Gärtner-Ellis Theorem and applications.

Elena Kosygina

July 25, 2018

In this lecture we turn to the non-i.i.d. case and discuss Gärtner-Ellis theorem. As an application, we study Curie-Weiss model with random external magnetic fields.

We have discussed Cramér's theorem for empirical means of i.i.d. random variables which take values in $\mathbb{R}$. This theorem under exactly the same condition[1] can be shown to hold for empirical means of i.i.d. random vectors with values in $\mathbb{R}^d$ (see [2]Corollary 6.1.6). The result we consider in this note is much more general: measures μ_n are not required to be distributions of empirical means, they can be arbitrary distributions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ as long as the limit

$$\Lambda(t) := \lim_{n\to\infty} \frac{1}{n} \ln \Lambda_n(nt) := \lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathbb{R}^d} e^{\langle nt,x\rangle} d\mu_n \in [-\infty, \infty] \qquad (1)$$

exists for all $t \in \mathbb{R}^d$.

[1] The condition is $0 \in D^o_\Lambda$, where Λ is the logarithmic MGF of the common distribution μ on $\mathbb{R}^d$.

[2] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 1998

A TYPICAL SITUATION which arises in applications is the following: we have a sequence of random vectors $(Y_n)_{n\in\mathbb{N}}$ on possibly different probability spaces $(\Omega_n, \mathcal{G}_n, \mathbb{P}_n)$, $Y_n : \Omega_n \to \mathbb{R}^d$, and

$$\mu_n(A) := \mathbb{P}_n(Y_n \in A) \quad \text{for every } A \in \mathcal{B}(\mathbb{R}^d).$$

We can always think of these random variables as defined on a common probability space so that their distributions are exactly those that we want (e.g. construct them on an infinite product space) but it is not necessary.

IN THE SPECIAL CASE when μ_n is the distribution of empirical means of i.i.d. random vectors $(X_i)_{i\in\mathbb{N}}$ with a common distribution μ we have

$$\Lambda(t) = \frac{1}{n} \ln \Lambda_n(nt) = \frac{1}{n} \ln \left[\mathbb{E}\left(e^{\langle t,X_i\rangle}\right)\right]^n = \ln \int_{\mathbb{R}^d} e^{\langle t,x\rangle} d\mu$$

- just the familiar logarithmic MGF of μ.

Gärtner-Ellis Theorem

Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Assume that for all $t \in \mathbb{R}^d$ a possibly infinite limit $\Lambda(t)$ in (1) exists. The convexity of

$$\Lambda_n(t) := \ln \int_{\mathbb{R}^d} e^{\langle t,x\rangle} d\mu_n$$

for each $n \in \mathbb{N}$ and the limit definition of Λ immediately imply that

$$\Lambda \text{ is convex.}$$

Similarly to the assumption in Cramér's theorem, we shall assume throughout this note that $0 \in D^o_\Lambda$. This will ensure, in particular, that

$$\Lambda > -\infty.$$

Indeed, note that as $\Lambda_n(0) = 0$ for all n, so $\Lambda(0) = 0$. If for some t we had $\Lambda(t) = -\infty$ then by convexity we would have for all $\alpha \in (0, 1]$

$$\Lambda(\alpha t) = \Lambda(\alpha t + (1-\alpha)0) \leq \alpha\Lambda(t) + (1-\alpha)\Lambda(0) = -\infty.$$

But then

$$0 = \Lambda(0) = \Lambda\left(\frac{1}{2}(\alpha t) + \frac{1}{2}(-\alpha t)\right) \leq \frac{1}{2}\Lambda(\alpha t) + \frac{1}{2}\Lambda(-\alpha t),$$

and we would also have $\Lambda(-\alpha t) = \infty$ for all $\alpha \in (0, 1]$. This contradicts the assumption $0 \in D^o_\Lambda$.

We shall also need the following definition.

Definition 1. *Let $\mathcal{I}$ be the Legendre-Fenchel transform of Λ, i.e.*

$$\mathcal{I}(x) = \sup_{x\in\mathbb{R}^d} (\langle t, x\rangle - \Lambda(t)).$$

A point $x \in D_{\mathcal{I}} := \{x \in \mathbb{R}^d : \mathcal{I}(x) < \infty\}$ is said to be exposed for $\mathcal{I}$ if there is a $\eta \in \mathbb{R}^d$ such that

$$\mathcal{I}(y) - \mathcal{I}(x) > \langle \eta, y - x\rangle \quad \text{for all } y \neq x.$$

The hyperplane $h_x(y) = \mathcal{I}(x) + \langle \eta, y - x\rangle$ is called an exposing hyperplane to the graph of $\mathcal{I}$ at x. For a given $(x, \mathcal{I}(x))$, it is characterized by its normal η. With a slight abuse of terminology, η itself will be referred to as an exposing hyperplane.

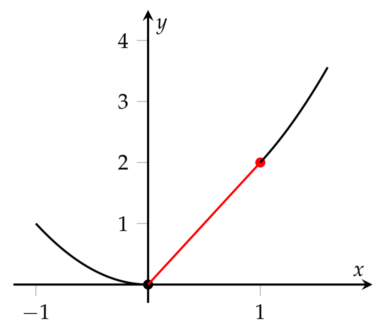


Figure 1: Points $x \in (0, 1]$ are not exposed for the pictured function.

Since $\mathcal{I}$ is the Legendre-Fenchel transform of Λ, $\mathcal{I}$ is convex and satisfies all conditions of a rate function, i.e. it is non-negative and has compact sub-level sets. A justification of the last two claims can be given along the same lines as in lecture notes 1 (right after Cramér theorem).

Theorem 1 (Gärtner-Ellis). *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Assume that for all $t \in \mathbb{R}^d$ a possibly infinite limit $\Lambda(t)$ in* (1) *exists and that $0 \in D^o_\Lambda$. Then*

(i) for every closed set $C \subset \mathbb{R}^d$,

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

Jurgen Gärtner. On large deviations from an invariant measure. *Teor. Verojatnost. i Primenen.*, 22(1):27–42, 1977

Richard S. Ellis. Large deviations for a general class of random vectors. *Ann. Probab.*, 12(1):1–12, 1984 *(ii) for every open set $O \subset \mathbb{R}^d$,*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mu_n(O) \geq -\inf_{x\in O\cap E} \mathcal{I}(x),$$

where E is the set of those exposed points for $\mathcal{I}$ which have an exposing hyperplane in D^o_Λ.

Suppose in addition that Λ is lower semi-continuous on $\mathbb{R}^d$, differentiable on D^o_Λ, and either $D_\Lambda = \mathbb{R}^d$ or Λ is steep, i.e.

$$\lim_{n\to\infty} |\nabla\Lambda(t_n)| = \infty$$

whenever $t_n \in D^o_\Lambda$, $t_n \to t \in \partial D^o_\Lambda$ as $n \to \infty$. Then $(\mu_n)_{n\in\mathbb{N}}$ satisfies the LDP with rate function $\mathcal{I}$.

THIS THEOREM is rather general but still it does not capture all the cases in which a sequence of measures on $\mathbb{R}^d$ satisfies a LDP. Our concern is, of course, only about the lower bound. Here is a simple example borrowed from [3], p. 45.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 1998

Example 1. Let $\mu_n((-\infty, x]) = (1 - e^{-nx})\mathbb{1}_{[0,\infty)}(x)$ (exponential distribution with parameter n), $x \in \mathbb{R}$. Then

$$\Lambda(t) = \lim_{n\to\infty} \frac{1}{n} \ln\left(\int_0^\infty n e^{ntx-nx}\, dx\right) = \begin{cases} 0, & \text{if } t < 1; \\ \infty, & \text{if } t \geq 1. \end{cases}$$

$$\mathcal{I}(x) = \sup_{t\in\mathbb{R}}(tx - \Lambda(t)) = \begin{cases} x, & \text{if } x \geq 0; \\ \infty, & \text{if } x < 0. \end{cases}$$

We see that $E = \{0\}$ while $D_{\mathcal{I}} = [0,\infty)$, and for each open set O with $O \cap E = \emptyset$ Gärtner-Ellis theorem gives only a trivial lower bound $-\infty$.

But it is easy to see directly that for every open set O for which $O \cap D_{\mathcal{I}} \neq \emptyset$

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \ln \mu_n(O) &= \lim_{n\to\infty} \frac{1}{n} \ln \int_{O\cap[0,\infty)} n e^{-nx}\, dx \\ &\geq -\inf\{x, x \in O \cap [0,\infty)\} = -\inf_{x\in O} \mathcal{I}(x). \end{aligned}$$

Exercise. Give the details of this computation. *Hint:* Every open set in $\mathbb{R}$ is a countable union of disjoint open intervals. Use the leftmost interval which intersects $D_{\mathcal{I}}$.

This says that $(\mu_n)_{n\in\mathbb{N}}$ satisfy a LDP with rate $\mathcal{I}$.

WE NOTE THAT Gärtner-Ellis Theorem readily implies Cramér theorem on $\mathbb{R}^d$, $d \geq 1$, when the logarithmic MGF of μ is finite on all of $\mathbb{R}^d$. Otherwise we have to impose an additional condition of steepness (see Theorem 1). As we have already mentioned above, the inclusion $0 \in D_\Lambda^o$ is sufficient for the result in Cramér theorem to hold. The following exercise shows that Theorem 1 does not fully include Cramér theorem if $D_\Lambda \neq \mathbb{R}^d$.

Exercise 1 (Exercise 2.3.17(a) in A. Dembo, O. Zeitouni)**.** *Let μ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with a density proportional to $e^{-|x|}/(1+|x|^{d+2})$ and Λ be its logarithmic MGF.*

(a) Find D_Λ^o.

(b) Show that Λ is not steep.

Conclude that in this example Cramér theorem yields the full LDP for empirical means while Gärtner-Ellis theorem does not.

A general limitation of Gärtner-Ellis theorem is that the differentiability and steepness conditions on Λ are often hard to check in applications. This is due to the fact that the existence of the limit which defines Λ is often obtained by "soft methods" which do not give a usable formula for Λ.

Let us look again at the already familiar Curie-Weiss model and see what Gärtner-Ellis theorem can and can not do in this case. This example also provides a good illustration of notions introduced in this note.

Curie-Weiss model II: what does Gärtner-Ellis theorem give us?

For each spin configuration $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in \Sigma_n = \{-1,1\}^n$ we define

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

where

$$H_n(\sigma) = -\frac{J}{2n}\sum_{i,j=1}^{n} \sigma_i\sigma_j \quad \text{and} \quad Z_{n,\beta} = \sum_{\sigma\in\Sigma_n} e^{-\beta H_n(\sigma)}.$$

For each n we have a different probability space $(\Sigma_n, \mathcal{G}_n, \mathbb{P}_{n,\beta})$. We study random variables $\overline{\sigma}_n = \sum_{i=1}^n \sigma_i \in [-1,1]$. Thus, the relevant probability measures $\mu_{n,\beta}$ (the distributions of $\overline{\sigma}_n$ under $\mathbb{P}_{n,\beta}$), $n \in \mathbb{N}$, are measures on the common measurable space $([-1,1], \mathcal{B}([-1,1]))$. Thus,

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

We obtained two main results for $(\mu_{n,\beta})_{n\in\mathbb{N}}$: the 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

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

and weak convergence to a limiting measure. These results clearly show the existence of a phase transition as β crosses J^{-1}.

Let us apply Gärtner-Ellis theorem for this case and draw conclusions. We know from the start that when $\beta > J^{-1}$ we should run into trouble, since in this regime $\mathcal{I}_\beta$ is not convex while every rate function in Gärtner-Ellis theorem is convex.

To start, we need to compute $\Lambda_\beta(t)$. As we did before, we rewrite everything in terms of $\mu_{n,0}$ ($\beta=0$) which assigns a binomial probability to each possible value of $\overline{\sigma}_n \in \{-1,-1+2/n,\ldots,1-2/n,1\}$,

$$\mu_{n,0}(\{-1+2m/n\}) = 2^{-n}\binom{n}{m}$$

and use the LDP for these measures (with the rate function $\mathcal{I}_0$) and Varadhan's lemma to get

$$\begin{aligned}
\Lambda_\beta(t) &= \lim_{n\to\infty}\frac{1}{n}\ln\int_{\mathbb{R}} e^{ntx}\,d\mu_{n,\beta} \\
&= \lim_{n\to\infty}\frac{1}{n}\ln\frac{E_{\mu_{n,0}}\left(e^{n(t\overline{\sigma}_n+\frac{\beta J}{2}(\overline{\sigma}_n)^2)}\right)}{E_{\mu_{n,0}}\left(e^{n\frac{\beta J}{2}(\overline{\sigma}_n)^2}\right)} = \sup_{x\in[-1,1]}\left(tx+\frac{\beta J}{2}\,x^2-\mathcal{I}_0(x)\right) - \sup_{y\in[-1,1]}\left(\frac{\beta J}{2}\,y^2-\mathcal{I}_0(y)\right) \\
&= \sup_{x\in[-1,1]}\left[tx+\left(\frac{\beta J}{2}\,x^2-\mathcal{I}_0(x)-\sup_{y\in[-1,1]}\left(\frac{\beta J}{2}\,y^2-\mathcal{I}_0(y)\right)\right)\right] = \sup_{x\in[-1,1]}(tx-\mathcal{I}_\beta(x)).
\end{aligned}$$

Since we take the supremum of continuous functions over $[-1,1]$, $\Lambda_\beta(t) < \infty$ for all t, i.e. $D_{\Lambda_\beta} = \mathbb{R}$. By Theorem 1 (i) we get a LDP upper bound with the rate function Λ^*_β.

IF $\beta \le J^{-1}$ THEN $\mathcal{I}_\beta$ is convex and $\Lambda^*_\beta = \mathcal{I}^{**} = \mathcal{I}_\beta$ with $D_{\mathcal{I}_\beta} = [-1,1]$. Moreover, $\mathcal{I}_\beta$ is strictly convex so that every point in $(-1,1)$ is exposed with an exposing hyperplane in D_{Λ_β} simply because $D_{\Lambda_\beta} = \mathbb{R}$. By Theorem 1 (ii) we get a LDP lower bound: for every open set O

$$\begin{aligned}
\liminf_{n\to\infty}\frac{1}{n}\ln\mu_{n,\beta}(O) &\ge -\inf_{x\in O\cap(-1,1)}\mathcal{I}_\beta(x) \\
&= -\inf_{x\in O\cap[-1,1]}\mathcal{I}_\beta(x) = -\inf_{x\in O}\mathcal{I}_\beta(x).
\end{aligned}$$

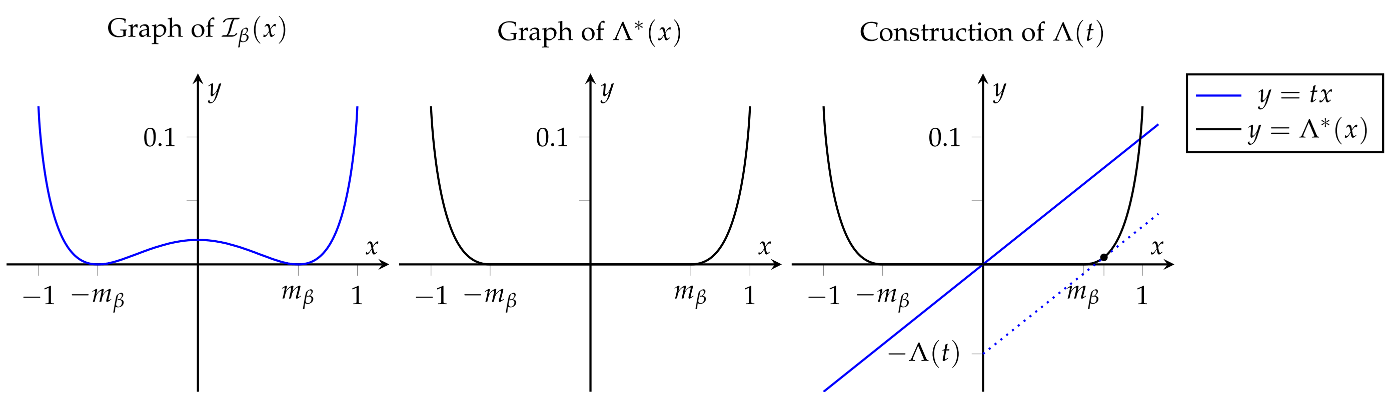
The next to the last equality is due to continuity of $\mathcal{I}_\beta$ on $[-1,1]$. The last equality holds because $\mathcal{I}_\beta(x) = \infty$ outside of $[-1,1]$. We recover an earlier obtained LDP in this case. Alternatively, we could have used the last statement of Theorem 1: Λ_β satisfies all additional conditions needed for a full LDP.

IF $\beta > J^{-1}$ THEN Λ^*_β is convex while $\mathcal{I}_\beta$ is not and

$$\Lambda^*_\beta(x) = \begin{cases} \mathcal{I}_\beta, & \text{if } |x| \geq m_\beta; \\ 0, & \text{if } |x| < m_\beta, \end{cases}$$

where $\pm m_\beta \in (-1,0) \cup (0,1)$ are the points where $\mathcal{I}_\beta$ attains its minimal value 0.

In lecture note 3 we have shown that m_β is the unique solution in $(0,1)$ of the equation $x = \tanh(\beta J x)$.



Our first observation is that $\Lambda^*_\beta < \mathcal{I}_\beta$ on $(-m_\beta, m_\beta)$ and an upper bound provided by Theorem 1 is strictly larger than the one we obtained earlier: for every closed set $G \subset (-m_\beta, m_\beta)$

$$-\inf_{x \in G} \mathcal{I}_\beta(x) < -\inf_{x \in G} \Lambda^*_\beta(x).$$

We also see that $\Lambda^*_\beta \equiv 0$ on $[-m_\beta, m_\beta]$. This implies that $\Lambda_\beta(t) = \mathcal{I}^*_\beta(t) = \Lambda^{**}_\beta(t) \geq |t| m_\beta$ and, since $\Lambda_\beta(0) = 0$, we conclude that Λ_β has a corner at the origin.

The differentiability condition fails at 0, and Theorem 1 only gives us the following lower bound: for all open sets O

$$\liminf_{n \to \infty} \frac{1}{n} \ln \mu_{n,\beta}(O) \geq -\inf_{x \in O \cap E} \Lambda^*_\beta(x),$$

where the set of exposed points E is $(-1, m_\beta) \cup (m_\beta, 1)$ (we have to exclude the set where Λ^*_β is not strictly convex). If $O \subset [-m_\beta, m_\beta]$ then Theorem 1 gives us only a trivial lower bound $-\infty$ (the infimum is taken over the empty set), while our earlier results give us more information in this case.

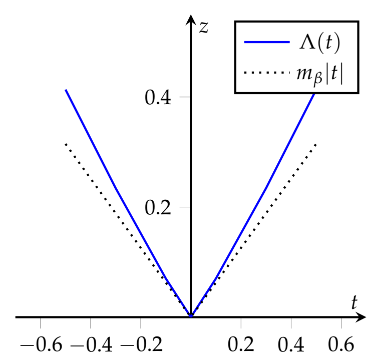


Figure 2: Graph of $\Lambda(t) = \Lambda^{**}(t)$.

Now it is the right time for an example where Gärtner-Ellis theorem provides an efficient and quick way of getting a LDP.

Curie-Weiss model III: adding a random external field

This subsection follows [4] and shows how to get a LDP for Curie-Weiss model in the case when a constant global external magnetic field h is replaced by a random local field $\mathbf{h} := (h_i)_{i\in\mathbb{N}}$ satisfying an appropriate averaging assumption (see **Assumption** below).

[4] Matthias Löwe, Raphael Meiners, and Felipe Torres. Large deviations principle for Curie-Weiss models with random fields. *J. Phys. A*, 46(12):125004, 10 pp., 2013

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbf{h} = (h_i)_{i\in\mathbb{N}}$ be a sequence of random variables on it. We shall consider random probability measures on $(\Sigma_n, \mathcal{G}_n)$ defined by

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

where

$$H_{n,\mathbf{h}}(\sigma) = -\frac{J}{2n}\sum_{i,j=1}^{n}\sigma_i\sigma_j - \sum_{i=1}^{n} h_i\sigma_i, \quad Z_{n,\beta,\mathbf{h}} = \sum_{\sigma\in\Sigma_n} e^{-\beta H_{n,\mathbf{h}}(\sigma)}.$$

An important difference with the original deterministic model is that the measure $\mathbb{P}_{n,\beta,\mathbf{h}}(\sigma)$ is not anymore completely determined by the value of $\overline{\sigma}_n$.

For each $\omega \in \Omega$, let $\mu_{n,\beta,\mathbf{h}}$ be the distribution of $\overline{\sigma}_n$. It is a measure on $[-1,1]$. Our goal will be to state and prove a LDP for $(\mu_{n,\beta,\mathbf{h}})_{n\in\mathbb{N}}$. For this we shall need an assumption on a random field $\mathbf{h}$.

Assumption: Let $\mathbf{h} = (h_i)_{i\in\mathbb{N}}$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that for almost every $\omega \in \Omega$

$$f_n(t,\omega) := \frac{1}{n}\sum_{i=1}^{n}\ln\cosh(t+\beta h_i(\omega)) \to f(t) \text{ as } n \to \infty$$

for every $t \in \mathbb{R}$ and some differentiable function $f : \mathbb{R} \to \mathbb{R}$.

Exercise 2. *Prove that if f satisfies the above assumption then necessarily $|f'(t)| \leq 1$ for all $t \in \mathbb{R}$ and, therefore, $f^*(x) := \sup_{t\in\mathbb{R}}(tx - f(t))$ is equal to ∞ for all $|x| > 1$.*

The most basic example is given by a sequence $(h_i)_{i\in\mathbb{N}}$ of i.i.d. random variables with a finite first moment, $\mathbb{E}[|h_i|] < \infty$. The convergence follows by the strong law of large numbers (and a small additional argument, see the exercise below) and $f(t) = \mathbb{E}[\ln\cosh(t+\beta h_i)]$.

Note that $|\ln\cosh t| \leq |t|$ and, thus, $\mathbb{E}[|\ln\cosh(t+\beta h_i)|] \leq |t| + \beta\mathbb{E}[|h_i|] < \infty$.

Exercise 3. *The strong law of large numbers implies that for every t the convergence holds outside of a set of probability 0. But $\mathbb{R}$ is uncountable. Start with rational t and show rigorously that the assumption is indeed satisfied.*

ANOTHER EXAMPLE, now with dependence, arises when $(h_i)_{i\in\mathbb{N}}$ is an irreducible positive recurrent Markov chain on a countable subset S of $\mathbb{R}$. Such Markov chain has a unique stationary distribution π on S. Assuming that

$$\sum_{s\in S} |\ln\cosh(t+\beta s)|\pi(s) < \infty,$$

the ergodic theorem for Markov chains (see Theorem 4.1 in [5]) implies that for any initial distribution with probability 1

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \ln\cosh(t+\beta h_i) = f(t) := \sum_{s\in S} \ln\cosh(t+\beta s)\pi(s).$$

Note that when the initial distribution is different from π the random variables $(h_i)_{i\in\mathbb{N}}$ are not identically distributed.

[5] Pierre Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues

FOR OUR LAST EXAMPLE let $(h_i)_{i\in\mathbb{N}\cup\{0\}}$ be a stationary ergodic sequence on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. More precisely, let $T:\Omega\to\Omega$ be a measure preserving transformation, i.e. T be $\mathcal{F}$-measurable and $\mathbb{P}(T^{-1}A)=\mathbb{P}(A)$ for all $A\in\mathcal{F}$. In addition, assume that T is ergodic, that is $T^{-1}A = A$ implies that $\mathbb{P}(A)\in\{0,1\}$. Let h be any random variable on $(\Omega,\mathcal{F},\mathbb{P})$ with a finite first moment. Define $h_i(\omega) = h(T^i\omega)$, $i\in\mathbb{N}\cup\{0\}$. The ergodicity assumption and the finiteness of the first moment of h ensure (by Birkhoff ergodic theorem) that for almost every $\omega\in\Omega$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \ln\cosh(t+\beta h_i(\omega)) = f(t) := \mathbb{E}[\ln\cosh(t+\beta h(\omega))].$$

A sequence $(h_i)_{i\in\mathbb{N}\cup\{0\}}$ is said to be stationary if for every $k\in\mathbb{N}\cup\{0\}$ and $n\in\mathbb{N}$ the random vector $(h_n,h_{n+1},\ldots,h_{n+k})$ has the same distribution as $(h_0,h_1,\ldots,h_n)$.

In simple words, ergodicity of T means that there are no non-trivial T-invariant sets.

This general framework, in fact, includes the previous two examples provided that the Markov chain in the second example starts from the stationary distribution.

Now we are ready to state the main theorem of this subsection.

Theorem 2. *Suppose that $\mathbf{h} = (h_i)_{i\in\mathbb{N}}$ satisfies the assumption above. Then for almost every $\omega\in\Omega$ measures $(\mu_{n,\beta,\mathbf{h}})_{n\in\mathbb{N}}$ satisfy a LDP with deterministic rate function*

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

where $f^*(x) := \sup_{t\in\mathbb{R}}(tx - f(t))$.

Remark 1. Note that if $h_i \equiv h$ then $f(t) = \ln\cosh(t+\beta h)$ and after a calculation we get that $f^*(x) = \mathcal{I}_0(x) - \beta h x$, where $\mathcal{I}_0$ is given by (2). Combining this result with Theorem 2 we get the rate function for the last exercise in lecture notes 3.

A useful identity: $\ln\cosh(\operatorname{arctanh} x) = -\frac{1}{2}\ln(1-x^2)$.

Proof. The plan is to introduce a family of convenient reference measures which satisfies a LDP and then use tilting to get a desired LDP for the original measures.

In our previous considerations measures $(\mu_{n,0})_{n\in\mathbb{N}}$ were such reference measures. They were the distributions of $\overline{\sigma}_n$ under the uniform measure $\mathbb{P}_{n,0}$ on all spin configurations $\{-1,1\}^n$. Under this uniform measure spins σ_i, $1 \le i \le n$, were independent and identically distributed with $\mathbb{P}_{n,0}(\sigma : \sigma_i = 1) = 1/2$.

We consider measures $\mathbb{Q}_{n,\beta,\mathbf{h}}$ under which individual spins are still independent but not identically distributed:

$$\mathbb{Q}_{n,\beta,\mathbf{h}}(\sigma \in \Sigma_n : \sigma_i = \pm 1) = \frac{e^{\pm\beta h_i}}{2\cosh(\beta h_i)}.$$

Let $\nu_{n,\beta,\mathbf{h}}$ be the distribution of $\overline{\sigma}_n$ under $\mathbb{Q}_{n,\beta,\mathbf{h}}$. These are our new reference measures. We are going to apply Gärtner-Ellis theorem to $(\nu_{n,\beta,\mathbf{h}})_{n\in\mathbb{N}}$ to get a LDP. We need to compute $\Lambda_\beta(t)$. Using our assumption we find that for almost every $\omega \in \Omega$

$$\begin{aligned}
\Lambda_\beta(t) &:= \lim_{n\to\infty}\frac{1}{n}\ln\int_{\mathbb{R}} e^{ntx}\,d\nu_{n,\beta,\mathbf{h}} \\
&= \lim_{n\to\infty}\frac{1}{n}\ln\sum_{\sigma\in\Sigma_n} e^{t\sum_{i=1}^n\sigma_i}\,\mathbb{Q}_{n,\beta,\mathbf{h}}(\sigma) = \lim_{n\to\infty}\frac{1}{n}\ln\prod_{i=1}^n\frac{e^{\beta h_i + t} + e^{-\beta h_i - t}}{2\cosh(\beta h_i)} \\
&= \lim_{n\to\infty}(f_n(t) - f_n(0)) = f(t) - f(0)
\end{aligned}$$

for all $t \in \mathbb{R}$. Therefore, $D_{\Lambda_\beta} = \mathbb{R}$ and Λ_β is differentiable on $\mathbb{R}$. By Gärtner-Ellis theorem $(\nu_{n,\beta,\mathbf{h}})_{n\in\mathbb{N}}$ satisfy a LDP with rate

$$\Lambda_\beta^*(x) = f^*(x) + f(0).$$

According to Exercise 2, $D_{\Lambda_\beta^*} \subset [-1,1]$.

A simpler way to see this is to note that $\overline{\sigma}_n \in [-1,1]$, and, therefore, the rate function has to be infinite outside of $[-1,1]$.

Our next step is tilting. For every $A \in \mathcal{B}([-1,1])$

$$\begin{aligned}
\mu_{n,\beta,\mathbf{h}}(A) &= \sum_{\sigma:\overline{\sigma}_n\in A}\mathbb{P}_{n,\beta,\mathbf{h}}(\sigma) = \sum_{\sigma:\overline{\sigma}_n\in A}\frac{\mathbb{P}_{n,\beta,\mathbf{h}}(\sigma)}{\mathbb{Q}_{n,\beta,\mathbf{h}}(\sigma)}\mathbb{Q}_{n,\beta,\mathbf{h}}(\sigma) \\
&= \frac{2^n\prod_{i=1}^n\cosh(h_i\sigma_i)}{Z_{n,\beta,\mathbf{h}}}\sum_{\sigma:\overline{\sigma}_n\in A} e^{\frac{n\beta J}{2}(\overline{\sigma}_n)^2}\mathbb{Q}_{n,\beta,\mathbf{h}}(\sigma) \\
&= \frac{2^n\prod_{i=1}^n\cosh(h_i\sigma_i)}{Z_{n,\beta,\mathbf{h}}}\int_A e^{\frac{n\beta J}{2}x^2}d\nu_{n,\beta,\mathbf{h}} =: \frac{1}{Z^G_{n,\beta,\mathbf{h}}}\int_A e^{nG(x)}d\nu_{n,\beta,\mathbf{h}}, \quad\text{where} \\
G(x) &:= \frac{\beta J}{2}x^2,\ x\in[-1,1];\quad Z^G_{n,\beta,\mathbf{h}} := \frac{Z_{n,\beta,\mathbf{h}}}{2^n\prod_{i=1}^n\cosh(h_i\sigma_i)} = \int_{[-1,1]} e^{nG(x)}d\nu_{n,\beta,\mathbf{h}}.
\end{aligned}$$

By part (b) of Lemma 0.3 from lecture notes 3 we conclude that for almost every $\omega \in \Omega$ measures $(\mu_{n,\beta,\mathbf{h}})_{n\in\mathbb{N}}$ satisfy a LDP with rate function

$$\mathcal{I}^G(x) = \Lambda^*_\beta(x) - G(x) - \inf_{y\in[-1,1]} (\Lambda^*_\beta(y) - G(y)) = \mathcal{I}_{\beta,f}(x)$$

as claimed. □

References

Pierre Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.

Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 1998.

Richard S. Ellis. Large deviations for a general class of random vectors. *Ann. Probab.*, 12(1):1–12, 1984.

Jurgen Gärtner. On large deviations from an invariant measure. *Teor. Verojatnost. i Primenen.*, 22(1):27–42, 1977.

Matthias Löwe, Raphael Meiners, and Felipe Torres. Large deviations principle for Curie-Weiss models with random fields. *J. Phys. A*, 46 (12):125004, 10 pp., 2013.