

# Kapitel 1

# Mathematische Grundlagen der Statistischen Mechanik

*Thomas Filk; 30.03.2024*

Wie schon der Name andeutet, handelt es sich bei der Statistischen Mechanik um eine „statische“ Theorie, d.h., man interessiert sich für die gemittelten Eigenschaften von Systemen mit vielen Freiheitsgraden bzw. von einem Ensemble von vielen gleichartigen Systemen. Die Thermodynamik ist eine phänomenologische Theorie für den Grenzfall sehr vieler (im Idealfall unendlich vieler) Freiheitsgrade bzw. Systemkomponenten. In diesem Grenzfall wird unter sehr allgemeinen Bedingungen das Verhalten von Mittelwerten exakt.

Das zentrale mathematische Rüstzeug für die Statistische Mechanik umfasst somit die Wahrscheinlichkeitstheorie bzw. die Statistik, und der Schwerpunkt beruht speziell auf dem Verhalten von Kenngrößen von Wahrscheinlichkeitsverteilungen im Grenzfall unendlich großer Systeme (oder von Systemen mit unendlich vielen Freiheitsgraden). Dieses Werkzeug soll in diesem Kurzttext kurz angerissen werden.

## 1.1 Wahrscheinlichkeiten

Das Thema „Wahrscheinlichkeit“ wurde zu einem Zweig der Mathematik, als man in der Zeit zwischen dem 15. und 17. Jahrhundert mehr und mehr erkannte, dass sich die relativen Häufigkeiten, mit der bestimmte Ereignisse bei Glücksspielen auftreten, mathematisch bestimmen lassen. Nach langen Schwierigkeiten bei der Suche nach einer Definition, was genau Wahrscheinlichkeit eigentlich sei, konnte Andrei Nikolajewitsch Kolmogorow (1903–1987) 1933 den Begriff der Wahrscheinlichkeit axiomatisch fassen. Diese Formulierung umgeht die philosophische Frage nach dem Wesen der Wahrscheinlichkeit und ersetzt sie durch einen Satz von Bedingungen, die von einer mathematischen Struktur erfüllt sein müssen, um von Wahrscheinlichkeiten im herkömmlichen Sinne sprechen zu können.

Die Hauptschwierigkeit eines solchen Formalismus bezieht sich auf die mathematisch exakte Beschreibung von Wahrscheinlichkeiten bei kontinuierlichen Ereignismengen. Als Beispiel sei die Wahrscheinlichkeit genannt, auf einer (mathematisch idealisierten) Dart-Scheibe mit einer (mathematisch idealisierten) Pfeilspitze einen (mathematisch idealisierten) Punkt zu treffen. Die erste Schwierigkeit besteht darin, dass wir elementaren Ereignissen keine von null verschiedenen Wahrscheinlichkeiten zuschreiben können, sondern nur Ereignismengen (beispielsweise einem Flächenausschnitt der Dart-Scheibe). Die zweite, für die Physik allerdings meist weniger wichtige Schwierigkeit liegt dar-

in, dass manchen Teilmengen des Ereignisraums überhaupt keine Wahrscheinlichkeiten (weder null noch von null verschieden) zugeschrieben werden können. Heute zählt die Wahrscheinlichkeitstheorie in der Mathematik zur Maßtheorie. Allerdings spielt gerade bei endlichen Ereignisräumen oft die Kombinatorik eine wichtigere Rolle.

### 1.1.1 Ereignisse als Teilmengen eines Ereignisraumes

Wenn wir im Alltag im Zusammenhang mit Wahrscheinlichkeiten von einem Ereignis sprechen, meinen wir meist sogenannte *Elementarereignisse*, die sich auf eine einzelne konkrete Realisierung beziehen. Bekannte Beispiele sind das Ereignis, mit einem Würfel bei einem Spiel eine bestimmte Zahl zu würfeln oder bei einem Münzwurf „Kopf“ zu erhalten. Haben wir Grund zu der Annahme, dass jedes solche Ereignis gleich wahrscheinlich ist, können wir die Wahrscheinlichkeit für ein solches Ereignis bestimmen, wenn wir die Gesamtzahl aller Ereignisse kennen. Beim Würfel gibt es sechs Zahlen, daher ist die Wahrscheinlichkeit für das Auftreten einer bestimmten Zahl (bei einem „ehrlichen“ Würfel) gleich  $1/6$ . Bei einer Münze gibt es zwei mögliche Elementarereignisse und wenn die Münze „fair“ ist, sollte jedes Ereignis mit der Wahrscheinlichkeit  $1/2$  auftreten.

In vielen Fällen verstehen wir unter einem Ereignis aber auch eine ganze Klasse von Elementarereignissen, beispielsweise wenn wir bei einem Würfelspiel mit zwei Würfeln nach der Wahrscheinlichkeit fragen, ein „Mäxle“ (eine Eins und eine Zwei) zu würfeln oder eine bestimmte Gesamtaugenanzahl zu erreichen (dies ist z.B. bei dem Spiel „Siedler von Catan“ wichtig). Das Ereignis „Gesamtaugenanzahl Sieben“ besteht somit aus mehreren Elementarereignissen:  $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ , wobei  $(p,q)$  das Elementarereignis bezeichnet, mit dem ersten Würfel (der z.B. durch seine Farbe markiert sein kann) die Zahl  $p$  und mit dem zweiten Würfel die Zahl  $q$  zu würfeln. Sind die Elementarereignisse gleich wahrscheinlich, erhalten wir die Wahrscheinlichkeit für ein solches zusammengesetztes Ereignis nach der *Laplace'schen Formel*

$$\text{Wahrscheinlichkeit} = \frac{\text{Menge der günstigen Ereignisse}}{\text{Menge der möglichen Ereignisse}}. \quad (1.1)$$

Bei kontinuierlichen Ereignismengen (z.B. dem oben angesprochenen mathematisch idealisierten Dart-Spiel) können wir den Einzelereignissen im Allgemeinen nur noch die Wahrscheinlichkeiten null zuschreiben, obwohl es nicht unmöglich ist, einen bestimmten Punkt zu treffen. In einem solchen Fall können wir nur Mengen von Elementarereignissen eine nicht-verschwindende Wahrscheinlichkeit zuschreiben. Diese Mengen bestehen auch nicht aus endlich vielen oder abzählbar unendlich vielen einzelnen Punkten, sondern sind typischerweise zusammenhängende Gebiete.

### 1.1.2 Borel-Mengen bzw. Borel-Algebren

Während man in der Physik in den meisten Fällen damit zufrieden ist, dass solche Gebiete ein endliches Maß haben (je nach Ereignismenge eine endliche Länge bzw. Dauer, Fläche, Volumen, etc.), möchte man in der Mathematik diese Gebiete genauer charakterisieren können. Insbesondere zeigt sich, dass es Teilmengen von kontinuierlichen Mengen gibt, denen man überhaupt kein sinnvolles Maß zuordnen kann.

Zur Umgehung dieses Problems definiert man zu einer *Ereignismenge*  $\Omega$  eine bestimmte Klasse von Teilmengen, die bezüglich der Bildung von Komplementen sowie abzählbaren Vereinigungen abgeschlossen ist. Eine solche Menge von Teilmengen bezeichnet man als  $\sigma$ -Algebra.<sup>1</sup> Eine  $\sigma$ -Algebra bezeichnet man als Borel-Menge oder Borel-Algebra, wenn sie die offenen Teilmengen von  $\Omega$  enthält. Insbesondere enthält  $\sigma$  in diesem Fall auch  $\Omega$  selbst und die leere Menge.

<sup>1</sup>Der Begriff der Algebra mag hier überraschen. In der Mathematik bezeichnet man ganz allgemein eine Menge (oder einen Satz von Mengen), auf der mehr als eine Verknüpfungsrelation definiert ist, als eine Algebra.

Die mathematischen Details bei allgemeinen topologischen Räumen spielen hier keine Rolle. Bei den reellen Zahlen bzw. dem  $\mathbb{R}^n$  sind die relevanten Teilmengen alle Mengen, die man aus Intervallen bzw. offenen Kugeln durch abzählbar viele Vereinigungen und Bildung von Komplementmengen erhält. Damit sind pathologische Fälle, beispielsweise die Äquivalenzklassen von reellen Zahlen, deren Differenz eine rationale Zahl ist, ausgeschlossen. Im Folgenden spreche ich manchmal von „sinnvollen“ Teilmengen, wenn die Elemente einer Borel-Algebra gemeint sind.

### 1.1.3 Das Axiomensystem von Kolmogorow

Das Kolmogorow'sche Axiomensystem beschreibt die Voraussetzungen, um von Wahrscheinlichkeiten sprechen zu können.

*Definition:* Ein Tripel  $(\Omega, \sigma, \omega)$  heißt Wahrscheinlichkeitsraum, wenn  $\sigma$  eine Borel-Menge von  $\Omega$  ist (also eine Menge von sinnvollen Teilmengen von  $\Omega$ ) und  $\omega : \sigma \rightarrow \mathbb{R}$  eine Abbildung, die folgende Bedingungen erfüllt:

$$\begin{aligned} \omega(\Omega) &= 1 \\ \omega(A) &\geq 0 \quad \text{für alle } A \in \sigma \\ \omega\left(\bigcup_i A_i\right) &= \sum_i \omega(A_i) \quad \text{sofern } A_i \in \sigma \text{ und } A_i \cap A_j = \emptyset \ (i \neq j). \end{aligned}$$

$\omega$  heißt Wahrscheinlichkeitsfunktion auf  $\sigma$ .

Die erste Bedingung bedeutet, dass die Wahrscheinlichkeit für das Ereignis  $\Omega$  (also dass irgendein Elementarereignis aus  $\Omega$  realisiert wird) gleich 1 ist. Die zweite Bedingung bedeutet, dass Wahrscheinlichkeiten positive Zahlen sind. Die nicht-triviale Bedingung ist die dritte Bedingung: Seien  $\{A_i\}$  höchstens abzählbar viele und paarweise disjunkte Teilmengen von  $\sigma$ , dann ist die Wahrscheinlichkeit für das Eintreten eines Elementarereignisses in irgendeiner dieser Ereignismengen gleich der Summe der Wahrscheinlichkeiten, dass dieses Ereignis in einer bestimmten dieser Mengen liegt. Anschaulich bedeutet es: Die Wahrscheinlichkeit für das Auftreten eines von mehreren unabhängigen Ereignissen ist gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse. Aus diesem Gesetz ergibt sich z.B. unmittelbar die oben erwähnte Laplace'sche Formel.

## 1.2 Zufallsvariable und Kenngrößen

Die Natur des Ereignisraumes  $\Omega$  kann sehr unterschiedlich sein. In den meisten Fällen möchten wir jedoch den Elementarereignissen Zahlen zuordnen und dann die Frage stellen können, mit welcher Wahrscheinlichkeit bestimmte Zahlen auftreten.

Dazu definieren wir das Konzept einer *Zufallsvariablen*:

*Definition:* Eine *Zufallsvariable*  $X$  zu einem Wahrscheinlichkeitsraum  $(\Omega, \sigma, \omega)$  ist eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$ , sodass das Urbild jeder messbaren Menge aus  $\mathbb{R}$  in  $\sigma$  liegt.

Konkret bedeutet dies: Die Menge aller Elementarereignisse  $a \in \Omega$ , für die  $X(a)$  in einem bestimmten Intervall in  $\mathbb{R}$  liegt, ist eine Menge, der man eine Wahrscheinlichkeit zuschreiben kann. Zu jeder Zufallsvariablen  $X$  definieren wir nun eine *Verteilungsfunktion*  $P_X$  auf  $\mathbb{R}$ :

$$P_X : \mathbb{R} \rightarrow [0,1] \quad \text{mit} \quad P_X(x) = \omega\left(\{a \in \Omega \mid X(a) < x\}\right). \quad (1.2)$$

$P(x)$  ist somit die Wahrscheinlichkeit, dass bei einer Realisierung ein Ereignis  $a$  auftritt, für das die Zufallsvariable  $X(a)$  einen Wert kleiner als  $x$  hat. Formal können wir nun einer Zufallsvariablen eine

Wahrscheinlichkeitsdichte zuordnen, indem wir die Ableitung von  $P(x)$  betrachten:

$$w_X(x) = \frac{dP_X}{dx}. \quad (1.3)$$

Allerdings ist diese Wahrscheinlichkeitsdichte nicht immer eine gewöhnliche Funktion auf den reellen Zahlen sondern eher eine Distribution, d.h., sinnvoll sind nur Integrale über messbare Teilmengen von  $\mathbb{R}$ .

Wie in der Physik oft üblich, werde ich nicht immer explizit von einem Wahrscheinlichkeitsraum  $(\Omega, \sigma, \omega)$  sowie einer Zufallsvariablen  $X$  und ihrer Verteilungsfunktion  $P_X$  sprechen, sondern einfacher von einer (reellen) Variablen  $x$  und ihrer Wahrscheinlichkeitsdichte  $w(x)$ . Ganz grob möchte ich die Zusammenhänge angeben, mit denen wir es in der Statistischen Mechanik zu tun haben werden:  $\Omega$  wird der Raum aller Mikrozustände sein, das ist klassisch der Phasenraum aller beteiligten Teilchen und in der Quantenmechanik meist die Menge der Eigenzustände zum Energieoperator.  $\sigma$  besteht aus (sinnvollen) Teilmengen im Phasenraum und  $\omega$  ist eine Wahrscheinlichkeit (in der QM) oder eine Wahrscheinlichkeitsdichte (über dem klassischen Phasenraum), die angibt, mit welcher Wahrscheinlichkeit ein bestimmter Mikrozustand vorliegt. Eine Zufallsvariable ist dann eine Observable – beispielsweise die Energie – auf dem Phasenraum: Sie ordnet jedem Mikrozustand eine Zahl zu.  $w$  ist die Wahrscheinlichkeitsdichte für das Auftreten eines bestimmten Messwerts, wenn an einem konkreten System diese Observable gemessen wird.

Bei einem diskreten System ( $\Omega$  endlich oder abzählbar) besteht  $\sigma$  aus allen Teilmengen von  $\Omega$  und enthält damit auch die Elementarereignisse  $i \in \Omega$ . Damit genügt es bei diskreten Systemen, für jedes Elementarereignis  $i$  seine Wahrscheinlichkeit  $\omega(i)$  anzugeben. Die Wahrscheinlichkeit für ein Ereignis  $A \subset \Omega$  ist dann einfach:

$$\omega(A) = \sum_{i \in A} \omega(i). \quad (1.4)$$

Da es sich nun um diskrete Ereignisse handelt, wird die Verteilung  $P$  zu einer Stufenfunktion (die an einer Stelle  $x$  um einen diskreten Sprung zunimmt, wenn es ein Ereignis  $a$  gibt, sodass  $X(a) = x$ ). Die Ableitung  $w(x)$  wird damit zu einer Summe über  $\delta$ -Funktionen. Das ist auch richtig, wenn man die Summe über diskrete Ereignisse als Integral schreiben möchte. Im Allgemeinen ist das aber nicht notwendig: Man spricht von Wahrscheinlichkeiten und berechnet diese durch die Bildung von Summen über diskrete Ereignismengen.

Ist eine Wahrscheinlichkeitsdichte gegeben, können wir verschiedene Kenngrößen bestimmen. Diese charakterisieren eine Verteilung und reichen oftmals für einfache Zusammenhänge aus. Für eine allgemeine (messbare, d.h. lokal integrierbare) Funktion  $f(x)$  kann man den Erwartungswert berechnen:

$$\langle f(x) \rangle = \int f(x)w(x) dx. \quad (1.5)$$

Besondere Erwartungswerte sind der *Mittelwert*

$$\bar{x} := \langle x \rangle = \int xw(x) dx \quad (1.6)$$

und die *Varianz*

$$\sigma_x^2 := \langle (x - \bar{x})^2 \rangle = \int (x - \bar{x})^2 w(x) dx. \quad (1.7)$$

Die Wurzel aus der Varianz, also  $\sigma_x$ , bezeichnet man als *Standardabweichung*. Allgemein sind die *Momente* der Verteilung durch

$$\langle x^n \rangle = \int x^n w(x) dx \quad (1.8)$$

gegeben, für die man auch die *erzeugende Funktion*

$$\tilde{f}(k) = \langle e^{ikx} \rangle = \int e^{ikx} w(x) dx \quad (1.9)$$

definiert. Die Momente erhält man aus den Ableitungen von  $\tilde{f}$ :

$$\langle x^n \rangle = (-i)^n \left. \frac{d^n \tilde{f}(k)}{dk^n} \right|_{k=0}. \quad (1.10)$$

Bis auf Faktoren ist  $\tilde{f}(k)$  die Fourier-Transformierte der Wahrscheinlichkeitsdichte  $w(x)$ .

Bei mehreren Zufallsvariablen  $X_i$  ( $i = 1, \dots, n$ ) definiert man eine gemeinsame Verteilungsfunktion

$$P(x_1, \dots, x_n) = \omega \left( \{a \in \Omega \mid X_1(a) < x_1, X_2(a) < x_2, \dots, X_n(a) < x_n\} \right), \quad (1.11)$$

und durch Ableitungen nach  $x_i$  ergibt sich die gemeinsame Wahrscheinlichkeitsdichte

$$w(x_1, \dots, x_n) = \frac{d^n P(x_1, \dots, x_n)}{dx_1 \cdots dx_n}. \quad (1.12)$$

Daraus erhält man z.B. die *Kovarianzmatrix*

$$C_{ij} = \int (x_i - \bar{x}_i)(x_j - \bar{x}_j) w(x_1, \dots, x_n) d^n x, \quad (1.13)$$

die ein guter Indikator für Korrelationen zwischen den beiden Variablen  $x_i$  und  $x_j$  ist. Die Diagonale der Kovarianzmatrix enthält die Varianzen zu den einzelnen Größen.

Gilt für die Wahrscheinlichkeitsdichte von zwei Größen  $x$  und  $y$  die Beziehung

$$w(x, y) = w(x)w(y), \quad (1.14)$$

so bezeichnet man  $x$  und  $y$  als *statistisch unabhängig*. In diesem Fall faktorisieren alle Erwartungswerte von Produkten der beiden Größen zu Produkten von den entsprechenden Erwartungswerten. Insbesondere ist die Kovarianzmatrix eine Diagonalmatrix, d.h., die Nicht-Diagonalelemente sind null. Allerdings gilt im Allgemeinen die Umkehrung nicht: Wenn die Kovarianzen verschwinden bedeutet das nicht notwendigerweise, dass die zugehörigen Größen statistisch unabhängig sind.

### 1.2.1 Die Gauß-Verteilung

Unter den kontinuierlichen Wahrscheinlichkeitsverteilungen spielt die Gauß-Verteilung eine besondere Rolle. Sie ist durch die Angabe ihres Mittelwerts  $\bar{x}$  und ihrer Standardabweichung  $\sigma_x$  bzw. ihrer Varianz  $\sigma_x^2$  festgelegt:

$$w(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma_x^2}\right). \quad (1.15)$$

Allgemeiner kann man für eine positive symmetrische Matrix  $A_{ij}$  (positiv bedeutet in diesem Fall, dass sämtliche Eigenwerte positiv sind) die Verteilungsfunktion

$$w(x_1, \dots, x_n) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^n (x_i - \bar{x}_i) A_{ij} (x_j - \bar{x}_j)\right) \quad (1.16)$$

definieren. Die Kovarianzmatrix dieser Verteilung ist

$$C_{ij} = (A^{-1})_{ij}, \quad (1.17)$$

wobei  $A^{-1}$  die inverse Matrix zu  $A$  ist.

## 1.3 Grenzwertsätze

In der Wahrscheinlichkeitstheorie unterscheidet man viele Arten von Grenzwertsätzen, die häufig auch noch in starker und schwacher Form vorkommen können. Ihnen allen gemein ist, dass Aussagen über die statistischen Eigenschaften von Funktionen von sehr vielen Zufallsvariablen (im Grenzfall unendlich vieler Zufallsvariablen) gemacht werden. An dieser Stelle möchte ich nur auf zwei Grenzwertsätze eingehen, und auch bei diesen nur eine schwache Version angeben, die für physikalische Anwendungen ausreicht: das „Gesetz der großen Zahlen“ und den „zentralen Grenzwertsatz“. Beide Grenzwertsätze machen Aussagen zu der Wahrscheinlichkeitsverteilung der Summe bzw. dem Mittelwert von vielen Zufallszahlen.

### 1.3.1 Das Gesetz der großen Zahlen

Das Gesetz der großen Zahlen besagt, dass es für den Mittelwert von Zufallsvariablen zunehmend (mit wachsender Anzahl der Zufallsvariablen) unwahrscheinlicher wird, von dem Durchschnitt der Mittelwerte dieser Zufallsvariablen wesentlich abzuweichen. Konkretisiert werden nun die Begriffe „zunehmend unwahrscheinlicher“ und „wesentlich abweichen“.

Generell gilt für die Summe von zwei Zufallsvariablen  $S = X_1 + X_2$ , dass ihr Mittelwert gleich der Summe der Mittelwerte von  $X_1$  und  $X_2$  ist. Das folgt unmittelbar aus der Definition des Mittelwerts und den Normierungseigenschaften der Wahrscheinlichkeitsdichten:

$$\bar{s} = \int (x_1 + x_2)w(x_1, x_2) d^2x \quad (1.18)$$

$$= \int x_1 w(x_1, x_2) dx_1 dx_2 + \int x_2 w(x_1, x_2) dx_1 dx_2 = \bar{x}_1 + \bar{x}_2. \quad (1.19)$$

Im Folgenden betrachten wir den Mittelwert  $M$  und die einfache Summe  $S$  von  $N$  Zufallsvariablen:

$$M = \frac{1}{N} \sum_i X_i \quad \text{und} \quad S = \sum_i X_i, \quad (1.20)$$

die sich nur in dem Normierungsfaktor  $1/N$  unterscheiden. Beide Größen sind selbst wieder Zufallsvariable. Für ihre Mittelwerte gilt allgemein:

$$\bar{m} = \frac{1}{N} \sum_i \bar{x}_i \quad \text{und} \quad \bar{s} = \sum_i \bar{x}_i. \quad (1.21)$$

Bisher haben wir noch keine Annahme über die Varianzen oder die statistische Unabhängigkeit der Zufallsvariablen gemacht. Nun betrachten wir einen Satz  $X_1, \dots, X_N$  von  $N$  Zufallsvariablen, die verschiedene Verteilungsfunktionen  $P_i(x_i)$  und damit verbundene Wahrscheinlichkeitsdichten  $w_i(x_i)$  mit zugehörigen Mittelwerten  $\bar{x}_i$  haben können, deren Varianzen  $\sigma_{x_i}^2$  aber endlich und durch eine Konstante beschränkt sein sollen. Außerdem nehmen wir an, dass die Zufallsvariablen statistisch unabhängig sein sollen. Das bedeutet, die gemeinsame Wahrscheinlichkeitsdichte faktorisiert:

$$w(x_1, \dots, x_N) = w_1(x_1)w_2(x_2) \cdots w_N(x_N). \quad (1.22)$$

Die eigentliche Aussage des Gesetzes der großen Zahlen ist, dass die Wahrscheinlichkeit, dass der Mittelwert von  $N$  Zufallszahlen um mehr als ein vorgegebenes  $\epsilon$  von dem Mittelwert der Mittelwerte abweicht, im Grenzfall  $N \rightarrow \infty$  gegen null geht. Es wird also für genügend große Werte von  $N$  beliebig unwahrscheinlich, dass der Mittelwert der Summe um mehr als eine beliebig kleine vorgegebene Konstante vom Mittelwert der Mittelwerte abweicht. Dieses Gesetz der großen Zahlen ist gleichzeitig ein Beweis für die häufig axiomatisch eingeführte Behauptung, dass die relative Häufigkeit bei genügend vielen Realisierungen eines Ereignisses gegen die Wahrscheinlichkeit für dieses Ereignis geht.

Für den Beweis machen wir die vereinfachende Annahme, dass alle Mittelwerte verschwinden; dies lässt sich immer durch eine additive Verschiebung der Zufallsvariablen erreichen. Damit verschwinden nicht nur die Einzelmittelwerte  $\bar{x}_i = 0$  sondern auch der Gesamtmittelwert  $\bar{m} = 0$ . Weiterhin seien alle Varianzen  $\sigma_i$  durch eine Konstante  $C$  beschränkt, also  $\sigma_i \leq C$ . Wir betrachten nun die Varianz des Mittelwerts der Zufallsvariablen:

$$\sigma_m^2 = \left\langle \left( \frac{1}{N} \sum_i x_i \right)^2 \right\rangle = \frac{1}{N^2} \left\langle \sum_{i,j} x_i x_j \right\rangle = \frac{1}{N^2} \sum_{i,j} \langle x_i x_j \rangle. \quad (1.23)$$

Nun nutzen wir die Unabhängigkeit der Zufallsvariablen aus, d.h. für  $i \neq j$  gilt  $\langle x_i x_j \rangle = 0$ . Damit erhalten wir:

$$\sigma_m^2 = \frac{1}{N^2} \sum_i \langle x_i^2 \rangle = \frac{1}{N^2} \sum_i \sigma_i^2 \leq \frac{1}{N} C. \quad (1.24)$$

Im Grenzfalle  $N \rightarrow \infty$  geht die rechte Seite gegen null, d.h., die Varianz zum Mittelwert der Zufallsvariablen wird in diesem Grenzfalle kleiner als jedes vorgegebene  $\epsilon$ . Für die Summe der Zufallsvariablen erhalten wir entsprechend:

$$\sigma_s^2 = \sum_i \langle x_i^2 \rangle = \sum_i \sigma_i^2. \quad (1.25)$$

Hinter dieser Formulierung des Gesetzes der großen Zahlen stecken mehrere wichtige Anwendungen. Die Gleichsetzung von relativer Häufigkeit und Wahrscheinlichkeit haben wir schon erwähnt. Für die Standardabweichung (Wurzel aus der Varianz) der Summe von Zufallsvariablen finden wir nach obiger Herleitung:

$$\sigma_s = \sqrt{\sum_i \sigma_i^2}. \quad (1.26)$$

Dies ist das bekannte Fehlerfortpflanzungsgesetz: Setzt sich der Gesamtfehler bei einer Messung als Summe von Einzelfehlern zusammen, die als unkorreliert angenommen werden können, dann ergibt sich der Gesamtfehler aus der Summe der Quadrate der Einzelfehler.

### 1.3.2 Der zentrale Grenzwertsatz

Während das Gesetz der großen Zahlen eine Aussage über die Varianz des Mittelwerts bzw. der Summe von vielen Zufallszahlen macht, geht der zentrale Grenzwertsatz noch einen Schritt weiter: Er macht eine Aussage über die Verteilungsfunktion der Summe bzw. des Mittelwerts von Zufallszahlen:

Für  $N$  statistisch unabhängige Zufallszahlen  $X_i$  mit Varianzen  $\sigma_i$ , die alle endlich und durch eine Konstante beschränkt sein sollen, wird die Verteilungsfunktion des Mittelwerts (bzw. der Summe) zu einer Gaußverteilung mit den durch das Gesetz der großen Zahlen gegebenen Mittelwerten und Standardabweichungen.

Etwas präziser behauptet der zentrale Grenzwertsatz: Für die normierte Variable

$$Z = \frac{S - Nm}{\sigma_m \sqrt{N}} \quad (1.27)$$

( $S$ ,  $m$  und  $\sigma_m$  wie oben) wird die Verteilungsfunktion im Grenzfalle  $N \rightarrow \infty$  zur Verteilungsfunktion der Standardnormalverteilung, also

$$\lim_{N \rightarrow \infty} P(Z < z) = \Phi(z) \quad (1.28)$$

wobei

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt \quad (1.29)$$

die Verteilungsfunktion der Standardnormalverteilung ist. (Die Standardnormalverteilung hat den Mittelwert 0 und die Standardabweichung 1).

Die herausragende Bedeutung der Normalverteilung (also der Gauß-Verteilung) ergibt sich unter anderem aus diesem Satz. Unabhängig von den Einzelverteilungen irgendwelcher Zufallsvariabler wird die Verteilung ihrer Summe zu einer Gauß-Verteilung. Die beiden getroffenen Annahmen – die statistische Unabhängigkeit der Zufallsvariablen und die Beschränktheit ihrer Varianzen – sind dabei wesentlich. Viele thermodynamische Größen sind die Summe mikroskopischer Größen. In manchen Fällen (beispielsweise an sogenannten Phasenübergängen) sind diese mikroskopischen Größen derart korreliert, dass der zentrale Grenzwertsatz nicht mehr gültig ist. Tatsächlich treten in diesen Fällen oft Nicht-Gauß'sche Verteilungsfunktionen auf.

Auch die Endlichkeit der Varianzen ist wichtig: Es gibt normierbare Wahrscheinlichkeitsdichten, beispielsweise die Lorentz- bzw. Cauchy-Verteilung

$$w(x) = \frac{1}{\pi} \frac{a}{x^2 + a^2}, \quad (1.30)$$

mit einem endlichen Mittelwert, deren Varianz aber unendlich ist. In diesen Fällen sind auch sehr große Schwankungen in den Realisationen der Zufallsvariablen nicht selten. Bei Zufallsvariablen mit solchen Verteilungen muss die Verteilung der Summe dieser Zufallsvariablen nicht gegen eine Gauß-Funktion gehen.

Eine Variante des zentralen Grenzwertsatzes ist die Brown'sche Bewegung bzw. der Random Walk. Die Zufallsvariablen  $X_i$  sind dabei Verteilungen für Einzelschritte, und die Wahrscheinlichkeitsdichte der Summe  $S$  beschreibt die Wahrscheinlichkeit, ein im Ursprung gestartetes Teilchen nach  $N$  Schritten an einem Punkt  $x$  bzw. in einem Abstand  $r$  zu finden.