Chapter 1

Elementary examples of equilibrium states

Equilibrium states, a concept originating from statistical mechanics, are probability measures on topological spaces that are characterized by variational principles. They maximize the sum (or difference) of an entropy and an energy like quantity. Depending on the special choice of the variational problem these measures can have various interesting properties. Some of them are discussed in this introductory chapter on a rather elementary level. Starting from the roots of these ideas in equilibrium statistical mechanics, we outline the connection between equilibrium states and the theory of large deviations and introduce the Ising model of ferromagnetism on a finite lattice as a more concrete example (Sections 1.1 and 1.2). Then we indicate how Markov measures on finite alphabets can be characterized by a variational principle (Section 1.3), and the final section deals with the role played by equilibrium states in the ergodic theory of dynamical systems. That section also furnishes the background for a first discussion of Birkhoff's ergodic theorem.

1.1 Equilibrium states in finite systems

A physical system consisting of many particles can be described on two levels: Microscopically it is determined by its *configuration*, i.e., by the positions and momenta of all particles. Knowing the configuration of a system which obeys the laws of classical mechanics and which is not influenced from outside allows one in principle to determine its exact configuration at any future time. Of course, this fact is of little practical relevance, because the configuration of a realistic large system (e.g., the positions and momenta of all $2.7 \cdot 10^{22}$ molecules of an ideal gas in a one litre container) cannot even approximately be known. On the other hand, a good description of the macroscopic properties of such a system is provided by a relatively small number of observable parameters like total energy, temperature, entropy, etc. Mathematically, these macroscopic quantities are parameters associated with probability distributions on the space of all configurations. We call such a distribution a *state*.

This point of view was developed towards the end of the 19th century in the works of Ludwig Boltzmann (1844–1906) and John Willard Gibbs (1839–1903) in their attempt to reconcile the irreversibility of (macroscopic) thermodynamic processes with the reversibility of the supposed underlying mechanical motions. A historical document that summarizes the debates of this period with great clarity is the article [18] by P. and T. Ehrenfest. Even today it is a pleasure to read this account of the pioneering ideas of Boltzmann and Gibbs that, although written at a time where measure theory and modern probability theory were not yet invented, describes exactly the mathematical shortcomings of the theory at that time. Today, nearly one century later, a broad mathematical discipline called *ergodic theory* provides a strong foundation for a better understanding of (not only) these ideas.

If a physical system that is confined to a finite volume is not influenced from outside, it is driven by its internal fluctuations towards an *equilibrium state*, i.e., it prefers configurations which are compatible with the macroscopic parameters of this state (and thus do not reveal any additional useful information about the system). In order to make these very vague considerations more precise, let us consider an elementary probabilistic model that reflects the two “Fundamental Theorems of Thermodynamics”:

1. The energy of a closed system is constant.
2. The entropy of such a system is maximized by its equilibrium states.

In this section we consider an elementary probabilistic setting that allows us to define quantities called entropy and energy which are associated with probability measures, and we describe those measures that maximize the entropy under the constraint of keeping the energy constant. Despite its simplicity this setting already contains germs of many of the technically more difficult definitions and proofs we shall meet in later chapters.

Let Ω be a finite set, our (abstract) configuration space.

▷ A *state* is a probability vector $\mu = (\mu(\omega) \mid \omega \in \Omega)$. The set of all states is denoted by $\mathcal{M}$.

▷ The *entropy* of the state μ is defined as $H(\mu) := -\sum_{\omega\in\Omega} \mu(\omega) \log \mu(\omega)$. It is continuous as a function of $\mu \in \mathcal{M} \subset \mathbb{R}^{\Omega}$.

1.1.1 Remark At this point the reader should not try to interpret the expression for $H(\mu)$ as the classical thermodynamic entropy. Instead, we take it as a measure for the amount of uncertainty that the observer is left with when he knows that the system is in state μ. The connection between the information theoretic and the thermodynamic aspects of entropy is discussed later.

Let $A \subseteq \Omega$. With the event "$\omega \in A$" we want to associate a positive real number I_A that can be interpreted as its amount of information. If one requires that I_A is a continuous function of the probability $\mu(A) := \sum_{\omega \in A} \mu(\omega)$ only and that $I_{A \cap B} = I_A + I_B$ for statistically independent events A and B, the only possible choice is $I_A = -\log \mu(A)$ where the logarithm can be taken to any base. For convenience we shall always use the natural logarithm. Now $H(\mu) = \sum_{\omega \in \Omega} \mu(\omega) I_\omega$ is just the average amount of information of the elementary events ω. In other words, $H(\mu)$ is the expected amount of information that can be gained from further observations on the system, if its present state is known to be μ. That the maximal value of $H(\mu)$ is $\log |\Omega|$ can be seen as follows:

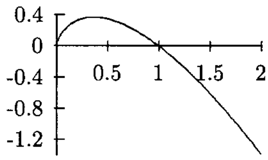


Figure 1.1: The function φ

Let $\varphi : [0,1] \to [0,\infty)$, $\varphi(t) = -t \log t$, see Figure 1.1. φ is continuous and concave and $H(\mu) = \sum_\omega \varphi(\mu(\omega))$. Throughout this book, φ will always denote this function.

Set $n := |\Omega|$. By an elementary version of Jensen's inequality (see A.2.2),

$$\frac{1}{n}\sum_\omega \varphi(\mu(\omega)) \le \varphi\left(\frac{1}{n}\sum_\omega \mu(\omega)\right) = \varphi\left(\frac{1}{n}\right)$$

so that

$$H(\mu) = \sum_\omega \varphi(\mu(\omega)) \le n \cdot \varphi\left(\frac{1}{n}\right) = H\left(\left(\frac{1}{n}, \dots, \frac{1}{n}\right)\right) = \log n\,.$$

Let μ and μ' be probability vectors on spaces Ω and Ω' respectively. The additivity of the information $I_{A \cap B} = I_A + I_B$ for independent events implies

$$H(\mu \times \mu') = H(\mu) + H(\mu')\,.$$

$\diamond$

We continue specifying our model:

- ▷ Each configuration $\omega \in \Omega$ is assigned an energy value $u(\omega) \in \mathbb{R}$ such that in state μ the system has mean energy $\mu(u) := \sum_{\omega \in \Omega} \mu(\omega) u(\omega)$. By $V_\mu(u) := \mu(u^2) - (\mu(u))^2$ we denote the variance of u under μ.
- ▷ $Z(\beta) := \sum_{\omega \in \Omega} \exp(-\beta u(\omega))$ is the *partition function* of u. Although it is defined for complex β, we are mostly interested in its values for real arguments.

▷ For real parameters β the *Gibbs measure* μ_β on Ω is defined by

$$\mu_\beta(\omega) := \frac{1}{Z(\beta)} \exp(-\beta u(\omega)) .$$

As $\frac{\mu_\beta(\omega)}{\mu_\beta(\omega')} \to 0$ for $\beta \to +\infty$ if $u(\omega) > u(\omega')$, the measures μ_β converge to the equidistribution on $\Omega_{min} := \{\omega : u(\omega) = \min_\Omega u\}$ if $\beta \to +\infty$. Since an analogous argument holds for the limit $\beta \to -\infty$, it follows in particular that

$$\lim_{\beta\to+\infty} \mu_\beta(u) = \min_\omega u(\omega) \quad \text{and} \quad \lim_{\beta\to-\infty} \mu_\beta(u) = \max_\omega u(\omega) . \tag{1.1}$$

1.1.2 Lemma *$\beta \mapsto \log Z(\beta)$ is a real analytic map and*

$$(\log Z)'(\beta) = -\mu_\beta(u) \quad \text{and} \quad (\log Z)''(\beta) = \mathrm{V}_{\mu_\beta}(u) \geq 0$$

with equality if and only if the function u is constant. In particular, $\beta \mapsto \log Z(\beta)$ is convex.

Proof: $Z(\beta) = \sum_{\omega\in\Omega} \exp(-\beta u(\omega))$ is analytic and nonzero for $\beta \in \mathbb{R}$ such that $\log Z(\beta)$ is real analytic. Elementary differentiation yields

$$(\log Z)'(\beta) = \frac{Z'(\beta)}{Z(\beta)} = -\frac{1}{Z(\beta)} \sum_{\omega\in\Omega} u(\omega) e^{-\beta u(\omega)} = -\mu_\beta(u)$$

and

$$\begin{aligned}(\log Z)''(\beta) &= \frac{Z''(\beta)}{Z(\beta)} - \left(\frac{Z'(\beta)}{Z(\beta)}\right)^2 = \frac{1}{Z(\beta)} \sum_{\omega\in\Omega} u^2(\omega) e^{-\beta u(\omega)} - (\mu_\beta(u))^2 \\ &= \mathrm{V}_{\mu_\beta}(u) .\end{aligned}$$

□

The following theorem is an elementary prototype of a much more general statement that is proved in a later section and that characterizes Gibbs measures by means of a variational principle.

1.1.3 Theorem (Variational principle)
Each Gibbs measure μ_β with $\beta \in \mathbb{R}$ satisfies

$$H(\mu_\beta) + \mu_\beta(-\beta u) = \log Z(\beta) = \sup_{\nu\in\mathcal{M}} (H(\nu) + \nu(-\beta u)) . \tag{1.2}$$

A measure ν for which this supremum is attained is called an equilibrium state *for $-\beta u$. Thus Gibbs measures are equilibrium states. In fact, μ_β is the only equilibrium state for $-\beta u$.*

Proof: Of course this theorem can be proved by elementary calculus using Lagrange multipliers. We prefer to give a proof based on a convexity argument (Jensen inequality, see A.2.2) applied to the concave function $x \mapsto \log x$. For $\nu \in \mathcal{M}$ we have

$$\begin{aligned} H(\nu) + \nu(-\beta u) &= -\sum_{\omega\in\Omega} \nu(\omega)(\log \nu(\omega) + \beta u(\omega)) \\ &= \sum_{\omega\in\Omega} \nu(\omega) \log \frac{e^{-\beta u(\omega)}}{\nu(\omega)} \\ &\leq \log \sum_{\omega\in\Omega} \nu(\omega) \frac{e^{-\beta u(\omega)}}{\nu(\omega)} \\ &= \log Z(\beta) \end{aligned}$$

with equality if and only if the random variable $\omega \mapsto \frac{\exp(-\beta u(\omega))}{\nu(\omega)}$ is constant, i.e., if $\nu = \mu_\beta$. □

1.1.4 Remark Although *equilibrium* is rather a dynamical notion that can hardly be separated from the idea of a dynamical process evolving towards it (in the case of a stable equilibrium) or away from it (if the equilibrium is unstable), the variational principle makes no statement about the time evolution of non-equilibrium initial states. It just characterizes the equilibrium on the basis of physical principles that are beyond the scope of this text. ◇

1.1.5 Corollary *If E^* is a real number and if $\min_\omega u(\omega) < E^* < \max_\omega u(\omega)$, then there is a unique parameter $\beta^* \in \mathbb{R}$ such that μ_{β^*} has energy $\mu_{\beta^*}(u) = E^*$ and maximizes entropy among all states ν with the same energy E^*.*

Proof: It follows from Lemma 1.1.2 that $\beta \mapsto \mu_\beta(u)$ is a continuous and strictly decreasing function of β. As $\lim_{\beta\to-\infty} \mu_\beta(u) = \max_\omega u(\omega)$ and $\lim_{\beta\to+\infty} \mu_\beta(u) = \min_\omega u(\omega)$ by (1.1), there is a unique parameter β^* with $\mu_{\beta^*}(u) = E^*$. Hence, for any state ν with $\nu(u) = E^*$ the variational principle implies

$$H(\nu) - \beta^* E^* = H(\nu) + \nu(-\beta^* u) \leq H(\mu_{\beta^*}) + \mu_{\beta^*}(-\beta^* u) = H(\mu_{\beta^*}) - \beta^* E^* ,$$

i.e., $H(\nu) \leq H(\mu_{\beta^*})$, and equality holds if and only if $\nu = \mu_{\beta^*}$. □

1.1.6 Remark In a physical context, $T = 1/\beta$ denotes the temperature of a system and $\mu_{1/T}$ describes the equilibrium of the system at temperature T. (For simplicity we set the Boltzmann constant $k = 1$.) Using the notation $F(\nu) := \nu(u) - T \cdot H(\nu)$ for the *free energy* of the state ν, we can reformulate the variational principle in the following way:

$$F(\nu) \geq F(\mu_{1/T}) = -T \log Z(1/T) \text{ with equality if and only if } \nu = \mu_{1/T}. \quad (1.3)$$

Even more can be said: In the next lemma we prove $\frac{d}{d\beta}H(\mu_\beta) = \beta \cdot \frac{d}{d\beta}\mu_\beta(u)$. It follows immediately that

$$\frac{d}{dT}H(\mu_{1/T}) = \frac{1}{T} \cdot \frac{d}{dT}\mu_{1/T}(u)$$

where $\mu_{1/T}(u)$ is the energy of the system at temperature T. Compare this to the usual thermodynamic definition of the entropy S of a system by its differential $dS = \frac{dQ}{T}$, where Q denotes the heat content of the system and T is its absolute temperature. ◇

1.1.7 Lemma

$$\frac{d}{d\beta}H(\mu_\beta) = \beta \cdot \frac{d}{d\beta}\mu_\beta(u)\ .$$

Proof: Recall from the variational principle that $H(\mu_\beta) = \log Z(\beta) + \beta\mu_\beta(u)$. By Lemma 1.1.2 we have $\frac{d}{d\beta}\log Z(\beta) = -\mu_\beta(u)$. Hence

$$\frac{d}{d\beta}H(\mu_\beta) = -\mu_\beta(u) + \mu_\beta(u) + \beta\frac{d}{d\beta}\mu_\beta(u) = \beta\frac{d}{d\beta}\mu_\beta(u)\ .$$

□

1.1.8 Remark Recall that $\mu_\beta(\omega) = \frac{1}{Z(\beta)}e^{-\beta u(\omega)}$. For high temperatures $T \to +\infty$ (i.e., $\beta \searrow 0$) the equilibrium states μ_β converge to the *equidistribution* on Ω, which is the state maximizing the entropy (i.e., maximizing the disorder of the system). At low temperatures $T \searrow 0$ (i.e., $\beta \to +\infty$) the μ_β converge towards the *ground state*, i.e., towards the equidistribution on $\Omega_{min} := \{\omega :\ u(\omega) = \min_\Omega u\}$, see (1.1). ◇

1.1.9 Remark In later sections we mostly fix $-\beta$ and include it in the function u. In other words, we study the case $\beta = -1$ for which (1.2) becomes

$$p(u) := \sup_{\nu\in\mathcal{M}} (H(\nu) + \nu(u)) = \log Z(-1) = \log \sum_{\omega\in\Omega} e^{u(\omega)}$$

where the supremum is attained if and only if $\nu(\omega) = e^{u(\omega)-p(u)}$. ◇

1.1.10 Exercise By a slight modification of the proof of Theorem 1.1.3 show that for each $\nu \in \mathcal{M}$

$$H(\nu) = \inf_{u:\Omega\to\mathbb{R}} (p(u) - \nu(u))\ .$$

◇

1.2 Systems on finite lattices

In more realistic physical models configurations usually have some spatial structure. Mathematically this means that there is a group of isometries acting on an underlying position space. We consider the following elementary example of such a situation:

- The configuration space is of the form $\Omega = \Sigma^G$, where the additive group $G = (\mathbb{Z}/\ell\mathbb{Z})^d$ is geometrically interpreted as a finite grid, and Σ is the finite set of possible values of a configuration at a given site. Elements $\omega \in \Omega$ are written as $\omega = (\omega_g)_{g\in G}$.
- The group G operates on Ω by an action $\mathcal{T} = (T^g : g \in G)$ where the maps $T^g : \Omega \to \Omega$ act as *shift* transformations on Ω, namely $(T^g\omega)_i := \omega_{i+g}$.
- $\mathcal{M}(\mathcal{T})$ denotes the set of all probability vectors ν on Ω that are invariant under $\mathcal{T}$, i.e., for which $\nu(T^{-g}\omega) = \nu(\omega)$ for all $g \in G$ and all $\omega \in \Omega$. If $\nu \in \mathcal{M}(\mathcal{T})$, then $\nu(u \circ T^g) = \sum_\omega \nu(\omega)u(T^g\omega) = \sum_\omega \nu(T^{-g}\omega)u(\omega) = \sum_\omega \nu(\omega)u(\omega) = \nu(u)$.
- The energy function u has the form $u(\omega) = \sum_{g\in G} \psi(T^g\omega)$ with a *local energy function* $\psi : \Omega \to \mathbb{R}$. In this case $u(T^g\omega) = u(\omega)$ for all $g \in G$. As $\mu_\beta(\omega) = \frac{1}{Z(\beta)}e^{-\beta u(\omega)}$, it follows that $\mu_\beta \in \mathcal{M}(\mathcal{T})$ (i.e., the Gibbs distribution is spatially homogeneous). In particular $\nu(u) = |G|\,\nu(\psi)$ for each $\nu \in \mathcal{M}(\mathcal{T})$.
- Define the *pressure* of $-\beta\psi$ as

  $$p(-\beta\psi) := \sup_{\nu\in\mathcal{M}(\mathcal{T})} \left(\frac{1}{|G|}H(\nu) + \nu(-\beta\psi)\right) .$$

  Then the variational principle can be written as

  $$\frac{1}{|G|}H(\mu_\beta) + \mu_\beta(-\beta\psi) = \frac{1}{|G|}\log Z(\beta) = p(-\beta\psi) . \tag{1.4}$$

  From a physical point of view the word *pressure* should not be taken too literally, because

  $$p(-\beta\psi) = -\frac{1}{T}\frac{F(\mu_{1/T})}{|G|}$$

  in terms of the free energy $F(\nu) = \nu(u) - TH(\nu)$, whereas in a thermodynamic context $p = -\frac{\partial F}{\partial V}$ where V denotes the volume of the system and p its pressure. Therefore it might be more appropriate to call $\frac{1}{\beta}p(-\beta\psi)$ the pressure (this is $p(-\psi)$ if $\beta = T = 1$), but this is not the common usage in ergodic theory.

  Note also that the term $\frac{1}{|G|}H(\mu_\beta)$ entering the definition of pressure can be interpreted as the entropy per lattice site of the measure μ_β.

1.2.1 Example (Bernoulli measures and large deviations) Suppose that the local energy depends on a single lattice site (which can be assumed without loss of generality to be $g=0$). So $\psi(\omega)=f(\omega_0)$ for some $f:\Sigma\to\mathbb{R}$. For $\sigma\in\Sigma$ let

$$q_\beta(\sigma):=\frac{e^{-\beta f(\sigma)}}{\sum_{\tau\in\Sigma}e^{-\beta f(\tau)}}$$

and write $N:=|G|$. Then $(q_\beta(\sigma)\mid\sigma\in\Sigma)$ is a Gibbs measure on Σ with $\rho_\beta:=\max_\sigma q_\beta(\sigma)<1$. As $Z(\beta)$ takes the form $Z(\beta)=(\sum_{\tau\in\Sigma}e^{-\beta f(\tau)})^N$, we have for each $\omega\in\Omega$

$$\mu_\beta(\omega)=\frac{1}{Z(\beta)}\exp\Big(-\beta\sum_{g\in G}f(\omega_g)\Big)=\prod_{g\in G}q_\beta(\omega_g)\,,$$

i.e., $\mu_\beta=q_\beta^{\times G}$ is a Bernoulli measure. In particular, $\mu_\beta(\omega)\le\rho_\beta^N$. Hence, the probability $\mu_\beta(\omega)$ of each configuration is exponentially small in the system size N, and it does not make any sense to think of the actually observed configuration of a system in state μ_β as the most probable one. Instead, following an idea of Boltzmann's, we shall argue that μ_β strongly favours such configurations for which the *empirical distribution* π_ω is very close to the one-dimensional marginal distribution of μ_β. Here π_ω is the probability distribution on Σ defined by $\pi_\omega(\sigma)=N^{-1}\operatorname{card}\{g\in G:\omega_g=\sigma\}$. (Section 12b of the Ehrenfests' article [18] is an excellent account of this argument.)

As q_β is the Gibbs measure for $-\beta f$ on Σ, we have

> $H(q_\beta)-\beta\,q_\beta(f)\ge H(\pi)-\beta\,\pi(f)$ for each probability vector π on Σ with equality if and only if $\pi=q_\beta$.

So we may measure the distance from a probability vector π to q_β by

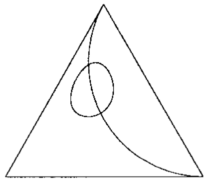
$$d_\beta(\pi):=(H(q_\beta)-\beta\,q_\beta(f))-(H(\pi)-\beta\,\pi(f))$$

and ask for the probability under μ_β of the event $U_{\beta,r}:=\{\omega\in\Omega:d_\beta(\pi_\omega)<r\}$. (Geometrically this is the event that π_ω belongs to a convex neighbourhood of q_β in the simplex of probability vectors on Σ, because $d_\beta(\pi)$ is a convex function of π. See Theorem 3.3.2 for the convexity and Figure 1.2 for an illustration in the case $|\Sigma|=3$.) The result we are going to prove is: if $0<r<H(q_\beta)-\beta\,q_\beta(f)$, then

$$\lim_{N\to\infty}\frac{1}{N}\log\mu_\beta(\Omega\setminus U_{\beta,r})=-r\,. \tag{1.5}$$

This is a simple example of the *large deviations property* of Gibbs measures. It says that under the probability distribution μ_β roughly a fraction e^{-rN} of all configurations does not belong to the neighbourhood $U_{\beta,r}$ of μ_β.

Figure 1.2: Gibbs measures q_β and a convex neighbourhood $\{\pi : d_\beta(\pi) < 0.05\}$ with $\Sigma = \{1,2,3\}$, $f(1) = 1, f(2) = 2, f(3) = 3$ and $\beta = 0.5$.



In order to evaluate $\mu_\beta(U_{\beta,r})$ observe that $H(\mu_\beta) = N\,H(q_\beta)$ and therefore

$$\begin{aligned}\log \mu_\beta(\omega) &= -\beta \sum_{g\in G} f(\omega_g) - \log Z(\beta)\\ &= -N\beta\pi_\omega(f) - (H(\mu_\beta) - N\beta\mu_\beta(\psi))\\ &= -N\cdot\Big((H(q_\beta) - \beta q_\beta(f)) - (H(\pi_\omega) - \beta\pi_\omega(f))\Big) - N\cdot H(\pi_\omega)\\ &= -N\cdot d_\beta(\pi_\omega) - N\cdot H(\pi_\omega)\,. \end{aligned}\tag{1.6}$$

Let $K := \{k \in \mathbb{N}^\Sigma : \sum_{\sigma\in\Sigma} k(\sigma) = N\}$. Then each π_ω is of the form $\pi_\omega = k\,N^{-1}$ for some $k \in K$. Let

$$N(k) := \operatorname{card}\{\omega \in \Omega : \pi_\omega = k\,N^{-1}\} = \frac{N!}{\prod_{\sigma\in\Sigma} k(\sigma)!}\,.$$

We approximate the numbers $N(k)$ by Stirling's formula $n! = \sqrt{2\pi n}\,e^{-n}\,n^n\,e^{r_n}$ where $\lim_{n\to\infty} r_n = 0$. Then

$$N(k) = \text{const}\cdot N^{-\frac12(|\Sigma|-1)}\cdot \prod_{\substack{\sigma\in\Sigma\\ k(\sigma)\neq 0}} \left(k(\sigma)\,N^{-1}\right)^{-k(\sigma)-\frac12}\cdot\gamma(N,k)$$

where $\gamma(N,k)$ is bounded away from 0 and ∞ uniformly in N and k. Hence

$$\log N(k) = N\cdot H(k\,N^{-1}) + O(\log N)\,, \tag{1.7}$$

and we can estimate

$$\begin{aligned}\frac1N \log\mu_\beta(\Omega\setminus U_{\beta,r}) &= \frac1N \log \sum_{\omega\in\Omega\setminus U_{\beta,r}} \mu_\beta(\omega)\\ &\le \frac1N \log \sum_{\omega\in\Omega\setminus U_{\beta,r}} \exp\left(-Nr - NH(\pi_\omega)\right) \quad \text{by (1.6)}\\ &\le \frac1N \log \sum_{k\in K} N(k)\cdot\exp\left(-Nr - NH(k\,N^{-1})\right)\\ &\le \frac1N \log\left(|K|\cdot\exp(-Nr + O(\log N))\right)\\ &= -r + O\left(\frac{\log N}{N}\right) \quad \text{as } |K| \le N^{|\Sigma|},\end{aligned}$$

i.e., $1 - \mu_\beta(U_{\beta,r})$ is exponentially small in the lattice size N. This shows the "$\leq$"-direction of equality (1.5).

On the other hand, let $r' > r$. As $r < H(q_\beta) - \beta\, q_\beta(f)$, there is, for sufficiently large N, some $k \in K$ such that $d_\beta(k\,N^{-1}) \in [r, r']$. Therefore (1.6) and (1.7) lead to

$$\begin{aligned}
\frac{1}{N}\log \mu_\beta(\Omega \setminus U_{\beta,r}) &\geq \frac{1}{N}\log \sum_{\substack{\omega\in\Omega \\ \pi_\omega = k\,N^{-1}}} \mu_\beta(\omega) \\
&\geq \frac{1}{N}\log\left(N(k)\cdot \exp(-Nr' - NH(k\,N^{-1}))\right) \\
&= -r' + O\left(\frac{\log N}{N}\right).
\end{aligned}$$

As $r' > r$ is arbitrary, this finishes the proof of (1.5). ◇

1.2.2 Example (Ising model) A classical example where the local energy ψ depends not only on one lattice site but on local configurations involving several sites is the *Ising model*. It was designed by the physicist Lenz around 1920 to explain ferromagnetism and was named after his student Ising who contributed to its theory. The idea is that iron atoms are situated at the sites of a lattice G and behave like small magnets that may be oriented upwards (denoted by $+1$) or downwards (denoted by -1). Physically, two magnets that are close to one another need less energy to be oriented in the same sense than in an opposite sense. This leads to the following simplified model: $\Sigma = \{-1, +1\}$ represents the possible orientations of the magnetization, $\psi(\omega) = -\sum_{j=1}^{d}(\omega_0\omega_{e_j} + \omega_0\omega_{-e_j})$ is the local energy, where $e_1, \ldots, e_d$ denote the canonical unit vectors in G. At positive values of β (i.e., at positive temperatures) $u(\omega)$ is small for rather homogeneous configurations ω. Indeed, it is minimal if ω is constant $+1$ or -1. The *ground state* (also frozen state) is $\lim_{\beta\to+\infty}\mu_\beta = \frac{1}{2}(\delta_{(+1)^G} + \delta_{(-1)^G})$. Here δ_ω denotes the unit point mass in $\omega \in \Omega$, and $(+1)^G$ and $(-1)^G$ are the constant configurations consisting of $+1$'s (respectively -1's) only.

A rather nontechnical introduction to the Ising model is presented in [49, Chapter 1]. We come back to this model in Section 5.5. Figure 1.3 gives an impression of typical Ising configurations. ◇

1.2.3 Exercise Determine $\lim_{\beta\to-\infty}\mu_\beta$ for the Ising model. (This is the ground state of the so called anti-ferromagnetic Ising model which is obtained if the function ψ in the Ising model is replaced by $-\psi$.) ◇

Statistical properties of the equidistribution ($\beta = 0$) and quite often also of the ground state ($\beta = \infty$) are rather easy to describe. This is no longer true for