Sanov's theorem

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 1. *(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

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

□

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 $\nu_n(f) = \mu_n(f 1_{\bar{G}}) / \mu_n(\bar{G})$ for all f bounded Borel on $\mathcal{K}$. The set $\bar{G}$ is another compact metrizable space.

Corollary 2. *Assume $\min_{\bar{G}} I < +\infty$. The sequence of conditional measures $(\nu_n)_n$ obey the LDP with rate function $J: \bar{G} \rightarrow [0, +\infty]$ 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 $\|\hat{f} 1_{\bar{G}}\|_{L^n(\mu^n)} \rightarrow \max_{\bar{G}} (|\hat{f}| e^{-I})$ so $\|f\|_{L^n(\nu_n)} = \|\hat{f} 1_{\bar{G}}\|_{L^n(\mu^n)} / \|1_{\bar{G}}\|_{L^n(\mu^n)} \rightarrow \max_{\bar{G}} (|f| e^{-I}) / \max_{\bar{G}} (e^{-I})$. □

Tensorization and projections

Theorem 3. *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 $\|f_1 \otimes f_2\|_{L^{n_k}(\nu_{n_k})} = \|f_1\|_{L^{n_k}(\mu_{n_k}^1)} \|f_2\|_{L^{n_k}(\mu_{n_k}^2)}$ 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 $f_i(x_i) = \exp(-Nd_i(x_i, z_i))$ 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 1. 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 4. (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 $G_k(x) = (g_1(x), \ldots, g_k(x))$. Assume that for all $k \geqslant 1$ the laws $\mu_n^k = (G_k)_* \mu_n$ of the vector $(g_1, \ldots, g_k)$ 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 $C(\mathcal{K}_0)$ 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 $I = I'$. □

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 $\mu, \nu \in \Pi(\{0, \ldots, N\})$ 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 2. Prove that $(\mu_n)_n$ satisfy the LDP with rate function

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

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 $(X_n)_{n\geqslant 1}$ of iid variables on the finite set $\mathcal{K} = \{1, \ldots, N\}$ 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 5. (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 $(Y_n^1, \ldots, Y_n^N) \colon \Omega \to \{0,1\}^N$ 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 $\Lambda \colon \mathbb{R}^N \to \mathbb{R}$ of the log mgf of the vector Y_1, but

$$\Lambda(\lambda_1, \ldots, \lambda_N) = \log \mathbb{E}(e^{\lambda_1 Y_1^1 + \cdots + \lambda_N Y_1^N}) = \log \sum_{i=1}^{N} e^{\lambda_i} \rho_i$$

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

$$I(x_1, \ldots, x_N) = \sup_{\lambda_1, \ldots, \lambda_N} [\lambda_1 x_1 + \cdots + \lambda_N x_N - \log \sum_{i=1}^{N} e^{\lambda_i} \rho_i]$$

$$= H((x_i)_{1\leqslant i\leqslant N} | (\rho_i)_{1\leqslant i\leqslant N}) = \sum_{i=1}^{N} \log \frac{x_i}{\rho_i} x_i .$$

Remark 6. 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 7. (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 $\mathcal{F}_k=\sigma(\varphi_1,\ldots,\varphi_k)$ 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 $F_k(\nu)=(\nu(\varphi_1)/\|\varphi_1\|,\ldots,\nu(\varphi_k)/\|\varphi_k\|)$. By Cramer's theorem the image law $(\mu_n^k=(F_k)_*\mu_n)$ satisfy the LDP on $[-1,1]^k$ with rate function $I_k(x_1,\ldots,x_k)$ given by

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

Then by the Dawson-Gartnër theorem

$$I(\nu)=\sup_k I_k(F_k(\nu))=\sup_k\sup_{\lambda_1,\ldots,\lambda_k}[\nu(\lambda_1\varphi_1+\cdots+\lambda_k\varphi_k)-\log\mathbb{E}[e^{\sum_i\lambda_i\varphi_k(X_1)}]]$$

$$=\sup_{\varphi}[\nu(\varphi)-\log\mathbb{E}[e^{\varphi(X_1)}]]=H(\nu|\rho).$$

□

Sanov's theorem can be used to prove Cramérs theorem. Indeed the empirical mean $\hat{S}_n$ of a vector $(X_1,\ldots,X_n)$ is a function of the empirical distribution L_n: $\hat{S}_n=\int xL_n(\mathrm{d}x)=m(L_n)$. 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 $J(x)=\inf\{H(\nu|\rho)\colon\nu\in\Pi(\mathcal{K}),\ \mu(\nu)=x\}$. 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 $J(y)=\inf\{H(\nu|\rho_\lambda)\colon\nu\in\Pi(\mathcal{K}),\ \mu(\nu)=y\}+\lambda y-\log\Lambda_\rho(\lambda)=\lambda y-\log\Lambda_\rho(\lambda)=\sup_\lambda[\lambda y-\log\Lambda_\rho(\lambda)]$ (think why). Conclusion : J is the Fenchel-Legendre transform of Λ_ρ.

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 $(X_1,\ldots,X_k)$ conditional of an event involving $L_n=n^{-1}\sum_{i=1}^n\delta_{X_i}$, the empirical measure of the vector $(X_1,\ldots,X_n)$:

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

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 3. 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 $\inf_{B^o} H_\rho = \min_B H_\rho = H_\rho(\hat{\nu})$ 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 $\mathbb{E}[\varphi(X_1)] < e < \sup_{\mathcal{K}} f$. 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 4. 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 5. 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 6. For B of the form $B = \{L_n(f) \in \mathcal{I}\}$ derive the LDP for $(\tilde{L}_i)_{i\geqslant 1}$ 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 $\{\nu \in \Pi(\mathcal{K}^k) \colon \nu(\tilde{f}) \in \mathcal{I}\}$ 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 $f \colon \mathcal{K} \to \mathbb{R}$: $\sum_i f(X_i)/n \simeq e$ (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 $n \simeq 10^{23}$) 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.

Large deviations for processes

Let $(X_n)_{n\geqslant 1}$ 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 $\mathcal{K}=\{f\in L^\infty([0,1]):\|f\|_{L^1}\leqslant 1\}$ 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 $\{\varphi_k\}_{k\geqslant 1}$ of the unit ball of L^1 and letting

$$d(f,g)=\sum_{k\geqslant 1}\frac{|\varphi_k(f)-\varphi_k(g)|}{2^k}$$

Another possible metric is given by $d(f,\ g)\ =\ \sup_{0\leqslant t\leqslant 1}\ |\ \int_0^t\ (f(\theta)\ -\ g(\theta))\mathrm{d}\theta|$. Let $J_p(x)\ =\ H(\mathrm{Ber}(x)|\mathrm{Ber}(p))$. Then we have the following result

Theorem 8. (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 $G_k(f)=(f_{k,1},\ldots,f_{k,k})$ 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 $I_k(x_1,\ldots,x_k)=\sum_{i=1}^k J_p(x_i)$ 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 $I(g)=I'(g)$. □