

# Boltzmanns Gesetz $S = k \log W$

von Peter Eichelsbacher und Matthias Löwe, Bielefeld

**Zusammenfassung:** Wir erarbeiten uns die grundlegenden Aussagen aus Ludwig Boltzmanns Artikel "Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht" aus dem Jahre 1877 mit den Hilfsmitteln elementarer Mathematik. Wir zeigen die Analogie zwischen dem Boltzmannschen Entropiebegriff in einem einfachen  $N$ -Teilchen Modell und dem Shannonschen Entropiebegriff aus der Informationstheorie auf. Die im Beweis verwendeten Methoden erlauben einen kleinen Einblick in die sogenannte "Theorie der großen Abweichungen".

## 1 Einleitung

*"Niemand, auch nicht Maxwell und Gibbs, hat so tief über die Einseitigkeit der Naturvorgänge und ihre wahrscheinlichkeitstheoretische Begründung nachgedacht wie Boltzmann. Ein Großteil seiner Arbeit ist diesem Problem gewidmet."*

(Sommerfeld, 1944)

Wir wollen hier mit elementarer Mathematik die große Leistung L. Boltzmanns darstellen, den Zusammenhang zwischen Entropie und Wahrscheinlichkeit richtig erkannt zu haben.

In der klassischen Mechanik wird der Zustand eines Systems mehrerer Teilchen durch Punkte im Phasenraum beschrieben, in dem man die Orts- und Impulskoordinaten aufgeführt hat. Die Bewegung des Systems wird durch ein System gewöhnlicher Differentialgleichungen (Lagrange, Hamilton) beschrieben. Schon Avogadro wußte, daß die Teilchenzahl pro Mol in der Größenordnung von  $10^{23}$  Partikeln liegt. Dies führt zu einem Differentialgleichungssystem, das keiner vernünftigen Behandlung mehr zugänglich ist.

Die Thermodynamik, die in der Mitte des letzten Jahrhunderts entstand, hat das Ziel, das Verhalten eines Systems mit Hilfe makroskopischer Variablen, sogenannter Zustandsgrößen, z. B. Druck, Volumen, Temperatur,

innere Energie, oder die 1865 von R. Clausius eingeführte Entropie, zu beschreiben.

Die grundlegende Beobachtung von Clausius, gestützt auf Arbeiten Carnots, war, daß für einen reversiblen, d.h. zeitlich umkehrbaren, thermodynamischen Kreisprozeß das (notwendigerweise entlang einer geschlossenen Kurve verlaufende) Integral über die Änderung der Wärme  $dQ$  pro Temperatur  $T$  verschwindet, also in Formeln  $\oint \frac{dQ}{T} = 0$ . Mathematisch impliziert das (über eine Form des Hauptsatzes der Integral- und Differentialrechnung für Vektorfelder) die Existenz einer Stamm- oder Potentialfunktion für den Integranden, die eine Zustandsfunktion des zugrunde liegenden Systems ist. Diese Zustandsfunktion nannte Clausius nach dem griechischen  $\epsilon\nu\tau\rho\omicron\pi\eta$  (Umkehr) Entropie. Eine wesentliche Eigenschaft der Entropie ist, daß sie für nicht reversible Prozesse stets positiv ist (und für reversible Prozesse – wie oben erwähnt – verschwindet). Diese Beobachtung führte zur Formulierung des zweiten Hauptsatzes der Thermodynamik:

*Prozesse in einem abgeschlossenen thermodynamischen System verlaufen stets so, daß sich die Entropie des Systems vergrößert.*

Eine Begründung der thermodynamischen Gesetze auf der Basis der Atomhypothese liefert die statistische Mechanik. Deren Wurzeln wurden mit Hilfe der sich entwickelnden Wahrscheinlichkeitstheorie von L. Boltzmann und J.W. Gibbs gelegt.

Betrachten wir dazu eine Teilchenkonstellation zu einem festen Zeitpunkt, eine sogenannte Konfiguration. Boltzmann ordnete jeder Konfiguration eine Wahrscheinlichkeit zu (und zwar jeder Konfiguration die gleiche) und fragte wieviele Konfigurationen dasselbe makroskopische Bild liefern, also denselben Zustand beschreiben. Er erkannte, daß die wesentlichen Zustände, also diejenigen die man beobachtet, diejenigen mit maximaler Wahrscheinlichkeit sind, sogenannte Gleichgewichtszustände. Ein System tendiert stets zu seinem wahrscheinlichsten Zustand hin, um dann um ihn zu fluktuieren. Bereits 1872 beschrieb er das Verhältnis von Wahrscheinlichkeitstheorie und Mechanik mit den Worten:

*“Lediglich dem Umstand, daß selbst die regellosesten Vorgänge, wenn sie unter denselben Verhältnissen vor sich gehen, doch jedes Mal dieselben Durchschnittswerte liefern, ist es zuzuschreiben, daß wir auch im Verhalten warmer Körper ganz bestimmte Gesetze wahrnehmen. Denn die Moleküle der Körper sind ja so zahlreich und ihre Bewegungen so rasch, daß*

*uns nie etwas anderes als jene Durchschnittswerte wahrnehmbar sind. Die Bestimmung der Durchschnittswerte ist Aufgabe der Wahrscheinlichkeitsrechnung.”*

Im Mittelpunkt dieser Arbeit soll die Herleitung des Boltzmannschen Prinzips und der Entropie stehen, so wie sie Boltzmann in seiner Arbeit aus dem Jahre 1877 einführte. Diese ist für das Verständnis des Wesens der Entropie von großem Wert: Die Definition der Entropie nach Clausius ist noch höchst unanschaulich – noch heute stößt ja die Einführung des Entropiebegriffs beispielsweise in der Schulphysik auf Schwierigkeiten, und in der Schulmathematik wird er größtenteils gar nicht thematisiert (obwohl sich dies, wie wir uns gleich überzeugen wollen, durchaus anbietet). Andererseits ist mit dem zweiten Hauptsatz ein fundamentales Gesetz der Thermodynamik und der gesamten Physik gefunden. Diese Bedeutung erlangt es nicht zuletzt daher, daß es – im Gegensatz zu den Gesetzen der klassischen Mechanik, die invariant sind unter Zeitumkehr – eine Zeitrichtung auszeichnet und somit mit der Alltagserfahrung einer in einer bestimmten Richtung verlaufenden Zeit übereinstimmt.

Wir wollen im zweiten Abschnitt das anschaulichere Bild der Entropie nach Boltzmann vorstellen. Dieses soll zunächst in dem Begriffsapparat geschehen, den C.E. Shannon 1948 bei der Einführung der Entropie in die Mathematik, genauer die Informationstheorie verwandte. Als ein Nebenprodukt unserer Überlegungen erhalten wir somit die Übereinstimmung von Shannonschem und Boltzmannschem Entropiebegriff. In Kapitel 3 wollen wir Boltzmanns einfaches Modell eines idealen Gases vorstellen. Die elementaren Rechnungen zu diesem Modell, die neben einer Verifikation der Beziehung  $S = k \log W$  auch eine Beschreibung der Gleichgewichtszustände liefern, führen wir in Abschnitt 4 aus. Dabei benötigen wir nur den Begriff der konvexen Funktion, sowie eine einfache, erste Approximation für  $n!$ . Da unsere Rechnungen eine einfache Version eines Prinzips großer Abweichungen darstellen, werden wir in einem Anhang kurz auf die recht neue Theorie der großen Abweichungen, ein Teilgebiet der Wahrscheinlichkeitstheorie, eingehen.

## 2 Die Entropie nach Boltzmann und Shannon

Das Verständnis des Entropiebegriffs wurde wesentlich vorangetrieben durch L. Boltzmann. Zurückgehend im wesentlichen auf seine Definition gibt es

in der Mathematik mittlerweile ein Vielzahl unterschiedlicher Entropiebegriffe (wir verweisen auf den Übersichtsartikel Flachsmeyer, Ollech, 1995, der ein paar dieser Begriffe auf einem gut verständlichen Niveau vorstellt).

Boltzmanns wichtigster neuer Gedanke scheint uns die Idee zu sein, daß man für gewöhnlich die Zustände mit maximaler Entropie beobachtet und daß sich ein abgeschlossenes System in Richtung Zustände größerer Wahrscheinlichkeit verändert und schließlich um den wahrscheinlichsten, den sogenannten Gleichgewichtszustand, fluktuiert. Nun stellte er spekulativ den Zusammenhang zwischen Wahrscheinlichkeit und der Entropie nach Clausius her: auch die makroskopische Größe der Entropie strebt nach ihrem maximalen Wert. Da die Entropie von zwei Systemen gleich der Summe der einzelnen Entropien ist, und die Wahrscheinlichkeit der beiden Systeme im Falle der Unabhängigkeit multiplikativ ist, sollte der Zusammenhang zwischen Entropie und Wahrscheinlichkeit logarithmisch sein:

$$S = k \log W,$$

wobei  $S$  die Entropie des Systems ist,  $W$  seine Wahrscheinlichkeit und  $k$  schließlich ein Proportionalitätsfaktor, die sogenannte Boltzmannkonstante. Boltzmann bestimmte  $k$  anhand eines idealen Gases und erhielt den Wert  $k = 1,38 \cdot 10^{-23} \text{ J/K}$ . Die Boltzmannkonstante  $k$  ist eine fundamentale Naturkonstante. Der zweite Hauptsatz nimmt in der Boltzmannschen Formulierung ungefähr folgende Form an:

1. Die Energie eines abgeschlossenen Systems bleibt erhalten.
2. Im Gleichgewicht ist die Entropie des Systems maximal.

Da die statistische Mechanik auf die zu ihrer Entstehungszeit noch unbewiesene Atomhypothese zurückgreift, hatte sie anfänglich nur wenig Anhänger. Einer ihrer prominentesten Gegner, der Begründer der Quantenmechanik Max Planck, lieferte aber nach seiner "Bekehrung" wesentliche Argumente, um Boltzmanns Gesetz zu stützen. So schreibt er 1906 in der "Wärmestrahlung":

*"Welche mechanische oder elektrodynamische Größe stellt nun aber die Entropie eines Zustandes dar? Um diese fundamentale Frage zu beantworten, braucht man nur auf das Wesen des zweiten Hauptsatzes, des Prinzips der Vermehrung der Entropie etwas näher einzugehen. Wie die Thermodynamik lehrt, wird von zwei verschiedenen mit gleicher Energie ausgestat-*

*teten Zuständen eines physikalischen Systems derjenige in der Natur bevorzugt, der die größere Entropie besitzt, in dem Sinne, daß eine Umwandlung, ohne anderweitige Kompensationen, nur in der Richtung wachsender Entropie stattfinden kann. Bedenken wir andererseits, daß eine solche Umwandlung stets im Sinne des Ausgleichs bestehender Geschwindigkeitsdifferenzen (durch Reibung), Temperaturunterschiede (durch Wärmeleitung oder Wärmestrahlung), Konzentrationsunterschiede (durch Diffusion) erfolgt, so ergibt sich, daß die Größe der Entropie in demselben Maße wächst als dieser Ausgleich fortschreitet. Da nun aber, wie wir gesehen haben, der Begriff der Entropie eng verknüpft ist mit dem der elementaren Unordnung und dem Mangel jeglicher Einzelkontrolle, so kann die mit der Vermehrung der Entropie verbundene Tendenz zum Ausgleich vorhandener Differenzen für den mechanischen oder elektrodynamischen Beobachter kaum einen anderen Sinn haben als den, daß eine gleichmäßige Verteilung ungeordneter Elemente von vornherein wahrscheinlicher ist als jede andere Verteilung, und daß die Vermehrung der Entropie nichts anderes bedeutet als den Übergang in einen Zustand größerer Wahrscheinlichkeit."*

Im Modell des folgenden Kapitels wollen wir neben einer Herleitung des Boltzmannschen Prinzips noch einen weiteren Aspekt betrachten. Die Beschreibung des Gleichgewichtszustandes nach Boltzmann trägt noch nicht der mechanischen Gleichgewichtsbedingung der Minimierung der sogenannten potentiellen Energie Rechnung. Um diesen Themenkreis ein wenig zu diskutieren, geben wir eine endliche Menge von möglichen potentiellen Energien vor und fragen uns, welches die wahrscheinlichste Verteilung – also die Verteilung mit der größten Entropie – auf diesen Energien ist, vorausgesetzt die mittlere Energie ist auf einen bestimmten Wert festgesetzt. Die Lösung dieses Maximierungsproblems führt auf die sogenannte Gibbs-Verteilung, die grob gesprochen großen Energiewerten exponentiell kleine Wahrscheinlichkeiten zuweist.

Wir wollen nun den Ausdruck für die Entropie, so wie er im Modell von Boltzmann auftreten wird, vorstellen. Dafür machen wir einen kleinen Exkurs in ein anderes Teilgebiet der Mathematik: die Informationstheorie. Der Informationstheoretiker C.E. Shannon nämlich war es, der 1948 die Bedeutung des Boltzmannschen Entropiebegriffs erkannte und ihn als Maß für die Ungewißheit erstmals in einer mathematischen Theorie verwandte. Sein Entropiebegriff läßt sich in groben Zügen so motivieren: Angenommen wir haben einen Zeichensatz (in der Informationstheorie ist auch der Na-

me "Alphabet" geläufig, wir verwenden beide Begriffe synonym) bestehend aus  $r$  Zeichen oder Buchstaben  $\{s_1, \dots, s_r\}$  gegeben, und die Buchstaben treten in einem Text mit verschiedenen Wahrscheinlichkeiten  $p_1, \dots, p_r$  auf (bei dieser Modellierung wird vernachlässigt, daß diese Wahrscheinlichkeiten natürlich "lokal" stark variieren können, so ist z.B. nach einem  $c$  ein  $h$  sehr viel wahrscheinlicher als sonst). Wir wollen nun ein Maß dafür finden, wie überrascht wir sein müssen, wenn der Buchstabe  $s_k$ ,  $1 \leq k \leq r$  auftritt. Sicherlich gibt es viele mögliche Kandidaten für eine solche "Überraschungsfunktion" (quasi jede Funktion, die in  $1/p_k$  monoton ist). Eine davon – die dazu noch positiv ist – ist  $\log 1/p_k = -\log p_k$  (wobei bei uns  $\log$  stets den natürlichen Logarithmus bezeichnet, jedoch sind unsere Resultate weitgehend unabhängig von der Basis des Logarithmus). Der Mittelwert dieser Überraschung, also die mittlere Überraschung, ergibt sich dann als  $-\sum_{k=1}^r p_k \log p_k$ . Diese mittlere Überraschung ist der Shannonsche Entropiebegriff: Für eine Wahrscheinlichkeitsverteilung  $P = \{p_1, \dots, p_r\}$  auf einer  $r$ -elementigen Menge definiert man als

$$H(P) := -\sum_{k=1}^r p_k \log p_k$$

die Entropie von  $P$ .

Um das Wesen der Funktion  $H(\cdot)$  besser zu verstehen, soll hier kurz auf den Begriff der konvexen bzw. konkaven Funktionen eingegangen werden. Man nennt eine Funktion  $f : (a, b) \rightarrow \mathbb{R}$  konvex, wenn sie in jedem Punkt linksgekrümmt ist, und dies ist dann der Fall, wenn ihr Graph stets unterhalb der Sehne durch den Graphen verläuft, in Formeln

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \text{für } x, y \in (a, b) \text{ und } \lambda \in [0, 1].$$

Falls  $f$  zweimal stetig differenzierbar ist, so folgt aus der Verbindung zwischen Linkskrümmung und zweiter Ableitung, daß  $f$  genau dann konvex ist, falls  $f'' \geq 0$  (siehe z.B. Forster, 1983). Elementar ist nun die Jensensche Ungleichung für konvexe Funktionen  $f : (a, b) \rightarrow \mathbb{R}$ . Sie besagt, daß für beliebige  $x_1, \dots, x_n \in (a, b)$  und beliebige positive Zahlen  $\lambda_1, \dots, \lambda_n$  mit  $\lambda_1 + \dots + \lambda_n = 1$

$$f(\lambda_1 x_1 + \dots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \dots + \lambda_n f(x_n)$$

gilt.

Für  $n = 2$  ist diese Ungleichung nichts anderes als die Definition der Konvexität. Für  $n \geq 3$  folgt sie aus mehrfacher Anwendung dieser per Induktion. Beispielsweise gilt für den Fall  $n = 3$ :

$$\begin{aligned} f(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3) &\leq \lambda_1 f(x_1) + (1 - \lambda_1) f\left(\frac{\lambda_2 x_2 + \lambda_3 x_3}{1 - \lambda_1}\right) \\ &\leq \lambda_1 f(x_1) + (1 - \lambda_1) \frac{\lambda_2}{1 - \lambda_1} f(x_2) + (1 - \lambda_1) \frac{\lambda_3}{1 - \lambda_1} f(x_3) \\ &= \lambda_1 f(x_1) + \lambda_2 f(x_2) + \lambda_3 f(x_3). \end{aligned}$$

Man vergewissert sich schnell, daß  $H(\cdot)$  positiv und konkav ist (was nach Definition nichts anderes bedeutet, als daß  $-H(\cdot)$  konvex ist), indem man durch zweimaliges Ableiten überprüft, daß  $L(x) := x \log x$  konvex ist. Weiterhin gilt:  $H(P)$  ist genau dann maximal, wenn  $P$  die Gleichverteilung ist, und in diesem Fall gilt  $H(P) = \log r$ . Dies folgt aus der Jensenschen Ungleichung angewandt auf die Funktion  $L(x)$ :

$$\sum_{k=1}^r p_k \log \left(\frac{p_k}{q_k}\right) = \sum_{k=1}^r q_k L\left(\frac{p_k}{q_k}\right) \geq L(1) = 0$$

und daher

$$\sum_{k=1}^r p_k \log p_k \geq \sum_{k=1}^r p_k \log q_k. \quad (1)$$

Für  $q_k = 1/r$  folgt

$$-\sum_{k=1}^r p_k \log p_k \leq \sum_{k=1}^r p_k \log r = \log r$$

mit Gleichheit für  $p_k = 1/r$  für alle  $k = 1, \dots, r$ .

Eine Variante des Shannonschen Entropiebegriffs erhalten wir, wenn wir von  $H(P)$  den möglichen Maximalwert  $\log r$  abziehen und es somit quasi normieren. Wir definieren:

$$\begin{aligned} h(P) &:= H(P) - \log r \\ &= -\sum_{k=1}^r p_k \log p_k - \log r \\ &= -\left(\sum_{k=1}^r p_k \log p_k - \sum_{k=1}^r p_k \log 1/r\right) \\ &= -\sum_{k=1}^r p_k \log \frac{p_k}{1/r}. \end{aligned}$$

In der Literatur ist  $h(\cdot)$  als relative Entropie bzgl. der Gleichverteilung oder als Kullback-Leibler Information bekannt. Es mißt quasi (mit negativem Vorzeichen) den Abstand einer Verteilung  $P$  zur Gleichverteilung (ohne daß allerdings die Dreiecksungleichung erfüllt wäre). Aus den obigen Eigenschaften von  $H(\cdot)$  liest man leicht die Eigenschaften von  $h(\cdot)$  ab:

1.  $h$  ist negativ und konkav.
2.  $h(P)$  ist genau dann gleich 0, wenn  $P$  die Gleichverteilung ist.

Wir wollen nun zeigen, daß in einem einfachen Modell diese Variante des Shannonschen Entropiebegriffs im wesentlichen mit dem Boltzmannschen übereinstimmt.

### 3 Ein einfaches Modell eines idealen Gases

Viele physikalische Gesetzmäßigkeiten für reale Gase wurden zunächst an einem vereinfachten Modell, dem sogenannten idealen Gas hergeleitet. Beim idealen Gas handelt es sich insofern um eine Vereinfachung, als daß man sich dieses aus einer sehr großen Anzahl von Atomen oder Molekülen aufgebaut vorstellt, die Wechselwirkung unter diesen kleinsten Teilchen jedoch vernachlässigt.

Wir wollen dieses Modell noch etwas weiter vereinfachen, indem wir annehmen, daß das Gas in einem Behälter lebt, den wir in  $d$  Zellen unterteilt haben (sogenanntes "coarse graining"). Die Zellen seien so eingeteilt, daß innerhalb der Zellen eine physikalische Größe, beispielsweise die Energie, konstant ist. Für jedes Atom schauen wir nun nach, in welcher der Zellen es sich befindet. Setzen wir voraus, daß die einzelnen Atome ununterscheidbar sind, so läßt sich der Zustand des Systems beschreiben als ein  $d$ -Tupel natürlicher Zahlen  $(x_1, \dots, x_d)$ ,  $x_i \geq 0$  und  $\sum_{i=1}^d x_i = n$ . Hierbei ist  $x_i$  die Besetzungszahl der  $i$ -ten Zelle, d.h. die Anzahl der Atome dort, und  $n$  steht für die Gesamtzahl der Atome im Behälter. Uns interessiert nun die Wahrscheinlichkeit einer bestimmten Besetzung

$$P(X_1 = x_1, \dots, X_d = x_d),$$

wobei  $X_i$  für die Zufallsvariable steht, die die Besetzungszahl in der  $i$ -ten Zelle angibt. Die sogenannte Maxwell-Boltzmann-Statistik geht davon aus, daß jedes Teilchen in jede Zelle plaziert werden kann. Dann ist

$$P(X_1 = x_1, \dots, X_d = x_d) = \frac{n!}{x_1! \cdots x_d!} p_1^{x_1} \cdots p_d^{x_d},$$

wobei die  $p_i$  die Wahrscheinlichkeit beschreiben, daß ein Teilchen in die  $i$ -te Zelle fällt. Man sagt dann auch, daß  $(X_1, \dots, X_d)$  multinomialverteilt ist. Im Spezialfall, der Gleichverteilung der Energieniveaus, d.h. für den Fall, daß alle Zellen die gleiche Wahrscheinlichkeit  $1/d$  besitzen, reduziert sich der obige Ausdruck zu:

$$P(X_1 = x_1, \dots, X_d = x_d) = \frac{n!}{x_1! \cdots x_d!} \left(\frac{1}{d}\right)^n.$$

Wählt man hier nun speziell  $d = 2$ , zerlegt also den Behälter in zwei gleichwahrscheinliche Zellen, und interessiert sich für die Besetzungszahlen dort, so ist die Verteilung die wohlbekannte Binomialverteilung und das Modell ist äquivalent zum  $n$ -fachen fairen Münzwurf, einem der gut studierten stochastischen Modelle der Schulmathematik.

Wir möchten diesen Abschnitt mit einem Kommentar zu unserem Modell schließen:

Die Annahme der Multinomialverteilung der Besetzungszahlen, also insbesondere Unabhängigkeit der Wahl der Zellen für verschiedenen Teilchen, ist in der klassischen Physik gerechtfertigt, da beliebig viele Teilchen in jeder Zelle auftreten können. Für Elementarteilchen trifft diese Hypothese i.a. nicht zu. Unter Berücksichtigung des sogenannten Pauli-Verbots bzw. der Ununterscheidbarkeit der einzelnen Teilchen kommt man zu den sogenannten Fermi-Dirac bzw. Bose-Einstein-Statistiken (siehe auch Dinges, Rost, 1982, S.63).

### 4 Große Abweichungen und Boltzmanns Gesetz

Zunächst wollen wir für das in Kapitel 3 vorgestellte Modell eine Bestätigung der Boltzmannschen Beziehung  $S = k \log W$  herleiten. Wir werden dies tun, indem wir für große Teilchenzahlen (es sei daran erinnert, daß ein reales Gas in der Natur etwa  $10^{23}$  Teilchen pro Kubikzentimeter besitzt) den Logarithmus der Wahrscheinlichkeit berechnen, daß sich das System in einem bestimmten Zustand befindet. Wie schon weiter oben erwähnt, läßt sich der Zustand des Systems durch ein  $d$ -Tupel  $(x_1, \dots, x_d)$  kodifizieren. Da zusätzlich aber gelten muß  $\sum_{i=1}^d x_i = n$ , hängt somit nicht nur der Zustand selbst, sondern der gesamte Zustandsraum, d.h. die Menge aller möglichen Zustände, von der Teilchenzahl  $n$  ab. Dies ist für Limesbetrachtungen, wie wir sie im Auge haben, ausgesprochen unglücklich, und daher

modifizieren wir unsere Darstellung eines Zustandes ein wenig, indem wir durch die Teilchenzahl teilen. Offenbar läßt sich ein Zustand ebenso gut beschreiben durch ein  $d$ -Tupel  $(x_1/n, \dots, x_d/n)$  mit  $x_i \in \mathbb{N}_0$  (wie vorher) und  $\sum_{i=1}^d \frac{x_i}{n} = 1$ . Diese Beschreibung ist zwar ebenso wenig von  $n$  unabhängig wie die vorhergehende, ist aber für unsere Zwecke dienlicher, da sich Konfigurationen nun unabhängig von  $n$  in der Menge  $\mathcal{M}$  aller  $d$ -Tupel  $(\varrho_1, \dots, \varrho_d)$  mit  $\sum_{i=1}^d \varrho_i = 1$  befinden, d.h. der Menge aller Wahrscheinlichkeitsverteilungen auf der Menge  $\{1, \dots, d\}$ . In Analogie zu der obigen Bezeichnungsweise werden wir die Wahrscheinlichkeit, daß sich das System im Zustand  $(\varrho_1, \dots, \varrho_d)$  befindet mit  $P((\varrho_1, \dots, \varrho_d))$  bezeichnen.

Wir wollen nun – wie schon angekündigt – den Logarithmus der Wahrscheinlichkeit berechnen, daß sich das System bei einer sehr großen Teilchenzahl in einem bestimmten Zustand befindet. In der Sprechweise der modernen Wahrscheinlichkeitstheorie führt das zu einem sogenannten Prinzip großer Abweichungen.

**Satz 4.1** Für das oben beschriebene Modell eines idealen Gases mit  $p_i = 1/d$  für jedes  $i = 1, \dots, d$  gilt für jedes  $\varrho := (\varrho_1, \dots, \varrho_d) \in \mathcal{M}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\sim (\varrho_1, \dots, \varrho_d)) = h(\varrho). \quad (2)$$

Hierbei bezeichnet  $h(\varrho)$  die Kullback–Leibler Information

$$h(\varrho) = - \sum_{i=1}^d \varrho_i \log\left(\frac{\varrho_i}{1/d}\right).$$

Die Schreibweise  $P(\sim (\varrho_1, \dots, \varrho_d))$  weist dabei auf folgendes Problem hin: Für die meisten  $\varrho \in \mathcal{M}$  ist  $P(\varrho) = 0$  für fast jede Teilchenzahl  $n$ , da sich ein  $\varrho$  mit positiver Wahrscheinlichkeit für ein bestimmtes  $n$  in der Form  $(x_1/n, \dots, x_r/n)$  schreiben lassen muß. Also sollte man die Aussage des Satzes eigentlich formulieren, indem man auf der linken Seite von (2)  $\varrho$  durch einen kleinen Ball um  $\varrho$  ersetzt und auf der rechten Seite  $h(\varrho)$  durch das Supremum dieser Funktion auf diesem Ball. Obwohl ein solcher Satz natürlich prinzipiell bewiesen werden kann, werden wir auf diesen mathematischen Ballast verzichten und – abgekürzt durch die Schreibweise  $P(\sim (\varrho_1, \dots, \varrho_d))$  – eine Folge  $\varrho^{(n)}$  betrachten mit  $\varrho^{(n)} = (x_1^{(n)}/n, \dots, x_r^{(n)}/n)$ ,  $x_i^{(n)} \in \mathbb{N}_0$  und  $x_i^{(n)}/n \rightarrow \varrho_i$  für alle  $i = 1, \dots, d$ .

**Beweis des Satzes:** Das wesentliche Hilfsmittel beim Beweis des Satzes ist eine Approximation der Fakultäten im Multinomialkoeffizienten. Natürlich denkt man dabei an die Stirlingsche Formel, deren Herleitung jedoch mit einigen Schwierigkeiten verbunden ist. Wir können uns jedoch mit einem einfachen ersten Schritt begnügen: Sei also  $m \in \mathbb{N}$ , dann ist

$$\log m! = \log 2 + \log 3 + \dots + \log m$$

und daher

$$\log(m-1)! < \int_1^m \log x \, dx < \log m!.$$

Da andererseits

$$\begin{aligned} \int_1^m \log x \, dx &= m \log m - m + 1 \\ &= \log(m^m) - \log(e^m) + \log e = \log\left(e\left(\frac{m}{e}\right)^m\right), \end{aligned}$$

folgt

$$(m-1)! < e\left(\frac{m}{e}\right)^m < m!.$$

Multipliziert man die linke Ungleichung mit  $m$ , so ergibt sich

$$m! < me\left(\frac{m}{e}\right)^m$$

und daher zusammen

$$e\left(\frac{m}{e}\right)^m < m! < me\left(\frac{m}{e}\right)^m$$

(siehe beispielsweise Treiber, 1995).

Logarithmiert man die obige Ungleichungskette und dividiert durch  $m$ , so ergibt sich

$$\frac{\log e}{m} + \log m - \log e < \frac{1}{m} \log m! < \frac{\log me}{m} + \log m - \log e,$$

und unter Verwendung der Konvergenz von  $\frac{\log m}{m} \rightarrow_{m \rightarrow \infty} 0$

$$\frac{1}{m} \log m! \approx \log m - 1.$$

Somit berechnen wir

$$\frac{1}{n} \log P(\sim (\varrho_1, \dots, \varrho_d))$$

$$\begin{aligned}
&\approx \frac{1}{n} \log \left( \frac{n!}{n \varrho_1! \cdots n \varrho_d!} \left(\frac{1}{d}\right)^n \right) \\
&\approx (\log n - 1) - \sum_{i=1}^d \varrho_i \log \varrho_i - \sum_{i=1}^d \varrho_i - \sum_{i=1}^d \varrho_i \log n - \log d \\
&\approx H(\varrho) - \log d = h(\varrho).
\end{aligned}$$

□

#### Bemerkungen 4.2

1. Es ist leicht nachzurechnen, daß im Falle beliebiger  $p_i$  der Logarithmus der Wahrscheinlichkeit die Gestalt  $-\sum_{i=1}^d \varrho_i \log(\varrho_i/p_i)$  annimmt.
2. Der Satz besagt nichts anderes als Boltzmanns Prinzip  $S = k \log W$  für das vereinfachte Modell eines idealen Gases: Untypische Zustände, d.h. Zustände deren Besetzungszahlen wesentlich von  $n/d$  abweichen, sind exponentiell unwahrscheinlich, und die Rate der Konvergenz ist gegeben durch die Entropie, oder kurz: Die Entropie ist der Logarithmus der Wahrscheinlichkeit (hierbei kann die Boltzmannkonstante ruhigen Gewissens gleich eins gesetzt werden). Da die in diesem Gesetz auftauchende Entropie eine Form der Shannonschen Entropie ist, ist damit auch der Zusammenhang zwischen Boltzmannscher und Shannonscher Entropie aufgezeigt.

Wir analysieren nun noch den ebenfalls von Boltzmann betrachteten Fall, eine Verteilung zu finden, die die Entropie unter Vorgabe eines festen Energiewertes maximiert. Dazu ordnen wir jedem  $k \in \{1, \dots, d\}$  einen festen Energiewert  $u_k$  zu, setzen also die Energie auf den einzelnen Zellen fest. Die Gesamtenergie  $E$  einer Konfiguration  $(p_1, \dots, p_d)$  sei der gewichtete Mittelwert der  $u_k$ , also  $E = \sum_{k=1}^d p_k u_k$ . Sei nun  $E_0$  ein vorgegebener Energiewert mit

$$\min_k u_k < E_0 < \max_k u_k.$$

Wir suchen eine Verteilung  $P = (p_1, \dots, p_d)$  mit  $E_0 = \sum_{k=1}^d p_k u_k$ , so, daß  $H(P)$  und somit auch  $h(P)$  maximal wird.

Wir müssen uns hier darauf beschränken, das maximierende  $P$  anzugeben und zu verifizieren, daß es wirklich das Gewünschte liefert. Die Herleitung

des maximierenden Maßes würde nämlich das Lösen einer mehrdimensionalen Extremwertaufgabe mit Nebenbedingung erfordern, und somit beispielsweise die Theorie der Lagrangeschen Multiplikatoren benutzen.

Wir setzen also

$$p_{k,\beta} := \frac{\exp(-\beta u_k)}{Z_\beta}$$

mit  $Z_\beta := \sum_{k=1}^d \exp(-\beta u_k)$ . Zunächst zeigen wir, daß zu jedem Energiewert  $E_0$  ein Wert  $\beta$  existiert, so daß die Verteilung  $P_\beta := (p_{k,\beta})_k$  gerade den Energiewert  $E_0$  liefert. Dies folgt unmittelbar aus dem Zwischenwertsatz angewandt auf die Funktion  $F(\beta) := \sum_{k=1}^d u_k \frac{\exp(-\beta u_k)}{Z_\beta}$ . Wegen

$$\frac{\exp(-\beta u_k)}{Z_\beta} = \frac{\exp(-\beta(u_k - \hat{u}))}{\sum_{k:u_k=\hat{u}} \exp(-\beta(u_k - \hat{u})) + \sum_{k:u_k \neq \hat{u}} \exp(-\beta(u_k - \hat{u}))}$$

mit  $\hat{u} = \min_{k=1, \dots, d} u_k$  ist nämlich

$$\lim_{\beta \rightarrow \infty} \frac{\exp(-\beta u_k)}{Z_\beta} = \begin{cases} \frac{1}{|\{j: u_j = \hat{u}\}|} & u_k = \hat{u} \\ 0 & \text{sonst,} \end{cases}$$

und somit  $\lim_{\beta \rightarrow \infty} F(\beta) = \min_k u_k$ . Analog zeigt man  $\lim_{\beta \rightarrow -\infty} F(\beta) = \max_k u_k$ . Wegen der Stetigkeit der Funktion  $F(\beta)$  gibt es also nach dem Zwischenwertsatz zu obigem  $E_0$  ein  $\beta_0$ , so daß  $P_{\beta_0}$  den Energiewert  $E_0$  liefert. Daß dieses  $P_{\beta_0}$  auch die Entropie auf der Energieschale  $E_0$  maximiert, sieht man so: Sei  $q_1, \dots, q_d$  eine weitere Wahrscheinlichkeitsverteilung auf den  $d$  Zellen mit  $\sum_{k=1}^d q_k u_k = E_0$ , so gilt nach (1)

$$\begin{aligned}
-\sum_{k=1}^d q_k \log q_k &\leq -\sum_{k=1}^d q_k \log p_{k,\beta_0} \\
&= \sum_{k=1}^d q_k (\log Z_{\beta_0} + \beta_0 u_k) \\
&= \log Z_{\beta_0} + \beta_0 E_0.
\end{aligned}$$

Andererseits ist

$$-\sum_{k=1}^d p_{k,\beta_0} \log p_{k,\beta_0} = \log Z_{\beta_0} + \beta_0 E_0,$$

also wird durch die Verteilung  $P_{\beta_0}$  die Entropie maximal, wenn man die Energie  $E_0$  vorgibt. Diese Verteilung fand auch Boltzmann für das ideale Gas, man bezeichnet sie aber als Gibbsverteilung, da Gibbs sie in größerer Allgemeinheit einfuhrte.

## Anhang A: Historisches zu den großen Abweichungen

Hier wollen wir noch einmal kurz die in Abschnitt 4 verwendeten Techniken beleuchten. Wie schon dort erwähnt, ist unser Satz von einem mathematischen Standpunkt aus ein Prinzip der großen Abweichungen. Von einem solchen Prinzip spricht man – grob gesprochen – wenn in einem Gesetz der großen Zahlen die Konvergenzgeschwindigkeit exponentiell ist. Die ersten solchen exponentiellen Konvergenzsätze wurden bereits von Cramér und Chernoff in den 30er und 50er Jahren dieses Jahrhunderts entdeckt. Allerdings waren diese durch statistische Fragestellungen motiviert, beispielsweise durch Fragen über die Gültigkeit des zentralen Grenzwertsatzes “weit draußen“.

Die zentrale Bedeutung der Theorie der großen Abweichungen für die Physik, insbesondere der statistischen Mechanik, kristallisierte sich in einer Serie von Arbeiten der Mathematiker M.D. Donsker und S.R.S. Varadhan heraus. Im Gegensatz zu der in diesem Artikel beschriebenen gewissermaßen statischen Situation, wird in den Donsker/Varadhan Arbeiten überwiegend eine dynamische Situation analysiert. Die Zusammenhänge zwischen der statistischen Mechanik und der Theorie der großen Abweichungen sind auf den verschiedensten Ebenen beschrieben in dem Buch von R. Ellis, 1985.

Heutzutage sind Prinzipien großer Abweichungen für eine Vielzahl von Situationen analysiert. Ihre Anwendungen reichen von der statistischen Mechanik, in dem hier beschriebenen Sinne, bis zur Informationstheorie. Eine schöne Einführung in die großen Abweichungen, eine Übersicht über den aktuellen Stand der Forschung und eine breite Palette möglicher Anwendungen bietet das Buch von A. Dembo und O. Zeitouni, 1993. Dort findet sich auch eine nahezu erschöpfende Liste der wichtigsten Literatur.

## Literatur

- [1] L. Boltzmann; Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht; Wiss. Abhandlungen, Bd II, S. 163-223 (1877)

- [2] A. Dembo, O. Zeitouni; Large Deviations and Applications; Jones and Bartlett; Boston, MA (1993)
- [3] H. Dinges, H. Rost; Prinzipien der Stochastik; Teubner Studienbücher Mathematik; Stuttgart (1982)
- [4] R. Ellis; Entropy, Large Deviations and Statistical Mechanics; Grundlehren der mathematischen Wissenschaften 271; Springer; New York (1985)
- [5] J.Flachsmeyer, A. Ollech; Was ist Entropie; Didaktik der Mathematik 2,1995, S. 81-92 (1995)
- [6] O. Forster; Analysis I; Vieweg Studium , Grundkurs Mathematik 24, 4. Auflage; Vieweg; Braunschweig (1983)
- [7] D. Treiber; Wieviele Dezimalstellen hat n-Fakultät; Didaktik der Mathematik 2,1995, S.93-98 (1995)

## Adresse der Autoren:

Peter Eichelsberger  
Matthias Löwe  
Fakultät für Mathematik  
Universität Bielefeld  
Postfach 100 131  
33501 Bielefeld