

1 Einleitung

1.1 Deterministische und statistische Physik

In den bisherigen Bänden dieser Lehrbuchreihe zur Theoretischen Physik wurden die Grundgleichungen der Mechanik, der Elektrodynamik und der Quantentheorie vorgestellt und analysiert. Die Newton'schen Bewegungsgleichungen oder als Äquivalent die Lagrange'schen oder die Hamilton'schen Bewegungsgleichungen, die Maxwell-Gleichungen oder die quantenmechanischen Evolutionsgleichungen sind deterministische Gleichungen (siehe Tab. 1.1). Solche Gleichungen zeichnen sich dadurch aus, dass sie einen definierten Anfangszustand eindeutig auf einen späteren Zustand abbilden¹.

Tab. 1.1 Determinismus in der klassischen und Quantenmechanik

	klassische Mechanik	Quantenmechanik
Zahl der Variablen	2f	f
Anfangszustand	$\{q_i(0), p_i(0) : i = 1, \dots, f\}$	$\psi(q_1, \dots, q_f, t = 0)$
Bewegungsgleichungen	$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$ und $\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$	$i\hbar \frac{\partial \psi}{\partial t} = \hat{H}\psi$
Endzustand	$\{q_i(t), p_i(t) : i = 1, \dots, f\}$	$\psi(q_1, \dots, q_f, t)$

Unter Determinismus wollen wir deshalb eine eindeutige Abbildung eines Anfangszustandes auf einen späteren Endzustand über die Bewegungsglei-

1) Dabei ist zu beachten, dass vor allem für Probleme der Elektrodynamik oder der Quantenmechanik – wo diese Grundgleichungen im Wesentlichen partielle Differentialgleichungen sind – die Randbedingungen als Bestandteil der vollständig definierten Feldgleichungen bzw. quantenmechanischer Evolutionsgleichungen betrachtet werden.

chungen entsprechend dem Schema

$$\text{Anfangszustand} \rightarrow \text{Bewegungsgleichungen} \rightarrow \text{Endzustand} \quad (1.1)$$

verstehen. Da alle mikroskopischen Bewegungsgleichungen invariant gegenüber der Zeitumkehr $t \rightarrow -t$ sind, erzeugen dieselben Bewegungsgleichungen aber auch aus jedem Endzustand wieder den zugehörigen Anfangszustand². Diese Eigenschaft nennt man *Reversibilität*³:

$$\text{Endzustand} \rightarrow \text{Bewegungsgleichungen} \rightarrow \text{Anfangszustand}. \quad (1.2)$$

Typische Probleme der klassischen Mechanik bzw. der Quantenmechanik befassen sich gewöhnlich nur mit Systemen aus wenigen Massenpunkten. Zum Beispiel sind der harmonische Oszillator, das Wasserstoffatom oder das Kepler-Problem bekannte und exakt behandelbare Einteilchenprobleme. Ebenso gehören Doppelsternsysteme, das Wasserstoffmolekül oder das Heliumatom zu den Lehrbuchbeispielen für Zweikörperprobleme.

Wächst die Zahl der Freiheitsgrade an, dann wird die mathematische Behandlung immer komplizierter und die Lösungsmannigfaltigkeit nimmt enorm zu. Schon das Dreikörperproblem mit Gravitationswechselwirkung kennt keine explizit formulierbare allgemeine Lösung. Für eine hinreichend große Anzahl von Freiheitsgraden gibt es nur noch ganz wenige Systeme, die eine analytische Lösung erlauben⁴.

Bei Gasen, Flüssigkeiten und Festkörpern hat man es mit Systemen zu tun, die größenordnungsmäßig 10^{23} Freiheitsgrade haben. Trotzdem können über Eigenschaften von solchen Vielteilchensystemen Aussagen getroffen werden, die scheinbar über die Möglichkeiten der Mechanik bzw. Quantenmechanik hinausgehen⁵. Solche Aussagen, die z. B. den Druck eines Gases oder die kinetische Energie aller Partikel betreffen, sind Gegenstand der statistischen Physik. Diese Teildisziplin der Physik stellt außerdem eine Verbindung zwischen der Mechanik bzw. der Quantenmechanik einerseits und der Thermodynamik andererseits her, da z. B. der mechanisch bzw. quantenmechanisch definierte Druck auch als thermische Zustandsgröße fungiert oder die mittlere kinetische Energie der Partikel mit der Temperatur des Systems zusammenhängt.

- 2) Man beachte, dass eine Zeitumkehr üblicherweise mit einer Umkehr der Impulse bzw. einer komplexen Konjugation der Wellenfunktion verbunden ist.
- 3) manchmal auch mechanische bzw. quantenmechanische Reversibilität, um eine Unterscheidung zu den später diskutierten thermodynamisch reversiblen Prozessen zu haben
- 4) z. B. die harmonische Kette mit Wellenlösungen, siehe Band IV, Kap. 4, oder die Toda-Kette mit Solitonenlösungen
- 5) Erhaltungsgrößen wie Gesamtenergie, -impuls oder -drehimpuls sind natürlich bereits im Rahmen der klassischen Mechanik bzw. der Quantenmechanik zugänglich.

Historisch gesehen hat sich die Thermodynamik als ein von der Mechanik unabhängiges Gebiet entwickelt⁶. Sie ist eine phänomenologische Theorie, die auf Erfahrungstatsachen aufbaut, welche durch Experimente vielfach bestätigt wurden. Die Erfahrungstatsachen wurden in zwei sog. Hauptsätzen formuliert, die besagen, dass

1. Wärme eine Energieform ist, welche zusammen mit den anderen im System auftretenden Energien einem Erhaltungssatz genügt;
2. Wärme insofern eine Sonderrolle innerhalb der verschiedenen Energieformen spielt, als sie nicht beliebig in andere Energieformen umgewandelt werden kann, während umgekehrt andere Energieformen vollständig in Wärme umgewandelt werden können.

Ausgehend von diesen Hauptsätzen lassen sich dann Beziehungen zwischen den Zustandsvariablen, z. B. Druck p , Volumen V und Temperatur T , eines Systems herleiten. Es fällt sofort auf, dass solche Variablen das jeweilige System als Ganzes charakterisieren. Man spricht deshalb auch davon, dass thermodynamische Zustandsvariable den Makrozustand eines Systems beschreiben. Ein Zusammenhang mit der atomaren Mikrostruktur der Materie lässt sich allerdings im Rahmen der Thermodynamik nicht herstellen. Deshalb können thermodynamisch auch keine Materialeigenschaften, welche der Mikrostruktur des jeweiligen Systems entspringen, hergeleitet werden. Auch eine Begründung der erwähnten Hauptsätze ist im Rahmen der Thermodynamik nicht möglich. Diese beiden Hauptsätze bilden gewissermaßen die axiomatische Basis der Thermodynamik⁷.

Genaugenommen ist die Thermodynamik eine Theorie für Vielteilchensysteme mit einer unendlich großen Anzahl von Freiheitsgraden. In diesem Sinne ist auch der Begriff „thermodynamischer Limes“ zu verstehen, den wir des Öfteren gebrauchen werden, um von zunächst mikroskopisch formulierten Erkenntnissen auf thermodynamische Zusammenhänge zu kommen. Die Erfahrung zeigt, dass in der Praxis bereits eine sehr kleine Menge eines Materials genügt, um dieses mit den Gesetzen der Thermodynamik ausreichend zu beschreiben. Allerdings sind heute auch Experimente an Systemen mit vielleicht einigen hundert oder tausend Freiheitsgraden möglich. Diese lassen sich einerseits nur sehr aufwändig und in vielen Fällen – trotz aller modernen Rechentechnik – nicht besonders genau mit den deterministischen quantenmechanischen oder mechanischen Gleichungen beschreiben. Andererseits zeigt

- 6) Zur Zeit der Etablierung der Quantenmechanik war bereits der Zusammenhang zwischen Mechanik und Thermodynamik weitgehend verstanden, so dass der Weg von der Quantenmechanik über die Statistik zur Thermodynamik praktisch schon vorbestimmt war.
- 7) Zwei weitere Hauptsätze, der nullte und der dritte Hauptsatz, werden zur axiomatischen Begründung der Thermodynamik nicht benötigt.

sich, dass auf dieser mesoskopischen Ebene auch die Gesetze der Thermodynamik nicht mehr voll gelten und modifiziert werden müssen.

Die Verbindung zwischen der mikroskopischen – mechanischen oder quantenmechanischen – Beschreibung eines Vielteilchensystems und der zugehörigen thermodynamischen Beschreibung wird durch die statistische Physik hergestellt. Da der mikroskopische Zustand eines hinreichend großen Vielteilchensystems nicht genau bestimmt werden kann, setzt die statistische Physik von vornherein auf eine wahrscheinlichkeitstheoretische Beschreibung. Es wird davon ausgegangen, dass jeder denkbare, mikroskopisch detailliert zu beschreibende Zustand vom System mit einer gewissen Wahrscheinlichkeit angenommen werden kann. Das heißt aber auch umgekehrt, dass man jedem Experiment einen über seine Zeitdauer mittelnden Charakter zuordnet: Gemessen wird also nicht eine physikalische Größe zu einem aktuellen Zustand eines Systems, sondern der Mittelwert über alle erlaubten Realisierungen der Zustände. Dieses Vorgehen bedingt eine Reihe von Interpretationen, die im Rahmen der Mechanik bzw. der Quantenmechanik so keine Rolle spielen.

Um überhaupt zu erklären, wie die Statistik ins Spiel kommt, benutzt man den Begriff des *Ensembles*. Darunter versteht man eine unendliche Menge physikalisch äquivalenter Systeme, die sich nur durch ihre Zustände unterscheiden. Prinzipiell sorgt die Dynamik jedes einzelnen Systems dafür, dass dieses sämtliche erreichbaren Zustände im Laufe der Zeit auch durchläuft. Zu jedem Zeitpunkt befinden sich die Systeme des Ensembles dann in einem der erlaubten Zustände. Damit kann man eine Häufigkeitsverteilung der Zustände bestimmen, die im thermodynamischen Gleichgewicht sogar zeitlich unveränderlich ist. Mit Hilfe dieser Verteilung ist es nun möglich, den Wert einer physikalischen Größe nicht als individuelle Charakteristik eines Systems, sondern über den Mittelwert als Kenngröße des ganzen Ensembles zu verstehen. Diese Ensemblemittelwerte stehen in direkter Beziehung zu den thermodynamischen Größen.

Natürlich ist das Ensemble nur ein geeignetes Modell zur Interpretation der statistischen Methode. In der Realität wird man thermodynamische Messungen an einem einzelnen System vornehmen und nicht an einem Ensemble gleichartiger Systeme, da bei jeder thermodynamischen Messung an den einzelnen Systemen des Ensembles stets das gleiche Ergebnis zu erwarten ist. Zur Rechtfertigung der in der Theorie und im Experiment konzeptionell unterschiedlichen Vorgehensweisen verwendet man das Argument der bei vorhandener Ergodizität⁸ gültigen Gleichheit von Zeit- und Ensemblemittelwert. Dabei lässt man sich von der Vorstellung leiten, dass das Ergebnis einer Messung als Mittelwert über die Messzeit zu verstehen ist. In diesem Sinne wäre dann der Messwert mit dem Ensemblemittelwert gleichzusetzen.

8) siehe hierzu die Abschnitte 2.4.3 und 2.4.4

Allerdings hat auch diese Interpretation einen Haken. Der empirisch bestimmte Zeitmittelwert stimmt formal nur dann mit dem Ensemblemittelwert überein, wenn man eine unendlich lange Messzeit voraussetzt. Das ist aber gewöhnlich nicht der Fall. Tatsächlich spielt ein mathematisches Theorem, nämlich das Gesetz der großen Zahlen, zum Verständnis eine entscheidende Rolle. Wir werden uns noch in diesem Kapitel kurz mit dieser Thematik befassen. Bereits jetzt kann aber gesagt werden, dass alle thermodynamischen Größen die theoretische Eigenschaft haben, um ihren Zeitmittelwert zu schwanken. Nur ist die Schwankungsbreite im Verhältnis zum Mittelwert so klein, dass diese praktisch nicht gemessen werden kann und folglich im thermodynamischen Limes irrelevant wird. Damit wird aber der Zeitmittelwert und damit auch der Ensemblemittelwert repräsentativ für jedes einzelne Vielteilchensystem.

1.2

Aufbau des Bandes Statistische Physik und Thermodynamik

Dieses Lehrbuch befasst sich fast ausschließlich mit Fragen der statistischen Physik und der Thermodynamik des Gleichgewichts. Dabei wird der Zugang möglichst parallel und weitgehend gleichberechtigt sowohl von der Seite der klassischen Mechanik als auch von der Seite der Quantenmechanik gesucht.

Nachdem noch in diesem Kapitel einige Grundlagen der mathematischen Statistik mehr wiederholend, aber schon mit Bezug auf die in diesem Band behandelten Themen, dargestellt werden, widmet sich das nächste Kapitel der mechanischen und quantenmechanischen Begründung einer statistischen Beschreibung makroskopischer Systeme. Dabei geht es im klassischen Teil darum, das durch unzureichende Informationen über den mikroskopischen Zustand entstehende Defizit in einer statistischen Alternative zu den Newton'schen Bewegungsgleichungen, nämlich der Liouville-Gleichung, zu berücksichtigen. Im Weiteren geht es dann vor allem darum zu klären, warum und wie ein sich selbst überlassenes isoliertes System aus einem beliebigen Anfangszustand in einen Gleichgewichtszustand gelangt oder genauer in einen Zustand strebt, der makroskopisch nicht mehr von einem Gleichgewichtszustand unterschieden werden kann. Auf der quantenmechanischen Ebene wird in diesem Kapitel die von Neumann-Gleichung abgeleitet, als deren klassisches Analogon die schon erwähnte Liouville-Gleichung fungiert.

In Kapitel 3 wird das mikrokanonische Ensemble im Rahmen der statistischen Physik analysiert. Ein solches Ensemble enthält abgeschlossene Systeme. Wir werden im Rahmen dieses Kapitels den Begriff Entropie klären und erste grundlegende Relationen zur Thermodynamik herstellen.

Obwohl das mikrokanonische Ensemble eine vollständige statistische Beschreibung erlaubt und auch den Anschluss an die Thermodynamik vollstän-

dig herstellt, ist es für praktische Rechnungen nicht immer geeignet. Weitaus besser handhabbar sind kanonische Ensemble, mit denen wir uns in Kapitel 4 befassen. Hier werden wir neben grundsätzlichen Relationen vor allem die Theorie des idealen Gases diskutieren.

Ebenfalls relativ gut behandelbar sind großkanonische Ensemble. Wir werden die Anwendbarkeit dieses Konzepts in Kapitel 5 vor allem bei der Beschreibung quantenmechanischer Vielteilchensysteme aus Elektronen, Phononen oder Photonen demonstrieren.

Gegenstand des Kapitels 6 ist die Untersuchung wechselwirkender Teilchen und die Darstellung der Grundzüge von Phasenübergängen auf der statistischen Ebene.

Die thermodynamische Behandlung von Vielteilchensystemen, die dieser physikalischen Beschreibung eigenen Argumentationen und deren Beziehungen zur statistischen Physik finden sich in Kapitel 7.

1.3

Grenzen der Statistik und Thermodynamik des Gleichgewichts

In diesem Band werden wir, abgesehen von einigen wenigen grundsätzlichen Überlegungen, ausschließlich Probleme der Gleichgewichtsthermodynamik und -statistik untersuchen.

Damit sind natürliche alle Fragen, die eine thermodynamische Prozessführung oder auch nur die Störung des Gleichgewichts durch äußere Einflüsse betreffen, gar nicht oder nur rudimentär beantwortbar. Immerhin kann man im Rahmen der Thermodynamik verschiedene Ungleichungen ableiten, die in gewisser Hinsicht thermodynamische Prozesse eingrenzen. Aber solche typischen Fälle wie die Berechnung von Leitfähigkeiten oder die Beschreibung von Diffusionsprozessen erfordern eine dynamische Beschreibung, die erst im Rahmen einer Nichtgleichgewichtsstatistik formuliert werden kann.

Eine Ausnahme sei allerdings erwähnt. Im Rahmen der statistischen Physik stationärer Zustände können auch sogenannte Fließgleichgewichte behandelt werden. Dabei handelt es sich um makroskopische Systeme, in denen durch eine geeignete Kopplung an die Umgebung bzw. an externe Felder permanent ein stationärer Nichtgleichgewichtszustand aufrechterhalten wird. Beispielsweise führt man einem Laser durch Pumplicht ständig Energie zu, die in Form der Laserstrahlung wieder abgegeben wird. Durch den Laser fließt also im stationären Zustand Energie. Andere Beispiele für Fließgleichgewichte findet man in der Chemie und der Biologie. Diesen Fließgleichgewichten ist gemeinsam, dass sich die meisten der vielen Freiheitsgrade des makroskopischen Systems im thermischen Gleichgewicht befinden und nur einige wenige Freiheitsgrade hoch angeregt sind. In diesem Band können wir aber auf die-

se äußerst interessanten Probleme, die auch Gegenstand aktueller Forschung sind, nicht eingehen.

Ebenfalls nicht berücksichtigt werden Fragen der relativistischen Thermodynamik, etwa die Transformationsgesetze für Temperaturen beim Übergang zwischen zwei Inertialsystemen oder die relativistische Formulierung der thermodynamischen Zustandsgleichungen. An einigen Stellen dieses Bandes benutzen wir zwar Erkenntnisse der relativistischen Physik, verbinden diese aber nicht zu einem thermodynamisch-relativistischen Gesamtkonzept.

Schließlich werden wir in diesem Band nur an wenigen Stellen auf mesoskopische Systeme eingehen. Dieses in der Forschung momentan so wichtige Gebiet, in dem sich deterministische und statistische Konzepte überlappen und eine Vielzahl numerischer Methoden Verwendung findet, ist derartig facettenreich, dass auch nur eine einigermaßen systematische Behandlung den Rahmen dieses Buches sprengen würde.

1.4

Grundbegriffe der Statistik

1.4.1

Wahrscheinlichkeit und Wahrscheinlichkeitsverteilung

Der Wahrscheinlichkeitsbegriff ist uns bereits aus der Quantenmechanik bekannt. Am einfachsten lässt er sich im Rahmen des Häufigkeitsbegriffs erläutern. Angenommen, ein Experiment hat N verschiedene Realisierungen, d. h. es können N verschiedene Messwerte A_l mit $l = 1, \dots, N$ realisiert werden. Der Ausgang des Experiments sei rein zufällig, also durch keine irgendwie gearteten vorhergegangenen Messungen erkennbar. Wir wollen dann dieses Experiment M -mal hintereinander ausführen. Jedes Experiment liefert ein bestimmtes Messergebnis A_l , das innerhalb der Statistik auch als Ereignis bzw. Zufallsereignis bezeichnet wird. Innerhalb der M Experimente wird man dann z. B. den Wert A_1 m_1 -mal messen, den Wert A_2 insgesamt m_2 -mal usw. Natürlich gilt $m_1 + m_2 + \dots + m_N = M$. Bildet man jetzt das Verhältnis

$$h_l = \frac{m_l}{M} \quad (1.3)$$

wobei die Normierung

$$\sum_{l=1}^N h_l = 1 \quad (1.4)$$

gilt, dann sind die so definierten *Häufigkeiten* h_l empirische Schätzungen der Wahrscheinlichkeit, den Messwert A_l in einem Experiment zu erhalten. Bezeichnet man mit p_l die Wahrscheinlichkeit, dass sich in einem Experiment

der l -te Messwert realisiert, dann hängt diese mit der entsprechenden Häufigkeit durch den Grenzübergang $M \rightarrow \infty$ zusammen

$$p_l = \lim_{M \rightarrow \infty} h_l \quad (1.5)$$

Man bezeichnet diese Häufigkeitsinterpretation der Wahrscheinlichkeit auch als *Gauß'sche Wahrscheinlichkeitsdefinition*. Neben dieser Wahrscheinlichkeitsinterpretation werden von der Internationalen Organisation für Standardisierung (ISO) noch zwei weitere Varianten aufgeführt⁹.

Einerseits handelt es sich dabei um die *kombinatorische Interpretation*, bei der das Verhältnis der Anzahl aller eine bestimmte Situation betreffenden Fälle im Verhältnis zur Gesamtzahl der Fälle berechnet wird. Dazu gehört etwa die Frage, wie groß die Wahrscheinlichkeit ist, mit k Würfeln die Gesamtaugenanzahl m mit $k \leq m \leq 6k$ zu erreichen¹⁰. Für die Augenzahl $k + 1$ gibt es z. B. k Möglichkeiten – nämlich der erste Würfel erzielt eine 2, alle anderen Würfel eine 1, oder der zweite Würfel erzielt eine 2 und alle anderen die 1 usw. Insgesamt gibt es 6^k Fälle, so dass die Wahrscheinlichkeit für die Augenzahl $k + 1$ bei k Würfeln gerade $6^{-k}k$ ist.

Die andere Version ist die *Interpretation nach Bayes*. Bei diesem Konzept wird die Wahrscheinlichkeit als ein Maß für die Sicherheit des Eintretens eines Ereignisses verstanden. Dieses Maß kann – und darin unterscheidet sich die Bayes'sche Interpretation von der Gauß'schen Variante – bereits vor der eigentlichen Messung feststehen. Es kann sich dabei sogar um eine rein subjektive Annahme oder Vermutung handeln, die a priori unter Heranziehung aller bekannten Informationen getroffen wird, im Gegensatz zu der a posteriori – also erst nach dem Experiment – in Form der Häufigkeitsverteilungen feststehenden Gauß'schen Wahrscheinlichkeit.

In ihrer mathematischen Behandlung unterscheiden sich alle drei Interpretationen nicht. Im Rahmen der statistischen Physik verwendet man meistens die Gauß'sche Interpretation, so z. B. bei der Ensembletheorie. In manchen Fällen ist aber auch die Bayes'sche Interpretation sinnvoll, zumal diese praktisch mit einem System auskommt, dessen mikroskopischen Zustand man nicht genau kennt, über dessen Realisierung man aber im Rahmen der physikalischen Gesetze sinnvolle Vermutungen anstellen kann. Wir werden auf diese Problematik in den Abschnitten 2.2 und 2.3.3 noch genauer zurückkommen.

Wahrscheinlichkeiten sind stets so definiert, dass sie für jedes Ereignis l einen Wert p_l mit $0 \leq p_l \leq 1$ besitzen. Zudem sind sie als Konsequenz von

⁹ Guide to the Expression of Uncertainty in Measurement, (ISO, Genf, 1993).

¹⁰ Die Lösung dieser Aufgabe bleibe dem Leser überlassen.

(1.4) auf 1 normiert

$$\sum_{l=1}^N p_l = 1 \quad (1.6)$$

In der statistischen Physik hat man es meistens mit sich gegenseitig ausschließenden Einzelereignissen zu tun. So liefert im obigen Beispiel das Experiment eindeutige Messergebnisse, d. h. bei einer einzelnen Messung wird genau einer der Messwerte A_l registriert, niemals jedoch gleichzeitig z. B. A_1 und A_2 . Nur in diesem Fall ist die Normierung in der Form (1.6) korrekt¹¹ und die Zusammenfassung sich ausschließender Ereignisse zu Ereignisgruppen ist in den Wahrscheinlichkeiten strikt additiv. So ist z. B. die Wahrscheinlichkeit, beim obigen Experiment entweder A_1 oder A_2 zu messen, gerade $p_1 + p_2$. In diesem Sinne kann die Normierung (1.6) auch so verstanden werden, dass die Wahrscheinlichkeit, überhaupt eines der zulässigen Ereignisse zu finden, gerade 1 ist.

Neben den Wahrscheinlichkeiten, die hauptsächlich bei diskreten, und damit meistens der Quantenmechanik nahestehenden, Ereignissen Verwendung finden, benutzt man in der statistischen Physik sehr häufig Wahrscheinlichkeitsdichten. Diese treten dann auf, wenn alle Punkte eines kontinuierlichen Gebietes mit zulässigen Ereignissen A verbunden werden können. Es ist dann sinnvoll, die Einzelereignisse nicht auf Punkte, sondern auf infinitesimal kleine Intervalle zu beziehen, in die man das Gebiet aller möglichen Realisierungen aufteilt. Dabei wird vorausgesetzt, dass die infinitesimalen Elemente einerseits das gesamte zulässige Gebiet überdecken, sich aber gegenseitig nicht überlappen. Damit gibt es wieder nur sich gegenseitig ausschließende Ereignisse, denn jede Realisierung fällt stets genau in ein infinitesimal kleines Intervall, niemals in zwei oder gar mehrere und niemals in kein Intervall.

Dann kann man fragen, wie groß die Wahrscheinlichkeit ist, dass ein Wert aus einem infinitesimalen Intervall dA um den Wert A realisiert wird. Diese Wahrscheinlichkeit wird selbst ebenfalls nur infinitesimal klein sein. Es kann deshalb folgender Zusammenhang erwartet werden

$$dP(A) = \rho(A)dA \quad (1.7)$$

Die hier auftretende Größe $\rho(A)$ bezeichnet man als *Wahrscheinlichkeitsdichte* oder auch *Wahrscheinlichkeitsverteilung* der kontinuierlichen *Zufallsvariablen* A . Da sich jede Realisierung irgendwo im zulässigen Gebiet befinden muss, ergibt sich damit auch die Normierungsbedingung für Wahrscheinlichkeitsdichten

$$\int \rho(A)dA = 1 \quad (1.8)$$

¹¹) Schließen sich Ereignisse nicht gegenseitig aus, ist es sinnvoller auf die mengentheoretische Beschreibung der Wahrscheinlichkeitstheorie auszuweichen, siehe z. B. [27].

Wahrscheinlichkeitsdichten sind stets nichtnegative, auf 1 normierte Funktionen ihrer Variablen. Bei einer Transformation $A \rightarrow A'$ der Variablen ändert sich auch die Wahrscheinlichkeitsdichte, $\rho \rightarrow \rho'$. Dagegen sind die mit ihr verbundenen infinitesimalen Wahrscheinlichkeiten invariant¹². Damit erhalten wir sofort für Wahrscheinlichkeitsdichten das Transformationsgesetz

$$\rho(A)dA = \rho'(A')dA' \quad (1.9)$$

Mit Hilfe der δ -Funktion kann man jederzeit auch die Wahrscheinlichkeiten für diskrete Ereignisse durch eine Wahrscheinlichkeitsdichte ausdrücken. So findet man z. B. für die zu Beginn des Abschnitts im Zusammenhang mit (1.5) beschriebene Situation die Dichte

$$\rho(A) = \sum_{l=1}^N p_l \delta(A_l - A) \quad (1.10)$$

und die Wahrscheinlichkeit, bei einem Experiment Messwerte aus einem bestimmten Intervall $A \in [A_{\min}, A_{\max}]$ zu finden, ergibt sich einfach durch Integration über dieses Intervall.

1.5

Multivariable und zusammengesetzte Ereignisse, bedingte Wahrscheinlichkeiten

Ein Ereignis A kann auch durch mehrere Variablen a_1, a_2, \dots, a_n gekennzeichnet werden. So kann man z. B. die Position x eines einzelnen Teilchens in einer Flüssigkeit beobachten. Da dieses ständig Stöße mit benachbarten Molekülen ausführt, bewegt es sich in einer zufälligen Zitterbewegung durch den Raum. Naiv könnte man deshalb fragen, mit welcher Wahrscheinlichkeit sich das Teilchen an einem bestimmten Ort aufhält, oder genauer mit welcher Wahrscheinlichkeit sich das Teilchen in einem infinitesimalen Volumen dV um den Punkt x aufhält. Wie im vorangegangenen Abschnitt lässt sich auch in diesem Fall über

$$dp(x) = \rho(x)dV \quad (1.11)$$

eine Wahrscheinlichkeitsdichte einführen, deren Argument jetzt aber ein dreidimensionaler Vektor ist. Da sich das beobachtete Teilchen irgendwo im Gesamtvolumen aufhalten muss, ergibt sich auch sofort als Normierungsbedingung

$$\int \rho(x)dV = \int \rho(x)d^3x = 1 \quad (1.12)$$

¹²⁾ Der Anteil aller Realisierungen, die aus physikalischen Gründen ein bestimmtes Gebiet treffen, darf sich nicht ändern, wenn man das Gebiet unverändert lässt und nur die mathematische Darstellung wechselt.

Verallgemeinert man diese Aussage, dann kommt man zu der Schlussfolgerung, dass jede nichtnegative Funktion $\rho(a_1, a_2, \dots, a_n)$ eine *multivariable Wahrscheinlichkeitsdichte* bildet, wenn sie entsprechend

$$\int \rho(a_1, a_2, \dots, a_n) da_1 da_2 \dots da_n = 1 \quad (1.13)$$

normiert ist.

Wir wollen noch zwei weitere, in diesem Band relativ häufig auftretende Wahrscheinlichkeitsdichten kurz charakterisieren. In der klassischen Physik ist der mechanische Zustand eines Massenpunktes nicht nur durch die Ortskoordinaten, sondern auch durch den Impuls definiert, so dass man dementsprechend auch eine Wahrscheinlichkeitsdichte $\rho(\mathbf{x}, \mathbf{p})$ finden kann. Hier besteht jetzt ein Einzelereignis darin, das Teilchen in einem 6-dimensionalen Volumenelement $d\mu = d^3x d^3p$ um den Ort \mathbf{x} und den Impuls \mathbf{p} zu finden. Dieser 6-dimensionale Raum wird innerhalb der statistischen Physik oft auch als μ -Raum bezeichnet (vgl. Abb. 1.1). Jeder Punkt dieses Raumes entspricht einem Ereignis der gerade beschriebenen Art, d. h. einem durch sechs Variable repräsentierten multivariablen Ereignis.

Man kann auch die Wahrscheinlichkeitsdichte dafür angeben, ein mechanisches System aus N Teilchen in einem Punkt $(x_1, \dots, x_N, p_1, \dots, p_N)$ des $6N$ -dimensionalen, aus allen Teilchenkoordinaten und -impulsen aufgespannten Phasenraumes zu finden. Wir bezeichnen, der bereits in Band I dieser Lehrbuchreihe eingeführten Konvention folgend, die Position eines $6N$ -dimensionalen Systems im *Phasenraum* (Γ -Raum) mit $\vec{\Gamma}$. Dann ist

$$dP(\vec{\Gamma}) = \rho(\vec{\Gamma}) d\Gamma \quad (1.14)$$

die infinitesimale Wahrscheinlichkeit, das System in einem $6N$ -dimensionalen Volumenelement $d\Gamma$ um den Punkt $\vec{\Gamma}$ zu finden.

Im Γ -Raum repräsentiert ein Punkt (dargestellt durch $6N$ Variable) ein System von N Massenpunkten. Dasselbe System kann auch im μ -Raum dargestellt werden, allerdings erhalten wir im μ -Raum N Punkte, einen für jeden Massenpunkt des Systems. Wenn wir im Γ -Raum eine Menge von Punkten vorfinden, so repräsentiert jeder dieser Punkte ein ganzes System, die Menge von Punkten also eine Menge von Systemen aus einem Ensemble oder sogar das ganze Ensemble. Für das weitere Verständnis ist die Unterscheidung zwischen μ - und Γ -Raum wichtig. Die Verhältnisse sind in Abb. 1.1 und Abb. 1.2 nochmals skizziert.

Wie im 6-dimensionalen μ -Raum entspricht auch im $6N$ -dimensionalen Phasenraum jeder Punkt einem Ereignis, allerdings wird jedes Ereignis jetzt durch $6N$ Variablen beschrieben und kennzeichnet den mechanischen Zustand des jeweiligen Systems. Will man dasselbe Ereignis im μ -Raum beschreiben, so sind nach obiger Diskussion dazu N Punkte notwendig. Dieses Bei-

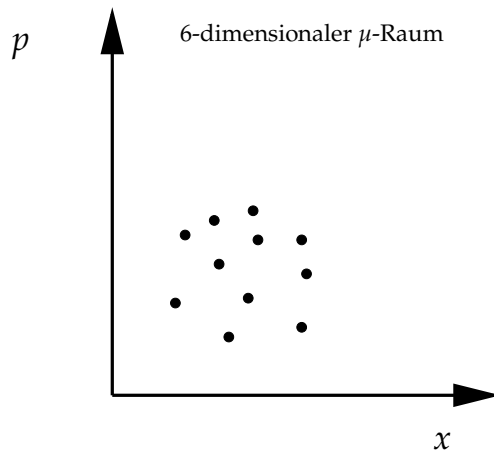


Abb. 1.1 Formale Darstellung eines Systems aus mehreren Massenpunkten im 6-dimensionalen μ -Raum. Jeder Punkt repräsentiert einen Massenpunkt mit jeweils drei Ortskoordinaten und drei Impulskordinaten

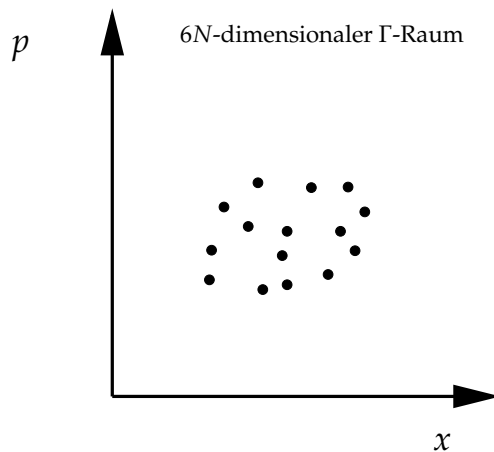


Abb. 1.2 Formale Darstellung eines Ensembles von Punktesystemen im $6N$ -dimensionalen Γ -Raum. Jeder Punkt repräsentiert jetzt ein System aus N Massenpunkte mit jeweils drei Ortskoordinaten und drei Impulskordinaten

spiel legt nahe, dass man ein multivariables Ereignis grundsätzlich auch als Kombination von elementareren Ereignissen verstehen kann. So könnte man das Ereignis, ein System im Punkt $\vec{\Gamma}$ des Phasenraums zu finden, auch als ein aus den Elementarereignissen Teilchen 1 in (x_1, p_1) des μ -Raums *und* Teilchen 2 in (x_2, p_2) des μ -Raums *und ...und* Teilchen N in (x_N, p_N) des μ -Raums zusammengesetzt denken. Aber auch diese Ereignisse ließen sich noch wei-

ter „atomisieren“, etwa durch Zerlegung in die einzelnen Koordinaten. Es ist stets eine Frage der Interpretation, ob man eine Wahrscheinlichkeitsdichte der Form $\rho(a_1, a_2, \dots, a_n)$ auf multivariable Einzelereignisse (der Dimension n) oder auf die simultane Realisierung von n atomaren, durch die Variablen a_i gekennzeichneten, Ereignissen beziehen will.

Interessiert man sich nicht für die spezielle Realisierung bestimmter atomarer Ereignisse, dann kann man diese aus der Wahrscheinlichkeitsdichte kombinierter Ereignisse durch Summation bzw. Integration über die entsprechenden Variablen eliminieren. So findet man die reduzierten Wahrscheinlichkeiten durch

$$\rho(a_1, a_2, \dots, a_{n-1}) = \int \rho(a_1, a_2, \dots, a_{n-1}, a_n) da_n \quad (1.15a)$$

$$\begin{array}{ccc} \vdots & \vdots & \vdots \\ \rho(a_1) & = & \int \rho(a_1, a_2, \dots, a_{n-1}, a_n) da_2 \dots da_{n-1} da_n \end{array} \quad (1.15b)$$

Letztendlich bedeutet z. B. die erste Gleichung, dass wir in $\rho(a_1, a_2, \dots, a_{n-1})$ alle kombinierten Ereignisse $(a_1, a_2, \dots, a_{n-1}, a_n)$ berücksichtigen, deren atomare Ereignisse $(a_1, a_2, \dots, a_{n-1})$ fest vorgegeben sind und bei denen das atomare Ereignis a_n einen beliebigen zulässigen Wert annimmt.

Auf diese Weise kann man aus der Dichte $\rho(x, p)$ eines Teilchens im μ -Raum auf die Wahrscheinlichkeitsdichte im Orts- bzw. im Impulsraum durch Integration über die Impuls- bzw. Ortskoordinaten gelangen

$$\rho(x) = \int \rho(x, p) d^3p \quad \text{bzw.} \quad \rho(p) = \int \rho(x, p) d^3x \quad (1.16)$$

Ebenso ist es möglich, aus der Wahrscheinlichkeitsverteilung $\rho(\vec{\Gamma})$ die Wahrscheinlichkeitsdichte $\rho(x, p)$ durch Integration über die Variablen aller übrigen Teilchen zu gewinnen.

Neben den Wahrscheinlichkeitsdichten einfacher und kombinierter Ereignisse spielen *bedingte Wahrscheinlichkeiten* bzw. Wahrscheinlichkeitsdichten eine wichtige Rolle. Darunter versteht man, dass das Auftreten eines Ereignisses an eine Bedingung geknüpft ist. Diese Bedingung wird formal durch einen Strich von dem eigentlich beschriebenen Ereignis getrennt. So bedeutet $\rho(a_1|a_2)$ die Wahrscheinlichkeitsdichte für das Auftreten des Ereignisses a_1 unter der Bedingung, dass das Ereignis a_2 sicher vorliegt. Kombinierte und bedingte Wahrscheinlichkeiten sind direkt miteinander verbunden. Es gilt

$$\rho(a_1, a_2) = \rho(a_1|a_2)\rho(a_2) \quad (1.17)$$

Dieses Theorem ist anschaulich leicht zu verstehen. Es gilt sowohl für Wahrscheinlichkeitsdichten als auch bei diskreten Realisierungen für Wahrscheinlichkeiten, wobei man in diesem Fall $p(a_1, a_2) = p(a_1|a_2)p(a_2)$

schreiben würde. Man kann sich diese Relation selbst leicht auf der Basis kombinatorischer Überlegungen, etwa unter Verwendung der Häufigkeitsdefinition, erklären.

Beispiel

Wir wollen diesen Zusammenhang anhand eines einfachen Beispiels erläutern. Dazu nehmen wir ein aus zwei Elementarereignissen a_1 und a_2 bestehendes kombiniertes Ereignis an. Jedes atomare Ereignis hat nur die Realisierungen $+$ und $-$. Nach $M = 100$ Experimenten findet man die in Tab. 1.2 dargestellten Resultate. Hieraus kann man sofort die Häufigkeiten bestimm-

Tab. 1.2 Beispiel: Anzahl des Auftretens der kombinierten (+,-)-Ereignisse. Aus der Tabelle folgt außerdem $m(a_1 = +) = 50$, $m(a_1 = -) = 50$, $m(a_2 = +) = 70$ und $m(a_2 = -) = 30$.

a_1	a_2	$m(a_1, a_2)$
+	+	40
+	-	10
-	+	30
-	-	20

men. Es ist $h(a_1 = +) = 1/2$, $h(a_1 = -) = 1/2$, $h(a_2 = +) = 7/10$ und $h(a_2 = -) = 3/10$. Außerdem findet man für die kombinierten Ereignisse $h(a_1 = +, a_2 = +) = 2/5$, $h(a_1 = +, a_2 = -) = 1/10$, $h(a_1 = -, a_2 = +) = 3/10$ und $h(a_1 = -, a_2 = -) = 1/5$. Um schließlich die bedingten Häufigkeiten zu berechnen, darf man nur die Fälle beachten, die jeweils die gegebene Bedingung erfüllen. So ist $h(a_1 = + | a_2 = +) = 4/7$ und $h(a_1 = - | a_2 = +) = 3/7$, aber $h(a_1 = + | a_2 = -) = 1/3$ und $h(a_1 = - | a_2 = -) = 2/3$. Damit kann man nun auch sofort (1.17) bestätigen. So ist

$$h(a_1 = +, a_2 = +) = h(a_1 = + | a_2 = +)h(a_2 = +) = \frac{4}{7} \times \frac{7}{10} = \frac{2}{5} \quad (1.18a)$$

$$h(a_1 = +, a_2 = -) = h(a_1 = + | a_2 = -)h(a_2 = -) = \frac{1}{3} \times \frac{3}{10} = \frac{1}{10} \quad (1.18b)$$

$$h(a_1 = -, a_2 = +) = h(a_1 = - | a_2 = +)h(a_2 = +) = \frac{3}{7} \times \frac{7}{10} = \frac{3}{10} \quad (1.18c)$$

$$h(a_1 = -, a_2 = -) = h(a_1 = - | a_2 = -)h(a_2 = -) = \frac{2}{3} \times \frac{3}{10} = \frac{1}{5} \quad (1.18d)$$

Da formal als Argumente der kombinierten Wahrscheinlichkeit elementare Ereignisse auftreten, die nur in ihrer Kombination¹³ zu verstehen sind, spielt deren Reihenfolge keine Rolle. Deshalb können wir auch schreiben

$$\rho(a_1, a_2) = \rho(a_2, a_1) = \rho(a_2|a_1)\rho(a_1) \quad (1.19)$$

und gelangen deshalb mit (1.17) zu der wichtigen Relation

$$\rho(a_1|a_2)\rho(a_2) = \rho(a_2|a_1)\rho(a_1) \quad (1.20)$$

Natürlich können sowohl das freie Ereignis a_1 als auch die Bedingung a_2 in $\rho(a_1|a_2)$ durch kombinierte (multivariable) Ereignisse ersetzt werden, so dass z. B. Ausdrücke der Form $\rho(a_1, \dots, a_m|a_{m+1}, \dots, a_{m+n})$ als bedingte Wahrscheinlichkeitsdichten auftreten können. In diesem Fall erweitert sich (1.17) zu

$$\rho(a_1, \dots, a_m, a_{m+1}, \dots, a_{m+n}) = \rho(a_1, \dots, a_m|a_{m+1}, \dots, a_{m+n})\rho(a_{m+1}, \dots, a_{m+n}) \quad (1.21)$$

Häufig werden in der statistischen Physik Systeme formal in Subsysteme (Teilsysteme) zerlegt. Da bei dieser oft willkürlichen Zerlegung die vorhandene Kopplung unverändert bleibt, bleibt auch die gegenseitige Einflussnahme der Subsysteme bestehen. Deshalb kann z. B. der durch $\vec{\Gamma}_1$ beschriebene mechanische Zustand des einen Subsystems vom Zustand des anderen Systems beeinflusst werden. Dann ist die bedingte Wahrscheinlichkeitsdichte $\rho(\vec{\Gamma}_1|\vec{\Gamma}_2)$ verschieden von der (unbedingten) Wahrscheinlichkeitsdichte $\rho(\vec{\Gamma}_1)$, die nur Aussagen über die Wahrscheinlichkeitsverteilung der mechanischen Zustände des ersten Systems macht und dabei die Zustände des zweiten Systems gar nicht berücksichtigt.

Auch für bedingte Wahrscheinlichkeitsdichten gilt die *Normierungsregel*. So ist

$$\int \rho(a_1|a_2) da_1 = 1 \quad (1.22)$$

oder allgemeiner

$$\int \rho(a_1, \dots, a_m|a_{m+1}, \dots, a_{m+n}) da_1 \dots da_m = 1 \quad (1.23)$$

Eine analoge Regel gilt im diskreten Fall natürlich auch für die bedingten Wahrscheinlichkeiten.

1.5.1

Zeitabhängige Wahrscheinlichkeiten und Wahrscheinlichkeitsdichten

Wahrscheinlichkeiten und Wahrscheinlichkeitsdichten können auch von der Zeit abhängig sein. Betrachtet man zum Beispiel die Wahrscheinlichkeitsverteilung für die Atome eines Gases in einem zeitlich veränderlichen Volumen,

¹³) also z. B. Realisierung von a_1 und a_2

dann ist – ohne die explizite Struktur dieser Verteilungsfunktion jetzt schon zu kennen – sofort klar, dass die funktionale Struktur dieser Verteilungsfunktion zeitabhängig sein muss. Liegt keine Zeitabhängigkeit vor, dann spricht man von *Stationarität* oder in besonderen Fällen auch vom *thermischen Gleichgewicht*. Wann genau welcher der beiden Begriffe anzuwenden ist, werden wir erst später klären.

Wir wollen aber jetzt schon bemerken, dass auch in stationären Systemen oder im thermischen Gleichgewicht Zeitabhängigkeiten auftreten können. So kann man in einer Flüssigkeit ein Teilchen beobachten und die kombinierte Wahrscheinlichkeit $\rho(x_1, t_1; x_2, t_2; \dots; x_r, t_r)$ angeben. Diese beschreibt jetzt die Wahrscheinlichkeit, das Teilchen zur Zeit t_1 am Ort x_1 , zur Zeit t_2 am Ort x_2 usw. zu finden. Auch hier ist bereits intuitiv klar, dass mit abnehmenden Zeitabständen die Wahrscheinlichkeitsdichte nur für solche Sätze (x_1, x_2, \dots, x_r) essentielle Beiträge liefert, die räumlich möglichst nahe beieinander stehen.

Ebenso können bedingte Wahrscheinlichkeiten bzw. Wahrscheinlichkeitsdichten in stationären Systemen zeitabhängig sein. So ist z. B. in dem eben benutzten Beispiel $\rho(x_1, t_1 | x_0, t_0)$ die Wahrscheinlichkeitsdichte dafür, dass das beobachtete Teilchen sich zur Zeit t_1 am Ort x_1 befindet unter der Bedingung, dass es zur Zeit t_0 im Ort x_0 war. Auch hier ist sofort verständlich, dass mit abnehmender Zeitdifferenz $t_1 - t_0$ die bedingte Wahrscheinlichkeitsdichte immer schärfer um den Punkt x_0 lokalisiert ist und für $t_1 \rightarrow t_0$ sogar gilt¹⁴

$$\lim_{t_1 \rightarrow t_0} \rho(x_1, t_1 | x_0, t_0) = \delta(x_1 - x_0) \quad (1.24)$$

Stationarität bedeutet bei solchen Zeitabhängigkeiten stets, dass die Wahrscheinlichkeitsdichten gegenüber beliebigen Translationen der Zeit invariant sind. So liegt z. B. Stationarität vor, wenn für ein beliebiges τ

$$\rho(x_1, t_1; x_2, t_2; \dots; x_r, t_r) = \rho(x_1, t_1 + \tau; x_2, t_2 + \tau; \dots; x_r, t_r + \tau) \quad (1.25)$$

gilt. Analoge Forderungen gelten für bedingte Wahrscheinlichkeiten, also z. B.

$$\rho(x_1, t_1 | x_0, t_0) = \rho(x_1, t_1 + \tau | x_0, t_0 + \tau) \quad (1.26)$$

Einzig im Fall der einfachen unbedingten Wahrscheinlichkeit $\rho(x, t)$ führt die Stationaritätsforderung sofort auf $\rho(x, t) = \rho(x)$, also auf die vollständige zeitliche Unabhängigkeit der Wahrscheinlichkeitsverteilung.

14) Das Teilchen kann nicht zur gleichen Zeit an zwei Orten sein.

1.5.2

Repräsentative Werte1.5.2.1 **Erwartungswerte**

Der empirische Mittelwert einer Folge von Werten ist die Summe dieser Werte geteilt durch die Anzahl der Werte. An dieser Definition orientiert sich auch die exakte Definition des Erwartungswertes. Betrachtet man wieder wie zu Beginn von Abschnitt 1.4.1 ein Experiment mit N verschiedenen Ausgängen A_1, A_2, \dots, A_N , die jeweils bei insgesamt M Wiederholungen des Experiments m_1 -mal, m_2 -mal, \dots , m_N -mal gemessen wurden, dann ist der empirische Mittelwert gegeben durch

$$\bar{A} = \frac{1}{M} \sum_{l=1}^N m_l A_l = \sum_{l=1}^N \frac{m_l}{M} A_l = \sum_{l=1}^N h_l A_l \quad (1.27)$$

wobei wir im letzten Schritt die Häufigkeit (1.3) eingeführt haben. Führt man jetzt noch den Grenzübergang $M \rightarrow \infty$ aus, dann gelangt man mit (1.5) zu der Basisdefinition des *Erwartungswertes* diskreter Ereignisse

$$\bar{A} = \sum_{l=1}^N p_l A_l \quad (1.28)$$

Man kann diese Definition sofort auf kontinuierliche Variable übertragen. Betrachten wir z. B. wieder ein Teilchen eines Systems im μ -Raum, dann können die Erwartungswerte beliebiger Funktionen dieser Variablen bestimmt werden. Im Vergleich zu der Definition im diskreten Fall ist dabei nur die Summation durch eine entsprechende Integration und die Wahrscheinlichkeit durch die Wahrscheinlichkeitsdichte zu ersetzen

$$\bar{A} = \int A \rho(A) dA \quad (1.29)$$

Der Erwartungswert ist ein repräsentatives Maß für die Realisierungen der durch die Wahrscheinlichkeiten p_l oder die Wahrscheinlichkeitsdichten $\rho(A)$ charakterisierten Zufallsgröße A . In der statistischen Physik wird der in (1.28) bzw. (1.29) definierte Erwartungswert auch oft als Mittelwert bezeichnet, vor allem dann, wenn eine Verwechslung mit dem auf Messergebnissen beruhenden empirischen Mittelwert ausgeschlossen ist. Im Sprachgebrauch des vorliegenden Bandes werden wir deshalb sowohl den Begriff *Mittelwert* als auch den Begriff *Erwartungswert* für die durch (1.28) und (1.29) definierten Größen verwenden.

Natürlich kann man auch die Erwartungswerte beliebiger Funktionen der Realisierungen von A bestimmen. So ist im diskreten Fall

$$\overline{f(A)} = \sum_{l=1}^N p_l f(A_l) \quad (1.30)$$

und für kontinuierliche Ereignisse

$$\overline{f(A)} = \int f(A)\rho(A)dA \quad (1.31)$$

So gilt z. B. für den Erwartungswert der über dem μ -Raum erklärten Funktion $f(x, \mathbf{p})$

$$\overline{f(x, \mathbf{p})} = \int \rho(x, \mathbf{p})f(x, \mathbf{p})d^3p d^3x \quad (1.32)$$

In analoger Weise kann man auch die Erwartungswerte im ganzen Phasenraum bestimmen. Solche Mittelwerte sind sinnvoll für Größen, die das Gesamtsystem beschreiben, z. B. die innere Energie des Systems. Hier würde man dann

$$\overline{f(\vec{\Gamma})} = \int \rho(\vec{\Gamma})f(\vec{\Gamma})d\vec{\Gamma} \quad (1.33)$$

verwenden.

1.5.2.2 Wahrscheinlichkeitsverteilung für abhängige Variable

Da die Variable A zufällige Ereignisse repräsentiert, ist auch jede nichtkonstante Funktion $f(A)$ eine Zufallsgröße, die aber aus statistischer Sicht nicht unabhängig von A ist. Es liegt deshalb nahe, eine Verteilungsfunktion $\rho(F)$ für $F = f(A)$ zu finden. Das ist vor allem dann sinnvoll, wenn A eine multivariable Größe, z. B. der Phasenraumvektor $\vec{\Gamma}$ eines Systems, F dagegen ein skalarer Wert, z. B. die innere Energie des Systems, ist.

Im diskreten Fall muss man zur Bestimmung von $p(F)$ alle Ereignisse A_l , die über $f(A_l)$ zu dem Wert F führen, zusammenzählen. Das liefert uns dann sofort

$$p(F) = \sum_{l=1}^N p_l \delta_{F, f(A_l)} \quad (1.34)$$

Analog findet man für Wahrscheinlichkeitsdichten

$$\rho(F) = \int \delta(F - f(A))\rho(A)dA \quad (1.35)$$

Wir machen darauf aufmerksam, dass bei dieser Schreibweise $\rho(A)$ und $\rho(F)$ normalerweise völlig unterschiedliche Funktionen sind. Als ein typisches Beispiel geben wir die Wahrscheinlichkeitsverteilung einer skalaren Variable F an, die sich als Funktion $f(\vec{\Gamma})$ der Position eines Systems im Phasenraum ergibt

$$\rho(F) = \int d\Gamma \rho(\vec{\Gamma})\delta(F - f(\vec{\Gamma})) = \overline{\delta(F - f(\vec{\Gamma}))} \quad (1.36)$$

Unter Verwendung von (1.35) kann man sich leicht davon überzeugen, dass auch $\rho(F)$ normiert ist

$$\int \rho(F)dF = \int dF \int dA \delta(F - f(A))\rho(A) = \int \rho(A)dA = 1 \quad (1.37)$$

und ferner, wie es auch sein muss, der Erwartungswert von F mit $\overline{f(A)}$ übereinstimmt

$$\bar{F} = \int F\rho(F)dF = \int dF \int dA F\delta(F - f(A))\rho(A) \quad (1.38a)$$

$$= \int f(A)\rho(A)dA = \overline{f(A)} \quad (1.38b)$$

Mit (1.35) ist es jederzeit möglich, komplizierte, multivariable Wahrscheinlichkeitsverteilungen auf einfachere Strukturen zurückzuführen, die sich leichter mit experimentellen Ergebnissen vergleichen und interpretieren lassen.

1.5.2.3 Median und wahrscheinlichster Wert

Neben dem Erwartungswert einer Zufallsgröße A gibt es noch zwei weitere, oft zur Charakterisierung eines Experiments herangezogene repräsentative Werte einer Wahrscheinlichkeitsverteilung. Der *Median* $A_{1/2}$ ist so definiert, dass 50% aller Realisierungen der Größe A kleiner als $A_{1/2}$ sind und dementsprechend auch 50% aller Realisierungen größer sind. Daher ist

$$\int_{-\infty}^{A_{1/2}} \rho(A)dA = \frac{1}{2} \quad (1.39)$$

Als untere Grenze dieser Definitionsgleichung wurde hier $-\infty$ gewählt. Ist der Definitionsbereich von A nach unten beschränkt, dann muss man (1.39) entsprechend modifizieren.

Der *wahrscheinlichste Wert* A_w entspricht dem globalen Maximum der Wahrscheinlichkeitsverteilung. Damit ist als notwendige Bedingung verknüpft, dass A_w die Extremwertgleichung

$$\frac{d\rho(A)}{dA} = 0 \quad (1.40)$$

erfüllt. Die Kenntnis von Median und wahrscheinlichstem Wert ist für die Einschätzung der Aussagekraft experimenteller Ergebnisse oft von großer Bedeutung. Ist z. B. der wahrscheinlichste Wert sehr viel kleiner als der Erwartungswert, dann kann man bei einer experimentellen Überprüfung der Resultate davon ausgehen, dass die entscheidenden Beiträge zum Erwartungswert hauptsächlich von sehr großen, dafür aber relativ selten auftretenden Ereignissen rühren. Mit großer Wahrscheinlichkeit befindet sich unter einer kleinen Menge von Messergebnissen kein einziger dieser großen Werte, da die meisten Messungen zu Resultaten in der Nähe des wahrscheinlichsten Wertes führen. Folglich fällt in der Regel der aus wenigen Messdaten bestimmte empirische Mittelwert viel kleiner als der theoretisch bestimmte Erwartungswert aus. Da der empirische Wert mit wachsender Anzahl von Messungen allmählich gegen den Erwartungswert streben muss, kann zudem der Eindruck

entstehen, dass das Experiment nicht stabil ist oder in der untersuchten Probe nichtstationäre Prozesse ablaufen. Gerade in solchen Fällen zeigt sich, dass eine sorgfältige Analyse experimenteller Ergebnisse besonders wichtig ist.

1.5.3

Schwankungen

1.5.3.1 Varianz

Die im vorangegangenen Abschnitt diskutierten repräsentativen Werte geben zwar in vielen Fällen wichtige Auskünfte über die Zufallsvariable, sie geben aber keine Vorstellung davon, wie stark die Realisierungen der Zufallsvariablen um diese Größen schwanken. Ein einfaches und relativ praktikables Maß zur Charakterisierung der Schwankungsbreite ist die *Varianz*, definiert durch

$$(\delta A)^2 = \overline{(A - \bar{A})^2} = \overline{A^2} - \bar{A}^2 = \int (A - \bar{A})^2 \rho(A) dA \quad (1.41)$$

Aus der Varianz berechnet man die *relative Schwankung* oder *Streuung*

$$s_A = \frac{\delta A}{\bar{A}} \quad (1.42)$$

die ein Maß für die Abweichung der Zufallsgröße A von ihrem statistischen Mittelwert ist. Je kleiner sie ist, desto seltener befindet sich ein System in einem Zustand für den die Größe A wesentlich von ihrem Mittelwert abweicht.

1.5.3.2 Spread

Ein alternatives Maß zur Bestimmung der Schwankungen einer Zufallsgröße ist der *Spread*

$$D_A = \int |A - A_{1/2}| \rho(A) dA \quad (1.43)$$

Dieses Maß ist vor allem dann sinnvoll, wenn die Varianz einer Zufallsgröße nicht existiert. Das ist insbesondere dann der Fall, wenn die Wahrscheinlichkeitsverteilung $\rho(A)$ für $|A| \rightarrow \infty$ langsamer als $|A|^{-3}$ gegen 0 konvergiert. Falls die Konvergenz aber noch stärker als $|A|^{-2}$ ist, können immerhin Erwartungswert und Spread berechnet werden, so dass man zur Bestimmung einer relativen Schwankung anstelle δA auch D_A heranziehen kann.

1.5.4

Statistische Unabhängigkeit, Korrelationen

1.5.4.1 Statistische Unabhängigkeit

Zwei Zufallsgrößen A und B heißen *statistisch unabhängig*, falls die kombinierte Wahrscheinlichkeitsdichte $\rho(A, B)$ in ein Produkt separiert werden kann

$$\rho(A, B) = \rho(A)\rho(B) \quad (1.44)$$

Dabei können A und B sowohl einfache als auch multivariable Ereignisse repräsentieren. Die Definition (1.44) der statistischen Unabhängigkeit ist wegen (1.21) äquivalent zu der Aussage

$$\rho(A|B) = \rho(A) \quad (1.45)$$

d. h. es liegt statistische Unabhängigkeit vor, wenn die Wahrscheinlichkeit der Realisierung eines Ereignisses nicht durch die Präsenz von Bedingungen verändert wird. Sind die Ereignisse A und B statistisch unabhängig, dann zerfallen die Erwartungswerte der Produkte $f(A)g(B)$ stets in Produkte der Erwartungswerte von $f(A)$ und $g(B)$

$$\begin{aligned} \overline{f(A)g(B)} &= \int \rho(A,B)f(A)g(B)dAdB = \int \rho(A)\rho(B)f(A)g(B)dAdB \\ &= \int \rho(A)f(A)dA \int \rho(B)g(B)dB = \overline{f(A)} \overline{g(B)} \end{aligned} \quad (1.46)$$

Wegen dieser Eigenschaft können statistisch unabhängige Systeme getrennt behandelt werden, was oft zu einer beträchtlichen Reduktion des zur Behandlung eines speziellen Problems notwendigen Arbeitsaufwandes führt.

1.5.4.2 Kovarianz und linearer Korrelationskoeffizient

Ein geeignetes Maß zur Charakterisierung der statistischen Abhängigkeit von Zufallsgrößen ist der Korrelationskoeffizient. Dieser basiert auf der *Kovarianz*, die für zwei Elementarereignisse A und B wie folgt definiert ist

$$C_{AB} = \overline{(A - \bar{A})(B - \bar{B})} = \int dA dB (A - \bar{A})(B - \bar{B})\rho(A,B) \quad (1.47)$$

Wegen (1.46) verschwindet C_{AB} , wenn A und B statistisch unabhängig sind. Die Definition (1.47) lässt sich auch auf die Kovarianz von Funktionen der Zufallsgrößen erweitern

$$C_{FG} = \overline{(f(A) - \overline{f(A)})(g(B) - \overline{g(B)})} \quad (1.48)$$

Da $F = f(A)$ und $G = g(B)$ selbst Zufallsgrößen sind, kann man mit dem in Abschnitt 1.5.2.2 beschriebenen Verfahren (siehe auch Abschnitt 1.5.4.3) aus $\rho(A,B)$ eine Wahrscheinlichkeitsdichte $\rho(F,G)$ erzeugen und mit dieser den Korrelationskoeffizienten entsprechend (1.47) bestimmen.

Die Kovarianz ist so definiert, dass Fluktuationen um den jeweiligen Mittelwert in die gleiche Richtung positive Beiträge, entgegengesetzte Fluktuationen jedoch negative Beiträge liefern. Als Nachteil erweist sich jedoch, dass die Kovarianz nicht normiert ist und so schlecht zu Vergleichen herangezogen werden kann. So kann eine kleine positive Kovarianz z. B. dadurch entstehen, dass große Fluktuationen sowohl mit gleicher als auch entgegengesetzter Richtung auftreten und sich die Beiträge gegenseitig fast eliminieren, oder

dass zwei völlig synchron laufende Prozesse mit relativen kleinen Fluktuationen um den Mittelwert vorliegen. Um dieses Defizit zu beheben, führt man den Korrelationskoeffizienten ein

$$\kappa_{AB} = \frac{C_{AB}}{\delta A \delta B} \quad (1.49)$$

Verwendet man die reduzierten Zufallsgrößen

$$\tilde{A} = \frac{A - \bar{A}}{\delta A} \quad \text{und} \quad \tilde{B} = \frac{B - \bar{B}}{\delta B} \quad (1.50)$$

dann kann man anstelle (1.49) auch schreiben

$$\kappa_{AB} = \overline{\tilde{A}\tilde{B}} \quad (1.51)$$

Man überzeugt sich schnell davon, dass $C_{AA} = (\delta A)^2$ und $C_{BB} = (\delta B)^2$ gilt, woraus sofort $\kappa_{AA} = \kappa_{BB} = 1$ folgt. Andererseits kann (1.51) auch als Skalarprodukt zwischen \tilde{A} und \tilde{B} verstanden werden. Wegen $\kappa_{AA} = \kappa_{BB} = 1$ handelt es sich dabei sogar um ein Skalarprodukt zwischen zwei jeweils auf 1 normierten Größen, d. h. es muss daher immer

$$-1 \leq \kappa_{AB} \leq +1 \quad (1.52)$$

gelten. Andererseits ist der Korrelationskoeffizient invariant gegenüber linearen Transformationen. So führt die Abbildung $A \rightarrow A' = \lambda A + \alpha$ mit $\lambda > 0$ zu keiner Änderung des Korrelationskoeffizienten. Zum Beweis betrachtet man zunächst den Mittelwert, für den $\bar{A}' = \lambda \bar{A} + \alpha$ gilt. Folglich ist $A' - \bar{A}' = \lambda(A - \bar{A})$ und deshalb auch $\delta A' = \lambda \delta A$. Hieraus erhalten wir dann die Invarianz der reduzierten Zufallsgröße $\tilde{A}' = \tilde{A}$ und damit auch die Invarianz des Korrelationskoeffizienten. Wegen dieser Eigenschaft wird der Korrelationskoeffizient auch oft als linearer Korrelationskoeffizient bezeichnet. Man spricht für $\kappa_{AB} > 0$ allgemein von positiver Korrelation und verbindet damit eine mehr oder weniger starke Synchronität der Fluktuationen der beiden Zufallsgrößen. Umgekehrt liegt für $\kappa_{AB} < 0$ eine Antikorrelation mit überwiegend entgegengerichtet verlaufenden Fluktuationen vor. Im Fall $\kappa_{AB} = 0$ sind die Fluktuationen unkorreliert oder asynchron. Im Fall statistisch unabhängiger Zufallsvariablen erhält man stets $\kappa_{AB} = 0$.

1.5.4.3 Wahrscheinlichkeitsverteilung funktional abhängiger Variablen

Wir betrachten jetzt zwei durch die Funktionen $f(A)$ und $g(A)$ derselben Zufallsvariablen¹⁵ gebildete Zufallsgrößen F und G . Sind $f(A)$ und $g(A)$ keine linearen Funktionen von A , dann ist der Korrelationskoeffizient κ_{FG} nicht

¹⁵ Dabei muss A kein Skalar sein, sondern kann durchaus multivariable Ereignisse beschreiben.

mehr zwingend ± 1 , sondern kann beliebige Werte im Intervall $[-1, +1]$ annehmen. Deshalb ist es sinnvoll, die Wahrscheinlichkeitsdichte für die Verteilung der funktional abhängigen Variablen F und G zu bestimmen. Dazu nutzt man wieder die δ -Funktion als eine Art Filter

$$\rho(F, G) = \int dA \delta(F - f(A)) \delta(G - g(A)) \rho(A) \quad (1.53)$$

Auch hier ist die Normierbarkeit wieder gesichert. Sinnvoll ist diese Methode wieder dann, wenn mit A ein multivariables Ereignis hoher Dimension verbunden ist und die funktionale Struktur von f und g dementsprechend kompliziert wird. In diesem Fall ist es besser, statt der Basisverteilung $\rho(A)$ die einfachere, zweikomponentige Verteilungsfunktion $\rho(F, G)$ zur Diskussion heranzuziehen, auch wenn diese nur für die spezielle Struktur der Funktionen f und g gültig ist. Ist A z. B. der mechanischen Zustand eines Vielteilchensystems, d. h. $A = \vec{\Gamma}$, dann bekommen wir für (1.53)

$$\rho(F, G) = \int d\Gamma \delta(F - f(\vec{\Gamma})) \delta(G - g(\vec{\Gamma})) \rho(\vec{\Gamma}) \quad (1.54)$$

Mit (1.53) kann auch sofort die Kovarianzmatrix entsprechend

$$\begin{aligned} C_{FG} &= \int dF dG (F - \bar{F})(G - \bar{G}) \rho(F, G) \\ &= \int dA dF dG (F - \bar{F})(G - \bar{G}) \delta(F - f(A)) \delta(G - g(A)) \rho(A) \\ &= \int dA (f(A) - \bar{F})(g(A) - \bar{G}) \rho(A) \end{aligned} \quad (1.55)$$

wobei \bar{F} durch

$$\begin{aligned} \bar{F} &= \int dF dG F \rho(F, G) \\ &= \int dA dF dG F \delta(F - f(A)) \delta(G - g(A)) \rho(A) \\ &= \int dA f(A) \rho(A) \end{aligned} \quad (1.56)$$

und \bar{G} analog darstellbar ist.

1.5.5

Das Gesetz der großen Zahlen

Ein makroskopisches Vielteilchensystem besteht aus einer riesigen Anzahl $N \gg 1$ von Partikeln. Wollen wir jetzt für dieses System die gesamte kinetische Energie bestimmen, dann setzt sich diese additiv aus den kinetischen Energien der einzelnen Partikel zusammen. Sind die Partikel alle gleich, dann

sollte auch der Mittelwert der kinetischen Energie jedes Teilchens den gleichen Wert \bar{E} haben. Somit ist der Mittelwert der kinetischen Energie aller Teilchen $N\bar{E}$. Ferner kann man bei den meisten Vielteilchensystemen davon ausgehen, dass die Kovarianz der kinetischen Energien unterschiedlicher Teilchen verschwindet, d. h. die kinetischen Energien der einzelnen Teilchen sind statistisch voneinander unabhängig¹⁶.

Bei einem entsprechenden Experiment wird aber nicht die mittlere kinetische Energie, sondern die Summe der tatsächlichen Realisierungen, also

$$E_{\text{ges}} = \sum_{i=1}^N E_i \quad (1.57)$$

bestimmt. Entscheidend ist jetzt die Frage, wie genau man E_{ges} messen kann. Normalerweise nimmt bei einem Experiment die absolute Messgenauigkeit mit wachsender Größe des Messwertes zu, die relative Messgenauigkeit bleibt dagegen näherungsweise erhalten. Wir nehmen deshalb an, die Messgenauigkeit werde durch die dimensionslose Größe $\lambda \ll 1$ in der Weise bestimmt, dass Messfehler der Größenordnung

$$\Delta E = \lambda E_{\text{ges}} \quad (1.58)$$

nicht mehr wahrgenommen werden können. Wir betrachten jetzt die Differenz

$$e_N = \frac{E_{\text{ges}}}{N\bar{E}} - 1 = \frac{1}{\bar{E}} \frac{1}{N} \sum_{i=1}^N (E_i - \bar{E}) \quad (1.59)$$

Die auf der rechten Seite auftretende Summe

$$s_N = \frac{1}{\bar{E}} \frac{1}{N} \sum_{i=1}^N (E_i - \bar{E}) \quad (1.60)$$

kann als empirischer Mittelwert über alle Fluktuationen $E_i - \bar{E}$ verstanden werden, wobei die E_i selbst identisch verteilt sind, d. h. wegen der vorausgesetzten Gleichheit der Partikel auch die gleiche Wahrscheinlichkeitsverteilung besitzen und die Kovarianz der Energien unterschiedlicher Teilchen verschwindet. Dann gilt das *schwache Gesetz der großen Zahlen* [27], nämlich

$$\lim_{N \rightarrow \infty} p(|s_N| > \varepsilon) = 0 \quad (1.61)$$

d. h. die Wahrscheinlichkeit, dass $|s_N|$ größer als ein beliebig kleiner Wert ε wird, verschwindet für $N \rightarrow \infty$. Mit anderen Worten, mit Sicherheit wird für $N \rightarrow \infty$ der Wert von $|s_N|$ gegen null streben. Dann ist aber auch

$$\frac{|E_{\text{ges}} - N\bar{E}|}{N\bar{E}} = s_N < \varepsilon \quad (1.62)$$

16) Es handelt sich hierbei um eine Annahme, die sehr gut für Gase und mit geringen Einschränkungen auch für Flüssigkeiten zutrifft, dagegen für die Atome eines Kristalls weniger sinnvoll ist.

wobei für ein hinreichend großes N der Wert von ε beliebig klein, also insbesondere kleiner als λ gewählt werden kann. Folglich ist

$$|E_{\text{ges}} - N\bar{E}| < \varepsilon N\bar{E} \approx \varepsilon E_{\text{ges}} < \lambda E_{\text{ges}} = \Delta E \quad (1.63)$$

d. h. die Abweichung zwischen dem Messwert E_{ges} der kinetischen Energien aller Teilchen und dem zugehörigen Erwartungswert fällt unter den Messfehler. Mithin wird das Experiment innerhalb seiner Messgenauigkeit stets $N\bar{E}$ messen. Das bedeutet letztendlich, dass in einem hinreichend großen Vielteilchensystem mögliche und sicher auch vorhandene Fluktuationen experimentell nicht registriert werden können.

Natürlich könnte man einwenden, dass die vorliegende Argumentation wesentlich auf der verschwindenden Kovarianz zwischen den kinetischen Energien der Teilchen beruht. Diese Forderung wird nicht streng auf andere Größen übertragbar sein. Deshalb werden wir in den Abschnitten 3.3.3 und 7.9 mehr physikalisch motivierte Begründungen für die Tatsache geben, dass auf makroskopischen Skalen Fluktuationen praktisch irrelevant werden.

Ergänzend sei noch bemerkt, dass unter den gegebenen Voraussetzungen auch das strengere *starke Gesetz der großen Zahlen*

$$p\left(\lim_{N \rightarrow \infty} \sup |s_N| = 0\right) = 1 \quad (1.64)$$

anwendbar gewesen wäre.

