The Boltzmann-Sanov
Large Deviation Principle and
Applications to Statistical Mechanics

Richard S. Ellis [1]
rsellis@math.umass.edu

Shlomo Ta'asan [2]
shlomo@andrew.cmu.edu

[1] Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003

[2] Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213

October 21, 2014

Abstract

In this review paper we illustrate the basic ideas of the theory of large deviations in the context of a simple model of a discrete ideal gas studied by Ludwig Boltzmann. These ideas include microscopic and macroscopic levels of description, Cramér's Theorem on the large deviation behavior of the sample mean of independent, identically distributed (i.i.d.) random variables, and a theorem first proved by Boltzmann and then generalized by Sanov describing the large deviation behavior of the empirical vectors of i.i.d. random variables. The rate of decay in the Boltzmann-Sanov Theorem is described by the relative entropy. Boltzmann used the relationship between the empirical vector and the relative entropy to describe the equilibrium distribution of the discrete ideal gas. In the second half of this paper we show how Boltzmann's ideas guided us in carrying out the large deviation analysis of a droplet model involving dependencies and having significant applications to technologies using sprays and powders.

1

1 Introduction

This review paper is an introduction to the theory of large deviations, which studies the exponential decay of probabilities in random systems. The theory is perfectly suited for applications to statistical mechanics, where it gives insight into physical phenomena such as equilibrium distributions and phase transitions. We illustrate the theory in the context of a simple model of a discrete ideal gas studied by Ludwig Boltzmann in 1877 [3]. This model has no interactions and is defined in terms of independent, identically distributed (i.i.d.) random variables. We then show how Boltzmann's ideas can be used to study the asymptotic behavior of a much more complicated droplet model that is defined in terms of dependent random variables. The droplet model is the subject of our paper [16].

In his paper of 1877, revolutionary and far ahead of its time because it was based on what was then the radical assumption that matter is composed of atoms [21], Boltzmann discovered a statistical interpretation of entropy when he calculated the equilibrium distribution of energy levels in the discrete ideal gas model. This distribution is now known as the Maxwell–Boltzmann distribution [20, §4.2–4.3], [22, 23]. The discrete ideal gas is a simple model because, having no interactions, it is defined in terms of a sequence of i.i.d. random variables. Despite the simplicity of the model, Boltzmann's calculation — based on what was then the radical use of probabilistic methods to derive equilibrium properties of matter — is the seed that engendered not only statistical mechanics, but also the theory of large deviations, two luxurious gardens that have continued to cross-fertilize each other since the time of Boltzmann.

Depressed and demoralized by the harsh criticism of his radical innovations in physics, Boltzmann paid the ultimate price by committing suicide in 1906. Ironically, during the previous year Einstein vindicated Boltzmann's belief in the atomic composition of matter when he proved the existence of atoms by using the Maxwell–Boltzmann distribution to explain Brownian motion. Einstein's work also provided the theoretical basis for the calculation of Avogadro's number three years later. Boltzmann died a broken man. However, future generations have been the beneficiaries of his scintillating genius.

In this paper we will explain Boltzmann's calculation of the Maxwell-Boltzmann distribution from the viewpoint of modern probability theory, emphasizing his derivation of a local version of Sanov's Theorem. This basic result is stated in Theorem 3.1. It gives a local large

deviation estimate, in terms of the relative entropy, of the probability that a sequence of empirical measures of i.i.d. random variables equals a sequence of possible asymptotic distributions. The fact that this probability is expressed as a multinomial coefficient reveals the essentially combinatorial nature of Boltzmann's proof. Starting with Boltzmann's work on the discrete ideal gas, combinatorial methods have remained an important tool in both statistical mechanics and in the theory of large deviations, offering insights into a wide variety of physical and mathematical phenomena via techniques that are elegant, powerful, and often elementary. In applications to statistical mechanics, this state of affairs is explained by the observation that "many fundamental questions . . . are inherently combinatorial, . . . including the Ising model, the Potts model, monomer-dimer systems, self-avoiding walks and percolation theory" [27]. For the two-dimensional Ising model and other exactly soluble models the classical references [2] and [24] are recommended.

A similar situation holds in the theory of large deviations. For example, section 2.1 of [8] discusses combinatorial techniques for finite alphabets and points out that because of the concreteness of these applications the large deviation principles are proved under much weaker conditions than the corresponding results in the general theory, into which the finite-alphabet results give considerable insight. The text [10] devotes several early sections to large deviation results for i.i.d. random variables having a finite state space and proved by combinatorial methods, including a sophisticated, level-3 result for the empirical pair measure.

Boltzmann's work on the discrete ideal gas has given rise to what has become a standard paradigm in the many fields of science that use the methods of statistical mechanics. As he observed, the Maxwell-Boltzmann distribution, in particular, and the equilibrium distributions of a wide variety of physical systems, in general, are each the global minimum point of an entropy function over all distributions satisfying the constraint defining the system, which in the case of an isolated system such as the discrete ideal gas is the energy. The Maxwell-Boltzmann distribution has the form of a normalized exponential parametrized by a quantity conjugate to the energy, which is the inverse temperature. It is described in Theorem 4.1. Our derivation of the Maxwell-Boltzmann distribution is much more precise than the derivations found in the physics literature, examples of which appear in [20, §4.3] and [26, §3.2].

These ideas have become the bedrock of classical statistical mechanics. Their beauty and their power can be seen when they are applied to new systems, for which they can provide a systematic procedure for deducing their equilibrium behavior. In our paper [16] we generalize the procedure developed by Boltzmann for the discrete ideal gas to carry out the asymptotic analysis of a droplet model involving dependencies. The question that motivated our research is natural and is simply stated. Given $b \in \mathbb{N}$ and $c > b$, K particles are placed, each with equal probability $1/N$, onto the N sites of a lattice. Assuming that $K/N = c$ and that each site is occupied by a minimum of b particles, we determine the equilibrium distribution, as $N \to \infty$, of the number of particles per site. As we discuss in subsection 6c, this equilibrium distribution

3

is a Poisson distribution $\rho_{b,\alpha_b(c)}$ restricted to $\mathbb{N}_b = \{n \in \mathbb{Z} : n \geq b\}$, where the parameter $\alpha_b(c)$ is chosen so that the mean of $\rho_{b,\alpha_b(c)}$ equals c.

Our description of the equilibrium distribution is one of the main results in [16]. As we describe in section 8, it is derived as a consequence of a large deviation principle (LDP) for a sequence of random probability measures, called number-density measures, which are the empirical measures of dependent random variables that count the droplet sizes in the model. Our proof of the LDP is self-contained and starts from first principles, using techniques that are familiar in statistical mechanics. The centerpiece of the proof is the local large deviation estimate in Theorem 3.1 of [16], the proof of which uses combinatorics, Stirlings formula, and Laplace asymptotics.

The droplet model has significant applications to technologies using sprays and powders. Sprays are liquids broken up into minute droplets and blown, ejected into, or falling through the air [7]; powders are dry, bulk solids composed of a large number of very fine particles that may flow freely when shaken or tilted [6]. Sprays and powders are prevalent throughout our society, in science, technology, industry, consumer products, and many other areas. An important problem for which a rigorous theory seems to be lacking is to determine the equilibrium size-distribution in sprays and powders [25, 29]. In the present paper we show, despite the difference in complexity between the discrete ideal gas and the droplet model, how Boltzmann's ideas provide a road map for the asymptotic analysis of the droplet model and explain how this analysis yields the equilibrium behavior of the dependent random variables that count the droplet sizes. The discovery of this equilibrium distribution fills a major gap in the literature on sprays and powders.

We will also explain the following four deep connections between the asymptotic analysis of the droplet model and that of Boltzmann's discrete ideal gas. Other connections are pointed out at the end of subsection 6b.

1. As Boltzmann did for the discrete ideal gas, we first prove a local large deviation estimate for a sequence of random, number-density measures that are the empirical measures of dependent random variables that count the droplet sizes in the model. This is done by a combinatorial analysis involving a product of two multinomial coefficients, of which the first is closely related to the multinomial coefficient that underlies Boltzmann's derivation of the Maxwell-Boltzmann distribution.

2. Boltzmann's local large deviation estimate for the discrete ideal gas can be lifted to a global result via a two-step procedure. Known as the Boltzmann-Sanov large deviation principle (LDP), it is a basic theorem in the modern theory. In the same way, the local large deviation estimate mentioned in item 1 for the droplet model can also be lifted to a global LDP via an analogous two-step procedure.

3. In analogy with Boltzmann's characterization of the Maxwell-Boltzmann distribution, the LDP for the droplet model mentioned in item 2 allows us to characterize its equilibrium distribution as the global minimum point of a suitable relative entropy function. As we already pointed out, this equilibrium distribution is a Poisson distribution describing the droplet sizes.

4. Because they involve independent random variables, the Boltzmann-Sanov LDP and related LDPs cannot be applied as stated to deduce the asymptotic behavior of the number-density measures in the droplet model, which involve dependent random variables. Nevertheless, the Boltzmann-Sanov LDP can be applied to give a formal derivation of the LDP for these measures, which suggests a Poisson distribution as the equilibrium distribution and which agrees asymptotically with the LDP mentioned in item 2 as the parameter c defining the droplet model converges to ∞.

We comment on item 2 in this list. The first step in the two-step procedure for lifting the local large deviation estimate to a global LDP is to use the local large deviation estimate and an approximation argument to prove large deviation limits for open balls. The second step is to use the large deviation limits for open balls to prove the large deviation upper bound for closed sets and the large deviation lower bound for open sets. These two steps involve techniques that can be applied to many problems in large deviations. An important example involving the second step is the proof of Cramér's Theorem in Polish spaces, which, as shown in [8, §6.1], is based on an elegant subadditivity argument that gives large deviation limits for open, convex subsets. In the case of the droplet model the two steps in this procedure are given a general formulation in Theorems 4.2 and 4.3 in [15], a companion paper to [16] containing a number of routine proofs omitted from [16] as well as additional background material. It is hoped that the reader who is new to the theory of large deviations will benefit from the presentation of the procedure in the much more elementary setting of the discrete ideal gas in section 3 of the present paper.

In section 2 we introduce the discrete ideal gas, state Cramér's Theorem 2.1 on the large deviation behavior of the sample means of the i.i.d. random variables in terms of which the discrete ideal gas is defined, and state the Boltzmann-Sanov Theorem 2.3 on the large deviation behavior of the empirical measures of these i.i.d. random variables. In both cases the large deviation behavior is described by an LDP. In section 3 we present a procedure for proving the Boltzmann-Sanov Theorem that is based on Boltzmann's local large deviation estimate in Theorem 3.1. The Boltzmann-Sanov Theorem is then applied in section 4 to derive the Maxwell-Boltzmann equilibrium distribution of the discrete ideal gas. Section 5 introduces the droplet model. In section 6 we describe the LDP for a sequence of random, number-density measures arising in this model that are the analogs of the empirical measures in the discrete ideal gas model. Section 6 ends by explaining how the LDP for the number-density measures

yields the form of the equilibrium distribution for the model. In Section 7 we show how Boltzmann's work on the discrete ideal gas suggests a formal procedure that motivates the LDP for the number-density measures. The paper ends with section 8. Here we outline the procedure for proving the LDP for the number-density measures. This procedure generalizes the procedure described in section 3 for proving the Boltzmann-Sanov Theorem for the discrete ideal gas.

One of the main themes of this paper is the power of the Boltzmann-Sanov LDP, and the power of the large deviation results that it has inspired, for studying the asymptotic behavior of statistical mechanical systems, including the discrete ideal gas and the droplet model. There are numerous other systems for which these ideas have been used. They include the following three lattice spin models: the Curie-Weiss spin system, the Curie-Weiss-Potts model, and the mean-field Blume-Capel model, which is also known as the mean-field BEG model. As explained in the respective sections 6.6.1, 6.6.2, and 6.6.3 of [11], the large deviation analysis shows that each of these three models has a different phase transition structure. Details of the analysis for the three models are given in the respective references [10, §IV.4], [17, 5], and [14]. Section 9 of [12] outlines how large deviation theory, and in particular a spatialized form of the Boltzmann-Sanov LDP, can be applied to determine equilibrium structures in statistical models of two-dimensional turbulence. Details of this analysis are given in [4].

We end this introduction by returning to the theme with which we started. Boltzmann's work is put in historical context by William Everdell, who traces the development of the modern consciousness in 19th and 20th century thought [19]. Chapter 3 focuses on the mathematicians of Germany in the 1870s — 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 person 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).

2 Discrete ideal Gas and Boltzmann-Sanov LDP

Our first topic is a probabilistic model for the discrete ideal gas. We will then consider microscopic and macroscopic levels of description of the model, which are basic to the theory of large deviations.

2a Probabilistic model for discrete ideal gas

Let $r \geq 2$ be an integer and consider r real numbers $y_1 < y_2 < \ldots < y_r$. The discrete ideal gas consists of n particles, each of which has a random energy lying in the set $\Lambda = \{y_1, y_2, \ldots, y_r\}$. In the absence of any information on the distribution of the random energies, we assume that they are independent and uniformly distributed on Λ.

The standard probabilistic model corresponding to this physical description is defined in terms of the configuration space $\Omega_n = \Lambda^n$. The quantities $\omega \in \Omega_n$ are called microstates, and each has the form $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, where ω_i denotes the energy of the i'th particle. It is convenient to introduce random variables representing these energy values. This is done by defining $X_i(\omega) = \omega_i$ for $i \in \{1, 2, \ldots, n\}$ and $\omega \in \Omega_n$. Let ρ denote the probability vector in $\mathbb{R}^r$ defined by $\rho = (\rho_1, \rho_2, \ldots, \rho_r)$, where each $\rho_j = 1/r$. The same notation ρ is used to denote the probability measure $\rho = \sum_{j=1}^{r}(1/r)\delta_{y_j}$ on Λ. Since ρ assigns equal probability $1/r$ to each $y \in \Lambda$, we call ρ the uniform distribution on Λ. We also consider the probability measure P_n that assigns to each $\omega \in \Omega_n$ the probability $P_n\{\omega\} = 1/r^n$. Since

$$P_n\{\omega\} = \frac{1}{r^n} = \prod_{i=1}^{n} \frac{1}{r} = \prod_{i=1}^{n} \rho\{\omega_i\},$$

we identify P_n with the n-fold product measure ρ^n having equal one-dimensional marginals ρ. P_n is extended to a probability on the set of all subsets of Ω_n by defining for any subset B

$$P_n\{B\} = \sum_{\omega \in B} P_n\{\omega\} = \sum_{\omega \in B} \frac{1}{r^n} = \frac{1}{r^n} \cdot \mathrm{card}(B), \tag{2.1}$$

where $\mathrm{card}(B)$ denotes the cardinality of B.

All of the mathematical results stated in sections 2, 3, and 4 are valid for any probability measure $\rho = \sum_{j=1}^{r} \rho_j \delta_{y_j}$ on Λ for which each $\rho_j > 0$. We focus on the uniform distribution ρ because this distribution is the appropriate choice in the absence of information on the distribution of the energy values in the discrete ideal gas.

It follows from the definition of P_n that for any $i \in \{1, 2, \ldots, n\}$ and any $y \in \Lambda$

$$P_n\{\omega \in \Omega_n : X_i(\omega) = y\} = P_n\{X_i = y\} = \frac{1}{r}$$

and that for any subsets $B_1, B_2, \ldots, B_n$ of Λ

$$P_n\{X_i \in B_i \text{ for } i = 1, 2, \ldots, n\} = \prod_{i=1}^{n} P_n\{X_i \in B_i\}.$$

We conclude that the random variables X_i are independent and uniformly distributed on Λ.

We now turn to the large deviation analysis of the discrete ideal gas, which we introduce by focusing on the levels of description of this model. There is a microscopic level and a sequence of macroscopic levels, of which we consider the first two. Statistical mechanics and the theory of large deviations focus on the interplay among these levels of description. The first macroscopic level is defined in terms of the sample mean, and the second is defined in terms of the empirical vector. At the end of the discussion of the sample mean, we introduce the microcanonical ensemble by conditioning the product measure P_n on a constraint that involves the sample mean and models the conservation of energy. In section 4 we will use the empirical vector to determine the Maxwell-Boltzmann distribution, which is the equilibrium distribtion of the random energy values X_i of the discrete ideal gas with respect to the microcanonical ensemble.

2b   Microscopic level

The configuration space Ω_n contains r^n microstates. Each microstate ω gives a microscopic description of the discrete ideal gas by specifying the energy values of all n particles. Although the microscopic level of description is precise, it is much too detailed to give useful information.

2c   Macroscopic level 1: sample mean

The most basic macroscopic level involves the sample mean. Given $\omega \in \Lambda^n$, define

$$S_n(\omega) = \sum_{i=1}^{n} X_i(\omega) = \sum_{i=1}^{n} \omega_i.$$

The **sample mean** S_n/n, or the average random energy, takes values in the closed interval $[y_1, y_r]$. This macroscopic variable summarizes, in terms of a single quantity, the r^n degrees of freedom in the microscopic description $\omega \in \Omega_n$. We also define the average energy

$$\bar{y} = \frac{1}{r} \cdot \sum_{k=1}^{r} y_k.$$

This quantity equals the mean of X_1 with respect to P_n, which is expressed by the expectation $E^{P_n}\{X_1\}$.

We next give a formal overview of the asymptotic behavior of S_n/n, which will soon be reformulated in a rigorous way. The **weak law of large numbers** states that

$$\lim_{n \to \infty} P_n\{S_n/n \sim \bar{y}\} = 1.$$

Cramér's LDP, stated in Theorem 2.1, expresses the large deviation behavior of S_n/n. It shows that there exists a function I mapping $[y_1, y_r]$ into $[0, \infty)$ with the property that for any $z \in (y_1, y_r)$

$$P_n\{S_n/n \sim z\} \approx \exp[-nI(z)] \text{ as } n \to \infty.$$

We now combine the law of large numbers with the large deviation estimate in the last equation. If $z = \bar{y}$, then $P_n\{S_n/n \sim \bar{y}\} \to 1$, and if $z \neq \bar{y}$, then $P_n\{S_n/n \sim z\} \to 0$. Thus if $z = \bar{y}$, then we expect that $I(z) = 0$. On the other hand, if $z \neq \bar{y}$, then we expect that $I(z) > 0$, implying that

$$P_n\{S_n/n \sim z\} \approx \exp[-nI(z)] \to 0 \text{ exponentially fast as } n \to \infty. \tag{2.2}$$

If $z \neq \bar{y}$, then we call the event $\{S_n/n \sim z\}$ a large deviation event. The function I is called a **rate function** or an **entropy function**.

In order to understand the statistical mechanical implication of (2.2), we use (2.1) to write

$$P_n\{S_n/n \sim z\} = \frac{1}{r^n} \cdot \text{card}\{\omega \in \Lambda^n : S_n(\omega)/n \sim z\}.$$

It follows that

$$\text{card}\{\omega \in \Lambda^n : S_n(\omega)/n \sim z\} \approx r^n \exp[-nI(z)] = \text{card}(\Omega_n) \cdot \exp[-nI(z)].$$

This formula yields the following appealing interpretation of the rate function. $I(z)$ records the multiplicity of microstates ω consistent with the macrostate z through the macroscopic variable S_n/n. This interpretation of the rate function is consistent with Boltzmann's insight of 1877 concerning the role of entropy in statistical mechanics. His insight is the following.

> Entropy is a bridge between a microscopic level, on which physical systems are defined in terms of the interactions among the individual constituent particles, and a macroscopic level, on which the laws describing the behavior of the system are formulated.

We now give a rigorous formulation of the asymptotic behavior of S_n/n. Given $z \in (y_1, y_r)$ and any $a > 0$ define the closed interval

$$F_{z,a} = \begin{cases} [z - a, z] & \text{if } y_1 < z < \bar{y} \\ [z - a, z + a] & \text{if } z = \bar{y} \\ [z, z + a] & \text{if } \bar{y} < z < y_r. \end{cases} \tag{2.3}$$

In all three cases $a > 0$ is chosen so small that the respective interval is a subset of (y_1, y_r). If $z = \bar{y}$, then $F_{z,a}$ is a symmetric interval with center $\bar{y}$, and the weak law of large numbers states that

$$\lim_{n \to \infty} P_n\{S_n/n \in F_{z,a}\} = 1.$$

If $z \neq \bar{y}$, then since $\bar{y}$ does not lie in the close interval $F_{z,a}$, the weak law of large numbers implies that

$$\lim_{n \to \infty} P_n\{S_n/n \in F_{z,a}\} = 0. \tag{2.4}$$

In (2.8) we show that the probability $P_n\{S_n/n \in F_{z,a}\}$ decays to 0 exponentially fast with an exponential rate depending on z. This gives a rigorous formulation of the heuristic limit (2.2).

For $t \in \mathbb{R}$ define $c(t) = \log E^{P_n}\{\exp(tX_1)\}$, which is the logarithm of the moment generating function of ρ, also known as the cumulant generating function of ρ. Since X_1 is distributed by ρ, we can write

$$c(t) = \log\left(\frac{1}{r}\sum_{k=1}^{r}\exp(ty_k)\right). \tag{2.5}$$

Cramér's Theorem describes the large deviation behavior of S_n/n in terms of a rate function I defined in (2.6) in terms of c. The statement of this theorem includes properties of I, the large deviation upper bound in part (a), and the large deviation lower bound in part (b). Cramér's Theorem is proved in [10, §VII.2, §VII.5] and [8, §2.2]. A general version is proved in [8, §6.1].

Theorem 2.1 (Cramér). *For $x \in \mathbb{R}$ we define*

$$I(x) = \sup_{t \in \mathbb{R}}\{tx - c(t)\}, \tag{2.6}$$

which expresses I as the Legendre-Fenchel transform of c. I is finite and continuous on $[y_1, y_r]$ and satisfies

$$I(x) > I(\bar{y}) = 0 \text{ for all } x \in [y_1, y_r] \text{ satisfying } x \neq \bar{y}. \tag{2.7}$$

Thus I attains its infimum of 0 over $[y_1, y_r]$ at the unique point $\bar{y}$. In addition, I is a strictly convex function on $[y_1, y_r]$; i.e., for any $w \neq x \in [y_1, y_r]$ and any $\lambda \in (0, 1)$, $I(\lambda w + (1-\lambda)x) < \lambda I(w) + (1 - \lambda)I(x)$.

We have the following large deviation bounds for S_n/n; for A any subset of $[y_1, y_r]$, $I(A)$ denotes the quantity $\inf_{x \in A} I(x)$.

(a) For any closed subset F of $[y_1, y_r]$

$$\limsup_{n \to \infty} \frac{1}{n} \log P_n\{S_n/n \in F\} \leq -I(F).$$

10

(b) *For any open subset G of $[y_1, y_r]$*

$$\limsup_{n \to \infty} \frac{1}{n} \log P_n\{S_n/n \in G\} \geq -I(G).$$

Let z be a point in (y_1, y_r) satisfying $z \neq \bar{y}$. It is not difficult to deduce from Cramér's Theorem that for the closed interval $F_{z,a}$ defined in (2.3) we obtain the large deviation limit

$$\lim_{n \to \infty} \frac{1}{n} \log P_n\{S_n/n \in F_{z,a}\} = -I(F_{z,a}).$$

Since z is the closest point in $F_{z,a}$ to $\bar{y}$, which does not lie in $F_{z,a}$, and since I is a finite, strictly convex function on $[y_1, y_r]$, the property of I stated in (2.7) implies that $I(F_{z,a}) = \inf_{x \in F_{z,a}} I(x) = I(z) > 0 = I(\bar{y})$. Hence the last equation becomes

$$\lim_{n \to \infty} \frac{1}{n} \log P_n\{S_n/n \in F_{z,a}\} = -I(z) < 0. \tag{2.8}$$

This equation shows that $P_n\{S_n/n \in F_{z,a}\} \to 0$ with exponential rate $\exp[-nI(z)]$ as $n \to 0$. Since as a decreases to 0, the interval $F_{z,a}$ decreases to $\{z\}$, it is a rigorous formulation of the heuristic limit noted in (2.2).

The discrete ideal gas is a simple model because it is defined in terms of a finite sequence of independent random variables X_i that are uniformly distributed on the finite set Λ consisting of the energy values $y_j, j = 1, 2, \ldots, r$. Surprisingly, the model has significant physical content, which can be seen by conditioning the product measure P_n on the energy constraint that S_n/n lies in the closed interval $F_{z,a}$. This conditioned measure defines the **microcanonical ensemble**, which for subsets B of Ω_n has the form $P_n(B|S_n/n \in F_{z,a}\}$ for subsets B of Ω_n. The constraint $S_n/n \in F_{z,a}$ is introduced to model conservation of energy in the discrete ideal gas.

In Theorem 4.2 we determine the Maxwell-Boltzmann distribution of the discrete ideal gas with respect to the microcanonical ensemble in the limit $n \to \infty$, which defines the **thermodynamic limit**. As we point out before the statement of Theorem 4.1, values of z lying in the low-energy interval $(y_1, \bar{y})$ correspond to the physically relevant region of positive temperatures while values of z lying in the high-energy interval $[\bar{y}, y_r)$ correspond to the physically nonrelevant region of negative temperatures or infinite temperature.

We now turn to macroscopic level 2 of description in terms of the empirical vector. The large deviation behavior of the empirical vector is expressed in the Boltzmann-Sanov LDP in Theorem 2.3, which will be applied in section 4 when we study the Maxwell-Boltzmann distribution.

11

2d   Macroscopic level 2: empirical vector

The most elementary macroscopic level of description is in terms of the sample mean

$$S_n(\omega)/n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i(\omega) = \frac{1}{n} \cdot \sum_{i=1}^{n} \omega_i.$$

This macroscopic variable summarizes the r^n degrees of freedom in the microscopic description $\omega \in \Omega_n$ in terms of a single quantity. A more refined macroscopic level of description is in terms of the empirical vector. For $\omega \in \Omega_n$ and $y \in \Lambda$ we define

$$L_n(y) = L_n(\omega, y) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(\omega)}\{y\}.$$

Thus $L_n(\omega, y)$ counts the relative frequency with which y appears in the configuration ω; in symbols, $L_n(\omega, y) = n^{-1} \cdot \text{card}\{i \in \{1, \ldots, n\} : \omega_i = y\}$. We then define the empirical vector

$$
\begin{aligned}
L_n &= L_n(\omega) = (L_n(\omega, y_1), \ldots, L_n(\omega, y_r)) \qquad (2.9) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\delta_{X_i(\omega)}\{y_1\}, \ldots, \delta_{X_i(\omega)}\{y_r\} \right).
\end{aligned}
$$

L_n equals the sample mean of the i.i.d. random vectors $(\delta_{X_i(\omega)}\{y_1\}, \ldots, \delta_{X_i(\omega)}\{y_r\})$. It takes values in the set of probability vectors

$$\mathcal{P}_r = \left\{ \theta = (\theta_1, \theta_2, \ldots, \theta_r) \in \mathbb{R}^r : \theta_j \geq 0, \sum_{j=1}^{r} \theta_j = 1 \right\}.$$

We use the same notation to denote the empirical measure

$$L_n = L_n(\omega) = \frac{1}{n} \sum_{j=1}^{r} L_n(\omega, y_j) \delta_{y_j},$$

which for $\omega \in \Omega_n$ is a probability measure on Λ. Similarly any probability vector $\theta \in \mathcal{P}_r$ can be identified with the quantity $\sum_{j=1}^{r} \theta_j \delta_{y_j}$, which is a probability measure on Λ.

The macroscopic variables L_n and S_n/n are closely related. In fact, the mean of the empirical measure L_n equals the sample mean S_n/n; in symbols, for each $\omega \in \Omega_n$, $\sum_{j=1}^{r} y_j L_n(y_j, \omega) = S_n(\omega)/n$. This fact is proved right after (4.3) and is used in the proof of part (a) of Theorem 4.2.

We now give a formal overview of the asymptotic behavior of L_n, which will soon be reformulated in a rigorous way. The limiting behavior of L_n is straightforward to determine. Since the X_i have the common distribution $\rho = (\rho_1, \rho_2, \ldots, \rho_r)$ with each $\rho_j = 1/r$, for each $y_k \in \Lambda$

$$E^{P_n}\{L_n(y_k)\} = E^{P_n}\left\{\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}\{y_k\}\right\} = \frac{1}{n}\sum_{i=1}^{n}P_n\{X_i = y_k\} = \rho_k = 1/r.$$

Hence by the weak law of large numbers for the sample means of i.i.d. random vectors

$$\lim_{n\to\infty} P_n\{L_n \sim \rho\} = 1.$$

The Boltzmann-Sanov LDP, stated in Theorem 2.3, expresses the large deviation behavior of L_n. It shows that there exists a function $R(\cdot|\rho)$ mapping $\mathcal{P}_r$ into $[0, \infty)$ and having the property that for any $\theta \in \mathcal{P}_r$

$$P_n\{L_n \sim \theta\} \approx \exp[-nR(\theta|\rho)] \text{ as } n \to \infty.$$

We now combine the law of large numbers with the large deviation estimate in the last equation. If $z = \rho$, then $P_n\{L_n \sim \rho\} \to 1$, and if $\theta \neq \rho$, then $P_n\{L_n \sim \theta\} \to 0$. Thus if $\theta = \rho$, then we expect that $R(\theta|\rho) = 0$. On the other hand, if $\theta \neq \rho$, then we expect that $R(\theta|\rho) > 0$, implying that

$$P_n\{L_n \sim \theta\} \approx \exp[-nR(\theta|\rho)] \to 0 \text{ exponentially fast as } n \to \infty. \qquad (2.10)$$

As in level 1, the last equation shows that $R(\theta|\rho)$ records the multiplicity of microstates ω consistent with the macrostate θ through the macroscopic variable L_n.

So far this discussion of the asymptotic behavior of L_n parallels the discussion of the asymptotic behavior of the sample mean in level 1. In the presentation of level 1, the rate function I is defined by the Legendre-Fenchel transform (2.6). Unfortunately, for general r the rate function I cannot be expressed in explicit form. By contrast, the rate function $R(\cdot|\rho)$ for level 2 has the explicit form of the relative entropy, which is of fundamental importance in probability theory, information theory, and other areas of mathematics and science. The definition and basic properties of the relative entropy are given in the next theorem. The function $x \log x$, well defined for $x > 0$, is extended to $x = 0$ by continuity; thus for $x = 0$ we define $x \log x = 0$.

Theorem 2.2. *The relative entropy of $\theta \in \mathcal{P}_r$ with respect to ρ is defined by*

$$R(\theta|\rho) = \sum_{j=1}^{r}\theta_j \log\frac{\theta_j}{\rho_j} = \sum_{j=1}^{r}\theta_j \log(r\theta_j) = \log r + \sum_{j=1}^{r}\theta_j \log\theta_j.$$

This function has the following properties.

(a) $R(\cdot|\rho)$ *is finite and continuous on* $\mathcal{P}_r$. *In addition* $R(\cdot|\rho)$ *is strictly convex; i.e., for any* $\theta \neq \mu \in \mathcal{P}_r$ *and any* $\lambda \in (0,1)$, $R(\lambda\theta + (1-\lambda)\mu|\rho) < \lambda R(\theta|\rho) + (1-\lambda)R(\mu|\rho)$.

(b) *We have*

$$R(\theta|\rho) > R(\rho|\rho) = 0 \text{ for all } \theta \in \mathcal{P}_r \text{ satisfying } \theta \neq \rho.$$

Thus $R(\theta|\rho)$ *attains its infimum of* 0 *over* $\mathcal{P}_r$ *at the unique measure* $\theta = \rho$.

Proof. (a) The finiteness and continuity of $R(\cdot|\rho)$ on $\mathcal{P}_r$ follows from the explicit formula defining $R(\cdot|\rho)$. The strict convexity of this function is a consequence of the strict convexity of $x \log x$ for $x \geq 0$.

(b) If $\theta = \rho$, then by definition $R(\theta|\rho) = 0$. To prove that $R(\cdot|\rho)$ attains its infimum over $\mathcal{P}_r$ at the unique measure ρ, we use a global, convexity-based inequality rather than calculus. This inequality is that for $x \geq 0$, $x \log x \geq x - 1$ with equality if and only if $x = 1$. The proof of this statement follows from the fact that the graph of the strictly convex function $x \log x$ for $x \geq 0$ has the tangent line $y = x - 1$ at $x = 1$. It follows that for any $\theta \in \mathcal{P}_r$

$$\frac{\theta_j}{\rho_j} \log \frac{\theta_j}{\rho_j} \geq \frac{\theta_j}{\rho_j} - 1$$

with equality if and only if $\theta_j = \rho_j = 1/r$. Multiplying this inequality by ρ_j and summing over j yields

$$R(\theta|\rho) = \sum_{j=1}^{r} \theta_j \log \frac{\theta_j}{\rho_j} \geq \sum_{j=1}^{r} (\theta_j - \rho_j) = 0.$$

In order to complete the proof of part (b), we must show that if $R(\theta|\rho) = 0$, then $\theta = \rho$. Assume that $R(\theta|\rho) = 0$. Then

$$\begin{aligned}
0 &= \sum_{j=1}^{r} \theta_j \log \frac{\theta_j}{\rho_j} = \sum_{j=1}^{r} \left(\theta_j \log \frac{\theta_j}{\rho_j} - (\theta_j - \rho_j) \right) \\
&= \sum_{j=1}^{r} \rho_j \left(\frac{\theta_j}{\rho_j} \log \frac{\theta_j}{\rho_j} - \left(\frac{\theta_j}{\rho_j} - 1 \right) \right).
\end{aligned}$$

We again use the fact that for $x \geq 0$, $x \log x \geq x - 1$ with equality if and only if $x = 1$. It follows that for each j, $\theta_j = \rho_j$ and thus that $\theta = \rho$. This completes the proof that $R(\theta|\rho) > R(\rho|\rho) = 0$ for all $\theta \in \mathcal{P}_r$ satisfying $\theta \neq \rho$. The proof of the theorem is complete. ∎

We next state the Boltzmann-Sanov Theorem, which is the LDP for L_n with rate function $R(\cdot|\rho)$.

14

Theorem 2.3 (**Boltzmann-Sanov**). *We have the following large deviation bounds for the empirical vector L_n; for A any subset of $\mathcal{P}_r$, $R(A|\rho)$ denotes the quantity $\inf_{\theta \in A} R(\theta|\rho)$.*

(a) *For any closed subset F of $\mathcal{P}_r$*

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

(b) *For any open subset G of $\mathcal{P}_r$*

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

As we point out in (2.9), L_n equals the sample of the independent, identically distributed (i.i.d.) random vectors $(\delta_{X_i}\{y_1\}, \ldots, \delta_{X_i}\{y_r\}$ for $j = 1, 2, \ldots, n$. The easiest way to prove the Boltzmann-Sanov LDP is to apply Cramér's Theorem for i.i.d. random vectors and to evaluate the Legendre-Fenchel transform defining the rate function, obtaining the relative entropy. This is carried out in [10, §VIII.2].

We are using the term Boltzmann-Sanov LDP to refer to the LDP for empirical vectors defined in terms of i.i.d. random variables and taking values in a finite set. A much more general form of this LDP is known for empirical measures based on i.i.d. random vectors taking values in infinite sets such as $\mathbb{N}$, $\mathbb{R}^d$ for $d \geq 2$, and general complete, separable metric spaces [8, §6.2], [9, Thm. 2.2]. In this more general context the LDP is known as Sanov's Theorem in honor of I. N. Sanov, one of the pioneers of the theory of large deviations [28].

In the next section we will take a more circuitous route to proving the Boltzmann-Sanov LDP by deriving it from a local large deviation estimate due to Boltzmann that studies the asymptotic behavior of a certain multinomial coefficient. The virtue of this approach is that it serves as a model for the proof of the LDP for the droplet model, for which standard formulations of the Boltzmann-Sanov Theorem cannot be applied. The droplet model is introduced in section 5, and the procedure for proving the LDP is outlined in section 8.

3   Proof of Theorem 2.3 from Boltzmann's Local Estimate

We prove the Boltzmann-Sanov LDP in Theorem 2.3 in four steps. Let $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^r$. If $\theta^{(n)}$ is a sequence in $\mathcal{P}_r$ and $\theta \in \mathcal{P}_r$, then we write $\theta^{(n)} \to \theta$ to denote the convergence $\|\theta - \theta^{(n)}\| \to 0$ as $n \to \infty$.

1. **Local estimate.** In Theorem 3.1 we prove Boltzmann's local large deviation (LD) estimate. This states that for $\theta^{(n)} \in \mathcal{P}_r$ lying in the range of L_n

$$\frac{1}{n} \log P_n\{L_n = \theta^{(n)}\} = -R(\theta^{(n)}|\rho) + \varepsilon_n(\theta^{(n)}),$$

where $\varepsilon_n(\theta^{(n)}) \to 0$ uniformly for $\theta^{(n)}$ in the range of L_n as $n \to \infty$.

2. **Approximation result.** In Theorem 3.2 we show that for any probability vector $\theta \in \mathcal{P}_r$ there exists a sequence $\theta^{(n)}$ in the range of L_n such that $\theta^{(n)} \to \theta$ and $R(\theta^{(n)}|\rho) \to R(\theta|\rho)$.

3. **Large deviation limit for open balls.** In Theorem 3.3 we show how to use the approximation result in step 2 to lift the local estimate in step 1 to the large deviation limit for L_n lying in open balls.

4. **Large deviation upper and lower bounds**. We end the section by showing how to lift the large deviation limit for L_n lying in open balls in step 3 to the large deviation upper bound for L_n lying in closed subsets of $\mathcal{P}_r$ and the large deviation lower bound for L_n lying in open sets of $\mathcal{P}_r$. This will complete the proof of the Boltzmann-Sanov Theorem.

We introduce notation that will help us formulate these four steps. Let $\mathbb{N}_0$ denote the set of nonnegative integers. We define A_n to be set of $\nu = (\nu_1, \nu_2, \ldots, \nu_r) \in \mathbb{N}_0^r$ satisfying $\sum_{j=1}^r \nu_j = n$, and we define D_n to be the range of $L_n(\omega)$ for $\omega \in \Omega_n$. D_n is a subset of the set of probability vectors $\mathcal{P}_r$ in $\mathbb{R}^r$. By definition of the empirical vector, any probability vector $\theta^{(n)} = (\theta_1^{(n)}, \theta_2^{(n)}, \ldots, \theta_r^{(n)})$ lying in the range of $L_n(\omega)$ for $\omega \in \Omega_n$ has the form for $j = 1, 2, \ldots, r$

$$\theta_j^{(n)} = \frac{\nu_j}{n} \text{ for some } \nu \in A_n.$$

We are now ready to state Boltzmann's local large deviation estimate. The proof is based on the asymptotic analysis of a certain multinomial coefficient.

Theorem 3.1 (**Boltzmann's local LD estimate**). *Let D_n denote the range of L_n. For $\theta^{(n)} \in D_n$*

$$\frac{1}{n} \log P_n\{L_n = \theta^{(n)}\} = -R(\theta^{(n)}|\rho) + \varepsilon_n(\theta^{(n)}),$$

where $\varepsilon_n(\theta^{(n)}) \to 0$ uniformly for $\theta^{(n)} \in D_n$ as $n \to \infty$.

Proof. Let $M_n = \{i \in \mathbb{N} : 1 \le i \le n\}$. Each component $\theta_j^{(n)}$ of $\theta^{(n)}$ has the form $\theta_j^{(n)} = \nu_j/n$ for some $\nu \in A_n$. Since each component ρ_j of ρ equals $1/r$, we have by (2.1) and elementary combinatorics

$$P_n\{L_n = \theta^{(n)}\} \tag{3.1}$$
$$= P_n\left\{ \omega \in \Omega_n : L_n(\omega) = \frac{1}{n}(n\theta_1^{(n)}, n\theta_2^{(n)}, \ldots, n\theta_r^{(n)}) \right\}$$
$$= P_n\{\omega \in \Omega_n : \text{card}\{i \in M_n : \omega_i = y_j\} = \nu_j \text{ for } j = 1, 2, \ldots, r\}$$
$$= \text{card}\{\omega \in \Omega_n : \text{card}\{i \in M_n : \omega_i = y_j\} = \nu_j \text{ for } j = 1, 2, \ldots, r\} \cdot \frac{1}{r^n}$$
$$= \frac{n!}{\prod_{j=1}^r \nu_j!} \cdot \frac{1}{r^n}.$$

16

We now apply a weak form of Stirling's approximation, which states that for all $n \in \mathbb{N}$ satisfying $n \geq 2$ and for all $\kappa \in \mathbb{N}_0$ satisfying $0 \leq \kappa \leq n$

$$1 \leq \log(\kappa!) - (\kappa \log \kappa - \kappa) \leq 2 \log n. \tag{3.2}$$

The straightforward proof begins as follows: for $\kappa \in \mathbb{N}$ satisfying $2 \leq \kappa \leq n$

$$
\begin{aligned}
\kappa \log \kappa - \kappa + 1 &= \int_1^\kappa \log x \, dx \leq \log(\kappa!) = \sum_{j=2}^\kappa \log j \\
&\leq \int_2^\kappa \log x \, dx + \log \kappa \\
&= \kappa \log \kappa - \kappa + 2(1 - \log 2) + \log \kappa.
\end{aligned}
$$

This implies that

$$1 \leq \log(\kappa!) - (\kappa \log \kappa - \kappa) \leq 2(1 - \log 2) + \log \kappa.$$

Since $\log 2 = 0.69\ldots > 2/3$, we obtain (3.2) for $2 \leq \kappa \leq n$. Because (3.2) is also valid for $\kappa = 0$ and $\kappa = 1$, the proof of the weak form of Stirling's approximation is complete. We summarize (3.2) by writing

$$\log(\kappa!) = \kappa \log \kappa - \kappa + \mathrm{O}(\log n) \;\forall n \in \mathbb{N}, n \geq 2 \text{ and } \forall \kappa \in \{0, 1, \ldots, n\}. \tag{3.3}$$

By (3.2) the term denoted by $\mathrm{O}(\log n)$ satisfies $1 \leq \mathrm{O}(\log n) \leq 2 \log n$.

For any $\nu \in A_n$ and each $j = 1, 2, \ldots, r$, since the components ν_j satisfy $0 \leq \nu_j \leq n$, we have

$$\log(\nu_j!) = \nu_j \log \nu_j - \nu_j + \mathrm{O}(\log n) \text{ for all } n \geq 2.$$

Using the fact that $\sum_{j=1}^r \nu_j = n$, we obtain

$$
\frac{1}{n} \log \left(\frac{n!}{\prod_{j=1}^r \nu_j!} \cdot \frac{1}{r^n} \right) \tag{3.4}
$$

$$
= \frac{1}{n} \log(n!) - \frac{1}{n} \sum_{j=1}^r \log(\nu_j!) - \log r
$$

$$
= \frac{1}{n}(n \log n - n + \mathrm{O}(\log n)) - \frac{1}{n} \sum_{j=1}^r (\nu_j \log \nu_j - \nu_j + \mathrm{O}(\log n)) - \log r
$$

$$
= -\sum_{j=1}^r (\nu_j/n) \log(\nu_j/n) - \log r + \frac{\mathrm{O}(\log n)}{n} - \frac{1}{n} \sum_{j=1}^r \mathrm{O}(\log n)
$$

$$
= -\sum_{j=1}^r \theta_j^{(n)} \log \theta_j^{(n)} - \log r + \zeta_n^{(1)} - \zeta_n^{(2)}(\theta^{(n)}),
$$

17

where $\zeta_n^{(1)} = [\mathrm{O}(\log n)]/n \to 0$ as $n \to \infty$ and

$$\zeta_n^{(2)}(\theta^{(n)}) = \frac{1}{n} \sum_{j=1}^{r} \mathrm{O}(\log n).$$

By the inequality noted after (3.3)

$$0 \leq \max_{\theta^{(n)} \in D_n} \zeta_n^{(2)}(\theta^{(n)}) \leq \max_{\theta^{(n)} \in D_n} \frac{2}{n} \sum_{j=1}^{r} \log n \leq \frac{2r \log n}{n}.$$

Since $(r \log n)/n \to 0$ as $n \to \infty$, we conclude that $\zeta_n^{(2)}(\theta^{(n)}) \to 0$ uniformly for $\theta^{(n)} \in D_n$ as $n \to \infty$.

Since $\sum_{j=1}^{r} \theta_j^{(n)} = 1$, we can write (3.4) in the form

$$
\begin{aligned}
\frac{1}{n} \log \left(\frac{n!}{\prod_{j=1}^{r} \nu_j!} \cdot \frac{1}{r^n} \right) &= -\sum_{j=1}^{r} \theta_j^{(n)} \log(r\theta_j^{(n)}) + \varepsilon_n(\theta^{(n)}) \\
&= -R(\theta^{(n)}|\rho) + \varepsilon_n(\theta^{(n)}),
\end{aligned}
$$

where $\varepsilon_n(\theta^{(n)}) = \zeta_n^{(1)} - \zeta_n^{(2)}(\theta^{(n)}) \to 0$ uniformly for $\theta^{(n)} \in D_n$ as $n \to \infty$. Substituting this equation into (3.1) completes the proof of the theorem. ∎

Theorem 3.1 is a local large deviation estimate. Using Theorems 3.2 and 3.3, we will convert this local estimate into the global estimate given in the Boltzmann-Sanov LDP in Theorem 2.3. Before proceeding with this proof, we give a formal argument showing how one can obtain a global estimate for $P_n\{L_n \in \Gamma\}$, where Γ is a Borel subset of $\mathcal{P}_r$. If we summarize the conclusion of Theorem 3.1 by the formal notation introduced in (2.10), then we can write

$$P_n\{L_n \in \Gamma\} = \sum_{\theta \in \Gamma \cap D_n} P_n\{L_n \sim \theta\} \approx \sum_{\theta \in \Gamma \cap D_n} \exp[-nR(\theta|\rho)].$$

The range of $L_n(\omega)$ for $\omega \in \Omega_n$ is the set D_n of probability vectors having the form ν/n, where $\nu \in \mathbb{R}^r$ has nonnegative integer coordinates summing to n; hence the cardinality of D_n does not exceed $(n+1)^r$. Since for any $\theta \in \Gamma$ we have $R(\theta|\rho) \geq R(\Gamma|\rho)$,

$$\exp[-nR(\Gamma|\rho)] \leq \sum_{\theta \in \Gamma \cap D_n} \exp[-nR(\theta|\rho)] \leq (n+1)^r \exp[-nR(\Gamma|\rho)],$$

one expects that to exponential order the following holds:

$$P_n\{L_n \in \Gamma\} \approx \exp[-nR(\Gamma|\rho)] \text{ as } n \to \infty. \tag{3.5}$$

18

This is a heuristic formulation of Theorem 2.3, which has separate bounds for open subsets and closed subsets of $\mathcal{P}_r$.

The next step in the proof of the Boltzmann-Sanov LDP in Theorem 2.3 is to prove the following approximation result, which will allow us to lift the local estimate in Theorem 3.1 to the large deviation limit for balls in Theorem 3.3.

Theorem 3.2. *Let θ be any probability measure in $\mathcal{P}_r$ and let D_n denote the range of $L_n(\omega)$ for $\omega \in \Omega_n$. Then there exists a sequence $\theta^{(n)} \in D_n$ for which the following properties hold.*
    (a) $\theta^{(n)} \to \theta$ *as $n \to \infty$.*
    (b) $R(\theta^{(n)}|\rho) \to R(\theta|\rho)$ *as $n \to \infty$.*

Proof. It suffices to determine a sequence $\theta^{(n)} \in D_n$ that satisfies part (a). If $\theta^{(n)} \in D_n$ satisfies part (a), then part (b) follows by the continuity of $R(\cdot|\rho)$ on $\mathcal{P}_r$. For $x \in \mathbb{R}$ we denote by $\lfloor x \rfloor$ the largest integer less than or equal to x. Given $\theta = (\theta_1, \theta_2, \ldots, \theta_r) \in \mathcal{P}_r$ we determine a sequence $\nu^{(n)} \in A_n$ such that the probability vectors $\theta^{(n)} \in \mathcal{P}_r$ with coordinates $\theta_j^{(n)} = \nu_j^{(n)}/n$ satisfy $\theta^{(n)} \to \theta$ as $n \to \infty$. For $j = 2, 3, \ldots, r$ the definition of these components is $\nu_j^{(n)} = \lfloor n\theta_j \rfloor$. We then define

$$\nu_1^{(n)} = n - \sum_{j=2}^{r} \nu_j^{(n)} = n - \sum_{j=2}^{r} \lfloor n\theta_j \rfloor.$$

For $j = 2, 3, \ldots, r$ each $\nu_j^{(n)}$ is a nonnegative integer, and since $n\theta_j - 1 \leq \nu_j^{(n)} \leq n\theta_j$, it follows that

$$\lim_{n\to\infty} \theta_j^{(n)} = \lim_{n\to\infty} \frac{\nu_j^{(n)}}{n} = \lim_{n\to\infty} \frac{\lfloor n\theta_j \rfloor}{n} = \theta_j.$$

In addition, since

$$\nu_1^{(n)} = n - \sum_{j=2}^{r} \nu_j^{(n)} \geq n - n\sum_{j=2}^{r} \theta_j = n\left(1 - \sum_{j=2}^{r} \theta_j\right) = n\theta_1 \geq 0,$$

we see that $\nu_1^{(n)}$ is also a nonnegative integer and that

$$\lim_{n\to\infty} \theta_1^{(n)} = \lim_{n\to\infty} \frac{\nu_1^{(n)}}{n} = 1 - \lim_{n\to\infty} \sum_{j=2}^{r} \frac{\nu_j^{(n)}}{n} = 1 - \sum_{j=2}^{r} \theta_j = \theta_1.$$

We conclude that $\nu^{(n)} \in A_n$, $\theta^{(n)} \in D_n$, and $\theta^{(n)} \to \theta$ as $n \to \infty$. This completes the proof of the theorem. ∎

We now use the approximation result in Theorem 3.2 to lift the local large deviation estimate in Theorem 3.1 to the following large deviation limit for L_n lying in open balls. For $\theta \in \mathcal{P}_r$ and $\varepsilon > 0$ we define the open ball with center θ and radius ε by

$$B(\theta, \varepsilon) = \{\mu \in \mathcal{P}_r : \|\theta - \mu\| < \varepsilon\}.$$

Theorem 3.3. *Let θ be a probability vector in $\mathcal{P}_r$ and take $\varepsilon > 0$. Then for any open ball $B(\theta, \varepsilon)$ we have the large deviation limit*

$$\lim_{n \to \infty} \frac{1}{n} \log P_n\{L_n \in B(\theta, \varepsilon)\} = -R(B(\theta, \varepsilon)|\rho) = - \inf_{\nu \in B(\theta, \varepsilon)} R(\nu|\rho).$$

Proof. To ease the notation we write B for the open ball $B(\theta, \varepsilon)$. By the local large deviation estimate in Theorem 3.1

$$P_n\{L_n \in B\} = \sum_{\theta^{(n)} \in B \cap D_n} P_n\{L_n = \theta^{(n)}\} = \sum_{\theta^{(n)} \in B \cap D_n} \exp[-n(R(\theta^{(n)}|\rho) - \varepsilon_n(\theta^{(n)}))].$$

For the last sum in this equation we have the bounds

$$\max_{\theta^{(n)} \in B \cap D_n} \exp[-n(R(\theta^{(n)}|\rho) - \varepsilon_n(\theta^{(n)}))]$$
$$\leq \sum_{\theta^{(n)} \in B \cap D_n} \exp[-n(R(\theta^{(n)}|\rho) - \varepsilon_n(\theta^{(n)}))]$$
$$\leq \operatorname{card}(D_n) \cdot \max_{\theta^{(n)} \in B \cap D_n} \exp[-n(R(\theta^{(n)}|\rho) - \varepsilon_n(\theta^{(n)}))].$$

In addition, for the term $\max_{\theta^{(n)} \in B \cap D_n} \exp[-n(R(\theta^{(n)}|\rho) - \varepsilon_n(\theta^{(n)}))]$ we have the bounds

$$\exp\left[-n\left(R(B \cap D_n|\rho) + \max_{\theta^{(n)} \in B \cap D_n} \varepsilon_n(\theta^{(n)})\right)\right]$$
$$= \exp\left[-n\left(\min_{\theta^{(n)} \in B \cap D_n} R(\theta^{(n)}|\rho) + \max_{\theta^{(n)} \in B \cap D_n} \varepsilon_n(\theta^{(n)})\right)\right]$$
$$\leq \max_{\theta^{(n)} \in B \cap D_n} \exp[-n(R(\theta^{(n)}|\rho) - \varepsilon_n(\theta^{(n)}))]$$
$$\leq \exp\left[-n\left(\min_{\theta^{(n)} \in B \cap D_n} R(\theta^{(n)}|\rho) - \max_{\theta^{(n)} \in B \cap D_n} \varepsilon_n(\theta^{(n)})\right)\right]$$
$$= \exp\left[-n\left(R(B \cap D_n|\rho) - \max_{\theta^{(n)} \in B \cap D_n} \varepsilon_n(\theta^{(n)})\right)\right].$$

It follows that

$$-R(B \cap D_n|\rho) - \max_{\theta^{(n)} \in B \cap D_n} \varepsilon_n(\theta^{(n)})$$

$$\leq \quad \frac{1}{n} \log P_n\{L_n \in B\}$$

$$\leq \quad -R(B \cap D_n|\rho) + \max_{\theta^{(n)} \in B \cap D_n} \varepsilon_n(\theta^{(n)}) + \frac{\log(\mathrm{card}(D_n))}{n}.$$

The last term in the last display converges to 0 as $n \to \infty$ because the cardinality of D_n does not exceed $(n+1)^r$. Since $\varepsilon_n(\theta^{(n)}) \to 0$ uniformly for $\theta^{(n)} \in D_n$, the proof is done once we show that

$$\lim_{n \to \infty} R(B \cap D_n|\rho) = R(B|\rho). \tag{3.6}$$

Since $B \cap D_n \subset B$, we have $R(B|\rho) \leq R(B \cap D_n|\rho)$, which implies that

$$R(B|\rho) \leq \liminf_{n \to \infty} R(B \cap D_n|\rho).$$

The limit in (3.6) is proved if we can show that

$$\limsup_{n \to \infty} R(B \cap D_n|\rho) \leq R(B|\rho). \tag{3.7}$$

For any $\delta > 0$ there exists $\theta^\star \in B$ such that $R(\theta^\star|\rho) \leq R(B|\rho) + \delta$. Theorem 3.2 guarantees the existence of a sequence $\theta^{(n)} \in D_n$ such that $\theta^{(n)} \to \theta^\star$ and $R(\theta^{(n)}|\rho) \to R(\theta^\star|\rho)$. Since for all sufficiently large n we have $\theta^{(n)} \in B \cap D_n$, it follows that $R(B \cap D_n|\rho) \leq R(\theta^{(n)}|\rho)$. Hence

$$\limsup_{n \to \infty} R(B \cap D_n|\rho) \leq \lim_{n \to \infty} R(\theta^{(n)}|\rho) = R(\theta^\star) \leq R(B|\rho) + \delta.$$

Taking $\delta \to 0$ gives (3.7) and thus proves the limit (3.6). This completes the proof of the theorem. ∎

We are now ready to prove the Boltzmann-Sanov LDP in Theorem 2.3, for which the main tool is the large deviation limit for open balls in Theorem 3.3. The proof of the large deviation lower bound for open sets is straightforward. The more challenging proof of the large deviation upper bound for closed sets is based on a covering argument involving open balls.

Proof of Boltzmann-Sanov LDP in Theorem 2.3. We first prove the large deviation lower bound in part (b) of Theorem 2.3. Let G be any open subset of $\mathcal{P}_r$. For any point $\theta \in G$ there exists $\varepsilon > 0$ such that the open ball $B(\theta, \varepsilon)$ is a subset of G. Theorem 3.3 implies that

$$\liminf_{n \to \infty} \frac{1}{n} \log P_n(L_n \in G) \geq \lim_{n \to \infty} \frac{1}{n} \log P_n(L_n \in B(\theta, \varepsilon))$$

$$= -R(B(\theta, \varepsilon)|\rho) \geq -R(\theta|\rho).$$

21

Since θ is an arbitrary point in G, it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(L_n \in G) \geq -\inf_{\theta\in G} R(\theta|\rho) = -R(G|\rho).$$

This completes the proof of the large deviation lower bound for any open subset G of $\mathcal{P}_r$.

We now prove the large deviation upper bound in part (a) of Theorem 2.3. Let F be any closed subset of $\mathcal{P}_r$. Since $\mathcal{P}_r$ is a compact subset of $\mathbb{R}^r$, F is also compact. We start the proof of the large deviation upper bound by showing that for each $\theta \in F$

$$\liminf_{\varepsilon\to 0^+} R(B(\theta,\varepsilon)|\rho) \geq R(F|\rho). \tag{3.8}$$

Let ε_n be any positive sequence converging to 0, and take any $\delta > 0$. For any $n \in \mathbb{N}$ there exists $\theta^{(n)} \in B(\theta,\varepsilon_n)$ such that $R(B(\theta,\varepsilon_n)|\rho) + \delta \geq R(\theta^{(n)}|\rho)$. Since $\theta^{(n)} \to \theta$, the continuity of $R(\cdot|\rho)$ on $\mathcal{P}_r$ and the fact that $\theta \in F$ imply that

$$\liminf_{n\to\infty} R(B(\theta,\varepsilon_n)|\rho) + \delta \geq \lim_{n\to\infty} R(\theta^{(n)}|\rho) = R(\theta|\rho) \geq R(F|\rho).$$

Sending $\delta \to 0$ yields (3.8) because ε_n is an arbitrary positive sequence converging to 0.

We now prove the large deviation upper bound in part (a) of Theorem 2.3. Take any $\eta > 0$. By (3.8) for each $\theta \in F$ there exists $\varepsilon_\theta > 0$ such that

$$R(B(\theta,\varepsilon_\theta)|\rho) \geq R(F|\rho) - \eta.$$

The open balls $\{B(\theta,\varepsilon_\theta), \theta \in F\}$ cover F. Since F is compact, there exist $M < \infty$ and finitely many points $\theta^{(i)} \in F, i = 1, 2, \ldots, M$, such that $F \subset \bigcup_{i=1}^{M} B(\theta^{(i)}, \varepsilon_i)$, where $\varepsilon_i = \varepsilon_{\theta_i}$. It follows that

$$\min_{i=1,2,\ldots,M} R(B(\theta^{(i)}, \varepsilon_i)|\rho) \geq R(F|\rho) - \eta.$$

We now apply to $B = B(\theta^{(i)}, \varepsilon_i)$ the large deviation limit for L_n lying in open balls proved in Theorem 3.3, obtaining

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n\{L_n \in F\} \tag{3.9}$$

$$\leq \limsup_{n\to\infty} \frac{1}{n} \log P_n\left(L_n \in \bigcup_{i=1}^{M} B(\theta^{(i)}, \varepsilon_i)\right)$$

$$\leq \limsup_{n\to\infty} \frac{1}{n} \log \left(\sum_{i=1}^{M} P_n(L_n \in B(\theta^{(i)}, \varepsilon_i))\right)$$

$$= \max_{i=1,2,\ldots,M} \left(\limsup_{n\to\infty} \frac{1}{n} \log P_n(L_n \in B(\theta^{(i)}, \varepsilon_i))\right)$$

$$= -\min_{i=1,2,\ldots,M} R(B(\theta^{(i)}, \varepsilon_i)\rho) \leq -R(F|\rho) + \eta.$$

22

The next-to-last step is a consequence of the easily verified fact [8, Lem. 1.2.15] that for any positive sequence of real numbers $a_{n,i}, i = 1, 2, \ldots, M$

$$\limsup_{n \to \infty} \frac{1}{n} \log \left(\sum_{i=1}^{M} a_{n,i} \right) = \max_{i=1,2,\ldots,M} \left(\limsup_{n \to \infty} \frac{1}{n} \log a_{n,i} \right).$$

Sending $\eta \to 0$ in the last line of (3.9), we obtain

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

This completes the proof of the large deviation upper bound for any closed subset F of $\mathcal{P}_r$. The proof of the theorem is done. ∎

In the next section we apply the Boltzmann-Sanov LDP in Theorem 2.3 to determine the Maxwell-Boltzmann equilibrium distribution of the energy values in the discrete ideal gas.

4   Derivation of Maxwell-Boltzmann Distribution

The Maxwell-Boltzmann distribution is the equilibrium distribution of energy values in the discrete ideal gas with respect to the microcanonical ensemble, which is defined by $P_n(B|S_n/n \in F_{z,a}\}$ for subsets B of Ω_n. We recall that the energy levels of the n particles in the discrete ideal gas are represented by a sequence of independent random variables $X_i, i = 1, 2, \ldots, n$, each having the uniform distribution $\rho = \sum_{j=1}^{r} (1/r) \delta_{y_j}$; the quantities y_j represent the r possible energy values. In the definition of the microcanonical ensemble we impose the energy constraint that the average random energy S_n/n lies in the interval $F_{z,a}$ defined in (2.3). The quantity z satisfies $y_1 < z < \bar{y}$, $z = \bar{y}$, or $\bar{y} < z < y_r$, and $a > 0$ is chosen so small that $F_{z,a}$ is a subset of (y_1, y_r). By (2.8), when $y_1 < z < \bar{y}$ or $\bar{y} < z < y_r$ the probability of the event $\{S_n/n \in F_{z,a}\}$ converges to 0 exponentially fast as $n \to \infty$.

In subsections 4a and 4b we present two derivations of the Maxwell-Boltzmann distribution. The first is a heuristic formulation due to Boltzmann and is based on the local large deviation estimate in Theorem 3.1. The second derivation is a rigorous formulation based on the Boltzmann-Sanov LDP in Theorem 2.3. In both derivations we use the notation $\langle \theta \rangle$ for $\theta \in \mathcal{P}_r$ to denote the mean of the probability measure $\theta = \sum_{j=1}^{r} \theta_j \delta_{y_j}$; in symbols

$$\langle \theta \rangle = \sum_{j=1}^{r} y_j \theta_j.$$

23

4a Boltzmann's formulation of Maxwell-Boltzmann distribution

In modern terminology, Boltzmann's idea for characterizing the Maxwell-Boltzmann distribution can be expressed as follows: the equilibrium distribution ρ^* of the discrete ideal gas is the distribution maximizing the probability $P_n\{L_n = \theta | S_n/n \in F_{z,a}\}$ over $\theta \in \mathcal{P}_r$. A non-probabilistic discussion of Boltzmann's formulation is given in [20, §4.2–4.3].

Boltzmann's idea is not rigorous because in general a fixed distribution $\theta \in \mathcal{P}_r$ is not in the range of L_n for a fixed n. Even if θ were in the range of L_n, the formulation would be imprecise because it seems to depend on the value of n while in fact the equilibrium distribution is an asymptotic phenomenon defined in the thermodynamic limit $n \to \infty$. Our goal is to use the local large deviation estimate in Theorem 3.1 to re-express Boltzmann's idea in a rigorous way. As we will see, this rigorous formulation of the Maxwell-Boltzmann distribution is closely related to the rigorous formulation based on the Boltzmann-Sanov LDP and presented in subsection 4b.

Let D_n denote the range of $L_n(\omega)$ for $\omega \in \Omega_n$. The local large deviation estimate in Theorem 3.1 states that for any $\theta^{(n)} \in D_n$

$$\frac{1}{n}\log P_n\{L_n = \theta^{(n)}\} = -R(\theta^{(n)}|\rho) + \varepsilon_n(\theta^{(n)}), \tag{4.1}$$

where $\varepsilon_n(\theta^{(n)}) \to 0$ uniformly for $\theta^{(n)} \in D_n$. If the error term $\varepsilon_n(\theta^{(n)})$ in (4.1) is dropped and we write θ in place of $\theta^{(n)}$, then the local large deviation estimate in Theorem 3.1 can be expressed as

$$P_n\{L_n = \theta\} \approx \exp[-nR(\theta|\rho)].$$

This suggests that if A is a subset of $\mathcal{P}_r$, then

$$\max_{\theta \in A} P_n\{L_n = \theta\} \approx \max_{\theta \in A}\exp[-nR(\theta|\rho)] = \exp\left[-n\min_{\theta \in A} R(\theta|\rho)\right]. \tag{4.2}$$

In turn this suggests that $P_n\{L_n = \theta\}$ is a maximum over $\theta \in A$ if and only if $R(\theta|\rho)$ is a minimum over $\theta \in A$.

How shall choose A so that maximizing $P_n\{L_n = \theta | S_n/n \in F_{z,a}\}$ over $\theta \in \mathcal{P}_r$ can be expressed in terms of minimizing $R(\theta|\rho)$ over $\theta \in A$? An insight leading to an answer is that for $\omega \in \Omega_n$ the sample mean $S_n(\omega)/n$ equals the mean of the empirical measure $L_n(\omega)$; in symbols

$$S_n(\omega)/n = \langle L_n(\omega)\rangle. \tag{4.3}$$

24

Here is a quick proof. By definition of $L_n(\omega, y_j)$

$$
\begin{aligned}
S_n(\omega)/n &= \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \\
&= \sum_{j=1}^{r} y_j \cdot \frac{1}{n}\text{card}\{i \in \{1, 2, \ldots, n\} : X_i(\omega) = y_j\} \\
&= \sum_{j=1}^{r} y_j L_n(\omega, y_j) = \langle L_n(\omega)\rangle.
\end{aligned}
$$

The third equality follows from the definition of $L_n(\omega, y_j)$. Using (4.3), we can write

$$
\begin{aligned}
P_n\{L_n = \theta | S_n/n \in F_{z,a}\} &= P_n\{L_n = \theta | \langle L_n\rangle \in F_{z,a}\} \\
&= P_n\{L_n = \theta, \langle L_n\rangle \in F_{z,a}\}/P_n\{\langle L_n\rangle \in F_{z,a}\} \\
&= P_n\{L_n = \theta, \langle \theta\rangle \in F_{z,a}\}/P_n\{\langle L_n\rangle \in F_{z,a}\} \\
&= \begin{cases} P_n\{L_n = \theta\}/P_n\{\langle L_n\rangle \in F_{z,a}\} & \text{if } \langle \theta\rangle \in F_{z,a} \\ 0 & \text{if } \langle \theta\rangle \notin F_{z,a}. \end{cases}
\end{aligned}
$$

We can now answer the question raised at the beginning of the preceding paragraph. When combined with (4.2), the last calculation suggests that if we choose $A = \{\theta \in \mathcal{P}_r : \langle \theta\rangle \in F_{z,a}\}$, then maximizing $P_n\{L_n = \theta | S_n/n \in F_{z,a}\}$ over $\theta \in \mathcal{P}_r$ can be expressed in terms of minimizing $R(\theta|\rho)$ over A; more precisely,

$$
\begin{aligned}
&\max_{\theta \in \mathcal{P}_r} P_n\{L_n = \theta | S_n/n \in F_{z,a}\} \\
&= \frac{1}{P_n\{\langle L_n\rangle \in F_{z,a}\}} \cdot \max\{P_n\{L_n = \theta\} : \theta \in \mathcal{P}_r, \langle \theta\rangle \in F_{z,a}\} \\
&\approx \frac{1}{P_n\{\langle L_n\rangle \in F_{z,a}\}} \cdot \exp[-n \cdot \min\{R(\theta|\rho) : \theta \in \mathcal{P}_r, \langle \theta\rangle \in F_{z,a}\}]
\end{aligned}
$$

This relationship between the maximum of $P_n\{L_n = \theta | S_n/n \in F_{z,a}\}$ over $\theta \in \mathcal{P}_r$ and the minimum of $R(\theta|\rho)$ over $\theta \in \mathcal{P}_r$ satisfying $\langle \theta\rangle \in F_{z,a}$ allows us to express Boltzmann's formulation of the Maxwell-Boltzmann distribution in the following form.

Boltzmann's formulation of Maxwell-Boltzmann distribution. The Maxwell-Boltzmann distribution ρ^* for the discrete ideal gas is the measure at which the relative entropy $R(\theta|\rho)$ attains its minimum over $\theta \in \mathcal{P}_r$ satisfying $\langle \theta\rangle \in F_{z,a}$; in symbols

$$
R(\rho^*|\rho) = \min\{R(\theta|\rho) : \theta \in \mathcal{P}_r, \langle \theta\rangle \in F_{z,a}\}.
$$

To show that the Maxwell-Boltzmann distribution is well defined, one must show that there exists a unique probability measure $\rho^* \in \mathcal{P}_r$ satisfying the constrained minimization problem in the last display. That this is the case is part of the content of the following theorem. It gives the form of the Maxwell-Boltzmann distribution, which we denote by $\rho^{(\beta)}$. This distribution is parametrized by $\beta = \beta(z) \in \mathbb{R}$, which is conjugate to the quantity z parametrizing the microcanonical ensemble. The parameter β is identified with the inverse temperature $1/T$. According to part (c) of the theorem, $\beta(z) > 0$ or $T > 0$ corresponds to $z \in (y_1, \bar{y})$, $\beta = 0$ or $T = \infty$ corresponds to $z = \bar{y}$, and $\beta(z) < 0$ corresponds to $z \in (\bar{y}, y_r)$. It follows that values of z lying in the low energy interval $(y_1, \bar{y})$ correspond to the physically relevant region of positive temperatures while values of z lying in the high energy interval $[\bar{y}, y_r)$ correspond to the physically irrelevant region of negative temperatures or infinite temperature.

Theorem 4.1. *Let $z \in (y_1, y_r)$ be given, and define the closed interval $F_{z,a}$ in (2.3). The following conclusions hold.*

(a) There exists a unique $\rho^{(\beta)} \in \mathcal{P}_r$ satisfying

$$R(\rho^{(\beta)}|\rho) = \min\{R(\theta|\rho) \,:\, \theta \in \mathcal{P}_r, \langle\theta\rangle \in F_{z,a}\}. \tag{4.4}$$

(b) The components of $\rho^{(\beta)}$ have the form for $j = 1, 2, \ldots, r$

$$\rho_j^{(\beta)} = \frac{1}{\sum_{k=1}^r \exp[-\beta y_k]\rho_k} \cdot \exp[-\beta y_j]\rho_j, \tag{4.5}$$

where $\rho_j = 1/r$ for each j. The quantity $\beta = \beta(z)$ is the unique value of β for which

$$\langle\rho^{(\beta)}\rangle = \sum_{j=1}^r y_j \rho_j^{(\beta)} = z.$$

(c) $\beta(z) > 0$ corresponds to $z \in (y_1, \bar{y})$, $\beta(z) = 0$ corresponds to $z = \bar{y}$, and $\beta < 0$ corresponds to $z \in (\bar{y}, y_r)$.

The proof of this theorem shows the close connection between the Maxwell-Boltzmann distribution and Cramér's LDP in Theorem 2.1. We recall that the rate function in Cramér's LDP is defined for $x \in \mathbb{R}$ by

$$I(x) = \sup_{t \in \mathbb{R}}\{tx - c(t)\}.$$

In this formula $c(t)$ is the cumulant generating function of ρ defined in (2.5). The connection with the Maxwell-Boltzmann distribution arises if we calculate the derivative of $c(t)$, obtaining

$$c'(t) = \frac{1}{\sum_{k=1}^r \exp[ty_k]\rho_k} \cdot \sum_{j=1}^r y_j \exp[ty_j]\rho_j,$$

where $\rho_j = 1/r$ for each j. For $t \in \mathbb{R}$ let $\rho^{(t)}$ denote the probability measure on Λ having the same components as $\rho^{(\beta)}$ in (4.5) with $-\beta$ replaced by t. It follows that

$$c'(t) = \sum_{j=1}^{r} y_j \rho_j^{(t)} = \langle \rho^{(t)} \rangle.$$

In Lemma 6.4.2 in [11] it is proved that $c(t)$ has the following properties:

1. $c''(t) > 0$ for all t; i.e., c is strictly convex on $\mathbb{R}$, and $c'(t)$ is strictly increasing for $t \in \mathbb{R}$.

2. $c'(0) = \sum_{k=1}^{r} y_k \rho_k = \bar{y}$.

3. $c'(t) \to y_1$ as $t \to -\infty$ and $c'(t) \to y_r$ as $t \to \infty$.

4. The range of $c'(t)$ for $t \in \mathbb{R}$ is the open interval (y_1, y_r), which is the interior of the smallest interval containing the support $\{y_1, y_2, \ldots, y_r\}$ of ρ.

It follows from these properties that $c'(t)$ is a strictly increasing function of $t \in \mathbb{R}$ and thus defines a bijection of $\mathbb{R}$ onto (y_1, y_r). In particular, given $z \in (y_1, y_r)$ there exists a unique value of $t = t(z)$ for which $c'(t) = \langle \rho^{(t)} \rangle = z$. Defining $\beta = \beta(z) = -t(z)$, we obtain part (b) of Theorem 4.1. Part (c) is a consequence of these mapping properties of c' and the fact that $c'(0) = \bar{y}$.

The proof of part (a) of Theorem 4.1 is an object of great beauty, using only properties of the relative entropy.

Proof of part (a) of Theorem 4.1. Given $z \in (y_1, y_r)$ define

$$\Gamma(z, a) = \{\theta \in \mathcal{P}_r : \langle \theta \rangle = \sum_{j=1}^{r} y_j \theta_j \in F_{z,a}\}$$

The proof is more elegant if we work with the measure $\rho^{(t)}$ having the same components as $\rho^{(\beta)}$ in (4.5) with $-\beta$ replaced by t. For each $j \in \{1, \ldots, r\}$

$$\frac{\rho_j^{(t)}}{\rho_j} = \frac{1}{\sum_{k=1}^{r} \exp[ty_k] \rho_k} \cdot \exp[ty_j] = \frac{1}{\exp[c(t)]} \cdot \exp[ty_j].$$

Our goal is to prove that $R(\theta|\rho)$ attains its infimum over $\Gamma(z, a)$ at the unique measure $\rho^{(t)}$, where $t = t(z)$ is the unique value of t for which $\langle \rho^{(t)} \rangle = z$.

We first consider the case when $z = \bar{y}$. Then $t(z) = -\beta(\bar{y}) = 0$, $F_{z,a} = [z - a, z + a]$, and the infimum in (4.4) is attained at the unique measure $\theta = \rho$ [Thm. 2.2(b)], which coincides with $\rho^{(t)}$ with $t(z) = 0$. This completes the proof when $z = \bar{y}$.

We now consider the case when $z \in (y_1, \bar{y})$. The proof for $z \in (\bar{y}, y_r)$ is carried out similarly. For any $\theta \in \Gamma(z, a)$

$$
\begin{aligned}
R(\theta|\rho) &= \sum_{j=1}^{r} \theta_j \log \frac{\theta_j}{\rho_j} = \sum_{j=1}^{r} \theta_j \log \frac{\theta_j}{\rho_j^{(t)}} + \sum_{j=1}^{r} \theta_j \log \frac{\rho_j^{(t)}}{\rho_j} \\
&= R(\theta|\rho^{(t)}) + t \sum_{j=1}^{r} y_j \theta_j - c(t) = R(\theta|\rho^{(t)}) + t\langle\theta\rangle - c(t).
\end{aligned}
$$

Since $R(\rho^{(t)}|\rho^{(t)}) = 0$ and $\langle\rho^{(t)}\rangle = z$, it follows that

$$
R(\rho^{(t)}|\rho) = R(\rho^{(t)}|\rho^{(t)}) + t\langle\rho^{(t)}\rangle - c(t) = tz - c(t). \tag{4.6}
$$

Now consider any $\theta \in \Gamma(z, a)$ satisfying $\theta \neq \rho^{(t)}$. Since $z \in (y_1, \bar{y})$, we have $t < 0$, $\langle\theta\rangle \leq z$, and $R(\theta|\rho^{(t)}) \geq 0$ with equality if and only if $\theta = \rho^{(t)}$ [Thm. 2.2(b)]. Hence

$$
\begin{aligned}
R(\theta|\rho) &= R(\theta|\rho^{(t)}) + t\langle\theta\rangle - c(t) \\
&> t\langle\theta\rangle - c(t) \geq tz - c(t) = R(\rho^{(t)}|\rho).
\end{aligned}
$$

We conclude that for any $\theta \in \Gamma(z, a)$, $R(\theta|\rho) \geq R(\rho^{(t)}|\rho)$ with equality if and only if $\theta = \rho^{(t)}$. Thus $R(\theta|\rho)$ attains its infimum over $\Gamma(z, a)$ at the unique measure $\rho^{(t)}$, equivalently at the unique measure $\rho^{(\beta)}$ with $\beta = -t$. This completes the proof of part (a) of Theorem 4.1. ∎

The proof of part (a) of Theorem 4.1 reveals another unexpected connection between the Maxwell-Boltzmann distribution and Cramér's Theorem. For $z \in (y_1, y_r)$ we show in the proof of part (a) of Theorem 4.1 that $R(\theta|\rho)$ attains its infimum for $\theta \in \mathcal{P}_r$ satisfying $\langle\theta\rangle \in F_{z,a}$ at the unique measure $\rho^{(t)}$, where $t = t(z)$ is the unique value of t for which $c'(t) = \langle\rho^{(t)}\rangle = z$. The measure $\rho^{(t)}$ has the same components as $\rho^{(\beta)}$ in (4.5) with $-\beta$ replaced by t. Combining this result with (4.6), we obtain

$$
\min\{R(\theta|\rho) : \theta \in \mathcal{P}_r, \langle\theta\rangle \in F_{z,a}\} = R(\rho^{(t(z))}|\rho) = t(z)z - c(t(z)).
$$

From Theorem 2.1 we recall that the rate function in Cramér's Theorem is defined by $I(z) = \sup_{t \in \mathbb{R}}\{tz - c(t)\}$. Since c is a strictly convex function on $\mathbb{R}$, the supremum in this definition is obtained at the unique value $t = t(z)$ satisfying $c'(t) = z$. It follows from the last display that

$$
\min\{R(\theta|\rho) : \theta \in \mathcal{P}_r, \langle\theta\rangle \in F_{z,a}\} = t(z)z - c(t(z)) = I(z).
$$

28

This result relating the level-2 rate function $R(\cdot|\rho)$ and the level-1 rate function I is known a **contraction principle**. It is usually stated for $a = 0$, taking the form

$$\min\{R(\theta|\rho) : \theta \in \mathcal{P}_r, \langle \theta \rangle = z\} = I(z).$$

A general formulation of the contraction principle is stated in [9, Thm. 1.3.2].

This completes our discussion of Boltzmann's formulation of the Maxwell-Boltzmann distribution. Our next topic is the rigorous formulation of this distribution, which is based on the Boltzmann-Sanov LDP in Theorem 2.3.

4b   Rigorous formulation of Maxwell-Boltzmann distribution

The rigorous formulation of the Maxwell-Boltzmann distribution $\rho^{(\beta)}$ is proved in part (a) of the next theorem. There we show that $\rho^{(\beta)}$ is the equilibrium distribution of the random energy values X_i of the discrete ideal gas in the following sense: in the limit $n \to \infty$, $\rho^{(\beta)}$ is the limiting marginal distribution of X_i with respect to the microcanonical ensemble $P_n\{\cdot|S_n/n \in F_{z,a}\}$. Since the random variables X_i are identically distributed, it suffices to prove this statement for $i = 1$. In part (b) we state the surprising result that for any $r \in \mathbb{N}$ satisfying $r \geq 2$ the finite product measure on Λ^r with equal one dimensional marginals $\rho^{(\beta)}$ is the limiting marginal distribution of $X_1, X_2, \ldots, X_r$ with respect to the microcanonical ensemble as $n \to \infty$. This finite product measure is called the **canonical ensemble**.

To see why part (b) of the next theorem is a surprise, we recall that the random variables $X_1, X_2, \ldots, X_r$, though independent with respect to the product measure P_n, are dependent with respect to the microcanonical ensemble because of the conditioning on $S_n/n \in F_{z,a}$. Part (b) shows that as $n \to \infty$, these random variables recover their independence, a property known as **propagation of chaos**. The result in part (b) shows that in the thermodynamic limit $n \to \infty$ the microcanonical ensemble and the canonical ensemble are equivalent. This is a simple example of the important topic of **equivalence of ensembles** in statistical mechanical models [11, §6.7], [13].

Theorem 4.2. *Let $z \in (y_1, y_r)$ be given, and define the closed interval $F_{z,a}$ in (2.3). Let $\rho^{(\beta)} = \rho^{(\beta(z))}$ denote the corresponding Maxwell-Boltzmann distribution having coordinates given in (4.5). Also let $P_r^{(\beta)} = P_r^{\beta(z)}$ denote the finite product measure on Λ^r with equal one-dimensional marginals $\rho^{(\beta)}$. The following conclusions hold.*
    (a) *For any $j = 1, 2, \ldots, r$*

$$\lim_{n\to\infty} P_n\{X_1 = y_j|S_n/n \in F_{z,a}\} = \rho_j^{(\beta)}.$$

(b) *For any $r \in \mathbb{N}$ satisfying $r \geq 2$ and any $y_{j_i} \in \Lambda$ for $i = 1, 2, \ldots, r$*

$$\lim_{n \to \infty} P_n\{X_i = y_{j_i}, i = 1, 2, \ldots, r | S_n/n \in F_{z,a}\}$$

$$= \prod_{i=1}^{r} \rho_{j_i}^{(\beta)} = P_r^{(\beta)}\{X_i = y_{j_i}, i = 1, 2, \ldots, r\}.$$

A complete proof of part (a) is given in Theorem 6.4.1 in [11]. We motivate part (a) of Theorem 4.2 by a heuristic argument. Let us assume that we can prove that for any $\varepsilon > 0$

$$\lim_{n \to \infty} P_n\{L_n \in B(\rho^{(\beta)}, \varepsilon) | L_n \in \Gamma(z, a)\} = 1, \tag{4.7}$$

where $B(\rho^{(\beta)}, \varepsilon)$ denotes the ball with center $\rho^{(\beta)}$ and radius ε and

$$\Gamma(z, a) = \{\theta \in \mathcal{P}_r : \langle \theta \rangle \in F_{z,a}\}.$$

We first motivate that the limit in (4.7) yields the limit in part (a) of Theorem 4.2. As noted in (4.3), for any $\omega \in \Omega_n$ the sample mean $S_n(\omega)/n$ equals the mean of the empirical measure $L_n(\omega)$; in symbols, $S_n(\omega)/n = \langle L_n(\omega) \rangle$. Hence the limit in (4.7) is equivalent to the limit

$$\lim_{n \to \infty} P_n\{L_n \in B(\rho^{(\beta)}, \varepsilon) | S_n/n \in F_{z,a}\} = 1.$$

Take any $j = 1, 2, \ldots, r$. Given this limit, it is reasonable to expect that for all large n we have with probability close to 1

$$\begin{aligned}
\rho_j^{(\beta)} &= E^{P_n}\{\rho_j^{(\beta)} \mid S_n/n \in F_{z,a}\} \approx E^{P_n}\{L_n(y_j) \mid S_n/n \in F_{z,a}\} \\
&= \frac{1}{n} \sum_{i=1}^{n} E^{P_n}\{\delta_{X_i}(y_j) \mid S_n/n \in F_{z,a}\} \\
&= \frac{1}{n} \sum_{i=1}^{n} P_n\{X_i = y_j \mid S_n/n \in F_{z,a}\} \\
&= P_n\{X_1 = y_j \mid S_n/n \in F_{z,a}\}.
\end{aligned}$$

The last line follows from the fact that the random variables X_i all have the same distribution. This completes the motivation of part (a) of Theorem 4.2 from the limit (4.7).

In order to motivate the limit (4.7), we use the formal global asymptotic result (3.5) for Borel subsets Γ of $\mathcal{P}_r$. If we apply this to the conditional probability on the left side of (4.7), then we obtain

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

30

Thus one should obtain the conditioned limit (4.7) if

$$R(B(\rho^{(\beta)}, \varepsilon) \cap \Gamma(z, a)|\rho) = R(\Gamma(z, a)|\rho).$$

The last equation is in fact valid since $\rho^{(\beta)}$ is the unique measure at which $R(\theta|\rho)$ attains its minimum over all $\theta \in \mathcal{P}_r$ satisfying $\langle\theta\rangle \in F_{z,a}$; equivalently, over all $\theta \in \mathcal{P}_r$ satisfying $\theta \in \Gamma(z, a)$. The latter property is proved in part (a) of Theorem 4.1. This completes the motivation of part (a) of Theorem 4.2.

The limit in part (a) of Theorem 4.2 is proved in Theorem 6.4.1 in [11] as a consequence of the limit (4.7), which in turn is proved by applying two results in the present paper: part (a) of Theorem 4.1 and the Boltzmann-Sanov LDP in Theorem 2.3. Together these results show that when conditioned on $L_n \in \Gamma(z, a)$, the conditional probability that L_n lies in the complement $[B(\rho^{(\beta)}, \varepsilon)]^c$ of the open ball $B(\rho^{(\beta)}, \varepsilon)$ converges to 0; in symbols

$$\lim_{n\to\infty} P_n\{L_n \in [B(\rho^{(\beta)}, \varepsilon)]^c | L_n \in \Gamma(z, a)\} = 0. \tag{4.8}$$

The limit (4.7) now follows. Because of the limit (4.7) it is reasonable to call $\rho^{(\beta)}$ the equilibrium distribution of the empirical vectors L_n with respect to the microcanonical ensemble.

The proof of part (b) of Theorem 4.2 follows the same pattern of proof as the proof of part (a) of Theorem 4.2 except that the empirical vector L_n is replaced by a more general random quantity, and the Boltzmann-Sanov Theorem for L_n is replaced by the LDP for this more general random quantity. Because part (b) of Theorem 4.2 is not used elsewhere in this paper, the proof is omitted.

We have now completed the discussion of the discrete ideal gas. In the next section we introduce the much more complicated droplet model, the analysis of which was inspired by the analysis of the discrete ideal gas.

5   Description of Droplet Model

In this section we introduce the droplet model and summarize the results obtained in [16] on the asymptotic behavior of the model. This includes the derivation of the equilibrium distribution of dependent random variables that count the droplet sizes. This equilibrium distribution coincides with the equilibrium distribution of random probability measures, called number-density measure, which are the empirical measures of the dependent droplet-size random variables. As we explain, this equilibrium distribution is derived by first proving an LDP for the number-density measures, which is an analogue of the Boltzmann-Sanov LDP used in the preceding section to derive the Maxwell-Boltzmann equilibrium distribution for the discrete ideal gas.

31

The definition of the droplet model depends on a nonnegative integer b and a parameter $c \in (b, \infty)$. K distinguishable particles are placed, each with equal probability $1/N$, onto the N sites of the lattice $\Lambda_N = \{1, 2, \ldots, N\}$. The large deviation limit — or in statistical mechanical terminology, the thermodynamic limit — is defined by taking $K \to \infty$ and $N \to \infty$ with K/N equal to c. The ratio K/N equals the average number of particles per site or the average size of a droplet. The question that motivated our research is natural and is simply stated. Given that each site is occupied by a minimum of b particles, what is the equilibrium distribution of the number of particles per site in the thermodynamic limit? As we explain in subsection 6c, this equilibrium distribution is a Poisson distribution $\rho_{b,\alpha_b(c)}$ restricted to $\mathbb{N}_b = \{n \in \mathbb{Z} : n \geq b\}$, where the parameter $\alpha_b(c)$ is chosen so that the mean of $\rho_{b,\alpha_b(c)}$ equals c.

In order to determine the form of the equilibrium distribution, we introduce a standard probabilistic model. The configuration space is the set $\Omega_N = \Lambda_N^K$ consisting of all $\omega = (\omega_1, \omega_2, \ldots, \omega_K)$, where ω_i denotes the site in Λ_N occupied by the i'th particle. The cardinality of Ω_N equals N^K. Because of the description of the droplet model to this point, it is consistent to introduce the uniform probability measure P_N that assigns equal probability $1/N^K$ to each of the N^K configurations $\omega \in \Omega_N$. For subsets A of Ω_N, $P_{N,b,m}(A) = \text{card}(A)/N^K$, where card denotes cardinality.

We next define the following two random variables, which are functions of the configuration $\omega \in \Omega_N$: for $\ell \in \Lambda_N$, $K_\ell(\omega)$ denotes the number of particles occupying the site ℓ in the configuration ω; for $j \in \mathbb{N} \cup \{0\}$, $N_j(\omega)$ denotes the number of sites $\ell \in \Lambda_N$ for which $K_\ell(\omega) = j$. In (6.2) we summarize the LDP for a sequence of random probability measures defined in terms of these random variables.

We now specify how the probabilistic model incorporates the nonnegative integer b, first considering the case where b is a positive integer. The case where $b = 0$ is discussed later. Given a positive integer b, we focus on the subset of Ω_N consisting of all configurations ω for which every site of Λ_N is occupied by at least b particles. Because of this requirement $N_j(\omega)$ is indexed by $j \in \mathbb{N}_b = \{n \in \mathbb{Z} : n \geq b\}$. It is useful to think of each particle as having one unit of mass and of the set of particles at each site ℓ as defining a droplet. With this interpretation, for each configuration ω, $K_\ell(\omega)$ denotes the mass or size of the droplet at site ℓ. The j'th droplet class has $N_j(\omega)$ droplets and mass $jN_j(\omega)$. Because the number of sites in Λ_N equals N and the sum of the masses of all the droplet classes equals K, the following conservation laws hold for such configurations:

$$\sum_{j \in \mathbb{N}_b} N_j(\omega) = N \ \text{ and } \ \sum_{j \in \mathbb{N}_b} jN_j(\omega) = K. \tag{5.1}$$

Because of these additive constraints, the random variables N_j are dependent.

In order to carry out the asymptotic analysis of the droplet model, we introduce a quantity $m = m(N)$ that converges to ∞ sufficiently slowly with respect to N; specifically, we require

that $m(N)^2/N \to 0$ as $N \to \infty$. In terms of b and m we define the subset $\Omega_{N,b,m}$ of Ω_N consisting of all configurations ω for which every site of Λ_N is occupied by at least b particles and at most $m(N)$ of the quantities $N_j(\omega)$ are positive. This second constraint, which restricts the number of positive components $N_j(\omega)$, is a useful technical device that allows us to control the errors in several estimates. In appendix D of [15] we present evidence supporting the conjecture that this restriction can be eliminated.

When $b = 0$, the constraint that every site of Λ_N is occupied by at least b particles disappears because we allow sites to be occupied by 0 particles and thus remain empty. Therefore, when $b = 0$, $N_j(\omega)$ is indexed by $j \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. However, in the definition of $\Omega_{N,0,m}$ we retain the constraint that at most $m(N)$ of the quantities $N_j(\omega)$ are positive. Because the choice $b = 0$ allows sites to be empty, we lose the interpretation of the set of particles at each site as being a droplet. However, for $\Omega_{N,0,n}$ the two conservation laws (5.1) continue to hold.

For the remainder of this paper, we work with any fixed, nonnegative integer b. The probability measure $P_{N,b,m}$ defining the droplet model is obtained by restricting the uniform measure P_N to the set of configurations $\Omega_{N,b,m}$. Thus $P_{N,b,m}$ equals the conditional probability $P_N(\cdot \,|\, \Omega_{N,b,m})$. For subsets A of $\Omega_{N,b,m}$, $P_{N,b,m}(A)$ takes the form

$$
\begin{aligned}
P_{N,b,m}(A) &= P_N(A \,|\, \Omega_{N,b,m}) = \frac{1}{P_N(\Omega_{N,b,m})} \cdot P_N(A) &\text{(5.2)} \\
&= \frac{1}{\text{card}(\Omega_{N,b,m})} \cdot \text{card}(A).
\end{aligned}
$$

The second line of this formula follows from the fact that P_N assigns equal probability $1/N^K$ to every $\omega \in \Omega_{N,b,m}$. $P_{N,b,m}$ defines the **microcanonical ensemble**, which incorporates the conservation laws for number and mass expressed in (5.1).

Having defined the droplet model, we introduce the random probability measures whose large deviations we will study. For $\omega \in \Omega_{N,b,m}$ these measures are the **number-density measures** $\Theta_{N,b}$ that assign to $j \in \mathbb{N}_b$ the probability $N_j(\omega)/N$. Thus for any subset A of $\mathbb{N}_b$

$$
\Theta_{N,b}(\omega, A) = \sum_{j \in \mathbb{N}_b} \Theta_{N,b;j}(\omega)\delta_j(A) = \sum_{j \in A} \Theta_{N,b;j}(\omega), \text{ where } \Theta_{N,b;j}(\omega) = \frac{N_j(\omega)}{N}.
$$

Because of the two conservation laws in (5.1) and because $K/N = c$, for $\omega \in \Omega_{N,b,m}$

$$
\sum_{j \in \mathbb{N}_b} \Theta_{N,b;j}(\omega) = 1 \text{ and } \sum_{j \in \mathbb{N}_b} j\Theta_{N,b;j}(\omega) = \frac{1}{N}\sum_{j \in \mathbb{N}_b} jN_j(\omega) = \frac{K}{N} = c. \qquad \text{(5.3)}
$$

Thus for $\omega \in \Omega_{N,b,m}$, $\Theta_{N,b}(\omega)$ is a probability measure on $\mathbb{N}_b$ having mean c.

There are obvious analogies as well as differences between the droplet model and the discrete ideal gas. The probability measure P_N for the droplet model, which assigns equal probability $1/N^K$ to each of the N^K configurations $\omega \in \Omega_N = \Lambda_N^K$, corresponds to the probability measure P_n for the discrete ideal gas, which assigns equal probability $1/r^n$ to each of the r^n configurations $\omega \in \Omega_n = \Lambda^n$. However, in the droplet model both K, the number of particles, and N, the number of sites, tend to ∞, in contrast to the discrete ideal gas, where the number of possible energy values y_j is a fixed number r. While the microcanonical ensemble $P_{N,b,m}$ for the droplet model is an analog of the microcanonical ensemble for the discrete ideal gas, $P_{N,b,m}$ incorporates more information corresponding to the more complicated definition of the droplet model.

Because of the analogies between the two models, it is important to point out that the number density measures $\Theta_{N,b}$ are the empirical measures of the random variables K_ℓ, which count the droplet sizes at the sites $\ell \in \Lambda_N$; i.e., for $\omega \in \Omega_{N,b,m}$, $\Theta_{N,b}(\omega)$ assigns to subsets A of $\mathbb{N}_b$ the probability

$$\Theta_{N,b}(\omega, A) = \frac{1}{N} \sum_{\ell=1}^{N} \delta_{K_\ell(\omega)}\{A\}.$$

This characterization of $\Theta_{N,b}$ follows from the fact that the empirical measure of K_ℓ assigns to $j \in \mathbb{N}_b$ the probability

$$\frac{1}{N} \sum_{\ell=1}^{N} \delta_{K_\ell(\omega)}(\{j\}) = \frac{1}{N} \cdot \text{card}\{\ell \in \Lambda_N : K_\ell(\omega) = j\} = \frac{N_j(\omega)}{N} = \Theta_{N,b;j}(\omega). \tag{5.4}$$

Thus $\Theta_{N,b}$ is an obvious analog of the empirical vectors L_n for the discrete ideal gas, which are the empirical measures of the random variables X_i. This analogy, however, is deceptive because of a fundamental difference between $\Theta_{N,b}$ and L_n. The empirical vector L_n is the empirical measure of the random variables X_i, which are independent and identically distributed. On the other hand, the empirical measure $\Theta_{N,b}$ is the empirical measure of the random variables K_ℓ, which denote the number of particles occupying the sites $\ell \in \Lambda_N$. Hence for all $\omega \in \Omega_{N,K,m}$, $\sum_{\ell \in \Lambda_N} K_\ell(\omega) = K$. Although they are identically distributed, the random variables K_ℓ are not independent because of this equality constraint. To give an extreme example, if site 1 is occupied by the maximum number of particles $K - (N-1)b$, then each of the $N-1$ other sites must be occupied by the minimum of b particles. This lack of independence makes the asymptotic analysis of $\Theta_{N,b}$ much more complicated than that of the empirical vector L_n.

In the next section we discuss the LDP for the number-density measures and describe the equilibrium distribution of these measures and of the droplet-size random variables K_ℓ.

6 Asymptotic Analysis of the Droplet Model

The asymptotic analysis of the droplet model involves the LDP for the number-density measures $\Theta_{N,b}$, which is summarized in (6.2), and the description of the equilibrium distribution, which is summarized in (6.4) both for the number-density measures $\Theta_{N,b}$ and for the droplet-size random variables K_ℓ. Our proof of the LDP is outlined in section 8.

6a Preliminaries

We first introduce two sets of probability measures that arise in this asymptotic analysis. $\mathcal{P}_{\mathbb{N}_b}$ denotes the set of probability measures on $\mathbb{N}_b = \{n \in \mathbb{Z} : n \geq b\}$. Thus $\theta \in \mathcal{P}_{\mathbb{N}_b}$ has the form $\sum_{j \in \mathbb{N}_b} \theta_j \delta_j$, where the components θ_j satisfy $\theta_j \geq 0$ and $\theta\{\mathbb{N}_b\} = \sum_{j \in \mathbb{N}_b} \theta_j = 1$. We say that a sequence of measures $\{\theta^{(n)}, n \in \mathbb{N}\}$ in $\mathcal{P}_{\mathbb{N}_b}$ converges weakly to $\theta \in \mathcal{P}_{\mathbb{N}_b}$, and write $\theta^{(n)} \Rightarrow \theta$, if for any bounded function f mapping $\mathbb{N}_b$ into $\mathbb{R}$

$$\lim_{n \to \infty} \int_{\mathbb{N}_b} f d\theta^{(n)} = \int_{\mathbb{N}_b} f d\theta.$$

$\mathcal{P}_{\mathbb{N}_b}$ is topologized by the topology of weak convergence. There is a standard technique for introducing a metric structure on $\mathcal{P}_{\mathbb{N}_b}$ in terms of a metric known as the Prohorov metric and denoted by π. This metric has the following two properties:

- Convergence with respect to the Prohorov metric is equivalent to weak convergence [18, Thm. 3.3.1]; i.e., $\theta^{(n)} \Rightarrow \theta$ if and only if $\pi(\theta^{(n)}, \theta) \to 0$ as $N \to \infty$.

- With respect to the Prohorov metric, $\mathcal{P}_{\mathbb{N}_b}$ is a complete, separable metric space [18, Thm. 3.1.7].

We denote by $\mathcal{P}_{\mathbb{N}_b,c}$ the set of measures in $\mathcal{P}_{\mathbb{N}_b}$ having mean c. Thus $\theta \in \mathcal{P}_{\mathbb{N}_b,c}$ has the form $\sum_{j \in \mathbb{N}_b} \theta_j \delta_j$, where the components θ_j satisfy $\theta_j \geq 0$, $\sum_{j \in \mathbb{N}_b} \theta_j = 1$, and $\int_{\mathbb{N}} x \theta(dx) = \sum_{j \in \mathbb{N}_b} j \theta_j = c$. By (5.3) the number-density measures $\Theta_{N,b}$ take values in this set. Not only is $\mathcal{P}_{\mathbb{N}_b,c}$ the smallest convex set containing the range of $\Theta_{N,b}$ for all $N \in \mathbb{N}$, but also the union over $N \in \mathbb{N}$ of the range of $\Theta_{N,b}$ is dense in $\mathcal{P}_{\mathbb{N}_b,c}$. Hence $\mathcal{P}_{\mathbb{N}_b,c}$ is the most natural space in which to formulate the LDP for these measures.

A natural question is to determine two equilibrium distributions of the droplet model: the equilibrium distribution $\rho^\star$ of the number-density measures and the equilibrium distribution $\rho^{\star\star} = \sum_{j \in \mathbb{N}_b} \rho_j^{\star\star} \delta_j$ of the droplet-size random variables K_ℓ. These distributions are defined by the following two limits: for any $\varepsilon > 0$, any $\ell \in \Lambda_N$, and all $j \in \mathbb{N}_b$

$$\lim_{N \to \infty} P_{N,b,m}\{\Theta_{N,b} \in B_\pi(\rho^*, \varepsilon)\} = 1 \text{ and } \lim_{N \to \infty} P_{N,b,m}\{K_\ell = j\} = \rho_j^{\star\star}.$$

$B_\pi(\rho^*, \varepsilon))$ denotes the open ball with center ρ^* and radius ε in $\mathcal{P}_{\mathbb{N}_b,c}$ defined with respect to the Prohorov metric π; in symbols

$$B_\pi(\rho^*, \varepsilon)) = \{\theta \in \mathcal{P}_{\mathbb{N}_b,c} : \pi(\rho^*, \theta) < \varepsilon\}.$$

The two limits in the last display are analogs of the limits defining the equilibrium distribution of the discrete ideal gas, the first corresponding to (4.7), which defines the equilibrium distribution of the empirical vectors L_n, and the second corresponding to the limit in part (a) of Theorem 4.2, which defines the equilibrium distribution of X_1. As we show in (4.8) and in part (a) of Theorem 4.2, the equilibrium distributions of L_n and of X_1 coincide and are equal to the Maxwell-Boltzmann distribution $\rho^{(\beta)}$. In part (a) of Theorem 4.1, the Maxwell-Boltzmann distribution is characterized as the unique measure at which the relative entropy $R(\theta|\rho)$ attains its minimum for all $\theta \in \mathcal{P}_r$ satisfying $\langle\theta\rangle \in F_{z,a}$.

Because of the analogy with the discrete ideal gas, the following observations concerning the equilibrium distributions for the droplet model should not be surprising.

1. The equilibrium distributions of $\Theta_{N,b}$ and K_ℓ coincide.

2. We first determine the equilibrium distribution ρ^* of $\Theta_{N,b}$ and then prove that ρ^* is also the equilibrium distribution of K_ℓ.

3. As in many statistical mechanical models, an efficient way to determine the equilibrium distribution of $\Theta_{N,b}$ is to prove an LDP for $\Theta_{N,b}$, which is summarized in (6.2).

4. The equilibrium distribution ρ^* is characterized as the unique measure at which the relative entropy $R(\theta|\rho_{b,\alpha_b(c)})$ attains its minimum for all $\theta \in \mathcal{P}_{\mathbb{N}_b,c}$.

We first describe the LDP for the number-density measures $\Theta_{N,b}$.

6b   LDP for number density measures $\Theta_{N,b}$

The LDP for $\Theta_{N,b}$ is proved in Theorem 2.1 in [16]. The proof is self-contained and starts from first principles, using techniques that are familiar in statistical mechanics. The centerpiece of the proof is the local large deviation estimate in Theorem 3.1 in [16], the proof of which uses combinatorics, Stirlings formula, and Laplace asymptotics.

The content of Theorem 2.1 in [16] is the following: as $N \to \infty$ the sequence of number-density measures $\Theta_{N,b}$ satisfies the LDP on $\mathcal{P}_{\mathbb{N}_b,c}$ with respect to the measures $P_{N,b,m}$. Thus there is a lower large deviation bound for $P_{N,b,m}\{\Theta_{N,b} \in G\}$, where G is an open subset of $\mathcal{P}_{\mathbb{N}_b,c}$, and there is an upper large deviation bound for $P_{N,b,m}\{\Theta_{N,b} \in F\}$, where F is a closed

subset of $\mathcal{P}_{\mathbb{N}_b,c}$. The rate function in this LDP is the relative entropy with respect to a Poisson distribution $\rho_{b,\alpha}$ restricted to $\mathbb{N}_b = \{j \in \mathbb{Z} : j \geq b\}$ and having components

$$\rho_{b,\alpha;j} = \frac{1}{Z_b(\alpha)} \cdot \frac{\alpha^j}{j!} \text{ for } j \in \mathbb{N}_b. \tag{6.1}$$

In this formula α equals the quantity $\alpha_b(c)$ defined in the next paragraph. In addition $Z_b(\alpha)$ is the normalization that makes $\rho_{b,\alpha}$ a probability measure. Thus we have $Z_0(\alpha) = e^\alpha$ while for $b \in \mathbb{N}$

$$Z_b(\alpha) = e^\alpha - \sum_{j=0}^{b-1} \alpha^j/j!.$$

For $\theta = \sum_{j\in\mathbb{N}_b} \theta_j \delta_j \in \mathcal{P}_{\mathbb{N}_b,c}$, the relative entropy of θ with respect to $\rho_{b,\alpha_b(c)}$ is defined by

$$R(\theta|\rho_{b,\alpha_b(c)}) = \sum_{j\in\mathbb{N}_b} \theta_j \log(\theta_j/\rho_{b,\alpha_b(c);j}).$$

If $\theta_j = 0$, then $\theta_j \log(\theta_j/\rho_{b,\alpha_b(c);j}) = 0$. For subsets Γ of $\mathcal{P}_{\mathbb{N}_b,c}$ we summarize the LDP for $\Theta_{N,b}$ by the notation

$$P_{N,b,m}(\Theta_{N,b} \in \Gamma) \approx \exp[-NR(\Gamma|\rho_{b,\alpha_b(c)})], \tag{6.2}$$

where $R(\Gamma|\rho_{b,\alpha_b(c)})$ denotes the infimum of $R(\theta|\rho_{b,\alpha_b(c)})$ over $\theta \in \Gamma$.

In (6.1) α equals the quantity $\alpha_b(c)$ having the property that $\rho_{b,\alpha_b(c)}$ has mean c and thus lies in the space $\mathcal{P}_{\mathbb{N}_b,c}$. We first consider $b = 0$. In this case $\rho_{0,\alpha}$ is a standard Poisson distribution on $\mathbb{N}_0$ having mean α. It follows that $\alpha_0(c) = c$ is the unique value for which $\rho_{0,\alpha_0(c)}$ has mean c. We now consider $b \in \mathbb{N}$. In this case $\rho_{b,\alpha}$ is a probability measure on $\mathbb{N}_b$ having mean

$$
\begin{aligned}
\langle \rho_{b,\alpha} \rangle &= \sum_{j\in\mathbb{N}_b} j\rho_{b,\alpha;j} = \frac{1}{Z_b(\alpha)} \cdot \sum_{j\in\mathbb{N}_b} \frac{\alpha^j}{(j-1)!} \\
&= \frac{1}{Z_b(\alpha)} \cdot \alpha \sum_{j=b-1}^{\infty} \frac{\alpha^j}{j!} = \frac{1}{Z_b(\alpha)} \cdot \alpha Z_{b-1}(\alpha).
\end{aligned} \tag{6.3}
$$

Thus $\rho_{b,\alpha}$ has mean c if and only if α satisfies $\gamma_b(\alpha) = c$, where $\gamma_b(\alpha) = \alpha Z_{b-1}(\alpha)/Z_b(\alpha)$. In Theorem A.2 in [16] we prove that $\gamma_b(\alpha) = c$ has a unique solution $\alpha_b(c) \in (0, \infty)$ for all $b \in \mathbb{N}$ and any $c > b$. In contrast to the straightforward proof for $b = 1$ [15, Thm. C.2(a)], the proof for general $b \in \mathbb{N}$ is much more difficult. In Theorem C.2 in [15] we show other properties of $\alpha_b(c)$ including the facts that $\alpha_b(c)$ satisfies the inequalities $c > \alpha_b(c) > c - b$ and thus is asymptotic to c as $c \to \infty$; i.e., $\lim_{c\to\infty} \alpha_b(c)/c = 1$.

For $b \in \mathbb{N}$ the distribution $\rho_{b,\alpha_b(c)}$ differs from a standard Poisson distribution because the former has 0 mass at $0, 1, \ldots, b-1$ while the latter has positive mass at these points. In fact, as shown in part (d) of Theorem C.1 in [15] $\rho_{b,\alpha_b(c)}$ can be identified as the distribution of a Poisson random variable $\Xi_{\alpha_b(c)}$ with parameter $\alpha_b(c)$ conditioned on $\Xi_{\alpha_b(c)} \in \mathbb{N}_b$.

We point out an unavoidable subtlety in the statement of the LDP for $\Theta_{N,b}$ in Theorem 2.1 in [16]. This subtlety arises because $\mathcal{P}_{\mathbb{N}_b,c}$ is not a closed subset of the complete separable metric space $\mathcal{P}_{\mathbb{N}_b}$, necessitating a different form of the large deviation upper bound for compact subsets of $\mathcal{P}_{\mathbb{N}_b,c}$, stated in part (b) of that theorem, and for closed noncompact subsets of $\mathcal{P}_{\mathbb{N}_b,c}$, stated in part (c) of that theorem. The interested reader is referred to section 2 of [16] for a complete discussion. Another consequence of the fact that $\mathcal{P}_{\mathbb{N}_b,c}$ is not a closed subset of $\mathcal{P}_{\mathbb{N}_b}$ is that many of the standard results in the theory of large deviations are not directly applicable, making our self-contained proof based on combinatorics and the local large deviation estimate more attractive.

The discussion concerning the quantity $\alpha_b(c)$ in the third paragraph above points to other significant connections between the discrete ideal gas and the droplet model. In the discrete ideal gas the microcanonical ensemble is parametrized by the energy parameter $z \in (y_1, y_r)$, and the canonical ensemble is parametrized by β, which equals the inverse temperature $1/T$. The canonical ensemble is a finite product measure with one-dimensional marginals equal to the Maxwell-Boltzmann distribution $\rho^{(\beta)}$; the parameter β equals $\beta(z) = -t(z)$, where $t = t(z)$ is the unique solution of $c'(t) = z$, $c(t)$ being the cumulant generating function defined in (2.5). As discussed after the statement of Theorem 4.1, $c'(t)$ is a strictly increasing function of $t \in \mathbb{R}$ and thus a strictly convex function on $\mathbb{R}$. Furthermore, c' defines a bijection of $\mathbb{R}$ onto (y_1, y_r), the open interval in which the energy parameter z lies.

All of these ideas have close analogs in the droplet model. In this model the microcanonical ensemble is parametrized by c, which equals the ratio K/N, the average number of particles per site. The quantity c satisfies $c > b$, where $b \in \mathbb{N} \cup \{0\}$ is the minimum number of particles per site. The equilibrium distribution $\rho_{b,\alpha_b(c)}$ is the analog of the Maxwell-Boltzmann distribution $\rho^{(\beta)}$ in the discrete ideal gas. The distribution $\rho_{b,\alpha_b(c)}$ — and presumably the canonical ensemble, which we did not consider — is parametrized by $\alpha = \alpha_b(c)$, which is the unique solution of $\gamma_b(\alpha) = c$. In the proof in Theorem A.2 in [16] that this equation has a unique solution, we prove that $\gamma_b(\alpha)$ is a strictly increasing function of $\alpha \in (0, \infty)$ and that γ_b defines a bijection of $(0, \infty)$ onto (b, ∞), the open interval in which c lies. Interestingly, in the proof of these properties of $\gamma_b(\alpha)$ there arises a cumulant generating function, the strict convexity of which is used in the proof.

This completes our discussion of the LDP for $\Theta_{N,b}$. Our next topic is the fact that $\rho_{b,\alpha_b(c)}$ is the equilibrium distribution of the droplet model.

6c Equilbrium distribution $\rho_{b,\alpha_b(c)}$

The LDP for $\Theta_{N,b}$ is summarized in (6.2) and is proved in Theorem 2.1 in [16]. As we now explain, this LDP implies that $\rho_{b,\alpha_b(c)}$ is the equilibrium distribution of both $\Theta_{N,b}$ and K_ℓ, satisfying for any $\varepsilon > 0$, any $\ell \in \Lambda_N$, and all $j \in \mathbb{N}_b$

$$\lim_{N \to \infty} P_{N,b,m}\{\Theta_{N,b} \in B_\pi(\rho_{b,\alpha_b(c)}, \varepsilon)\} = 1 \ \text{ and } \ \lim_{N \to \infty} P_{N,b,m}\{K_\ell = j\} = \rho_{b,\alpha_b(c);j}. \qquad (6.4)$$

We sketch how the first of these limits is proved. By applying the large deviation upper bound for $\Theta_{N,b}$ lying in the closed set $[B(\rho_{b,\alpha_b(c)}, \varepsilon)]^c$ and by using the fact that $R(\theta | \rho_{b,\alpha_b(c)})$ attains its infimum of 0 over $\theta \in \mathcal{P}_{\mathbb{N}_b,c}$ at the unique measure $\theta = \rho_{b,\alpha_b(c)}$, we prove that

$$\lim_{N \to \infty} P_{N,b,m}\{\Theta_{N,b} \in [B_\pi(\rho_{b,\alpha_b(c)}, \varepsilon)]^c\} = 0.$$

This limit yields the first limit in (6.4), thus showing that $\rho_{b,\alpha_b(c)}$ is the equilibrium distribution of $\Theta_{N,b}$. Details of this proof are given in Theorem 2.2 in [16]. The proof that $\rho_{b,\alpha_b(c)}$ is also the equilibrium distribution of K_ℓ is proved in Corollary 2.3 in [16] by applying Theorem 2.2. These proofs concerning the equilibrium distribution $\rho_{b,\alpha_b(c)}$ of $\Theta_{N,b}$ and K_ℓ in the droplet model parallel the proofs concerning the Maxwell-Boltzmann equilibrium distribution in the discrete ideal gas although the proofs for the droplet model involve many more technicalities.

The fact that the rate function in the LDP for the empirical measures $\Theta_{N,b}$ is the relative entropy suggests a possible connection between this LDP and Sanov's Theorem. This connection is explored in the next section.

7 Motivating the LDP for $\Theta_{N,b}$ via Sanov's Theorem

For $\omega \in \Omega_N$ the number-density measures $\Theta_{N,b}(\omega)$ take values in the space $\mathcal{P}_{\mathbb{N}_b,c}$ consisting of probability measures on $\mathbb{N}_b$ that have mean c. As we verify in (5.4), $\Theta_{N,b}$ is the empirical measure of the random variables K_ℓ, which count the droplet sizes at the sites $\ell \in \Lambda_N$. The LDP for $\Theta_{N,b}$ is summarized in (6.2). The random variables K_ℓ are identically distributed but are not independent because $\sum_{\ell=1}^{N} K_\ell(\omega) = K$ for each N. In addition, because the distributions of K_ℓ depend on N, these random variables form a triangular array. Hence, although the rate function in the LDP for the empirical measures $\Theta_{N,b}$ is the relative entropy, Sanov's Theorem for i.i.d. random variables cannot be applied as stated to prove this LDP.

Despite this state of affairs, there is a surprise: one can motivate the LDP for $\Theta_{N,b}$ by a calculation based on Sanov's Theorem for i.i.d. random variables. As we now show, this application of Sanov's Theorem yields an LDP having a rate function that is the relative entropy with respect to a Poisson distribution $\rho_{b,c}$ on $\mathbb{N}_b$ having parameter c. Although the mean of

$\rho_{b,c}$ is larger than c for $b \in \mathbb{N}$, as $c \to \infty$ the mean of $\rho_{b,c}$ is asymptotic to c, which is the mean of the Poisson distribution $\rho_{b,\alpha_b(c)}$ appearing in the rate function in the LDP for $\Theta_{N,b}$. It is reasonable to conjecture that as $c \to \infty$ the random variables K_ℓ exhibit an asymptotic independence property that is worth exploring.

In order to motivate the LDP for $\Theta_{N,b}$ we replace K_ℓ by a suitable sequence of independent random variables $\{\overline{K}_\ell, \ell \in \Lambda_N\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $\mathbb{N}_b$. We denote by $\overline{\Theta}_{N,b}$ the empirical measure of these random variables. For $\omega \in \Omega$, $\overline{\Theta}_{N,b}$ assigns to subsets A of $\mathbb{N}_b$ the probability

$$\overline{\Theta}_{N,b}(\omega, A) = \frac{1}{N} \sum_{\ell \in \Lambda_N} \delta_{\overline{K}_\ell(\omega)}\{A\}.$$

$\overline{\Theta}_{N,b}$ takes values in $\mathcal{P}_{\mathbb{N}_b}$, the set of probability measures on $\mathbb{N}_b$.

In order to determine the form of the LDP for $\overline{\Theta}_{N,b}$, we calculate the common distribution of the dependent random variables K_ℓ and then let the independent sequence $\overline{K}_\ell$ have the same common distribution. By the definition of the droplet model K distinguishable particles are placed, each with equal probability $1/N$, onto the N sites of the lattice $\Lambda_N = \{1, 2, \ldots, N\}$. The ratio K/N equals the given parameter c, and if $b \in \mathbb{N}$, then each site is required to be occupied by at least b particles. If $b = 0$, then empty sites are allowed. Because the distribution of K_ℓ depends on N, these random variables form a triangular array having a common distribution that is denoted by $\sigma^{(N)}$.

Let us assume for a moment that the random variables K_ℓ are not a triangular array, but a fixed sequence with common distribution $\rho_{b,c}$. As shown in Theorem 6.2.10 in [8], Sanov's Theorem would then imply that the empirical measures $\overline{\Theta}_{N,b}$, defined in terms of the i.i.d. random variables $\overline{K}_\ell$ having the same common distribution as K_ℓ, would satisfy the LDP on $\mathcal{P}_{\mathbb{N}_b}$ with rate function the relative entropy with respect to $\rho_{b,c}$; in symbols, for subsets Γ of $\mathcal{P}_{\mathbb{N}_b}$

$$P\{\overline{\Theta}_{N,b} \in \Gamma\} \approx \exp[-NR(\Gamma|\rho_{b,c})]. \tag{7.1}$$

One expects that this LDP could be modified in the present situation where we are dealing with triangular arrays. If K_ℓ and thus $\overline{K}_\ell$ have the common distribution $\sigma^{(N)}$, then formally for subsets Γ of $\mathcal{P}_{\mathbb{N}_b}$ the empirical measures $\overline{\Theta}_{N,b}$ would have the asymptotic behavior

$$P\{\overline{\Theta}_{N,b} \in \Gamma\} \approx \exp[-NR(\Gamma|\sigma^{(N)})].$$

If $\sigma^{(N)}$ converges weakly to $\rho_{b,c}$ as $N \to \infty$, then we expect that the LDP in (7.1) would hold.

The LDP conjectured in the preceding paragraph and summarized in (7.1) is true. It is a consequence of Theorem 5 in [1].

As we will see in Theorem 7.1, the N-dependent distributions of K_ℓ converge weakly as $N \to \infty$ to a Poisson distribution $\rho_{b,c}$ restricted to $\mathbb{N}_b$ and having parameter c; $\rho_{b,c}$ has components

$$\rho_{b,c;j} = \frac{1}{Z_b(c)} \cdot \frac{c^j}{j!} \text{ for } j \in \mathbb{N}_b. \tag{7.2}$$

In this formula $Z_b(c)$ is the normalization that makes $\rho_{b,c}$ a probability measure. Thus $Z_0(c) = e^c$ while for $b \in \mathbb{N}$

$$Z_b(c) = e^c - \sum_{j=0}^{b-1} c^j/j!.$$

Theorem 7.1 appears at the end of this section.

The results in this section for the droplet model have a different form for $b = 0$ and for $b \in \mathbb{N}$. We first consider $b = 0$. Because the N-dependent distributions of K_ℓ converge weakly to $\rho_{0,c}$, Theorem 5 in [1] implies the following: as $K \to \infty$ and $N \to \infty$ with $K/N = c$, the empirical measures $\overline{\Theta}_{N,0}$ satisfy the LDP on $\mathcal{P}_{\mathbb{N},0}$ with rate function the relative entropy with respect to $\rho_{0,c}$, which is a probability measure on $\mathbb{N} \cup \{0\}$. This verifies (7.1) for $b = 0$.

For $b = 0$ we now compare the LDP for $\overline{\Theta}_{N,0}$ proved in Theorem 5 in [1] and summarized in (7.1) with the LDP for $\Theta_{N,0}$ proved in Theorem 2.1 in [16] and summarized in (6.2). The rate function in the LDP for $\Theta_{N,0}$ is the relative entropy with respect to the Poisson distribution $\rho_{0,\alpha_0(c)}$ on $\mathbb{N} \cup \{0\}$, the components of which are defined in (6.1) with $\alpha = \alpha_0(c)$. Since $\alpha_0(c)$ has the property that the mean of $\rho_{0,\alpha_0(c)}$ equals c, it follows that $\alpha_0(c) = c$ and thus that $\rho_{0,\alpha_0(c)}$ equals $\rho_{0,c}$ defined in (7.2). Remarkably the rate functions in the LDP (7.1) for $\overline{\Theta}_{N,0}$ and in the LDP (6.2) for $\Theta_{N,0}$ are exactly the same.

We now discuss the results in this section for $b \in \mathbb{N}$. Because the N-dependent distributions of K_ℓ converge weakly to $\rho_{b,c}$, Theorem 5 in [1] implies the following: as $K \to \infty$ and $N \to \infty$ with $K/N = c$, the empirical measures $\overline{\Theta}_{N,b}$ satisfy the LDP on $\mathcal{P}_{\mathbb{N}_b}$ with rate function the relative entropy with respect to $\rho_{b,c}$, which is a probability measure on $\mathbb{N}_b = \{n \in \mathbb{Z} : n \geq b\}$. This verifies (7.1) for $b \in \mathbb{N}$.

For $b \in \mathbb{N}$ we now compare the LDP for $\overline{\Theta}_{N,b}$ proved in Theorem 5 in [1] and summarized in (7.1) with the LDP for $\Theta_{N,b}$ proved in Theorem 2.1 in [16] and summarized in (6.2). The rate function in the LDP for $\Theta_{N,b}$ is the relative entropy with respect to the Poisson distribution $\rho_{b,\alpha_b(c)}$ on $\mathbb{N}_b$, the components of which are defined in (6.1) with $\alpha = \alpha_b(c)$. By the choice of $\alpha_b(c)$, $\rho_{b,\alpha_b(c)}$ has mean c. By contrast the Poisson distribution $\rho_{b,c}$ has mean

$$\langle \rho_{b,c} \rangle = \sum_{j \in \mathbb{N}} j \rho_{b,c;j} = \frac{c Z_{b-1}(c)}{Z_b(c)}.$$

This follows from (6.3) if α is replaced by c. Since $Z_{b-1}(c) > Z_b(c)$, it follows that

$$\langle \rho_{b,c} \rangle > c = \langle \rho_{b,\alpha} \rangle.$$

Thus in contrast to the situation for $b = 0$, for $b \in \mathbb{N}$ the distribution $\rho_{b,c}$ appearing in the rate function in the LDP (7.1) for $\overline{\Theta}_{N,b}$ differs from the distribution $\rho_{b,\alpha_b(c)}$ appearing in the rate function in the LDP (6.2) for $\Theta_{N,b}$.

Despite the fact that for $b \in \mathbb{N}$ the two distributions do not coincide, they are both Poisson distributions and their means are related. As we now show, as $c \to \infty$, the mean of $\rho_{b,c}$ is asymptotic to the mean of $\rho_{b,\alpha_b(c)}$, which equals c. To prove this, we use the last two displays to write

$$
\begin{aligned}
c \;<\; \langle \rho_{b,c} \rangle &= \frac{cZ_{b-1}(c)}{Z_b(c)} = c + \frac{c(Z_{b-1}(c) - Z_b(c))}{Z_b(c)} \\
&= c + \frac{c^b/(b-1)!}{Z_b(c)} < c + \frac{c^b/(b-1)!}{c^b/b!} = c + b.
\end{aligned}
$$

It follows that $1 \le \langle \rho_{b,c} \rangle/c < 1 + b/c$. This implies that $\lim_{c \to \infty} \langle \rho_{b,c} \rangle/c = 1$, as claimed.

In order to complete our motivation of the LDP for $\Theta_{N,b}$ via Sanov's Theorem, we now prove that as $N \to \infty$ the N-dependent distributions of K_ℓ converge weakly to the Poisson distribution $\rho_{b,c}$ having components (7.2).

Theorem 7.1. *Fix $b \in \mathbb{N} \cup \{0\}$ and $c > b$. In the limit $K \to \infty$ and $N \to \infty$ with $K/N = c$ the distributions $P_{N,b,m}\{K_\ell \in \cdot\}$ converge weakly to $\rho_{b,c}$.*

Proof. Throughout this proof we write $\lim$ to denote the limit as $K \to \infty$ and $N \to \infty$ with $K/N = c$. We start by determining the N-dependent distribution of K_ℓ, first for $b = 0$. For $j \in \mathbb{N}_0$ satisfying $0 \le j \le K$ and for any $\ell \in \Lambda_N$, the event $\{K_\ell = j\}$ occurs if and only if j of the K particles occupy the site ℓ and $N - K$ particles occupy any of the other sites. Thus K_ℓ has the distribution of a binomial random variable $B_{K,1/N}$ based on K independent Bernoulli trials each having the probability of success $1/N$. Thus for $j \in \mathbb{N}_0$ satisfying $0 \le j \le K$

$$
P_{N,b,m}\{K_\ell = j\} = P\{B_{K,1/N} = j\} = \frac{K!}{j!(K-j)!} \left(\frac{1}{N}\right)^j \left(1 - \frac{1}{N}\right)^{K-j}. \tag{7.3}
$$

The N-dependent distribution of K_ℓ for general $b \in \mathbb{N}$ is more complicated. In this case, because each site must be occupied by a minimum of b particles, K_ℓ has the distribution of the binomial random variable $B_{K,1/N}$ conditioned on $B_{K,1/N} \ge b$. Thus for $j \in \mathbb{N}_b$ satisfying

$b \leq j \leq K$ the common distribution of K_ℓ is given by

$$
\begin{aligned}
P_{N,b,m}\{K_\ell = j\} &= P\{B_{K,1/N} = j | B_{K,1/N} \geq b\} \quad\quad (7.4)\\
&= \frac{1}{P\{B_{K,1/N} \geq b\}} \cdot P\{B_{K,1/N} = j\}\\
&= \frac{1}{1 - \sum_{k=0}^{b-1} P\{B_{K,1/N} = k\}} \cdot P\{B_{K,1/N} = j\}.
\end{aligned}
$$

We now prove that the N-dependent distributions of K_ℓ converge weakly to $\rho_{b,c}$, first for $b = 0$. In this case K_ℓ has the same distribution as $B_{K,1/N}$, and $\rho_{0,c}$ is a standard Poisson distribution on $\mathbb{N} \cup \{0\}$. The weak convergence of the distributions of $B_{K,1/n}$ to $\rho_{0,c}$ is a classical result known as the Poisson Theorem [30, Thm. III.3.4]. To prove it one uses characteristic functions and the Continuity Theorem [30, Thm. III.3.1] to show that for each $t \in \mathbb{R}$

$$
\lim E\{\exp(itB_{K,1/N})\} = \exp[c(e^{it} - 1)] = \int_{\mathbb{N} \cup \{0\}} \exp(itx)\rho_{0,c}(dx). \quad\quad (7.5)
$$

This completes the proof that for $b = 0$ the N-dependent distribution of K_ℓ converges to $\rho_{0,c}$.

We now prove that the N-dependent distributions of K_ℓ converge weakly to $\rho_{b,c}$ for $b \in \mathbb{N}$. In this case K_ℓ has the same distribution as $B_{K,1/N}$ conditioned on $B_{K,1/N} \geq b$, and $\rho_{b,c}$ is a Poisson distribution restricted to $\mathbb{N}_b$. We start by calculating the characteristic function of $\rho_{b,c}$. For $t \in \mathbb{R}$

$$
\int_{\mathbb{N}_b} e^{itx} \rho_{b,c}(dx) = \frac{1}{Z_b(c)} \sum_{j=b}^{\infty} e^{itj} \frac{c^j}{j!} = \frac{1}{Z_b(c)} \left(\exp(ce^{it}) - \sum_{k=0}^{b-1} \frac{(ce^{it})^k}{k!} \right). \quad\quad (7.6)
$$

The next step is to calculate the characteristic function of K_ℓ. For $t \in \mathbb{R}$ we have by (7.4)

$$
\begin{aligned}
&E\{\exp(itK_\ell)\} \quad\quad (7.7)\\
&= \sum_{j=b}^{K} e^{itj} P\{B_{K,1/N} = j | B_{K,1/N} \geq b\}\\
&= \frac{1}{P\{B_{K,1/N} \geq b\}} \cdot \sum_{j=b}^{K} e^{itj} P\{B_{K,1/N} = j\}\\
&= \frac{1}{P\{B_{K,1/N} \geq b\}} \cdot \left(\sum_{j=0}^{K} e^{itj} P\{B_{K,1/N} = j\} - \sum_{k=0}^{b-1} e^{itk} P\{B_{K,1/N} = k\} \right)\\
&= \frac{1}{1 - \sum_{k=0}^{b-1} P\{B_{K,1/N} = k\}} \cdot \left(E\{\exp(itB_{K,1/n})\} - \sum_{k=0}^{b-1} e^{itk} P\{B_{K,1/N} = k\} \right)
\end{aligned}
$$

43

The weak convergence of the distributions of $B_{K,1/N}$ to $\rho_{0,c}$, which is a consequence of the limit (7.5) for all $t \in \mathbb{R}$, implies that for each $k \in \mathbb{N} \cup \{0\}$ satisfying $0 \leq k \leq K$

$$\lim P\{B_{K,1/N} = k\} = \rho_{0,c;k} = e^{-c}\frac{c^k}{k!}.$$

Combining this with (7.5)–(7.7), we see that for $t \in \mathbb{R}$

$$\lim E\{\exp(itK_\ell)\}$$
$$= \frac{1}{1 - \sum_{k=0}^{b-1} e^{-c}c^k/k!} \cdot \left(\exp[c(e^{it} - 1)] - \sum_{k=0}^{b-1} e^{itk}e^{-c}c^k/k! \right)$$
$$= \frac{1}{e^c - \sum_{k=0}^{b-1} c^k/k!} \cdot \left(\exp(ce^{it}) - \sum_{k=0}^{b-1} (ce^{it})^k/k! \right)$$
$$= \frac{1}{Z_b(c)} \left(\exp(ce^{it}) - \sum_{k=0}^{b-1} \frac{(ce^{it})^k}{k!} \right) = \int_{\mathbb{N}_b} e^{itx} \rho_{b,c}(dx).$$

This shows that for all $t \in \mathbb{R}$ the characteristic functions of K_ℓ converge to the characteristic function of $\rho_{b,c}$. Again, the Continuity Theorem [30, Thm. III.3.1] implies that the N-dependent distributions of K_ℓ converge weakly to $\rho_{b,c}$ for $b \in \mathbb{N}$. This completes the proof of the theorem. ∎

Our motivation of the LDP for $\Theta_{N,b}$ by applying Sanov's Theorem is now done. We end this paper by outlining our approach to proving the LDP for $\Theta_{N,b}$

8 Our Approach to Proving the LDP for $\Theta_{N,b}$

Our approach to proving the LDP for $\Theta_{N,b}$ involves the following four steps, which are exact analogs of the four steps used in section 3 to prove the Boltzmann-Sanov LDP in Theorem 2.3. It is thus fair to say that Boltzmann's work on the discrete ideal gas was both an inspiration and a road map for our large deviation analysis of the droplet model.

1. **Local estimate.** Inspired by Boltzmann's calculation of the Maxwell–Boltzmann distribution, step 1 is to derive a local large deviation estimate for $\Theta_{N,b}$, which is stated in part (b) of Theorem 3.1 in [16]. This local estimate, one of the centerpieces of the paper, gives information not available in the LDP for $\Theta_{N,b}$, which involves global estimates. It states

44

that in the limit as $K \to \infty$ and $N \to \infty$ with $K/N = c$, for any probability measure $\theta^{(N)}$ in the range of the number-density measure $\Theta_{N,b}$

$$\frac{1}{N} \log P_{N,K,m}(\Theta_{N,b} = \theta^{(N)}) = -R(\theta|\rho_{b,\alpha_b(c)}) + \varepsilon_N(\theta^{(N)}),$$

where $\varepsilon_N(\theta^{(N)}) \to 0$ uniformly for all measures $\theta^{(N)}$ in the range of $\Theta_{N,b}$. Showing that the parameter of the Poisson distribution in the local large deviation estimate equals $\alpha_b(c)$ is one of the crucial elements of the proof; it was inspired by the derivation of the Maxwell–Boltzmann distribution as explained in part (a) of Theorem 4.1 in the present paper. The proof of this local estimate involves the asymptotic analysis of a product of two multinomial coefficients. The first of these is an analog of the multinomial coefficient appearing in the last line of (3.1) in the proof of the local large deviation estimate for the discrete ideal gas in Theorem 3.1 in this paper.

2. **Approximation result.** In Theorem B.1 in [15] we prove that for any probability measure $\theta \in \mathcal{P}_{\mathbb{N}_b,c}$ there exists a sequence $\theta^{(N)}$ in the range of $\Theta_{N,b}$ for which the following properties hold: $\theta^{(N)} \Rightarrow \theta$ as $N \to \infty$ and if $R(\theta|\rho_{b,\alpha}) < \infty$, then $R(\theta^{(N)}|\rho_{b,\alpha}) \to R(\theta|\rho_{b,\alpha})$ as $N \to \infty$.

3. **Large deviation limit for open balls and other subsets.** In Theorem 4.1 in [15] we show how to use the approximation result in step 2 to lift the local estimate in step 1 to the large deviation limit for $\Theta_{N,b}$ lying in open balls and in certain other subsets of $\mathcal{P}_{\mathbb{N}_b,c}$. Theorem 4.1 in [15] is derived as a consequence of the general formulation given in Theorem 4.2 in [15].

4. **Large deviation upper and lower bounds.** Theorem 4.3 in [15] presents a general procedure that we apply to lift the large deviation limit for $\Theta_{N,b}$ lying in open balls and in certain other subsets of $\mathcal{P}_{\mathbb{N}_b,c}$ in step 3 to the large deviation upper and lower bounds for $\Theta_{N,K}$ stated in Theorem 2.1 in [16].

This completes our discussion of how Boltzmann's work on the discrete ideal gas guided our large deviation analysis of the droplet model.

References

[1] John R. Baxter and Naresh C. Jain. A comparison principle for large deviations. *Proceedings of the American Mathematical Society* 103:1235–1240, 1988.

[2] R. J. Baxter. *Exactly Solved Models in Statistical Mechanics*, Academic Press, 1982.

[3] Ludwig Boltzmann. Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung, respective den Sätzen über das Wärmegleichgewicht [On the relationship between the second main theorem of mechanical heat theory and the probability calculation with respect to the results about the heat equilibrium]. *Sitzb. d. Kaiserlichen Akademie der Wissenschaften*, mathematich-naturwissen Cl. LXXVI, Abt. II, 373-435, 1877.

[4] Christopher Boucher, Richard S. Ellis, and Bruce Turkington. Derivation of maximum entropy principles in two-dimensional turbulence via large deviations. *Journal of Statistical Physics* 98:1235–1278, 2000.

[5] Marius Costeniuc, Richard S. Ellis, and Hugo Touchette. Analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model. *Journal of Mathematical Physics* 46:063301 (25 pages), 2005.

[6] Definition of "powder" retrieved May 16, 2014 from Wikipedia.com website: http://en.wikipedia.org/wiki/Powder_substance.

[7] Definition of "spray" retrieved March 4, 2014 from Dictionary.com website: http://dictionary.reference.com/browse/spray?s=t.

[8] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, second edition, Springer, New York, 1998.

[9] Paul Dupuis and Richard S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, New York, 1997.

[10] Richard S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*, Springer, New York, 1985. Reprinted in 2006 in *Classics of Mathematics*.

[11] Richard S. Ellis. The theory of large deviations and applications to statistical mechanics. *Long-Range Interacting Systems: Les Houches 2008 Session XC*, 227–277. Edited by T. Dauxois, S. Ruffo, and L. F. Cugliandolo. Oxford University Press (New York), 2010. Posted at http://www.math.umass.edu/~rsellis/pdf-files/Les-Houches-paper.pdf.

[12] Richard S. Ellis. The theory of large deviations: from Boltzmann's 1877 calculation to equilibrium macrostates in 2D turbulence. *Physica D* 133:106–136, 1999.

[13] Richard S. Ellis, Kyle Haven, and Bruce Turkington. Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles. *Journal of Statistical Physics* 101: 999–1064, 2000.

[14] Richard S. Ellis, Peter Otto, and Hugo Touchette. Analysis of phase transitions in the mean-field Blume-Emery-Griffiths model. *Annals of Applied Probability* 15:2203–2254, 2005.

[15] Richard S. Ellis and Shlomo Ta'asan. Detailed large deviation analysis of a droplet model having a Poisson equilibrium distribution. Unpublished. Posted at http://arxiv.org/abs/1405.5091v3.

[16] Richard S. Ellis and Shlomo Ta'asan. Large deviation analysis of a droplet model having a Poisson equilibrium distribution. Submitted for publication (2014). Posted at http://www.math.umass.edu/~rsellis/pdf-files/ldp-droplet-model.pdf.

[17] Richard S. Ellis and Kongming Wang. Limit theorems for the empirical vector of the Curie-Weiss-Potts model. *Stochastic Processes and Their Applications* 35:59–79, 1990.

[18] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York, 1986.

[19] William R. Everdell. *The First Moderns: Profiles in the Origins of Twentieth-Century Thought*. University of Chicago Press, Chicago, 1997.

[20] Kerson Huang. *Statistical Mechanics*, John Wiley & Sons, New York, 1963.

[21] David Lindsay. *Boltzmann's Atom: The Great Debate That Launched a Revolution in Physics*, The Free Press, New York, 2001.

[22] Maxwell–Boltzmann distribution. Retrieved March 1, 2014 from http://en.wikipedia.org /wiki/Maxwell-Boltzmann-distribution.

[23] Maxwell–Boltzmann statistics. Retrieved March 1, 2014 from http://en.wikipedia.org /wiki/Maxwell-Boltzmann-statistics.

[24] Barry McCoy and Tai Tsun Wu. *The Two-Dimensional Ising Model*, Harvard University Press, 1973.

[25] R. A. Mugele and H. D. Evans. Droplet size distribution in sprays, *Ind. Eng. Chem.* 43:1317–1324, 1951.

[26] R. K. Parthia and Paul D. Beale. *Statistical Mechanics*, second edition, Academic Press, 2011.

[27] Dana Randall. Combinatorial Methods for Statistical Physics Models. Special Topics Course, Winter 1999. Retrieved October 2, 2014 from http://people.math.gatech.edu/ randall/topics2.html.

[28] I. N. Sanov. On the probability of large deviations of random variables. In Russian, 1957. English translation from *Mat. Sb.* (42) in *Selected Translations in Mathematical Statistics and Probability I*, 1961, pp. 213–244.

[29] R. W. Sellens and T. A. Brzustowski. A prediction of the drop size distribution in a spray from first principles. *Atomisation and Spray Technology* 1:89–102, 1985.

[30] A. N. Shiryaev. *Probability*, second edition, Springer, 1996.