Varadhan's lemma and applications. Curie-Weiss model.

Elena Kosygina

July 21, 2018

We start by recalling a LDP and presenting a useful tool: Varadhan's lemma and Bryc's inverse of it as an equivalent formulation of LDP. We apply this reformulation to show how a known LDP can be transformed to a new LDP by discussing the contraction principle and tilting. Then we apply some of these tools to Curie-Weiss model of ferromagnetism.

Varadhan's lemma and Bryc's inverse

Recall the definition of LDP from the first lecture.

Definition 1 (LDP)**.** *A sequence of probability measures on a Polish space $\mathbb{X}$ equipped with the Borel σ-algebra $\mathcal{X}$ satisfies a LDP with a rate function $\mathcal{I} : \mathbb{X} \to [0, \infty]$ if*

(i) *$\mathcal{I}$ has compact sub-level sets $\{x \in \mathbb{X} : \ \mathcal{I}(x) \le \ell\}$ for all $\ell \in [0, \infty)$;*

(ii) *for every closed set C*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(C) \le -\inf_{x\in C} \mathcal{I}(x); \tag{1}$$

(iii) *for every open set O*

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mu_n(O) \ge -\inf_{x\in O} \mathcal{I}(x). \tag{2}$$

The following theorem, on the one hand, gives an equivalent formulation of a LDP and, on the other hand, is one of the very useful tools for applications. The theorem consists of two results,[1] and [2]. Each of them has a stronger version (see, for example Sections 4.3 and 4.4 in the already quoted book book by A. Dembo and O. Zeitouni). Following F. Rezakhanlou, *Lectures on the Large Deviation Principle*, we chose the conditions which allow us to write the theorem as an equivalence. For a slightly more general form of the theorem below and its proof we refer to Theorem 1.1 in Chapter 2 of https://math.berkeley.edu/~rezakhan/LD.pdf.

[1] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966

[2] Włodzimierz Bryc. Large deviations by the asymptotic value method. In *Diffusion processes and related problems in analysis, Vol. I (Evanston, IL, 1989)*, volume 22, pages 447–472. Birkhäuser Boston, Boston, MA, 1990

Theorem 1 (Varadhan's lemma $\Rightarrow$, Bryc's inverse $\Leftarrow$). *Let $(\mathbb{X}, \mathcal{X})$ be a Polish space, $(\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{X}, \mathcal{X})$, and $\mathcal{I} : \mathbb{X} \to [0, \infty]$ satisfy (i) of Definition 1. Then (ii) and (iii) of Definition 1 are equivalent to the following statement: for every $F \in C_b(\mathbb{X})$*

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathbb{X}} e^{nF}\, d\mu_n = \sup_{x\in\mathbb{X}} \left(F(x) - \mathcal{I}(x)\right) =: \Lambda_F. \tag{3}$$

Moreover, $\mathcal{I}(x) = \sup_{F\in C_b(\mathbb{X})} \left(F(x) - \Lambda_F\right)$.

The set of all continuous bounded functions on $\mathbb{X}$ is denoted by $C_b(\mathbb{X})$.

IT IS INSTRUCTIVE at this point to recall equivalent definitions of weak convergence of probability measures and take a note of similarities and differences with the definition of LDP and Theorem 1.

Theorem 2 (Portmanteau Theorem). *Let $(\mathbb{X}, \mathcal{X})$ be a Polish space and μ_∞, $(\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{X}, \mathcal{X})$. The following statements are equivalent:*

(i) $\int_{\mathbb{R}} f(x)\, d\mu_n(x) \to \int_{\mathbb{R}} f(x)\, d\mu_\infty$ *as* $n \to \infty$ *for all continuous bounded functions* $f : \mathbb{R} \to \mathbb{R}$.

(ii) $\limsup_{n\to\infty} \mu_n(C) \leq \mu_\infty(C)$ *for every closed set C.*

(iii) $\liminf_{n\to\infty} \mu_n(O) \geq \mu_\infty(O)$ *for every open set O.*

(iv) $\lim_{n\to\infty} \mu_n(A) = \mu_\infty(A)$ *for every Borel set A such that $\mu(\partial A) = 0$.*

If any one of them holds, then we say that $\mu_n \Rightarrow \mu_\infty$ as $n \to \infty$.

See, for example, p. 15 of

Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York, second edition, 1999

Borel sets such that $\mu_\infty(\partial A) = 0$ are called μ_∞-continuity sets.

Statement *(i)* of Theorem 2 and Theorem 1 express weak convergence and LDP respectively in the language of asymptotics of integrals using continuous bounded functions instead of sets. This is very convenient as we shall see later.

Note that in Theorem 2 statements *(ii)* and *(iii)* are trivially equivalent by taking complements while in the LDP principle the corresponding parts *(ii)* and *(iii)* constitute separate statements, and these statements are not equivalent to each other. This is not surprising because we are now looking not at measures of sets but at scaled logarithms of these measures.

Recall that in the first lecture we restated the LDP as follows: if $\mathcal{I}$

satisfies *(i)* of Definition 1 then the LDP holds iff for every $B \in \mathcal{X}$

$$-\inf_{x \in B^o} \mathcal{I}(x) \le \liminf_{n\to\infty} \frac{1}{n} \ln \mu_n(B) \le \limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(B) \le -\inf_{x \in \overline{B}} \mathcal{I}(x).$$

In particular,

$$\lim_{n\to\infty} \frac{1}{n} \ln \mu_n(B) \text{ exists if } \inf_{x \in B^o} \mathcal{I}(x) = \inf_{x \in \overline{B}} \mathcal{I}(x).$$

This restatement fits well with *(iv)* of Theorem 2.

Continuous mappings and Tilts

Theorem 1 can be used to derive new LDPs from known LDPs. Below are two representative results: the contraction principle and tilting.

Lemma 3. *Let $(\mathbb{X}, \mathcal{X})$ be a Polish space, $(\mu_n)_{n\in\mathbb{N}}$ be probability measures on $(\mathbb{X}, \mathcal{X})$ which satisfy a LDP with a rate function $\mathcal{I}$.*

(a) (Contraction Principle) If $(\mathbb{Y}, \mathcal{Y})$ is another Polish space and $g : \mathbb{X} \to \mathbb{Y}$ is a continuous function, then the family of probability measures

$$\nu_n(A) := \mu_n(g^{-1}(A)), \quad A \in \mathcal{Y},\ n \in \mathbb{N},$$

satisfies a LDP with the rate function $\mathcal{J}(y) = \inf\{\mathcal{I}(x) : g(x) = y\}$.

(b) (Tilting) If $G \in C_b(\mathbb{X})$ then the family of probability measures

$$\mu_n^G(A) = \frac{1}{Z_n^G} \int_A e^{nG} d\mu_n, \quad \textit{where } Z_n^G = \int_{\mathbb{X}} e^{nG} d\mu_n,$$

satisfies a LDP with the rate function

$$\mathcal{I}^G(x) = \mathcal{I}(x) - G(x) - \inf_{y \in \mathbb{X}} (\mathcal{I}(y) - G(y)).$$

Proof. *(a)* Let us first check that $\mathcal{J}$ satisfies *(i)* of Definition 1. We claim that for all $\ell \in \mathbb{R}$

$$\{y \in \mathbb{Y} : \mathcal{J}(y) \le \ell\} = g\left(\{x \in \mathbb{X} : \mathcal{I}(x) \le \ell\}\right). \tag{4}$$

Indeed, if $\mathcal{J}(y) \le \ell$ then by the definition of the infimum for every $m \in \mathbb{N}$ there is an $x_m \in \mathbb{X}$ such that $g(x_m) = y$ and $\mathcal{I}(x_m) \le \ell + 1/m$. Since $\{x_m, m \in \mathbb{N}\} \subset \{x \in \mathbb{X} : \mathcal{I}(x) \le \ell + 1\}$ and the latter is compact, there is a converging subsequence $x'_m \to x$ as $m \to \infty$. By continuity of g, $g(x) = y$ and by lower semi-continuity of $\mathcal{I}$, $\mathcal{I}(x) \le \ell$. Therefore,

$$\{y \in \mathbb{Y} : \mathcal{J}(y) \le \ell\} \subset g\left(\{x \in \mathbb{X} : \mathcal{I}(x) \le \ell\}\right).$$

To see the other inclusion, note that if $y \in g\left(\{x \in \mathbb{X} : \mathcal{I}(x) \le \ell\}\right)$ then for some $x_0 \in \{x \in \mathbb{X} : \mathcal{I}(x) \le \ell\}$, $y = g(x_0)$ and

$$\mathcal{J}(y) = \inf\{\mathcal{I}(x) : g(x) = y\} \le \mathcal{I}(x_0) \le \ell.$$

Now that we established (4), we see that the sub-level sets of $\mathcal{J}$ are compact as continuous images of compact sets.

Let $F \in C_b(\mathbb{Y})$. By the definition of ν_n and (3) applied to the family $(\mu_n)_{n\in\mathbb{N}}$ and $F \circ g \in C_b(\mathbb{X})$,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathbb{Y}} e^{nF} d\nu_n = \lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathbb{X}} e^{n(F\circ g)} d\mu_n &= \sup_{x\in\mathbb{X}} ((F\circ g)(x) - \mathcal{I}(x)) \\
&= \sup_{y\in\mathbb{Y}} \sup_{x:\, g(x)=y} ((F\circ g)(x) - \mathcal{I}(x)) \\
&= \sup_{y\in\mathbb{Y}} \left(F(y) - \inf_{x: g(x)=y} \mathcal{I}(x) \right) = \sup_{y\in\mathbb{Y}} (F(y) - \mathcal{J}(y)).
\end{aligned}$$

By Bryc's part of Theorem 1 we conclude that the family $(\nu_n)_{n\in\mathbb{N}}$ satisfies a LDP with a rate function $\mathcal{J}$.

(b) We shall again use Theorem 1. For every $F \in C_b(\mathbb{X})$

$$\begin{aligned}
\lim_{n\to\infty} & \frac{1}{n} \ln \int_{\mathbb{X}} e^{nF} d\mu_n^G \\
&= \lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathbb{X}} e^{n(F+G)} d\mu_n - \lim_{n\to\infty} \frac{1}{n} \ln Z_n^G \\
&= \sup_{x\in\mathbb{X}} (F(x) + G(x) - \mathcal{I}(x)) - \sup_{x\in\mathbb{X}} (G(x) - \mathcal{I}(x)) \\
&= \sup_{x\in\mathbb{X}} (F - \mathcal{I}^G(x)).
\end{aligned}$$

All we need to do is to check that $\mathcal{I}^G$ satisfies *(i)* of Definition 1. Sub-level sets of $\mathcal{I}^G$ are closed, since $\mathcal{I}^G$ as a sum of a lower semi-continuous and continuous functions is lower semi-continuous. Moreover, for each $\ell \in \mathbb{R}$

$$\{x \in \mathbb{X} : \mathcal{I}^G(x) \le \ell\} \subset \{x \in \mathbb{X} : \mathcal{I}(x) \le L\},$$

where $L = \ell + \sup_{\mathbb{X}} G - \sup_G (F - \mathcal{I})$. Thus, $\{x \in \mathbb{X} : \mathcal{I}^G(x) \le \ell\}$ is compact as a closed subset of a compact set. □

Exercise 1. *Let $(Y_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of lognormal random variables with parameters μ and σ and ν_n be the distribution of geometric means, $\tilde{Y}_n = \left(\prod_{i=1}^n Y_i\right)^{1/n}$. Is there a LDP for ν_n? If yes, then what is the rate function?*

Y is said to have a *lognormal* distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if $\ln Y$ is normal with mean μ and variance σ^2.

Exercise 2. *Use the contraction principle to derive Cramér theorem on a finite probability space from Sanov theorem (see the second lecture).*

Exercise 3. *Let $(\mu_n)_{n\in\mathbb{N}}$ be the distribution of empirical means of a sequence of i.i.d. Bernoulli variables with parameter $1/2$. Use tilting (part (b) of Lemma 3) to obtain a LDP for the distributions $(\nu_n)_{n\in\mathbb{N}}$ of empirical means of a sequence of i.i.d. Bernoulli variables with parameter $p \neq 1/2$ directly from the LDP for $(\mu_n)_{n\in\mathbb{N}}$ (see the first lecture).*

Curie-Weiss model I

Let $\Sigma_n = \{-1,1\}^n$ and $\mathcal{G}_n$ be all subsets of Σ_n. A spin configuration $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in \Sigma_n$ is a collection of n spins, each of which can be either 1 or -1. Assigning weight $1/2^n$ to each configuration gives us a uniform measure $\mathbb{P}_n$ on $(\Sigma_n, \mathcal{G}_n)$. This measure corresponds to a non-interacting case: n spins σ_i, $i = 1, 2, \dots, n$, are independent and take value 1 or -1 with equal probabilities. A simple toy model which allows interactions is Curie-Weiss model. Given parameters $\beta \geq 0$, $h \in \mathbb{R}$, and $J > 0$, each configuration $\sigma \in \Sigma_n$ is given a weight

This is a very popular introductory model. A nice exposition can be found, for example, in Chapter 2 of

Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems*. Cambridge University Press, Cambridge, 2018

β is called the inverse temperature, h represents an external magnetic field, and J is a coupling constant.We shall fix J and drop it from the notation.

$$\mathbb{P}_{n,\beta,h}(\sigma) = \frac{e^{-\beta H_{n,h}(\sigma)}}{Z_{n,\beta,h}},$$

where

$$H_{n,h}(\sigma) := -\frac{J}{2n} \sum_{i,j=1}^{n} \sigma_i \sigma_j - h \sum_{j=1}^{n} \sigma_j \tag{5}$$

is called a *Hamiltonian* and

$$Z_{n,\beta,h} := \sum_{\sigma\in\Sigma_n} e^{-\beta H_{n,h}(\sigma)} \tag{6}$$

is a normalization factor which makes $\mathbb{P}_{n,\beta,h}$ into a probability measure on $(\Sigma_n, \mathcal{G}_n)$. $Z_{n,\beta,h}$ is commonly referred to as a *partition sum*. In the special case when $\beta = 0$ (infinite temperature) we have $\mathbb{P}_{n,0,h} = \mathbb{P}_n$. From now one we shall assume that $\beta > 0$.

The first part of the Hamiltonian represents interaction between spins. We can write

$$\frac{J}{2n} \sum_{i,j=1}^{n} \sigma_i \sigma_j = \frac{J}{2} \sum_{i=1}^{n} \sigma_i \left(\frac{1}{n} \sum_{j=1}^{n} \sigma_j \right)$$

and interpret the term

$$J\sigma_i \left(\frac{1}{n} \sum_{j=1}^{n} \sigma_j \right)$$

as an interaction of "strength" J between σ_i and the average over all spins (magnetization density). This model is a representative of a class of *mean field models*: there is no geometry, each spin interacts with all spins.

One remarkable feature of Curie-Weiss model is that $H_{n,h}(\sigma)$ can be written as a function of the spin average, $\overline{\sigma}_n := \frac{1}{n}\sum_{i=1}^{n} \sigma_i \in \mathcal{R}_n := \{-1, -1+2/n, \ldots, 1-2/n, 1\} \subset [-1,1]$,

$$H_{n,h}(\sigma) = -n \left(\frac{J}{2} \, (\overline{\sigma}_n)^2 + h\overline{\sigma}_n \right).$$

This tells us, in particular, that all configurations with the same magnetization density $\overline{\sigma}_n$ have the same weight, i.e. the conditional distribution of $\mathbb{P}_{n,\beta,h}$ given $\overline{\sigma}_n = m$ is uniform. Therefore, we shall concentrate on the study of the behavior of $\overline{\sigma}_n$ for large n. Denote by $\mu_{n,\beta,h}$ the distribution of $\overline{\sigma}_n$ under $\mathbb{P}_{n,\beta,h}$,

$$\mu_{n,\beta,h}(A) := \mathbb{P}_{n,\beta,h}(\sigma : \overline{\sigma}_n \in A), \quad A \subset [-1,1].$$

Note that while $\mathbb{P}_{n,\beta,h}$ "live" on different measurable spaces $(\Sigma_n, \mathcal{G}_n)$, probability measures $(\mu_{n,\beta,h})_{n\in\mathbb{N}}$ can be conveniently defined on the same space $([-1,1], \mathcal{B}([-1,1]))$.

We shall first consider the case $h = 0$. To ease up the notation we drop the subscript h to indicate that $h = 0$. Theorems 4 and 5 contain the main results.

A more realistic *Ising model* can be defined in a large box on $\mathbb{Z}^d$ (number of sites in the box will correspond to our n), where spins would interact only with its nearest $2d$ neighbors: the interaction term (see the first term in (5)) is equal to

$$-\sum_{i\sim j} \sigma_i\sigma_j = -\frac{1}{2}\sum_i \left(\sigma_i \sum_{j\sim i} \sigma_j \right),$$

where $j \sim i$ means that j is a neighbor of i on the lattice $\mathbb{Z}^d$. Now σ_i interacts only with the local magnetization density,

$$\sigma_i \sum_{j\sim i} \sigma_j = 2d\sigma_i \left(\frac{1}{2d} \sum_{j\sim i} \sigma_j \right).$$

Since the "neighboring" relation is symmetric, we write a factor $1/2$ when summing over all i in the box. Replacing the average over nearest neighbors with the average over all spins in the box gives the Curie-Weiss model with $J = 2d$.

Theorem 4. *For each $\beta > 0$ the family of measures $(\mu_{n,\beta})_{n\in\mathbb{N}}$ satisfies a LDP with the rate function*

$$\mathcal{I}_\beta(x) = \mathcal{I}_0(x) - \frac{\beta J}{2}\, x^2 - \inf_{y\in[-1,1]} \left[\mathcal{I}_0(y) - \frac{\beta J}{2}\, y^2 \right],$$

where

$$\mathcal{I}_0(x) = \begin{cases} \ln 2 + \frac{1+x}{2} \ln \frac{1+x}{2} + \frac{1-x}{2} \ln \frac{1-x}{2}, & \text{if } |x| < 1; \\ \ln 2, & \text{if } |x| = 1; \\ \infty, & \text{if } |x| > 1. \end{cases}$$

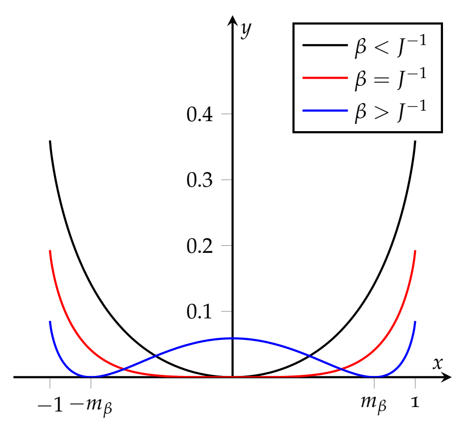

Figure 1: Graph of $\mathcal{I}_\beta(x)$ in different regimes. (The picture uses $\beta J = 2/3, 1, 3/2$.)

It is worth noticing that the theorem also holds for $\beta = 0$. In the first lecture we discussed empirical averages of i.i.d. sequences of Bernoulli random variables. The non-interacting case $\beta = 0$ puts us in that setting. The only difference is that our spins take values 1 and -1 instead of 1 and 0. Thus, a simple linear transformation gives us a LDP for the case $\beta = 0$. The rate function is exactly $\mathcal{I}_0$.

IT IS NOT OBVIOUS from the statement of Theorem 4 that the behavior of the model undergoes any changes as we "cool" the system, i.e. increase the inverse temperature β from 0 to ∞. We shall see the changes only when we take a more careful look at the rate function $\mathcal{I}_\beta$. Note that $\mathcal{I}_\beta$ is an even function. We shall see that when $\beta \leq J^{-1}$ it has a unique strict minimum (equal to 0) at the origin and is convex but when $\beta > J^{-1}$, the minimum is attained at two points, $\pm m_\beta \in (-1,0) \cup (0,1)$, and (surprise!) $\mathcal{I}_\beta$ is not anymore convex. The next theorem shows clearly that the model has two phases depending on the value of β.

In more realistic models, such as Ising model, the rate function is always convex. The non-convexity of the rate function in Curie-Weiss model could be attributed to its mean field nature (no geometry).

Theorem 5. *If $0 \leq \beta \leq J^{-1}$ then*

$$\mu_{n,\beta} \Rightarrow \delta_0 \quad \text{as } n \to \infty.$$

For $\beta > J^{-1}$ denote by $m_\beta = m_\beta(J)$ the unique solution in $(0,1)$ of the equation $x = \tanh(\beta J x)$. Then

$$\mu_{n,h} \Rightarrow \frac{1}{2}\delta_{m_\beta} + \frac{1}{2}\delta_{-m_\beta} \quad \text{as } n \to \infty.$$

Proof of Theorem 4. Since $\mathcal{I}_\beta$ is continuous on $[-1,1]$, it satisfies (i) of Definition 1: for each $c \geq 0$ the sub-level set $\{x \in \mathbb{R} : \mathcal{I}_\beta(x) \leq c\} \subset [-1,1]$ is closed and bounded; thus, it is compact.

We would like to apply Bryc's part of Theorem 1. For this we need to check that for every $F \in C_b([-1,1])$

Since $\mu_{n,\beta}([-1,1]^c) = 0$, it is enough to consider functions on $[-1,1]$ instead of $\mathbb{R}$.

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{[-1,1]} e^{nF(x)}\, d\mu_{n,\beta} = \sup_{x\in[-1,1]} [F(x) - \mathcal{I}_\beta(x)].$$

Equivalently, the integral can be written simply as a sum

$$\sum_{x\in\mathcal{R}_n} e^{nF(x)} \mu_{n,\beta}(\{x\}),$$

where $\mathcal{R}_n = \left\{-1, -1+\frac{2}{n}, \ldots, 1\right\}$.

We shall change the measure from $\mu_{n,\beta}$ to $\mu_{n,0}$ ($\beta = 0$). The point is that, as discussed above, for $\mu_{n,0}$ we already have a LDP. And our plan is to use Varadhan's lemma (see Theorem 1). We write

$$\frac{d\mu_{n,\beta}}{d\mu_{n,0}}(x) = \frac{\mu_{n,\beta}(x)}{\mu_{n,0}(x)} = \frac{2^n e^{\frac{\beta J n}{2} x^2}}{Z_{n,\beta}}$$

and get

$$\begin{aligned}\frac{1}{n}\ln\int_{[-1,1]} e^{nF(x)}\,d\mu_{n,\beta} &= \frac{1}{n}\ln\int_{[-1,1]} e^{nF(x)}\frac{d\mu_{n,\beta}}{d\mu_{n,0}}\,d\mu_{n,0}\\ &= \frac{1}{n}\ln\int_{[-1,1]} e^{n\left(F(x)+\frac{\beta J}{2}x^2\right)}\,d\mu_{n,0} - \frac{1}{n}\ln\frac{Z_{n,\beta}}{2^n}\\ &= \frac{1}{n}\log\int_{[-1,1]} e^{n\left(F(x)+\frac{\beta J}{2}x^2\right)}\,d\mu_{n,0} - \frac{1}{n}\log\int_{[-1,1]} e^{\frac{n\beta J}{2}x^2}\,d\mu_{n,0}.\end{aligned}$$

Functions $F(x)$, $\frac{\beta J}{2}x^2$ are continuous and bounded on $[-1,1]$, and $\mu_{n,0}$ satisfies a LDP with the rate function $\mathcal{I}_0$. By Varadhan's lemma, we can take the limit in the right hand side of the above expression and conclude that

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\ln\int_{[-1,1]} e^{nF(x)}\,d\mu_{n,\beta}\\ &= \sup_{x\in[-1,1]}\left(F(x)+\frac{\beta J}{2}x^2-\mathcal{I}_0(x)\right) - \sup_{y\in[-1,1]}\left(\frac{\beta J}{2}y^2-\mathcal{I}_0(y)\right)\\ &= \sup_{x\in[-1,1]}\left[F(x)-\left(\mathcal{I}_0(x)-\frac{\beta J}{2}x^2-\inf_{y\in[-1,1]}\left(\mathcal{I}_0(y)-\frac{\beta J}{2}y^2\right)\right)\right]\\ &= \sup_{x\in[-1,1]}\left[F(x)-\mathcal{I}_\beta(x)\right].\end{aligned}$$

Bryc's part of Theorem 1 yields the desired LDP. □

Proof of Theorem 5. We shall use the equivalent definition (ii) of weak convergence from Portmanteau Theorem. We need to show that for every closed set $C\subset[-1,1]$

$$\limsup_{n\to\infty}\mu_{n,\beta}(C)\le\begin{cases}\delta_0(C), & \text{if } 0<\beta\le J^{-1};\\ \frac{1}{2}\delta_{m_\beta}(C)+\frac{1}{2}\delta_{-m_\beta}(C), & \text{if } \beta>J^{-1}.\end{cases}$$

Case $\beta\in(0,J^{-1}]$. If $0\in C$ then $\delta_0(C)=1$, and the upper bound holds trivially. Suppose now that $0\notin C$. By Theorem 4 and part *(ii)* of Definition 1,

$$\limsup_{n\to\infty}\frac{1}{n}\ln\mu_{n,\beta}(C)\le-\inf_{x\in C}\mathcal{I}_\beta(x)<-\epsilon<0$$

for some $\epsilon>0$. This is due to the fact that $0\notin C$ and $\mathcal{I}_\beta$ is continuous and attains its unique strict minimum 0 at $x=0$. This implies that for all sufficiently large n, $\mu_{n,\beta}(C)\le e^{-n\epsilon/2}$ and, thus,

$$\limsup_{n\to\infty}\mu_{n,\beta}(C)=0=\delta_0(C).$$

Case $\beta>J^{-1}$. If $\{-m_\beta,m_\beta\}\subset C$ then $\frac{1}{2}\delta_{-m_\beta}(C)+\frac{1}{2}\delta_{m_\beta}(C)=1$ and the upper bound holds trivially.

Recall that the rate function $\mathcal{I}_\beta$ attains its minimum value 0 at $\pm m_\beta$, $m_\beta \in (0,1)$. If $\{-m_\beta, m_\beta\} \cap C = \emptyset$ then, just as above, by the LPD upper bound for some $\epsilon > 0$ and all sufficiently large n, $\mu_{n,\beta}(C) \leq e^{-n\epsilon/2}$ and

$$\limsup_{n\to\infty} \mu_{n,\beta}(C) = 0 = \frac{1}{2}\delta_{-m_\beta}(C) + \frac{1}{2}\delta_{m_\beta}(C).$$

Moreover, by symmetry, for any $\epsilon > 0$

$$\begin{aligned} 1 = \lim_{n\to\infty} \mu_{n,\beta}((-m_\beta - \epsilon, -m_\beta + \epsilon) \cup (m_\beta - \epsilon, m_\beta + \epsilon)) \\ = 2 \lim_{n\to\infty} \mu_{n,\beta}((-m_\beta - \epsilon, -m_\beta + \epsilon)), \end{aligned}$$

and, thus, the last limit exists and is equal to 1/2. If $m_\beta \in C$ and $-m_\beta \notin C$ then for some sufficiently small $\epsilon > 0$ we have that

$$C \subset [-1,1] \setminus (-m_\beta - \epsilon, -m_\beta + \epsilon)$$

and

$$\begin{aligned} \limsup_{n\to\infty} \mu_{n,\beta}(C) \leq 1 - \liminf_{n\to\infty} \mu_{n,\beta}((-m_\beta - \epsilon, -m_\beta + \epsilon)) \\ = \frac{1}{2} = \frac{1}{2}\delta_{-m_\beta}(C) + \frac{1}{2}\delta_{m_\beta}(C). \end{aligned}$$

The case when $-m_\beta \in C$ and $m_\beta \notin C$ is similar. □

Exercise 4. *State and prove the LDP and the limit theorem for $h \neq 0$.*

References

Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York, second edition, 1999.

Włodzimierz Bryc. Large deviations by the asymptotic value method. In *Diffusion processes and related problems in analysis, Vol. I (Evanston, IL, 1989)*, volume 22, pages 447–472. Birkhäuser Boston, Boston, MA, 1990.

Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems*. Cambridge University Press, Cambridge, 2018.

S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.