Notes on Large Deviations, Γ-convergence and Entropy.

BY M. GUBINELLI

CEREMADE - UMR 7534
Université Paris-Dauphine

Email: m.gubinelli@ceremade.dauphine.fr
Web: http://www.ceremade.dauphine.fr/~mgubi

Abstract

This are the notes for some lectures on large deviations they are mainly a re-elaboration of lecture notes of Tsirelson. Like in the weak convergence approach of Dupuis and Ellis and in the idempotent probability approach of Puhalskii the theory is presented from the point of view of Varadhan lemma which inspire the idea of LD convergence. We describe also the interpretation of large deviations as the Γ-convergence of relative entropies due to M. Mariani. The basic facts and the main examples of applications are presented in the setting of compact metric spaces. Extension to non-compact spaces is left as an exercise.

(v.2 20111221)[1]

Table of contents

1 Introduction 2

2 The Large Deviation Principle 2

2.1 LD-convergence 2
2.2 Back to probabilities 4
2.3 The Gärtner-Ellis theorem and Cramérs theorem in $\mathbb{R}^d$ 5
2.4 The rate function is the Γ-limit of relative entropy 6
2.5 Properties of LD-converging sequences 7

3 Sanov's theorem 8

3.1 Restriction and conditioning 8
3.2 Tensorization and projections 9
3.3 Large deviations for coin tossing and Boltzmann discovery 9
3.4 Gibbsian conditioning 11

4 Large deviations for processes 12

4.1 Mogulskii theorem 12
4.2 Large deviations for random walks 13
4.3 Large deviations for Brownian motion 14

5 Other applications 15

5.1 Curie-Weiss model 15
5.2 Further Gibbsian conditioning 16

1. This paper has been typesetted with TEXMACS (www.texmacs.org)

1 Introduction

Laplace principle says that if $A \subseteq \mathbb{R}^d$ and φ is a measurable function, then

$$\lim_{n\to\infty} \log\left[\int_A e^{-n\varphi(x)}\mathrm{d}x\right]^{1/n} = -\mathrm{essinf}_{x\in A}\varphi(x)$$

that is asymptotically only the minima of the exponent contribute to the integral. Large deviation theory provide similar asymptotic for a more general class of measures and base spaces. The basic setting is that of a probability space $\mathcal{K}$ (Polish space with the Borel σ-algebra) endowed with a family of probability measures $(\mu_n)_{n\geqslant 1}$ where we are interested in the behavior of this family when $n \to \infty$. It can happen that these probabilities weakly converge to some degenerate probability (sometimes just a Dirac mass, other times maybe some convex combination). They describe a system which, as n grows, behaves more and more deterministically. The remarkable phenomenon is that sensible deviation to this deterministic behavior will take place at some characteristic scale with rates which decay exponentially as some function of n. These are large deviation and they are interesting both in the pure and applied math point of view. For example they describe the most-likely way a statistical physical system can deviate from the equilibrium behavior, how a chemical compound can dissociate and at which rate it happens, what are the most-likely ways an industrial plant or any other complex system can undergo a major failure or any global condition like congestion. Statistical mechanics can be seen as large deviation theory applied to a large assembly of simple mechanical elements. Large deviations describe irreversible phenomena and surprisingly also the stability of peculiar hydrodynamic systems like Jupiter's Red Spot. In these notes we describe large deviation theory starting from the Laplace principle. In other words we will consider particular averages of continuous functions, instead of the most conventional approach of looking at probabilities of small events. This has the advantage of simplifying the statement of the Large Deviation Principle (LDP). The key observation is a sequential compactness property: any sequence of probability measures has subsequences which satisfy the LDP with some rate function. Indeed large deviation theory is structurally very similar to the theory of weak convergence. This point of view has been exploited systematically by Dupuis and Ellis in their monograph. Given this compactness, establishing an LDP for the full sequence usually reduces to proving the uniqueness of the limiting rate function. Another very interesting relation (which explain many structural properties of large deviations) is due to Mauro Mariani: the LDP is equivalent to the Γ-convergence of the relative entropies of the sequence of measures and the rate function is just the corresponding Γ-limit. We end this paper with some applications in the form of exercises for the reader. Most of the material is a reelaboration of the very nice lecture notes of Boris Tsirelson for a course on Large Deviations which can be found on his home-page.

2 The Large Deviation Principle

Here we investigate general properties of large deviation phenomena. For the moment we will restrict ourselves to probability measures on a *compact* metrizable space $\mathcal{K}$ (a compact Polish space). We denote with $\Pi(\mathcal{K})$ the space of probability measures on $\mathcal{K}$ when endowed with the Borel σ-algebra and with $C(\mathcal{K})$ the space of (real valued) continuous function on $\mathcal{K}$.

2.1 LD-convergence

For any $\mu \in \Pi(\mathcal{K})$ we define the semi-norm $\|f\|_{L^p(\mu)} = [\mu(|f|^p)]^{1/p}$ on $C(\mathcal{K})$. It satisfy $\|1\| = 1$,

$$|f| \leqslant |g| \Rightarrow \|f\|_{L^p(\mu)} \leqslant \|g\|_{L^p(\mu)} \tag{1}$$

and

$$f, g \geqslant 0 \Rightarrow \|\max(f,g)\|_{L^p(\mu)} \leqslant 2^{1/p}\max(\|f\|_{L^p(\mu)}, \|g\|_{L^p(\mu)}). \tag{2}$$

Exercise 1. Prove these inequalities.

Exercise 2. Fix $1 \leqslant p < +\infty$. Show that the following two statements are equivalent:

a) $\forall f \in C(\mathcal{K}), \|f\|_{L^p(\mu_n)} \to \|f\|_{L^p(\mu)}$

b) $\mu_n \to \mu$ weakly

for all μ_n, $\mu \in \Pi(\mathcal{K})$. Convergence of L^p norms for all continuous functions is then equivalent to weak convergence. (b) $\Rightarrow$ (a) is easy. For the reverse implication use $f = |g|^p - |h|^p$.

Let us be given a sequence $(\mu_n)_{n\geqslant 1}$ for probabilities, assume that there exists a sequence of positive numbers $(p_n)_{n\geqslant 1}$ with $p_n \to +\infty$ such that

$$\lim_n \|f\|_{L^{p_n}(\mu_n)}$$

exists for all $f \in C(\mathcal{K})$. Then this limit is another semi-norm on $C(\mathcal{K})$ which we denote by $\|f\|_*$. It will satisfy $\|1\|_* = 1$, the inequality (1) and

$$\|\max(f, g)\|_* \leqslant \max(\|f\|_*, \|g\|_*)$$

for all $f, g \in C(\mathcal{K})$.

Definition 1. *A sequence $(\mu_n)_{n\geqslant 1}$ of probability measures on a compact metrizable space $\mathcal{K}$ is said* LD*-convergent if the limit*

$$\lim_n \|f\|_{L^n(\mu_n)} = \|f\|_*$$

exists for all $f \in C(\mathcal{K})$.

Theorem 2. *Let $\|\cdot\|_*$ be a semi-norm on $C(\mathcal{K})$ satisfying for all $f, g \in C(\mathcal{K})$*

$$\|1\|_* \leqslant 1,\ |f| \leqslant |g| \Rightarrow \|f\|_* \leqslant \|g\|_* \text{ and } \|\max(f, g)\|_* \leqslant \max(\|f\|_*, \|g\|_*). \tag{3}$$

Let $\Pi: \mathcal{K} \to [0, 1]$ such that $1/\Pi(x) = \sup_{\|f\|_* \leqslant 1} f(x)$ then the function Π is usc and

$$\|f\|_* = \max_{x\in\mathcal{K}} [|f(x)|\Pi(x)] \tag{4}$$

(supremum is reached by upper semi-continuity).

Proof. $1/\Pi(x)$ is the supremum of continuous functions so it is lsc. Let us prove eq. (4). It is enough to consider $f \geqslant 0$. By definition $1 \geqslant \Pi(x) f(x)$ for all $x \in \mathcal{K}$ and $f \geqslant 0$ such that $\|f\|_* \leqslant 1$ this implies that $\|f\|_* \geqslant \Pi(x) f(x)$ for all f and then $\|f\|_* \geqslant \sup_x \{\Pi(x) f(x)\}$. To prove the reverse inequality let $C_f = \max_{x\in\mathcal{K}} [f(x)\Pi(x)]$ so for all x, $f(x) \leqslant C_f/\Pi(x)$ and so there exist g with $\|g\|_* \leqslant 1$ such that $f(x) < C_f g(x) + \varepsilon$. By continuity of f and g this continues to hold in a neighborhood of x and by repeating the argument and by the compactness of $\mathcal{K}$ it is possible to find a finite number of functions $\{g_k\}_{k=1,\dots,n}$ such that $\|g_k\|_* \leqslant 1$ and $f(x) < C_f \max_k g_k(x) + \varepsilon$ for all $x \in \mathcal{K}$. Now by the properties of the semi-norm we have

$$\|f\|_* \leqslant C_f \|\max_k g_k\|_* + \varepsilon \leqslant C_f \max_k \|g_k\|_* + \varepsilon \leqslant C_f + \varepsilon$$

and by the arbitrariness of ε we can conclude. □

Exercise 3. Assume that $\Pi_1, \Pi_2: \mathcal{K} \to [0, 1]$ are upper semi-continuous. Prove that if $\max_{x\in\mathcal{K}} [|f(x)|\Pi_1(x)] = \max_{x\in\mathcal{K}} [|f(x)|\Pi_2(x)]$ for all $f \in C(\mathcal{K})$ then $\Pi_1 = \Pi_2$. (Hint: use $f(x) = \exp(-Md(x, x_0))$ for large M assuming that $\Pi_1(x_0) < \Pi_2(x_0)$).

It is suggestive to introduce the lower semi-continuous function $I: \mathcal{K} \to \mathbb{R}_+$ such that $\Pi(x) = e^{-I(x)}$. Such a function is called a *rate function*. It defines a semi-norm on $C(\mathcal{K})$ by

$$\|f\|_I = \max_{x\in\mathcal{K}} [|f(x)|e^{-I(x)}].$$

Definition 3. *The sequence $(\mu_n)_{n\geqslant 1}$ satisfy a LDP (large deviation principle) with a rate function $I: \mathcal{K} \to \mathbb{R}_+$ (lower semi-continuous) if, for all $f \in C(\mathcal{K})$,*

$$\lim_n \|f\|_{L^n(\mu_n)} = \|f\|_I.$$

Then we have proved that

Corollary 4. *If $(\mu_n)_{n\geqslant 1}$ is LD-convergent then it satisfy the LDP with one and only one rate function I given by $I(x)=\log\sup_{\|f\|_*\leqslant 1} f(x)$.*

Theorem 5. *Let $\{N_n\}_{n\geqslant 1}$ be a sequence of semi-norms satisfying $N_n(1)\leqslant 1$ and $|f|\leqslant |g| \Rightarrow N_n(f)\leqslant N_n(g)$. There there exists a converging subsequence $\{N'_n\}\subseteq\{N_n\}$, i.e. such that $\lim_n N'_n(f)=N'_\infty(f)$ exists for every $f\in C(\mathcal{K})$.*

Proof. Take a sequence N_n of such semi-norms. Note that $N_n(f)\leqslant\|f\|_\infty$ since $|f|\leqslant\|f\|_\infty$. Let $\{f_k\}_{k\geqslant 1}$ be a countable dense set in $C(\mathcal{K})$ for the sup norm. Take a subsequence $\{N'_n\}\subseteq\{N_n\}$ so that $S(f_k)=\lim_n N'_n(f_k)$ exists for every k. For every f take ℓ such that $\|f-f_\ell\|_\infty<\varepsilon$ and note that $|N'_n(f)-N'_k(f)|\leqslant|N'_n(f-f_\ell)|+|N'_n(f_\ell)-N'_k(f_\ell)|+|N'_k(f_\ell-f)|\leqslant 2\varepsilon+|N'_n(f_\ell)-N'_k(f_\ell)|\to 2\varepsilon$ for $n,k\to\infty$. The sequence $\{N'_n(f)\}$ is Cauchy for every f and we denote $N'_\infty(f)=\lim_n N'_n(f)$ its limit. □

Remark 6. In fact the space of all semi-norms $N\colon C(\mathcal{K})\to\mathbb{R}_+$ satisfying $N(1)\leqslant 1$ and $|f|\leqslant|g|\Rightarrow N(f)\leqslant N(g)$ is a compact metrizable space. (Without proof)

Exercise 4. Let $\mathcal{K}=[0,1]$ and μ_n the Lebesgue measure for every $n\geqslant 1$. Prove that $(\mu_n)_{n\geqslant 1}$ satisfy the LDP with rate function $I(x)=0$.

Exercise 5. Let $\mathcal{K}=[0,1]$ and $\mu_\alpha\in\Pi(\mathcal{K})$ given by $\mu_\alpha(\mathrm{d}x)=(\alpha+1)\,x^\alpha\mathrm{d}x$. Prove that each of the following sequences are LD-convergent and find the related rate functions: $(\mu_n)_n$, $(\mu_{2n})_n$, $(\mu_{n^2})_n$, $(\mu_{\sqrt{n}})_n$.

Exercise 6. Prove that if $(\mu_n)_n$ is LD-convergent with rate I, then $(\mu_{2n})_n$ is LD-convergent with rate $2I$. (Hint: $\|f\|_{L^n(\mu_{2n})}=\||f|^{1/2}\|_{L^{2n}(\mu_{2n})}$)

Exercise 7. Let $\mathcal{K}=[0,1]$ and $(\mu_n)_n$ a sequence satisfying the LDP with rate function $I(x)=\log(1/x)$. Prove that $\mu_n([0,0.5])<0.6^n$ for all n sufficiently large. (Hint: take f such that $f(x)=1$ in $[0,0.5]$ and $f(x)=0$ in $[0.55,1]$).

Exercise 8. Prove that $\inf_{\mathcal{K}} I=0$ (Hint: try $f=1$). Prove that for every $\varepsilon>0$, $\mu_n(\{x\colon I(x)\leqslant\varepsilon\})\to 1$ for $n\to\infty$. (Hint: first prove it for I continuous using the Markov inequality and $f=e^I$, then use the fact that any lsc function is the point-wise limit of an increasing sequence of continuous functions).

2.2 Back to probabilities

Lemma 7. *Let $\varphi_n,\varphi\colon\mathcal{K}\to\mathbb{R}$ and $\varphi_n\uparrow\varphi$ point-wise, then $\sup_{\mathcal{K}}\varphi_n\uparrow\sup_{\mathcal{K}}\varphi$. If $\varphi_n\downarrow\varphi$ and φ_n,φ are usc then $\max_{\mathcal{K}}\varphi_n\downarrow\max_{\mathcal{K}}\varphi$.*

Proof. First statement. For every $\varepsilon>0$ take x such that $\varphi(x)>\sup_{\mathcal{K}}\varphi-\varepsilon/2$ and n such that $\varphi_n(x)>\varphi(x)-\varepsilon/2$ then $\sup_{\mathcal{K}}\varphi_n>\sup_{\mathcal{K}}\varphi-\varepsilon$. Second statement. Let $c=\lim_n\max_{\mathcal{K}}\varphi_n$. For every $\varepsilon>0$ the sets $\{x\colon\varphi_n(x)\geqslant c-\varepsilon\}$ form a decreasing sequence of non-empty closed sets. By compactness some x belongs to all the sets, thus $\varphi(x)=\lim_n\varphi_n(x)\geqslant c-\varepsilon$ and $\max_{\mathcal{K}}\varphi>c-\varepsilon$. □

Exercise 9. Semicontinuity is necessary in the second statement of the previous lemma. Find a counter-example.

Lemma 8. *Let $f\colon\mathcal{K}\to\mathbb{R}$, $f\geqslant 0$*

a) If f is lsc then $\liminf_n\|f\|_{L^n(\mu_n)}\geqslant\sup_{\mathcal{K}}(fe^{-I})$;

b) If f is usc then $\limsup_n\|f\|_{L^n(\mu_n)}\leqslant\max_{\mathcal{K}}(fe^{-I})$.

Proof. a) Take $f_n\in C(\mathcal{K})$, $0\leqslant f_n\uparrow f$. As $j\to\infty$

$$\liminf_n\|f\|_{L^n(\mu_n)}\geqslant\liminf_n\|f_j\|_{L^n(\mu_n)}=\max_{\mathcal{K}}(f_j\,e^{-I})\uparrow\sup_{\mathcal{K}}(fe^{-I}).$$

Part b) is similar using the semi-continuity of the limit f. □

Corollary 9.

a) $\liminf_n(\mu_n(G))^{1/n}\geqslant\exp(-\inf_G I)$ for every open set $G\subset\mathcal{K}$;

b) $\limsup_n(\mu_n(F))^{1/n}\leqslant\exp(-\inf_F I)$ for every closed set $F\subset\mathcal{K}$;

c) If an open set $G \subset \mathcal{K}$ satisfy $\inf_G I = \inf_{\bar{G}} I$ then

$$\lim_n (\mu_n(G))^{1/n} = \lim_n (\mu_n(\bar{G}))^{1/n} = \exp(-\inf_G I) = \exp(-\min_{\bar{G}} I).$$

Exercise 10. Note that $\inf_G I = \inf_{\bar{G}} I$ for open G does not imply that $\mu_n(G)/\mu_n(\bar{G}) \to 1$. Find a counterexample. (Hint: try $\mathcal{K} = [0,1]$, $G = (0,1]$ and choose μ_n as a mixture of Lebesgue measure and an atom in 0 with appropriate weights).

Choose a metric d on $\mathcal{K}$ and consider open and closed balls $B_{x,r-} = \{y \in \mathcal{K} : d(x,y) < r\}$ and $B_{x,r+} = \{y \in \mathcal{K} : d(x,y) \leqslant r\}$. We can describe the rate function I in terms of probability decay of such balls. In particular the following holds.

Proposition 10. *For every $x \in \mathcal{K}$ there exists a function $N : (0,1) \to \mathbb{N}$ such that*

$$\lim_{r \to 0+} \sup_{n \geqslant N(r)} \left| \frac{1}{n} \log \mu_n(B_{x,r\pm}) + I(x) \right| = 0.$$

Proof. By the probability decay rate estimates we have that there exists $N(r)$ such that for all $n \geqslant N(r)$ we have

$$\exp(-\inf_{B_{x,r-}} I) - r \leqslant (\mu_n(B_{x,r-}))^{1/n} \leqslant (\mu_n(B_{x,r-}))^{1/n} \leqslant \exp(-\inf_{B_{x,r+}} I) + r.$$

By lsc-ity of I, $\inf_{B_{x,r\pm}} I \to I(x)$ for $r \to 0+$, therefore the statement. □

In fact decay of small balls determine LD-convergence.

Proposition 11. *Assume that for every $x \in \mathcal{K}$*

$$\liminf_{r \to 0+} \liminf_n \log(\mu_n(B_{x,r-}))^{1/n} = \limsup_{r \to 0+} \limsup_n \log(\mu_n(B_{x,r+}))^{1/n} = -I(x)$$

then the sequence $(\mu_n)_n$ obey the LDP with rate function I.

Proof. By compactness we can extract a sub-sequence obeying the LDP with rate function I' but by the above proposition

$$-I(x) \leqslant \lim_{r \to 0+} \liminf_{n_k} \log(\mu_{n_k}(B_{x,r-}))^{1/n_k} = -I'(x) = \lim_{r \to 0+} \limsup_{n_k} \log(\mu_{n_k}(B_{x,r+}))^{1/n_k} \leqslant -I(x).$$

□

2.3 The Gärtner-Ellis theorem and Cramérs theorem in $\mathbb{R}^d$

Let us give a first application of the sequential compactness result above: LD-convergence needs convergence of $\|f\|_{L^n(\mu_n)}$ for all continuous functions. Here we will see that in the vector space case and with enough regularity it is enough to check convergence of exponentials of linear functionals.

Theorem 12. (Gärtner-Ellis) *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be compact and let X_n be $\mathcal{K}$-valued r.v.s such that for all $\lambda \in \mathbb{R}^d$*

$$\lim_n \frac{1}{n} \log \mathbb{E}[e^{n \langle \lambda, X_n \rangle}] = G(\lambda)$$

where $G : \mathbb{R}^d \to \mathbb{R}_+$ is a differentiable (convex) function. Then $\{X_n\}_{n \geqslant 1}$ obey the LDP with the (convex) rate function $I(x) = \sup_{\lambda \in \mathbb{R}^d} [\langle \lambda, x \rangle - G(\lambda)]$.

Proof. Given the existence of the limit above, convexity of G is just a consequence of Hölder inequality. By compactness there exists a subsequence $\{X_{n_k}\}$ which satisfy the LDP for some rate function $I' : \mathbb{R}^d \to \mathbb{R}_+$ which can be taken equal to $+\infty$ outside $\mathcal{K}$. Then by definition

$$G(\lambda) = \lim_k \frac{1}{n_k} \log \mathbb{E}[e^{n_k \langle \lambda, X_{n_k} \rangle}] = \sup_{x \in \mathcal{K}} (\langle \lambda, x \rangle - I'(x)) = \sup_{x \in \mathbb{R}^d} (\langle \lambda, x \rangle - I'(x))$$

To conclude that $I = I'$ we need to prove that the Fenchel-Legendre transform is invertible at G. Our hypothesis guarantees that for each x there exists $\lambda_x \in \mathbb{R}^d$ such that $I(x) = \langle \lambda_x, x\rangle - G(\lambda_x)$ and that $I(y) > I(x) + \langle \lambda_x, y - x\rangle$ for all $y \neq x$ (prove it!). Then for all $y \in \mathbb{R}^d$, $G(\lambda_x) \geqslant \langle \lambda_x, y\rangle - I'(y)$ so that $I'(x) \geqslant \langle \lambda_x, x\rangle - G(\lambda_x) = I(x)$. On the other hand, by compactness, there exists $\hat{y} \in \mathcal{K}$ such that, if $\hat{y} \neq x$,

$$\langle \lambda_x, \hat{y} - x\rangle - I'(\hat{y}) = G(\lambda_x) - \langle \lambda_x, x\rangle = -I(x) > -I(\hat{y}) + \langle \lambda_x, \hat{y} - x\rangle$$

which means that $I'(\hat{y}) < I(\hat{y})$, in contradiction to the fact that $I' \geqslant I$ on $\mathcal{K}$. So we must have $x = \hat{y}$ but then $I(x) = \langle \lambda_x, x\rangle - G(\lambda_x) = I'(x)$ which concludes the proof. □

Of course a basic corollary is the multidimensional Cramér theorem.

Corollary 13. (Cramér-Varadhan) *Let $(X_n)_{n\geqslant 1}$ be an iid sequence with values in the compact $\mathcal{K} \subseteq \mathbb{R}^d$. Let $\Lambda(\lambda) = \log \mathbb{E}[\exp\langle \lambda, X_1\rangle]$, $I(x) = \sup_\lambda [\langle \lambda, x\rangle - \Lambda(\lambda)]$ and μ_n be the law of the empirical mean $S_n = (X_1 + \cdots + X_n)/n$. Then the sequence $(\mu_n)_n$ satisfy the LDP with the convex rate function I.*

Proof. Exercise. You need only to verify that $G(\lambda) = \Lambda(\lambda)$ and justify its smoothness. □

2.4 The rate function is the Γ-limit of relative entropy

Given a probability measure μ on the compact $\mathcal{K}$ we can define the relative entropy $H_\mu : \Pi(\mathcal{K}) \to [0, +\infty]$ as

$$H_\mu(\nu) = \sup_\varphi [\nu(\varphi) - \log \mu(e^\varphi)]$$

where the supremum is taken wrt all continuous functions on $\mathcal{K}$. Then

$$\mathrm{H}_\mu(\nu) = \begin{cases} \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\mathrm{d}\mu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

The function H_μ is strictly convex and $\mathrm{H}_\mu(\nu) = 0$ iff $\nu = \mu$, moreover

$$\log \mu(e^\varphi) = \sup_\nu [\nu(\varphi) - H_\mu(\nu)]$$

where the supremum runs over all probability measures on $\mathcal{K}$. Consider a sequence $(\mu_n)_n$ and define $H_n(\nu) = H_{\mu_n}(\nu)/n$, then $H_n(\nu) = \sup_\varphi [\nu(\varphi) - \log \|e^\varphi\|_{L^n(\mu_n)}]$ and

$$\log \|e^\varphi\|_{L^n(\mu_n)} = \sup_{\nu \in \Pi(\mathcal{K})} [\nu(\varphi) - H_n(\nu)]$$

Assume that $(\mu_n)_n$ obey the LDP with rate I, then we have

$$\lim_n \log \|e^\varphi\|_{L^n(\mu_n)} = \sup_{x \in \mathcal{K}} (\varphi(x) - I(x)) = \sup_{\nu \in \Pi(\mathcal{K})} [\nu(\varphi) - \nu(I)]$$

where we introduce a maximization over probabilities to stress the similarity with the variational representation at finite n.

Rate function can be seen as asymptotic entropy in Γ-convergence sense. Γ-convergence is a variational convergence introduced by E. de Giorgi. In a metric space $\mathcal{M}$ the Γ-convergence of a sequence of positive functionals $I_n : \mathcal{M} \to \mathbb{R}_+$ is defined via

$$\Gamma\liminf_n I_n(x) = \inf\{\liminf_n I_n(x_n) : x_n \to x\}$$

and

$$\Gamma\limsup_n I_n(x) = \inf\{\limsup_n I_n(x_n) : x_n \to x\}$$

When $\Gamma\limsup I_n = \Gamma\liminf I_n$ we denote the common limit as $\Gamma\lim I_n$. The Γ limits are lsc functions. Here, as before, we restrict ourselves to a compact space $\mathcal{M}$. We will be interested in the case where $\mathcal{M} = \Pi(\mathcal{K})$ for some compact metrizable $\mathcal{K}$.

Theorem 14. *For every $J:\mathcal{M}\to\mathbb{R}$ and $x\in\mathcal{M}$*

a) $J(x)\leqslant \Gamma\liminf_n I_n(x) \Leftrightarrow \forall x_n\to x:\ \liminf_n I_n(x_n)\geqslant J(x)$;

b) $\Gamma\limsup_n I_n(x)\leqslant J(x)\Leftrightarrow \exists x_n\to x:\limsup_n I_n(x_n)\leqslant J(x)$.

$\Gamma\liminf$ *and* $\Gamma\limsup$ *are respectively the largest and smallest lsc functions satisfying the above properties.*

Then $\Gamma\lim_n I_n = I$ if and only if for all x_n, x such that $x_n\to x$ we have $\liminf_n I_n(x_n)\geqslant I(x)$ and moreover for every x there exists $x_n\to x$ such that $\limsup_n I_n(x_n)\leqslant I(x)$.

Theorem 15. *Properties of Γ-convergence :*

a) $\min_{\mathcal{M}} I = \lim_n (\min_{\mathcal{M}} I_n)$

b) *If $I_n \overset{\Gamma}{\to} I$, x_n minimizer of I_n and $x_n\to x \Rightarrow x$ minimizer of I*

c) *If G is continuous and $I_n\overset{\Gamma}{\to} I$ then $I_n+G\overset{\Gamma}{\to} I+G$*

d) *If $\mathcal{M}$ is separable from every sequence I_n we can extract a Γ-converging subsequence.*

Example 16. Let $\mathcal{M}=[0,1]$ and $I_n(x)=\sin^2(2\pi n\, x)$ then $\Gamma\lim_n I_n(x)=0$ for all $x\in\mathcal{M}$.

Theorem 17. *(μ_n) obey the LDP with rate function I iff $\Gamma\lim_n H_n = I$.*

Proof. Let us prove that LDP $\Rightarrow$ Γ-convergence. By the variational characterization $H_n(\nu_n)\geqslant \nu_n(\varphi)-\log\|e^{\varphi}\|_n$ for every $\varphi\in C(\mathcal{K})$. If $\nu_n\to\nu$ then, by LD convergence, $\liminf_n H_n(\nu_n)\geqslant \nu(\varphi)-\sup_{\mathcal{K}}(\varphi-I)$. Since I is lsc there exists a sequence of continuous functions $\varphi_n\uparrow I$. Then

$$\liminf_n H_n(\nu_n)\geqslant \lim_n\,[\nu(\varphi_n)-\sup_{\mathcal{K}}(\varphi_n-I)]=\nu(I)$$

since $\sup_{\mathcal{K}}(\varphi_n-I)\to 0$. Now for every φ there exists $(\nu_n)_n$ such that $\|e^{\varphi}\|_n=\nu_n(\varphi)-H_n(\nu_n)$. By compactness we can take a subsequence still denoted ν_n such that $\nu_n\to\nu$ for some ν. By LD-convergence we get $\lim_n H_n(\nu_n)=\nu(\varphi)-\sup_{\mathcal{K}}(\varphi-I)$. But by the previous bound $\nu(\varphi)-\sup_{\mathcal{K}}(\varphi-I)\geqslant\nu(I)$ which implies that $\nu(\varphi-I)\geqslant\sup_{\mathcal{K}}(\varphi-I)$, so the limit point ν must be concentrated on the maxima of $\varphi-I$. Take φ such that this function has a unique maximum at x to obtain that for every $x\in\mathcal{K}$ there exists $\nu_n\to\delta_x$ for which $\lim_n H_n(\nu_n)\leqslant I(x)$. By convexity, for all $\nu\in\Pi(\mathcal{K})$, $\Gamma\limsup_n H_n(\nu) = \Gamma\limsup_n H_n(\int\ \delta_y\ \nu(\mathrm{d}y))\leqslant\int\nu(\mathrm{d}y)\ \Gamma\limsup_n H_n(\delta_y)\leqslant\nu(I)$. The reverse implication is a direct consequence of the variational properties of Γ-convergence:

$$\sup_{\nu}(\nu(\varphi)-\nu(I))=\limsup_n\sup_{\nu}(\nu(\varphi)-H_n(\nu))=\lim_n\log\|e^{\varphi}\|_n$$

for every $\varphi\in C(\mathcal{K})$ which by density implies LD-convergence with rate function I. □

Remark 18. Compactness of LD-convergence can be obtained as consequence of the compactness of Γ-convergence.

2.5 Properties of LD-converging sequences

Theorem 19. *(Contraction Principle) Let $\mathcal{K}'$ be another compact metrizable space and F: $\mathcal{K}\to\mathcal{K}'$ be a continuous function, $(\mu_n)_n$ a sequence of measures on $\mathcal{K}$ satisfying the LDP with rate function I, and $\nu_n=F_*\mu_n\in\Pi(\mathcal{K}')$ the image measure of μ_n wrt. F, i.e. $\nu_n(A)=\mu_n(F^{-1}(A))$ for all Borel sets $A\subseteq\mathcal{K}'$. Then $(\nu_n)_n$ satisfies the LDP with rate function*

$$I'(y)=\min_{F^{-1}(y)} I=\min\{I(x):x\in F^{-1}(\{y\})\}$$

where the minimum is by definition $+\infty$ if $F^{-1}(\{y\})=\varnothing$. Otherwise the minimum is attained since $F^{-1}(\{y\})$ is compact and I is lsc.

Proof. Exercise. Use the fact that for all $g\in C(\mathcal{K}')$ the change of variables formula gives $\|g\|_{L^n(\nu_n)}=\|g\circ F\|_{L^n(\mu_n)}$. □

Actually, in the contraction principle only continuity of F at the points where $I<+\infty$ matters:

Theorem 20. *(CONTRACTION PRINCIPLE II) Let $F:\mathcal{K}\to\mathcal{K}'$ be a function which is continuous on $\mathcal{K}_0\subseteq\mathcal{K}$. Let $(\mu_n)_n\subseteq\Pi(\mathcal{K})$ obey the LDP with rate function I such that $I(x)=+\infty$ for every $x\notin\mathcal{K}_0$. Then $\nu_n=F_*\mu_n$ satisfy the LDP with rate function $I'(y)=\min_{F^{-1}(y)}I$.*

Proof. Take a function $g>0$ continuous on $\mathcal{K}_0$ and consider the functions $f=g\circ F$ and

$$(f)_k^+(x)=\sup_{x'\in\mathcal{K}}\left[e^{-kd(x,x')}f(x')\right],\qquad (f)_k^-(x)=\inf_{x'\in\mathcal{K}}\left[e^{kd(x,x')}f(x')\right]$$

The functions $f_k^\pm$ are continuous, $f_k^+\geqslant f\geqslant f_k^-$ and $f_k^+\downarrow f$, $f_k^-\uparrow f$ point-wise on $\mathcal{K}_0$ as $k\to\infty$ (check it: use the triangular inequality for d). We have $\|(f)_k^-\|_{L^n(\mu_n)}\leqslant\|g\circ F\|_{L^n(\nu_n)}\leqslant\|(f)_k^+\|_{L^n(\mu_n)}$ so it remains to prove that

$$\sup_\varepsilon\sup_{\mathcal{K}}\left[(f)_k^-e^{-I}\right]=\inf_\varepsilon\sup_{\mathcal{K}}\left[(f)_k^+e^{-I}\right]=\sup_{\mathcal{K}_0}\left[(g\circ F)\,e^{-I'}\right]=\sup_{\mathcal{K}'}\left[ge^{-I'}\right]$$

to get existence of the limit and the identification of the rate function I'. Now you can check that by Lemma 7 we have $\sup_{\mathcal{K}}\left[(f)_k^+e^{-I}\right]=\sup_{\mathcal{K}_0}\left[(f)_k^+e^{-I}\right]\downarrow\sup_{\mathcal{K}_0}\left[e^{-I}f\right]$ since the functions $(f)_k^+$, f are continuous on $\mathcal{K}_0$ and $\sup_{\mathcal{K}}\left[(f)_\varepsilon^-e^{-I}\right]=\sup_{\mathcal{K}_0}\left[(f)_\varepsilon^-e^{-I}\right]\uparrow\sup_{\mathcal{K}_0}\left[fe^{-I}\right]$ by the (increasing) point-wise convergence of $(f)_k^-$ towards f. □

Theorem 21. *(CHANGE OF MEASURE) Let $(\mu_n)_n$ and $(\nu_n)_n$ two sequences of probabilities on $\mathcal{K}$ such that, for all n,*

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}=c_n\,e^{-nh}$$

for some $h\in C(\mathcal{K})$ and $(c_n)_n$. Then if $(\mu_n)_n$ obey the LDP with rate function I then $(\nu_n)_n$ obey the LDP with rate function J given by

$$J=I+h-\min_{\mathcal{K}}(I+h)=I+h-\lim_n\frac{\log c_n}{n}.$$

Proof. Exercise. Use the fact that $\|f\|_{L^n(\nu_n)}=\|fe^{-h}\|_{L^n(\mu_n)}/\|e^{-h}\|_{L^n(\mu_n)}$. □

3 Sanov's theorem

3.1 Restriction and conditioning

Consider a sequence $(\mu_n)_n$ satisfying the LDP with rate function I and an open set $G\subset\mathcal{K}$ such that $\inf_G I=\inf_{\bar{G}}I$.

Proposition 22. *(RESTRICTION) For every $f\in C(\mathcal{K})$*

$$\lim_n\|1_G f\|_{L^n(\mu_n)}=\lim_n\|1_{\bar{G}}f\|_{L^n(\mu_n)}=\sup_G(|f|e^{-I})=\max_{\bar{G}}(|f|e^{-I}).$$

Proof. We can restrict ourselves to $f\geqslant 0$ and by density assume that $f>0$. We thus write $f=e^{-h}$ with $h\in C(\mathcal{K})$ and define new probability measures by $\nu_n=c_n\,e^{-h}\mu_n$ for suitable constants c_n. The sequence $(\nu_n)_n$ satisfy the the LDP with rate function $J=I+h-\min_{\mathcal{K}}(I+h)$, since h is continuous $\inf_G J=\inf_{\bar{G}}J$ so

$$\|1_G f\|_{L^n(\mu_n)}=\|f\|_{L^n(\mu_n)}(\nu_n(G))^{1/n}\to\exp\left(-\inf_G J-\min_{\mathcal{K}}(I+h)\right)=\exp\left(-\inf_G(I+h)\right).$$

□

If $I(x)<+\infty$ for some $x\in\bar{G}$ then $\mu_n(\bar{G})>0$ for n large enough and we can introduce conditional measures ν_n such that $\nu_n(f)=\mu_n(f1_{\bar{G}})/\mu_n(\bar{G})$ for all f bounded Borel on $\mathcal{K}$. The set $\bar{G}$ is another compact metrizable space.

Corollary 23. *Assume $\min_{\bar{G}}I<+\infty$. The sequence of conditional measures $(\nu_n)_n$ obey the LDP with rate function $J:\bar{G}\to[0,+\infty]$ given by $J(x)=I(x)-\min_{\bar{G}}I$ for all $x\in\bar{G}$.*

Proof. Take $f \in C(\bar{G})$ and let $\hat{f} \in C(\mathcal{K})$ any continuous extension of f (which exists for example due to separability of $C(\mathcal{K})$, think about it). Then $\|\hat{f}1_{\bar{G}}\|_{L^n(\mu^n)} \to \max_{\bar{G}}(|\hat{f}|e^{-I})$ so $\|f\|_{L^n(\nu_n)} = \|\hat{f}1_{\bar{G}}\|_{L^n(\mu^n)}/\|1_{\bar{G}}\|_{L^n(\mu^n)} \to \max_{\bar{G}}(|f|e^{-I})/\max_{\bar{G}}(e^{-I})$. □

3.2 Tensorization and projections

Theorem 24. *Consider two compact Polish spaces $\mathcal{K}_1$ and $\mathcal{K}_2$. Let $(\mu_n^1)_n$ and $(\mu_n^2)_n$ be sequences resp in $\Pi(\mathcal{K}_1)$ and $\Pi(\mathcal{K}_2)$ which obey the LDP with rate functions I_1 and I_2. Then the sequence $(\nu_n = \mu_n^1 \times \mu_n^2)_n$ in $\Pi(\mathcal{K}_1 \times \mathcal{K}_2)$ obey the LDP with rate function $I(x_1, x_2) = I_1(x_1) + I_2(x_2)$.*

Proof. Take $f(x_1, x_2) = (f_1 \otimes f_2)(x_1, x_2) = f_1(x_1) f_2(x_2)$. Then for any LD-converging sub-sequence of $(\nu_n)_n$ we have, for some rate function I',

$$\lim_k \|f_1 \otimes f_2\|_{L^{n_k}(\nu_{n_k})} = \sup_{x \in \mathcal{K}_1 \times \mathcal{K}_2} (f_1(x_1) f_2(x_2) e^{-I'(x_1, x_2)}).$$

On the other hand $\|f_1 \otimes f_2\|_{L^{n_k}(\nu_{n_k})} = \|f_1\|_{L^{n_k}(\mu^1_{n_k})} \|f_2\|_{L^{n_k}(\mu^2_{n_k})}$ and then

$$\sup_{x \in \mathcal{K}_1 \times \mathcal{K}_2} (f_1(x_1) f_2(x_2) e^{-I'(x_1, x_2)}) = \sup_{x_1 \in \mathcal{K}_1} (f_1(x_1) e^{-I_1(x_1)}) \sup_{x_2 \in \mathcal{K}_2} (f_2(x_2) e^{-I_2(x_2)})$$

For some fixed $z \in \mathcal{K}_1 \times \mathcal{K}_2$ choose $f_i(x_i) = \exp(-N d_i(x_i, z_i))$ for $i = 1, 2$. Letting $N \to \infty$ we get

$$I'(z_1, z_2) = I_1(z_1) + I_2(z_2) = I(z)$$

(prove it!) thus all possible accumulation points of $(\nu_n)_n$ have the same rate functions so the whole sequence satisfy the LDP with rate function I. □

Exercise 11. Prove that $(\mu_n^1 \times \mu_n^2)_n$ is LD-convergent if and only if $(\mu_n^1)_n$ and $(\mu_n^2)_n$ are LD-convergent.

Theorem 25. (Dawson-Gartnër) *Consider a sequence of measures $(\mu_n)_n$. Let $\{g_k\}_{k \geqslant 1} \subseteq C(\mathcal{K})$ be a family of continuous functions which separates the points of $\mathcal{K}$. Define G_k: $\mathcal{K} \to \mathbb{R}^k$ as $G_k(x) = (g_1(x), ..., g_k(x))$. Assume that for all $k \geqslant 1$ the laws $\mu_n^k = (G_k)_* \mu_n$ of the vector $(g_1, ..., g_k)$ obey the LDP with rate function I_k on the compact set $G_k(\mathcal{K})$. Then $(\mu_n)_n$ satisfy the LDP with rate function*

$$I(x) = \sup_k I_k(G_k(x)).$$

Proof. By the Stone-Weierstrass theorem the functions of the form $f = g(G_k)$ are dense in $C(\mathcal{K})$ and the limit $\lim_n \|g \circ G_k\|_n$ exists. Convergence of $\|f\|_n$ for a dense set of f imply LD-convergence. Let us call I' the rate function, then

$$I'(x) = \log \sup \{f(x) : \lim_n \|f\|_n \leqslant 1\} \geqslant \log \sup \{g(G_k(x)) : \lim_n \|g \circ G_k\|_n \leqslant 1\} = I_k(G_k(x))$$

so $I'(x) \geqslant \sup_k I_k(G_k(x)) = I(x)$. Now for every $f \in C(\mathcal{K})$ such that $\lim_n \|f\| \leqslant 1$ choose k and g such that $\|f - g \circ G_k\|_\infty \leqslant \varepsilon$. Then

$$f(x) \leqslant g(G_k(x)) + \varepsilon \leqslant \lim_n \|g \circ G_k\|_n e^{I_k(G_k(x))} + \varepsilon \leqslant (1 + \varepsilon) e^{I(x)} + \varepsilon$$

so $I'(x) \leqslant \log[(1 + \varepsilon) e^{I(x)} + \varepsilon]$ for arbitrary $\varepsilon > 0$. Then $I = I'$. □

3.3 Large deviations for coin tossing and Boltzmann discovery

Let $(X_n)_{n \geqslant 1}$ be an iid sequence with law Bernoulli(p) for some $p \in [0, 1]$. Consider the r.v. $N_n = \sum_{k=1}^n X_n = \#\{X_k = 1 : 1 \leqslant k \leqslant n\}$ which counts the number of ones in the sequence. Of course $N_n \sim B(n, p)$ and if μ_n is the law of N_n/n we have

$$\|f\|_n = \left[\sum_{k=0}^n |f(k/n)|^n \binom{k}{n} p^k (1-p)^{n-k} \right]^{1/n}$$

Recall that given $\mu,\nu\in\Pi(\{0,...,N\})$ the relative entropy of ν wrt μ is given by

$$H(\nu|\mu)=\sum_{i=0}^{N}\nu(i)\log\frac{\nu(i)}{\mu(i)}.$$

Exercise 12. Prove that $(\mu_n)_n$ satisfy the LDP with rate function

$$I(x)=x\log(x/p)+(1-x)\log((1-x)/(1-p))=H(\mathrm{Ber}(x)|\mathrm{Ber}(p)).$$

Hints:

a) Prove that $\|f\|_n\sim\max_{0\leqslant k\leqslant n}(|f(k/n)|\binom{k}{n}^{1/n}p^{k/n}(1-p)^{1-k/n})$ using the fact that the cardinality of the summation in the definition of $\|f\|_n$ is of order n.

b) Prove that, uniformly in $0\leqslant k\leqslant n$,

$$\binom{k}{n}^{1/n}\sim\left(\frac{k}{n}\right)^{-k/n}\left(1-\frac{k}{n}\right)^{-(1-k/n)}$$

by observing that the bound

$$\int_{1/k}^{1}\log x\,\mathrm{d}x\leqslant\frac{1}{k}\sum_{m=1}^{k}\log(m/k)\leqslant\int_{0}^{1}\log x\,\mathrm{d}x$$

imply $(k!)^{1/k}\sim(k/e)$ as $k\to+\infty$ and then conclude that $(k!)^{1/n}\sim(k/e)^{k/n}$ uniformly in k by using different arguments for small and large k.

For sequences $(X_n)_{n\geqslant1}$ of iid variables on the finite set $\mathcal{K}=\{1,...,N\}$ with common law $\rho\in\Pi(\mathcal{K})$ we can define the *empirical vector* L_n with values in the compact metrizable space $\Pi(\mathcal{K})=\{p\in[0,1]^N:p_1+\cdots+p_N=1\}$ as

$$L_n(i)=\frac{1}{n}\sum_{k=1}^{n}1_{X_k=i}=\frac{\#\{1\leqslant k\leqslant n:X_n=i\}}{n}$$

and let μ_n to be the law on L_n (thus $\mu_n\in\Pi(\Pi(\mathcal{K}))$).

Theorem 26. (Boltzmann, 1877) *The sequence $(\mu_n)_n$ satisfy the LDP on $\Pi(\mathcal{K})$ with (convex) rate function $I(\nu)=H(\nu|\rho)$.*

The key point of a direct proof of this theorem is that the set of all possible empirical vectors of a sample of size n is of cardinality not larger than $(n+1)^N$ (each of the N components can take at most $n+1$ values). This magnitude disappear in the LD limit since it is sub-exponential in n. Only the asymptotic size of the set of the microscopic configurations compatible with a given empirical vector will contribute to the rate function, as in the coin tossing ($N=2$) case.

Another possible proof of this theorem goes via Cramér theorem on $\mathbb{R}^N$. Replace each X_n by the vector of Bernoulli variables $(Y_n^1,...,Y_n^N):\Omega\to\{0,1\}^N$ where $Y_n^i=1_{X_n=i}$ and observe that $L_n(i)=n^{-1}\sum_{i=1}^{n}Y_n^i$ so that empirical measure becomes an empirical mean. Then Cramérs theorem gives that the rate function on $\Pi(\mathcal{K})$ is given by the Fenchel-Legendre transform $\Lambda:\mathbb{R}^N\to\mathbb{R}$ of the log mgf of the vector Y_1, but

$$\Lambda(\lambda_1,...,\lambda_N)=\log\mathbb{E}\,(e^{\lambda_1Y_1^1+\cdots+\lambda_NY_1^N})=\log\sum_{i=1}^{N}e^{\lambda_i}\,\rho_i$$

so, for every $x_1,...,x_N\in[0,1]$ with $x_1+\cdots+x_n=1$ we have

$$\begin{aligned}I(x_1,...,x_N)&=\sup_{\lambda_1,...,\lambda_N}[\lambda_1x_1+\cdots+\lambda_Nx_N-\log\sum_{i=1}^{N}e^{\lambda_i}\,\rho_i]\\&=H((x_i)_{1\leqslant i\leqslant N}|(\rho_i)_{1\leqslant i\leqslant N})=\sum_{i=1}^{N}\log\frac{x_i}{\rho_i}x_i\,.\end{aligned}$$

Remark 27. Boltzmann discovered this asymptotic probability in 1877 during his attempt to ground thermodynamics of the perfect gas on a microscopic statistical theory of a system of free particles. His proof, using Stirling asymptotic formula for $n!$, was not completely rigorous but nonetheless right. His work showed that the physical entropy of thermodynamics is linked with the mathematical entropy of a probability distribution as a measure of its "unevenness".

Now we are ready to generalize Sanov theorem to compactly supported measures on $\mathbb{R}^d$, not necessarily discrete. Let $\mathcal{K}$ a compact subset of $\mathbb{R}^d$ and $(X_n)_{n\geqslant 1}$ an iid sequence with values in $\mathcal{K}$ with law $\rho = (X_1)_*\,\mathbb{P} \in \Pi(\mathcal{K})$ and $\mu_n \in \Pi(\Pi(\mathcal{K}))$ the law of the empirical measure L_n defined via $L_n(f) = n^{-1}\sum_{i=1}^n f(X_i)$ for all bounded measurable $f\colon \mathcal{K} \to \mathbb{R}$ (alternatively $L_n(\mathrm{d}x) = n^{-1}\sum_i \delta_{X_i}(\mathrm{d}x)$).

Theorem 28. (Sanov) *The sequence $(\mu_n)_n$ obey the LDP on $\Pi(\mathcal{K})$ with rate function $I(\nu) = H(\nu|\rho)$.*

Proof. Let $\{\varphi_k\}_{k\geqslant 1}$ be a countable dense set in $C(\mathcal{K})$ and let $\mathcal{F}_k = \sigma(\varphi_1, ..., \varphi_k)$ the associated filtration of the Borel σ-algebra $\mathcal{B}(\mathcal{K})$. Let $F_k\colon \Pi(\mathcal{K}) \to [-1,1]^k$ be the continuous function $F_k(\nu) = (\nu(\varphi_1)/\|\varphi_1\|, ..., \nu(\varphi_k)/\|\varphi_k\|)$. By Cramer's theorem the image law $(\mu_n^k = (F_k)_*\,\mu_n)$ satisfy the LDP on $[-1,1]^k$ with rate function $I_k(x_1, ..., x_k)$ given by

$$I_k(x_1, ..., x_k) = \sup_{\lambda_1, ..., \lambda_k} \left[\lambda_1 x_1 + \cdots + \lambda_k x_k - \log \mathbb{E}[e^{\sum_i \lambda_i \varphi_k(X_1)/\|\varphi_k\|}]\right].$$

Then by the Dawson-Gartnër theorem

$$\begin{aligned} I(\nu) = \sup_k I_k(F_k(\nu)) &= \sup_k \sup_{\lambda_1, ..., \lambda_k} \left[\nu(\lambda_1\varphi_1 + \cdots + \lambda_k\varphi_k) - \log \mathbb{E}[e^{\sum_i \lambda_i \varphi_k(X_1)}]\right] \\ &= \sup_{\varphi} \left[\nu(\varphi) - \log \mathbb{E}[e^{\varphi(X_1)}]\right] = H(\nu|\rho)\,. \end{aligned}$$

□

Sanov's theorem can be used to prove Cramérs theorem. Indeed the empirical mean $\hat{S}_n$ of a vector $(X_1, ..., X_n)$ is a function of the empirical distribution L_n: $\hat{S}_n = \int x\, L_n(\mathrm{d}x) = m(L_n)$. If the variables take values on a compact subset of $\mathbb{R}^d$ then $m\colon \Pi(\mathcal{K}) \to \mathbb{R}$ is a continuous function and its image its compact. By the contraction principle the laws σ_n of $\hat{S}_n$ obey the LDP with rate function $J(x) = \inf\{H(\nu|\rho) : \nu \in \Pi(\mathcal{K}),\, \mu(\nu) = x\}$. Introduce the probability measures $\rho_\lambda = e^{\lambda x - \log \Lambda_\rho(\lambda)}\rho$ and observe that for any $y \in \operatorname{supp}\rho$ we can find λ such that $y = m(\rho_\lambda)$ and that

$$H(\nu|\rho_\lambda) = \int \log \frac{\mathrm{d}\nu}{\mathrm{d}\rho_\lambda} \mathrm{d}\nu = H(\nu|\rho) - \lambda m(\nu) + \log \Lambda_\rho(\lambda)$$

so $J(y) = \inf\{H(\nu|\rho_\lambda) : \nu \in \Pi(\mathcal{K}), \mu(\nu) = y\} + \lambda y - \log \Lambda_\rho(\lambda) = \lambda y - \log \Lambda_\rho(\lambda) = \sup_\lambda [\lambda y - \log \Lambda_\rho(\lambda)]$ (think why). Conclusion : J is the Fenchel-Legendre transform of Λ_ρ.

3.4 Gibbsian conditioning

Sanov's theorem allow us to discuss another physical phenomenon related to "Gibbsian" distributions. Let $(X_n)_{n\geqslant 1}$ be an iid sequence with values in $\mathcal{K}$ and law $\rho \in \Pi(\mathcal{K})$. Fix some integer $k \geqslant 1$ and consider the law $\mu_n \in \Pi(\mathcal{K}^k)$ of $(X_1, ..., X_k)$ conditional of an event involving $L_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$, the empirical measure of the vector $(X_1, ..., X_n)$:

$$\mu_n(f) = \int_{\mathcal{K}^k} f(x)\, \mu_n(\mathrm{d}x) = \mathbb{E}[f(X_1, ..., X_k)|L_n \in B]$$

where $A \in \mathcal{B}(\mathcal{K}^k)$ and $B \in \mathcal{B}(\Pi(\mathcal{K}))$. We will work with $k = 1$ generalization to higher k being easy.

Exercise 13. Let $k = 1$, show that

$$\mu_n(f) = \mathbb{E}[f(X_1)|L_n \in B] = \mathbb{E}[L_n(f)|L_n \in B] = \int_{\Pi(\mathcal{K})} \sigma(f)\, \nu_n(\mathrm{d}\sigma)$$

where ν_n is the law of L_n conditional to the event B.

Assume that B is closed and that $\inf_{B^\circ} H_\rho = \min_B H_\rho = H_\rho(\hat{\nu})$ for a unique minimum point $\hat{\nu} \in B$. Then the LDP for $(L_n)_{n\geqslant 1}$ and the restriction theorem imply that the sequence $(\nu_n)_{n\geqslant 1}$ obey the LDP on $B \subseteq \Pi(\mathcal{K})$ with rate function $J(\nu) = H_\rho(\nu) - H_\rho(\hat{\nu})$ and that $\nu_n \to \delta_{\hat{\nu}}$ as $n \to \infty$. Then, for every continuous $f \in C(\mathcal{K})$, $\mu_n(f) = \int_{\Pi(\mathcal{K})} \sigma(f)\, \nu_n(\mathrm{d}\sigma) \to \hat{\nu}(f)$ and we have proved that the sequence $(\mu_n)_{n\geqslant 1}$ converge weakly to the probability measure $\hat{\nu}$ which is the solution of the minimization of H_ρ over B.

Interesting case is where the conditioning set B is of the form $B=\{\nu\in\Pi(\mathcal{K}):\nu(\varphi)\in[e,e+\delta]\}$ for some small $\delta>0$ and $e\in\mathbb{R}$ is such that $\mathbb{E}[\varphi(X_1)]<e<\sup_{\mathcal{K}}\varphi$. This last condition is to render the event $\{L_n\in B\}=\{L_n(\varphi)\in[e,e+\delta]\}$ atypical : by the LLN we have $L_n(\varphi)\to\mathbb{E}[\varphi(X_1)]$ a.s. In this case $\hat\nu$ can be described explicitly as an exponential perturbation of ρ. Let $\lambda\in\mathbb{R}$ and introduce the "tilted" measures $\rho_\lambda=e^{\lambda f}\rho/Z(\lambda)$ with $Z(\lambda)=\rho(e^{\lambda f})$ and observe that $H_\rho(\nu)=H_{\rho_\lambda}(\nu)+\lambda\nu(f)-\log Z(\lambda)$.

Exercise 14. Show that there exists $\lambda>0$ for which $\rho_\lambda(f)=e$ and that $H_\rho(\rho_\lambda)=\lambda e-\log Z(\lambda)$.

Now

$$H_\rho(\rho_\lambda)=\lambda e+\log Z(\lambda)=\min_{\nu:\nu(f)\in[e,e+\delta]}[H_{\rho_\lambda}(\nu)+\lambda\nu(f)]+\log Z(\lambda)=\min_B H_\rho$$

so $\hat\nu=\rho_\lambda$.

Let us extend the result to $k>1$. It is enough to consider new "block" variables $\tilde X_i=(X_{1+k(i-1)},\ldots,X_{k+k(i-1)})$ and observe if $\tilde L_n\in\Pi(\mathcal{K}^k)$ is the empirical law of $(\tilde X_1,\ldots,\tilde X_n)$ then $L_{nk}(f)=\tilde L_n(\tilde f)$ where $\tilde f(x_1,\ldots,x_k)=(f(x_1)+\cdots+f(x_k))/k$.

Exercise 15. Let $\nu\in\Pi(\mathcal{K}^2)$, show that $H_{\rho\otimes\rho}(\nu)\geqslant H_{\rho\otimes\rho}(\nu_1\otimes\nu_2)=H_\rho(\nu_1)+H_\rho(\nu_2)$ where ν_1,ν_2 are the marginals of ν. Hint: in the variational formula for $H_{\rho\otimes\rho}(\nu)$ take test functions φ of the form $\varphi_1\otimes\varphi_2$.

Exercise 16. For B of the form $B=\{L_n(f)\in\mathcal{I}\}$ derive the LDP for $(\tilde L_i)_{i\geqslant1}$ on $\Pi(\mathcal{K}^k)$ and obtain that the law of Y_1 conditional on B is given by the minimum $\tilde\nu\in\Pi(\mathcal{K}^k)$ of the functional $H_{\tilde\rho}$ over $\{\nu\in\Pi(\mathcal{K}^k):\nu(\tilde f)\in\mathcal{I}\}$ where $\tilde\rho=\rho\otimes\cdots\otimes\rho$ is the k-fold product measure with marginals ρ. Conclude that if $\hat\nu$ is the unique minimum of H_ρ over $\{\nu\in\Pi(\mathcal{K}):\nu(f)\in\mathcal{I}\}$ then $\hat\nu\otimes\cdots\otimes\hat\nu$ is the unique minimum of the variational problem on $\Pi(\mathcal{K}^k)$ and observe that this imply the independence of $(X_1,\ldots,X_k)$ in the limit law.

The physical interpretation of this phenomenon goes as follows: consider an assembly of n independent particles each of them characterized by some quantity X_i, $i=1,\ldots,n$ taking values in $\mathcal{K}$ (e.g. energy, momentum, position, etc...) and assume that the allowed configurations of the whole system are those compatible with a given mean value of some function $f:\mathcal{K}\to\mathbb{R}:\sum_i f(X_i)/n\simeq e$ (e.g. energy per particle, density, etc..). This constraint is macroscopic in the sense that involves only an average over all the particles. Then in the limit of a infinite system ($n\to\infty$, in reality $n\simeq10^{23}$) the configurations of a very small subsystem of size k (in our model k is fixed as $n\to\infty$) are described by iid configurations, each particle distributes as ρ_λ, the Gibbs distribution compatible with the macroscopic constraint.

4 Large deviations for processes

4.1 Mogulskii theorem

Let $(X_n)_{n\geqslant1}$ be an iid sequence of Bernoulli(p) r.v. For every n the vectors $X_{\leqslant n}=(X_1,\ldots,X_n)$ are random elements in $\{0,1\}^n$ which we will embed in $L^\infty([0,1])$ as follows : for each n let

$$F_n(x_1,\ldots,x_n)(\theta)=\sum_{i=1}^n x_i\,1_{\theta\in[(i-1)/,i/n)}$$

so that $F_n(X_{\leqslant n})$ is a random element in $\mathcal{K}=\{f\in L^\infty([0,1]):\|f\|_{L^1}\leqslant1\}$ and we denote by μ_n its law. On $\mathcal{K}$ we consider the weak-$*$ topology, i.e. the smallest topology which renders all the linear maps $f\mapsto g(f)=\int_0^1 f(\theta)\,g(\theta)\mathrm{d}\theta$ continuous for every $g\in L^1([0,1])$. With this topology $\mathcal{K}$ is compact and metrizable. A possible metric is obtained by taking a countable dense subset $\{\varphi_k\}_{k\geqslant1}$ of the unit ball of L^1 and letting

$$d(f,g)=\sum_{k\geqslant1}\frac{|\varphi_k(f)-\varphi_k(g)|}{2^k}$$

Another possible metric is given by $d(f,g)=\sup_{0\leqslant t\leqslant1}|\int_0^t(f(\theta)-g(\theta))\mathrm{d}\theta|$. Let $J_p(x)=H(\mathrm{Ber}(x)|\mathrm{Ber}(p))$. Then we have the following result

Theorem 29. (Mogulskii) *The sequence $(\mu_n)_n$ obey the LDP on $\mathcal{K}$ with rate function*

$$I(f) = \int_0^1 J_p(f(\theta))\mathrm{d}\theta.$$

Proof. We only need to uniquely identify the rate function I' of possible accumulation points. For each k define $Q_{k,l} = ((l-1)/k, l/k]$ and $G_k \colon \mathcal{K} \to [-1,1]^k$ as $G_k(f) = (f_{k,1}, ..., f_{k,k})$ where $f_{k,l} = \int_{Q_{k,l}} f/|Q_{k,l}|$ is the mean of f over $Q_{k,l}$ so that $\pi_k(f) = F_k(G_k(f)) \to f$ in $\mathcal{K}$ (why?). By Cramérs theorem the laws μ_n^k of $G_k(F_{kn}(X_{\leqslant kn}))$ on $[-1,1]^k$ satisfy the LDP with speed n/k and rate function $I_k(x_1, ..., x_k) = \sum_{i=1}^k J_p(x_i)$ taking into account the change of speed we have that, for every $g \in \mathcal{K}$ and for every k,

$$\min\{I'(f) : G_k(f) = G_k(g)\} = \frac{1}{k}\sum_{i=1}^k J_p(G_k(g)_i) = \int_0^1 J_p(\pi_k(g)(\theta))\mathrm{d}\theta = I_k(g)$$

Now using Fatou lemma it is easy to compute the Γ-limit of the functional $\int_0^1 J_p(\pi_k(g)(\theta))\mathrm{d}\theta$ as

$$\Gamma\lim_k \int_0^1 J_p(\pi_k(g)(\theta))\mathrm{d}\theta = \int_0^1 J_p(g(\theta))\mathrm{d}\theta = I(g)$$

(exercise) while another easy argument gives $\Gamma\lim_k \min\{I'(f) : G_k(f) = G_k(g)\} = I'(g)$ since I' is lsc. Then we can conclude that $I(g) = I'(g)$. □

4.2 Large deviations for random walks

Let $(X_n)_{n\geqslant 1}$ be a sequence of iid Bernoulli(p) random variables. Consider the process $S_n = X_1 + \cdots + X_n$ with $S_0 = 0$. Define a continuous random function φ_n on $[0,1]$ by

$$\varphi_n(t) = \frac{S_k}{n} + \frac{(S_{k+1} - S_k)}{n}(t-k) \qquad \text{for } k \leqslant t < k+1.$$

Let $\mathcal{K}$ be the subset of $C([0,1])$ such that $f \in \mathcal{K}$ if and only if $f(0) = 0$ and $|f(t) - f(s)| \leqslant |t-s|$ for all $0 \leqslant s \leqslant t \leqslant 1$. Observe that $\varphi_n \in \mathcal{K}$ and that using the notations of the section "Large deviations for processes" of the 4th lecture notes of the course we have

$$\varphi_n(t) = \int_0^t F_n(X_{\leqslant n})(s)\,\mathrm{d}s$$

where for every n the vectors $X_{\leqslant n} = (X_1, ..., X_n)$ are random elements in $\{0,1\}^n$ and

$$F_n(x_1, ..., x_n)(\theta) = \sum_{i=1}^n x_i\, 1_{\theta \in [(i-1)/, i/n)} \in L^\infty([0,1]).$$

Observe also that $\varphi_n(t)$ is a piecewise linear function for which

$$\varphi_n(k/n) = S_k/n.$$

Recall that J_p has been defined in Poly 4 (section "Large deviations for processes") as

$$J_p(x) = H(\mathrm{Ber}(x)|\mathrm{Ber}(p)) = x\log\frac{x}{p} + (1-x)\log\frac{1-x}{1-p}$$

the relative entropy of a Bernoulli law of parameter x with respect to the Bernoulli law of parameter p.

a) Prove that the set $\mathcal{K}$ with the norm of uniform convergence is compact.

b) Prove that the laws μ_n of φ_n on $\mathcal{K}$ satisfy a large deviation principle with rate function

$$I(f) = \int_0^1 J_p(f'(s))\mathrm{d}s$$

where $f'(s)$ is the derivative of $f \in \mathcal{K}$ (which exists almost everywhere since f is Lipshitz). Hint: use Theorem 8 (Mogulskii theorem) of Poly 4 and the contraction principle (Thm 19 of Poly 3).

c) With $p = 1/2$, use the result of point b) to prove that, if we set

$$p_{n,\varepsilon} = \mathbb{P}\left(\left| \frac{S_k}{n} - \frac{1}{2}\left(\frac{k}{n}\right)^2 \right| \leqslant \varepsilon \quad \text{for all } k = 1, ..., n \right)$$

then

$$\limsup_{n \to +\infty} \left| (p_{n,\varepsilon})^{1/n} - \frac{\sqrt{e}}{2} \right| \to 0 \qquad \text{as } \varepsilon \to 0+ .$$

Hint: use Proposition 10 of Poly 3.

d) With $p = 1/2$, use the result of point b) to prove that, if we set

$$p_n = \mathbb{P}\left(\frac{S_k}{n} \geqslant c\left(\frac{k}{n}\right)^2 \quad \text{for all } k = 1, ..., n \right)$$

then

$$(p_n)^{1/n} \to \begin{cases} 1 & \text{for } c \in [0, 1/2] \\ \frac{1}{2c^c(1-c)^{1-c}} & \text{for } c \in [1/2, 1] \end{cases}$$

Hint: use Corollary 9 of Poly 3 to convert the limit computation in a variational problem over the rate function, guess the shape of the extremal function and then use the fact that the function $J_{1/2}$ is convex and Jensen's inequality gives

$$\int_0^1 J_{1/2}(\varphi(s))\mathrm{d}s \geqslant J_{1/2}\Big(\int_0^1 \varphi(s)\mathrm{d}s\Big).$$

To understand fully this problem it is a good idea to give a look at the paper

`http://www.math.tau.ac.il/~tsirel/Courses/LargeDev/lect5.pdf`

in particular Section 5a.

4.3 Large deviations for Brownian motion

We would like to establish large deviation results for Brownian motion. The first difficulty is that the space of continuous trajectories is not compact so the theory developed during the course cannot be directly applied.

a) Look at the paper

`http://www.math.tau.ac.il/~tsirel/Courses/LargeDev/lect9.pdf`

in particular at sections 9a,9c,9d and extend the large deviation principle to non-compact spaces $\mathcal{K}$ introducing the property of exponential tightness.

b) Look at the paper

`http://www-math.cudenver.edu/~puhalski/ld/course_1.pdf`

and in particular at Theorems 2.5, 2.6 and 2.7 to help you prove the following statement. Consider the space $\mathcal{X} = C([0,1];\mathbb{R})$ and let μ_n be the law of the process $(B_t/n)_{t\in[0,1]}$ on $\mathcal{X}$ where B_t is a standard Brownian motion of unit variance i.e. $\mathrm{Var}(B_t) = t$. Then μ_n describe a rescaled Brownian motion on $[0,1]$ with variance $1/n^2$. Prove that the family $\{\mu_n\}_{n\geqslant 1}$ is exponentially tight and that it satisfy a large deviation principle on $\mathcal{X}$ with rate function

$$I(f) = \int_0^1 (f'(s))^2 \mathrm{d}s$$

if $f' \in L^2([0,1])$ and $I(f) = +\infty$ otherwise.

5 Other applications

5.1 Curie-Weiss model

(Reference : Chapter 9.1 of the paper
`http://www.math.umass.edu/~rsellis/pdf-files/Les-Houches-lectures.pdf`)

This is a model of interacting components where it is possible to use large deviation theory to understand the collective behavior of the system. Fix $\beta \geqslant 0$ and let $\Omega_n = \{-1,1\}^n$ and consider the probability measure P_n over Ω_n given by

$$P_n(\omega_1,...,\omega_n) = \frac{1}{2^n}\frac{e^{\beta H_n(\omega)}}{Z_n}, \qquad \omega = (\omega_1,...,\omega_n) \in \Omega_n$$

where

$$H_n(\omega) = \frac{1}{n}\sum_{i,j=1}^{n} \omega_i \omega_j$$

and where Z_n is a normalization constant such that

$$P_n(\Omega_n) = \sum_{\omega\in\Omega_n} P_n(\omega_1,...,\omega_n) = 1$$

that is

$$Z_n = \sum_{\omega\in\Omega_n} \frac{\exp\left(\beta H_n(\omega)\right)}{2^n}.$$

Observe that $H_n(\omega) = n\beta F(M_n(\omega))$ where $F(x) = x^2$ and $M_n(\omega) = (\omega_1 + \cdots + \omega_n)/n$ is the empirical mean of the random variables ω_1, ..., $\omega_n = \pm 1$. Let μ_n be the law of M_n under the probability measure P_n. Observe that $M_n : \Omega_n \to [-1,1]$ so that μ_n is a probability over $[-1,1]$ for all $n \geqslant 1$ and that the law μ_n is defined by

$$\int_{[-1,1]} \varphi(x)\mu_n(\mathrm{d}x) = \sum_{\omega\in\Omega_n} \varphi(M_n(\omega))P_n(\omega) = \sum_{\omega\in\Omega_n} \varphi((\omega_1+\cdots+\omega_n)/n)\frac{\exp\left(\beta H_n(\omega)\right)}{2^n\, Z_n}$$

The family $\{\mu_n\}_{n\geqslant 1}$ satisfy a large deviation principle on $[-1,1]$ with rate function

$$I(x) = \frac{1+x}{2}\log(1+x) + \frac{1-x}{2}\log(1-x) + \beta x^2 - C$$

where the constant C is chosen such that $\inf_{x\in[-1,1]} I(x) = 0$.

a) Prove it. Hint: Observe that the measure μ_n can be written as

$$\int_{[-1,1]} \varphi(x)\mu_n(\mathrm{d}x) = \int_{[-1,1]} \varphi(x)\frac{e^{\beta n F(x)}}{Z_n}\mu_n^0(\mathrm{d}x)$$

where μ_n^0 is the law of M_n under P_n when $\beta = 0$, that is when each $\omega_1,...,\omega_n$ is an independent random variable such that $P_n(\omega_n = \pm 1) = 1/2$. So that by Cramers theorem $\{\mu_n^0\}_{n\geqslant 1}$ satisfy a large deviation principle with suitable rate function. Use then Theorem 21 (Change of measure) of Poly 3.

b) Prove that for $\beta \in]0,1]$ the unique minimizer of the function $I : [-1,1] \to \mathbb{R}_+$ is 0 and that if $\beta > 1$ then there exists a function $m(\beta)$ such that the function I has exactly two minima at the points $\pm m(\beta)$.

c) Use the above results to prove that, when $\beta \leqslant 1$ the family μ_n weakly converges to the Dirac mass in 0 and that when $\beta > 1$ it weakly converge to the probability measure

$$\frac{1}{2}\delta_{-m(\beta)} + \frac{1}{2}\delta_{m(\beta)}$$

that is the discrete measure that assign probability 1/2 to the points $\pm m(\beta)$. Hint: when $\beta \leqslant 1$ consider the probability $\mu_n(]-\varepsilon,\varepsilon[^c)$ for some $\varepsilon > 0$ and use the large deviation estimate to show that it goes to zero as $n \to \infty$. Use a similar strategy in the case $\beta > 1$.

To understand fully this problem it is a good idea to give a look at the paper

`http://www.math.umass.edu/~rsellis/pdf-files/Les-Houches-lectures.pdf`

in particular Chapter 9.1.

5.2 Further Gibbsian conditioning

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space on which we consider a sequence of iid discrete random variables $(X_n)_{n\geqslant 1}$ with law $\mathbb{P}(X_n = \pm 1) = 1/2$. Let $M_n = (X_1 + \cdots + X_n)/n$ the empirical mean of the first n variables. Fix $\varepsilon > 0$ and $m \in]-1, 1[$ and let B_n be the set $B_n = \{\omega \in \Omega : M_n(\omega) \in [m, m+\varepsilon]\}$. We want to study the limit law as $n \to \infty$ of the k-ple $(X_1, ..., X_k)$ when $(X_1, ..., X_n)$ is conditioned on the event B_n. More precisely fix $k \geqslant 1$ and let μ_n be the law of $(X_1, ..., X_k)$ conditional on B_n:

$$\mu_n(x_1, ..., x_k) = \mathbb{P}(X_1 = x_1, ..., X_k = x_k | B_n) = \frac{\mathbb{P}(X_1 = x_1, ..., X_k = x_k, B_n)}{\mathbb{P}(B_n)}.$$

Our aim is to prove that the family $\{\mu_n\}_{n\geqslant 1}$ converge weakly to the law ρ_λ^k on $\{-1, 1\}^k$ for which all the components are independent and

$$\rho_\lambda^k(x_1, ..., x_k) = \rho_\lambda(x_1) \cdots \rho_\lambda(x_k)$$

where ρ is the discrete probability on $\{-1, 1\}$ given by

$$\rho_\lambda(x) = \frac{e^{\lambda x}}{e^\lambda + e^{-\lambda}} \qquad \text{for } x = -1, 1$$

where $\lambda \in \mathbb{R}$ is a fixed number which is determined by the fact that the mean of the measure ρ_λ should be m:

$$m = \sum_{x=-1,1} x \rho_\lambda(x) = \frac{e^\lambda - e^{-\lambda}}{e^\lambda + e^{-\lambda}} = \tanh(\lambda).$$

To prove this weak convergence result you need to understand the discussion on "Gibbsian conditioning" in the "Poly 4" of the lecture notes and proceed as follows:

a) Start by proving the statement for $k = 1$. Note that for any continuous function

$$\sum_{x_1=\pm 1} f(x_1)\, \mu_n(x_1) = \frac{\mathbb{E}[f(X_1) 1_{B_n}]}{\mathbb{E}[1_{B_n}]} = \frac{\mathbb{E}[L_n(f) 1_{B_n}]}{\mathbb{E}[1_{B_n}]}$$

where $L_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_i)$ is the mean of f with respect to the empirical measure of the random vector $(X_i)_{1\leqslant i\leqslant n}$. Observe also that $B_n = \{L_n(h) \in [m, m+\varepsilon]\}$ where $h: \{-1, +1\} \to \mathbb{R}$ is the identity function given by $h(x) = x$. Then

$$\sum_{x_1=\pm 1} f(x_1)\, \mu_n(x_1) = \frac{\mathbb{E}[L_n(f) 1_{L_n(h)\in[m,m+\varepsilon]}]}{\mathbb{E}[1_{L_n(h)\in[m,m+\varepsilon]}]}$$

Use Sanov theorem (Theorem 7 of Poly 4) and Proposition 1 and Corollary 2 of Poly 4 to deduce a large deviation principle for μ_n. Conclude that $\mu_n \to \rho_\lambda$.

b) Follow the discussion on "Gibbsian conditioning" in the "Poly 4" of the lecture notes to extend the argument to $k > 1$. For the purpose of the exam it is enough to prove the statement for $k = 2$: that is we want to prove that conditionally on B_n, the pair (X_1, X_2) converge weakly to a pair of independent variables each of them with law ρ_λ. (with λ depending on m).