NOTES ON LARGE DEVIATIONS

NIKOS ZYGOURAS

1. Introduction

Large deviations should be thought as a principle of the theory of probability. We are already familiar with the two basic ones, the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT).

The clasical LLN says that if $(X_i)_{i\geq 1}$ is a sequence of i.i.d. variables, on a probability space $(\Omega, \mathcal{F}, P)$, with $E[X_1] = \mu$, $E[|X_1|] < \infty$, then

$$\frac{X_1 + \cdots + X_n}{n} \to \mu, \quad P - a.s.$$

The clasical CLT says that, under the additional assumption that the variance exists, say equals 1, then

$$P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}} > x\right) \cong \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy.$$

That is it tells us what is the asymptotic probability that the sum $S_n = X_1 + \cdots + X_n$ has a typical deviation from its mean $n\mu$ of order $\sqrt{n}$.

Large deviations study the asymptotic probability of a (large) deviation of S_n from the mean $n\mu$ of order n. This can be summarised by

$$P\left(X_1 + \cdots + X_n - n\mu \cong n\,x\right) \cong e^{-nI(x)}.$$

The function $I(\cdot)$ that appears is called the *rate function* and governs the asymptotics of such probabilities.

Example Let $(X_i)_{i\geq 1}$, such that $P(X_i = 1) = P(X_i = 0) = 1/2$. Then for all $a > 1/2$,

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(S_n \geq na\right) = -I(a),$$

with

$$I(z) = \begin{cases} \log 2 + z\log z + (1-z)\log(1-z) & , \quad z \in (0,1), \\ \infty & , \quad z \notin (0,1) \end{cases}$$

Proof. It is easy to verify the case that $a > 1$, so we will restrict to the case $1/2 < a < 1$. The proof goes by a direct combinatorial computation.

$$P(S_n \geq na) = \sum_{k\geq na} \binom{n}{k} \frac{1}{2^n}.$$

Use know the easy and useful fact that, for every sequence of positive numbers $(a_n)_{n\geq 1}$,

elem

$$\lim_n \frac{1}{n} \log \sum_{1\leq i\leq n} a_n = \lim_n \frac{1}{n} \log \max_{1\leq i\leq n} a_n, \tag{1.1}$$

to get that

$$\lim_n \frac{1}{n} \log P(S_n \geq na) = \lim_n \frac{1}{n} \log \binom{n}{na} \frac{1}{2^n}$$

Use now Stirling's formula $n! \sim \sqrt{2\pi n}\, n^n e^{-n}$ to get that $\binom{n}{na} \sim \frac{1}{\sqrt{2\pi\, n}} a^{-na}(1-a)^{-n(1-a)}$. The result now follows. □

Knowing the large deviations is also very important when we want to evaluate asymptotically exponential integrals. Consider for example the case of a sequence of measures $\mu_n(dx)$. You can think of the case $P\left(\frac{S_n}{n} \in dx\right)$. Supose we want to evaluate the integrals

$$\frac{1}{n} \log \int e^{n\theta x} \mu_n(dx).$$

Think again of $\frac{1}{n} \log E[e^{\theta S_n}]$. If $\mu_n(dx) \cong e^{-nI(x)}$, then by the Laplace asymptotics, we have that the integral is asymptotically equivallent to

$$\frac{1}{n} \log \int e^{n\theta x - nI(x)} dx \cong \sup_x \left(\theta x - I(x)\right).$$

Let us give a first definition of Large Deviation Principle (LDP), but don't worry about the details now.

Definition 1.1. *Let X be a Polish space (complete, separable, metric space). We say that the sequence of measures μ_n satisfies a LDP with rate function I, if for every Borel set Γ*

LDP

$$-\inf_{x \in \Gamma^o} I(x) \leq \liminf_n \frac{1}{n} \log \mu_n(\Gamma) \leq \limsup_n \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x) \tag{1.2}$$

We want the rate function $I : X \to [0, \infty)$ to be lower semicontinuous. Often I is called good rate function if the level sets $\{x : I(x) \leq L\}$ are compact.

REMARK: 1. For the moment think of $\mu_n(\Gamma)$ as $P\left(\frac{S_n}{n} \in \Gamma\right)$.

2. If $\inf_{x \in \Gamma^o} I(x) = \inf_{x \in \bar{\Gamma}} I(x)$, then $\lim_n \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \Gamma} I(x)$

3.We need the distinction between $\Gamma^o, \Gamma, \bar{\Gamma}$ to deal with pathological cases e.g. when μ_n are non-atomic, i.e. $\mu_n(\{x\}) = 0$. In this case (1.2) cannot hold for $\Gamma = \{x\}$, without considering Γ^o. This formulation takes also into account the possibility of concentration of measure on the boundaries of Γ.

4. Notice that the reason of the presence of the inf in the formulation lies is relation (1.1). a way to see it formally is that $\lim \frac{1}{n} \log \mu_n(\Gamma)$" = " $\lim \frac{1}{n} \log \sum_{x \in \Gamma} \mu_n(x)$ " = " $\lim \frac{1}{n} \log \max_{x \in \Gamma} \mu_n(x)$

2. TWO EXAMPLES

We will know show two examples which are very important to understand the set of ideas used to prove LDP's. The first one is Crámer's Theorem.

2.1. **CRAMER'S THEOREM.**

Theorem 2.1. *Let* $(X_i)_{i\geq 1}$ *a sequence of i.i.d variables. Assume that* $\phi(t) := E[e^{tX_1}]$ *is finite for every* $t \in \mathbb{R}$. *Then the measures* $\mu_n(\cdot) := P\left(\frac{S_n}{n} \in \cdot\right)$ *satisfy a LDP with rate function* $I(x) = \sup_{t\in\mathbb{R}}\{xt - \log\phi(t)\}$.

Before the proof let us remark on the rate function.

Definition 2.2. *For any real function* ϕ, *we define the Legendre transform* $\phi^*(x) := \sup_{t\in\mathbb{R}}\{xt - \log\phi(t)\}$. *This definiton also generalise to many dimensions (even infinite), by* $\phi^*(x) := \sup_{t\in\mathbb{R}}\{< x, t > - \log\phi(t)\}$.

Lemma 2.3. *Let* $\phi(t) := E[e^{tX_1}]$ *and* $\phi^*(t)$ *its Legendre transform. Then*

1. $t \to \log\phi(t)$ *is convex.*
2. ϕ^* *is nonnegative, convex and lower semicontinuous.*
3. *If* $\mu = E[X_1]$, *then* $\phi^*(\mu) = 0$.
4. *For* $x \geq \mu$, $x \to \phi^*(x)$ *is nondecreasing, and for* $x \leq \mu$ *it is nonincreasing.*
5. *For* $x \geq \mu$, $\phi^*(x) = \sup_{t\geq 0}\{xt - \log\phi(t)\}$, *and for* $x \leq \mu$, $\phi^*(x) = \sup_{t\leq 0}\{xt - \log\phi(t)\}$.

Proof. 1. and 2. are trivial (for 1. just use Jensen's inequality).

3. By Jensen's inequality we have that $\log E[e^{tX_1}] \geq tE[X_1] = t\mu$, for every t. So $\phi^*(\mu) = \sup_t\{\mu t - \log\phi(t)\} \leq 0$, which implies the result by the nonnegativity of ϕ^*.

4. It follows from the convexity of ϕ^* and the fact that $\phi^*(\mu) = 0$.

5. Notice that the slope of $\log\phi(t)$ at the origin is equal to $E[X_1]$. The result now follows by the convexity of $\log\phi(t)$.

□

We are now ready to prove Crámer's Theorem.

Proof. Without loss of generality we will assume that $E[X_1] = 0$.

THE UPPER BOUND. The upper bound is an optimization over a family of exponential Chebyshev's inequalities. Let us first bound the quantity, let $t \geq 0$

$$\begin{aligned} P\left(\frac{S_n}{n} \geq x\right) &\leq e^{-n\,tx} E\left[e^{tS_n}\right] \\ &= e^{-n\,tx}\left(E[e^{t\,X_1}]\right)^n = \exp\{-n(tx - \log\phi(t))\}. \end{aligned}$$

Since this bound is true for all $t \geq 0$ we have that

$$\begin{aligned} P\left(\frac{S_n}{n} \geq x\right) &\leq \exp\{-n\sup_{t\geq 0}(tx - \log\phi(t))\} \\ &= \exp\{-n\phi^*(x)\}. \end{aligned}$$

where the last equality is due to part 5. of the previous lemma. To conclude the upper bound notice that $\phi^*(x) = \inf_{y\geq x} I(y)$, by part 4. of the previous lemma.

To pass from the event $\{\frac{S_n}{n} \geq x\}$ to a general event $\{\frac{S_n}{n} \in \Gamma\}$ we do

$$\begin{aligned} P\left(\frac{S_n}{n} \in \Gamma\right) &\leq P\left(\frac{S_n}{n} \in \bar{\Gamma}\right) \\ &\leq P\left(\frac{S_n}{n} \in \bar{\Gamma} \cap [0,\infty)\right) + P\left(\frac{S_n}{n} \in \bar{\Gamma} \cap (-\infty, 0]\right) \\ &\leq e^{-n\phi^*(x_+)} + e^{-n\phi^*(x_-)}, \end{aligned}$$

where $x_+ = \inf\{x \in \bar{\Gamma} \cap [0,\infty)\}$ and analogously for x_-. We then conclude, using (1.1) that $\limsup_n \frac{1}{n} \log P\left(\frac{S_n}{n} \in \Gamma\right) \leq -\phi^*(x_+) \wedge \phi^*(x_-)$, and by part 4. of the previous lemma it is equal to $-\inf_{x\in\bar{\Gamma}} \phi^*(x)$.

THE LOWER BOUND. It is clear that it is enough to prove that for every x and $\epsilon > 0$, we have $\liminf_n \frac{1}{n} \log P\left(\frac{S_n}{n} \in (x-\epsilon, x+\epsilon)\right) \geq \phi^*(x)$.

The idea (which is very important) to prove this is the following: Remember that $E[X_1] = 0$, so by the LLN $S_n/n \to 0$, $P\,a.s.$. This is the *typical* event, i.e. $P\left(\frac{S_n}{n} \in (-\epsilon, \epsilon)\right) \to 1$. We want though to compute the probability of the *atypical* event $P\left(\frac{S_n}{n} \in (x-\epsilon, x+\epsilon)\right)$. To do this we will introduce a new measure $\hat{P}$, such that $\hat{E}[X_1] = x$ and thus $\hat{P}\left(\frac{S_n}{n} \in (-\epsilon, \epsilon)\right) \to 1$. In other words the atypical event becomes under the new measure typical. The asymptotic probability that we are after will be captured by this change of measure.

Let's fix the ideas. Let $F(dy)$ the distribution function related to P. Define the new measure $\hat{P}$ by

$$\hat{F}(dy) = \frac{e^{\tau y}}{\phi(\tau)} F(dy).$$

The value of τ will be chosen as the unique value (if it exists), for which $\hat{E}[X_1] = \int y\hat{F}(dy) = x$. Notice that

$$\int y\hat{F}(dy) = (\log\phi(\tau))'.$$

So we are looking for a value of τ for which $(\log\phi(\tau))' = x$. Such a τ exists if the $\sup_t (tx - \log\phi(t))$ is achieved. Suppose that this is the case (we will deal with the case that this is not the case separately), then $\phi^*(x) = \tau x - \log\phi(\tau)$. Now compute, let $\delta < \epsilon$ and also sup'pose that $\tau \geq 0$ (the case $\tau \leq 0$ is handled similarly).

$$\begin{aligned}
P\left(\frac{S_n}{n} \in (x-\delta, x+\delta)\right) &= \int_{\{\frac{S_n}{n} \in (x-\delta,x+\delta)\}} F(dy_1)\cdots F(dy_n) \\
&= \int_{\{\frac{S_n}{n} \in (x-\delta,x+\delta)\}} e^{-\tau(y_1+\cdots+y_n)}\,\phi(\tau)^n\,\hat{F}(dy_1)\cdots\hat{F}(dy_n) \\
&\geq e^{-n\tau(x+\delta)}\phi(\tau)^n \int_{\{\frac{S_n}{n} \in (x-\delta,x+\delta)\}} \hat{F}(dy_1)\cdots\hat{F}(dy_n) \\
&= e^{-n(x\tau+\delta\tau-\log\phi(\tau))}\,\hat{P}\left(\frac{S_n}{n} \in (x-\delta, x+\delta)\right).
\end{aligned}$$

Finally we have (we will use the fact that $\hat{P}\left(\frac{S_n}{n} \in (x-\delta, x+\delta)\right) \to 1$, as $n \to \infty$)

$$\begin{aligned}
&\liminf_n \frac{1}{n}\log P\left(\frac{S_n}{n} \in (x-\epsilon, x+\epsilon)\right) \geq \\
&\quad\geq \liminf_n \frac{1}{n}\log P\left(\frac{S_n}{n} \in (x-\delta, x+\delta)\right) \\
&\quad= -(x\tau+\delta\tau-\log\phi(\tau)) + \liminf_n \frac{1}{n}\log\hat{P}\left(\frac{S_n}{n} \in (x-\delta, x+\delta)\right) \\
&\quad= -(x\tau+\delta\tau-\log\phi(\tau)) \\
&\quad= -\phi^*(x) - \delta\tau.
\end{aligned}$$

Now, we just need to let $\delta \to 0$.

There only remains the case that the $\sup_t\{xt - \log\phi(t)\}$ is not achieved. In this case because of the convexity of $\log\phi(t)$, and the fact that $\phi(t) < \infty$, for every t, there must be a sequence $t_n \to +\infty$ (assume $x > 0$, similarly for $x < 0$), such that

$$\begin{aligned}\phi^*(x) &= \lim_{t_n\to\infty}\left(xt_n - \log\phi(t_n)\right)\\ &= -\lim_{t_n\to\infty}\log\int e^{t_n(y-x)}F(dy)\\ &= -\lim_{t_n\to\infty}\log\int_{y\geq x} e^{t_n(y-x)}F(dy).\end{aligned}$$

If now $F\left((x,\infty)\right) > 0$, then by the monotone convergence the last integral converges to $+\infty$, and this will imply that $\phi^*(x) = -\infty$, which is false by the positivity of ϕ^*. Thus we get that $F(\{x\}) = e^{-\phi^*(x)}$. Then

$$\begin{aligned}&\liminf_n \frac{1}{n}\log P\left(\frac{S_n}{n}\in(x-\epsilon,x+\epsilon)\right)\\ &\qquad\geq \liminf_n \frac{1}{n}\log P\left(X_1 = \cdots X_n = x\right) = \log P\left(X_1 = x\right) = -\phi^*(x)\end{aligned}$$

This completes the proof of Crámer's theorem. □

REMARK: Crámer's theorem holds also without the assumption that $\phi(t) < \infty$, for every t. In fact it holds even in the case that $D_\phi := \{t\colon \phi(t) < \infty\} = \{0\}$, although in this case the rate function ϕ^* might be trivial. To prove it in this case one proves it first for the measures $\nu_n^M(\cdot) := P\left(\frac{S_n}{n}\in\cdot\middle||X_i| < M\right)$, which reduces to the case that we considered and then passes to the limit $M \to \infty$.

2.2. **EMPIRICAL MEASURES-SANOV'S THEOREM-ENTROPY.**

Let us consider the simplest case. Let $(X_i)_{i\geq 1}$ i.i.d. variables with marginal distribution ρ_s, i.e. $P(X_i = s) = \rho_s$, for $1 \leq s \leq r$, and for every r, $\rho_r > 0$.

So far we have considered the case of LDP of the empirical average

$$\frac{X_1 + \cdot + X_N}{N}.$$

Consider now the empirical measure

$$L_n = \frac{\delta_{X_1} + \cdot + \delta_{X_n}}{n}.$$

This is a random measure on $\Gamma = \{1,\ldots,r\}$ such that

$$\begin{aligned}L_N(s) &= \frac{1_{X_1=s} + \cdots + 1_{X_N=s}}{N}\\ &= \frac{\#\{i \leq N\colon X_i = s\}}{N}.\end{aligned}$$

Notice that this is a generalisation of the example considered in the introduction. If we know the asymptotic behavior of L_N then we can deduce the asymptotics of

$$\frac{f(X_1) + \cdots + f(X_n)}{n} = \int f(y)L_n(dy).$$

EXAMPLE: Let's compute the asymptotic probability

$$P\left(L_N(s) \cong k_s, \text{for } s \in \{1,\ldots,r\}\right),$$

where $k_s \geq 0$ and $\sum_{s=1}^{r} k_s = 1$. In other words $\nu = (k_1, \ldots, k_r)$ is a probability measure on Γ.

By elementary combinatorics we have that

$$\begin{aligned} P\left(L_N(s) \cong k_s, \text{for } s \in \{1, \ldots, r\}\right) &= \binom{N}{k_1 N\, k_2 N \cdots k_r N} \rho_1^{k_1 N} \cdots \rho_r^{k_r N} \\ &= \frac{N!}{(k_1 N)! \cdots (k_r N)!} \rho_1^{k_1 N} \cdots \rho_r^{k_r N}, \end{aligned}$$

and using Stirling's formula again we have that it is asymptotically equal to

$$\left(\frac{1}{\sqrt{2\pi\, N}}\right)^{r-1} k_1^{-Nk_1} \cdots k_r^{-Nk_r}\, \rho_1^{k_1 N} \cdots \rho^{k_r N}.$$

So

$$\frac{1}{N} P\left(L_N \cong \nu\right) \cong -\sum_{s=1}^{r} k_s \log k_s + \sum_{s=1}^{r} k_s \log \rho_s - \frac{r-1}{N} \log \sqrt{2\pi\, N}$$

$$= \quad -\sum_{s=1}^{r} k_s \log \frac{k_s}{\rho_s} + o(1)$$

Definition 2.4. *Consider a measurable space X and two probability distributions on it μ, ν. Define the relative entropy of ν with respect to μ as the quantity*

$$H(\nu \big| \mu) = \{ \begin{array}{lll} \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu & , & \nu \ll \mu \\ \infty & , & \textit{if not} \end{array}$$

Notice that $\int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu = \int \log \frac{d\nu}{d\mu} d\nu$.

We have thus shown that

$$\frac{1}{N} \log P\left(L_N \cong \nu\right) \cong H(\nu \big| \rho).$$

To state the result precisely we need to specify the space and the metric. The space will be the space $\mathcal{M}_1(\Gamma)$ of probability measures on Γ with the variational metric, that is

$$d(\mu, \nu) - \frac{1}{2} \sum_{r} |\, \mu_s - \nu_s \,|.$$

Notice that by the LLN of the ergodic theorem we have that $d(L_N, \rho) \to 0$.

Theorem 2.5. *Let (X_i) i.i.d variables. Then*

$$\frac{1}{n} \log P\left(L_n \in B_a^c(\rho)\right) \to -\inf_{\{\nu \in B_a^c(\rho)\}} I_\rho(\nu),$$

where $I_\rho(\nu) = H(\nu \big| \mu)$, and $B_a(\rho)$ is the ball of radius a around the measure ρ.

For the proof look at den Hollander's book. It is basically the previous argument combined with density and the continuouity of relative entropy.

CONNECTION BETWEEN THE TWO EXAMPLES.

sp_contraction **Proposition 2.6.**

$$\phi^*(x) = \inf_{\{\nu :\ \int y\, \nu(dy) = x\}} H(\nu \big| \mu)$$

Proof.

$$\phi^*(x) = \sup_t \{xt - \log \phi(t)\} = \sup_t \{xt - \log \int e^{ty} \mu(dy)\}$$

Also by Jensen's inequality

$$\begin{aligned} \log \int e^{ty} \mu(dy) &= \log \int e^{ty} \frac{d\mu}{d\nu}(y)\, \nu(dy) \\ &\geq t \int y\, \nu(dy) + \int \log \frac{d\mu}{d\nu}\, \nu(dy) = t \int y\, \nu(dy) - H(\nu \big| \mu). \end{aligned}$$

So

$$\begin{aligned} H(\nu \big| \mu) &\geq t \int y\, \nu(dy) - \log \phi(t) \\ &= tx - \log \phi(t). \end{aligned}$$

if $\int y\, \nu(dy) = x$. So

$$\inf_{\{\nu:\ \int y\, \nu(dy) = x\}} H(\nu \big| \mu) \geq \sup_t \{xt - \log \phi(t)\}$$

Equality holds when $\frac{d\mu}{d\nu}(y) e^{ty} = C$, where C will be the normalizing constant, or

$$d\nu = \frac{e^{ty}}{\int e^{ty} d\mu}\, d\mu.$$

Notice that $\int y\, \nu(dy) = \frac{\int y e^{ty} \mu(dy)}{\int e^{ty} \mu(dy)} = \big(\log \phi(t)\big)'$. If we choose t s.t. $\big(\log \phi(t)\big)' = x$ (suppose we can do it) then we can construct a measure ν, with $\int y \nu(dy) = x$ and for which equality holds. this implies the result.

In the case that the equation doesn't have a solution, we take a sequence of t_n along which we approximate the $\sup_t \{xt - \log \phi(t)\}$, and in this way we construct an approximating sequence of $H(\nu_{t_n} \big| \mu)$. □

REMARK: The above proposition is a special case of a more general principle, called the Contraction Principle. We will state it later.

2.3. **LARGE DEVIATIONS FOR EMPIRICAL MEASURE OF 2-LETTER WORDS.** Again (X_i) is a sequence of i.i.d. variables on $\Gamma = \{1, \dots, r\}$, and consider the empirical measure

$$L_n^{(2)} = \frac{\sum_{i=1}^n \delta_{(X_i, X_{i+1})}}{n}.$$

We assume periodic boundary condition so that there is no ambiguity. Notice that $L_n^{(2)} \in \tilde{\mathcal{M}}_1(\Gamma \times \Gamma)$, the subspace of $\mathcal{M}_1(\Gamma \times \Gamma)$ such that $\sum_t \nu_{st} = \sum_t \nu_{ts}$. Consider also the metric

$$d(\mu, \nu) = \frac{1}{2} \sum_{s,t} |\, \mu_{st} - \nu_{st}\, |.$$

under this metric the space $\mathcal{M}_1(\Gamma \times \Gamma)$ is Polish. by the ergodic theorem we have that $d(L_n^{(2)}, \rho \times \rho) \to 0$. The large deviations are summarised in

Theorem 2.7. *If (X_i) are i.i.d. then*

$$\lim_n \frac{1}{n} \log P\left(L_n^{(2)} \in B^c_{\rho\times\rho}(a)\right) = -\inf_{\nu\in B^c_\rho(a)} I_\rho^{(2)}(\nu),$$

where $I_\rho^{(2)}(\nu) = H\left(\nu \middle| \bar{\nu}\times\rho\right)$, and $\bar{\nu}_s = \sum_t \nu_{st}$. Finally $B_\rho(a)$ is the ball of radius a.

Proof. We need to compute $P\left(L_n^{(2)} \cong \nu\right)$, We can first do it for ν such that $\nu_{st} = \frac{k_s t}{n}$, with $0 \le k_{st} \le n$, $\sum_{st} = n$ and k_{st} integers. Tp count we make a directed graph whose vertices are the elements of Γ and an edge connects s and t if there is the word st in the sequence of $X_i's$. Because of periodicity one can go through all the edges of the graph passing only ones from each one. We call this Euler circuit. So

$$P\left(L_n^{(2)} = \nu\right) = O(n)\,\frac{\#\{\text{Euler circuits}\}}{\prod_{s,t}\nu_{st}!}\prod_s \rho_s^{\bar{\nu}_s\, n}.$$

The factor $O(n)$ is because the circuit can start anywhere in the sequence of $X_i's$. To count the number of Euler circuits notice that this is almost $\prod \bar{\nu}_s!$.

We can now use the Stirling's formula to conclude the result as in the case of simple empirical measure.

To prove the LDP just write

$$\max_{\nu\in B^c_\rho(a)\cap K_n} P\left(L_n^{(2)} = \nu\right) \le P\left(L_n^{(2)} \in B^c_\rho(a)\right) \le |K_n| \max_{\nu\in B^c_\rho(a)\cap K_n} P\left(L_n^{(2)} = \nu\right)$$

and $|K_n| = n^{r^2}$. Finally we need the following lemma □

Lemma 2.8. *The rate function $I_\rho^{(2)}$ is finite, continuous, convex and affine on segments $\alpha\nu + (1-\alpha)\nu'$, where $\frac{\nu_{st}}{\bar{\nu}_s} = \frac{\nu'_{st}}{\nu'_s}$.*

Also it is nonnegative and equal to 0 only if $\nu = \rho\times\rho$.

2.4. **WORDS OF LENGTH N..** We can extend the previous result to the empirical measures

$$L_n^N = \frac{1}{n}\sum_{i=1}^n \delta_{X_i,\dots,X_{i+n-1}},$$

with periodic conditions, The periodicity makes the empirical measure an element of the space $\tilde{\mathcal{M}}_1(\Gamma^N) = \{\nu \in \mathcal{M}_1(\Gamma)\colon \sum_{s_n}\nu_{s_1,\dots,s_N} = \sum_{s_n}\nu_{s_N,s_1,\dots,s_{N-1}}\}$ We consider the metric

$$d(\mu,\nu) = \frac{1}{2}|\,\mu_{s_1,\dots,s_N} - \nu_{s_1\dots,s_N}|.$$

The LDP in this case reads as

Theorem 2.9. *If (X_i) are i.i.d the*

$$\lim_n \frac{1}{n} P\left(L_n^N \in B^c_\rho(a)\right) = -\inf_{\nu\in B^\rho_c(a)} I_\rho^N(\nu),$$

where $I_\rho^N(\nu) = H(\nu|\bar{\nu}\times\rho)$, with $\bar{\nu}$ the $(N-1)$-dimensional marginal of ν.

Again we have that

Lemma 2.10. *the rate function I_ρ^N is finite, continuous, convex and affine on segment $\alpha\nu + (1-\alpha)\nu'$, with $\nu_{s_1\dots s_n}/\bar{\nu}_{s_1\dots s_{N-1}} = \nu'_{s_1\dots s_n}/\bar{\nu'}_{s_1\dots s_{N-1}}$. It is also nonnegative and equal to 0 only if $\nu = \rho^N$.*

2.5. **THE EMPIRICAL PROCESS..** We now consider words of length n themselves i.e.

$$L_n^n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i,\cdots,X_{i+n-1}}.$$

it is useful to consider a periodic sequence as

$$\mathbb{X}^n = X_1, X_2, \ldots, X_n, X_1, X_2 \ldots$$

we can also view the measure L_n^n as a measure on the space of sequences as

$$R_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\sigma^i(\mathbb{X}^n}.$$

where σ is the usual left shift. It is easy to see that $R_n\sigma^{-1} = R_n$, i.e. R_n is shift invariant. denote by $\hat{M}_1(\Gamma^{\mathbb{N}})$ the space of shift invariant measures on the space of sequences. The metric that we will consider will be

$$d(\mu,\nu) = \sum_N 2^{-N} d_N(\pi_N\nu, \pi_N\mu),$$

where π_N is the projection to Γ^N.

Notice that $\pi_n\nu \in \tilde{\mathcal{M}}_1(\Gamma^N)$.

A simple application of the ergodic theorem shows that $P - a.s.$, R_n converges weakly to $\rho^{\mathbb{N}}$, where the weak convergence is understood as convergence on finite dimensional cylinders. The LDP for the empirical process is described in the following theorem

Theorem 2.11. *Let (X_i) i.i.d. sequence on Γ. For $a > 0$ let $B_a(\rho^{\mathbb{N}})$ the closed ball in $\hat{\mathcal{M}}_1$ of radius a around the measure $\rho^{\mathbb{N}}$. Define*

$$J(a) = \inf_{\nu\in B_a(\rho^{\mathbb{N}})} I_\rho^\infty(\nu),$$

with

$$I_\rho^\infty(\nu) = \sup_N H(\pi_n\nu \big| \pi_{N-1}\nu \times \rho).$$

Then

a. $\liminf_n \frac{1}{n}\log P(R_n \in B_a^c(\rho^{\mathbb{N}})) \geq -J(a)$

b.$\limsup_n \frac{1}{n}\log P(R_n \in B_a^c(\rho^{\mathbb{N}})) \leq -J(a^-)$.

The proof is fairly easy and we refer to den Hollander. It uses the following lemma

Lemma 2.12. *A. $N \to I_\rho^N(\pi_N\nu)$ is nondecreasing on $\hat{\mathcal{M}}_1(\Gamma^{\mathbb{N}})$.*

B. $a \to J(a)$ is right continuous and nondecreasing.

C. $\inf_{\nu\in\hat{\mathcal{M}}_1(\Gamma^{\mathbb{N}}):\, \pi_M\nu=\mu_M} I_\rho^\infty(\nu) = I_\rho^M(\nu_M)$, *for $M \in \mathbb{N}$ and $\mu_M \in \tilde{\mathcal{M}}_1(\Gamma^M)$*

Proof. The proof of A. uses Jensen's inequality together with the shift invariance of the measures. The proof of B. is elementary. The proof of C. is nontrivial.

We need to construct a measure μ on the space of sequences such that the projection to the M-dimensional subspace is μ_M and

ent_eq

$$I_\rho^\infty(\nu) = I_\rho^M(\nu_M). \tag{2.1}$$

This will imply that the left hand side of C. is less or equal to the right hand side. The other direction follows from A.

To construct this measure we will use Kolmogorov's constistncy theorem. We thus need to construct a consistent sequence of measures on any $N-$dimensional subspace, with $N \geq M$, i.e. $\pi_{N-1}\mu_N = \mu_{N-1}$. One cheap way to get the equality ent_eq (2.1) is to construct the measures μ_N in such a way that $I_\rho^{N+1}(\mu_{N+1}) = I_\rho^N(\mu_N)$. In fact, the solution of this equation will produce the correct measures. Notice that A. gives us that □

3. General Theory.

We will now present some general principles. First let us recall some definitions.

rate_function2 **Definition 3.1.** *A function $I : \mathcal{X} \to [0,\infty]$ is called a rate function if it is not identically equal to $+\infty$ and lower semicontinuous. It is often called a good rate function if its level sets are compact.*

Recall also the definition of the LDP. A very easy consequence of the definition of the LDP and the lower semicontinouity of the rate funtion is the following:

Proposition 3.2. *The rate function I is unique.*

Remark 3.3. *Once again let us point out that the distinction between the closure and the interior in the LDP is related to the fact that there might me concentration of measure on the boundary of the set. This is in the same spirit as in the weak convergence of measures. In particular, we know that a sequence of measures P_n converges weakly to a measure P if and only if $\limsup_n P_n(C) \leq P(C)$, for every closed set C, and $\limsup_n P_n(O) \geq P(O)$ for every open set O.*

exp_tight **Definition 3.4.** *A sequence of measures P_n on $\mathcal{X}$ is called exponentially tight if for every $M > 0$ one can find a compact set K_M such that*

$$\limsup_n \frac{1}{n} \log P_n\left(\mathcal{X} \setminus K_M\right) \leq -M.$$

Notice that a LDP with a good rate function implies exponential tightnes.

In many cases it is difficult to prove a full LDP directly. On the other hand it is easier to prove the following weak version of it. If also one has exponential tightness the one can derive the full LDP.

weakLDP **Definition 3.5.** *We say that the sequence P_n satisfies the weak LDP if the upper bound in the definition of the LDP is replaced by*

$$\limsup_n \frac{1}{n} \log P_n(K) \leq -\inf_{x\in K} I(x),$$

for every K compact.

It is easy to check that the weak LDP together with exponential tightness imply the full LDP.

The first basic result in the theory is Varadhan's lemma, which is a generalisation of the Laplace asymptotics.

Varadhan **Theorem 3.6.** *Let (P_n) satisfy the (full) LDP with good rate function I. Let also F be a continuous function on the Polish space $\mathcal{X}$, such that*

condition
$$\limsup_M \limsup_n \frac{1}{n} \log \int_{\{F>M\}} e^{nF(x)}\, P_n(dx) = -\infty. \tag{3.1}$$

Then

Laplace
$$\lim_n \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)}\, P_n(dx) = \sup_{x\in\mathcal{X}} \left(F(x) - I(x) \right). \tag{3.2}$$

Proof. Both for the upper and the lower bound we will be using the fact that if f is a l.s.c. function and $B_r(x)$ a ball (open/closed) around the point x, then

lsc
$$\lim_{r\to 0} \inf_{B_r(x)} f = f(x). \tag{3.3}$$

UPPER BOUND

Assume first that $F \le M$ for some M positive. Pick L large enough and write $\mathcal{X} = \{I \le L\} \cup \{I > L\}$. Then by the LDP

$$\limsup_n \frac{1}{n} \log \int_{\{I>L\}} e^{nF(x)} P_n(dx) \le M - \inf_{\{I>L\}} I(x) \le M - L,$$

and this can be done very small as $L \to \infty$. Thus, by (1.1), (3.2) is asymptotically equivalent to

$$\limsup_n \frac{1}{n} \log \int_{\{I\le L\}} e^{nF(x)} P_n(dx).$$

By the goodness of I, the set $\{I \le L\}$ is compact and so it can be covered by a finite number of *closed* balls, say N, $B_{r_i}(x_i)$ where x_i are points in this set and r_i are chosen so that $\inf_{B_{r_i}(x_i)} I > I(x_i) - \delta$ (this can be done by (3.3) and $\sup_{B_{r_i}(x_i)} F \le F(x_i) + \delta$ (this can be done by the continuouity of F We take *closed* balls instead of open, just to use the upper LDP bound. We now have

$$\begin{aligned}
&\limsup_n \frac{1}{n} \log \int_{\{I\le L\}} e^{nF(x)} P_n(dx) \\
&\le \limsup_n \frac{1}{n} \log \sum_i \int_{B_{r_i}(x_i)} e^{nF(x)} P_n(dx) \\
&\le \limsup_n \frac{1}{n} \log \sum_i e^{n(F(x_i)+\delta)} P_n(B_{r_i}(x_i)) \\
&\le \max_i \left(F(x_i) + \delta - \inf_{B_{r_i}(x_i)} I \right) \\
&\le \max_i \left(F(x_i) + \delta - I(x_i) + \delta \right) \\
&\le \sup_x (F - I) + 2\delta.
\end{aligned}$$

If F is not bounded from above, consider the function $F \wedge M$. The previous computation shows that

$$\begin{aligned}
\limsup_n \frac{1}{n} \log \int_{\mathcal{X}} e^{nF\wedge M\,(x)}\, P_n(dx) &\le \sup_{x\in\mathcal{X}} \left(F \wedge M(x) - I(x) \right) \\
&\le \sup_{x\in\mathcal{X}} \left(F(x) - I(x) \right)
\end{aligned}$$

Now, we just need to split the integral over the sets $\{F > M\}$, $\{F \le M\}$, for some large M and then we use (3.1) combined with the elementary equality (1.1).

In more detail

$$\begin{aligned}
&\limsup_n \frac{1}{n}\log\int_{\mathcal{X}} e^{nF}\,P_n(dx)\\
&=\max\left(\limsup_n \frac{1}{n}\log\int_{F>M} e^{nF}\,P_n(dx),\limsup_n \frac{1}{n}\log\int_{F\le M} e^{nF}\,P_n(dx)\right)\\
&=\limsup_n \frac{1}{n}\log\int_{F\le M} e^{nF}\,P_n(dx)\\
&\le \sup_{x\in\mathcal{X}}\big(\,F(x)-I(x)\,\big)
\end{aligned}$$

LOWER BOUND

$$\begin{aligned}
&\liminf_n \frac{1}{n}\log\int_{\mathcal{X}} e^{nF(y)}\,P_n(dy)\\
&\ge \liminf_n \frac{1}{n}\log\int_{B_r(x)} e^{nF(y)}\,P_n(dy)\\
&\ge \inf_{B_r(x)} F-\inf_{B_r(x)} I.
\end{aligned}$$

Since r is arbitrary, we can use the continuouity of F and the l.s.c. of I via (lsc 3.3) to let $r\to 0$ and get that

$$\liminf_n \frac{1}{n}\log\int_{\mathcal{X}} e^{nF(y)}\,P_n(dy)\ge F(x)-I(x).$$

. The result now follows by the arbitrariness of x. □

Remark 3.7. *Notice that in the proof we used the full LDP through the estimate*

$$\limsup_n \frac{1}{n}\log P_n(B_r(x))\le -\inf_{B_r(x)} I,$$

which might not be true if we only have the weak LDP, since closed balls in Banach spaces might not be compact.

Remark 3.8. *It is interesting to check the above Lemma in the case that $F(x)=\lambda x$, say $\mathcal{X}=\mathbb{R}$. In this case we have that*

$$\lim_n \frac{1}{n}\log\int_{\mathbb{R}} e^{n\,\lambda y}\,P_n(dy)=\sup_x\,(\lambda x-I(x))=I^*(\lambda)$$

Let's denote by $\Lambda(\lambda)$ the right hand side of the above equation. If we know that I is convex, then $I^{**}=I$ and so $I(x)=\Lambda^*(x)=\sup_\lambda(\lambda x-\Lambda(\lambda))$. Compare this with the form of the rate function in Crámer's theorem. This observation gives us a way to identify the rate function.

The following lemma provides a converse of Varadhan's lemma.

Theorem 3.9. *Let*

$$\Lambda_n(F)=\frac{1}{n}\log\int_{\mathcal{X}} e^{nF(x)}P_n(dx),$$

for $F\in C_b(\mathcal{X})$. If (P_n) is exponentially tight and $\Lambda(F)=\lim_n\Lambda_n(F)$ exists for every $F\in C_b(\mathcal{X})$, then (P_n) satisfies a LDP with rate function

$$I(x)=\sup_{F\in C_b(\mathcal{X})}(F(x)-\Lambda(F))$$

Proof. Both the upper and the lower bound follow the ideas in Crámer's theorem.

UPPER BOUND

For the upper bound we will imitate the exponential Chebyshev inequality. By the exponential tightness of the (P_n) it is enough to prove the weak LDP. Let C be a compact set. Then

$$P_n(C) = \int_C P_n(dx) \leq \int_C e^{n(F(x) - \inf_C F)} P_n(dx)$$

So,

$$\limsup_n \frac{1}{n} \log P_n(C) \leq -\inf_C (F(x) - \Lambda(F))$$

Since the choice of F was arbitrary we have that

$$\limsup_n \frac{1}{n} \log P_n(C) \leq -\sup_{F \in C_b(\mathcal{X})} \inf_C (F(x) - \Lambda(F))$$

So far we haven't used the fact that C ic compact. We now need to interchange the inf with the sup and here is where we will use the compactness of C. Because of its compactness we can cover it by a finite number of balls $B_{r_i}(x_i)$ where $x_i' s$ are points in C and $r_i' s$ are chosen such that $\inf_{B_{r_i}(x_i)} F \geq F(x_i) - \delta$, this can be done by the continuouity of F.

Using now the elementary equality (1.1) we have that

$$\begin{aligned}\limsup_n \frac{1}{n} \log P_n(C) \quad &\leq \quad \max_i \limsup_n \frac{1}{n} \log P_n(\bar{B}_{r_i}(x_i)) \\ &\leq -\min_i \sup_{F \in C_b(\mathcal{X})} \inf_{\bar{B}_{r_i}(x_i)} (F(x) - \Lambda(F)) \\ &\leq -\min_i \sup_{F \in C_b(\mathcal{X})} (F(x_i) + \delta - \Lambda(F)) \\ &\leq -\inf_C \sup_{F \in C_b(\mathcal{X})} (F(x) - \Lambda(F)) - \delta\end{aligned}$$

LOWER BOUND

Let O open set $x \in O$ and r small enough such that $B_r(x) \subset O$ Choose a function F_M which is bounded, continuous, nonpositive, $F_M(x) = 0$ and $F_M(x) = -M$ for $x \in B_r^c(x)$, the complement of a ball around x. Then

$$\begin{aligned}\int_{\mathcal{X}} e^{nF(x)} P_n(dx) \quad &= \quad e^{-Mn} P_n(B_r^c(x)) + \int_{B_r(x)} e^{nF(x)} P_n(dx) \\ &\leq \quad e^{-Mn} + P_n(B_r(x)) \leq e^{-Mn} + P_n(O).\end{aligned}$$

By the elementary equality (1.1) and the fact that $F(x) = 0$, we have

$$\begin{aligned}\max\left(-M\,, \liminf_n \frac{1}{n} \log P_n(O) \right) \quad &\geq \quad \lim_n \frac{1}{n} \int_{\mathcal{X}} e^{nF(x)} P_n(dx) \\ &= \quad -(\,F(x) - \Lambda(F)) \\ &\geq \quad \sup_{F \in C_b(\mathcal{X})} (\,F(x) - \Lambda(F))\,.\end{aligned}$$

To conclude let $M \to \infty$. □

Remark 3.10. *Notice that the rate function that appears in Bryc's lemma need not be convex.*

Now we will give the contraction principle. We already saw a version of it when we related the entropy with the Legendre tranform in Crámer's theorem.

Theorem 3.11. *Let (P_n) a sequence of measures on a Polish space $\mathcal{X}$ that satisfies a LDP with rate function I. Consider also another Polish space $\mathcal{Y}$ and a continuous maping $T : \mathcal{X} \to \mathcal{Y}$ as well as the induced sequence of measures $Q_n = P_n T^{-1}$. Then Q_n satisfies a LDP with rate function*

$$J(y) = \inf_{x \in \mathcal{X} \colon Tx = y} I(x). \tag{3.4}$$

Proof. We will only show the upper bound. The lower bound is the same. Let C closed set in $\mathcal{Y}$. Then, notice that $T^{-1}(C)$ is also closed,

$$\limsup_n \frac{1}{n} \log Q_n(C) = \limsup_n \frac{1}{n} \log P_n(T^{-1}(C))$$
$$\leq - \inf_{x \in T^{-1}(C)} I(x) = - \inf_{y \in C} \inf_{x \colon Tx = y} I(x).$$

□

Remark 3.12. *Proposition 2.6 is a special case of the contraction principle where $\mathcal{X} = \mathcal{M}_1(\Gamma)$, $\mathcal{Y} = \mathbb{R}$, $T\nu = \int x\nu(dx)$, $I(\nu) = H(\nu|\mu)$ and $J(x)$ is the Legendre transform of the moment generating function of μ.*

Remark 3.13. *We can also use the contraction principle to relate the LDP's for the empirical measures for words of different lengths. For example we can consider the $\mathcal{X} = \tilde{\mathcal{M}}_1(\Gamma^N)$ and $\mathcal{Y} = \tilde{\mathcal{M}}_1(\Gamma^M)$ for $M < N$ and the transformation $T \colon \tilde{\mathcal{M}}_1(\Gamma^N) \to \tilde{\mathcal{M}}_1(\Gamma^M)$ as the projection of an N-dimensional measure to the M-dimensional subspace. Then we can obtain that*

$$I_\rho^M(\nu_M) = \inf_{\nu \in \tilde{\mathcal{M}}_1(\Gamma^N) \colon \pi_M \nu = \nu_M} I_\rho^N(\nu).$$

In the same way one could immediatelly obtain the third conclusion of Proposition **??**. *The only problem would be that one needs to know a priori the LDP, while we used that property in order to obtain the LDP.*

3.1. **CONVEXITY.** We will now prove Crámer's theorem in an abstract setting that makes use of convexity considerations. We should expect that convexity plays an important role since the rate function in Crámer's theorem is a convex function, namely the Legendre transform of the log-moment generating function.

The setting we will be working is this of a Polish space $\mathcal{X}$. We will also assume

Assumption 3.14. *1. $\mathcal{X}$ is a convex subset of a linear space.*
2. $\mathcal{X}$ is locally convex.
3. $\lim_{\beta \to \alpha} d(\alpha x + (1-\alpha)y, \beta x + (1-\beta)y) = 0$.
4. In $\mathcal{X}$ the closed convex hull of a compact subset is compact.

Theorem 3.15. *Let $\mathcal{X}$ a polish space that satisfies the assumptions 3.14. Consider a sequence (Y_i) of i.i.d. variables in this space and define the sequence of measures (P_n) by*

$$P_n(A) = \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n Y_i \in A \right).$$

Then

A. For every open, convex set $A \subset \mathcal{X}$ we have that

$$\lim_n \frac{1}{n} \log P_n(A) = -\hat{I}(A),$$

exists and is finite iff $P_n(A) > 0$ for some n.

B. The function

$$I(x) = \sup_{A:\, x\in A, open, convex} \hat{I}(A)$$

is a rate function.

C. I is convex.

Proof. We will only give an outline of the proof. The details are in den Hollander's book.

A. Here the crucial thing is subadditivity. This is a general tool that often underlines LDPs. We will obtain the subadditivity by the convexity of A.

$$\begin{aligned} P_{n+m}(A) &= \mathbb{P}\left(\frac{1}{n+m}\sum_{i=1}^{n+m} Y_i \in A\right) \\ &\geq \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \in A; \frac{1}{n}\sum_{i=n}^{n+m} Y_i \in A\right) \\ &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \in A\right) \mathbb{P}\left(\frac{1}{n}\sum_{i=n}^{n+m} Y_i \in A\right) \\ &= P_n(A)\, P_m(A) \end{aligned}$$

So $-\log P_n(A)$ is subadditive. One now uses the elemntary fact that if a_n is a subadditive sequence of positive numbers, then $\lim_n a_n/n = \inf_n a_n/n$. One finally needs to check that if $P_n(A)$ is positive for one n then it is positve for all greater $n's$.

B. For the lower semicontinuouity we refer to den Hollander.

C. The convexity follows the same strategy as A. Let x_1, x_2 two points. Let A_1, A_2 open convex such that $x_1 \in A_1$ and $x_2 \in A_2$. Consider the set $A = 1/2A_1 + 1/2A_2$. Then

$$\begin{aligned} -\hat{I}(A) &= \lim_n \frac{1}{2n} \log \mathbb{P}\left(\frac{1}{2n}\sum_{i=1}^{2n} Y_i \in A\right) \\ &\geq \lim_n \frac{1}{2n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \in A_1\right) \cdot \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \in A_2\right) \\ &= -\frac{1}{2}\hat{I}(A_1) - \frac{1}{2}\hat{I}(A_2). \end{aligned}$$

So we get,

$$\hat{I}(A) \leq \frac{1}{2}\hat{I}(A_1) + \frac{1}{2}\hat{I}(A_2) \leq \frac{1}{2}I(x_1) + \frac{1}{2}I(x_2)$$

and now take the supremum over all open, convex $A's$ □

Theorem 3.16. *(***General Crámer's Theorem***) The sequence of measures (P_n) defined above satisfy a weak LDP with rate function I defined in the above proposition.*

Proof. UPPER BOUND

Consider the function $I^\delta(X) = \min(I(x) - \delta, \frac{1}{\delta})$. We take into account the possibility that I might equal infinity at some point, and that's why we truncate by $\frac{1}{\delta}$. By the definition of I, we can find open, convex set A^δ_x, that contains x, such that $\hat{I}(A^\delta_x) > I^\delta(x)$. Also, by the previous proposition we have that

$$\lim_n \frac{1}{n} \log P_n(A^\delta_x) = -\hat{I}(A^\delta_x) \le -I^\delta(x).$$

Let now C be a compact subset of $\mathcal{X}$. then, for every $\delta > 0$, there is a finite covering of C by sets of the form $A^\delta_{x_i}$, i.e. $C \subset \cup_i A^\delta_{x_i}$. Then using the elementary equality elem (1.1) we have that

$$\limsup_n \frac{1}{n} \log P_n(C) \le \max_i -I^\delta(x_i) \le -\inf_C I^\delta(x).$$

Let now $\delta \to 0$.

LOWER BOUND

Let O be an open set. The because we assume that $\mathcal{X}$ is locally convex we can find open, convex set $x \in A \subset O$. Then

$$\liminf_n \frac{1}{n} \log P_n(O) \ge \liminf_n \frac{1}{n} \log P_n(A) = -\hat{I}(A).$$

The result now follows by taking the supremum over such $A's$. □

4. Large Deviations for Path Processes- Exit Problems.

Consider the problem $\epsilon \frac{1}{2} u_{xx} + b(x) u_x = 0$ in $[-1, 1]$ with boundary conditions $u(-1) = A$ and $u(1) = B$. The problem is to determine what is the limit as $\epsilon \to 0$ of the solution $u^\epsilon(x)$. One could guess that the limit is the the soltuion of the equation $b(x)u_x = 0$ with boubdary values $u(-1) = A, u(1) = B$. But there is some ambiguity if $b(x) \neq 0$, and $A \neq B$, since then $u_x = 0$, which means that u is constant with different values at $1, -1$.

One needs to make a more carefull analysis and we will see that LDP for the measure P_ϵ, which corresponds to the distribution of the path $\sqrt{\epsilon}\beta(t)$, where $\beta(\cdot)$ is Brownian Motion, plays an important role.

The first result states this LDP

Schilder **Theorem 4.1.** *(Schilder) Let $\Omega = C\big([0, T]; \mathbb{R}^d\big)$, and P_ϵ the distribution of $\sqrt{\epsilon}\beta(\cdot)$. Then we have that*

$$\limsup_{\epsilon \to 0} \epsilon \log P_\epsilon(A) \le - \inf_{f \in A, f(0)=0} I(f)$$

for $A \subset C\big([0, T]; \mathbb{R}^d\big)$ closed, and

$$\liminf_{\epsilon \to 0} \epsilon \log P_\epsilon(G) \ge - \inf_{f \in G, f(0)=0} I(f),$$

The rate function $I(f) = \frac{1}{2} \int_0^T |\dot{f}(s)|^2 ds$ if $\dot{f} \in L^2\big([0, T]; \mathbb{R}^d\big)$ and infinity if not. The topology that we consider on the space of continuous functions is the uniform topology.

For the proof we will need some preliminaries on the Wiener measure. AS we know the Wiener measure has the property that the joint distribution of $(x(t_1),\dots,x(t_n))$ is joint Gaussian with covariance matrix $(t_i\wedge t_j)_{i,j}$. Another way to put this is that

$$E^P\Big[e^{\sum_{i=1}^n<\lambda_i,x(t_i)>}\Big]=e^{\frac{1}{2}\sum_{i,j}t_i\wedge t_i\,\lambda_i\lambda_j},$$

for any vector $\lambda_i\in\mathbb{R}^d$. This formula can be generalised by the taking approximation to

$$E^P\Big[e^{\sum_{i=1}^n\int_0^T x(t)\lambda(dt)}\Big]=e^{\frac{1}{2}\int_0^T t\wedge s\,\lambda(dt)\lambda(ds)},$$

for any $\mathbb{R}$ valued Borel measure $\lambda(dt)$ on $[0,T]$.

Let's define

$$\Lambda_W(\lambda)=\frac{1}{2}\int_0^T t\wedge s\,\lambda(dt)\lambda(ds),$$

the log-moment generating function of the Wiener measure. and let's define the Legendre transform of the log-moment generating function by

$$\begin{aligned}\Lambda_W^*(\phi)&=\sup_\lambda\{<\lambda,\phi>-\Lambda_W(\lambda)\}\\&=\sup_\lambda\{\int_0^T\phi(t)\lambda(dt)-\int_0^T\int_0^T t\wedge s\lambda(dt)\lambda(ds)\}.\end{aligned}$$

for any function $\phi\in C\big([0,T];\mathbb{R}^d\big)$.

Lemma 4.2. $\Lambda_W^*(\phi)=\frac{1}{2}\int_0^T|\dot\phi(s)|^2ds$, *where the interpretation of the integral is infinity if* $\dot\phi$ *does not belong in* $L^2\big([0,T];\mathbb{R}^d\big)$.

Proof. For the proof look at the book of Deutschel-Strock, on Large Deviation. □

Let's now give the proof of Schilders Theorem.

Proof. We will prove the weak LDP. To get the full LDP one needs to prove exonential tightness of P_ϵ. For this we refer to the book of Deutschel-Stroock. An alternative proof can be found in the book of Varadhan on Large devitaion. The proof we follow here follows very closely the steps of Cramer's Theorem.

UPPER BOUND. Let $B(\psi,r)=\{\phi\colon\|\phi-\psi\|_{C_b}<r\}$. Let's compute

$$\begin{aligned}P_\epsilon\big(\overline{B(\psi,r)}\big)&=P\big(\overline{B(\frac{\psi}{\sqrt\epsilon},\frac{r}{\sqrt\epsilon})}\big)\le\int_{\overline{B(\psi,r)}}e^{<\frac{\lambda}{\sqrt\epsilon},\theta>-\inf_{\phi\in\overline{B(\psi,\frac{r}{\sqrt\epsilon})}}<\frac{\lambda}{\sqrt\epsilon},\phi>}P(d\theta)\\&\le\sup_{\phi\in\overline{B(\psi,\frac{r}{\sqrt\epsilon})}}e^{-<\frac{\lambda}{\sqrt\epsilon},\phi>}\int_{\overline{B(\psi,r)}}e^{<\frac{\lambda}{\sqrt\epsilon},\theta>}P(d\theta)\\&\le\sup_{\phi\in\overline{B(\psi,\frac{r}{\sqrt\epsilon})}}e^{-<\frac{\lambda}{\sqrt\epsilon},\phi>}\int e^{<\frac{\lambda}{\sqrt\epsilon},\theta>}P(d\theta)\\&\le\sup_{\phi\in\overline{B(\psi,\frac{r}{\sqrt\epsilon})}}e^{-\frac{1}{\epsilon}(<\lambda,\psi>-r\|\lambda\|-\Lambda_W(\lambda))}\end{aligned}$$

If $\Lambda_W^*(\lambda)<\infty$, then choose λ such that $<\lambda,\psi>-\Lambda_W(\lambda)>\Lambda_W^*(\lambda)-\frac{\delta}{2}$, and $r-\frac{\delta}{2(1+\|\lambda\|)}$, to get that

$$\limsup_{\epsilon\to0}P_\epsilon(B(\psi,r))\le-\Lambda_W^*(\lambda)+\delta.$$

If $\Lambda_W^*(\lambda) = \infty$ then choose λ such that $< \lambda, \psi > -\Lambda_W(\lambda) > 1 + \frac{1}{\delta}$, and $r = \frac{1}{1+\|\lambda\|}$ to get that

$$\limsup_{\epsilon \to 0} P_\epsilon(B(\psi, r)) \leq -\frac{1}{\delta}.$$

Let K be a compact set. Then choose a finite covering of K with balls with center ψ_k's and radii r_k's such that the estimate just proved for $P_\epsilon(B(\psi_k, r_k))$ is true. Then the upper bound on $\limsup_{\epsilon \to 0} \epsilon \log P_\epsilon(K)$, follows the standard procedure.

LOWER BOUND. □