-16}\,\mathrm{eV\,s}$ verwendet wurde. Dies verdeutlicht, dass die Hubble-Konstante die charakteristische Expansionsrate des Universums beschreibt.

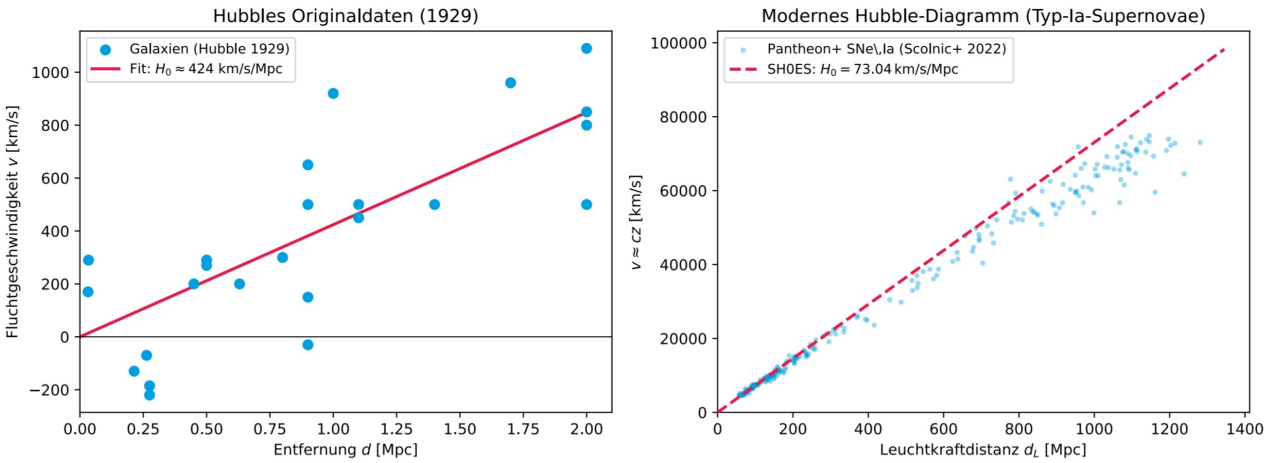


Abbildung 1.2: Vergleich vom originalen Hubble Diagramm mit dem Mordernen, entnommen aus dem Skript von Carlo Tasillo.

DAS STANDARDMODELL DER TEILCHENPHYSIK

Natürliche Einheiten und Zusammensetzung des primordialen Plasmas

von Mathilde, Amalia

2.1 Natürliche Einheiten

Im Zusammenhang mit Astronomie sind natürlichen Einheiten unverzichtbar, da sich durch diese Vereinheitlichung physikalische Zusammenhänge und Gleichungen vereinfachen lassen. Es sind Maßeinheiten, welche auf Naturkonstanten wie etwa der Lichtgeschwindigkeit, dem reduzierten Planck'schen Wirkungsquantum und der Boltzmann-Konstante beruhen, folglich nicht arbiträr festgelegt wurden, wie etwa der Meter. All diese Naturkonstanten werden gleich eins gesetzt:

$$c = \hbar = k_{\mathrm{B}} = 1\,. \tag{2.1}$$

Die Lichtgeschwindigkeit c ermöglicht das Umrechnen von Längeneinheiten in Zeitintervalle.

Indem das reduzierte Planck'sche Wirkungsquantum $\hbar$ gleich eins gesetzt wird, ist die Entsprechung eines Joule in Hertz bekannt. Vor dem Hintergrund, dass mit Joule große Energiemengen besser zu beschreiben sind, die deutlich höhere Frequenzen als etwa Radiowellen haben, bietet sich Elektronenvolt als alternative, nicht SI-Einheit an. Dies ist die Energie, die ein Elektron benötigt, um eine Potentialdifferenz von einem Volt zu überwinden. Des weiteren wird sie in der Physik als Basiseinheit verwendet, wie in Tabelle 2.1 zu sehen.

Die Temperatur eines Gases bzw. die thermische Energie der Teilchen im Gas wird mithilfe der Boltzmann-Konstante bestimmt. Sie hilft also auch, Kelvin in Elektronenvolt zu bestimmen. Zusätzlich belegt sie, dass Energie gleich Masse ist, wie Einstein es in $E = mc^2$ beschrieben hat.

Aus der Gleichung ergibt sich eine Reihe an Umrechnungesfaktoren von SI-Einheiten in Elektronenvolt.

Größe	**Einheit (geschrieben)**	**Wert in SI-Einheiten**
Energie	$1\ \mathrm{eV}$	$1.602 \cdot 10^{-19}\ \mathrm{J}$
Länge	$\frac{1}{1\ \mathrm{eV}}$	$1.973 \cdot 10^{-7}\ \mathrm{m}$
Zeit	$\frac{1}{1\ \mathrm{eV}}$	$6.582 \cdot 10^{-16}\ \mathrm{s}$
Masse	$1\ \mathrm{eV}$	$1.783 \cdot 10^{-36}\ \mathrm{kg}$
Dichte	$1\ \mathrm{eV}^4$	$2.320 \cdot 10^{-16}\ \mathrm{kg/m^3}$
Impuls	$1\ \mathrm{eV}$	$5.344 \cdot 10^{-28}\ \mathrm{N \cdot s}$
Temperatur	$1\ \mathrm{eV}$	$1.160 \cdot 10^{4}\ \mathrm{K}$

Tabelle 2.1: Umrechnungsfaktoren für die wichtigsten physikalischen Größen zwischen natürlichen Einheiten ($c = \hbar = k_B = 1$, Basiseinheit eV) und SI-Einheiten.

$$F_S = 6.63 \times 10^{-9}\,\frac{1}{\mathrm{eV}} = 6.63 \times 10^{-9} \times 1.973 \cdot 10^{-7}\ \mathrm{m}$$

Sowohl die Einheiten als auch die Umrechnungsfaktoren geben Hinweise über die Zusammenhänge und Proportionen der Einheiten: Wenn die Länge sich verändert, ändert sich das Zeitintervall proportional und die Energie antiproportional zu ihr. Als Beispiel einer Umrechnung kann die Distanz von der Erde zur Sonne $d_{SE} = 1.496 \times 10^8\,\mathrm{km}$ in s in eV berechnet werden.

$$
\begin{aligned}
d_{\mathrm{SE}} = 1.496 \times 10^{8}\,\mathrm{km} &= 499\,\mathrm{s} \\
\Leftrightarrow 499\,\mathrm{s} &= \frac{499}{6.582 \times 10^{-16}\,\mathrm{eV}} \\
\Leftrightarrow 499\,\mathrm{s} &= 7.581 \times 10^{-18}\,\mathrm{eV}
\end{aligned}
$$

Im frühen Universum gehen daher Temperaturen von 10^{32} Kelvin mit dementsprechend viel Energie und Masse in einer hohen Dichte einher, weswegen es sich auch in kurzer Zeit exponentiell schnell ausbreiten konnte.

2.2 Zusammensetzung des primordialen Plasmas

Das primordiale Plasma bezeichnet den extrem heißen und dichten Zustand der Materie nur wenige Sekundenbruchteile nach dem Urknall. Dieses Plasma bestand aus den Elementarteilchen, den fundamentalen Teilchen, aus denen unsere Mateire aufgebaut ist.

2.2.1 Die Elementarteilchen des Standardmodells

Lange Zeit betrachtete man Atome als die kleinste Einheit unseres Universums. Diese Annahme ist heutzutage widerlegt. Schon Ernest Rutherford belegte mit seinem bekannten Streuversuch, dass Atome einen positiven Kern haben müssen und es somit kleinere Teilchen als das Atom geben muss.

Heutzutage ist diese Erkenntnis triviales Wissen: Atome besitzen einen Atomkern, in dem sich Protonen und Neutronen befinden. Um den Atomkern herum befinden sich Elektronen. Protonen und Neutronen allerdings sind ebenfalls keine fundamentalen Teilchen. Sie bestehen aus den Teilchen, welche im Standardmodell der Teilchenphysik beschrieben werden.

Das Standardmodell beschreibt die Elementarteilchen. Sie werden in *Quarks*, *Leptonen*, *Eichbosonen* und in das *Higgs* eingeteilt. Quarks und Leptonen werden als *Fermionen* zusammengefasst und sind die Bausteine unserer Materie. Zu den Leptonen zählen Elektronen, Myonen, Tauen und die jeweiligen Neutrinos. Diese sind neutral geladen, besitzen eine sehr geringe Masse und gehen kaum Wechselwirkung mit anderen Elementarteilchen ein.

Zu den Quarks zählen das Up, Down, Charm, Strange, Top und Bottom. Aus Quarks bestehen beispielsweise Protonen und Neutronen. Beide bestehen aus Ups und Downs: Zwei Up-Quarks und ein Down-Quark bilden ein Proton, ein Neutron widerum besteht aus zwei Down-Quarks und einem Up-Quark. Eine solche Verbindung aus drei Quarks wird als Baryon bezeichnet, während ein Meson eine Verbindung aus zwei Quarks betitelt.

Zusammengehalten werden die Quarks durch die *starke Kernkraft*, auch starke Wechselwirkung genannt, vermittelt über das Gluon. Gluonen gehören zu den Eichbosonen. In diese Kategorie fällt auch das Photon. Dieses ist für die *elektromagnetische Kraft* verantwortlich, die Elektronen an ihren Kern bindet.

Die *schwache Kernkraft* wird über die $W^{\pm}$- und Z-Bosonen vermittelt. Diese Kraft kann zum Beispiel Protonen zu Neutronen umwandeln und umgekehrt. Sie ist auch verantwortlich für radioaktive Zerfälle. Zuletzt gibt es noch das Higgs-Boson. Dieses ist über das Higgs-Feld dafür verantwortlich, dass Teilchen Massen haben.

Zu jedem Teilchen gibt es noch ein zugehöriges Anti-Teilchen. Die Anti-Teilchen besitzen die gleichen Eigenschaften wie die zugehörigen Teilchen, haben jedoch eine genau umgekehrte Ladung. Diese Antimaterie bildet das Gegenstück zu unserer Materie.

2.2.2 Das primordale Plasma im Zusammenhang mit der Expansion des Universums

Das verbreitetste physikalische Modell, welches die Entstehung unseres Universums beschreibt, nennt sich die Urknall-Theorie oder *Big-Bang-Theorie*. Diese beschreibt die Expansion des Universums, ausgehend von einem dichten und heißen Ausgangszustand. Aufgrund dieser rapiden Expansion nahm die Materie-Energie-Dichte ρ so schnell ab, dass die Temperatur innerhalb von einer Sekunde von $150\,\mathrm{GeV}$ (Die Temperatur im *Elektroschwachen Phasenübergang*, etwa $10\,\mathrm{ps}$ nach dem Urknall) auf ungefär einen Wert von $1\,\mathrm{MeV}$ absank.

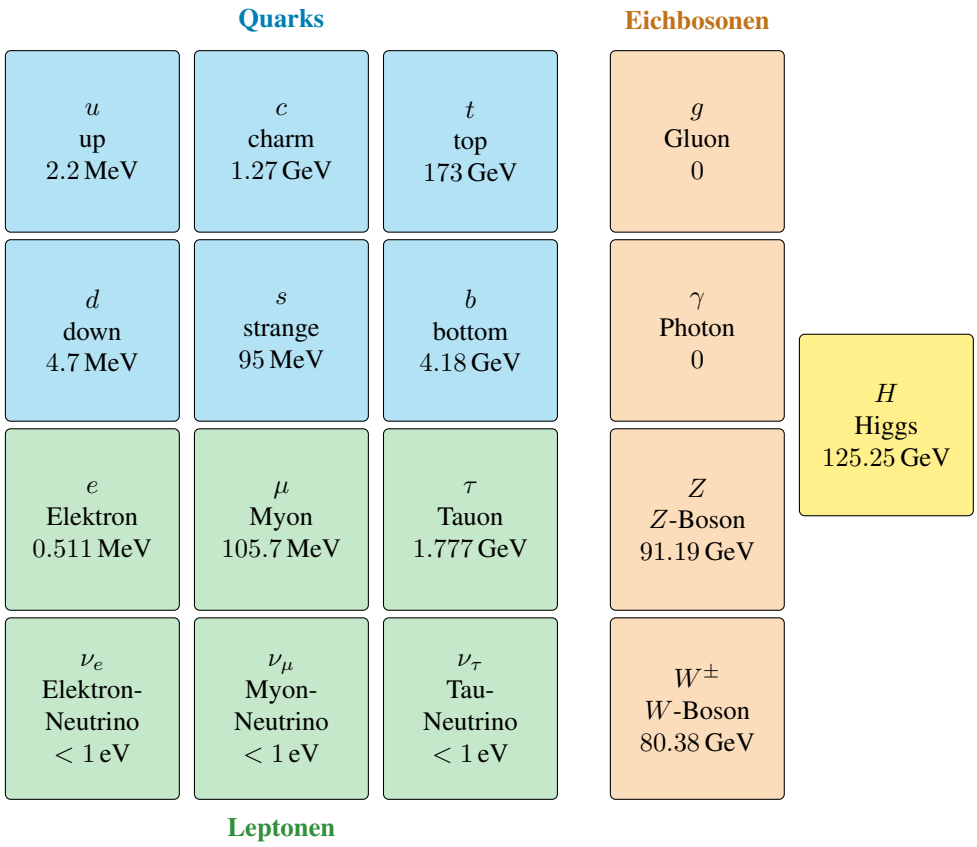


Abbildung 2.1: Die Elementarteilchen des Standardmodells mit ihren (Ruhe-)Massen. Quarks und Leptonen (links, je drei „Generationen“) sind Fermionen und bilden die Materie; Gluon, Photon, Z- und W-Bosonen (rechts) sind die Austauschteilchen der starken, elektromagnetischen und schwachen Wechselwirkung. Das Higgs-Boson ist für die Massen der anderen Teilchen verantwortlich.

Nun wollen wir uns der Zusammensetzung des primordialen Plasmas widmen, welche von der Temperatur abhängig ist. Die Temperatur steht im engen Zusammenhang mit der Zusammensetzung des Plasmas, da nur Teilchen mit einer Masse $m \leq T$ effektiv produziert werden können und so zur Materie-Energie-Dichte ρ beitragen. Teilchen mit einer Masse $m \ll T$ werden exponentiell unterdrückt.

So lässt sich feststellen, dass mit abnehmender Temperatur die Anzahl der verschiedenen Teilchenspezies, welche noch produziert werden können, ebenfalls abnimmt. Zu Beginn, direkt nach dem Ende der sogenannten Inflationsphase, bei $T = 10^{16}\,\text{GeV}$ bestand das Plasma aus allen bekannten Teilchenspezies. Ab einer Temperatur von $T = 10\,\text{GeV}$ besaßen das Higgs, das Top-Quark und die $W^{\pm}$- und Z-Bosonen eine zu hohe Masse und wurden exponentiell unterdrückt. Der *QCD-Phasenübergang* bezeichnet die Zeit etwa $20\,\mu\text{s}$ nach dem Urknall. Zu diesem Zeitpunkt lag eine Temperatur von etwa $150\,\text{MeV}$ vor. Er bezeichnet die Phase, in er Quarks und Gluonen sich mittels der starken Kernkraft zu Protonen und Neutronen verbanden. Man spricht hierbei von dem *Confinement*. Kurz vor der Elektron-Positron-Annihilation bei $T = 0.1\,\text{MeV}$ konnte nahezu keines der Elementarteilchen mehr produziert werden, mit Ausnahme des Photons und der Neutrinos. Aufgrund des kleinen Überschuss an Materie im Vergleich zur Antimaterie lagen außerdem noch freie Protonen und Neutronen im Plasma vor, welche sich dann zum Zeitpunkt der Rekombination ($T = 0.1\,\text{meV}$) zusammen mit den verbliebenen Elektronen zu neutralem Wasserstoff- und Helium-Gas verbanden. Ab diesem Zeitpunkt wurde das Universum elektrisch neutral und durchsichtig.

DISKRETE UND BEDINGTE WAHRSCHEINLICHKEITEN

von Mahika, Aylin

Um die Grundfragen des Kurses beantworten zu können benötigt man, wie so oft, nicht nur ein physikalisches Verständnis, sondern auch mathematische beziehungsweise statistische Grundkenntnisse.

3.1 Grundlagen und Wahrscheinlichkeitsräume

Bei einem Zufallsexperiment, wie dem zweimaligen Drehen eines Glücksrades, sind mehrere Ergebnisse möglich.

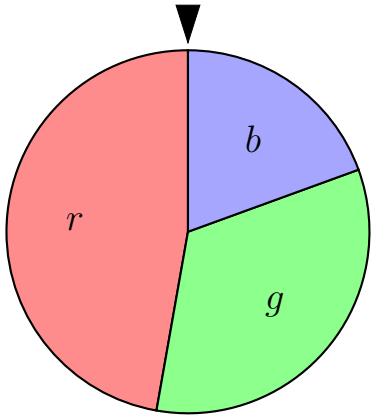


Abbildung 3.1: Ein Glücksrad mit drei unterschiedlich großen, farbig markierten Feldern – rot (r), grün (g) und blau (b) –, deren Flächenanteil jeweils der Wahrscheinlichkeit entspricht, mit der der Zeiger dort stehen bleibt (eigene Darstellung).

Man fasst die möglichen Ergebnisse des Zufallsexperiments in der Ergebnismenge $\Omega = \{r, g, b\}$ zusammen. Dabei steht r dafür, dass der Zeiger im roten Feld stehen bleibt; g und b sind entsprechend für das grüne bzw. blaue Feld zu verstehen. Jedem Ergebnis wird eine Wahrscheinlichkeit zugeordnet. Aus dieser Zuordnung ergibt sich mithilfe der Additivität (vgl. Kolmogorov-Axiom 3) auch für jedes Ereignis, also jede Teilmenge von Ω, eine Wahrscheinlichkeit. Wahrscheinlichkeiten sind mathematische Modelle für die Realität. Die Wahrscheinlichkeiten für r, g und b könnte zum Beispiel mittels des Flächenanteils modelliert werden.

Als Motivation dafür, wie man bei mehrstufigen Zufallsexperimenten die Ergebnismenge konstruiert, betrachten wir das eingangs erwähnte zweistufige Glücksrad-Experiment, bei dem das Rad zweimal hintereinander gedreht wird. Ein Ergebnis besteht dabei aus dem Paar der beiden Einzelergebnisse, sodass sich die Ergebnismenge als

$$\Omega = \{rr, rg, rb, gr, gg, gb, br, bg, bb\}$$

auffassen lässt. Auf dieselbe Weise – als Menge der Tupel aller möglichen Einzelergebnisse – lässt sich die Ergebnismenge auch für kompliziertere mehrstufige Experimente mit mehr Stufen oder größeren Ergebnismengen pro Stufe konstruieren.

Eine inhaltliche Definition der Wahrscheinlichkeit ist schwierig. Die Rechenregeln dagegen lassen sich präzise angeben. Grundlage ist daher ein axiomatischer Zugang: Die Wahrscheinlichkeit wird nicht inhaltlich bestimmt, sondern über einige wenige Grundregeln, aus denen sich alles Weitere ableiten lässt. Sie ist damit zunächst nichts weiter als eine Abbildung, die Ereignissen Zahlen zuordnet und dabei diesen Regeln genügt.

Der Vorteil dieses Vorgehens liegt darin, dass die Rechenregeln unabhängig von der inhaltlichen Deutung gelten. Die Axiome sagen also nichts darüber aus, was Wahrscheinlichkeit bedeutet, sondern nur, wie sie sich verhält.

3.1.1 Der Wahrscheinlichkeitsraum

Das zentrale mathematische Objekt der Stochastik ist der **Wahrscheinlichkeitsraum**, ein Tripel

$$(\Omega, \Sigma, P).$$

Er besteht aus folgenden drei Bestandteilen:

- Die **Grundmenge** Ω ist die Menge aller möglichen Ausgänge des betrachteten Zufallsexperiments. Für einen sechsseitigen Würfel ist etwa
$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$
- Das **Ereignissystem** Σ ist eine Menge von Teilmengen von Ω; ihre Elemente heißen **Ereignisse**. Ein Ereignis wie $\{1, 2, 3\}$ meint, es fällt eine 1, 2 oder 3. Im diskreten Fall wählt man für Σ üblicherweise die volle **Potenzmenge**
$$\Sigma = \mathcal{P}(\Omega) := \{\, U \mid U \subseteq \Omega \,\},$$
also die Menge *aller* Teilmengen von Ω. Zu jedem Ereignis A gehört sein **Komplementärereignis** $A^c = \Omega \setminus A$, das genau dann eintritt, wenn A eben nicht eintritt.
- Das **Wahrscheinlichkeitsmaß**
$$P : \Sigma \longrightarrow [0, 1]$$
ordnet jedem Ereignis eine Zahl zwischen 0 und 1 zu, seine Wahrscheinlichkeit.

3.1.2 Die Kolmogorov-Axiome

Damit P tatsächlich als Wahrscheinlichkeitsmaß gelten darf, muss es die drei Kolmogorov-Axiome erfüllen:

$$P(E) \geq 0 \qquad \text{für alle } E \in \Sigma, \tag{3.1}$$

$$P(\Omega) = 1, \tag{3.2}$$

$$P(A \cup B) = P(A) + P(B) \qquad \text{für } A, B \in \Sigma \text{ mit } A \cap B = \varnothing. \tag{3.3}$$

Anschaulich besagen sie: Wahrscheinlichkeiten sind niemals negativ (3.1); das sichere Ereignis Ω hat Wahrscheinlichkeit 1 (3.2); und die Wahrscheinlichkeiten zweier sich gegenseitig ausschließender (disjunkter) Ereignisse addieren sich (3.3). Aus diesen wenigen Regeln folgt unmittelbar $P(\varnothing) = 0$ sowie $P(A^c) = 1 - P(A)$.

Wie wichtig diese Grundsätze sind, zeigt sich am besten an einem konkreten Beispiel: einem unfairen Würfel, dessen Einzelwahrscheinlichkeiten wir allein aus einigen wenigen Beobachtungen und den obigen Axiomen vollständig bestimmen können.

Beispiel: Der unfaire Würfel.

Ein sechsseitiger Würfel sei manipuliert. Der Ergebnisraum ist $\Omega = \{1, 2, 3, 4, 5, 6\}$, und für die Einzelwahrscheinlichkeiten schreiben wir kurz $p_i := P(\{i\})$. Bekannt sei aus Beobachtungen Folgendes:

- Die geraden Augenzahlen sind untereinander gleich wahrscheinlich, ebenso die ungeraden: $p_2 = p_4 = p_6$ und $p_1 = p_3 = p_5$.
- Eine gerade Augenzahl ist doppelt so wahrscheinlich wie eine ungerade: $p_2 = 2\,p_1$.

Gesucht sind alle sechs Einzelwahrscheinlichkeiten. Da die Ereignisse $\{1\}, \ldots, \{6\}$ paarweise disjunkt sind, liefert wiederholte Anwendung von Axiom (3.3) zusammen mit der Normierung (3.2)

$$p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = P(\Omega) = 1. \tag{3.4}$$

Einsetzen der beiden Bedingungen ($p_1 = p_3 = p_5$ und $p_2 = p_4 = p_6 = 2p_1$) ergibt

$$3\,p_1 + 3\,p_2 = 3\,p_1 + 6\,p_1 = 9\,p_1 = 1, \qquad \text{also} \qquad p_1 = \tfrac{1}{9}. \tag{3.5}$$

Damit folgt sofort

$$p_1 = p_3 = p_5 = \tfrac{1}{9}, \qquad p_2 = p_4 = p_6 = \tfrac{2}{9}. \tag{3.6}$$

Alle Werte sind nicht negativ (3.1) und summieren sich zu 1, sind also mit den Axiomen verträglich.

Aus diesen Grundgrößen lassen sich nun weitere Wahrscheinlichkeiten allein über die Axiome berechnen. Für das Ereignis $G = \{2, 4, 6\}$ („gerade Augenzahl“) etwa gilt nach (3.3)

$$P(G) = p_2 + p_4 + p_6 = \tfrac{2}{3}, \tag{3.7}$$

und für sein Komplement ohne weitere Rechnung

$$P(G^c) = 1 - P(G) = \tfrac{1}{3}. \tag{3.8}$$

3.2 Bedingte Wahrscheinlichkeiten

In der Praxis liegt oft bereits Vorwissen über ein Experiment vor. Die bedingte Wahrscheinlichkeit $P(A|B)$ berechnet die Wahrscheinlichkeit für ein Ereignis A unter der Voraussetzung, dass das Ereignis B bereits eingetreten ist. Das Wissen über B schränkt den ursprünglichen Ergebnisraum auf B ein.

- Definition:
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{für } P(B) > 0$$
- Lemma:
$$P(A \cap B) = P(A|B) \cdot P(B)$$
- Ereignisse A und B sind stochastisch unabhängig, wenn $P(A \cap B) = P(A) \cdot P(B)$ gilt. Daraus folgt $P(A|B) = P(A)$ für $P(B) > 0$).

3.2.1 Beispiel: Monty-Hall Prblem

Das Ziegenproblem oder Monty Hall Problem, benannt nach dem Moderator der US-Gameshow „Lets Make a Deal “, ist eines der bekanntesten Paradoxon der Stochastik. Es veranschaulicht auf verblüffende Weise, wie stark unsere intuitive Wahrscheinlichkeitsrechnung von der mathematischen Realität abweichen kann.

Ein Kandidat seht vor drei verschlossenen Türen. Hinter genau einer Tür befindet sich als Hauptgewinn ein Neuwagen, hinter den anderen beiden jeweils eine Ziege als Niete (deshalb der Name).

Der Kandidat wählt zunächst eine Tür. Diese bleibt vorerst geschlossen. Die Gewinnwahrscheinlichkeit liegt hier exakt bei $\frac{1}{3}$.

Da der Moderator genau weiß, hinter welcher Tür das Auto steht, öffnet er einer der beiden verbleibenden Türen, hinter der sich garantiert eine Ziege befindet.

Nun entsteht ein Dilemma, da der Moderator dem Kandidaten anbietet, seine Entscheidung zu ändern und noch auf die andere noch verschlossene Tür zu wechseln.

Darauf stellt sich die Frage; erhöht ein Wechsel die Chance auf den Gewinn, oder bleibt die Wahrscheinlichkeit gleich?

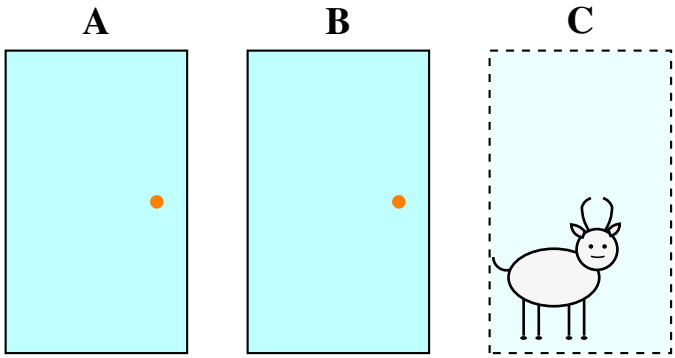
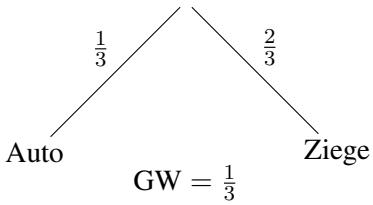


Abbildung 3.2: Zu Beginn stehen drei geschlossene Türen zur Auswahl; hinter genau einer davon befindet sich der Hauptgewinn, hinter den beiden anderen jeweils eine Ziege. Der Moderator öffnet daraufhin eine der beiden nicht gewählten Türen und zeigt eine Ziege, hier beispielhaft Tür C (eigene Darstellung).

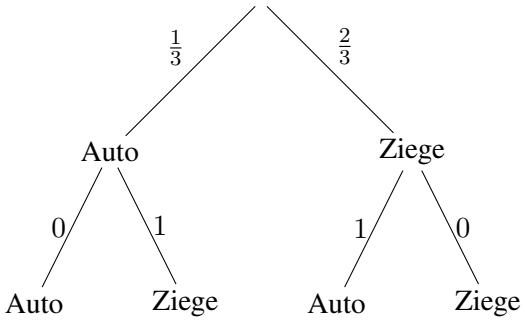
Gefühlsmäßig nehmen die meisten Menschen an, das nach dem Öffnen der ersten Ziegen-Tür eine 50:50 -Chance zwischen den zwei verbleibenden Türen besteht. Das ist jedoch ein Fehlschluss, da das Wissen des Moderators die Wahrscheinlichkeiten beeinflusst. Berühmt wurde das Problem 1990 durch Marilyn vos Savant, die als Frau mit dem weltweit höchsten gemessenen IQ ins Guinness-Buch der Rekorde einging. In ihrer Magazin-Kolumne erklärte sie mathematisch korrekt: Man sollte die Tür wechseln! Durch den Wechsel verdoppelt sich die Gewinnchance von $\frac{1}{3}$ auf $\frac{2}{3}$. Ihre Lösung löste damals eine riesige Welle der Empörung aus. Hunderte Mathematik-Professoren schrieben ihr Briefe und behaupteten sie liege falsch. Erst spätere Simulationen und stochastische Belege gaben ihr zweifelsfrei recht.

Strategie „Bleiben“



(GW = Gewinnwahrscheinlichkeit Bleiben)

Strategie „Wechseln“



(GW = Gewinnwahrscheinlichkeit Wechseln)

$$\text{GW} = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$$

Warum funktioniert das?: Beim ersten Tipp liegt man zu $\frac{2}{3}$ falsch. Wenn man eine Ziege gewählt hat, muss der Moderator die andere Ziege zeigen. Die verbleibende Tür enthält dann automatisch das Auto. Die Strategie

"Wechseln"gewinnt also immer nur dann, wenn man sich am Anfang falsch entschieden hat und das ist in zwei von drei Fällen der Fall.

STATISTISCHE GRUNDLAGEN

Satz von Bayes, Modellbildung und Bayes- vs. frequentistische Interpretation der Statistik

von Elio, Simon

4.1 Modellbildung

Physik lebt vom Wechselspiel zweier Dinge: einem (mathematischen) *Modell* der Welt – einem physikalischen Gesetz mit ein paar offenen Parametern – und *Daten*, also Messwerten aus dem Labor, von Teleskopen oder anderen Messinstrumenten. Die Wahrscheinlichkeitstheorie verknüpft beide, indem sie vorwärts denkt: Sind Modell und Parameter gegeben, sagt sie vorher, welche Daten wir mit welcher Wahrscheinlichkeit beobachten sollten. Die Statistik geht denselben Weg rückwärts: Ausgehend von den gemessenen Daten versucht sie herauszufinden, welches Modell bzw. welche Parameter eines Modells am besten zu ihnen passt.

Damit ein statistischer Rückschluss gelingen kann, müssen wir zunächst klären, was wir unter einem Modell in diesem Sinne überhaupt verstehen – also womit wir arbeiten und worauf wir unsere Daten beziehen wollen. Der Satz von Bayes hat dabei stillschweigend vorausgesetzt, dass P bereits eine feste, bekannte Größe ist – in der Praxis kennen wir höchstens die *Form* von P, mit einigen offenen Parametern, die erst aus Daten bestimmt werden sollen. Genau das erfasst der folgende Begriff:

Definition 4.1. Sei Ω die Menge aller möglichen Ausgänge eines Experiments und Σ eine geeignete Ereignismenge. Ein **statistisches Modell** ist ein Tupel

$$(\Omega, \Sigma, (P_\theta)_{\theta\in\Theta}) \ ,$$

wobei $(P_\theta)_{\theta\in\Theta} : \Sigma \to [0,1]$ eine Familie von Wahrscheinlichkeitsverteilungen *auf derselben Grundmenge* ist, indiziert durch einen Parameter (oder Parametervektor) θ aus einem Parameterraum Θ.

Für jeden festen Wert von θ ist $(\Omega, \Sigma, P_\theta)$ also ein Wahrscheinlichkeitsraum; welcher Wert von θ jedoch der „richtige“ ist, bleibt zunächst offen und soll aus Daten erschlossen werden. Entscheidend ist dabei, dass *alle* P_θ auf derselben Grundmenge Ω leben: Ein statistisches Modell legt also fest, *welche* Daten überhaupt beobachtet werden können, und modelliert, *mit welcher Wahrscheinlichkeit(sverteilung)* sie auftreten könnten. Ein Beispiel macht das greifbar:

Abbildung 4.1: 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.

Beispiel 4.1 (Die unfaire Münze)**.** Wir werfen eine Münze $n = 14$-mal und zählen die Anzahl k der Würfe, bei denen „Kopf“ erscheint. Die möglichen Ergebnisse sind $\Omega = \{0, 1, \ldots, 14\}$. Ist $p \in [0, 1]$ die (unbekannte) Kopfwahrscheinlichkeit eines einzelnen Wurfs – bei einer fairen Münze wäre $p = \frac{1}{2}$, doch a priori könnte die Münze auch unfair sein –, so folgt k einer Binomialverteilung

$$P_p(k) = \binom{14}{k} p^k (1-p)^{14-k} . \tag{4.1}$$

Das statistische Modell ist damit die Familie $(P_p)_{p\in[0,1]}$: Parameterraum $\Theta = [0, 1]$, und für jedes feste p liegt eine vollständig bestimmte Wahrscheinlichkeitsverteilung über die möglichen Kopfzahlen $k \in \Omega$ vor. Werfen wir die Münze tatsächlich 14-mal und beobachten z. B. $k = 10$-mal Kopf, so beginnt genau hier die Arbeit der Statistik: Welcher Wert von p passt am besten zu dieser Beobachtung?

Ähnliche statistische Modelle begegnen einem in der Physik ständig: 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 λ (Parameterraum $\Theta = (0, \infty)$); eine physikalische Messgröße x wird oft durch eine Normalverteilung $P_{(\mu,\sigma)}(x) = \mathcal{N}(x; \mu, \sigma^2)$ mit unbekanntem Mittelwert μ und unbekannter Streuung σ modelliert. Hier ist $\theta = (\mu, \sigma)$ bereits ein zweidimensionaler Parametervektor – der Parameterraum Θ muss also keineswegs ein einfaches Intervall sein.

Das Ziel der Statistik ist es nun, ausgehend von beobachteten Daten auf den plausibelsten Parameter θ zurückzuschließen – in unserem Münzbeispiel also auf den plausibelsten Wert von p. Wie genau man das anstellt, hängt jedoch entscheidend davon ab, welche der beiden folgenden Philosophien man vertritt.

4.2 Satz von Bayes

Ein zentrales Hilfsmittel, um von gegebenen Daten D auf die Modellparameter θ zurückzuschließen, ist der sogenannte *Satz von Bayes*.

Theorem 4.1. *Für zwei Ereignisse A, B mit $P(B) > 0$ gilt allgemein*

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

Setzt man speziell $A = \theta$ und $B = D$, so folgt für $P(D) > 0$

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{P(D)} . \tag{4.2}$$

Der Beweis benutzt nur die elementare Definition der bedingten Wahrscheinlichkeit:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{P(A\cap B)}{P(A)}\, P(A)}{P(B)} = \frac{P(B \mid A)\, P(A)}{P(B)}. \tag{4.3}$$

In der Gleichung (4.3) steht θ für die Modellparameter und D für die beobachteten Daten. Die Bayessche Statistik bezeichnet $P(\theta)$ als den *Prior*, das Vorwissen, welches bereits ohne die vorhandenen Daten erlangt wurde. Die *Likelihood* $P(D \mid \theta)$ gibt an, mit welcher Wahrscheinlichkeit die Daten D auftreten, wenn der Parameter θ gegeben ist. Analog zum Prior ist der *Posterior* $P(\theta \mid D)$ das erweiterte Wissen über θ unter Einbezug der Daten. Die *Evidenz* $P(D)$ wird häufig nicht explizit berechnet, da sie nicht explizit von den Modellparametern abhängt.

4.3 Bayessche und frequentistische Interpretation der Statistik

In der Statistik ist bis heute umstritten, was eine Wahrscheinlichkeit eigentlich bedeutet – und je nachdem, wie man diese Frage beantwortet, zieht man am Ende unterschiedliche Schlüsse aus denselben Daten.

In der klassischen frequentistischen Interpretation wird die Wahrscheinlichkeit eines Ereignisses als der Grenzwert seiner relativen Häufigkeit in einer (gedanklich) beliebig oft wiederholbaren Versuchsreihe verstanden. Dieses Wahrscheinlichkeitsmodell ist nach wie vor weit verbreitet. Das Bayessche Modell der Statistik hingegen versteht Wahrscheinlichkeiten anders, nämlich – motiviert durch den Satz von Bayes – als Grad der Überzeugung („degree of belief“), der auch für unwiederholbare Ereignisse berechnet werden kann.

Am Münzbeispiel aus Beispiel 4.1 (wir hatten $n = 14$ Würfe und $k = 10$-mal Kopf beobachtet) lässt sich das gut zeigen.

Fasst man Wahrscheinlichkeit frequentistisch auf, sucht man zunächst einen Punktschätzer für p, etwa den Maximum-Likelihood-Schätzer

$$\hat{p} = k/n = 10/14 \approx 0.71 \,,$$

und ergänzt ihn um ein Konfidenzintervall – hier näherungsweise $[0.48;\, 0.95]$ zum $95\,\%$-Niveau. Die Bedeutung dieses Intervalls ist auf den ersten Blick gewöhnungsbedürftig: p selbst liegt nicht mit $95\,\%$ Wahrscheinlichkeit darin; denn p ist ja fest, wenn auch unbekannt. Gemeint ist vielmehr, dass ein so berechnetes Intervall den wahren Wert von p in $95\,\%$ aller Fälle enthält, wenn man den Versuch immer wieder durchführt.

Der bayessche Blick beginnt dagegen mit einem Prior für p – zum Beispiel der Gleichverteilung $P(p) = 1$ auf $[0, 1]$, die ausdrückt, dass uns vor der Messung jeder Wert gleich plausibel erscheint. Über den Satz von Bayes ergibt sich daraus eine vollständige Posterior-Verteilung für p, eine Beta-Verteilung, aus der sich sowohl ein Erwartungswert

$$\mathbb{E}[p \mid D] = 11/16 \approx 0.69$$

als auch ein Kredibilitätsintervall von etwa $[0.47;\, 0.91]$ ablesen lassen. Die Interpretation ist hier direkter: Gegeben die Daten liegt p tatsächlich mit $95\,\%$ Wahrscheinlichkeit in diesem Bereich.

Interessant ist, dass beide Rechnungen auf recht ähnliche Zahlen kommen – das liegt am flachen Prior und wäre bei wenigen Daten oder einem informativeren Prior nicht mehr so. Der eigentliche Unterschied liegt also nicht in den Zahlen, sondern darin, wovon die Aussage handelt: einmal von den Daten bei angenommenem p, einmal von p selbst, gegeben die beobachteten Daten.

Keine der beiden Sichtweisen ist dabei grundsätzlich vorzuziehen. Die frequentistische Statistik kommt ohne subjektive Annahmen aus und passt gut zu Experimenten, die sich tatsächlich beliebig oft wiederholen lassen, etwa vielen gleichartigen Messungen im Labor. Die bayessche Statistik erlaubt es dagegen, Vorwissen einzubeziehen, und funktioniert auch dort, wo sich ein Ereignis nicht wiederholen lässt: wie wahrscheinlich ist zum Beispiel ein bestimmtes kosmologisches Modell, wenn es nur das eine Universum gibt, das wir beobachten können? Dafür nimmt man in Kauf, dass die Wahl des Priors das Ergebnis mitbestimmt, was oft als Kritikpunkt gegen die bayessche Methode vorgebracht wird. Tabelle 4.1 fasst die wichtigsten Gegenüberstellungen zusammen.

	Frequentistisch	**Bayesianisch**
Wahrscheinlichkeit ist	relative Häufigkeit	Grad der Überzeugung
Parameter θ	fest, unbekannt	selbst zufällig verteilt
Ergebnis	Punktschätzer $\hat{\theta}$	vollständige Posterior $p(\theta \mid D)$
Unsicherheit	Konfidenzintervall	Kredibilitätsintervall
Vorwissen	fließt nicht formal ein	explizit über den Prior

Tabelle 4.1: Gegenüberstellung der frequentistischen und der bayesschen Sichtweise auf Statistik.

ZENTRALER GRENZWERTSATZ

von Angelina, Valérie

5.1 Erwartungswert und Varianz

Eine wichtige Größe von diskreten Verteilungen, die für den Zentralen Grenzwertsatz benötigt werden, sind der Erwartungswert und die Varianz, die angeben, welchen Wert die Zufallsvariable am wahrscheinlichsten annimmt und wie viel sie um diesen Wert streut.

Definition 5.1. Der Erwartungswert einer diskreten Verteilung X ist gegeben durch

$$\langle X \rangle = \sum_{i=1}^{n} x_i \cdot P(X = x_i) \; .$$

Im Kontinuierlichen geht diese Summe über in das Integral

$$\langle X \rangle = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \; .$$

Die Varianz ist die quadrierte Standardabweichung und somit durch

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2$$

gegeben.

5.2 Beispiele für diskrete Verteilungen

Nicht nur für endlich viele Ergebnisse (wie beim Würfeln), sondern auch für unbeschränkt viele diskrete Werte ($k \in \{0, 1, 2, \dots\}$) existieren bewährte Standardverteilungen. Im Folgenden werden die Binomial- und die Poissonverteilung thematisiert.

5.2.1 Binomialverteilung

Führt man n unabhängige Versuchsdurchläufe mit konstanter Erfolgswahrscheinlichkeit p durch, ist die Anzahl der Treffer k binomialverteilt. Dabei kann das Experiment nur zwei mögliche Ergebnisse annehmen – einen Treffer oder eine Niete. Die Binomialverteilung wird bei n-mal *unabhängig* durchführbaren Experimenten angewandt, bei denen die Trefferwahrscheinlichkeit p konstant ist. Die Gesamtanzahl der Durchführungen wird somit durch n beschrieben. Eine Zufallsvariable $X \sim \mathrm{Bin}(n, p)$ ist binomialverteilt, wenn

$$P(X = k) = B(k; n, p) := \binom{n}{k} p^k (1-p)^{n-k} \; . \tag{5.1}$$

Hier beschreibt der Binomialkoeffizient $\binom{n}{k}$ (gespr.: „k aus n“ oder „n über k“) die Anzahl an Möglichkeiten, k Treffer aus einer Menge von n Elementen zu ziehen. Der Binomialkoeffizient ist dabei definiert über

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} \; ,$$

wobei $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$ die Fakultät von n bezeichnet.

In Gleichung (5.1) stellt p^k die Wahrscheinlichkeit für genau k erfolgreiche Versuche und $(1-p)^{n-k}$ die Wahrscheinlichkeit, dass die restlichen $n-k$ Versuche (unter Verwendung der Gegenwahrscheinlichkeit $1-p$) nicht erfolgreich sind dar.

Eine beispielhafte Binomialverteilung ist in Abbildung 5.1 zu sehen.

Weiterhin kann bei Binomialverteilungen der sog. Erwartungswert $\langle k \rangle = np$ ausgerechnet werden, welcher den Mittelwert bei häufiger Wiederholung des Experiments angibt. Bei einer fairen Münze ist der Erwartungswert beispielsweise $\langle k \rangle = 0.5$, denn beide Seiten der Münze haben die gleiche Wahrscheinlichkeit.

Zusätzlich kann auch die mittlere quadratische Abweichung – Varianz – aller Ergebnisse zum Erwartungswert mithilfe von $\sigma_k^2 = np(1-p)$ ausgerechnet werden. Damit wird die Streuung der Werte zum Erwartungswert beschrieben.

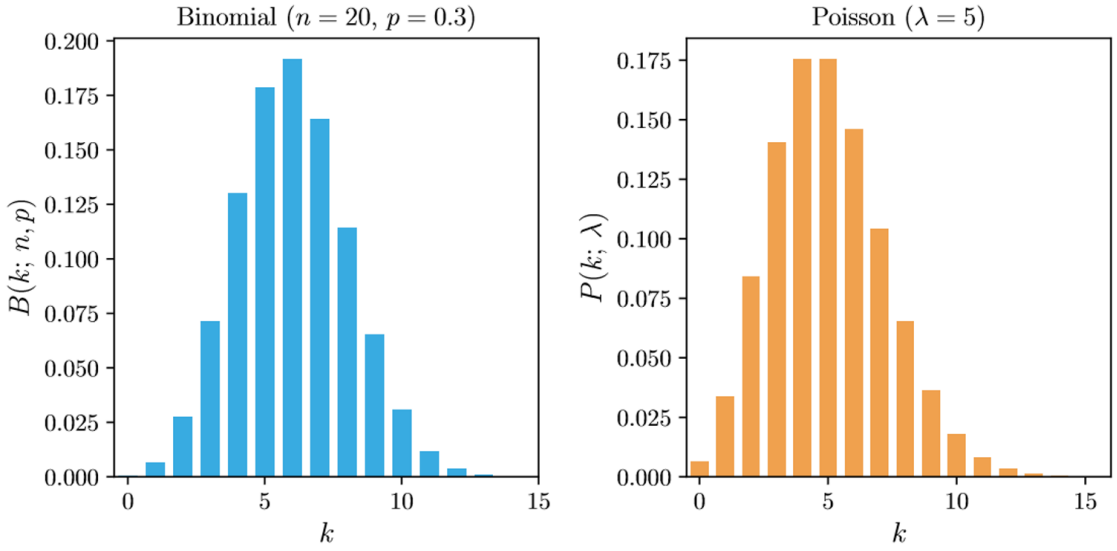


Abbildung 5.1: *Links*: Binomialverteilung $B(k; n = 20, p = 0.3)$. *Rechts*: Poissonverteilung $P(k; \lambda = 5)$ – ihr Grenzfall für seltene Ereignisse mit konstanter Rate. In beiden Histogrammen entspricht die Balkenhöhe der Wahrscheinlichkeit für die jeweilige Anzahl an k Treffern.

5.2.2 Poissonverteilung

Die Poissonverteilung ergibt sich als Grenzfall aus der Binomialverteilung. Dabei geht die Binomialverteilung für $n \to \infty$ und $p \to 0$ bei festem $\lambda = np$ in die Poissonverteilung über:

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

Allgemein wird die Poisson-Verteilung auch als Verteilung der seltenen Ereignisse bezeichnet, da die Trefferwahrscheinlichkeit p gegen Null und die Gesamtanzahl der Durchführungen n gegen Unendlich konvergieren. Dabei sind der Erwartungswert $\langle k \rangle = \lambda$ und die Varianz $\sigma_k^2 = \lambda$ gleich groß.

In Abbildung 5.1 ist eine beispielhafte Poissonverteilung dargestellt.

5.3 Zentraler Grenzwertsatz und Normalverteilung

Nun wird betrachtet, was für eine Verteilung sich ergibt, wenn man ein Zufallsexperiment mit gleichbleibender Erfolgswahrscheinlichkeit immer öfter wiederholt oder, anders ausgedrückt, wie sich die Binomialverteilung für $n \to \infty$ verhält.

Es ist zu erkennen, dass der Mittelwert und die Varianz ansteigen, wie es für $\langle k \rangle = np$ und $\sigma_{\mathrm{k}}^2 = np(1-p)$ für wachsendes n zu erwarten ist. Daher wird der Graph um $k - np$ verschoben, sodass das Maximum im Ursprung liegt und die x-Achse so skaliert wird, dass die Varianz 1 beträgt, wofür durch die Wurzel der Varianz, auch genannt Standardabweichung $\sqrt{np(1-p)}$, geteilt wird. Wird nun $z = \frac{k-np}{\sqrt{np(1-p)}}$ geplottet, ist zu erkennen, dass sich die Verteilung einer Glockenkurve annähert, die durch die Funktion $\mathcal{N}(x;0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ beschrieben wird. Dasselbe lässt sich nicht nur für eine Zufallsgröße, sondern auch für die Summe beliebig vieler Zufallsgrößen $X_1, X_2, \ldots$ zeigen, sofern sie unabhängig und identisch verteilt sind und die Varianz der Zufallsgrößen $\sigma^2 = \mathrm{Var}(X_i)$ jeweils endlich ist, d. h. $\sigma^2 < \infty$. Diese Aussage ist als *Zentraler Grenzwertsatz* (ZGS) bekannt:

Theorem 5.1 (Zentraler Grenzwertsatz)**.** *Seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsgrößen mit Erwartungswert $\langle X_i \rangle = \mu$ und endlicher Varianz $\sigma^2 = \mathrm{Var}(X_i) < \infty$. Für $S_n := \sum_{i=1}^n X_i$ und $Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$ gilt dann für jedes $a \in \mathbb{R}$*

$$\lim_{n\to\infty} P(Z_n \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \,.$$

Im Folgenden soll nachvollzogen werden, warum der Zentrale Grenzwertsatz gilt. Dazu wird die Verteilung um die Erwartungswerte der verschiedenen Potenzen der Zufallsvariablen charakterisiert, diese nennen wir *Momente*, wobei das k-te Moment durch $m_k := \langle Z_n^k \rangle$ gegeben ist. All diese Momente können von einer *momenterzeugenden Funktion* zusammengefasst werden, deren k-te Ableitung an der Stelle 0 jeweils als Wert das k-te Moment hat. Folglich gilt $M_{Z_n} := \langle e^{Z_n t} \rangle$. Die Idee ist nun, zu zeigen, dass die momenterzeugende Funktion von Z_n gegen die momenterzeugende Funktion von $\mathcal{N}(x;0,1)$ konvergiert. Daraufhin gilt es noch zu zeigen, dass aus der Konvergenz der momenterzeugenden Funktion auch Konvergenz der Verteilung folgt. Zuerst ermitteln wir also M_{Z_n}. Wir verschieben die Verteilung wenn nötig in den Ursprung, sodass $\mu = 0$ gilt und daher $Z_n := \frac{S_n}{\sigma\sqrt{n}}$. Es gilt nach den Potenzgesetzen:

$$M_{Z_n}(t) = \left\langle e^{Z_n t} \right\rangle = \left\langle e^{\frac{S_n}{\sigma\sqrt{n}} t} \right\rangle = M_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Nun wird $S_n = \sum_{i=1}^n X_i$ eingesetzt und verwendet, dass aufgrund der Potenzgesetze $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ und somit nach Induktion $M_{S_n}(t) = [M_X(t)]^n$ gilt. Es ergibt sich:

$$M_{Z_n}(t) = \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Im nächsten Schritt wird die Taylorentwicklung von $M_X\left(t/(\sigma\sqrt{n})\right)$ um die Stelle 0 betrachtet, wobei wir $e^x = 1 + x + \frac{x^2}{2} + \mathcal{O}(x^3)$ verwenden. Es ergibt sich

$$M_X(t) = 1 + t\langle X \rangle + \frac{t^2}{2}\langle X^2 \rangle + \mathcal{O}(t^3) \,.$$

Mit $\langle X \rangle = 0$ und der Definition der Varianz $\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2$ ergibt sich für $M_{Z_n}(t)$ durch Einsetzen von $t/(\sigma\sqrt{n})$ und Kürzen mit $\langle X^2 \rangle = \sigma^2$

$$M_{Z_n}(t) = \left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{\frac{-3}{2}})\right]^n$$

Als nächstes wird die Gleichung logarithmiert und die Taylor-Entwicklung $\ln(1+x) = x + \mathcal{O}(x^2)$ um $x = 0$ verwendet.

$$\ln[M_{Z_n}(t)] = n \ln\left[1 + \frac{t^2}{2n} + \mathcal{O}(n^{\frac{-3}{2}})\right] = n\left[\frac{t^2}{2n} + \mathcal{O}(n^{-\frac{3}{2}}) + \mathcal{O}(n^{-2})\right]$$

Für $n \to \infty$ gehen die Restterme gegen 0 und wendet man auf beiden Seiten die Exponentialfunktion an $M_{Z_n} \to \exp \frac{t^2}{2}$. Nun soll gezeigt werden, dass $\mathcal{N}(x; 0, 1)$ genau diese momenterzeugende Funktion hat. Dabei ist die momenterzeugende Funktion einer Verteilung deren Wahrscheinlichkeitsdichte durch die Funktion $f(x)$ bestimmt ist, gegeben durch $M_N(x) = \int_{-\infty}^{\infty} e^{tx} f(x) \mathrm{d}x$ wobei eine normalverteilte Zufallsvariable ist, was durch die Notation $N \sim \mathcal{N}(0, 1)$ ausgedrückt wird. Der im Exponenten auftretende Term wird quadratisch ergänzt $tx - \frac{1}{2}x^2 = -\frac{1}{2}(x - t)^2 + \frac{t^2}{2}$, sodass der Term im Integral zu einer verschobenen Normalverteilung wird, über die integriert sich der Wert 1 ergibt. Folglich gilt

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(tx - \frac{x^2}{2}\right) \mathrm{d}x \\ &= \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x - t)^2\right) \mathrm{d}x = \exp\left(\frac{t^2}{2}\right) \end{aligned}$$

Damit ist gezeigt, dass die momenterzeugende Funktion von Z_n, wenn Z_n entsprechend normiert und verschoben wurde, gegen die momenterzeugende Funktion der Normalverteilung $\mathcal{N}(0, 1)$ konvergiert. Es bleibt zu zeigen, dass daraus die Konvergenz der Verteilung folgt. Dazu schreiben wir die beiden momentenerzeugenden Funktionen mithilfe einer Taylor-Entwicklung als Potenzreihen. Es gilt, dass bei Potenzreihen mit gemeinsamen Konvergenzradius Ableitung und Grenzwert vertauscht werden dürfen. Aus $M_{Z_n}(t) \to M(t)$ folgt also $M_{Z_n}^{(k)}(t) \to M^{(k)}(t)$, was gleichbedeutend zu $\langle Z_n^k \rangle \to \langle N^k \rangle$ ist. Die beiden Verteilungen haben also dieselben Momente, solange die momentanerzeugende Funktion zumindest in einer Umgebung von 0 existiert, was die Grundannahme für diesen Beweis ist.

METROPOLIS-HASTINGS-ALGORITHMUS

von Gabriel, Geri

6.1 Fluch der Dimensionen

Der Metropolis-Hastings-Algorithmus ist ein Sampling-Algorithmus, der für ein vorgegebenes Modell Stichproben aus der Posterior-Verteilung zieht, statt sie direkt zu berechnen. Der Algorithmus sampelt anhand einer vorgegebenen Posterior-Funktion alle Parameter. Dieser Algorithmus wird verwendet, da ein systematisches Iterieren aller möglichen Werte für alle Parameter nicht vorteilhaft ist. Der Grund dafür besteht darin, dass die Anzahl der Evaluationen Ω^m beträgt, wobei m die Anzahl der Parameter (die Dimension) und Ω die Anzahl der Werte ist, die je ein Parameter annimmt. Mit Ω^m lässt sich der Rechenaufwand nur bei diskreten Verteilungen exakt angeben; bei kontinuierlichen Verteilungen ist ein Ausprobieren aller Möglichkeiten sowieso nicht anwendbar.

6.2 Algorithmus

6.2.1 Allgemein

Die Motivation hinter dem Metropolis-Hastings-Algorithmus ist es, den besten Parametersatz zu finden, damit das Modell bestmöglich zu den Daten passt. Dabei sucht der Algorithmus die Parameter nicht durch vollständiges Ausprobieren aller Möglichkeiten, sondern über einen randomisierten Suchprozess, bei dem jeder vorgeschlagene Schritt angenommen oder abgelehnt wird.

Dafür benötigt man den Begriff des Parameterraums: Der Parameterraum ist ein Raum, der aus allen Parametersätzen besteht. So sei ein Modell gegeben durch $ax^3 + bx^2 + cx + d$, dann sind die Parameter $\theta = (a, b, c, d)$. In diesem Fall hat man einen 4-dimensionalen Raum, wobei jeder Parameter eine Achse ist. Also beschreibt jedes θ einen Punkt in diesem Raum. Die Menge aller Punkte in diesem Raum heißt Θ, somit ist $\theta \in \Theta$. Der Algorithmus fängt also bei einem Punkt θ_{curr} an, wählt einen neuen zufälligen Punkt θ_{neu} in der Nähe mithilfe einer Normalverteilung und vergleicht, wie gut θ_{curr} und θ_{neu} jeweils zu den Daten passen (genauer: vergleicht ihre Posterior-Wahrscheinlichkeiten) und entscheidet dann, ob er zu θ_{neu} weitergeht oder stehen bleibt. Um den Posterior $P(\theta|D)$ für ein gegebenes $\theta \in \Theta$, bedingt auf die beobachteten Daten D, zu bestimmen, benutzt man den Satz von Bayes:

$$P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{P(D)} \,, \tag{6.1}$$

vergleiche auch das Kapitel zu den Grundlagen der Statistik. Angewandt auf den Algorithmus stellt D die Daten dar. Die Likelihood ($P(D|\theta)$) gibt an, wie gut die Daten zu einem gegebenen θ passen, der Prior ($P(\theta)$), wie wahrscheinlich dieses θ aufgrund von Vorwissen ist, und der Posterior ($P(\theta|D)$), wie gut das Modell mit diesem Parametersatz insgesamt zu den Daten passt – darüber bestimmt man also die Qualität des jeweiligen Parametersatzes.

6.2.2 Auswahlverfahren

Um entscheiden zu können, ob θ_{curr} oder θ_{neu} besser ist, schaut der Algorithmus zunächst, ob der Posterior von θ_{neu} größer ist als der von θ_{curr}:

$$P(\theta_{\text{neu}}|D) \geq P(\theta_{\text{curr}}|D) \tag{6.2}$$

Falls Gleichung (6.2) gilt, wird θ_{neu} als neues θ_{curr} übernommen; andernfalls wird θ_{neu} nur mit einer Wahrscheinlichkeit von α, der Standard-MH-Akzeptanzrate, übernommen. Somit ist α durch folgende Gleichung

beschrieben:

$$\alpha = \min\left(1, \frac{P(\theta_{\text{neu}} \mid D)}{P(\theta_{\text{curr}} \mid D)}\right) \tag{6.3}$$

Das ermöglicht eine freiere Bewegung im Parameterraum: Würde man immer nur zum bestmöglichen Punkt gehen, bliebe man eventuell in einem lokalen Maximum, ohne dabei das globale Maximum zu finden. Um algorithmisch ein Ereignis zu modellieren, das mit einer bestimmten Wahrscheinlichkeit passiert, nimmt man eine zufällig gleichverteilte Zahl u im Intervall $[0, 1)$ und prüft, ob sie kleiner als die Wahrscheinlichkeit α ist.

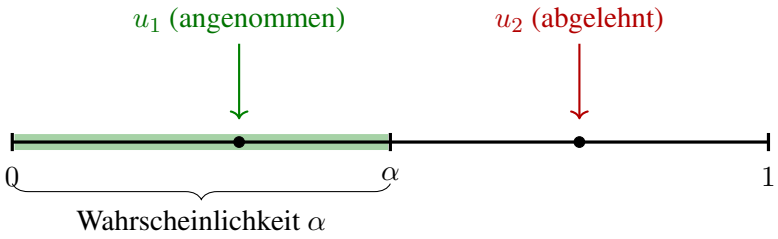


Abbildung 6.1: Veranschaulichung der Bernoulli-Entscheidung: Eine gleichverteilte Zufallszahl $u \in [0, 1)$ wird mit α verglichen. Landet u im Intervall $[0, \alpha)$ (grün, Beispiel u_1), wird der Vorschlag angenommen; landet u im Intervall $[\alpha, 1)$ (rot, Beispiel u_2), wird er abgelehnt. Da $[0, \alpha)$ genau die Länge α hat, passiert Ersteres mit Wahrscheinlichkeit α.

Ein Problem ist, dass die Posteriors sehr klein werden, weshalb man die Gleichung (6.3) logarithmiert. Der Grund dafür ist, dass der Logarithmus auf dem Intervall $(0, 1]$ Werte im Intervall $[-\infty, 0]$ annimmt, wodurch sich die sonst sehr kleinen Zahlen deutlich besser vergleichen lassen. Wenn man den Schritt ausführt, ergibt sich:

$$\log(u) < \log(P(\theta_{\text{neu}}|D)) - \log(P(\theta_{\text{curr}}|D)) \tag{6.4}$$

Damit hat man aber gleichzeitig beide Auswahlkriterien erfüllt, denn wenn $P(\theta_{\text{neu}}|D) \geq P(\theta_{\text{curr}}|D)$, dann ist α größer als 1 (Gleichung (6.3)), und da u nur Zahlen im Intervall $[0, 1)$ annehmen kann, wird u immer kleiner als α sein.

6.2.3 Vereinfachung der Gleichung

Wenn man den Logarithmus auf Gleichung (6.1) anwendet, erhält man folgende Gleichung:

$$\log(P(\theta|D)) = \log(P(D|\theta)) + \log(P(\theta)) - \log(P(D)). \tag{6.5}$$

Da man beide Terme zur Auswahlentscheidung voneinander abzieht (Gleichung (6.4)), fallen alle Konstanten heraus, sodass man sie nicht berechnen muss. $\log(P(D))$ ist eine solche Konstante, weswegen man ihn einfach aus der Gleichung ignorieren kann. Ähnlich verhält es sich beim Prior, wobei man zwischen zwei Fällen unterscheiden muss: Fall 1: Man hat kein Vorwissen über die Parameter, sodass sie alle gleichverteilt sind. Damit ist die Wahrscheinlichkeit unabhängig von der Zahl, mit anderen Worten: Der Prior wäre auch eine Konstante und man kann ihn ignorieren.

$$\log(P(\theta|D)) = \log(P(D|\theta)) \tag{6.6}$$

Fall 2: Man hat Vorwissen über einen oder mehrere Parameter. In diesem Fall sind diese Parameter nicht mehr gleichverteilt und man kann den Prior nicht mehr vernachlässigen. Aber auch in diesem Fall kann man ggf. eine Vereinfachung durchführen, denn in vielen Fällen gilt für den Prior, dass er einfach das Produkt der Wahrscheinlichkeiten der Parameter ist. Also:

$$P(\theta) = P(\theta_1) \cdot P(\theta_2) \cdot \cdots \cdot P(\theta_n). \tag{6.7}$$

Nun teilt man diesen Ausdruck in zwei Faktoren auf, indem man alle gleichverteilten Parameter miteinander multipliziert und alle ungleichverteilten Parameter ebenso, wie folgt:

$$P(\theta) = P(\theta_{\text{gleich}}) \cdot P(\theta_{\text{ungleich}}) \tag{6.8}$$

Nun logarithmiert man die Gleichung und erhält:

$$\log(P(\theta)) = \log(P(\theta_{\text{gleich}})) + \log(P(\theta_{\text{ungleich}})). \tag{6.9}$$

Auch hier hat man wieder lediglich eine Summe von Log-Wahrscheinlichkeiten, und wegen der Gleichungen (6.4) und (6.6) gelangt man auf die folgende Gleichung:

$$\log(p(\theta|D)) = \log(P(D|\theta)) + \log(P(\theta_{\text{ungleich}})). \tag{6.10}$$

6.2.4 Normalverteilung zum Ausrechnen des Posteriors

Jetzt ist aber noch offen, wie man $\log(P(D|\theta))$ und $\log(P(\theta_{\text{ungleich}}))$ ausrechnet. Dafür benutzt man die in Kapitel 5 genannte Normalverteilung, um die Likelihood und den Prior zu bestimmen. Für die Berechnung der Likelihoods muss man sie zunächst definieren. Die Likelihood gibt die Wahrscheinlichkeit an, die Daten zu beobachten, unter der Annahme, dass das Modell stimmt. Um die Wahrscheinlichkeit eines einzelnen Datenpunkts zu berechnen, geht man davon aus, dass die Abweichungen normalverteilt sind: Man definiert den Erwartungswert μ als den vom Modell vorhergesagten Wert und die Standardabweichung σ als die aus dem Versuch bestimmbare Unsicherheit dieses Werts. Außerdem benötigt man eine Variable d_i, die den i-ten gemessenen Datenpunkt repräsentiert. Somit kann man die Likelihood eines Datenpunkts wie folgt berechnen

$$P(d_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(d_i-\mu)^2}{2\sigma^2}} \tag{6.11}$$

Um die Likelihood für alle Datenpunkte zu berechnen, rechnet man sie für jeden Datenpunkt einzeln aus und multipliziert die Ergebnisse miteinander.

$$P(D|\theta) = \prod_{i=1}^{n} \left(\frac{1}{\sigma_i\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \frac{(d_i-\mu_i)^2}{\sigma_i^2} \right] \right) , \tag{6.12}$$

wobei n die Anzahl aller Datenpunkte bezeichnet. Vereinfacht wird es zu:

$$P(D|\theta) = \prod_{i=1}^{n} \left(\frac{1}{\sigma_i\sqrt{2\pi}} \right) \cdot \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{(d_i-\mu_i)^2}{\sigma_i^2} \right) \right] \tag{6.13}$$

Logarithmiert ergibt sich:

$$\log(P(D|\theta)) = \sum_{i=1}^{n} \log\left(\frac{1}{\sigma_i\sqrt{2\pi}} \right) + \left[-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{(d_i-\mu_i)^2}{\sigma_i^2} \right) \right] \tag{6.14}$$

Falls die Varianz σ^2 konstant für jeden Datenpunkt ist, kann man die Gleichung weiter vereinfachen:

$$\log(P(D|\theta)) = n \cdot \log\left(\frac{1}{\sigma\sqrt{2\pi}} \right) + \left[-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{(d_i-\mu_i)^2}{\sigma^2} \right) \right] \tag{6.15}$$

Jetzt benötigt man aber noch die Wahrscheinlichkeit des ungleichverteilten Priors: Wie schon in Gleichung (6.8) ist der relevante Teil des Priors ein Produkt aus allen Wahrscheinlichkeiten der relevanten Parameter. Es bietet sich an die Normalverteilung zu nutzen, wobei dann der Erwartungswert μ die Zahl ist, die man aufgrund des Vorwissens für den Parameter erwartet, und das σ die Unsicherheit dieses gewählten Erwartungswertes

beschreibt. Der Wert x entspricht in diesem Fall dem Parameter θ_i. Aus diesen Vorgaben und Gleichung (6.15) ergibt sich, dass der relevante Teil des Priors gegeben ist durch:

$$\log(P(\theta)) = \sum_{i=1}^{m} \log\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) + \left[-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{(\theta_i-\mu_i)^2}{\sigma_i^2}\right)\right] \tag{6.16}$$

Jetzt kann man den Posterior mithilfe der Gleichung (6.5) berechnen. Zusammenfassend funktioniert der Metropolis-Hastings-Algorithmus wie folgt:

Algorithm 1 Metropolis-Hastings-Algorithmus

Eingabe: Startpunkt $\theta_{\text{curr}} \in \Theta$, Anzahl Iterationen N, Schrittweite s
Ausgabe: Folge von Samples $\theta^{(1)}, \ldots, \theta^{(N)}$

for $k = 1, \ldots, N$ **do**
  Ziehe $\theta_{\text{neu}} \sim \mathcal{N}(\theta_{\text{curr}}, s^2 I)$ ▷ Normalverteilung um θ_{curr}
  Berechne $\log(P(\theta_{\text{neu}}|D))$ und $\log(P(\theta_{\text{curr}}|D))$ ▷ Gleichung (6.5)
  Ziehe $u \sim$ Gleichverteilung$[0, 1)$
  if $\log(u) < \log(P(\theta_{\text{neu}}|D)) - \log(P(\theta_{\text{curr}}|D))$ **then** ▷ Gleichung (6.4)
    $\theta_{\text{curr}} \leftarrow \theta_{\text{neu}}$ ▷ Vorschlag annehmen
  else
    θ_{curr} bleibt unverändert ▷ Vorschlag ablehnen
  Speichere $\theta^{(k)} \leftarrow \theta_{\text{curr}}$
return $\theta^{(1)}, \ldots, \theta^{(N)}$

6.3 Beispiel: Mehrdimensionale Normalverteilung

Mit dem vorhin dargestellten Metropolis-Hastings-Algorithmus kann man verschiedene Verteilungen samplen. Diese Verteilungen werden durch Funktionen mit mehreren, gegebenenfalls mehrdimensionalen Argumenten beschrieben. So kann man anhand des Metropolis-Hastings-Algorithmus eine 2-dimensionale Gauß-Verteilung samplen:

$$f(\vec{x}, \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\,\Sigma}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right] \tag{6.17}$$

mit

$$\vec{\mu} = \begin{pmatrix} 1.0 \\ -0.5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1.0 & 0.6 \\ 0.6 & 0.5 \end{pmatrix} \quad \text{und} \quad \vec{x} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

In diesem Beispiel ist die Zielverteilung bereits durch $f(\vec{x}, \vec{\mu}, \Sigma)$ gegeben und übernimmt direkt die Rolle des Posteriors aus Gleichung (6.5). Um die gewünschte Funktion des Posteriors (post) zu bestimmen, muss man sie logarithmieren. So ist:

$$\log(\text{post}) = \log(f(\vec{x}, \vec{\mu}, \Sigma)) \tag{6.18}$$

$$= \log\left(\frac{1}{(2\pi)^{n/2}\sqrt{\det\,\Sigma}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right]\right) \tag{6.19}$$

$$= \log\left(\frac{1}{(2\pi)^{n/2}\sqrt{\det\,\Sigma}}\right) + \log\left(\exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right]\right) \tag{6.20}$$

$$= \log(\text{konst.}) + \left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right) \tag{6.21}$$

Aus Gleichung (6.4) folgt:

$$\log(\text{post}) = -\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \tag{6.22}$$

Wenn man jetzt den Metropolis-Hastings-Algorithmus auf dieses Beispiel anwendet, bekommt man für den Best-Fit-Parametersatz

$$\vec{x} = \begin{pmatrix} 0.996 \\ -0.497 \end{pmatrix} ,$$

welcher sehr nah zum Erwartungswert/realen Wert μ liegt. Somit weist der Algorithmus eine hohe Genauigkeit auf, die man jedoch weiter optimieren kann, indem man die Anzahl der Iterationen erhöht und die Schrittweite verringert. Außerdem liefert er einen Corner-Plot, welcher darstellt, welche Parametersätze wie häufig als θ_{curr} vorkamen.

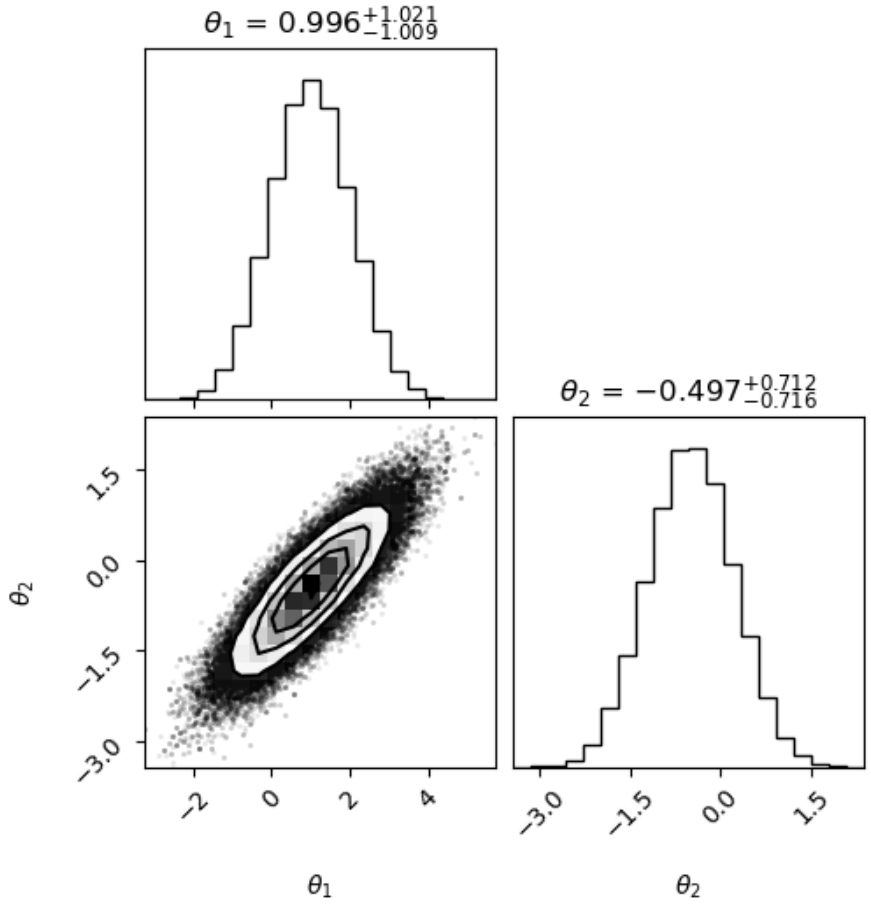


Abbildung 6.2: Corner-Plot des Beispiels: unten links sieht man das Streudiagramm, das alle angenommenen θ als Punkte zeigt; je dichter die Punkte liegen, desto wahrscheinlicher liegt der reale Wert dort. Oben und rechts lassen sich zwei Histogramme der jeweiligen Parameter θ_1 und θ_2 erkennen, mit den Werten, die sie während des Algorithmus angenommen haben.

PENDELVERSUCH

von Ela, Masuod

Die Schüler:innen und Kursleiter des Kurses 5.1 – Unendliche Weiten beim Experimentieren, aufgenommen von Simon.

7.1 Versuchsaufbau und Durchführung

In diesem Kapitel beschäftigen wir uns mit dem Forschungsproblem, ob es möglich sei, eine Höhe ohne die Verwendung eines Längenmessers (z.B. Lineal, Maßband, etc.) zu bestimmen. Stattdessen werden Anhand von experimentellen Messdaten und Modellen die Messdaten statistisch ausgewertet, um die Höhe abzuschätzen. Für das Experiment werden ein Schuh, ein Seil, ein Smartphone mit einer passenden Messapp, wie z.B. Physics Tool, und ein Computer mit Python benötigt. Das Experiment wird anhand des Beispiels der Höhe des Balkons im Innenhof durchgeführt. Dabei wird für den Aufbau der Schuh an das Seil gebunden, um als Pendel zu fungieren. Das Pendel wird vom Balkon herunter gelassen bis der Schuh sich knapp über den Boden befindet. Das Smartphone wird in den Schuh gelegt und die Messapp für die Messungen wird anschließend vorbereitet, um die Daten für die Beschleunigung auf den unterschiedlichen Achsen in Abhängigkeit der Zeit zu messen.Nun wird für die Messung das Pendel um einen moderaten Winkel von etwa $30°$ ausgelenkt, dann losgelassen, damit dieses hin- und herschwingen kann. Es ist wichtig, dass das Pendel parallel zur Wand schwingt und nicht durch äußere Gegenstände von der Bahn abkommt. Schließlich wird nach einer kurzen Zeit (ungefähr $20\,\mathrm{s}$) das Pendel angehalten. Die Daten werden dann auf einen Computer zur Weiterverarbeitung übertragen.

7.2 Datenanalyse

Zur sinnvollen Datenauslesung und -darstellung werden die Daten auf einem Computer mit Python verarbeitet. Bei der grafischen Darstellung dieser Daten entsteht der blaue Graph (siehe Abbildung 7.1), der sich durch die folgende Funktion annähern lässt:

$$a(t) = Ae^{-\gamma(t-t_0)} \cos\left(2\omega(t-t_0) + \psi_0\right) \tag{7.1}$$

Hierbei ist $t - t_0$ die Zeit seit Beginn der Datenaufnahme. Der Kosinus-Term wird für die Beschreibung der Periodizität der Daten genutzt. Zusätzlich beschreibt ψ_0 die Phasenverschiebung auf der t-Achse und A steht für die Anfangsamplitude. Da die Schwingungen des Pendels aufgrund von Luftwiderstand und Reibung mit der Zeit abklingen, nimmt die Amplitude kontinuierlich ab. Dieses Dämpfungsverhalten wird durch den Faktor $e^{-\gamma(t-t_0)}$ beschrieben. Dabei gibt γ die Stärke der Dämpfung an. Der orange Graph ist das Modell, welches die aufgestellte Modell-Likelihood maximiert, die Daten also am besten fittet. Dafür wurden zunächst händisch mögliche Parameterkombinationen und -intervalle ermittelt und daraufhin mithilfe des Metropolis-Hastings-Algorithmus aus Kapitel 6 die Likelihood maximiert. Die zentrale Informationsquelle für die Bestimmung der Höhe des vermessenen Balkons ist die Pendelfrequenz ω. Der Faktor 2 vor dem ω wurde eingeführt, weil das Smartphone nicht den Winkel des Pendels misst, sondern die Beschleunigung entlang seiner Sensorachse. Dadurch wiederholt sich das gemessene Signal innerhalb einer vollständigen Pendelschwingung zweimal (siehe Abbildung 7.2). Aus der gemessenen Kreisfrequenz ω kann die Pendellänge L aus der Gleichung

$$\omega = \sqrt{\frac{g}{L}} \quad \Longleftrightarrow \quad L = \frac{g}{\omega^2} \tag{7.2}$$

bestimmt werden. Hierbei ist $g = 9.81\,\mathrm{m/s^2}$ die Erdbeschleunigung, wodurch die ermittelte Kreisfrequenz direkte Rückschlüsse auf die Pendellänge L erlaubt, womit wir die Höhe des Balkons approximieren. Die für die Parameter verwendeten Priors wurden entweder aus der bestehenden Literatur ausgenommen ($g = 9.81\,\mathrm{m/s^2}$) oder anhand der anfangs manuellen Schätzung unter den vorher verfügbaren Daten (vgl. Abbildung 7.1) festgelegt. Im Folgenden wollen wir anhand verschiedener diagnostischer Grafiken die Konvergenz des Metropolis-Hastings-Algorithmus als auch die Korrelation der einzelnen Messgrößen illustrieren:

- **Traceplot:** Auf der horizontale Achse sind die Schritte, die der Metropolis-Hastings-Algorithmus durchgeführt hat, aufgetragen und die vertikale Achse zeigt die unterschiedlichen Werte von den Variablen, die es ausgerechnet hat. Die Oszillation um jeweils konstante Werte, wie im Trace-Plot-Diagramm von den Parameter dieses Experimentes, weist auf ein Konvergieren der Werte der Posteriore hin, die der Metropolis-Hastings-Alogrithmus generiert hat. Man spricht auch von der Konvergenz der Markov-Kette. Die gestrichelte Linie, auch genannt das „Ende des burn-in“, zeigt den Punkt, wo Einträge vorher in der

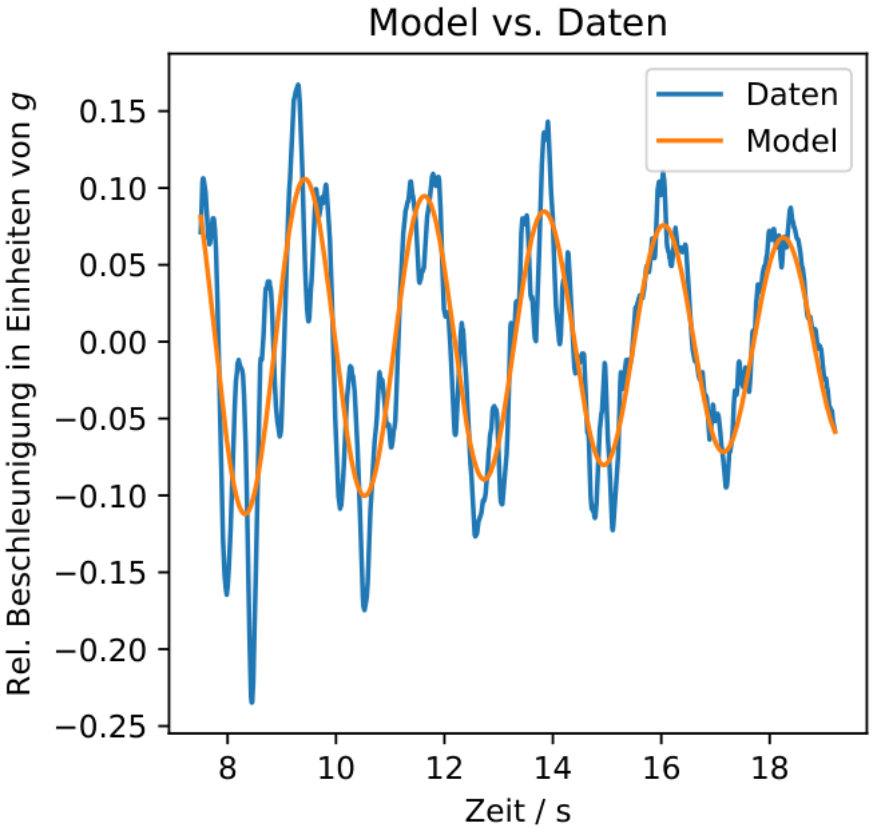


Abbildung 7.1: Die Messdaten des Pendelversuchs verglichen mit der Modell-Realisierung, welche die Likelihood maximiert.

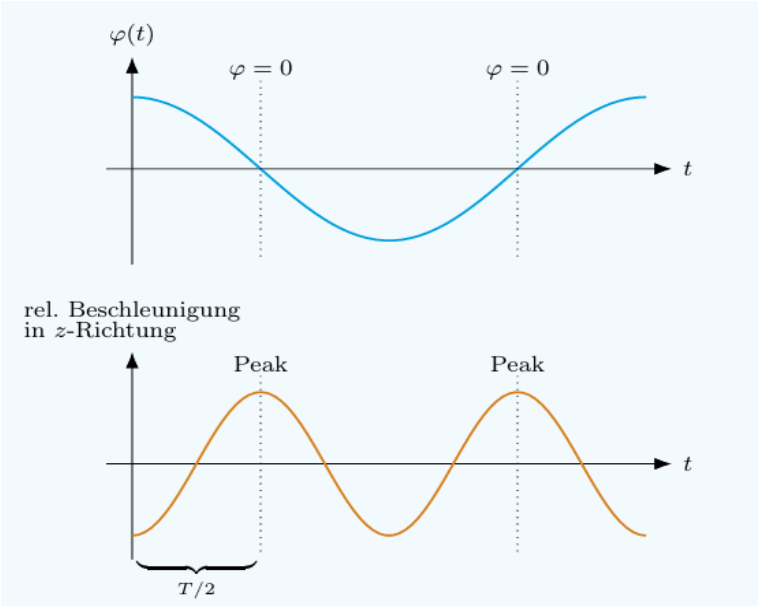


Abbildung 7.2: Der obere Graph zeigt den Winkelauslenkung (ϕ) am Seil während der untere das ϕ entlang der Beschleunigungsachse des Handys z-Richtung zeigt. Man sieht, dass der untere Graph die doppelte Frequenz (f) hat.

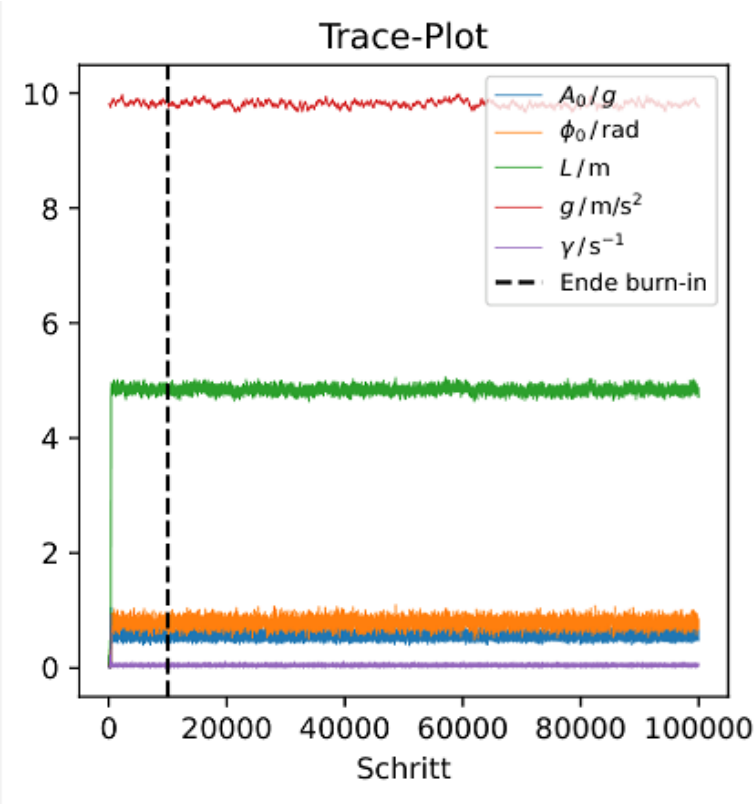


Abbildung 7.3: Trace-Plot der Markov-Kette aus der Datenanalyse auf dem die Daten aus dem Versuch basieren.

Markov-Kette ignoriert werden und nicht in die Datenauswertung einfließen (siehe Kapitel 6 für weitere Erklärungen).

- **Scatterplot:** Ein Scatterplot, auch Streudiagramm genannt, stellt eine Stichprobe von zwei Variablen in einem zweidimensionalen Diagramm dar. Dabei wird versucht eine Korrelation zwischen den beiden Variablen sichtbar zu machen. Aus den Daten dieses Experimentes wurden bei der Abbildung 7.4 Beschleunigung und Länge verglichen, wobei eine positive Korrelation zwischen den beiden erkennbar ist, da sich die Samples entlang der Diagonalen häufen. Eine enge Verteilung der Punkte bedeutet, dass der Wert der Variable genau bestimmt ist. Eine breite Verteilung der Punkte wiederum weist auf eine große Ungenauigkeit hin. Die Ursache der positiven Korrelation von L und g liegt bei $\omega = \sqrt{L/g}$, siehe Funktion (7.2). Denn wenn L größer wird, muss es g ihm gleichtun, damit die gemessene Kreisfrequenz ω konstant bleibt. Dies ist auch im umgekehrten Fall so, jedoch hat g bereits einen sehr exakten Richtwert, daher ist nur eine schmale Prior-Verteilung möglich.

- **Cornerplot:** Ein Cornerplot (Abbildung 7.5) besteht aus Histogrammen, welche die Werte der Variablen A_0, ψ_0 L, g und γ und deren Wahrscheinlichkeitsverteilung anzeigen. Die restlichen Darstellung sind mehreren Scatterplots, die aus allen möglichen Kombinationen der Variablen bestehen und als eine Kombinationstafel angeordnet sind. Darin kann man u.a. die Korrelation zwischen den einzelnen Variablen sehen.Beispielsweise ist zwischen A_0 und γ eine positive Korrelation sichtbar. Diese entsteht dadurch, dass eine größere Anfangsamplitude durch eine stärkere Dämpfung kompensiert werden muss, sodass ein ähnlicher Verlauf der Messdaten entsteht, vgl. Funktion (7.1). Diese Korrelation ist daher keine direkte physikalische Beziehung zwischen Anfangsauslenkung und Dämpfung dar.

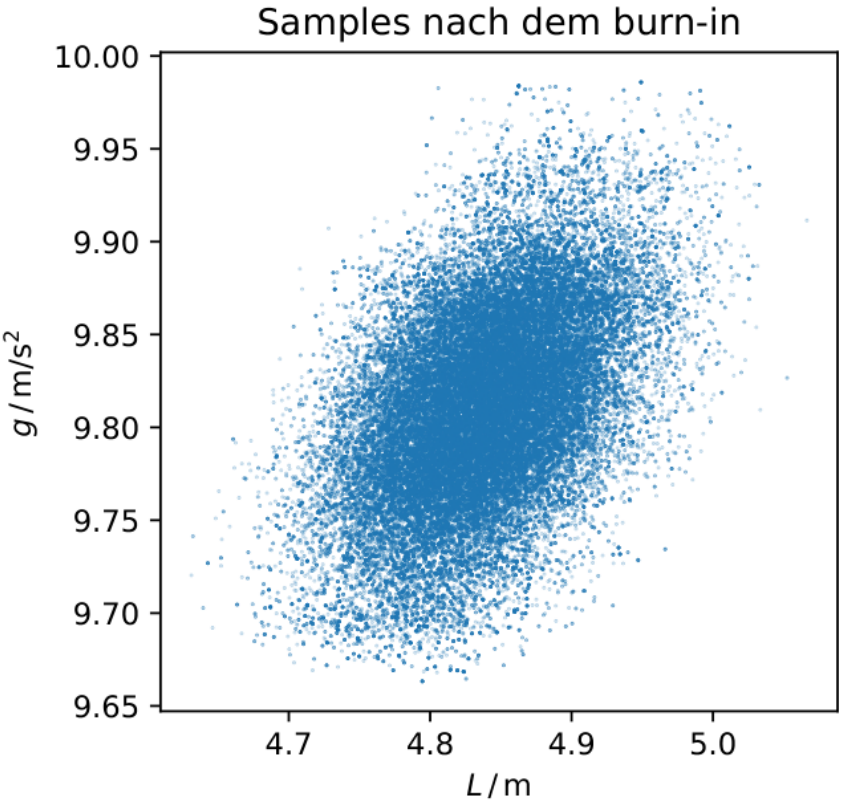


Abbildung 7.4: Sample-Verteilung nach dem Burn-in des Pendelversuchs.

7.3 Interpretation

Abschließend kann man sagen, dass das Forschungsproblem:

„(...), eine Höhe ohne Verwendung eines Längenmesser (z.B. Lineal, Maßband, etc.) zu messen."

lösbar ist, da die Ergebnisse, trotz einigen Vernachlässigungen wie von Elastizität, Dehnbarkeit, vereinfachter Dämpfung und Kleinwinkelnäherung, insgesamt sehr vielversprechend sind. Im Falle dieses Experiments ergibt sich für die Länge des Seil, aus dem das Pendel bestand, woraus sich die Höhe des Balkons schließen lässt. Der Wert des

$$L = (4.84 \pm 0.05)\,\text{m}. \tag{7.3}$$

und daher sehr präzise (vgl. Abbildung 7.5). Die relative Messgenauigkeit liegt im 1.03%-Bereich und die Unsicherheit erst in der zweiten Nachkommastelle. Vergleicht man den Wert der realen Länge des Seils ($\approx 4.8\,\text{m}$) mit dem aus den Messdaten bestimmten Wert, findet man eine sehr gute Übereinstimmung. Daraus lässt sich schließen, dass das verwendete Modell die wesentlichen Eigenschaften des Pendels beschreibt und die Daten angemessen verarbeitet wurden.

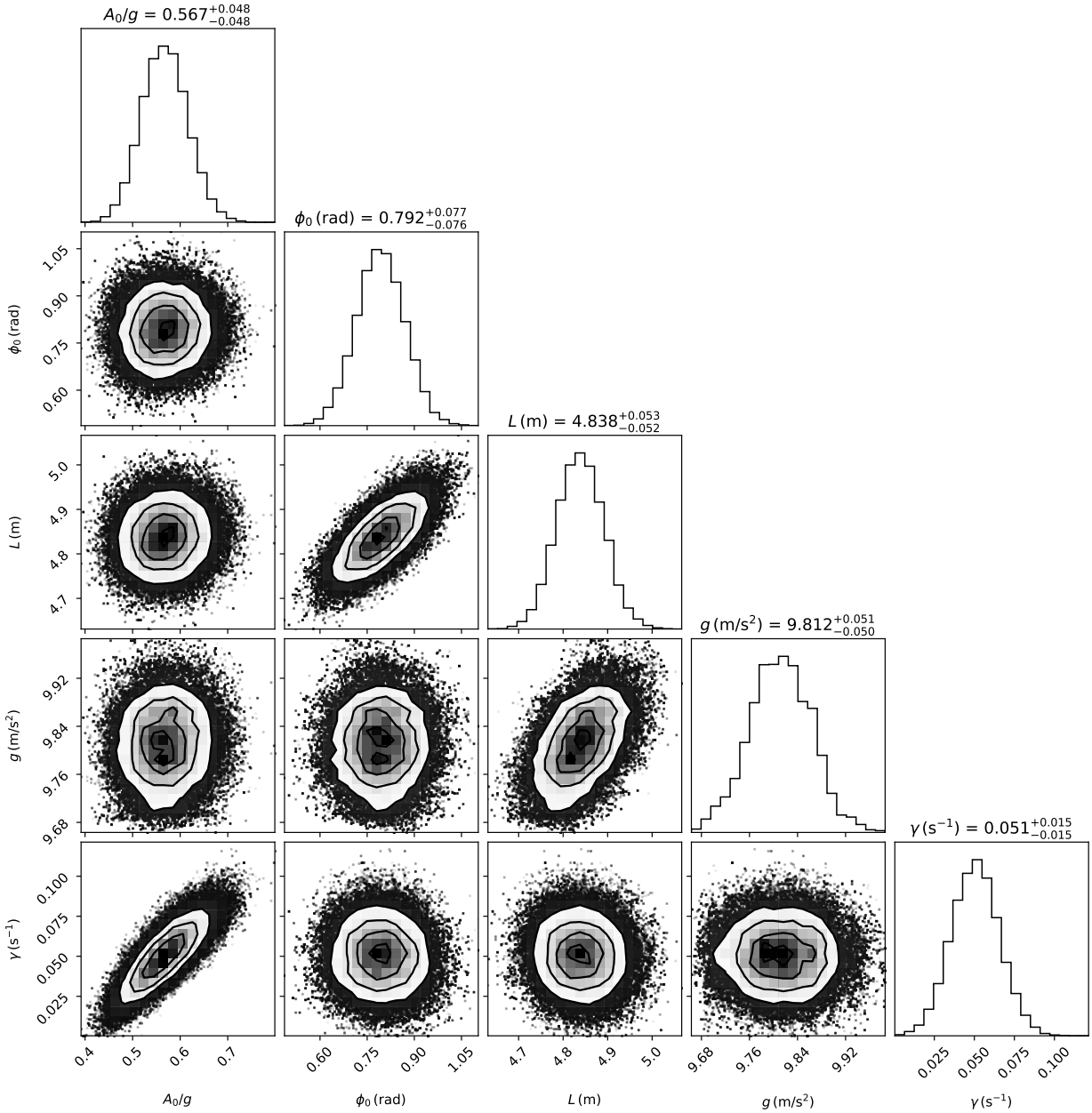


Abbildung 7.5: Corner-Plot der Parameterverteilungen.

GRAVITATIONSWELLEN UND PULSAR TIMING ARRAYS

von Victoria, Moritz

8.1 Pulsare

8.1.1 Entstehung von Pulsaren

Jeder Stern durchläuft einen Lebenszyklus, der im Tod des Sterns endet. Dieser Tod eines Sternes passiert dann, wenn die Wasserstoff- bzw. Helium-Atome, die dem Stern als Energiequelle dienten miteinander fusioniert wurden. In diesem Fall bestehen zwei Möglichkeiten, was mit dem Stern passiert. Bei Sternen mit einer Anfangsmasse unterhalb von ungefähr $8M_\odot$, wobei $M_\odot$ eine Sonnenmasse ist, expandieren sie zu Roten Riesen, stoßen ihre äußeren Gasschichten ab und kollabieren schließlich zu weißen Zwergen.

Sollte die Anfangsmasse eines Sternes größer sein als ungefähr $8M_\odot$, so implodiert der Stern und löst eine Supernova-Explosion aus. Zurück bleibt ein sehr dichter Neutronenstern oder bei noch höherer Anfangsmasse ein schwarzes Loch.

Beim Zusammenfallen in einen Neutronenstern bleibt der Drehimpuls eines Sterns enthalten. Da sich dessen Radius aber enorm verkleinert, erhöht sich aufgrund der Drehimpulserhaltung die Rotationsgeschwindigkeit des zurückbleibenden Sterns drastisch.

Beim Kollaps eines Roten Riesen zu einem Neutronenstern erhöht sich jedoch nicht nur Rotationsgeschwindigkeit, auch das Magnetfeld auf der Oberfläche des Sterns wird deutlich stärker. Dies sorgt dafür, dass der extrem schnell rotierende Neutronenstern auch ein extrem starkes Magnetfeld besitzt. Die Stärke des Magnetfeld befindet sich im Bereich von 10^8 Tesla bis hin zu 10^{11} Tesla bei sog. Magnetaren, welche eine Unter-Kategorie der Pulsare sind.

Das Magnetfeld auf der Oberfläche von rotierenden Neutronensternen kann so groß sein, dass es durch elektrische Induktion Elektronen aus der Oberfläche des Neutronensterns zieht. Diese Elektronen wirbeln als Plasma mit dem Magnetfeld um den Pulsar herum. Je weiter das Plasma vom Neutronenstern entfernt ist, desto schneller rotiert es. An dem Punkt wo sich die Rotationsgeschwindigkeit der Lichtgeschwindigkeit annähert, können die magnetischen Feldlinien nicht mehr geschlossen werden und Elektronen werden entlang der offenen Magnetfeldlinien ins Weltall geschleudert. Hierbei werden durch die sich auf Spiralen bewegenden, geladenen Teilchen auch Radiowellen ausgesendet. Da die Rotationsache nicht genau der Magnetfeldachse entspricht, rotieren die Magnetfeldachse und somit auch die Strahlen um die Rotationsachse. Das ist auch in Abbildung 8.1 dargestellt. Sollten diese Strahlen während ihrer Rotation an einem Punkt die Erde kreuzen, so bezeichnet man diesen Neutronenstern als Pulsar, weil der Radiowellenstrahl als Puls auf der Erde wahrgenommen wird. Aus diesem Grund werden Pulsare häufig als Sterne mit rotierenden Strahlen an beiden Polen dargestellt.

8.1.2 Messung von Pulsaren

Der erste Pulsar wurde 1967 von Jocelyn Bell entdeckt, die auf ein Radiosignal stieß, welches sich etwa alle 1.3 Sekunden wiederholte. Das erst als von Außerirdischen kommend vermutete Signal wurde bald als schnell rotierender Neutronenstern erkannt.

Da die Rotationsgeschwindigkeit eines Pulsars konstant ist und der Strahl immer der gleichen Bahn folgt, erscheint der Puls immer sehr präzise nach dem gleichen Zeitintervall, so genau, dass die Genauigkeit der von Atomuhren ähnelt. Dies ist auch der Grund warum man von Pulsaren oft als „kosmische Uhren“ spricht.

Da die Signale von Pulsaren meistens im Radiowellenbereich liegen, benötigt man ein Radioteleskop um sie aufzuzeichnen. Aus diesen Daten kann ein Timing Model gebildet werden, welches die nächsten Pulse vorhersagt.

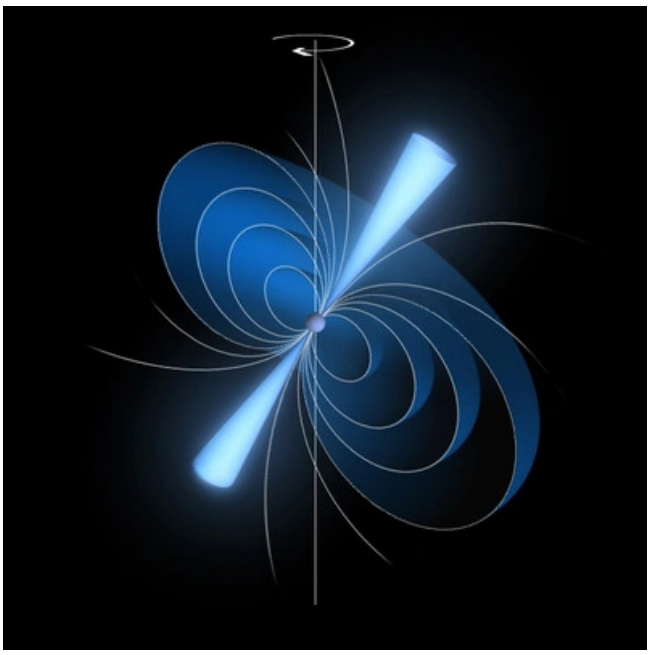


Abbildung 8.1: Darstellung aus dem Artikel „Wie kosmische Uhren ticken “ von der Website von „Welt der Physik “ (2013). Hier gezeigt ist eine künstlerische Darstellung eines Pulsars. Sichtbar ist der Neutronenstern als kleine Kugel, umgeben von den Feldlinien seines Magnetfeldes. Man sieht zwei Strahlen die entlang der Magnetfeldachsen aus dem Neutronenstern kommen. Auch entspricht die Magnetfeldachse nicht der Rotationsachse, was dafür sorgt, dass der Radiowellenstrahl um die Rotationsachse rotiert.

Beim Übereinanderlegen des Timing-Modell mit den tatsächlich aufgenommenen Pulsen kann man oft allerdings kleine Differenzen erkennen, die als Timing-Residuen bezeichnet werden, siehe dazu auch Abbildung 8.2. Gemessen werden die Timing-Residuen am Maximum des jeweiligen Pulses im Timing-Model im Vergleich zum Maximum des jeweiligen Pulses in den Aufnahmedaten. Für die Berechnungen der Timing-Residuen gilt also

$$\delta t = t_{\mathrm{D}} - t_{\mathrm{TM}}, \tag{8.1}$$

wenn t_{D} der Zeitpunkt des Maxima eines beliebigen Pulses k in den aufgezeichneten Daten und t_{TM} der Zeitpunkt des gleichen Pulses k laut des Timing Models ist.

8.2 Gravitationswellen

Wir widmen uns nun der Frage nach dem Ursprung der Timing-Residuen. Zum größten Teil liegt dies an Störfaktoren, wie Messfehlern. Ebenso können Störeffekte im Pulsar oder in der Propagation der Pulse Timing-Residuen verursachen. Allerdings kann die Ursache auch bei Gravitationswellen (GWn) liegen, wobei GWn Wellen in der Raumzeit sind. Bereits 1916 sagte Einstein, aufgrund seiner Allgemeinen Relativitätstheorie, die Existenz jener Wellen voraus, welche dann fast 100 Jahre später auch beobachtet werden konnten. GWn haben die Eigenschaft Abstände zwischen frei-fallenden Massen zu verändern. Dies gilt auch für den Abstand zwischen Erde und Pulsar, wenn eine GW den Raum zwischen diesen durchläuft. Dadurch werden minimale Timing-Residuen verursacht. Eine typische GW verschiebt zwei auffeinanderfolgende Pulse im Abstand von einer Millisekunde allerdings nur um ein paar Attosekunden (also 10^{-15} ms), weshalb die Daten über längere Zeiträume und von mehreren Pulsaren aufgezeichnet werden müssen.

Nachdem die Wirkung von GWn erläutert wurde, muss hier noch betont werden, dass GWn aus einem Frequenzspektrum bestehen und nicht etwa aus einer einzelnen Frequenz. Dies ist analog zu weißem Licht zu sehen, welches sich aus den Spektralfarben zusammensetzt und nicht etwa nur aus einer einzelnen Farbe. Die

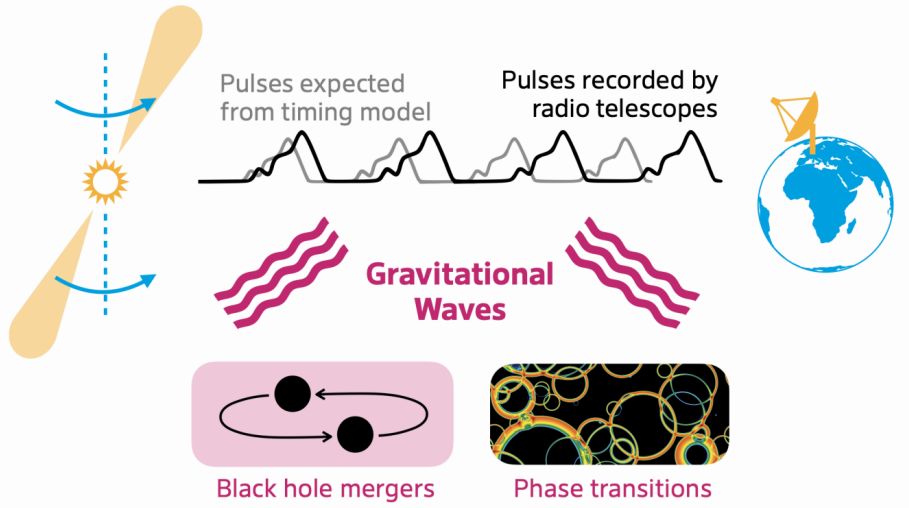


Abbildung 8.2: Schaubild aus einem Forschungshighlight der Universität Hamburg (Oktober 2024): "Das kosmische Brummen – woher kommt das Hintergrundrauschen aus Gravitationswellen?". Dargestellt sind die Pulsar-Signale anhand das Timing-Modell und die tatsächlich aufgenommenen Pulse. Anhand der Maxima der einzelnen Signale sieht man die Timing-Residuen, die hier im Laufe der Zeit immer größer werden.

Tatsache, dass eine GW durch ein Frequenzspektrum beschrieben werden kann, hilft uns verschiedene mögliche Quellen der GWn voneinander zu unterscheiden

Grundsätzlich gibt es zwei Phänomene, bei denen GWn entstehen können:

1. GWn entstehen dann, wenn sich zwei Massen (z.B. zwei stellare schwarze Löcher) umkreisen und letzlich verschmelzen. Bei diesem Prozess wird Energie abgestrahlt, welche sich in Form von GWn im Universum ausbreitet. Dieses Phänomen wird auf der Erde anhand des sog. LIGO-Detektors beobachtet, siehe Abb. 8.3. Die Apparatur besteht hierbei aus zwei senkrecht zueinander stehenden Armen, welche eine Länge von ca. 4 km aufweisen. Am Ende dieser Arme befindet sich jeweils ein Spiegel. Die Funktionsweise des Interferometers ist nun hierbei, dass ein Laserstrahl aufgespalten, durch beide Arme ausgesendet und schließlich durch die Spiegel auch wieder zurückreflektiert wird. Auf einem Photodetekor entsteht ein Interferenzmuster, welches sich ohne äußere Einflüsse (GWn) nicht verändert. Läuft allerdings eine GW durch den LIGO-Detektor, wird der eine Arm gestreckt, während der andere Arm gestaucht wird, was zu einer Veränderung des Interferenzmusters führt. Dadurch kann der Effekt der Gravitationswelle auf Materie vermessen werden, siehe Abb. 8.3.

2. Die zweite Quelle von Gravitationswellen sind stochastischer Natur, also die Überlagerung vieler unabhängiger Prozesse. Das entstehende Signal ist nicht zeitlich beschränkt sondern in Form eines stochastischen Gravtiationswellen-Hintergrund sichtbar. Hinweise auf diesen Gravitationswellen-Hintergrund entdeckte die NANOGrav-Kollaboration im Juni 2023. Die Quelle davon ist aktuell noch ungeklärt, wobei zurzeit zwei Modelle erforscht werden. Das eine Modell führt das Hintergrundrauschen auf die Verschmelzung supermassiver schwarzer Löcher zurück, während eine andere Interpretation der Daten die Quelle im Urknall und somit im frühen Universum sieht.

Würde das zweite Modell immer wahrscheinlicher werden, wäre dies revolutionär, da man so einen direkten Informationszugang zu dem frühen Universum hätte und darüber Informationen über das Universum vor dem Zeitpunkt der Rekombination gewinnen könnte. Die Rekombination ist hierbei der Zeitpunkt, an dem das Universum für elektromagnetische Strahlung durchsichtig wurde.

Um letzten Endes entscheiden zu können, welche der beiden Modelle am wahrscheinlichsten den Gravitationswellenhintergrund beschreiben würde, werden große Mengen an Daten benötigt, welche man durch sog. Pulsar Timing Arrays (PTA) erhält.

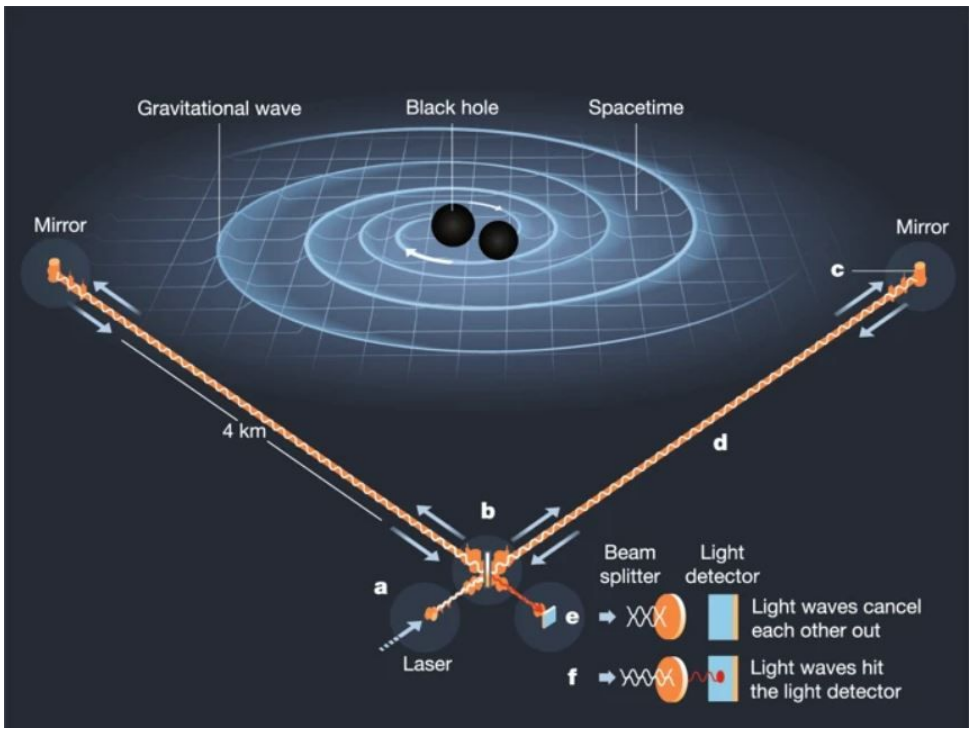


Abbildung 8.3: Hier wird eine schematische Darstellung des LIGO-Detektors von dem Artikel “The next frontier of gravitational waves” gezeigt, wobei der Artikel von M. Coleman Miller und Nicolás Yunes geschrieben und auf der Internetseite nature.com am 24.April 2019 veröffentlicht wurde.

Deshalb ist, die in der Praxis taugliche Lösung, die Verwendung von PTAs, um zwischen Gravitationswellenhintergrund und anderen Störfaktoren unterscheiden zu können . Dieses PTA besteht aus vielen Pulsaren, die durch die ganze Galaxie verteilt sind. Indem man die Daten und die jeweiligen Timing Residuen vergleicht, lässt sich so ein korreliertes Muster von Messfehlern oder Rauschen trennen. Das Signal einer Gravitationswelle bei Messung eines Pulsars wäre so gering, dass es von anderen Störfaktoren wie intrinsisches Rauschen eines Pulsars, Störungen durch den Weltraum, Messfehler oder anderes nicht zu differenzieren wäre. Die große Distanz zwischen der Erde und den PTAs erlaubt es uns also Gravitationswellen mit Frequenzen im nHz Bereich zu messen, wobei das PTA nicht in der Lage ist hohe Frequenzen aufzuzeichnen.

8.3 Datenanalyse der PTA-Daten

Nachdem wir uns der Erklärung der Funktionsweise von PTAs gewidmet haben, kommen wir nun zu der Auswertung der NANOGrav-Daten nach 15 Jahren Beobachtungsdauer. Über ein free spectrum haben wir die Daten in ein Diagramm geplottet, siehe Abb. 8.4, um das Gravitationswellen-Spektrum abzubilden. Dabei wird die Energiedichte der GWn (die Amplitude) in Abhängigkeit zu einem logarithmischen Frequenzintervall angegeben. Indem jene Energiedichte auf die kritische Energiedichte des Universums normiert ist, wird automatisch der Anteil der Energiedichte der GWn an der Gesamtenergiedichte des Universums angezeigt. Die kritische Energiedichte des Universums ist hierbei die Energiedichte die ein geometrisch flaches Universum benötigen würde.

Um nun die Daten mithilfe eines „Toy model“ beschreiben zu können, haben wir eine PTA-Likelihood konstruiert und zusätzlich den Metropolis-Hastings-Algorithmus, um die beste Parameter-Kombination für unser Toy model zu fnden, vgl. Abb. 8.5. Bei dem Model handelt es sich um ein gebrochenes Potenzgesetz, welches vier offene Parameter $\log_{10} A$, $\log_{10} f_{\text{peak}}$, a und c hat,

$$\log_{10}\left(h^2\Omega_{\text{gw}}(f)\right) = \log_{10} A + \log_{10}\left(\frac{x^a}{1+x^c}\right), \quad x = \frac{f}{f_{\text{peak}}}. \tag{8.2}$$

Es fungiert dabei als eine Parametrisierung einer physikalischen Signalquelle, wie der astrophysikalischen oder der kosmologischen Quelle des Signals.

Nachdem wir die PTA-Likelihood geplottet und den Metropolis-Hastings-Algorithmus verwendet hatten, wurden uns die wahrscheinlichsten Parameter für unser „Toy model“ in Form eines corner plots angezeigt, vgl. Abbildung 8.4. Dort sieht man, dass die wahrscheinlichste Amplitude bei etwa $10^{-8.63}$ und die wahrscheinlichste Frequenz bei etwa $10^{-8.30}$ liegt. Aus Abbildung 8.4 können auch die best-fit-Werte $A = 10^{-8.64} = 2.3 \cdot 10^{-9}$ und $f_{\text{peak}} = 10^{-8.03}\,\text{Hz} = 10.0\,\text{nHz}$, sowie die besten Werte für $a = 3.00$ und $c = 4.65$ entnommen werden. Hierbei steht a für die Steigung des Modells links des Peaks in Abbildung 8.5 und c für die Steigung des Modells rechts des Peaks. Alle diese Parameter konnten nun in das Gravitationswellen-Spektrum integriert werden, wodurch die finale Datenauswertung erhalten werden konnte, vgl. Abb. 8.5.

Indem jene Energiedichte auf die kritische Energiedichte des Universums normiert ist, wird automatisch der Anteil der Energiedichte der GWn an der Gesamtenergiedichte des Universums angezeigt. Die kritische Energiedichte des Universums ist hierbei die Energiedichte die ein geometrisch flaches Universum benötigen würde.

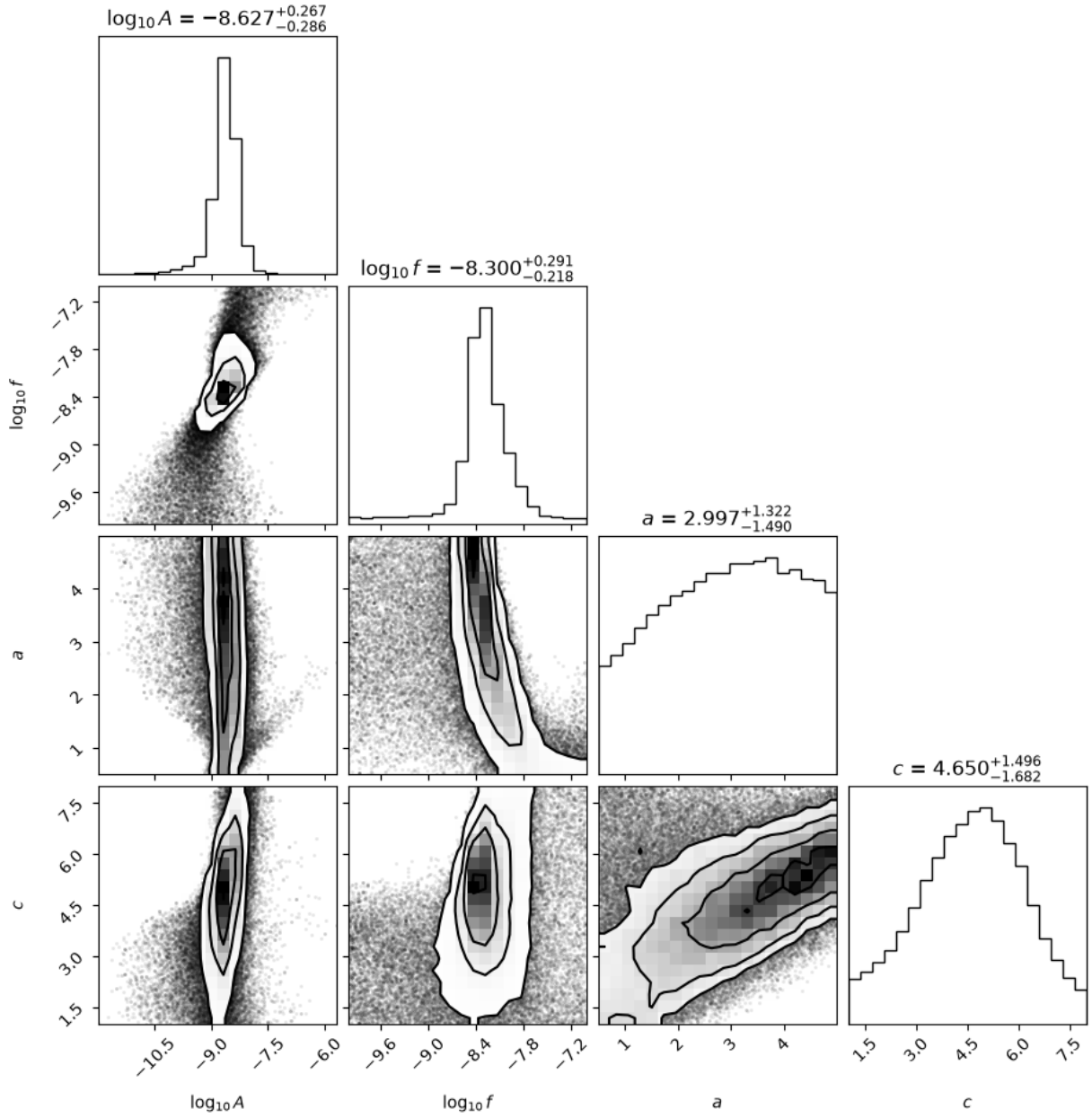


Abbildung 8.4: Hier ist ein in Python selbstprogrammierter corner plot sichtbar, wobei die 1σ, 2σ und 3σ-Umgebungen der Posterior-Verteilungen von $\log_{10} A$, $\log_{10} f_{\text{peak}}$, a und c sowie die dazugehörigen Streudiagramme dargestellt sind. Die Streudiagramme zeigen die Wahrscheinlichkeiten der möglichen Kombinationen der unterschiedlichen Parameter an, wobei die mittigen schwarzen Punkte die wahrscheinlichsten Kombinationen darstellen. In den Peaks der Normalverteilungen findet sich die wahrscheinlichste Kombination (schwarzer Punkt) auch wieder.

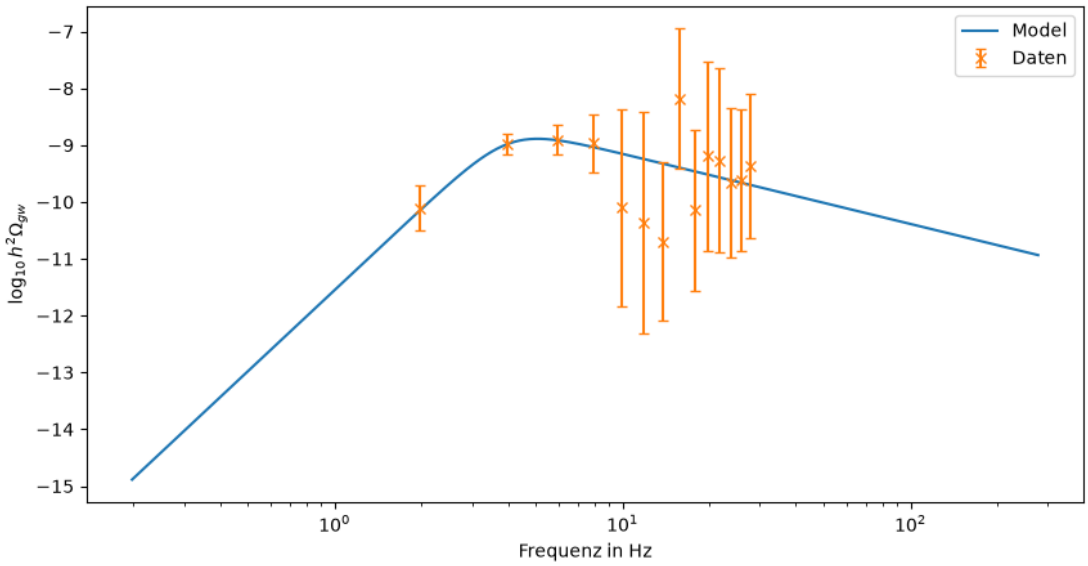


Abbildung 8.5: Dies ist eine Abbildung, die aus einem selbsterstelltem python plot stammt. Dabei stellen die orangenen Punkte die eigentlichen Daten mit dazugehörigem Fehler dar, während die blaue Kurve dem berechneten Modell für diese Daten entspricht.