Relative Entropy, Large Deviations, and Statistical Mechanics

Richard S. Ellis
Department of Mathematics and Statistics
University of Massachusetts
Amherst, MA 01003
rsellis@math.umass.edu

Predictability Conference
Center for Nonlinear Studies
Los Alamos National Laboratory
May 11-15, 1998

Outline of Talk

1. Basic ideas in large deviations.

2. Analysis of a basic model of ferromagnetism.

3. A general approach to the analysis of statistical mechanical models.

4. Derivation of maximum entropy principles in two-dimensional turbulence.

The theory of large deviations studies the exponential decay of probabilities in certain random systems (rare events, statistical outliers, ...). The theory gives precise, exponential-order estimates that are perfectly suited for the asymptotic analysis arising in statistical mechanics.

—(1)—

Two Basic Questions of Statistical Mechanics

1. How can one use probability theory to model physical systems such as an ideal gas, a ferromagnet, or a fluid?

- Boltzmann, Gibbs: One models physical systems via probability measures on configuration space known as Gibbs canonical ensembles or Gibbs states.
- The theory of large deviations can be used to derive the form of the Gibbs state for an ideal gas.

2. How can one use such probabilistic models to predict the equilibrium behavior of the physical systems?

- One wants to describe phenomena such as phase transitions: the liquid-gas transition or spontaneous magnetization in a ferromagnet.
- In a freely evolving, inviscid fluid, one wants to describe coherent states, which are steady, stable mean flows comprised of one or more vortices that persist amidst the turbulent fluctuations of the vorticity field.
- We will use the theory of large deviations to study ferromagnets, a model of two-dimensional turbulence, and other models. In each case we derive a maximum entropy principle for which the extremal points represent equilibrium macrostates.

—(2)—

The First Moderns

1877. In a physical context Ludwig Boltzmann calculates the asymptotics of the multinomial coefficients.

- A revolutionary moment in human culture.
- Statistical mechanics and the theory of large deviations are born.

William R. Everdell, *The First Moderns*, University of Chicago Press (1997). This book traces the development of the modern consciousness in 19th and 20th century thought. Modernism's most important concept: discontinuity.

- p. 31: "[T]he mathematicians of 1870's Germany [Georg Cantor, Richard Dedekind, Gottlob Frege] were about to change the world.... [T]hey would become the first creative thinkers in any field to look at the world in a fully twentieth-century manner."
- p. 48: In large measure, it was the contributions of Ludwig Boltzmann in stochastics and statistics "which made the work of Planck and Einstein possible.... *He was at the center of the change.*"

—(3)—

Discrete Probabilistic Model

$\Lambda \doteq \{y_1, y_2, \ldots, y_\alpha\}$	set of possible outcomes ($\alpha \geq 2$)
$\Omega_n \doteq \Lambda^n$	configuration space
$\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega_n$	a configuration (each $\omega_i \in \Lambda$)
$\rho_k > 0,\ \sum_{k=1}^{\alpha} \rho_k = 1$	probability of $y_k \in \Lambda$
$\rho \doteq (\rho_1, \rho_2, \ldots, \rho_\alpha)$	probability vector in $I\!R^\alpha$
$\rho \doteq \sum_{k=1}^{\alpha} \rho_k \delta_{y_k}$ ($\rho\{y\} = \rho_k$ if $y = y_k$)	probability measure on Λ
$\mathcal{P}_\alpha \doteq \left\{\gamma \in I\!R^\alpha : \gamma_k \geq 0, \sum_{k=1}^{\alpha} \gamma_k = 1\right\}$	set of probability vectors in $I\!R^\alpha$
$P_n\{\omega\} \doteq \prod_{j=1}^{n} \rho\{\omega_j\}$	probability of $\omega \in \Omega_n$
$P_n\{B\} \doteq \sum_{\omega \in B} P_n\{\omega\}$ for $B \subset \Omega_n$	probability of an event B
$X_j(\omega) \doteq \omega_j, j = 1, 2, \ldots, n$	i.i.d. random variables with dist. ρ

Examples. (a) Coin tossing: $\Lambda \doteq \{1, 2\}$, $\rho_1 = \rho_2 \doteq 1/2$.

(b) Dice tossing: $\Lambda \doteq \{1, 2, \ldots, 6\}$, each $\rho_k \doteq 1/6$.

(c) Ideal gas: $\Lambda \doteq \{y_1, \ldots, y_\alpha\}$, the discretized energies of the individual molecules; each $\rho_k \doteq 1/\alpha$.

—(4)—
Gibbs State

Definition. Let β be a positive parameter proportional to the inverse absolute temperature. Assume that the "energy" of a configuration $\omega \in \Omega_n$ is defined by a function $H_n(\omega)$ called the *Hamiltonian*. Then the *Gibbs canonical ensemble*, or the *Gibbs state*, is the probability measure $P_{n,\beta}$ on Ω_n defined by

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

where $Z_n(\beta)$ is the normalization factor that makes $P_{n,\beta}$ a probability measure: $\sum_{\omega\in\Omega_n} P_{n,\beta}\{\omega\} = 1$. Thus

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

We call $Z_n(\beta)$ the *partition function*. For $B \subset \Omega_n$ define

$$P_{n,\beta}\{B\} \doteq \sum_{\omega\in B} P_{n,\beta}\{\omega\}.$$

Example. *Ideal Gas.* Define $H_n(\omega) \doteq \sum_{j=1}^{n} \omega_j$. Then $P_{n,\beta}$ is product measure:

$$\begin{aligned} P_{n,\beta}\{\omega\} &\doteq \exp\left[-\beta \sum_{j=1}^{n} \omega_j\right] \prod_{j=1}^{n} \rho\{\omega_j\} \cdot \frac{1}{Z_n(\beta)} \\ &= \prod_{j=1}^{n} \left(\exp[-\beta\omega_j]\, \rho\{\omega_j\} \cdot \frac{1}{\sum_{y_k\in\Lambda} \exp[-\beta y_k]\, \rho_k} \right) \end{aligned}$$

—(5)—

Boltzmann's Discovery

For $\omega \in \Omega_n$ and $y \in \Lambda$ define

$$L_n(y) = L_n(\omega, y) \doteq \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}\{y\} = \frac{1}{n} \cdot \#\{j \in \{1, \ldots, n\} : \omega_j = y\}.$$

$L_n = L_n(\omega) \doteq (L_n(\omega, y_1), L_n(\omega, y_2), \ldots, L_n(\omega, y_\alpha))$ takes values in $\mathcal{P}_\alpha$:

$$L_n(y_k) \geq 0 \text{ and } \sum_{k=1}^{\alpha} L_n(y_k) = 1.$$

Example. Let $\Lambda \doteq \{-1, 0, 1\}$, $n \doteq 6$, $\omega \doteq (-1, 1, 0, -1, -1, 0)$. Then

$$L_n(\omega) = \frac{1}{6}(3, 2, 1).$$

Definition. *Relative Entropy.* Recall $\rho = (\rho_1, \ldots, \rho_\alpha) \in \mathcal{P}_\alpha$, $\rho > 0$. For any $\gamma \in \mathcal{P}_\alpha$ define the relative entropy of γ w.r.t. ρ by

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

$I_\rho(\gamma)$ measures the discrepency between γ and ρ:
$I_\rho(\gamma) \geq 0$ and $I_\rho(\gamma) = 0 \Leftrightarrow \gamma = \rho$. In addition, $I_\rho(\gamma)$ is strictly convex.

Boltzmann's Discovery. For $\gamma \in \mathcal{P}_\alpha$ and $n \to \infty$

$$P_n\{\omega \in \Omega_n : L_n(\omega) \sim \gamma\} \approx \exp[-nI_\rho(\gamma)].$$

Both statistical mechanics and large deviations start with this asymptotic result.

—(6)—

Proof of Boltzmann's Discovery

Boltzmann's Discovery. For $\gamma \in \mathcal{P}_\alpha$ and $n \to \infty$

$$P_n\{\omega \in \Omega_n : L_n(\omega) \sim \gamma\} \approx \exp[-nI_\rho(\gamma)].$$

Heuristic Proof. Combinatorics:

$$\begin{aligned} P_n\{L_n \sim \gamma\} &= P_n\Big\{L_n \sim \frac{1}{n}(n\gamma_1, n\gamma_2, \ldots, n\gamma_\alpha)\Big\} \\ &= P_n\{\#\{\omega_j\text{'s} = y_1\} \sim n\gamma_1, \ldots, \#\{\omega_j\text{'s} = y_\alpha\} \sim n\gamma_\alpha\} \\ &\approx \frac{n!}{(n\gamma_1)!(n\gamma_2)!\cdots(n\gamma_\alpha)!}\,\rho_1^{n\gamma_1}\rho_2^{n\gamma_2}\cdots\rho_\alpha^{n\gamma_\alpha}. \end{aligned}$$

Stirling's formula:

$$\begin{aligned} \frac{1}{n}\log P_n\{L_n \sim \gamma\} &\approx \frac{1}{n}\log\left(\frac{n!}{(n\gamma_1)!(n\gamma_2)!\cdots(n\gamma_\alpha)!}\right) + \sum_{k=1}^{\alpha}\gamma_k \log\rho_k \\ &= -\sum_{k=1}^{\alpha}\gamma_k\log\gamma_k + \mathrm{O}\Big(\frac{\log n}{n}\Big) + \sum_{k=1}^{\alpha}\gamma_k\log\rho_k \\ &= -\sum_{k=1}^{\alpha}\gamma_k\log\Big(\frac{\gamma_k}{\rho_k}\Big) + \mathrm{O}\Big(\frac{\log n}{n}\Big) \\ &= -I_\rho(\gamma) + \mathrm{O}\Big(\frac{\log n}{n}\Big). \end{aligned}$$

This completes the proof. ■

—(7)—

Large Deviation Principle for L_n

For $B \subset \mathcal{P}_\alpha$ define $I_\rho(B) \doteq \inf_{\gamma \in B} I_\rho(\gamma)$. Since

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

it is reasonable to expect that

$$\lim_{n\to\infty} \frac{1}{n} \log P_n\{L_n \in B\} = \sup_{\gamma \in B}(-I_\rho(\gamma)) = -\inf_{\gamma \in B} I_\rho(\gamma) = -I_\rho(B);$$

i.e.,

$$P_n\{L_n \in B\} \approx \exp[-nI_\rho(B)].$$

This is true for a large class of sets B. Follows from:

Sanov's Theorem. The sequence of random probability vectors $\{L_n, n \in I\!N\}$ satisfies the large deviation principle (LDP) on $\mathcal{P}_\alpha$ with rate function $I_\rho(\gamma)$ given by the relative entropy of γ w.r.t. ρ:
(a) For any closed subset F of $\mathcal{P}_\alpha$

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

(b) For any open subset G of $\mathcal{P}_\alpha$

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

Corollary. Let B be any Borel subset of $\mathcal{P}_\alpha$ such that $\overline{\text{int } B} = \overline{B}$ (e.g., B an open ball $B(\gamma, \varepsilon)$ with center $\gamma \in \mathcal{P}_\alpha$ and radius $\varepsilon > 0$). Then

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

—(8)—

Consequences of the Large Deviation Principle

Recall that $I_\rho(\gamma) > I_\rho(\rho) = 0$ for all $\gamma \neq \rho$.

- If B is any Borel subset of $\mathcal{P}_\alpha$ such that the closure of B does not contain ρ, then $I_\rho(B) > 0$ and $P_n\{L_n \in B\} \leq \exp[-nI_\rho(B)/2] \to 0$ as $n \to \infty$.
- For any $\varepsilon > 0$ $P_n\{L_n \in B(\rho, \varepsilon)\} \to 1$ as $n \to \infty$.

The last limit is the law of large numbers for L_n. It describes a "P_n-typical" configuration in $\mathcal{P}_\alpha$ for L_n:

$$\begin{aligned} P_n\{L_n \in B(\rho,\varepsilon)\} &\approx P_n\{L_n \sim \rho\} \\ &= P_n\{n\rho_1\, \omega_j\text{'s} \sim y_1, \ldots, n\rho_\alpha\, \omega_j\text{'s} \sim y_\alpha\} \approx 1. \end{aligned}$$

We call ρ the equilibrium value of L_n.

Maximum Entropy Principle 1. Suppose that each $\rho_k = 1/\alpha$. Then

$$\begin{aligned} &\gamma = \gamma_0 \text{ is the equilibrium value} \\ &\quad \Longleftrightarrow \gamma_0 \text{ minimizes } I_\rho(\gamma) = \log\alpha + \sum_{k=1}^{\alpha} \gamma_k \log \gamma_k \\ &\quad \Longleftrightarrow \gamma_0 \text{ maximizes } R(\gamma) \doteq -\sum_{k=1}^{\alpha} \gamma_k \log \gamma_k. \end{aligned}$$

—(9)—

A Conditioned Limit Theorem for L_n

Define

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

and

$$\bar{y} \doteq \sum_{k=1}^{\alpha} y_k \rho_k = \int_\Lambda X_1 \, d\rho.$$

For $a > 0$ fix a closed interval $[z, z+a] \subset [y_1, y_\alpha]$.

Question. Given that $S_n/n \in [z, z+a]$, what is the most likely configuration of L_n?

If $\bar{y} \in (z, z+a)$, then the answer is ρ. So assume that $[z, z+a] \subset (\bar{y}, y_\alpha]$.

Theorem. There exists a unique $\gamma^* \in \mathcal{P}_\alpha$ such that when conditioned on $S_n/n \in [z, z+a]$, the most likely configuration of L_n is γ^*. More precisely, for every $\varepsilon > 0$

$$\lim_{n\to\infty} P_n\{L_n \in B(\gamma^*, \varepsilon) | S_n/n \in [z, z+a]\} = 1.$$

$\gamma^* = (\gamma_1^*, \gamma_2^*, \ldots, \gamma_\alpha^*)$ has the form

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

where $\beta = \beta(z) \in I\!R$ is chosen so that $\sum_{k=1}^{\alpha} y_k \gamma_k^* = z$.

Maximum Entropy Principle for γ^*

Idea of the Proof. Define $A(z) \subset \mathcal{P}_\alpha$ by

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

Then

$$\{\omega \in \Omega_n : S_n(\omega) \in [z, z+a]\} = \{L_n(\omega) \in A(z)\}.$$

By Sanov's Theorem, for $B \subset \mathcal{P}_\alpha$

$$P_n\{L_n \in B\} \approx \exp[-nI_\rho(B)].$$

Hence

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

This is of order 1 provided

$$I_\rho(B(\gamma^*, \varepsilon) \cap A(z)) = I_\rho(A(z)).$$

Choose $\gamma^* \in A(z)$ such that $I_\rho(\gamma^*) = I_\rho(A(z))$. If $\rho_k = 1/\alpha$ for each k, then γ^* can be characterized by the following.

Maximum Entropy Principle 2. Conditioned on $S_n/n \in [z, z+a]$, the most likely configuration of L_n is γ^*, which is the unique $\gamma \in \mathcal{P}_\alpha$ that maximizes

$$R(\gamma) = -\sum_{k=1}^{\alpha} \gamma_k \log \gamma_k$$

subject to the constraint that $\gamma \in A(z)$.

—(11)—
Derivation of the Gibbs State for the Ideal Gas

Define the probability measure on Λ

$$\gamma^* \doteq \sum_{k=1}^{\alpha} \gamma_k^* \delta_{y_k}, \text{where } \gamma_k^* \doteq \exp[-\beta y_k]\, \rho_k \cdot \frac{1}{\sum_{j=1}^{\alpha} \exp[-\beta y_j]\, \rho_j};$$

$\beta = \beta(z)$ is chosen so that $\sum_{k=1}^{\alpha} y_k \gamma_k^* = z$. For $s \in I\!N$ rewrite the Gibbs canonical ensemble for the ideal gas consisting of s molecules:

$$\begin{aligned} P_{s,\beta}\{\omega\} &\doteq \exp\left[-\beta \sum_{j=1}^{s} \omega_j\right] \prod_{j=1}^{s} \rho\{\omega_j\} \cdot \frac{1}{Z_n(\beta)} \\ &= \prod_{j=1}^{s} \left(\exp[-\beta\omega_j]\, \rho\{\omega_j\} \cdot \frac{1}{\sum_{\omega_k \in \Lambda} \exp[-\beta\omega_k]\, \rho\{\omega_k\}} \right) \\ &= \prod_{j=1}^{s} \gamma^*\{\omega_j\} \text{ for } \omega \in \Omega_s \doteq \Lambda^s. \end{aligned}$$

Theorem. *Equivalence of ensembles.* Fix $s \in I\!N$ and $[z, z+a] \in (\bar{y}, y_\alpha]$. Then conditioned on

$$\frac{1}{n} S_n(\omega) = \frac{1}{n} \sum_{j=1}^{n} X_j(\omega) = \frac{1}{n} \sum_{j=1}^{n} \omega_j \in [z, z+a],$$

as $n \to \infty$ the marginal distribution of $(X_1, X_2, \ldots, X_s)$ w.r.t. P_n (Gibbs microcanonical ensemble) equals $P_{s,\beta}$, the Gibbs canonical ensemble. In other words, for subsets $B_1, B_2, \ldots, B_s$ of Λ

$$\lim_{n\to\infty} P_n\{\omega \in \Omega_n : X_1 \in B_1, X_2 \in B_2, \ldots, X_s \in B_s | S_n/n \in [z, z+a]\}$$
$$= P_{s,\beta}\{B_1 \times B_2 \times \cdots \times B_s\}.$$

Proof. $P_n\{L_n \in B\} \approx \exp[-n I_\rho(B)]$ and generalizations.

—(12)—

Definition of the Large Deviation Principle

For each $n \in I\!N$ let $(\Omega_n, \mathcal{F}_n, P_n)$ be a probability space.

- $\Omega_n = \Lambda^n$, where $\Lambda \doteq \{y_1, y_2, \ldots, y_\alpha\}$; $\mathcal{F}_n$ the set of all subsets of Ω_n;

$$P_n\{\omega\} \doteq \prod_{k=1}^{n} \rho\{\omega_k\} \text{ for } \omega \in \Omega_n.$$

Let $\mathcal{X}$ be a complete separable metric space (Polish space).

1. $\mathcal{X} = \mathcal{P}_\alpha$, the set of probability vectors in $I\!R^\alpha$.
2. $\mathcal{X}$ the closed bounded interval $[y_1, y_\alpha]$.

Let $\{Z_n, n \in I\!N\}$ be a sequence of random variables mapping $\{\Omega_n, n \in I\!N\}$ into $\mathcal{X}$.

1. With $\mathcal{X} = \mathcal{P}_\alpha$, let $Z_n = L_n$.
2. With $\mathcal{X} = [y_1, y_\alpha]$, let $Z_n = \dfrac{1}{n} \sum_{j=1}^{n} X_j$, where $X_j(\omega) = \omega_j$ for $\omega \in \Omega_n = \Lambda^n$.

A function $I : \mathcal{X} \mapsto [0, \infty]$ is called a *rate function* if for all $M < \infty$ $\{x \in \mathcal{X} : I(x) \leq M\}$ is compact (a technical regularity condition).

1. With $\mathcal{X} = \mathcal{P}_\alpha$, let $I(\gamma) = I_\rho(\gamma)$, the relative entropy of γ w.r.t. ρ.
2. With $\mathcal{X} = [y_1, y_\alpha]$, let $I : [y_1, y_\alpha] \mapsto [0, \infty)$ be continuous.

Let $\{Z_n, n \in I\!N\}$ be sequence of random variables mapping $\{(\Omega_n, \mathcal{F}_n, P_n), n \in I\!N\}$ into a Polish space $\mathcal{X}$. Let I be a rate function on $\mathcal{X}$. Then $\{Z_n\}$ satisfies the *large deviation principle* or *LDP* on $\mathcal{X}$ if for all closed subsets F of $\mathcal{X}$

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

and for all open subsets G of $\mathcal{X}$

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

We write the LDP as

$$P_n\{Z_n \in dx\} \asymp \exp[-nI(x)]\, dx.$$

—(13)—
Examples of LDP's

1. Sanov's Theorem for $\{L_n, n \in I\!N\}$:

$$P_n\{L_n \in d\gamma\} \asymp \exp[-n\, I_\rho(\gamma)]\, d\gamma \ \text{ on } \mathcal{P}_\alpha.$$

2. Let $\Lambda \doteq \{-1,1\}$, $\rho \doteq \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, $S_n(\omega) \doteq \sum_{j=1}^{n} \omega_j$, $\mathcal{X} \doteq [-1,1]$.
For $k \in \{-n, -n+1, \ldots, 0, \ldots, n\}$

$$\begin{aligned}
&P_n\{S_n/n = k/n\} \\
&\quad = P_n\{L_n(-1) = \tfrac{1}{2}(1-k/n), L_n(1) = \tfrac{1}{2}(1+k/n)\} \\
&\quad \asymp \exp[-nI_\rho(\tfrac{1}{2}(1-k/n),\tfrac{1}{2}(1+k/n))] \\
&\quad = \exp[-nI(k/n)],
\end{aligned}$$

where

$$I(k/n) \doteq \tfrac{1}{2}(1-k/n)\log(1-k/n) + \tfrac{1}{2}(1+k/n)\log(1+k/n).$$

Cramér's Theorem. $P_n\{S_n/n \in dx\} \asymp \exp[-nI(x)]\, dx$ on $[-1,1]$,
where

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

—(14)—

The LDP Is Equivalent to the Laplace Principle

$\{Z_n\}$ satisfies the large deviation principle on $\mathcal{X}$ with rate function I if and only if $\{Z_n\}$ satisfies the *Laplace principle* on $\mathcal{X}$ with rate function I: i.e., for all bounded continuous $h : \mathcal{X} \mapsto I\!R$

$$\lim_{n\to\infty} \frac{1}{n} \log E_n\{\exp[n\, h(Z_n)]\} = \sup_{x\in\mathcal{X}}\{h(x) - I(x)\}.$$

Heuristically, if we substitute $P_n\{Z_n \in dx\} \asymp \exp[-nI(x)]\, dx$, then

$$\begin{aligned}
\frac{1}{n} \log E_n\{\exp[n\, h(Z_n)]\} &= \frac{1}{n} \log \int_{\mathcal{X}} \exp[n\, h(x)]\, P_n\{Z_n \in dx\} \\
&\approx \frac{1}{n} \log \int_{\mathcal{X}} \exp[n\, h(x)]\, \exp[-n\, I(x)]\, dx \\
&\to \sup_{x\in\mathcal{X}}\{h(x) - I(x)\}.
\end{aligned}$$

A proof of the equivalence of the LDP and the Laplace principle is given in P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, 1997. In statistical mechanics the Laplace principle gives a variational formula for the specific Gibbs free energy.

—(15)—

Curie-Weiss Model

3 Levels of Large Deviations for I.I.D. Random Variables and Statistical Mechanical Models

Level	*Process*	*LD Theorem*	*Model*
1	sample mean	Cramér	Curie-Weiss
2	empirical measure	Sanov	Curie-Weiss-Potts
3	empirical field	D-V, F-O, O	Ising

The Curie-Weiss model is a spin system on $\Delta_n \doteq \{1, 2, \ldots, n\}$. For $\omega = \{\omega_i, i \in \Delta_n\}$ in the configuration space $\Omega_n \doteq \{1, -1\}^{\Delta_n}$, we define $P_n\{\omega\} \doteq 1/2^n$. Thus

$$P_n(d\omega) \doteq \prod_{i \in \Lambda_n} \rho(d\omega_i), \;\; \rho \doteq \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

The *Hamiltonian*, or energy of a configuration, is defined by

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

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

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

where $Z_n(\beta)$ is the *finite-volume partition function*

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

$P_{n,\beta}$ models a ferromagnet: aligned configurations have larger probability.

——(16)——

LDP and LLN for the Curie-Weiss Model

Our goal is to evaluate

$$\lim_{n\to\infty} P_{n,\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i \in dx\right\}.$$

Method: first derive LDP.

With $\rho \doteq \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, define for $x \in [-1,1]$

$$\begin{aligned} I(x) &\doteq \sup_{\alpha\in\mathbb{R}}\left\{\alpha x - \log\int_{\mathbb{R}} e^{\alpha y}\rho(dy)\right\} \\ &= \frac{1-x}{2}\log(1-x) + \frac{1+x}{2}\log(1+x). \end{aligned}$$

Since by Cramér's Theorem

$$P_n\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i \in dx\right\} \asymp \exp[-n\,I(x)]\,dx,$$

the Laplace principle yields

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\log Z_n(\beta) &= \lim_{n\to\infty}\frac{1}{n}\log\int_{\Omega_n}\exp[-\beta H_n(\omega)]\,P_n(d\omega) \\ &= \lim_{n\to\infty}\frac{1}{n}\log\int_{\Omega_n}\exp\left[\frac{\beta n}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\omega_i\right)^2\right]\,P_n(d\omega) \\ &= \lim_{n\to\infty}\frac{1}{n}\log\int_{[-1,1]}\exp\left[n\frac{\beta}{2}y^2\right]\,P_n\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i \in dy\right\} \\ &= \sup_{y\in[-1,1]}\left\{\frac{\beta}{2}y^2 - I(y)\right\} \;=\; -\inf_{y\in[-1,1]}\left\{I(y) - \frac{\beta}{2}y^2\right\}. \end{aligned}$$

Thus

$$\frac{1}{Z_n(\beta)} \asymp \exp\left[n\cdot\inf_{y\in[-1,1]}\left\{I(y)-\frac{\beta}{2}y^2\right\}\right].$$

Hence

$$\begin{aligned}
&P_{n,\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i \in dx\right\} \\
&\quad = \frac{1}{Z_n(\beta)} \cdot \exp\left[\frac{\beta n}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\omega_i\right)^2\right] \cdot P_n\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i \in dx\right\} \\
&\quad \asymp \frac{1}{Z_n(\beta)} \cdot \exp\left[\frac{n\beta}{2}x^2\right] \cdot \exp[-n\, I(x)]\, dx \\
&\quad \asymp \exp[-n\, I_\beta(x)]\, dx,
\end{aligned}$$

where

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

For $0 < \beta \le 1$ $I_\beta(x)$ attains its infimum of 0 at $x = 0$, while for $\beta > 1$ $\exists\, m(\beta) \in (0,1)$ such that $I_\beta(x)$ attains its infimum of 0 at $\pm m(\beta)$. Hence

$$P_{n,\beta}\left\{\frac{1}{n}\sum_{i=1}^{n}\omega_i \in dx\right\} \Longrightarrow \begin{cases} \delta_0 & \text{if} \quad 0 < \beta \le 1 \\ \frac{1}{2}\delta_{m(\beta)} + \frac{1}{2}\delta_{-m(\beta)} & \text{if} \quad \beta > 1. \end{cases}$$

We call $m(\beta)$ the *spontaneous magnetization* for the Curie-Weiss model and $\beta_c = 1$ the *critical inverse temperature*. The phase transition at β_c is *second order*: $m(\beta) \to 0$ as $\beta \to 1^+$ but $m'(\beta) \to \infty$ as $\beta \to 1^+$; in fact, $m(\beta) \sim [3(\beta-1)]^{1/2}$ as $\beta \to 1^+$.

Moral: points x^* satisfying $I_\beta(x^*) = 0$ or $I(x^*) - \beta(x^*)^2/2 = \inf_{x\in[-1,1]}\{I(x) - \beta x^2/2\}$ are equilibrium macrostates for the model.

—(17)—

Ising Model

Let $\{\Delta_n, n \in I\!N\}$ be a sequence of expanding hypercubes of side length n in the D-dimensional integer lattice $Z\!\!Z^D$, $D \geq 2$. Thus Δ_n contains $|\Delta_n| \doteq (n+1)^D$ sites of $Z\!\!Z^D$. The Ising model is a spin system on Δ_n. For $\omega = \{\omega_i, i \in \Delta_n\}$ in the configuration space $\Omega_n \doteq \{1, -1\}^{\Delta_n}$, we define $P_n\{\omega\} \doteq 1/2^{|\Delta_n|}$. Thus

$$P_n(d\omega) \doteq \prod_{i \in \Delta_n} \rho(d\omega_i), \ \ \rho \doteq \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}.$$

The *Hamiltonian*, or energy of a configuration, is defined by

$$H_n(\omega) \doteq -\frac{1}{2} \sum_{i,j \in \Delta_n} J(i-j)\omega_i\omega_j,$$

where

$$J(i-j) \doteq \begin{cases} 1 & \text{if} \quad \|i-j\| = 1 \\ 0 & \text{if} \quad \|i-j\| \neq 1. \end{cases}$$

The probability of a configuration corresponding to inverse temperature $\beta > 0$ is defined by the finite-volume Gibbs state

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

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

Although it is much more complicated, the asymptotics of the Ising model can be analyzed as in the Curie-Weiss model via large deviations. Rather than rewrite the Hamiltonian as in the Curie-Weiss model in terms of $\Sigma_{i=1}^n \omega_i/n$, we rewrite it in terms of the empirical field. Using the LDP for the empirical field proved by Donsker-Varadhan, Föllmer-Orey, and Olla, one shows that $\lim_{n\to\infty} |\Delta_n|^{-1} \log Z_n(\beta)$ is given by a variational formula over the space of translation invariant measures on $\{1, -1\}^{Z\!\!Z^D}$. In the Curie- Weiss model, the analogous variational formula is over $[-1, 1]$.

—(18)—

A General Large Deviations Approach to Models in Statistical Mechanics

We give a general approach to the large deviation analysis of models in statistical mechanics. We consider a model that is defined in terms of the following data.

Item	*Name*	*Curie-Weiss*	*Ising*
Ω_n	configuration spaces	$\{1,-1\}^{\Delta_n}$	$\{1,-1\}^{\Delta_n}$
$H_n(\omega)$	Hamiltonian	$-\frac{n}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\omega_i\right)^2$	$-\frac{1}{2}\sum_{i,j\in\Delta_n} J(i-j)\omega_i\omega_j$
a_n	norming constants	$\lvert\Delta_n\rvert = n$	$\lvert\Delta_n\rvert = (n+1)^D$
P_n	probability measure	$\prod_{i\in\Delta_n}\rho(d\omega_i)$	$\prod_{i\in\Delta_n}\rho(d\omega_i)$

Define for $n \in I\!N$ and $\beta > 0$ the partition function

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

and the Gibbs state

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

For $\beta > 0$ we define

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

if the limit exists; $-\beta^{-1}\varphi(\beta)$ is the *specific Gibbs free energy* for the model.

Our hope: $\varphi(\beta)$ is given by a variational formula of the form $\sup_{x\in\mathcal{X}}\{\beta\Gamma(x) - I(x)\}$ for some space $\mathcal{X}$ and functions Γ and I. We would like to interpret supremizing points $x^* \in \mathcal{X}$ as equilibrium macrostates for the model.

In order to carry out a large deviation analysis of the model, the following four items are needed.

(i) **Hidden space:** a complete separable metric space $\mathcal{X}$.

(ii) **Hidden process:** random variables Y_n mapping Ω_n into $\mathcal{X}$.

(iii) **Hamiltonian representation function:** a bounded continuous function Γ mapping $\mathcal{X}$ into $I\!R$ such that

$$-H_n(\omega) = a_n\,\Gamma(Y_n(\omega)) + \mathrm{o}(a_n) \text{ uniformly for } \omega \in \Omega_n;$$

i.e.,

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

(iv) **LDP for $\boldsymbol{Y_n}$:** $P_n\{Y_n \in dx\}$ satisfies the LDP on $\mathcal{X}$ with some rate function I and norming constants a_n: $P_n\{Y_n \in dx\} \asymp \exp[-a_n\, I(x)]\, dx$.

Model	$\mathcal{X}$	Y_n	Γ	I
Curie-Weiss	$[-1,1]$	$\sum_{i=1}^{n} \omega_i/n$	$x^2/2$	Cramér
Ising	$\mathcal{P}_{\mathrm{tr}}\Big(\{1,-1\}^{\mathbb{Z}^D}\Big)$	emp. field	$\frac{1}{2}\sum_{\lVert i\rVert=1} \int \omega_0\omega_i Q(d\omega)$	D-V, F-O, O

Since

$$P_{n,\beta}(d\omega) = \frac{1}{\int_{\Omega_n} \exp[-\beta\, H_n(\omega)]\, P_n(d\omega)} \cdot \exp[-\beta\, H_n(\omega)]\, P_n(d\omega),$$

$$P_{n,\beta}\{Y_n \in dx\} \asymp \frac{1}{\int_{\Omega_n} \exp[a_n\,\beta\,\Gamma(x)]\, P_n\{Y_n \in dx\}} \cdot \exp[a_n\,\beta\,\Gamma(x)]\, P_n\{Y_n \in dx\}.$$

We can read off the LDP for $P_{n,\beta}\{Y_n \in dx\}$.

Theorem. For each $\beta > 0$ the following hold.

(a) $\varphi(\beta) \doteq \lim_{n\to\infty} \dfrac{1}{a_n} \log Z_n(\beta) = \sup_{x\in\mathcal{X}}\{\beta\,\Gamma(x) - I(x)\}$.

(b) $P_{n,\beta}\{Y_n \in dx\}$ satisfies the LDP on $\mathcal{X}$ with rate function

$$I_\beta(x) \doteq I(x) - \beta\,\Gamma(x) - \inf_{y\in\mathcal{X}}\{I(y) - \beta\,\Gamma(y)\}$$

and norming constants a_n.

(c) The set of equilibrium macrostates

$$\mathcal{E}_\beta \doteq \{x^* \in \mathcal{X} : I_\beta(x^*) = 0\} = \left\{x^* \in \mathcal{X} : I(x^*) - \beta\Gamma(x^*) = \inf_{\mathcal{X}}\{I - \beta\Gamma\}\right\}$$

is a nonempty compact subset of $\mathcal{X}$. If $A \in \mathcal{B}(\mathcal{X})$ satisfies $\bar{A} \cap \mathcal{E}_\beta = \emptyset$, then

$$\lim_{n\to\infty} P_{n,\beta}\{Y_n \in A\} = 0.$$

(d) If $|\mathcal{E}_\beta| = \{\tilde{x}\}$ (absence of a phase transition), then

$$P_{n,\beta}\{Y_n \in dx\} \Rightarrow \delta_{\tilde{x}}(dx).$$

If $|\mathcal{E}_\beta| \geq 2$, then along a subsubsequence of any subsequence

$$P_{n',\beta}\{Y_{n'} \in dx\} \Rightarrow \Pi_\beta(dx), \text{ where } \Pi_\beta\{\mathcal{E}_\beta\} = 1.$$

LDP for $P_n\{Y_n \in dx\}$ and the Laplace principle give (a)-(b): for any $f \in \mathcal{C}_b(\mathcal{X})$

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{a_n} \log \int_{\mathcal{X}} \exp[a_n\, f(x)]\, P_{n,\beta}\{Y_n \in dx\} \\
&\quad = \lim_{n\to\infty} \frac{1}{a_n} \log \int_{\mathcal{X}} \exp[a_n(f(x) + \beta\Gamma(x))]\, P_n\{Y_n \in dx\} \\
&\qquad - \lim_{n\to\infty} \frac{1}{a_n} \log \int_{\mathcal{X}} \exp[a_n\, \beta\,\Gamma(x)]\, P_n\{Y_n \in dx\} \\
&\quad = \sup_{x\in\mathcal{X}}\{f(x) + \beta\Gamma(x) - I(x)\} - \sup_{y\in\mathcal{X}}\{\beta\Gamma(y) - I(y)\} \\
&\quad = \sup_{x\in\mathcal{X}}\{f(x) - I_\beta(x)\}.
\end{aligned}$$

—(19)—

Maximum Entropy Principles in Two-Dimensional Turbulence

Our aim is to use Gibbs states to predict the large-scale, long-lived order of coherent vortices that persist amidst the turbulent fluctuations of the vorticity field. We do this by applying a statistical equilibrium theory of the two-dimensional Euler equations, which govern the motion of an inviscid, incompressible fluid. These equations are an infinite-dimensional Hamiltonian system with Hamiltonian H having a family of other conserved quantities called generalized enstrophies.

This statistical equilibrium theory is defined via continuum limits of Gibbs measures of two lattice models for these continuum equations (the Miller-Robert model and the Turkington model). These models differ in how they discretize the continuum dynamics. The first model of this kind was due to Onsager, who studied point vortices and predicted that the equilibrium states with high enough energy have a negative temperature and represent large-scale, coherent vortices. This model was further developed in the 1970's, notably by Montgomery and Joyce. However, the point vortex model fails to incorporate all the conserved quantities for two-dimensional ideal flow.

A model that respects conservation of energy and also the generalized enstrophy constraints was developed by Miller (1990) and Robert (1991). A related model, which discretizes the continuum dynamics in a different way, was developed by Turkington (1998). These authors use formal arguments to derive maximum entropy principles that are equivalent to variational formulas for the equilibrium macrostates. In terms of these macrostates, coherent vortices of two-dimensional turbulence can be studied.

We apply large deviation theory to give a rigorous derivation of these variational formulas. References (2) and (5) discuss in detail the physical background of these models.

References.

(1) C. Boucher, R. S. Ellis, and B. Turkington. Spatializing random measures: doubly indexed processes and the large deviation principle. Submitted to *Ann. Probab.*, 1998.

(2) C. Boucher, R. S. Ellis, and B. Turkington. Derivation of maximum entropy principles in two-dimensional turbulence via large deviations. Submitted to *Commun. Math. Phys.*, 1998.

(3) C. Boucher, R. S. Ellis, and B. Turkington. Use of two-parameter processes to prove large deviation principles. Technical report, Univ. of Mass., 1998.

(4) R. S. Ellis. An overview of the theory of large deviations and applications to statistical mechanics. *Scand. Actuarial J.* Number 1:97-142, 1995.

(5) B. Turkington. Statistical equilibrium measures and coherent states in two-dimensional turbulence. To appear in *Commun. Math. Phys.*, 1998.

—(20)—

A Model of Two-Dimensional Turbulence

$T^2 \doteq$ the unit square $[0,1) \times [0,1)$ with periodic boundary conditions.

$g(x - x')$ the Green's function for $-\triangle$ on T^2.

$\mathcal{L}$ a uniform lattice of $n \doteq 2^{2z}$ sites in T^2, $z \in I\!N$.

Configuration space $\Omega_n \doteq \mathcal{Y}^n$, where $\mathcal{Y}$ is a compact subset of $I\!R$. Elements of Ω_n are denoted by a *vorticity field* $\zeta = \{\zeta(s), s \in \mathcal{L}\}$.

$P_n(d\zeta) \doteq \prod_{s \in \mathcal{L}} \rho(d\zeta(s))$, where ρ is a probability measure on $\mathcal{Y}$.

Hamiltonian mapping Ω_n into $I\!R$: $H_n(\zeta) \doteq \dfrac{1}{2n^2} \sum_{\substack{s,s' \in \mathcal{L} \\ s \neq s'}} g(s - s')\zeta(s)\zeta(s')$.

Continuous function $a : \mathcal{Y} \mapsto I\!R$.

Generalized enstrophy mapping Ω_n into $I\!R$: $A_{n,a}(\zeta) \doteq \dfrac{1}{n} \sum_{s \in \mathcal{L}} a(\zeta(s))$.

Partition function $Z(n, \beta, a) \doteq \int_{\Omega_n} \exp[-\beta H_n(\zeta) - A_{n,a}(\zeta)]\, P_n(d\zeta)$.

Gibbs state on Ω_n: $P_{n,\beta,a}(d\zeta) \doteq \dfrac{1}{Z(n,\beta,a)} \exp[-\beta H_n(\zeta) - A_{n,a}(\zeta)]\, P_n(d\zeta)$.

Goal: study the continuum limit obtained by sending $n \to \infty$ along the sequence $n = 2^{2z}$, $z \in I\!N$. We hope to evaluate $\varphi(\beta, a) \doteq \lim_{n \to \infty} \dfrac{1}{n} \log Z(n, n\beta, na)$ in terms of a variational formula over a hidden space.

—(21)— Large Deviation Analysis of the Model

In order to study the continuum limit of the model, we introduce macrocells and microcells in $\mathcal{L}$.

For even $r < 2z$ there are 2^r "macrocells" $\{D_{r,k}, k = 1, 2, \ldots, 2^r\}$, each macrocell containing $n/2^r$ lattice sites. Specifically,

$$D_{r,k} = [(i-1)/2^{r/2}, i/2^{r/2}) \times [(j-1)/2^{r/2}, j/2^{r/2}) \text{ for } i, j \in \{1, 2, \ldots, 2^{r/2}\}.$$

Each $D_{r,k}$ is partitioned dyadically into $n/2^r$ squares, called "microcells," each having area $1/n$ and containing one site of $\mathcal{L}$. For $s \in \mathcal{L}$ $M(s)$ denotes the unique microcell containing the site s.

We must find representation functions both for the Hamiltonian and for the generalized enstrophy.

(i) **Hidden space.** The space $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ of probability measures on $T^2 \times \mathcal{Y}$ with first marginal θ, where $\theta(dx) = dx$ is Lebesgue measure on T^2.

(ii) **Hidden process.** $Y_n : \Omega_n \mapsto \mathcal{P}_\theta(T^2 \times \mathcal{Y})$:

$$Y_n(dx \times dy) = Y_n(\zeta, dx \times dy) \doteq dx \otimes \sum_{s \in \mathcal{L}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy),$$

(iii) **Hamiltonian representation function.** $\tilde{H} : \mathcal{P}_\theta(T^2 \times \mathcal{Y}) \mapsto I\!R$:

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

$\tilde{H}$ is bounded and continuous and $\exists\, C < \infty$ such that

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

(iv) **Generalized enstrophy representation function.** $\tilde{A}_a : \mathcal{P}_\theta(T^2 \times \mathcal{Y}) \mapsto I\!R$:

$$\tilde{A}_a(\mu) \doteq \int_{T^2 \times \mathcal{Y}} a(y)\, \mu(dx \times dy) = \int_{\mathcal{Y}} a(y)\, \mu_2(dy),$$

where μ_2 denotes the second marginal of μ. $\tilde{A}_a$ is bounded and continuous and

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

(v) **LDP for $\boldsymbol{Y_n}$.** $P_n\{Y_n \in d\mu\}$ satisfies the LDP on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ with rate function the relative entropy

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

With $\mathcal{P}_\theta(T^2 \times \mathcal{Y}) \doteq \{\mu \in \mathcal{P}(T^2 \times \mathcal{Y}) : \mu_1 = \theta\}$ and

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

(iii) follows with some work and (iv) is immediate. However, the LDP in (v) is far from obvious. To prove (iii), since

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

and $\theta\{M(s)\} = 1/n$, it is plausible that

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

is a good approximation to $H_n(\zeta) \doteq \dfrac{1}{2n^2} \displaystyle\sum_{\substack{s, s' \in \mathcal{L} \\ s \neq s'}} g(s - s')\zeta(s)\zeta(s')$.

Concerning (iv), for $\zeta \in \Omega_n$

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

—(22)—

Summary

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

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

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

$$P_n\{Y_n \in d\mu\} \asymp \exp[-n\, R(\mu|\theta \times \rho)]\, d\mu \quad \text{on } \mathcal{P}_\theta(T^2 \times \mathcal{Y}).$$

Thus

$$\begin{aligned} Z(n, n\beta, na) &\doteq \int_{\Omega_n} \exp[-n\beta H_n(\zeta) - nA_{n,a}(\zeta)]\, P_n(d\zeta) \\ &\approx \int_{\Omega_n} \exp[-n(\beta \tilde{H}(Y_n(\zeta, \cdot)) + \tilde{A}_a(Y_n(\zeta, \cdot)))]\, P_n(d\zeta) \\ &= \int_{\mathcal{P}_\theta(T^2 \times \mathcal{Y})} \exp[-n(\beta \tilde{H}(\mu) + \tilde{A}_a(\mu))]\, P_n\{Y_n \in d\mu\}, \end{aligned}$$

and so

$$\lim_{n \to \infty} \frac{1}{n} \log Z(n, n\beta, na) = \sup_{\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})} \{-\beta \tilde{H}(\mu) - \tilde{A}_a(\mu) - R(\mu|\theta \times \rho)\}.$$

—(23)—

Rigorous Derivation of Well Known Maximum Entropy Principles

We next state the asymptotic behavior of $Z(n, n\beta, na)$ and $P_{n,n\beta,na}\{Y_n \in d\mu\}$. In order to apply the theory of large deviations, we must scale β and a by n.

Theorem (Boucher, Ellis, Turkington). For each $\beta \in I\!R$ and $a \in \mathcal{C}_b(\mathcal{Y})$ the following hold.

(a) $$\varphi(\beta, a) \doteq \lim_{n\to\infty} \frac{1}{n} \log Z(n, n\beta, na) = \sup_{\mu\in\mathcal{P}_\theta(T^2\times\mathcal{Y})} \left\{-\beta\tilde{H}(\mu) - \tilde{A}_a(\mu) - R(\mu|\theta\times\rho)\right\}.$$

(b) $P_{n,n\beta,na}\{Y_n \in d\mu\}$ satisfies the LDP on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ with rate function

$$\begin{aligned} I_{\beta,a}(\mu) &\doteq R(\mu|\theta\times\rho) + \beta\tilde{H}(\mu) + \tilde{A}_a(\mu) \\ &\quad - \inf_{\nu\in\mathcal{P}_\theta(T^2\times\mathcal{Y})} \left\{R(\nu|\theta\times\rho) + \beta\tilde{H}(\nu) + \tilde{A}_a(\nu)\right\}. \end{aligned}$$

(c) The set of equilibrium macrostates

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

is a nonempty compact subset of $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$. If $A \in \mathcal{B}(\mathcal{P}_\theta(T^2 \times \mathcal{Y}))$ satisfies $\bar{A} \cap \mathcal{E}_{\beta,a} = \emptyset$, then

$$\lim_{n\to\infty} P_{n,n\beta,na}\{Y_n \in A\} = 0.$$

—(24)—

Physical Significance of the Theorem

The asymptotic theorem on the preceding transparency gives a rigorous derivation of maximum entropy principles for two dimensional turbulence derived by formal arguments by J. Miller (1990) and R. Robert (1991) and in somewhat modified form by B. Turkington (1998). According to these authors, equilibrium macrostates are measures $\mu^* \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ which maximize $-R(\mu|\theta \times \rho)$ and satisfy certain energy and enstrophy constraints. This is equivalent to μ^* giving the supremum in the variational formula for the Gibbs free energy $\varphi(\beta, a)$ in part (a) of the theorem:

$$\begin{aligned}\varphi(\beta, a) &\doteq \lim_{n\to\infty} \frac{1}{n} \log Z(n, n\beta, na) \\ &= \sup_{\mu \in \mathcal{P}_\theta(T^2\times\mathcal{Y})} \left\{ -\beta \tilde{H}(\mu) - \tilde{A}_a(\mu) - R(\mu|\theta \times \rho) \right\}.\end{aligned}$$

Writing an equilbrium macrostate μ^* in the form

$$\mu^*(dx \times dy) = \theta(dx) \otimes \tau^*(x, dy),$$

we define the *mean vorticity field*

$$\bar{\zeta}(x) \doteq \int_{\mathcal{Y}} y\, \tau^*(x, dy).$$

The large-scale, long-lived order of the vorticity field can be studied in terms of $\bar{\zeta}(x)$. For examples, see B. Turkington and N. Whitaker, *Siam J. Sci. Comput.* 17:1414-1433, 1996.

LDP for $\boldsymbol{Y_n}$

We prove the LDP for

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

by an innovative technique. We approximate Y_n by a doubly indexed sequence of random measures $W_{n,r}$ for which the LDP is, at least formally, almost obvious.

$\mathcal{L}$ contains 2^{2z} sites. For even $r < 2z$ we introduce 2^r macrocells $\{D_{r,k}, k = 1, 2, \ldots, 2^r\}$, each macrocell containing $n/2^r$ lattice sites. Specifically,

$D_{r,k} = [(i-1)/2^{r/2}, i/2^{r/2}) \times [(j-1)/2^{r/2}, j/2^{r/2})$ for $i, j \in \{1, 2, \ldots, 2^{r/2}\}$.

Each $D_{r,k}$ is partitioned dyadically into $n/2^r$ squares, called "microcells," each having area $1/n$ and containing one site of $\mathcal{L}$. For $s \in \mathcal{L}$ $M(s)$ denotes the unique microcell containing the site s. We now define

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

Write

$$Y_n(\zeta, dx \times dy) \doteq dx \otimes \sum_{k=1}^{2^r} \sum_{s \in D_{r,k}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy).$$

$W_{n,r}$ is obtained by replacing, for each $s \in D_{r,k}$, $\delta_{\zeta(s)}$ by $(n/2^r)^{-1} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}$.

Example: $r = 2$, $z = 3$,
$2^{2z} = 64$ sites in $\mathcal{L}$

$\boldsymbol{W_{n,r}}$ Is a Good Approximation to $\boldsymbol{Y_n}$

Lemma. $d(Y_n, W_{n,r}) \leq \sqrt{2}/2^{r/2}$, where d denotes the dual-bounded-Lipschitz metric on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$.

Comments on the Proof. In Y_n introduce, for each $k \in \{1, 2, \ldots, 2^r\}$, a sum over the $n/2^r$ sites $s' \in D_{r,k}$. This yields

$$\begin{aligned} Y_n(dx \times dy) &= dx \otimes \sum_{s \in \mathcal{L}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy) \\ &= dx \otimes \sum_{k=1}^{2^r} \sum_{s \in D_{r,k}} \frac{1}{n/2^r} \sum_{s' \in D_{r,k}} 1_{M(s)}(x)\, \delta_{\zeta(s)}(dy). \end{aligned}$$

Since $D_{r,k} = \cup_{s' \in D_{r,k}} M(s')$,

$$\begin{aligned} W_{n,r}(dx \times dy) &= dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \frac{1}{n/2^r} \sum_{s \in D_{r,k}} \delta_{\zeta(s)}(dy) \\ &= dx \otimes \sum_{k=1}^{2^r} \sum_{s \in D_{r,k}} \frac{1}{n/2^r} \sum_{s' \in D_{r,k}} 1_{M(s')}(x)\, \delta_{\zeta(s)}(dy). \end{aligned}$$

Thus Y_n and $W_{n,r}$ differ, for each $s \in D_{r,k}$, only in the indicator functions $1_{M(s)}(x)$ versus $1_{M(s')}(x)$. The fact that the diameter of $D_{r,k}$ equals $\sqrt{2}/2^{r/2}$ is used to complete the proof.

LDP for $W_{n,r}$

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

Theorem. $P_n\{W_{n,r} \in d\mu\}$ satisfies the two parameter LDP on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ ($n \to \infty$, then $r \to \infty$) with rate function $R(\mu|\theta \times \rho)$ and scaling constants n.

Since $d(Y_n, W_{n,r}) \leq \sqrt{2}/2^{r/2}$, the LDP for $W_{n,r}$ implies the LDP for Y_n. We now motivate the LDP for $W_{n,r}$. For $k \in \{1, \ldots, 2^r\}$ define the empirical measures

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

where $\{\zeta(s), s \in \mathcal{L}\}$ is the i.i.d. vorticity field with distribution ρ. $\{L_{n,r,k}, k = 1, \ldots, 2^r\}$ is also i.i.d. Since $D_{r,k}$ contains $n/2^r$ lattice sites s, $L_{n,r,k}$ takes values in $\mathcal{P}(\mathcal{Y})$, and

$$W_{n,r}(dx \times dy) = dx \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x) \, L_{n,r,k}(dy).$$

Sanov's Theorem. *LDP for $L_{n,r,k}$.* For each r and k, as $n \to \infty$

$$P_n\{L_{n,r,k} \in d\nu\} \asymp \exp[-(n/2^r) R(\nu|\rho)] \, d\nu \;\text{ on } \mathcal{P}(\mathcal{Y}).$$

Let $\tau_1, \ldots, \tau_{2^r}$ be probability measures on $\mathcal{Y}$ and take $\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ of the form

$$\mu(dx \times dy) = dx \otimes \tau(x, dy), \text{ where } \tau(x, dy) \doteq \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \tau_k(dy).$$

Since

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

Sanov's Theorem yields

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log P\{W_{n,r} \sim \mu\} \\
&\quad = \lim_{n\to\infty} \frac{1}{n} \log P\{L_{n,r,k} \sim \tau_k, k = 1, \ldots, 2^r\} \\
&\quad = \frac{1}{2^r} \sum_{k=1}^{2^r} \lim_{n\to\infty} \frac{1}{n/2^r} \log P\{L_{n,r,k} \sim \tau_k\} \\
&\quad \approx -\frac{1}{2^r} \sum_{k=1}^{2^r} R(\tau_k|\rho) \\
&\quad = -\int_{T^2} R(\tau(x, \cdot)|\rho)\, dx \\
&\quad = -R(\mu|\theta \times \rho).
\end{aligned}$$

The last equality is a consequence of the chain rule and the form of μ. The LDP for $W_{n,r}$ is certainly plausible, especially since any measure $\mu \in \mathcal{P}_\theta(T^2 \times \mathcal{Y})$ can be well approximated, as $r \to \infty$, by a sequence of measures of the form $dx \otimes \Sigma_{k=1}^{2^r} 1_{D_{r,k}}(x)\, \tau_k(dy)$.

—(28)—

Comments and Plans for Future Work

The LDP for $W_{n,r}$ is a special case of a general LDP for doubly indexed sequences of random measures. In an equivalent but slightly different notation, this LDP assumes the following: $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces; I is a convex rate function on $\mathcal{P}(\mathcal{Y})$; for each $r \in I\!N$ $\{L_{q,k}, k = 1, 2, \ldots, 2^r\}$ is an i.i.d. sequence of random measures taking values in $\mathcal{P}(\mathcal{Y})$ and satisfying, for each k, the LDP on $\mathcal{P}(\mathcal{Y})$ with rate function I; θ is a probability measure on $\mathcal{X}$; and $\{D_{r,k}, k = 1, 2, \ldots, 2^r\}$ is a dyadic partition of $\mathcal{X}$ satisfying $\theta\{D_{r,k}\} = 1/2^r$ and two other natural conditions.

Theorem. $W_{r,q}(dx \times dy) \doteq \theta(dx) \otimes \sum_{k=1}^{2^r} 1_{D_{r,k}}(x)\, L_{q,k}(dy)$ satisfies the two parameter LDP on $\mathcal{P}_\theta(T^2 \times \mathcal{Y})$ ($q \to \infty$, then $r \to \infty$) with scaling constants $2^r q$ and rate function

$$J(\mu) \doteq \int_{\mathcal{X}} I(\tau(x, \cdot))\, \theta(dx) \quad [\mu(dx \times dy) = \theta(dx) \otimes \tau(x, dy)].$$

Plans for future work:

- Adapt the proof of the LDP for $W_{r,q}$ to prove the LDP for doubly indexed function space-valued processes in which $L_{r,k}$ is replaced by other random variables. These doubly indexed processes can be used to approximate singly indexed processes and thus to prove the LDP for such processes. In some cases, the use of doubly indexed approximations replaces the technically more complicated use of projective limits.
- Study continuum limits of other models of turbulence in dimension 2 and higher.
- Focus on the solutions, both analytic and numerical, of the variational formulas arising in the large deviation analyses of these models. This will give important insights into the continuum limit behavior of the models.