The theory of large deviations: from Boltzmann's 1877 calculation to equilibrium macrostates in 2D turbulence

Richard S. Ellis *

Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003, USA

Abstract

After presenting some basic ideas in the theory of large deviations, this paper applies the theory to a number of problems in statistical mechanics. These include deriving the form of the Gibbs state for a discrete ideal gas; describing probabilistically the phase transition in the Curie–Weiss model of a ferromagnet; and deriving variational formulas that describe the equilibrium macrostates in models of two-dimensional turbulence. A general approach to the large deviation analysis of models in statistical mechanics is also formulated. ©1999 Elsevier Science B.V. All rights reserved.

Keywords: Relative entropy; Large deviation principle; Statistical mechanics; Maximum entropy principle; Equilibrium macrostates

1. Introduction

The theory of large deviations studies the exponential decay of probabilities in certain random systems. It is being applied to a wide range of problems in which detailed information on rare events is required. One is often interested not only in the probability of rare events but also in the characteristic behavior of the system as the rare event occurs. For example, in applications to queueing theory and communication systems, the rare event could represent an overload or breakdown of the system. In this case, large deviation methodology can lead to an efficient redesign of the system so that the overload or breakdown does not occur. In applications to statistical mechanics, which will be the main focus of this paper, the theory of large deviations gives precise, exponential-order estimates that are perfectly suited for asymptotic analysis.

This paper will discuss a number of topics in the theory of large deviations and several applications to statistical mechanics, all united by the concept of relative entropy. This concept entered human culture through the first large deviation calculation in science, carried out by Boltzmann. Stated in a modern terminology, his discovery was that the relative entropy expresses the asymptotic behavior of multinomial probabilities. This statistical interpretation of entropy has the following crucial physical implication. Entropy is a bridge between a microscopic level, on which

* Tel.: +1-413-545-3125; fax: +1-413-545-1801

E-mail address: rsellis@math.umass.edu (R.S. Ellis)

physical systems are defined in terms of the complicated interactions among the individual constituent particles, and a macroscopic level, on which the laws describing the behavior of the system are formulated.

Building on the work of Boltzmann, Gibbs asked a fundamental question. How can one use probability theory to study equilibrium properties of physical systems such as an ideal gas, a ferromagnet, or a fluid? These properties include such phenomena as phase transitions; e.g., the liquid–gas transition or spontaneous magnetization in a ferromagnet. Another example arises in the study of freely evolving, inviscid fluids, for which one wants to describe coherent states. These are steady, stable mean flows comprised of one or more vortices that persist amidst the turbulent fluctuations of the vorticity field. Gibbs's answer, which led to the development of classical equilibrium statistical mechanics, is that one studies equilibrium properties via probability measures on configuration space known today as Gibbs canonical ensembles or Gibbs states. For background in statistical mechanics, I recommend [1–3], which cover a number of topics relevant to the contents of this paper.

One of our main purposes in this paper is to show the utility of the theory of large deviations by applying it to a number of statistical mechanical models. Our applications of the theory include a derivation of the form of the Gibbs state for a discrete ideal gas (Section 5); a probabilistic description of the phase transition in the Curie–Weiss model of a ferromagnet in terms of the breakdown of the law of large numbers for the spin per site (Section 7); and as an overview of recent work carried out in [4], a derivation of variational formulas that describe the equilibrium macrostates in models of two-dimensional turbulence (Section 9). In terms of these macrostates, coherent vortices of two-dimensional turbulence can be studied.

Boltzmann's calculation of the asymptotic behavior of multinomial probabilities in terms of relative entropy was carried out in 1877 as a key component of his paper that gave a probabilistic interpretation of the Second Law of Thermodynamics [5]. This fundamental calculation represents a revolutionary moment in human culture during which both statistical mechanics and the theory of large deviations were born. Boltzmann's work is put in historical context by Everdell in his book *The First Moderns*, which traces the development of the modern consciousness in 19th and 20th century thought [6]. Chapter 3 focuses on the mathematicians of Germany in the 1870's – namely, Cantor, Dedekind, and Frege – who "would become the first creative thinkers in any field to look at the world in a fully twentieth-century manner" (p. 31). Boltzmann is then presented as the man whose investigations in stochastics and statistics made possible the work of the two other great founders of twentieth-century theoretical physics, Planck and Einstein. "He was at the center of the change" (p. 48).

In this paper Boltzmann's discovery of the asymptotic behavior of multinomial probabilities in terms of relative entropy is described in Section 3 after a preliminary section that introduces a basic probabilistic model. Two related problems are then considered: the calculation of the probabilities of a loaded die in Section 4 and the calculation of the probabilities of the energy states of a discrete ideal gas in Section 5. The solutions of these problems motivate the form of the Gibbs canonical ensemble. The general concept of a large deviation principle and related ideas are presented in Section 6. In Section 7 the theory of large deviations is used to study equilibrium properties of a basic model of ferromagnetism known as the Curie–Weiss model. This leads in Section 8 to the formulation of a general procedure for applying the theory of large deviations to the analysis of an extensive class of statistical mechanical models. This general procedure is then used in Section 9 along with Sanov's Theorem, which generalizes Boltzmann's 1877 calculation, to derive variational formulas that describe the equilibrium macrostates in two models of two-dimensional turbulence; namely, the well known Miller–Robert theory and a modification of that theory recently proposed by Turkington. Because Sanov's Theorem plays a vital role in the derivation, this final application of the theory of large deviations brings our focus back home to Boltzmann, through whose research in the foundations of statistical mechanics the theory was born.

2. A basic probabilistic model

In later sections we will investigate a number of questions in the theory of large deviations in the context of a basic probabilistic model, which we now introduce. Let $\alpha \geq 2$ be an integer, $y_1 < y_2 < \cdots < y_\alpha$ a set of α real numbers, and $\rho_1, \rho_2, \ldots, \rho_\alpha$ a set of α positive real numbers summing to 1. We think of $\Lambda \doteq \{y_1, y_2, \ldots, y_\alpha\}$ as the set of possible outcomes of a random experiment in which each individual outcome y_k has the probability ρ_k of occurring. The vector $\rho \doteq (\rho_1, \rho_2, \ldots, \rho_\alpha)$ is an element of the set of probability vectors

$$\mathcal{P}_\alpha \doteq \left\{ \gamma \in \mathbb{R}^\alpha : \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_\alpha) \geq 0, \sum_{k=1}^\alpha \gamma_k = 1 \right\}.$$

Any vector $\gamma \in \mathcal{P}_\alpha$ also defines a probability measure on the set of subsets of Λ via

$$\gamma = \gamma(\mathrm{d}y) \doteq \sum_{k=1}^\alpha \gamma_k \, \delta_{y_k}(\mathrm{d}y),$$

where for $y \in \Lambda$ $\delta_{y_k}\{y\} = 1$ if $y = y_k$ and equals 0 otherwise. Thus for $B \subset \Lambda$, $\gamma\{B\} = \sum_{y_k \in B} \gamma_k$. For each integer n, the configuration space for n independent repetitions of the experiment is $\Omega_n \doteq \Lambda^n$, a typical element of which is denoted by $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$. For each $\omega \in \Omega_n$ we define

$$P_n\{\omega\} \doteq \prod_{j=1}^n \rho\{\omega_j\}$$

and extend this to a probability measure on the set of subsets of Ω_n by defining

$$P_n\{B\} \doteq \sum_{\omega \in B} P_n\{\omega\} \quad \text{for} \quad B \subset \Omega_n.$$

P_n is called the product measure with one-dimensional marginals ρ. With respect to P_n the coordinate functions $X_j(\omega) \doteq \omega_j$, $j = 1, 2, \ldots, n$, are independent, identically distributed (i.i.d.) random variables with common distribution ρ; that is, for any subsets $B_1, B_2, \ldots, B_n$ of Λ

$$P_n\{\omega \in \Omega_n : X_j(\omega) \in B_j \quad \text{for} \quad j = 1, 2, \ldots, n\} = \prod_{j=1}^n P_n\{\omega \in \Omega_n : X_j(\omega) \in B_j\} = \prod_{j=1}^n \rho\{B_j\}.$$

Example 1. Random phenomena that can be studied via this basic model include standard examples such as coin tossing and die tossing and also include a discrete ideal gas.

1. *Coin tossing.* In this case $\Lambda \doteq \{1, 2\}$ and $\rho_1 = \rho_2 \doteq 1/2$.
2. *Die tossing.* In this case $\Lambda \doteq \{1, 2, \ldots, 6\}$ and each $\rho_k \doteq 1/6$.
3. *Discrete ideal gas.* Consider a 'discrete ideal gas' consisting of n identical, noninteracting particles, each having α equally likely energy levels $y_1, y_2, \ldots, y_\alpha$; in this case each ρ_k equals $1/\alpha$. The coordinate functions X_j represent the random energy levels of the molecules of the gas. The statistical independence of these random variables reflects the fact that the molecules of the gas do not interact.

We will return to the discrete ideal gas in Section 5 after introducing some basic concepts in theory of large deviations.

3. Boltzmann's discovery

In its original form Boltzmann's discovery concerns the asymptotic behavior of multinomial coefficients. For the purpose of applications in this paper, it is advantageous to formulate it in terms of a probabilistic quantity known as the empirical vector. For $\omega \in \Omega_n$ and $y \in \Lambda$ define

$$L_n(y) = L_n(\omega, y) \doteq \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}\{y\}.$$

Thus $L_n(\omega, y)$ equals $n^{-1} \cdot \#\{j \in \{1, \dots, n\} : \omega_j = y\}$; it counts the relative frequency with which y appears in the configuration ω. We then define the empirical vector

$$L_n = L_n(\omega) \doteq (L_n(\omega, y_1), \dots, L_n(\omega, y_\alpha)) = \frac{1}{n} \sum_{j=1}^{n} (\delta_{X_j(\omega)}\{y_1\}, \dots, \delta_{X_j(\omega)}\{y_\alpha\}).$$

L_n takes values in $\mathcal{P}_\alpha$. By the last equality it equals the sample mean of the i.i.d. random variables $(\delta_{X_j(\omega)}\{y_1\}, \dots, \delta_{X_j(\omega)}\{y_\alpha\})$.

The limiting behavior of L_n is straightforward to determine. Let $\| \cdot \|$ denote the Euclidean norm on $\mathbb{R}^\alpha$. For any $\gamma \in \mathcal{P}_\alpha$ and $\varepsilon > 0$, we define the open ball

$$B(\gamma, \varepsilon) \doteq \{\nu \in \mathcal{P}_\alpha : \|\gamma - \nu\| < \varepsilon\}.$$

Since the X_j have the common distribution ρ, for each $y_k \in \Lambda$

$$E^{P_n}\{L_n(y_k)\} = E^{P_n}\left\{\frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}\{y_k\}\right\} = \frac{1}{n} \sum_{j=1}^{n} P_n\{X_j = y_k\} = \rho_k,$$

where E^{P_n} denotes expectation with respect to P_n. Hence by the weak law of large numbers for the sample means of i.i.d. random variables, for any $\varepsilon > 0$

$$\lim_{n \to \infty} P_n\{L_n \in B(\rho, \varepsilon)\} = 1. \tag{1}$$

It follows that for any $\gamma \in \mathcal{P}_\alpha$ not equal to ρ and for any $\varepsilon > 0$ satisfying $0 < \varepsilon < \|\rho - \gamma\|$

$$\lim_{n \to \infty} P_n\{L_n \in B(\gamma, \varepsilon)\} = 0. \tag{2}$$

As we will see, Boltzmann's discovery implies that these probabilities converge to 0 exponentially fast in n. The exponential decay rate is given in terms of the relative entropy, which we now define.

Definition 1 (Relative entropy). Let $\rho = (\rho_1, \dots, \rho_\alpha)$ denote the probability vector in $\mathcal{P}_\alpha$ in terms of which the basic probabilistic model is defined. The relative entropy of $\gamma \in \mathcal{P}_\alpha$ with respect to ρ is defined by

$$I_\rho(\gamma) \doteq \sum_{k=1}^{\alpha} \gamma_k \log \frac{\gamma_k}{\rho_k}.$$

Several properties of the relative entropy are given in the next lemma.

Lemma 1. *For $\gamma \in \mathcal{P}_\alpha$, $I_\rho(\gamma)$ measures the discrepancy between γ and ρ in the sense that $I_\gamma(\rho) \geq 0$ and $I_\gamma(\rho) = 0$ if and only if $\gamma = \rho$. Thus $I_\rho(\gamma)$ attains its infimum of 0 over $\mathcal{P}_\alpha$ at the unique measure $\gamma = \rho$. In addition, I_ρ is strictly convex on $\mathcal{P}_\alpha$.*

Proof. For $x \geq 0$ the graph of the strictly convex function $x\log x$ has the tangent line $y = x - 1$ at $x = 1$. Hence $x\log x \geq x - 1$ with equality if and only if $x = 1$. It follows that for any $\gamma \in \mathcal{P}_\alpha$

$$\frac{\gamma_k}{\rho_k}\log\frac{\gamma_k}{\rho_k} \geq \frac{\gamma_k}{\rho_k} - 1 \tag{3}$$

with equality if and only if $\gamma_k = \rho_k$. Multiplying this inequality by ρ_k and summing over k yields

$$I_\rho(\gamma) = \sum_{k=1}^\alpha \gamma_k \log\frac{\gamma_k}{\rho_k} \geq \sum_{k=1}^\alpha (\gamma_k - \rho_k) = 0.$$

$I_\rho(\gamma) = 0$ if and only if equality holds in Eq. (3) for each k; i.e., if and only if $\gamma = \rho$. This yields the first assertion in the proposition. This proof is typical of proofs of analogous results involving relative entropy [cf. Proposition 1] in that we use a global convexity inequality, $x\log x \geq x - 1$ with equality if and only if $x = 1$, rather than calculus to determine where I_ρ attains its infimum over $\mathcal{P}_\alpha$. Since

$$I_\rho(\gamma) = \sum_{k=1}^\alpha \rho_k \frac{\gamma_k}{\rho_k}\log\frac{\gamma_k}{\rho_k},$$

the strict convexity of I_ρ is a consequence of the strict convexity of $x\log x$ for $x \geq 0$. $\qquad\square$

We are now ready to give the first formulation of Boltzmann's discovery, which we state using a heuristic notation. However, the proof uses formal calculations that can easily be turned into a rigorous proof of an asymptotic theorem. That theorem is stated in Theorem 2. From Boltzmann's momentous discovery both the theory of large deviations and the Gibbsian formulation of equilibrium statistical mechanics grew.

Theorem 1 (Boltzmann's discovery – formulation 1). *For any $\gamma \in \mathcal{P}_\alpha$ and all sufficiently small $\varepsilon > 0$*

$$P_n\{L_n \in B(\gamma, \varepsilon)\} \approx \exp[-nI_\rho(\gamma)] \quad \text{as } n \to \infty.$$

Heuristic proof. By elementary combinatorics

$$P_n\{L_n \in B(\gamma, \varepsilon)\} = P_n\left\{\omega \in \Omega_n : L_n(\omega) \sim \frac{1}{n}(n\gamma_1, n\gamma_2, \ldots, n\gamma_\alpha)\right\}$$

$$\approx P_n\{\#\{\omega_j\text{'s} = y_1\} \sim n\gamma_1, \ldots, \#\{\omega_j\text{'s} = y_\alpha\} \sim n\gamma_\alpha\}$$

$$\approx \frac{n!}{(n\gamma_1)!(n\gamma_2)!\cdots(n\gamma_\alpha)!}\rho_1^{n\gamma_1}\rho_2^{n\gamma_2}\cdots\rho_\alpha^{n\gamma_\alpha}.$$

Stirling's formula in the weak form $\log(n!) = n\log n - n + \mathrm{O}(\log n)$ yields

$$\frac{1}{n}\log P_n\{L_n \in B(\gamma, \varepsilon)\} \approx \frac{1}{n}\log\left(\frac{n!}{(n\gamma_1)!(n\gamma_2)!\cdots(n\gamma_\alpha)!}\right) + \sum_{k=1}^\alpha \gamma_k\log\rho_k$$

$$= -\sum_{k=1}^\alpha \gamma_k\log\gamma_k + \mathrm{O}\left(\frac{\log n}{n}\right) + \sum_{k=1}^\alpha \gamma_k\log\rho_k = -\sum_{k=1}^\alpha \gamma_k\log\frac{\gamma_k}{\rho_k} + \mathrm{O}\left(\frac{\log n}{n}\right)$$

$$= -I_\rho(\gamma) + \mathrm{O}\left(\frac{\log n}{n}\right).$$

Theorem 1 has the following interesting consequence. Let γ be any vector in $\mathcal{P}_\alpha$ which differs from ρ. Since $I_\rho(\gamma) > 0$ (Lemma 1), it follows that for all sufficiently small $\varepsilon > 0$

$$P_n\{L_n \in B(\gamma, \varepsilon)\} \approx \exp[-nI_\rho(\gamma)] \to 0 \quad \text{as } n \to \infty,$$

a limit which, if rigorous, would imply Eq. (2).

Let A be a Borel subset of $\mathcal{P}_\alpha$; the class of Borel subsets includes all closed sets and all open sets. If ρ is not contained in the closure of A, then by the weak law of large numbers

$$\lim_{n\to\infty} P_n\{L_n \in A\} = 0,$$

and by analogy with the heuristic asymptotic result given in Theorem 1 we expect that these probabilities converge to 0 exponentially fast with n. This is in fact the case. In order to express the exponential decay rate of such probabilities in terms of the relative entropy, we introduce the notation $I_\rho(A) \doteq \inf_{\gamma \in A} I_\rho(\gamma)$. The range of $L_n(\omega)$ for $\omega \in \Omega_n$ is the set of probability vectors having the form $\mathbf{k}/n$, where $\mathbf{k} \in \mathbb{R}^\alpha$ has non-negative integer coordinates summing to n; hence the cardinality of the range does not exceed n^α. Since

$$P_n\{L_n \in A\} = \sum_{\gamma \in A} P_n\{L_n \sim \gamma\} \approx \sum_{\gamma \in A} \exp[-nI_\rho(\gamma)]$$

and

$$\exp[-nI_\rho(A)] \leq \sum_{\gamma \in A} \exp[-nI_\rho(\gamma)] \leq n^\alpha \exp[-nI_\rho(A)],$$

one expects that at least to exponential order

$$P_n\{L_n \in A\} \approx \exp[-nI_\rho(A)] \quad \text{as } n \to \infty. \tag{4}$$

As formulated in Corollary 1, this asymptotic result is indeed valid. It is a consequence of the following rigorous reformulation of Boltzmann's discovery, known as Sanov's Theorem, which expresses the large deviation principle for the empirical vectors L_n. That concept is defined in general in Definition 3.

Theorem 2 (Boltzmann's discovery – formulation 2). *The sequence of random probability vectors $\{L_n, n \in \mathbb{N}\}$ satisfies the large deviation principle on $\mathcal{P}_\alpha$ with rate function I_ρ in the following sense.*
  1. *Large deviation upper bound: for any closed subset F of $\mathcal{P}_\alpha$*

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n\{L_n \in F\} \leq -I_\rho(F).$$

  2. *Large deviation lower bound: for any open subset G of $\mathcal{P}_\alpha$*

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n\{L_n \in G\} \geq -I_\rho(G).$$

Comments on the proof. For $\gamma \in \mathcal{P}_\alpha$ and $\varepsilon > 0$, $B(\gamma, \varepsilon)$ denotes the open ball with center γ and radius ε and $\bar{B}(\gamma, \varepsilon)$ denotes the corresponding closed ball. Since $\mathcal{P}_\alpha$ is a compact subset of $\mathbb{R}^\alpha$, any closed subset F of $\mathcal{P}_\alpha$ is automatically compact. By a standard covering argument it is not hard to show that the large deviation upper bound holds for any closed set F provided one obtains the large deviation upper bound for any closed ball $\bar{B}(\gamma, \varepsilon)$:

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n\{L_n \in \bar{B}(\gamma, \varepsilon)\} \leq -I_\rho(\bar{B}(\gamma, \varepsilon)).$$

Likewise, the large deviation lower bound holds for any open set G provided one obtains the large deviation lower bound for any open ball $B(\gamma, \varepsilon)$:

$$\liminf_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in B(\gamma, \varepsilon)\} \geq -I_\rho(B(\gamma, \varepsilon)).$$

The bounds in the last two displays can be proved via combinatorics and Stirling's formula as in the heuristic proof of Theorem 1; one can easily adapt the calculations given in ([1], Section I.4). The details are omitted. $\qquad\square$

For a class of Borel subsets A of $\mathcal{P}_\alpha$ we can now derive a rigorous version of the asymptotic formula (4). This class consists of sets A such that $\overline{\text{int } A}$, the closure of the interior of A relative to $\mathcal{P}_\alpha$, equals $\bar{A}$, the closure of A. Any open ball $B(\gamma, \varepsilon)$ or closed ball $\bar{B}(\gamma, \varepsilon)$ satisfies this condition.

Corollary 1. *Let A be any Borel subset of $\mathcal{P}_\alpha$ such that $\overline{\text{int } A} = \bar{A}$. Then*

$$\lim_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in A\} = -I_\rho(A).$$

Proof. Since $\bar{A} \supset A \supset \text{int } A$,

$$-I_\rho(\bar{A}) \geq \limsup_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in \bar{A}\} \geq \limsup_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in A\} \geq \liminf_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in A\}$$

$$\geq \liminf_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in \text{int } A\} \geq -I_\rho(\text{int } A).$$

The continuity of I_ρ on $\mathcal{P}_\alpha$ implies that $I_\rho(\text{int } A) = I_\rho(\overline{\text{int } A})$. Hence by the condition on A, the extreme terms in this display are equal. The desired limit follows. $\qquad\square$

The next corollary of Theorem 2 allows one to conclude that a large class of probabilities involving L_n converge to 0. The analogue of this corollary in other large deviation settings is extremely useful in applications. For example, we will use it in Section 7 to analyze the Curie–Weiss model of ferromagnetism.

Corollary 2. *Let A be any Borel subset of $\mathcal{P}_\alpha$ such that $\bar{A}$ does not contain ρ. Then $I_\rho(\bar{A}) > 0$ and for some $C < \infty$*

$$P_n\{L_n \in A\} \leq C \exp[-nI_\rho(\bar{A})] \to 0 \quad \text{as } n \to \infty.$$

Proof. Since $I_\rho(\gamma) > I_\rho(\rho) = 0$ for any $\gamma \neq \rho$, the positivity of $I_\rho(\bar{A})$ follows from the continuity of I_ρ on $\mathcal{P}_\alpha$. The second assertion is an immediate consequence of the large deviation upper bound applied to $\bar{A}$ and the positivity of $I_\rho(\bar{A})$. $\qquad\square$

Take any $\varepsilon > 0$. Applying Corollary 2 to the complement of the open ball $B(\rho, \varepsilon)$ yields $P_n\{L_n \notin B(\rho, \varepsilon)\} \to 0$ or equivalently

$$\lim_{n \to \infty} P_n\{L_n \in B(\rho, \varepsilon)\} = 1.$$

Although this rederives the weak law of large numbers for L_n as already expressed in Eq. (1), this second derivation relates the order-1 limit for L_n to the point in $\mathcal{P}_\alpha$ – namely, ρ – where the rate function I_ρ attains its infimum. In this context we call ρ the 'equilibrium value' of L_n with respect to the measures P_n. This limit is the simplest example, and the first of several more complicated but related formulations to be encountered in this paper, of what is commonly called a 'maximum entropy principle'. Following the usual convention in the physical literature, we will continue to use this terminology in referring to such principles even though we are minimizing the relative entropy (equivalently, maximizing $-I_\rho(\gamma)$) rather than maximizing the physical entropy. When $\rho_k = 1/\alpha$ for each k, the two quantities differ by a minus sign and an additive constant.

Maximum Entropy Principle 1. $\gamma_0 \in \mathcal{P}_\alpha$ *is an equilibrium value of* L_n *with respect to* P_n *if and only if* γ_0 *minimizes* $I_\rho(\gamma)$ *over* $\mathcal{P}_\alpha$; *this occurs if and only if* $\gamma_0 = \rho$.

In the next section we will present a limit theorem for L_n whose proof is based on the precise, exponential-order estimates given by the large deviation principle in Theorem 2.

4. A conditioned limit theorem for L_n

You participate in a crooked gambling game being played with a loaded die. How can you determine the actual probabilities of each face 1, 2, ... , 6? The conditioned limit theorem to be introduced in this section not only gives an answer to this apparently ambiguous question, but also, with some additional work, has important statistical mechanical implications. As we will see in Section 5, it motivates the form of the Gibbs state for the discrete ideal gas and, by extension, for any statistical mechanical system characterized by conservation of energy. These unexpected theorems are the first indication of the power of Boltzmann's discovery, which gives precise exponential-order estimates for probabilities of the form $P_n\{L_n \in A\}$. The theorems have the following form. Suppose that one is given a particular set A and wants to determine a set B belonging to a certain class (e.g., open balls) such that the conditioned limit

$$\lim_{n \to \infty} P_n\{L_n \in B | L_n \in A\} = \lim_{n \to \infty} P_n\{L_n \in B \cap A\}\frac{1}{P_n\{L_n \in A\}} = 1$$

is valid. Since to exponential order

$$P_n\{L_n \in B \cap A\}\frac{1}{P_n\{L_n \in A\}} \approx \exp[-n(I_\rho(B \cap A) - I_\rho(A))],$$

one should obtain the conditioned limit if B satisfies $I_\rho(B \cap A) = I_\rho(A)$. If one can determine the point in A where the infimum of I_ρ is attained, then one picks B to contain this point. In the examples involving the loaded die and the discrete ideal gas, such a minimizing point can be determined. It will lead to a second maximum entropy principle for L_n with respect to the conditional probabilities $P_n\{\cdot|L_n \in A\}$.

We return to the question concerning the loaded die, using the basic probabilistic model introduced in Section 2 (Example 1, part 2). Upon entering the crooked gambling game, one assigns the equal probabilities $\rho_k = 1/6$ to each of the six faces because one has no additional information. One then observes the game for n tosses; probabilistically this corresponds to knowing a configuration $\omega \in \{1, \ldots, 6\}^n$. Based on the value of

$$S_n(\omega) \doteq \sum_{j=1}^{n} X_j(\omega) = \sum_{j=1}^{n} \omega_j,$$

one desires to recalculate the probabilities of the six faces. Being a mathematician rather than a professional gambler, I will carry this out in the limit $n \to \infty$.

If the die were fair, then the sample mean $S_n(\omega)/n$ should equal approximately the theoretical mean

$$\bar{y} \doteq \sum_{k=1}^{6} k\rho_k = 3.5.$$

Hence let us assume that $S_n/n \in [z - a, z]$, where a is a small positive number and $1 \le z - a < z < \bar{y}$; a similar result would hold if we assumed that $S_n/n \in [z, z + a]$, where $\bar{y} < z < z + a \le 6$. We can now

formulate the question concerning the loaded die as the following conditioned limit: determine positive numbers $\{\rho_k^*, k = 1, \ldots, 6\}$ summing to 1 such that

$$\rho_k^* = \lim_{n \to \infty} P_n \left\{ X_1 = k \Big| \frac{S_n}{n} \in [z - a, z] \right\}.$$

This will be seen to follow from the following more easily answered question: in the limit $n \to \infty$, conditioned on $S_n / n \in [z - a, z]$, determine the most likely configuration $\rho^* = (\rho_1^*, \ldots, \rho_6^*)$ of L_n. In other words, we want $\rho^* \in \mathcal{P}_\alpha$ such that for any $\varepsilon > 0$

$$\lim_{n \to \infty} P_n \left\{ L_n \in B(\rho^*, \varepsilon) \Big| \frac{S_n}{n} \in [z - a, z] \right\} = 1.$$

The form of ρ^* is given in the following theorem; it depends only on z, not on a.

We formulate the theorem for a general state space $\Lambda = \{y_1, \ldots, y_\alpha\}$ and a given positive vector $\rho = (\rho_1, \ldots, \rho_\alpha) \in \mathcal{P}_\alpha$. As above, define

$$S_n \doteq \sum_{j=1}^n X_j \quad \text{and} \quad \bar{y} \doteq \sum_{k=1}^\alpha y_k \rho_k$$

and for $a > 0$ fix a closed interval $[z - a, z] \subset [y_1, \bar{y})$. A theorem analogous to the following would hold if $[z - a, z] \subset [y_1, \bar{y})$ were replaced by $[z, z + a] \subset (\bar{y}, y_\alpha]$.

Theorem 3.

1. *There exists $\rho^{(\beta)} \in \mathcal{P}_\alpha$ such that for every $\varepsilon > 0$*

$$\lim_{n \to \infty} P_n \left\{ L_n \in B(\rho^{(\beta)}, \varepsilon) \Big| \frac{S_n}{n} \in [z - a, z] \right\} = 1. \tag{5}$$

   The quantity $\rho^{(\beta)} = (\rho_1^{(\beta)}, \ldots, \rho_\alpha^{(\beta)})$ has the form

$$\rho_k^{(\beta)} \doteq \frac{\exp[-\beta y_k] \rho_k}{\sum_{j=1}^\alpha \exp[-\beta y_j] \rho_j},$$

   where $\beta = \beta(z) \in \mathbb{R}$ is chosen so that $\sum_{k=1}^\alpha y_k \rho_k^{(\beta)} = z$.

2. *For any continuous function f mapping $\mathcal{P}_\alpha$ into $\mathbb{R}$*

$$\lim_{n \to \infty} E^{P_n} \left\{ f(L_n) \Big| \frac{S_n}{n} \in [z - a, z] \right\} = f(\rho^{(\beta)}).$$

3. *For each $k \in \{1, \ldots, \alpha\}$*

$$\rho_k^{(\beta)} = \lim_{n \to \infty} P_n \left\{ X_1 = y_k \Big| \frac{S_n}{n} \in [z - a, z] \right\}.$$

We first show that $\rho^{(\beta)}$ is well defined. For $r \in \mathbb{R}$ simple calculus gives the following properties of $c(r) \doteq \log(\sum_{k=1}^\alpha \exp[r y_k] \rho_k)$: $c''(r) > 0$, $c'(r) \to y_1$ as $r \to -\infty$, $c'(0) = \bar{y}$, and $c'(r) \to y_\alpha$ as $r \to \infty$. Hence there exists a unique $\beta = \beta(z)$ such that $c'(-\beta) = \sum_{k=1}^\alpha y_k \rho_k^{(\beta)} = z$, as claimed. Since $y_1 < z < \bar{y}$, $\beta = \beta(z)$ is positive.

In order to prove the limit (5) in part 1, we express the conditional probability in Eq. (5) in terms of the empirical vector L_n. Define the closed convex set

$$A(z) \doteq \left\{ \gamma \in \mathcal{P}_\alpha : \sum_{k=1}^{\alpha} y_k \gamma_k \in [z - a, z] \right\},$$

which contains $\rho^{(\beta)}$. Since for each $\omega \in \Omega_n$

$$\frac{1}{n} S_n(\omega) = \sum_{j=1}^{n} y_k L_n(\omega, y_k),$$

it follows that $\{\omega \in \Omega_n : S_n(\omega)/n \in [z - a, z]\} = \{\omega \in \Omega_n : L_n(\omega) \in A(z)\}$. Hence using the formal notation of Eq. (4), we have for large n

$$\begin{aligned}
P_n \left\{ L_n \in B(\rho^{(\beta)}, \varepsilon) | \frac{S_n}{n} \in [z - a, z] \right\} &= P_n \{ L_n \in B(\rho^{(\beta)}, \varepsilon) | L_n \in A(z) \} \\
&= P_n \{ L_n \in B(\rho^{(\beta)}, \varepsilon) \cap A(z) \} \frac{1}{P_n \{ L_n \in A(z) \}} \\
&\approx \exp[-n(I_\rho(B(\rho^{(\beta)}, \varepsilon) \cap A(z)) - I_\rho(A(z)))].
\end{aligned}$$

The last expression, and thus the probability in the first line of the display, are of order 1 provided

$$I_\rho(B(\rho^{(\beta)}, \varepsilon) \cap A(z)) = I_\rho(A(z)). \tag{6}$$

The next proposition shows that I_ρ attains its infimum over $A(z)$ at the unique point $\rho^{(\beta)}$. This gives Eq. (6) and motivates the fact that for large n, $P_n \{ L_n \in B(\rho^{(\beta)}, \varepsilon) | S_n/n \in [z - a, z] \} \approx 1$. It is not difficult to convert these formal calculations into a proof of the limit (5). The details are omitted.

Proposition 1. I_ρ *attains its infimum over* $A(z) \doteq \{ \gamma \in \mathcal{P}_\alpha : \sum_{k=1}^{\alpha} y_k \gamma_k \in [z - a, z] \}$ *at the unique point* $\rho^{(\beta)}$ *defined in part* 1 *of Theorem 3.*

Proof. We recall that $\beta = \beta(z) > 0$ and that for each $k \in \{1, \ldots, \alpha\}$

$$\frac{\rho_k^{(\beta)}}{\rho_k} = \frac{\exp[-\beta y_k]}{\exp[c(-\beta)]},$$

where for $r \in \mathbb{R} \ c(r) \doteq \log(\sum_{k=1}^{\alpha} \exp[r y_k] \rho_k)$. Hence for any $\gamma \in A(z)$

$$\begin{aligned}
I_\rho(\gamma) &= \sum_{k=1}^{\alpha} \gamma_k \log \frac{\gamma_k}{\rho_k} = \sum_{k=1}^{\alpha} \gamma_k \log \frac{\gamma_k}{\rho_k^{(\beta)}} + \sum_{k=1}^{\alpha} \gamma_k \log \frac{\rho_k^{(\beta)}}{\rho_k} \\
&= I_{\rho^{(\beta)}}(\gamma) - \beta \sum_{k=1}^{\alpha} y_k \gamma_k - c(-\beta) \geq I_{\rho^{(\beta)}}(\gamma) - \beta z - c(-\beta).
\end{aligned}$$

Since $I_{\rho^{(\beta)}}(\gamma) \geq 0$ with equality if and only if $\gamma = \rho^{(\beta)}$ (Lemma 1), it follows that for any $\gamma \in A(z)$

$$I_\rho(\gamma) \geq -\beta z - c(-\beta) = I_\rho(\rho^{(\beta)})$$

with equality if and only if $\gamma = \rho^{(\beta)}$. □

Combining this proposition with part 1 of Theorem 3 gives the second maximum entropy principle in the paper.

Maximum Entropy Principle 2. *Conditioned on* $S_n/n \in [z - a, z]$, *the asymptotically most likely configuration of* L_n *is* $\rho^{(\beta)}$, *which is the unique* $\gamma \in \mathcal{P}_\alpha$ *that minimizes* $I_\rho(\gamma)$ *subject to the constraint that* $\gamma \in A(z)$. *In statistical*

mechanical terminology, $\rho^{(\beta)}$ is the equilibrium macrostate of $\{L_n\}$ with respect to the conditional probabilities $P_n\{\cdot|S_n/n \in [z-a, z]\}$.

Part 2 of Theorem 3 states that for any continuous function f mapping $\mathcal{P}_\alpha$ into $\mathbb{R}$

$$\lim_{n\to\infty} E^{P_n}\left\{f(L_n)|\frac{S_n}{n} \in [z-a, z]\right\} = f(\rho^{(\beta)}).$$

This is an immediate consequence of part 1 and the continuity of f. Part 2 of Theorem 3 is another expression of the maximum entropy Principle 2.

Let y_i be any point in Λ. As in ([7], p. 87), we prove part 3 of Theorem 3 by relating the conditional probability $P_n\{X_1 = y_i|S_n/n \in [z-a, z]\}$ to the conditional expectation $E^{P_n}\{f(L_n)|S_n/n \in [z-a, z]\}$ considered in part 2. Given φ any function mapping Λ into $\mathbb{R}$, we define a continuous function on $\mathcal{P}_\alpha$ by

$$f(\gamma) \doteq \sum_{k=1}^{\alpha} \varphi(y_k)\gamma_k.$$

Since $f(L_n) = \sum_{k=1}^{\alpha}\varphi(y_k)L_n(y_k) = (1/n)\sum_{j=1}^{n}\varphi(X_j)$, by symmetry and part 2

$$\lim_{n\to\infty} E^{P_n}\left\{\varphi(X_1)|\frac{S_n}{n} \in [z-a, z]\right\} = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}E^{P_n}\left\{\varphi(X_j)|\frac{S_n}{n} \in [z-a, z]\right\}$$

$$= \lim_{n\to\infty} E^{P_n}\left\{f(L_n)|\frac{S_n}{n} \in [z-a, z]\right\} = f(\rho^{(\beta)}) = \sum_{k=1}^{\alpha}\varphi(y_k)\rho_k^{(\beta)}.$$

Setting $\varphi \doteq 1_{y_i}$ yields the limit in part 3 of Theorem 3:

$$\lim_{n\to\infty} P_n\left\{X_1 = y_i|\frac{S_n}{n} \in [z-a, z]\right\} = \rho_i^{(\beta)}.$$

With some additional work one can generalize part 1 of Theorem 3 by proving that with respect to the conditional probabilities $P_n\{\cdot|S_n/n \in [z-a, a]\}\{L_n\}$ satisfies the large deviation principle on $\mathcal{P}_\alpha$ with rate function

$$I(\gamma) \doteq \begin{cases} I_\rho(\gamma) - I_\rho(A(z)) & \text{if} \quad \gamma \in A(z) \\ \infty & \text{if} \quad \gamma \in \mathcal{P}_\alpha \backslash A(z). \end{cases}$$

This large deviation principle is not needed in the sequel, and the proof is omitted.

In the next section we will show how calculations analogous to those used to motivate Theorem 3 can be used to derive the form of the Gibbs state for the discrete ideal gas.

5. Gibbs states for models in statistical mechanics

The discussion in the previous section concerning a loaded die applies with minor changes to the discrete ideal gas, introduced in part 3 of Example 1. This system consists of n identical, noninteracting particles, each having α possible energy levels $y_1, y_2, \ldots, y_\alpha$. For $\omega \in \Omega_n = \Lambda^n$ we write $H_n(\omega)$ in place of $S_n(\omega) \doteq \sum_{j=1}^{n}\omega_j$; $H_n(\omega)$ denotes the total energy in the configuration ω. In the absence of further information, one assigns the equal probabilities $\rho_k = 1/\alpha$ to each of the y_k's. Defining $\bar{y} \doteq \sum_{k=1}^{\alpha} y_k\rho_k$, suppose that the energy per particle, H_n/n, is

conditioned to lie in an interval $[z - a, z]$, where a is a small positive number and $y_1 \leq z - a < z < \bar{y}$. According to part 3 of Theorem 3, for each $k \in \{1, \ldots, \alpha\}$

$$\rho_k^{(\beta)} = \lim_{n \to \infty} P_n \left\{ X_1 = y_k \Big| \frac{H_n}{n} \in [z - a, z] \right\},$$

where $\rho_k^{(\beta)} \doteq \exp[-\beta y_k] \rho_k / \sum_{j=1}^{\alpha} \exp[-\beta y_j] \rho_j$ and $\beta = \beta(z) \in \mathbb{R}$ is chosen so that $\sum_{k=1}^{\alpha} y_k \rho_k^{(\beta)} = z$.

Let $t \geq 2$ be a positive integer. The limit in the last display leads to a natural question. Conditioned on $H_n/n \in [z - a, z]$, as $n \to \infty$ what is the limiting conditional distribution of the random variables $X_1, \ldots, X_t$, which represent the energy levels of the first t particles? Although $X_1, \ldots, X_t$ are independent with respect to the original product measure P_n, this independence is lost when P_n is replaced by the conditional distribution $P_n\{\cdot | H_n/n \in [z - a, z]\}$. Hence the answer given in the next theorem is somewhat surprising: with respect to $P_n\{\cdot | H_n/n \in [z - a, z]\}$, the limiting distribution is the product measure on Ω_t with one-dimensional marginals $\rho^{(\beta)}$. In other words, in the limit $n \to \infty$ the independence of $X_1, \ldots, X_t$ is regained. The theorem leads to, and in a sense motivates, the concept of the Gibbs state of the discrete ideal gas. We will end the section by discussing Gibbs states for this and other statistical mechanical models. As in Theorem 3, a theorem analogous to the following would hold if $[z - a, z] \subset [y_1, \bar{y})$ were replaced by $[z, z + a] \subset (\bar{y}, y_\alpha]$.

Theorem 4. *Given $t \in \mathbb{N}$, $y_{k_1}, \ldots, y_{k_t} \in \Lambda$, and $[z - a, z] \subset [y_1, \bar{y})$,*

$$\lim_{n \to \infty} P_n \left\{ X_1 = y_{k_1}, \ldots, X_t = y_{k_t} \Big| \frac{H_n}{n} \in [z - a, z] \right\} = \prod_{j=1}^{t} \rho_{k_j}^{(\beta)}. \tag{7}$$

Comments on the proof. We consider $t = 2$; arbitrary $t \in \mathbb{N}$ can be handled similarly. For $\omega \in \Omega_n$ and $i, j \in \{1, \ldots, \alpha\}$ define

$$L_{n,2}(\{y_i, y_j\}) = L_{n,2}(\omega, \{y_i, y_j\}) \doteq \frac{1}{n} \left(\sum_{j=1}^{n-1} \delta_{X_j(\omega), X_{j+1}(\omega)} \{y_i, y_j\} + \delta_{X_n(\omega), X_1(\omega)} \{y_i, y_j\} \right).$$

This counts the relative frequency with which the pair $\{y_i, y_j\}$ appears in the configuration $(\omega_1, \ldots, \omega_n, \omega_1)$. We then define the empirical pair vector

$$L_{n,2} \doteq \{L_{n,2}(\{y_i, y_j\}), \quad i, j = 1, \ldots, \alpha\}.$$

This takes values in the set $\mathcal{P}_{\alpha,2}$ consisting of all $\tau = \{\tau_{i,j}, i, j = 1, \ldots, \alpha\}$ satisfying $\tau_{i,j} \geq 0$ and $\sum_{i,j=1}^{\alpha} \tau_{i,j} = 1$. Suppose one can show that $\tau^* \doteq \{\rho_i^{(\beta)} \rho_j^{(\beta)}, i, j = 1, \ldots, \alpha\}$ has the property that for every $\varepsilon > 0$

$$\lim_{n \to \infty} P_n \left\{ L_{n,2} \in B(\tau^*, \varepsilon) \Big| \frac{H_n}{n} \in [z - a, z] \right\} = 1. \tag{8}$$

Then as in Theorem 3, it will follow that

$$\lim_{n \to \infty} P_n \left\{ X_1 = y_i, X_2 = y_j \Big| \frac{H_n}{n} \in [z - a, z] \right\} = \rho_i^{(\beta)} \rho_j^{(\beta)}.$$

As the analogous limit in part 1 of Theorem 3 is derived, Eq. (8) can be proved by showing that the sequence $\{L_{n,2}, n \in \mathbb{N}\}$ satisfies the large deviation principle on $\mathcal{P}_{\alpha,2}$ ([1], Section I.5) and that the rate function attains its infimum over an appropriately defined, closed convex subset of $\mathcal{P}_{\alpha,2}$ at the unique point τ^* (cf. Eq. (6)). The details are omitted. □

The quantity appearing on the right side of Eq. (7) defines a probability measure $P_{t,\beta}$ on Ω_t which equals the product measure with one-dimensional marginals $\rho^{(\beta)}$. In the notation of Theorem 4,

$$P_{t,\beta}\{X_1 = y_{k_1}, \dots, X_t = y_{k_t}\} = \prod_{j=1}^{t} \rho_{k_j}^{(\beta)}.$$

$P_{t,\beta}$ can be written in terms of the total energy $H_t(\omega) \doteq \sum_{j=1}^{t} \omega_j$: for $\omega \in \Omega_t$

$$P_{t,\beta}\{\omega\} = \prod_{j=1}^{t} \rho^{(\beta)}\{\omega_j\} = \frac{1}{Z_t(\beta)} \exp[-\beta H_t(\omega)] \, P_t\{\omega\},$$

where $P_t\{\omega\} = \prod_{j=1}^{t} \rho\{\omega_j\} = 1/\alpha^t$,

$$Z_t(\beta) \doteq \sum_{\omega \in \Omega_t} \exp[-\beta H_t(\omega)] \, P_t\{\omega\} = \left(\sum_{k=1}^{\alpha} \exp[-\beta y_k] \rho_k \right)^t,$$

and $\beta = \beta(z) \in \mathbb{R}$ is the unique value of β for which $\sum_{k=1}^{\alpha} y_k \rho_k^{(\beta)} = z$ is valid.

Theorem 4 can be motivated by a non-large deviation calculation that we present using a formal notation [8]. Since $\bar{y} \doteq \sum_{k=1}^{\alpha} y_k \rho_k = E^{P_n}\{X_1\}$, by the weak law of large numbers $P_n\{H_n/n \sim \bar{y}\} \approx 1$ for large n. Since the conditioning is on a set of probability close to 1, one expects that

$$\lim_{n \to \infty} P_n \left\{ X_1 = y_{k_1}, \dots, X_t = y_{k_t} \Big| \frac{H_n}{n} \sim \bar{y} \right\} = \lim_{n \to \infty} P_n\{X_1 = y_{k_1}, \dots, X_t = y_{k_t}\} = \prod_{j=1}^{t} \rho_{k_j}$$

$$= P_t\{X_1 = y_{k_1}, \dots, X_t = y_{k_t}\}.$$

Now take $z \neq \bar{y}$ and for any $\beta > 0$ let $P_{n,\beta}$ denote the product measure on Ω_n with one-dimensional marginals $\rho^{(\beta)}$. A short calculation shows that for any $\beta > 0$

$$P_n \left\{ X_1 = y_{k_1}, \dots, X_t = y_{k_t} \Big| \frac{H_n}{n} \sim z \right\} = P_{n,\beta} \left\{ X_1 = y_{k_1}, \dots, X_t = y_{k_t} \Big| \frac{H_n}{n} \sim z \right\}.$$

If one picks $\beta = \beta(z)$ such that $z = \sum_{k=1}^{\alpha} y_k \rho_k^{(\beta(z))} = E^{P_{n,\beta(z)}}\{X_1\}$, then by the weak law of large numbers $P_{n,\beta(z)}\{H_n/n \sim z\} \approx 1$, and since the conditioning is on a set of probability close to 1, again one expects that

$$\lim_{n \to \infty} P_n \left\{ X_1 = y_{k_1}, \dots, X_t = y_{k_t} \Big| \frac{H_n}{n} \sim z \right\} = \lim_{n \to \infty} P_{n,\beta(z)} \left\{ X_1 = y_{k_1}, \dots, X_t = y_{k_t} \Big| \frac{H_n}{n} \sim z \right\}$$

$$= \lim_{n \to \infty} P_{n,\beta(z)}\{X_1 = y_{k_1}, \dots, X_t = y_{k_t}\} = \prod_{j=1}^{t} \rho_{k_j}^{(\beta(z))}$$

$$= P_{t,\beta(z)}\{X_1 = y_{k_1}, \dots, X_t = y_{k_t}\}.$$

This is consistent with Theorem 4.

For any subset B of Ω_t, Eq. (7) implies that

$$\lim_{n \to \infty} P_n \left\{ (X_1, \dots, X_t) \in B \Big| \frac{H_n}{n} \in [z - a, z] \right\} = P_{t,\beta}\{B\}. \tag{9}$$

Since $\sum_{\omega \in \Omega_t} [H_t(\omega)/t] \, P_{t,\beta}\{\omega\} = \sum_{k=1}^{\alpha} y_k \rho_k^{(\beta)}$, the constraint on $\beta = \beta(z)$ can be expressed as a constraint on $P_{t,\beta}$:

$$\text{choose } \beta = \beta(z) \text{ so that } \sum_{\omega \in \Omega_t} \left[\frac{H_t(\omega)}{t} \right] P_{t,\beta}\{\omega\} = z. \tag{10}$$

The conditional probability on the left side of Eq. (9) is known as the Gibbs microcanonical ensemble, and the probability on the right side of Eq. (9) as the Gibbs canonical ensemble or Gibbs state. This limit expresses the equivalence of the two ensembles provided β is chosen in accordance with Eq. (10). Since the Gibbs state has a much simpler form than the Gibbs microcanonical ensemble, one usually prefers to work with the former. One can interpret β as a parameter that is proportional to the inverse temperature.

This discussion motivates the definition of the Gibbs states for a wide class of statistical mechanical models that are defined in terms of an energy function. We will write the energy function, or Hamiltonian, and the corresponding Gibbs state as H_n and $P_{n,\beta}$ rather than as H_t and $P_{t,\beta}$, as we did in the preceding paragraph. The notation of Section 2 is used. Thus P_n is the product measure on the set of subsets of $\Omega_n \doteq \Lambda^n$ with one-dimensional marginals ρ.

Definition 2. Let H_n be a function mapping Ω_n into $\mathbb{R}$; $H_n(\omega)$ defines the energy of the configuration ω and is known as a Hamiltonian. Let β be a parameter proportional to the inverse temperature. Then the Gibbs canonical ensemble, or the Gibbs state, is the probability measure

$$P_{n,\beta}\{\omega\} \doteq \frac{1}{Z_n(\beta)} \exp[-\beta H_n(\omega)] \, P_n\{\omega\} \quad \text{for} \quad \omega \in \Omega_n,$$

where $Z_n(\beta)$ is the normalization factor that makes $P_{n,\beta}$ a probability measure. That is,

$$Z_n(\beta) \doteq \sum_{\omega \in \Omega_n} \exp[-\beta H_n(\omega)] \, P_n\{\omega\}.$$

We call $Z_n(\beta)$ the partition function. For $B \subset \Omega_n$ we have $P_{n,\beta}\{B\} = \sum_{\omega \in B} P_{n,\beta}\{\omega\}$.

Noninteracting systems such as the discrete ideal gas have Hamiltonians of the form $H_n(\omega) = \sum_{j=1}^{n} H_{n,j}(\omega_j)$. The equivalence of ensembles and related questions for interacting systems have been studied by a number of authors, including ([7], Section 7.3) [9–11].

One can also characterize Gibbs states in terms of a maximum entropy principle ([12], p. 6). Given $n \in \mathbb{N}$ and a Hamiltonian H_n, let $B_n \subset \mathbb{R}$ denote the smallest closed interval containing the range of $\{H_n(\omega)/n, \omega \in \Omega_n\}$. For each $z \in \text{int } B_n$ define $\Gamma_n(z)$ to be the set of probability measures Q on Ω_n satisfying the energy constraint $\sum_{\omega \in \Omega_n} [H_n(\omega)/n] \, Q\{\omega\} = z$.

Maximum Entropy Principle 3. *Let $n \in \mathbb{N}$ and a Hamiltonian $H_n : \Omega_n \mapsto \mathbb{R}$ be given. The following conclusions hold:*

1. *For each $z \in \text{int } B_n$ there exists a unique $\beta = \beta(z) \in \mathbb{R}$ such that $\sum_{\omega \in \Omega_n} [H_n(\omega)/n] \, P_{n,\beta}\{\omega\} = z$; i.e., such that $P_{n,\beta} \in \Gamma_n(z)$.*
2. *The relative entropy I_{P_n} attains its infimum over $\Gamma_n(z)$ at the unique measure $P_{n,\beta}$; $I_{P_n}(P_{n,\beta}) = -n(\beta z + c(-\beta)) = n I_\rho(\rho^\beta)$, where $c(r) \doteq \log(\sum_{k=1}^{\alpha} \exp[r y_k] \rho_k)$.*

Part 1 can be proved by a calculation similar to that given after the statement of Theorem 3 while part 2 can be proved like Proposition 1. We leave the details to the reader.

In the next section we formulate the general concept of a large deviation principle. Subsequent sections will apply the theory of large deviations to study interacting systems in statistical mechanics.

6. Definition of the Large Deviation Principle

In Theorem 2 we formulated Sanov's Theorem, which is the large deviation principle for the empirical vectors $\{L_n\}$ on the space $\mathcal{P}_\alpha$ of probability vectors in $\mathbb{R}^\alpha$. Applications of the theory of large deviations to models in statistical mechanics require large deviation principles in much more general settings. As we will see in the next section, analyzing the Curie–Weiss model of ferromagnetism involves a large deviation principle on the closed interval $[-1, 1]$ for the sample means of i.i.d. random variables. Analyzing the Ising model in dimensions $d \geq 2$ is much more complicated. It involves a large deviation principle on the space of translation invariant probability measures on $\{-1, 1\}^{\mathcal{Z}^d}$ ([13], Section 11). As we will see in Section 9, treating models of two-dimensional turbulence involves a large deviation principle on the space of probability measures on $T^2 \times \mathcal{Y}$, where T^2 is the unit torus in $\mathbb{R}^2$ and $\mathcal{Y}$ is a compact subset of $\mathbb{R}$.

In order to define the general concept of a large deviation principle, we need some notation. First, for each $n \in \mathbb{N}$ let $(\Omega_n, \mathcal{F}_n, Q_n)$ be a probability space. Thus Ω_n is a set of points, $\mathcal{F}_n$ is a σ-algebra of subsets of Ω_n, and Q_n is a probability measure on $\mathcal{F}_n$. An example is given by the basic model in Section 2, where $\Omega_n \doteq \Lambda^n = \{y_1, y_2, \ldots, y_\alpha\}^n$, $\mathcal{F}_n$ is the set of all subsets of Ω_n, and Q_n is the product measure with one-dimensional marginals ρ.

Second, let χ be a complete separable metric space or, as it is often called, a Polish space. Elementary examples are $\mathcal{X} \doteq \mathbb{R}^d$ for $d \in \mathbb{N}$; $\mathcal{X} \doteq \mathcal{P}_\alpha$, the set of probability vectors in $\mathbb{R}^\alpha$; and in the notation of the basic probabilistic model in Section 2, $\mathcal{X}$ equal to the closed bounded interval $[y_1, y_\alpha]$. A class of Polish spaces arising naturally in applications is obtained by taking a Polish space $\mathcal{Y}$ and considering the space $\mathcal{P}(\mathcal{Y})$ of probability measures on (the Borel subsets of) $\mathcal{Y}$. We say that a sequence $\{\Pi_n, n \in \mathbb{N}\}$ in $\mathcal{P}(\mathcal{Y})$ converges weakly to $\Pi \in \mathcal{P}(\mathcal{Y})$, and write $\Pi_n \Rightarrow \Pi$, if $\int f \, d\Pi_n \to \int f \, d\Pi$ for all bounded continuous functions f mapping $\mathcal{Y}$ into $\mathbb{R}$. A fundamental fact is that there exists a metric m on $\mathcal{P}(\mathcal{Y})$ such that $\Pi_n \Rightarrow \Pi$ if and only if $m(\Pi, \Pi_n) \to 0$ and $\mathcal{P}(\mathcal{Y})$ is a Polish space with respect to m ([14], Section 3.1).

Third, for each $n \in \mathbb{N}$ let Y_n be a random variable mapping Ω_n into $\mathcal{X}$. For example, with $\mathcal{X} \doteq \mathcal{P}_\alpha$ let $Y_n \doteq L_n$, or with $\mathcal{X} \doteq [y_1, y_\alpha]$ let $Y_n \doteq \sum_{j=1}^n X_j / n$, where $X_j(\omega) \doteq \omega_j$ for $\omega \in \Omega_n \doteq \Lambda^n$.

Finally, let I be a function mapping the Polish space $\mathcal{X}$ into $[0, \infty]$. I is called a rate function if I has compact level sets; i.e., for all $M < \infty \{x \in \mathcal{X} : I(x) \leq M\}$ is compact. This technical regularity condition implies that I is lower semicontinuous; if $\mathcal{X}$ is compact, then the lower semicontinuity of I implies that I has compact level sets. For any subset A of $\mathcal{X}$ we define $I(A) \doteq \inf_{x \in A} I(x)$. When $\mathcal{X} \doteq \mathcal{P}_\alpha$, an example of a rate function is the relative entropy I_ρ with respect to ρ; when $\mathcal{X} \doteq [y_1, y_\alpha]$, any continuous function I function mapping $[y_1, y_\alpha]$ into $[0, \infty)$ is a rate function.

Definition 3 (Large deviation principle). Let $\{(\Omega_n, \mathcal{F}_n, P_n), n \in \mathbb{N}\}$ be a sequence of probability spaces, $\mathcal{X}$ a Polish space, $\{Y_n, n \in \mathbb{N}\}$ a sequence of random variables such that Y_n maps Ω_n into $\mathcal{X}$, and I a rate function on $\mathcal{X}$. Then $\{Y_n\}$ satisfies the large deviation principle on $\mathcal{X}$ with rate function I if for any closed subset F of $\mathcal{X}$

$$\limsup_{n \to \infty} \frac{1}{n} \log Q_n \{Y_n \in F\} \leq -I(F)$$

and for any open subset G of $\mathcal{X}$

$$\liminf_{n \to \infty} \frac{1}{n} \log Q_n \{Y_n \in G\} \geq -I(G).$$

If $\{Y_n\}$ satisfies the large deviation principle with rate function I, then we summarize this by the formal notation

$$Q_n \{Y_n \in dx\} \approx \exp[-nI(x)] \, dx.$$

Evaluating the limit superior in Definition 3 for $F = \mathcal{X}$ and the limit inferior for $G = \mathcal{X}$ yields $I(\mathcal{X}) = 0$, and since I has compact level sets, the set of $x \in \mathcal{X}$ for which $I(x) = 0$ is nonempty and compact. The following result generalizes Corollary 2.

Theorem 5. *Suppose that $\{Y_n, n \in \mathbb{N}\}$ satisfies the large deviation principle on the Polish space $\mathcal{X}$ with rate function I. Define $\mathcal{E}$ to be the nonempty, compact set of $x \in \mathcal{X}$ for which $I(x) = 0$ and let A be a Borel subset of $\mathcal{X}$ such that $\bar{A} \cap \mathcal{E} = \emptyset$. Then $I(\bar{A}) > 0$ and for some $C < \infty$*

$$Q_n\{Y_n \in A\} \leq C\exp[-nI(\bar{A})/2] \to 0 \quad as \ n \to \infty.$$

For application in the next section, we state a special case of Cramér's Theorem, which is the large deviation principle for the sample means of i.i.d. random variables.

Theorem 6 (Cramér's Theorem). *In the basic probability model of Section 2, let $\Lambda \doteq \{-1, 1\}$ and let $\rho \doteq (1/2)\delta_{-1} + (1/2)\delta_1$. For $\omega \in \Omega_n$ define $S_n(\omega) \doteq \sum_{j=1}^{n} \omega_j$. Then the sequence of sample means $\{S_n/n, n \in \mathbb{N}\}$ satisfies the large deviation principle on $[-1, 1]$ with rate function*

$$I(x) \doteq \tfrac{1}{2}(1-x)\log(1-x) + \tfrac{1}{2}(1+x)\log(1+x).$$

This theorem is easy to motivate using the formal notation of Theorem 1. For any $x \in [-1, 1]$, $S_n(\omega)/n \sim x$ if and only if approximately $(n/2)(1-x)$ of the ω_j's equal -1 and approximately $(n/2)(1+x)$ of the ω_j's equal 1. Hence

$$P_n\left\{\frac{S_n}{n} \sim x\right\} \approx P_n\{L_n(-1) = \tfrac{1}{2}(1-x), L_n(1) = \tfrac{1}{2}(1+x)\} \approx \exp[-nI_\rho(\tfrac{1}{2}(1-x), \tfrac{1}{2}(1+x))]$$
$$= \exp[-nI(x)].$$

The book [7] presents Cramér's Theorem first in the setting of $\mathbb{R}^d$ and then in the setting of a Polish space.

For application in Section 9, we state a general version of Sanov's Theorem, which gives the large deviation principle for the sequence of empirical measures of i.i.d. random variables. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{Y}$ a Polish space, ρ a probability measure on $\mathcal{Y}$, and $\{X_j, j \in \mathbb{N}\}$ a sequence of i.i.d. random variables mapping Ω into $\mathcal{Y}$ and having the common distribution ρ. For $\omega \in \Omega$ and A any Borel subset of $\mathcal{Y}$ we define the empirical measure

$$L_n(A) = L_n(\omega, A) \doteq \frac{1}{n}\sum_{j=1}^{n}\delta_{X_j(\omega)}\{A\},$$

where for $y \in \mathcal{Y}$ $\delta_y\{A\}$ equals 1 if $y \in A$ and 0 if $y \notin A$. For each ω $L_n(\omega, \cdot)$ is a probability measure on $\mathcal{Y}$. Hence the sequence $\{L_n, n \in \mathbb{N}\}$ takes values in the Polish space $\mathcal{P}(\mathcal{Y})$.

Theorem 7 (Sanov's Theorem). *The sequence $\{L_n, n \in \mathbb{N}\}$ satisfies the large deviation principle on $\mathcal{P}(\mathcal{Y})$ with rate function given by the relative entropy with respect to ρ. For $\gamma \in \mathcal{P}(\mathcal{Y})$ this quantity is defined by*

$$I_\rho(\gamma) \doteq \begin{cases} \int_{\mathcal{Y}} (\log(d\gamma/d\rho)) \, d\gamma & if \quad \gamma \ll \rho \\ \infty & otherwise. \end{cases}$$

This theorem is proved, for example, in ([7], Section 6.2) and in ([15], Ch. 2). If the support of ρ is a finite set $\Lambda \subset \mathbb{R}$, then Theorem 7 reduces to Theorem 2.

The following concept will be useful in the analysis of statistical mechanical models.

Definition 4 (Laplace Principle). Let $\{(\Omega_n, \mathcal{F}_n, P_n), n \in \mathbb{N}\}$ be a sequence of probability spaces, $\mathcal{X}$ a Polish space, $\{Y_n, n \in \mathbb{N}\}$ a sequence of random variables such that Y_n maps Ω_n into $\mathcal{X}$, and I a rate function on $\mathcal{X}$. Then $\{Y_n\}$

satisfies the Laplace principle on $\mathcal{X}$ with rate function I if for all bounded continuous functions f mapping $\mathcal{X}$ into $\mathbb{R}$

$$\lim_{n\to\infty}\frac{1}{n}\log E^{Q_n}\{\exp[nf(Y_n)]\} = \lim_{n\to\infty}\frac{1}{n}\int_{\mathcal{X}}\exp[nf(x)]\,Q_n\{Y_n \in \mathrm{d}x\} = \sup_{x\in\mathcal{X}}\{f(x) - I(x)\}.$$

Suppose that $\{Y_n\}$ satisfies the large deviation principle on $\mathcal{X}$ with rate function I. Then substituting $Q_n\{Y_n \in \mathrm{d}x\} \approx \exp[-nI(x)]\,\mathrm{d}x$ gives

$$\frac{1}{n}\log E^{Q_n}\{\exp[nf(Y_n)]\} = \frac{1}{n}\log\int_{\mathcal{X}}\exp[nf(x)]\,Q_n\{Y_n \in \mathrm{d}x\} \approx \frac{1}{n}\log\int_{\mathcal{X}}\exp[nf(x)]\exp[-nI(x)]\,\mathrm{d}x.$$

By analogy with Laplace's method on $\mathbb{R}$, the main contribution to the last integral should come from the largest value of the integrand, and thus the following limit should hold:

$$\frac{1}{n}\log E^{Q_n}\{\exp[nf(Y_n)]\} = \sup_{x\in\mathcal{X}}\{f(x) - I(x)\}.$$

Hence it is plausible that $\{Y_n\}$ satisfies the Laplace principle with rate function I. In fact, it is not difficult to prove that $\{Y_n\}$ satisfies the large deviation principle on $\mathcal{X}$ with rate function I if and only if $\{Y_n\}$ satisfies the Laplace principle on $\mathcal{X}$ with rate function I ([15], Theorems 1.2.1 and 1.2.3). As we will see in the next section, where the Curie–Weiss model is studied, the Laplace principle gives a variational formula for the specific Gibbs free energy.

7. The Curie–Weiss model of ferromagnetism

The Curie–Weiss model of ferromagnetism is one of the simplest examples of an interacting system in statistical mechanics. Using the theory of large deviations to analyze it suggests how one can apply the theory to analyze much more complicated models. The Curie–Weiss model is a spin system on the configuration spaces $\Omega_n \doteq \{-1, 1\}^n$; the value -1 represents 'spin-down' and the value 1 'spin-up'. Let $\rho \doteq (1/2)\delta_{-1} + (1/2)\delta_1$ and let P_n denote the product measure on Ω_n with one-dimensional marginals ρ. Thus $P_n\{\omega\} = 1/2^n$ for each configuration $\omega = \{\omega_i, i = 1, \ldots, n\} \in \Omega_n$. The Hamiltonian, or energy, of a configuration ω is defined by

$$H_n(\omega) \doteq -\frac{1}{2n}\sum_{i,j=1}^{n}\omega_i\omega_j = -\frac{n}{2}\left(\frac{1}{n}\sum_{j=1}^{n}\omega_j\right)^2, \tag{11}$$

and the probability of a configuration corresponding to inverse temperature $\beta > 0$ is defined by the Gibbs state

$$P_{n,\beta}\{\omega\} \doteq \frac{1}{Z_n(\beta)}\exp[-\beta H_n(\omega)]\,P_n\{\omega\},$$

where $Z_n(\beta)$ is the partition function

$$Z_n(\beta) \doteq \int_{\Omega_n}\exp[-\beta H_n(\omega)]\,P_n(\mathrm{d}\omega) = \sum_{\omega\in\Omega_n}\exp[-\beta H_n(\omega)]\,\frac{1}{2^n}.$$

$P_{n,\beta}$ models a ferromagnet in the sense that the maximum of $P_{n,\beta}\{\omega\}$ over $\omega \in \Omega_n$ occurs at the two configurations having all coordinates ω_i equal to -1 or all coordinates equal to 1. Furthermore, as $\beta \to \infty$ all the mass of $P_{n,\beta}$ concentrates on these two configurations. The Curie–Weiss model is often used as a mean-field approximation to the much more complicated Ising model and related ferromagnetic models ([1], Section V.9).

A distinguishing feature of the Curie–Weiss model is its phase transition. Namely, the alignment effects incorporated in the Gibbs states $P_{n,\beta}$ persist in the limit $n \to \infty$. This is most easily seen by evaluating the $n \to \infty$ limit of the distributions $P_{n,\beta}\{S_n/n \in \mathrm{d}x\}$, where $S_n(\omega)/n$ equals the spin per site $\sum_{j=1}^{n} \omega_j/n$. We will see that for $\beta \le 1$ this limit acts like the classical weak law of large numbers, concentrating on the value 0. However, for $\beta > 1$ the analogy with the classical law of large numbers breaks down; the alignment effects are so strong that the limiting $P_{n,\beta}$-distribution of S_n/n concentrates on the two points $\pm m(\beta)$ for some $m(\beta) \in (0, 1)$. The analysis of the Curie–Weiss model to be presented below can be easily modified to handle an external magnetic field h. The resulting probabilistic description of the phase transition yields the predictions of mean field theory ([1], Section V.9) and ([12], Section 3.2).

We calculate the $n \to \infty$ limit of $P_{n,\beta}\{S_n/n \in \mathrm{d}x\}$ by establishing a large deviation principle for the spin per site with respect to $P_{n,\beta}$. For each n S_n/n takes values in $[-1, 1]$. By the equivalence between the Laplace principle and the large deviation principle mentioned at the end of the previous section, it suffices to find a rate function I_β on $[-1, 1]$ such that for any continuous function f mapping $[-1, 1]$ into $\mathbb{R}$

$$\lim_{n \to \infty} \frac{1}{n} \log E^{P_{n,\beta}} \left\{ \exp\left[nf\left(\frac{S_n}{n} \right) \right] \right\} = \sup_{x \in [-1,1]} \{f(x) - I_\beta(x)\}.$$

Substituting $H_n(\omega) = -(n/2)(S_n/n)^2$ gives

$$\frac{1}{n} \log E^{P_{n,\beta}} \left\{ \exp\left[nf\left(\frac{S_n}{n} \right) \right] \right\} = \frac{1}{Z_n(\beta)} \int_{\Omega_n} \exp\left[nf\left(\frac{S_n}{n} \right) + n\left(\frac{\beta}{2} \right)\left(\frac{S_n}{n} \right)^2 \right] P_n(\mathrm{d}\omega)$$

$$= \frac{1}{Z_n(\beta)} \int_{[-1,1]} \exp\left[nf(x) + n\left(\frac{\beta}{2} \right)x^2 \right] P_n\left\{ \frac{S_n}{n} \in \mathrm{d}x \right\}. \tag{12}$$

Noticing that

$$Z_n(\beta) = \int_{\Omega_n} \exp\left[n\left(\frac{\beta}{2} \right)\left(\frac{S_n}{n} \right)^2 \right] P_n(\mathrm{d}\omega) = \int_{[-1,1]} \exp\left[n\left(\frac{\beta}{2} \right)x^2 \right] P_n\left\{ \frac{S_n}{n} \in \mathrm{d}x \right\},$$

we apply Cramér's Theorem 6 twice, in the equivalent form of the Laplace principle. Thus

$$\lim_{n \to \infty} \frac{1}{n} \log Z_n(\beta) = \sup_{x \in [-1,1]} \left\{ \left(\frac{\beta}{2} \right)x^2 - I(x) \right\} \tag{13}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \log \int_{[-1,1]} \exp\left[nf(x) + n\left(\frac{\beta}{2} \right)x^2 \right] P_n\left\{ \frac{S_n}{n} \in \mathrm{d}x \right\} = \sup_{x \in [-1,1]} \left\{ f(x) + \left(\frac{\beta}{2} \right)x^2 - I(x) \right\}, \tag{14}$$

where $I(x) \doteq (1/2)(1 - x)\log(1 - x) + (1/2)(1 + x)\log(1 + x)$. For $x \in [-1, 1]$ define

$$I_\beta(x) \doteq I(x) - \left(\frac{\beta}{2} \right)x^2 - \inf_{y \in [-1,1]} \left\{ I(y) - \left(\frac{\beta}{2} \right)y^2 \right\}. \tag{15}$$

Equations (13) and (14) give the $n \to \infty$ limit of Eq. (12):

$$\lim_{n \to \infty} \frac{1}{n} \log E^{P_{n,\beta}} \left\{ \exp\left[nf\left(\frac{S_n}{n} \right) \right] \right\} = \sup_{x \in [-1,1]} \{f(x) - I_\beta(x)\}.$$

We conclude that with respect to $\{P_{n,\beta}\}$ the sequence $\{S_n/n\}$ satisfies the Laplace principle on $[-1, 1]$, and thus the large deviation principle, with rate function I_β.

The limiting behavior of the distributions $P_{n,\beta}\{S_n/n \in \mathrm{d}x\}$ is now determined by examining where I_β attains its infimum of 0 ([1], Section IV.4). Infimizing points x^* satisfy $I_\beta'(x^*) = 0$ or $I'(x^*) = \beta x^*$. Since $I''(0) = 1$ and I' is concave on $[-1, 0]$, convex on $[0, 1]$, for $0 < \beta \leq 1$ the only solution to this equation is $x^* = 0$. For $\beta > 1$ there are three solutions $0, m(\beta), -m(\beta)$, where $0 < m(\beta) < 1$; of these, 0 is a local maximum and $\pm m(\beta)$ are the infimizers. The function $m(\beta)$ is monotonically increasing on $(1, \infty)$; $m(\beta) \to 0$ as $\beta \to 1^+$ and $m(\beta) \to 1$ as $\beta \to \infty$. The equation $I'(x^*) = \beta x^*$ is equivalent to the well known mean field equation $x^* = (I')^{-1}(\beta x^*) = \tanh(\beta x^*)$ ([1], Section V.9) and ([12], Section 3.2).

With $Q_n \doteq P_{n,\beta}$ and $Y_n \doteq S_n/n$, we now apply Theorem 5 for $0 < \beta \leq 1$ to any closed subset $A \subset [-1, 1]$ that does not contain 0 and for $\beta > 1$ to any closed subset $A \subset [-1, 1]$ that does not contain $\pm m(\beta)$. Since $I_\beta(\bar{A}) > 0$ and $P_{n,\beta}\{S_n/n \in A\} \leq C\exp[-nI_\beta(\bar{A})/2] \to 0$ as $n \to \infty$, we are led to the following weak limits:

$$P_{n,\beta}\left\{ \frac{1}{n}\sum_{i=1}^n \omega_i \in \mathrm{d}x \right\} \Rightarrow \begin{cases} \delta_0 & \text{if} \quad 0 < \beta \leq 1 \\ (1/2)\delta_{m(\beta)} + (1/2)\delta_{-m(\beta)} & \text{if} \quad \beta > 1. \end{cases} \tag{16}$$

We call $m(\beta)$ the spontaneous magnetization for the Curie–Weiss model and $\beta_c = 1$ the critical inverse temperature ([1], Section IV.4).

For each $\beta > 0$ we define $\mathcal{E}_\beta \doteq \{x \in [-1, 1] : I_\beta(x) = 0\}$; thus, $\mathcal{E}_\beta \doteq \{0\}$ for $0 < \beta \leq 1$ and $\mathcal{E}_\beta \doteq \{\pm m(\beta)\}$ for $\beta > 1$. The limit (16) justifies calling $\mathcal{E}_\beta$ the set of equilibrium macrostates for the spin per site S_n/n in the Curie–Weiss model. It is not difficult to show that points $x^* \in \mathcal{E}_\beta$ have an equivalent characterization in terms of a maximum entropy principle. Because of the relatively simple nature of the model, this maximum entropy principle takes a rather trivial form. The details are omitted.

Before leaving the Curie–Weiss model, there are several crucial points that should be emphasized. The first is to understand what makes possible the large deviation analysis of the model. In Eq. (11) we write the Hamiltonian as a quadratic function of the spin per site S_n/n, which by Cramér's Theorem 6 satisfies the large deviation principle on $[-1, 1]$ with respect to the product measures P_n. The equivalent Laplace principle allows us to convert this large deviation principle into a large deviation principle with respect to the Gibbs states $P_{n,\beta}$. The form of the rate function I_β allows us to complete the analysis. As we will see in the next section, this insight is fundamental in understanding how the theory of large deviations can be applied to more complicated models.

The second crucial point involves the variational formula derived in Eq. (13). If $Z_n(\beta)$ is a partition function of a statistical mechanical model on the configuration space $\Omega_n = \{-1, 1\}^n$, then $(-1/\beta)$ times $\lim_{n\to\infty}(1/n)\log Z_n(\beta)$ defines a quantity known as the specific Gibbs free energy. There is an analogous definition for models on other configuration spaces. A general statistical mechanical principle characterizes the set of equilibrium macrostates as those that give the extremum in the variational formula for the specific Gibbs free energy. In the case of the Curie–Weiss model, this variational formula is given in Eq. (13); x^* gives the supremum of $(\beta/2)x^2 - I(x)$ over $[-1, 1]$ if and only if $I_\beta(x^*) = 0 = \inf_{x \in [-1,1]} I_\beta(x)$. This holds if and only if $x^* \in \mathcal{E}_\beta$. Our large deviation analysis of the phase transition in the Curie–Weiss model has the attractive feature that, rather than appeal to a general statistical mechanical principle, it directly motivates the physical importance of $\mathcal{E}_\beta$. This set is the support of the $n \to \infty$ limit of the distributions $P_{n,\beta}\{S_n/n \in \mathrm{d}x\}$. As we will see in the next section, an analogous fact is true for a large class of statistical mechanical models (Theorem 9).

The third crucial point is related to the second. The large deviation analysis of the Curie–Weiss model yields the limiting behavior of the $P_{n,\beta}$-distributions of S_n/n. This limit corresponds to the classical weak law of large numbers for the sample means of i.i.d. random variables and suggests examining the analogues of other classical limit results such as the central limit theorem. Such limit theorems are derived and their statistical mechanical implications are explained in ([1], Section V.9) and in [16–18]. Related work has been done for the Curie–Weiss–Potts model [19,20]

as well as for the Ising and other models. For the latter models, refined large deviations at the surface level have been studied; see ([7], p. 339) for references.

8. A general approach to the large deviation analysis of models in statistical mechanics

By abstracting the calculations in the last section, we can give a general approach to the large deviation analysis of models in statistical mechanics. This approach will be applied in the next section to two-dimensional turbulence. We consider a class of models that are defined in terms of the following data.

- A sequence of probability spaces $\{(\Omega_n, \mathcal{F}_n, P_n), n \in \mathbb{N}\}$; $\{\Omega_n\}$ are the configuration spaces.
- For each $n \in \mathbb{N}$ the Hamiltonian $H_n(\omega)$ of $\omega \in \Omega_n$; H_n is a bounded measurable function mapping Ω_n into $\mathbb{R}$.
- A sequence of positive scaling constants $\{a_n, n \in \mathbb{N}\}$ such that $a_n \to \infty$.

In terms of these quantities we define for each $n \in \mathbb{N}$, $\beta \in \mathbb{R}$, and set $B \in \mathcal{F}_n$ the partition function

$$Z_n(\beta) \doteq \int_{\Omega_n} \exp[-\beta H_n(\omega)] P_n(\mathrm{d}\omega),$$

which is well defined and finite, and the Gibbs state

$$P_{n,\beta}\{B\} \doteq \frac{1}{Z_n(\beta)} \int_B \exp[-\beta H_n(\omega)] P_n(\mathrm{d}\omega).$$

Although for spin systems one usually takes $\beta > 0$, in general $\beta \in \mathbb{R}$ is allowed; for example, negative values of β arise naturally in the study of two-dimensional turbulence. For $\beta \in \mathbb{R}$ we define

$$\varphi(\beta) \doteq \lim_{n \to \infty} \frac{1}{a_n} \log Z_n(\beta)$$

if the limit exists and is nontrivial. The function $-\beta^{-1}\varphi(\beta)$ is the specific Gibbs free energy for the model. As in the Curie–Weiss model, one of our goals is to express $\varphi(\beta)$ as a variational formula of the form

$$\varphi(\beta) = \sup_{x \in \mathcal{X}} \{-\beta \tilde{H}(x) - I(x)\} = -\inf_{x \in \mathcal{X}} \{\beta \tilde{H}(x) + I(x)\}, \tag{17}$$

where $\mathcal{X}$ is some Polish space, $\tilde{H}$ is a bounded continuous function mapping $\mathcal{X}$ into $\mathbb{R}$, and I is a rate function on $\mathcal{X}$. We would also like to use large deviation methodology to interpret probabilistically the points $x^* \in \mathcal{X}$ that give the extremum in such variational formulas.

Before continuing with the general analysis, we recall how we proceeded in the case of the Curie–Weiss model. For that model $\Omega_n = \{-1, 1\}^n$, $\mathcal{F}_n$ is the set of subsets of Ω_n, and P_n is the product measure with one-dimensional marginals $(1/2)\delta_{-1} + (1/2)\delta_1$. H_n is defined in Eq. (11), and $\varphi(\beta)$ is expressed in terms of the variational formula (13). In order to derive this formula as well as the large deviation principle for $\{S_n/n\}$ with respect to the Gibbs states, we rewrote H_n as a quadratic function of S_n/n and used the Laplace principle for $\{S_n/n\}$ given by Cramér's Theorem 6.

Numerous other models can be treated analogously. For example, the Curie–Weiss–Potts model of ferromagnetism is a spin system on the configuration spaces $\Omega_n \doteq \Lambda^n$, where $\Lambda \doteq \{1, 2, \ldots, q\}$ and $q \geq 3$; $\mathcal{F}_n$ is the set of subsets of Ω_n; and P_n is the product measure with one-dimensional marginals $(1/q)\sum_{i=1}^q \delta_i$. The Hamiltonian H_n for the model equals $-(1/2)\langle L_n, L_n \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{R}^q$ and L_n is the empirical vector on Λ. The elementary form of Sanov's Theorem given in Theorem 2 for the empirical vectors $\{L_n\}$ allows one to derive a variational formula for $\varphi(\beta)$ as in Eq. (17) as well as a large deviation principle for $\{L_n\}$ with respect to the Gibbs states ([13], Section 10).

Although much more complicated, the Ising model on $\mathcal{Z}^D$, $D \geq 2$, is also amenable to a large deviation analysis. Let Δ_n be the hypercube in $\mathcal{Z}^D$ consisting of sites $i = (i_1, \ldots, i_D)$ each coordinate of which satisfies $1 \leq i_j \leq n$; Δ_n contains n^D points. The Ising model is a spin system on the configuration spaces Ω_n consisting of $\omega = \{\omega_i, i \in \Delta_n\}$ such that each $\omega_i \in \{-1, 1\}$. We write $\Omega_n \dot{=} \{-1, 1\}^{\Delta_n}$. The scaling constants a_n equal the number of sites in Δ_n; thus $a_n \dot{=} n^D$. For the Ising model, $\mathcal{F}_n$ is the set of subsets of Ω_n and P_n is the product measure with one-dimensional marginals $(1/2)\delta_{-1} + (1/2)\delta_1$. The Hamiltonian H_n is a sum over nearest neighbor pairs in Δ_n. Up to a small error as in Eq. (20), one rewrites H_n in terms of an infinite dimensional generalization of the empirical measure known as the empirical field. Using the large deviation principle for the latter with respect to the product measures $\{P_n\}$ derived in either of the papers [21,22], one expresses $\varphi(\beta)$ in terms of a variational formula as in Eq. (17) and derives a large deviation principle for the empirical fields with respect to the Gibbs states. The argument is outlined in ([13], Section 11).

This discussion points the way to a general approach. First, we update two definitions given in Section 6. Let $\{(\Omega_n, \mathcal{F}_n, Q_n), n \in \mathbb{N}\}$ be a sequence of probability spaces, $\mathcal{X}$ a Polish space, $\{Y_n, n \in \mathbb{N}\}$ a sequence of random variables such that Y_n maps Ω_n into $\mathcal{X}$, and I a rate function on $\mathcal{X}$. Then $\{Y_n\}$ is said to satisfy the large deviation principle on $\mathcal{X}$ with scaling constants $\{a_n\}$ and rate function I if for any closed subset F of $\mathcal{X}$ the large deviation upper bound

$$\limsup_{n \to \infty} \frac{1}{a_n} \log Q_n\{Y_n \in F\} \leq -I(F) \tag{18}$$

is valid and for any open subset G of $\mathcal{X}$ the large deviation lower bound

$$\liminf_{n \to \infty} \frac{1}{a_n} \log Q_n\{Y_n \in G\} \geq -I(G) \tag{19}$$

is valid. $\{Y_n\}$ is said to satisfy the Laplace principle on $\mathcal{X}$ with scaling constants $\{a_n\}$ and rate function I if for all bounded continuous functions f mapping $\mathcal{X}$ into $\mathbb{R}$

$$\lim_{n \to \infty} \frac{1}{a_n} \log E^{Q_n}\{\exp[a_n f(Y_n)]\} = \lim_{n \to \infty} \frac{1}{a_n} \int_{\mathcal{X}} \exp[a_n f(x)]\, Q_n\{Y_n \in \mathrm{d}x\} = \sup_{x \in \mathcal{X}}\{f(x) - I(x)\}.$$

As pointed out in ([15], Theorems 1.2.1 and 1.2.3), $\{Y_n\}$ satisfies the large deviation principle with scaling constants $\{a_n\}$ and rate function I if and only if $\{Y_n\}$ satisfies the Laplace principle with scaling constants $\{a_n\}$ and rate function I.

As we will see, a large deviation analysis of the general model is possible provided the following can be determined.

- *Hidden space.* This is a Polish space $\mathcal{X}$.
- *Hidden process.* This is a sequence $\{Y_n, n \in \mathbb{N}\}$, where for each $n \in \mathbb{N}$, Y_n is a random variable mapping Ω_n into $\mathcal{X}$.
- *Hamiltonian representation function.* This is a bounded continuous function $\tilde{H}$ mapping $\mathcal{X}$ into $\mathbb{R}$ such that for each $n \in \mathbb{N}$

$$H_n(\omega) = a_n \tilde{H}(Y_n(\omega)) + \mathrm{o}(a_n) \ \text{ uniformly for } \ \omega \in \Omega_n; \tag{20}$$

i.e.,

$$\lim_{n \to \infty} \sup_{\omega \in \Omega_n} \frac{1}{a_n} |H_n(\omega) - a_n \tilde{H}(Y_n(\omega))| = 0.$$

- *Large deviation principle for the hidden process.* There exists a rate function I mapping $\mathcal{X}$ into $[0, \infty]$ such that with respect to $\{P_n\}$ the sequence $\{Y_n\}$ satisfies the large deviation principle on $\mathcal{X}$, or equivalently the Laplace principle on $\mathcal{X}$, with scaling constants $\{a_n\}$ and rate function I.

For example, in the case of the Curie–Weiss model a_n equals n, $\mathcal{X}$ equals $[-1, 1]$, Y_n equals the spin per site S_n/n, and for $x \in [-1, 1]$ the Hamiltonian representation function is given by $\tilde{H}(x) \doteq -(1/2)x^2$. Equation (20) holds without any error term; i.e., $H_n(\omega) = n\tilde{H}(Y_n)$. In the case of the Curie–Weiss–Potts model, a_n equals n, $\mathcal{X}$ equals $\mathcal{P}_q$, Y_n equals L_n, and for $\gamma \in \mathcal{P}_q$ the Hamiltonian representation function is given by $\tilde{H}(\gamma) \doteq -(1/2)\langle \gamma, \gamma \rangle$. As in the Curie–Weiss model, Eq. (20) holds without any error term. In the case of the Ising model on $\mathcal{Z}^D$, a_n equals n^D, the hidden process is the sequence of empirical fields on $\Omega_n \doteq \{-1, 1\}^{\Delta_n}$, and the hidden space is the space of translation invariant probability measures on $\{-1, 1\}^{\mathcal{Z}^d}$. The form of the Hamiltonian representation function is given in ([13], Section 11); Eq. (20) is valid with an error term that represents boundary effects. While for numerous other models the hidden space, the hidden process, and the Hamiltonian representation function can be identified, in general it is not obvious how to determine them. This explains our choice of the term 'hidden'.

We now return to the general case. The large deviation analysis of the general model is summarized in the next theorem. Part 1 states a variational formula for the specific Gibbs free energy and part 2 the large deviation principle for the hidden process with respect to the sequence of Gibbs states. Part 3 describes probabilistically the set $\mathcal{E}_\beta$ of equilibrium macrostates, which is the set of points at which the rate function in part 2 attains its infimum of 0.

Theorem 8. *We assume that there exists a hidden space $\mathcal{X}$, a hidden process $\{Y_n, n \in \mathbb{N}\}$, and a Hamiltonian representation function $\tilde{H}$ and that with respect to $\{P_n\}$ the hidden process satisfies the large deviation principle on $\mathcal{X}$ with scaling constants $\{a_n\}$ and some rate function I. For each $\beta \in \mathbb{R}$ the following conclusions hold,*

1. $\varphi(\beta) \doteq \lim_{n \to \infty} (1/a_n) \log Z_n(\beta)$ *exists and is given by*

$$\varphi(\beta) = -\inf_{x \in \mathcal{X}} \{\beta \tilde{H}(x) + I(x)\}.$$

2. *With respect to the Gibbs states $\{P_{n,\beta}\}$ the hidden process $\{Y_n\}$ satisfies the large deviation principle on $\mathcal{X}$ with scaling constants $\{a_n\}$ and rate function*

$$I_\beta(x) \doteq I(x) + \beta \tilde{H}(x) - \inf_{y \in \mathcal{X}} \{I(y) + \beta \tilde{H}(y)\}.$$

3. *We define the set of equilibrium macrostates*

$$\mathcal{E}_\beta \doteq \{x \in \mathcal{X} : I_\beta(x) = 0\}.$$

Then $\mathcal{E}_\beta$ is a nonempty, compact subset of $\mathcal{X}$. In addition, if A is a Borel subset of $\mathcal{X}$ such that $\bar{A} \cap \mathcal{E}_\beta = \emptyset$, then $I_\beta(\bar{A}) > 0$ and for some $C < \infty$

$$P_{n,\beta}\{Y_n \in A\} \le C \exp\left[-\frac{n I_\beta(\bar{A})}{2}\right] \to 0 \quad as \ n \to \infty.$$

Proof. The proofs of parts 1 and 2 follow the similar calculations for the Curie–Weiss model once we take into account the error between H_n and $a_n \tilde{H}(Y_n)$ expressed in Eq. (20).

1. By Eq. (20)

$$\left| \frac{1}{a_n} \log Z_n(\beta) - \frac{1}{a_n} \log \int_{\Omega_n} \exp[-\beta a_n \tilde{H}(Y_n)] \, dP_n \right|$$

$$= \left| \frac{1}{a_n} \log \int_{\Omega_n} \exp[-\beta H_n] \, dP_n - \frac{1}{a_n} \log \int_{\Omega_n} \exp[-\beta a_n \tilde{H}(Y_n)] \, dP_n \right|$$

$$\le |\beta| \frac{1}{a_n} \sup_{\omega \in \Omega_n} |H_n(\omega) - a_n \tilde{H}(Y_n(\omega))| \to 0 \quad \text{as } n \to \infty.$$

Since $\tilde{H}$ is a bounded continuous function mapping $\mathcal{X}$ into $\mathbb{R}$, the Laplace principle satisfied by $\{Y_n\}$ with respect to $\{P_n\}$ yields part 1:

$$\lim_{n\to\infty}\frac{1}{a_n}\log Z_n(\beta) = \lim_{n\to\infty}\frac{1}{a_n}\log\int_{\Omega_n}\exp[-\beta a_n\tilde{H}(Y_n)]\,\mathrm{d}P_n = -\inf_{x\in\mathcal{X}}\{\beta\tilde{H}(x)+I(x)\}.$$

2. We proceed as in the proof of part 1, but now with $P_{n,\beta}$ replacing P_n. For any bounded continuous function f mapping $\mathcal{X}$ into $\mathbb{R}$, again Eq. (20) and the Laplace principle satisfied by $\{Y_n\}$ with respect to $\{P_n\}$ yield

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{a_n}\log\int_{\Omega_n}\exp[a_n f(Y_n)]\,\mathrm{d}P_{n,\beta} &= \lim_{n\to\infty}\frac{1}{a_n}\log\int_{\Omega_n}\exp[a_n f(Y_n)-\beta H_n]\,\mathrm{d}P_n - \lim_{n\to\infty}\frac{1}{a_n}\log Z_n(\beta)\\
&= \lim_{n\to\infty}\frac{1}{a_n}\log\int_{\Omega_n}\exp[a_n(f(Y_n)-\beta\tilde{H}(Y_n))]\,\mathrm{d}P_n - \lim_{n\to\infty}\frac{1}{a_n}\log Z_n(\beta)\\
&= \sup_{x\in\mathcal{X}}\{f(x)-\beta\tilde{H}(x)-I(x)\} + \inf_{x\in\mathcal{X}}\{\beta\tilde{H}(x)+I(x)\}\\
&= \sup_{x\in\mathcal{X}}\{f(x)-I_\beta(x)\}.
\end{aligned}$$

By hypothesis, I has compact level sets and $\tilde{H}$ is bounded and continuous. Thus I_β has compact level sets. Since I_β maps $\mathcal{X}$ into $[0,\infty]$, I_β is a rate function. We conclude that with respect to $\{P_{n,\beta}\}$ the sequence $\{Y_n\}$ satisfies the Laplace principle, and thus the large deviation principle, with scaling constants $\{a_n\}$ and rate function I_β.

3. As pointed out before Theorem 5, since the infimum of I_β over $\mathcal{X}$ equals 0, $\mathcal{E}_\beta$ is the set of minimum points of I_β over $\mathcal{X}$ and is nonempty and compact. If $\bar{A}\cap\mathcal{E}_\beta=\emptyset$, then for each $x\in\bar{A}$ $I_\beta(x)>0$. Since I_β is a rate function, it follows that $I_\beta(\bar{A})>0$. The large deviation upper bound completes the proof of part 3. □

Part 3 of the theorem can be regarded as a concentration property of the $P_{n,\beta}$-distributions of Y_n which justifies calling $\mathcal{E}_\beta$ the set of equilibrium macrostates. With respect to these distributions, the probability of any Borel set A whose closure has empty intersection with $\mathcal{E}_\beta$ goes to 0 exponentially fast with a_n. This large deviation characterization of the equilibrium macrostates is an attractive feature of our approach.

The concentration property of the $P_{n,\beta}$-distributions of Y_n as expressed in part 3 of the theorem has a refinement that arises in our study of the Curie–Weiss model. From Section 7 we recall that $\mathcal{E}_\beta=\{0\}$ for $0<\beta\leq 1$ and $\mathcal{E}_\beta=\{\pm m(\beta)\}$ for $\beta>1$, where $m(\beta)$ is the spontaneous magnetization. According to Eq. (16), for all $\beta>0$ the weak limit of $P_{n,\beta}\{S_n/n\in\mathrm{d}x\}$ is concentrated on $\mathcal{E}_\beta$. While in the case of the general model treated in the present section one should not expect such a precise formulation, the next theorem gives considerable information, relating weak limits of subsequences of $P_{n,\beta}\{Y_n\in\mathrm{d}x\}$ to the set of equilibrium macrostates $\mathcal{E}_\beta$. For example, if one knows that $\mathcal{E}_\beta$ consists of a unique point $\tilde{x}$, then it follows that the entire sequence $\{P_{n,\beta}\{Y_n\in\mathrm{d}x\},n\in\mathbb{N}\}$ converges weakly to $\delta_{\tilde{x}}$. This situation corresponds to the absence of a phase transition. The proof of the theorem is technical and is omitted.

Theorem 9. *We fix $\beta\in\mathbb{R}$ and use the notation of Theorem 8. If $\mathcal{E}_\beta$ consists of a unique point $\tilde{x}$, then $P_{n,\beta}\{Y_n\in\mathrm{d}x\}\Rightarrow\delta_{\tilde{x}}$. If $\mathcal{E}_\beta$ does not consist of a unique point, then any subsequence of $\{P_{n,\beta}\{Y_n\in\mathrm{d}x\},n\in\mathbb{N}\}$ has a subsubsequence converging weakly to a probability measure Π_β on $\mathcal{X}$ that is concentrated on $\mathcal{E}_\beta$; i.e., $\Pi_\beta\{(\mathcal{E}_\beta)^c\}=0$.*

In the next section we apply the general large deviation procedure just presented to the analysis of models of two-dimensional turbulence.

9. Maximum entropy principles in two-dimensional turbulence

This section presents an overview of recent research, in which Gibbs states are used to predict the large-scale, long-lived order of coherent vortices that persist amid the turbulent fluctuations of the vorticity field in two dimensions [4]. This is done by applying a statistical equilibrium theory of the two-dimensional Euler equations, which govern the motion of an inviscid, incompressible fluid. As shown in [23,24], these equations are reducible to the vorticity transport equations

$$\frac{\partial \omega}{\partial t} + \frac{\partial \omega}{\partial x_1}\frac{\partial \psi}{\partial x_2} - \frac{\partial \omega}{\partial x_2}\frac{\partial \psi}{\partial x_1} = 0 \quad \text{and} \quad -\Delta \psi = \omega, \tag{21}$$

in which ω is the vorticity, ψ is the stream function, and $\Delta = \partial^2/\partial x_1^2 + \partial/\partial x_2^2$ denotes the Laplacian operator on $\mathbb{R}^2$. The two-dimensionality of the flow means that these quantities are related to the velocity field $v = (v_1, v_2, 0)$ according to $(0, 0, \omega) = \text{curl } v$ and $v = \text{curl}(0, 0, \psi)$. All of these fields depend upon the time variable $t \in [0, \infty)$ and the space variable $x = (x_1, x_2)$, which runs through a bounded domain in $\mathbb{R}^2$. Throughout this section we assume that this domain equals the unit torus $T^2 = [0, 1) \times [0, 1)$, and we impose doubly periodic boundary conditions on all the flow quantities.

The governing Eq. (21) can also be expressed as a single equation for the scalar vorticity field $\omega = \omega(x, t)$. The periodicity of the velocity field implies that $\int_{T^2} \omega \, dx = 0$. With this restriction on its domain, the Green's operator $G = (-\Delta)^{-1}$ taking ω into ψ with $\int_{T^2} \psi \, dx = 0$ is well-defined. More explicitly, G is the integral operator

$$\psi(x) = G\omega(x) = \int_{T^2} g(x - x')\omega(x') \, dx', \tag{22}$$

where g is the Green's function defined by the Fourier series

$$g(x - x') \doteq \sum_{0 \neq z \in \mathbb{Z}^2} |2\pi z|^{-2} e^{2\pi i \langle z, (x-x') \rangle}. \tag{23}$$

Consequently, Eq. (21) can be considered as an equation in ω alone.

Even though the initial value problem for Eq. (21) is known to be well-posed for weak solutions whenever the initial data $\omega^0 = \omega(\cdot, 0)$ belongs to $L^\infty(\mathcal{X})$ [24], it is well known that this deterministic evolution does not provide a useful description of the system over long time intervals. When one seeks to quantify the long-time behavior of solutions, therefore, one is compelled to shift from the microscopic, or fine-grained, description inherent in ω to some kind of macroscopic, or coarse-grained, description. We will make this shift by adopting the perspective of equilibrium statistical mechanics. That is, one views the underlying deterministic dynamics as a means of randomizing the microstate ω subject to the conditioning inherent in the conserved quantities for the governing Eq. (21), and one takes the appropriate macrostates to be the canonical Gibbs measures $P(d\omega)$ built from these conserved quantities. In doing so, of course, one accepts an ergodic hypothesis that equates the time averages with canonical ensemble averages. Given this hypothesis, one hopes that these macrostates capture the long-lived, large-scale, coherent vortex structures that persist amid the small-scale vorticity fluctuations. The characterization of these self-organized macrostates, which are observed in simulations and physical experiments, is the ultimate goal of the theory.

The models that we will consider build on earlier and simpler theories, the first of which was due to Onsager [25]. Studying point vortices, he predicted that the equilibrium states with high enough energy have a negative temperature and represent large-scale, coherent vortices. This model was further developed in the 1970's, notably by Montgomery and Joyce [26]. However, the point vortex model fails to incorporate all the conserved quantities for two-dimensional ideal flow.

These conserved quantities are the energy, or Hamiltonian functional, and the family of generalized enstrophies, or Casimir functionals [24]. Expressed as a functional of ω, the kinetic energy is

$$H(\omega) \doteq \frac{1}{2} \int_{T^2 \times T^2} g(x - x') \omega(x) \omega(x') \, dx \, dx'. \tag{24}$$

The so-called generalized enstrophies are the global vorticity integrals

$$A(\omega) \doteq \int_{T^2} a(\omega(x)) \, dx, \tag{25}$$

where a is an arbitrary continuous real function on the range of the vorticity. In terms of these conserved quantities, the canonical ensemble is defined by the formal Gibbs measure

$$P_{\beta,a}(d\omega) = Z(\beta, a)^{-1} \exp[-\beta H(\omega) - A(\omega)] \, \Pi(d\omega), \tag{26}$$

where $Z(\beta, a)$ is the associated partition function and $\Pi(d\omega)$ denotes some invariant product measure on some phase space of all admissible vorticity fields ω. Of course, this formal construction is not meaningful as it stands due to the infinite dimensionality of such a phase space. We therefore proceed to define a sequence of lattice models on T^2 in order to give a meaning to this formal construction.

One lattice model that respects conservation of energy and also the generalized enstrophy constraints was developed by Miller et. al. [27,28] and Robert et. al. [29,30]; we will refer to it as the Miller–Robert model. A related model, which discretizes the continuum dynamics in a different way, was developed by Turkington [31]. These authors use formal arguments to derive maximum entropy principles that are argued to be equivalent to variational formulas for the equilibrium macrostates. In terms of these macrostates, coherent vortices of two-dimensional turbulence can be studied. The purpose of this section is to outline how large deviation theory can be applied to give a rigorous derivation of these variational formulas. References [4] and [31] discuss in detail the physical background.

The variational formulas will be derived for the following lattice model that includes both the Miller–Robert model and the Turkington model as special cases. Let T^2 denote the unit torus $[0, 1) \times [0, 1)$ with periodic boundary conditions and let $\mathcal{L}$ be a uniform lattice of $n \doteq 2^{2m}$ sites s in T^2, where m is a positive integer. The intersite spacing in each coordinate direction is 2^{-m}. We make this particular choice of n to ensure that the lattices are refined dyadically as m increases, a property that is needed later when we study the continuum limit obtained by sending $n \to \infty$ along the sequence $n = 2^{2m}$. In correspondence with this lattice we have a dyadic partition of T^2 into n squares called microcells, each having area $1/n$. For each $s \in \mathcal{L}$ we denote by $M(s)$ the unique microcell having the site s in its lower left corner. Although $\mathcal{L}$ and $M(s)$ depend on n, this is not indicated in the notation.

The configuration spaces for the lattice model are the product spaces $\Omega_n = \mathcal{Y}^n$, where $\mathcal{Y}$ is a compact set in $\mathcal{R}$. Configurations in Ω_n are denoted by $\zeta = \{\zeta(s), s \in \mathcal{L}\}$, which represents the discretized vorticity field. Let ρ be a probability measure on $\mathcal{Y}$ and let P_n denote the product measure on Ω_n with one-dimensional marginals ρ. As discussed in [4], the Miller–Robert model and the Turkington model differ in their choices of the compact set $\mathcal{Y}$ and the probability measure ρ.

For $\zeta \in \Omega_n$ the Hamiltonian for the lattice model is defined by

$$H_n(\zeta) \doteq \frac{1}{2n^2} \sum_{s,s' \in \mathcal{L}} g_n(s - s') \zeta(s) \zeta(s'), \tag{27}$$

where g_n is the lattice Green's function defined by the finite Fourier sum

$$g_n(s - s') \doteq \sum_{0 \neq z \in \mathcal{L}^*} |2\pi z|^{-2} e^{2\pi i \langle z, s - s' \rangle} \tag{28}$$

over the finite set $\mathcal{L}^* \doteq \{z = (z_1, z_2) \in \mathbb{Z}^2 : -2^{m-1} < z_1, z_2 \leq 2^{m-1}\}$. Let a be any continuous function mapping $\mathcal{Y}$ into $\mathcal{R}$. For $\zeta \in \Omega_n$ we also define functions known as the generalized enstrophies by

$$A_{n,a}(\zeta) \doteq \frac{1}{n} \sum_{s \in \mathcal{L}} a(\zeta(s)). \tag{29}$$

In terms of these quantities we define the partition function

$$Z_n(\beta, a) \doteq \int_{\Omega_n} \exp[-\beta H_n(\zeta) - A_{n,a}(\zeta)] \, P_n(\mathrm{d}\zeta) \tag{30}$$

and the Gibbs state $P_{n,\beta,a}$, which is the probability measure that assigns to a Borel subset B of Ω_n the probability

$$P_{n,\beta,a}\{B\} \doteq \frac{1}{Z_n(\beta, a)} \int_B \exp[-\beta H_n(\zeta) - A_{n,a}(\zeta)] \, P_n(\mathrm{d}\zeta). \tag{31}$$

These probability measures are parametrized by the constant $\beta \in \mathbb{R}$ and the function $a \in \mathcal{C}(\mathcal{Y})$. The dependence of Gibbs measures on the inverse temperature β is standard, while their dependence on the function a that determines the enstrophy functional is a novelty of this particular statistical equilibrium problem. The Miller–Robert model and the Turkington model also differ in their choices of the parameter β and the function a.

The main theorem in this section applies the theory of large deviations to derive the continuum limit $n \to \infty$ of the lattice model just introduced. Because the interactions $g_n(s - s')$ in the lattice model are long-range, one must replace β and a by $n\beta$ and na in order to obtain a nontrivial continuum limit [4,27,28]. Replacing β and a by $n\beta$ and na in the formulas for the partition function and the Gibbs state is equivalent to replacing H_n and A_n by nH_n and nA_n and leaving β and a unscaled. We carry out the large deviation analysis of the lattice model by applying the general procedure specified in the preceding section, making the straightforward modifications necessary to handle both the Hamiltonian and the generalized enstrophy. Thus, we seek a hidden space, a hidden process $\{Y_n\}$, representation functions $\tilde{H}$ and $\tilde{A}_a$ for the Hamiltonian and for the generalized enstrophy, and a large deviation principle for $\{Y_n\}$ with respect to $\{P_n\}$. Because of the replacement of H_n and A_n by nH_n and nA_n, the defining properties of the representation functions that we now present differ by a factor of n from what appears in the preceding section. The first marginal of a probability measure μ on $T^2 \times \mathcal{Y}$ is defined to be the probability measure $\mu_1\{A\} \doteq \mu\{A \times \mathcal{Y}\}$ for Borel subsets A of T^2.

- *Hidden space.* This is the space $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ of probability measures on $T^2 \times \mathcal{Y}$ with first marginal θ, where $\theta(\mathrm{d}x) = \mathrm{d}x$ is Lebesgue measure on T^2.
- *Hidden process.* For each $n \in \mathbb{N}$, $Y_n : \Omega_n \mapsto \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ is defined by

$$Y_n(\mathrm{d}x \times \mathrm{d}y) = Y_n(\zeta, \mathrm{d}x \times \mathrm{d}y) \doteq \mathrm{d}x \otimes \sum_{s \in \mathcal{L}} 1_{M(s)}(x) \delta_{\zeta(s)}(\mathrm{d}y).$$

Thus for Borel subsets A of $T^2 \times \mathcal{Y}$

$$Y_n\{A\} \doteq \sum_{s \in \mathcal{L}} \int_A 1_{M(s)}(x) \, \mathrm{d}x \, \delta_{\zeta(s)}(\mathrm{d}y).$$

Since $\sum_{s \in \mathcal{L}} 1_{M(s)}(x) = 1$ for all $x \in T^2$, the first marginal of Y_n equals $\mathrm{d}x$.

- *Hamiltonian representation function.* $\tilde{H} : \mathcal{P}_\theta(T^2 \times \mathcal{Y}) \mapsto \mathcal{R}$ is defined by

$$\tilde{H}(\mu) \doteq \frac{1}{2} \int_{(T^2 \times \mathcal{Y})^2} g(x - x') y y' \, \mu(\mathrm{d}x \times \mathrm{d}y) \mu(\mathrm{d}x' \times \mathrm{d}y'),$$

where $g(x - x')$ is defined by the Fourier series $\sum_{0 \neq z \in \mathcal{Z}^2} |2\pi z|^{-2} e^{2\pi i \langle z, x-x' \rangle}$. As proved in [4], $\tilde{H}$ is bounded and continuous and there exists $C < \infty$ such that

$$\sup_{\zeta \in \Omega_n} |H_n(\zeta) - \tilde{H}(Y_n(\zeta, \cdot))| \leq C \left(\frac{\log n}{n} \right)^{1/2} \quad \text{for all } n \in \mathbb{N}. \tag{32}$$

- *Generalized enstrophy representation function.* $\tilde{A}_a : \mathcal{P}_\theta(T^2 \times \mathcal{Y}) \mapsto \mathcal{R}$ is defined by

$$\tilde{A}_a(\mu) \doteq \int_{T^2 \times \mathcal{Y}} a(y) \, \mu(\mathrm{d}x \times \mathrm{d}y).$$

$\tilde{A}_a$ is bounded and continuous and

$$A_{n,a}(\zeta) = \tilde{A}_a(Y_n(\zeta, \cdot)) \quad \text{for all } \zeta \in \Omega_n. \tag{33}$$

- *Large deviation principle for the hidden process.* With respect to the product measures $\{P_n\}$, $\{Y_n\}$ satisfies the large deviation principle on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ with rate function the relative entropy

$$I_{\theta \times \rho}(\mu) \doteq \begin{cases} \int_{T^2 \times \mathcal{Y}} \left(\log \frac{\mathrm{d}\mu}{\mathrm{d}(\theta \times \rho)} \right) \mathrm{d}\mu & \text{if } \mu \ll \theta \times \rho \\ \infty & \text{otherwise.} \end{cases}$$

We first comment on the last item. The large deviation principle for the hidden process with respect to $\{P_n\}$ is far from obvious and in fact is one of the main contributions of [4]. We will address this issue after specifying the large deviation behavior of the model in Theorem 10. Concerning Eq. (32), since $\theta\{M(s)\} = 1/n$, it is plausible that

$$\tilde{H}(Y_n(\zeta, \cdot)) = \frac{1}{2} \sum_{s,s' \in \mathcal{L}} \int_{M(s) \times M(s')} g(x - x') \, \mathrm{d}x \, \mathrm{d}x' \, \zeta(s)\zeta(s')$$

is a good approximation to $H_n(\zeta) \doteq [1/(2n^2)] \sum_{s,s' \in \mathcal{L}} g_n(s - s') \zeta(s)\zeta(s')$. Concerning Eq. (33), for $\zeta \in \Omega_n$

$$\tilde{A}_a(Y_n(\zeta, \cdot)) = \int_{T^2 \times \mathcal{Y}} a(y) \, Y_n(\zeta, \mathrm{d}x \times \mathrm{d}y) = \frac{1}{n} \sum_{s \in \mathcal{L}} a(\zeta(s)) = A_{n,a}(\zeta).$$

The proofs of the boundedness and continuity of $\tilde{A}_a$ are straightforward.

Part 1 of Theorem 10 gives the asymptotic behavior of the scaled partition functions $Z_n(n\beta, na)$, and part 2 states the large deviation principle for the hidden process $\{Y_n\}$ with respect to the scaled Gibbs states $P_{n,n\beta,na}$. The rate function has the familiar form $I_{\beta,a} \doteq I_{\rho \times \theta} + \beta \tilde{H} + \tilde{A} -$ const; the relative entropy $I_{\rho \times \theta}$ arises from the large deviation principle for $\{Y_n\}$ with respect to $\{P_n\}$, and the other terms arise from Eqs. (32) and (33) and the form of $P_{n,n\beta,na}$. Part 3 of the theorem gives properties of the set $\mathcal{E}_{\beta,a}$ of equilibrium macrostates. $\mathcal{E}_{\beta,a}$ consists of measures μ at which the rate function $I_{\beta,a}$ in part 2 attains its infimum of 0 over $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$. The proof of the theorem is omitted since it is similar to the proof of Theorem 8. We also omit the analogue of Theorem 9 concerning the relationship between weak limits of the $P_{n,n\beta,na}$-distributions of Y_n and $\mathcal{E}_{\beta,a}$.

Theorem 10. *For each $\beta \in \mathcal{R}$ and $a \in \mathcal{C}(\mathcal{Y})$ the following conclusions hold.*

1. $\varphi(\beta, a) \doteq \lim_{n \to \infty} (1/n) \log Z_n(n\beta, na)$ *exists and is given by the variational formula*

$$\varphi(\beta, a) = - \inf_{\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})} \{ \beta \tilde{H}(\mu) + \tilde{A}_a(\mu) + I_{\rho \times \theta}(\mu) \}.$$

2. *With respect to the scaled Gibbs states* $\{P_{n,n\beta,na}\}$, $\{Y_n\}$ *satisfies the large deviation principle on* $\mathcal{X}$ *with scaling constants* $\{n\}$ *and rate function*

$$I_{\beta,a}(\mu) \doteq I_{\rho \times \theta}(\mu) + \beta \tilde{H}(\mu) + \tilde{A}_a(\mu) - \inf_{\nu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})} \{I_{\rho \times \theta}(\nu) + \beta \tilde{H}(\nu) + \tilde{A}_a(\nu)\}.$$

3. *We define the set of equilibrium macrostates*

$$\mathcal{E}_{\beta,a} \doteq \{\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y}) : I_{\beta,a}(\mu) = 0\}.$$

Then $\mathcal{E}_{\beta,a}$ *is a nonempty, compact subset of* $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$. *In addition, if A is a Borel subset of* $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ *such that* $\bar{A} \cap \mathcal{E}_{\beta,a} = \emptyset$, *then* $I_{\beta,a}(\bar{A}) > 0$ *and for all sufficiently large n*

$$P_{n,n\beta,na}\{Y_n \in A\} \le \exp\left[-\frac{nI_{\beta,a}(\bar{A})}{2}\right] \to 0 \quad as \ n \to \infty.$$

The paper [4] and a sequel currently in preparation discuss the physical implications of the theorem and the relationship between the following concepts in the context of the Miller–Robert model and the Turkington model: $\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ is an equilibrium macrostate (i.e., $\mu \in \mathcal{E}_{\beta,a}$) and μ satisfies a corresponding maximum entropy principle. In the Miller–Robert model, the maximum entropy principle takes the form of minimizing the relative entropy $I_{\theta \times \rho}(\mu)$ over $\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ subject to the constraints

$$\tilde{H}(\mu) = H(\omega^0) \quad \text{and} \quad \int_{T^2} \mu(\mathrm{d}x \times \cdot) = \int_{T^2} \delta_{\omega^0(x)}(\cdot)\,\mathrm{d}x,$$

where ω^0 is an initial vorticity field and $H(\omega^0)$ is defined in Eq. (24). In the Turkington model, the maximum entropy principle takes a somewhat related form in which the second constraint appearing in the Miller–Robert maximum entropy principle is relaxed to a family of convex inequalities parametrized by points in $\mathcal{Y}$. Understanding for each model the relationship between equilibrium macrostates μ and the corresponding maximum entropy principle allows one to identify a steady vortex flow with a given equilibrium macrostate μ. Through this identification, which is described in [4], one demonstrates how the equilibrium macrostates capture the long-lived, large-scale, coherent structures that persist amid the small-scale vorticity fluctuations.

We spend the rest of this section outlining how the large deviation principle is proved for the hidden process

$$Y_n(\mathrm{d}x \times \mathrm{d}y) \doteq \mathrm{d}x \otimes \sum_{s \in \mathcal{L}} 1_{M(s)}(x)\,\delta_{\zeta(s)}(\mathrm{d}y)$$

with respect to the product measures $\{P_n\}$. The proof is based on the innovative technique of approximating Y_n by a doubly indexed sequence of random measures $\{W_{n,r}\}$ for which the large deviation principle is, at least formally, almost obvious. This doubly indexed sequence, obtained from $\{Y_n\}$ by averaging over an intermediate scale, clarifies the physical basis of the large deviation principle and reflects the multiscale nature of turbulence. A similar large deviation principle is derived in [32,33] by an abstract approach that relies on a convex analysis argument. That approach obscures the role of spatial coarse-graining in the large deviation behavior.

In order to define $W_{n,r}$, we recall that $\mathcal{L}$ contains $n = 2^{2m}$ sites s. For even $r < 2m$ we consider a regular dyadic partition of T^2 into 2^r macrocells $\{D_{r,k}, k = 1, 2, \ldots, 2^r\}$. Each macrocell contains $n/2^r$ lattice sites and is the union of $n/2^r$ microcells $M(s)$, where $M(s)$ contains the site s in its lower left corner. We now define

$$W_{n,r}(\mathrm{d}x \times \mathrm{d}y) = W_{n,r}(\zeta, \mathrm{d}x \times \mathrm{d}y) \doteq \mathrm{d}x \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\,\frac{1}{n/2^r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(\mathrm{d}y).$$

$W_{n,r}$ is obtained from Y_n by replacing, for each $s \in D_{r,k}$, the point mass $\delta_{\zeta(s)}$ by the average $(n/2^r)^{-1}\sum_{s \in D_{r,k}}\delta_{\zeta(s)}$ over the $n/2^r$ sites contained in $D_{r,k}$.

We need the key fact that with respect to a suitable metric d on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ $d(Y_n, W_{n,r}) \leq \sqrt{2}/2^{r/2}$ for all $n = 2^{2m}$ and all even $r \in \mathbb{N}$ satisfying $r < 2m$. The proof of this approximation property uses the fact that the diameter of each macrocell $D_{r,k}$ equals $\sqrt{2}/2^{r/2}$ [4]. The next theorem states the two-parameter large deviation principle for $\{W_{n,r}\}$ with respect to the product measures $\{P_n\}$. By means of the approximation property, it is then straightforward to show that with respect to $\{P_n\}$, $\{Y_n\}$ satisfies the large deviation principle with the same rate function $I_{\theta \times \rho}$.

Theorem 11. *With respect to the product measures $\{P_n\}$, the sequence $\{W_{n,r}\}$ satisfies the following two-parameter large deviation principle on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ with rate function $I_{\theta \times \rho}$: for any closed subset F of $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$*

$$\limsup_{r \to \infty} \limsup_{n \to \infty} \frac{1}{n}\log P_n\{W_{n,r} \in F\} \leq -I_{\theta \times \rho}(F)$$

and for any open subset G of $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$

$$\liminf_{r \to \infty} \liminf_{n \to \infty} \frac{1}{n}\log P_n\{W_{n,r} \in G\} \geq -I_{\theta \times \rho}(G).$$

Our purpose in introducing the doubly indexed process $W_{n,r}$ is the following. The local averaging over the sets $D_{r,k}$ introduces a spatial scale that is intermediate between the macroscopic scale of the torus T^2 and the microscopic scale of the microcells $M(s)$. As a result, $W_{n,r}$ can be written in the form

$$W_{n,r}(\mathrm{d}x \times \mathrm{d}y) = \mathrm{d}x \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, L_{n,r,k}(\mathrm{d}y), \tag{34}$$

where

$$L_{n,r,k}(\mathrm{d}y) = L_{n,r,k}(\zeta, \mathrm{d}y) \doteq \frac{1}{n/2^r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(\mathrm{d}y).$$

Since each $D_{r,k}$ contains $n/2^r$ lattice sites s, with respect to $\{P_n\}$ the sequence $\{L_{n,r,k}, k = 1, \ldots, 2^r\}$ is a family of i.i.d. empirical measures. For each r and each $k \in \{1, \ldots, 2^r\}$ Sanov's Theorem 7 implies that as $n \to \infty$ $\{L_{n,r,k}\}$ satisfies the large deviation principle on $\mathcal{P}(\mathcal{Y})$ with scaling constants $n/2^r$ and rate function I_ρ.

It is easy to motivate the large deviation principle for $\{W_{n,r}\}$ stated in Theorem 11. Suppose that $\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ has finite relative entropy with respect to $\theta \times \rho$ and has the special form

$$\mu(\mathrm{d}x \times \mathrm{d}y) = \mathrm{d}x \otimes \tau(x, \mathrm{d}y), \quad \text{where } \tau(x, \mathrm{d}y) \doteq \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \tau_k(\mathrm{d}y) \tag{35}$$

and $\tau_1, \ldots, \tau_{2^r}$ are probability measures on $\mathcal{Y}$. The representation (34), Sanov's Theorem, and the independence of $L_{n,r,1}, \ldots, L_{n,r,2^r}$ suggest that

$$\lim_{n \to \infty} \frac{1}{n}\log P_n\{W_{n,r} \sim \mu\} = \lim_{n \to \infty} \frac{1}{n}\log P_n\{L_{n,r,k} \sim \tau_k, k = 1, \ldots, 2^r\}$$

$$= \frac{1}{2^r} \sum_{k=1}^{2^r} \lim_{n \to \infty} \frac{1}{n/2^r}\log P_n\{L_{n,r,k} \sim \tau_k\}$$

$$\approx -\frac{1}{2^r}\sum_{k=1}^{2^r} I_\rho(\tau_k) = -\int_{T^2} I_\rho(\tau(x,\cdot))\,dx$$

$$= -\int_{T^2}\int_{\mathcal{Y}}\left(\log\frac{d\tau(x,\cdot)}{d\rho(\cdot)}(y)\right)\tau(x,dy)\,dx$$

$$= -\int_{T^2\times\mathcal{Y}}\left(\log\frac{d\mu}{d(\theta\times\rho)}(x,y)\right)\mu(dx\times dy) = -I_{\theta\times\rho}(\mu).$$

The two-parameter large deviation principle for $W_{n,r}$ with rate function $I_{\theta\times\rho}$ is certainly plausible, in view of the fact that any measure $\mu \in \mathcal{P}_\theta(T^2\times\mathcal{Y})$ can be well approximated, as $r\to\infty$, by a sequence of measures of the form Eq. (35) ([34], Lemma 3.2). The reader is referred to [4] for an outline of the proof of this two-parameter large deviation principle. The large deviation principle for $\{W_{n,r}\}$ is a special case of a large deviation principle proved in [34] for an extensive class of random measures which includes $\{W_{n,r}\}$ as a special case.

This completes our application of the theory of large deviations to models of two-dimensional turbulence. The asymptotic behavior of these models is stated in Theorem 10. One of the main components of the proof is the large deviation principle for the hidden process $\{Y_n\}$, which in turn follows by approximating the hidden process by the doubly indexed sequence $\{W_{n,r}\}$ and proving the large deviation principle for this sequence. This proof relies on Sanov's Theorem, which generalizes Boltzmann's 1877 calculation of the asymptotic behavior of multinomial probabilities. Earlier in the paper we used the elementary form of Sanov's Theorem stated in Theorem 2 to derive the form of the Gibbs state for the discrete ideal gas and to motivate the version of Cramér's Theorem needed to analyze the Curie–Weiss model. It is hoped that both the importance of Boltzmann's 1877 calculation and the applicability of the theory of large deviations to problems in statistical mechanics have been amply demonstrated in this paper. It is also hoped that the paper will inspire the reader to discover new applications.

Acknowledgements

References

[1] R.S. Ellis, Entropy, Large Deviations, and Statistical Mechanics, Springer, New York, 1985.

[2] O.E. Lanford, Entropy and equilibrium states in classical statistical mechanics, in: A. Lenard (Ed.), Statistical Mechanics and Mathematical Problems, Lecture Notes in Physics, vol. 20, Springer, Berlin, 1973, pp. 1–113.

[3] A.S. Wightman, Convexity and the notion of equilibrium state in thermodynamics and statistical mechanics, in: Introduction to R.B. Israel, Convexity in the Theory of Lattice Gases, Princeton University Press, Princeton, 1979.

[4] C. Boucher, R.S. Ellis, B. Turkington, Derivation of maximum entropy principles in two-dimensional turbulence via large deviations, J. Stat. Phys. (1999), submitted.

[5] L. Boltzmann, Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respecktive den Sätzen über das Wärmegleichgewicht (On the relationship between the second law of the mechanical theory of heat and the probability calculus), Wiener Berichte 2(76) (1877) 373–435.

[6] W.R. Everdell, The First Moderns, University of Chicago Press, Chicago, 1997.

[7] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, 2nd ed., Springer, New York, 1998.

[8] P. Ney, Private Communication, 1997.

[9] J.-D. Deuschel, D.W. Stroock, H. Zessin, Microcanonical distributions for lattice gases, Comm. Math. Phys. 139 (1991) 83–101.

[10] A. Meda, P. Ney, A conditioned law of large numbers for Markov additive chains, Studia Scientiarum Mathematicarum Hungarica (1998), in press.

[11] A. Sznitman, Topics in propagation of chaos, in: P.L. Hennequin (Ed.), École d'Été de Probabilités de Saint-Flour XIX–1989, Lecture Notes in Mathematics 1464, Springer, New York, 1991, pp. 165–251.

[12] G. Parisi, Statistical Field Theory, Addison-Wesley, Redwood City, CA, 1988.

[13] R.S. Ellis, An overview of the theory of large deviations and applications to statistical mechanics, Scand. Actuarial J. 1 (1995) 97–142.

[14] S.N. Ethier, T.G. Kurtz, Markov Processes: Characterization and Convergence, Wiley, New York, 1986.

[15] P. Dupuis, R.S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations, Wiley, New York, 1997.

[16] R.S. Ellis, C.M. Newman, Limit theorems for sums of dependent random variables occurring in statistical mechanics, Z. Wahrsch. verw. Geb. 44 (1978) 117–139.

[17] R.S. Ellis, C.M. Newman, The statistics of Curie–Weiss models, J. Stat. Phys. 19 (1978) 149–161.

[18] R.S. Ellis, C.M. Newman, J.S. Rosen, Limit theorems for sums of dependent random variables occurring in statistical mechanics II: conditioning, multiple phases, and metastability, Z. Wahrsch. verw. Geb. 51 (1980) 153–169.

[19] R.S. Ellis, K. Wang, Limit theorems for the empirical vector of the Curie–Weiss–Potts model, Stoch. Proc. Appl. 35 (1990) 59–79.

[20] R.S. Ellis, K. Wang, Limit theorems for maximum likelihood estimators in the Curie–Weiss–Potts model, Stoch. Proc. Appl. 40 (1992) 251–288.

[21] H. Föllmer, S. Orey, Large deviations for the empirical field of a Gibbs measure, Ann. Probab. 16 (1987) 961–977.

[22] S. Olla, Large deviations for Gibbs random fields, Probab. Th. Rel. Fields 77 (1988) 343–359.

[23] A.J. Chorin, Vorticity and Turbulence, Springer, Berlin, 1994.

[24] C. Marchioro, M. Pulvirenti, Mathematical Theory of Incompressible Nonviscous Fluids, Springer, New York, 1994.

[25] L. Onsager, Statistical hydrodynamics, Suppl. Nuovo Cim. 6 (1949) 279–287.

[26] D. Montgomery, G. Joyce, Statistical mechanics of negative temperature states, Phys. Fluids 17 (1974) 1139–1145.

[27] J. Miller, Statistical mechanics of Euler equations in two dimensions, Phys. Rev. Lett. 65 (1990) 2137–2140.

[28] J. Miller, P. Weichman, M.C. Cross, Statistical mechanics, Euler's equations, and Jupiter's red spot, Phys. Rev. A 45 (1992) 2328–2359.

[29] R. Robert, A maximum-entropy principle for two-dimensional perfect fluid dynamics, J. Stat. Phys. 65 (1991) 531–553.

[30] R. Robert, J. Sommeria, Statistical equilibrium states for two-dimensional flows, J. Fluid Mech. 229 (1991) 291–310.

[31] B. Turkington, Statistical equilibrium measures and coherent states in two-dimensional turbulence, Comm. Pure Appl. Math. (1999), in press.

[32] J. Michel, R. Robert, Large deviations for Young measures and statistical mechanics of infinite dimensional dynamical systems with conservation law, Comm. Math. Phys. 159 (1994) 195–215.

[33] R. Robert, Concentration et entropie pour les mesures d'Young, C. R. Acad. Sci. Paris Série I 309 (1989) 757–760.

[34] C. Boucher, R.S. Ellis, B. Turkington, Spatializing random measures: doubly indexed processes and the large deviation principle, Ann. Probab. 27 (1999) 297–324.