④ The LDP for Abstract Empirical Measures

We now study LDPs in infinite-dimensional spaces. The general paradigm is to attempt to obtain an LDP by lifting to an infinite dimensional setting finite dimensional results. Sub-additivity is explored as an additional ingredient. The abstract versions of the LDP allow for explicit mixing conditions that are sufficient for the existence of an LDP to be given, even in the $\mathbb{R}^d$ case.

4.1 Cramér's Theorem in Polish spaces

(E, d) Polish, $\mu \in \mathcal{M}_1(E)$. Suppose that $X_1, X_2, \dots$ are i.i.d. random variables on E, each distributed according to the law μ. That is, their joint distribution μ^n is the product measure on $(E^n, \mathcal{B}_E^{\otimes n})$. i.e., e.g. $(X_1, \dots, X_n)$

We shall consider partial averages

$$\hat{S}_n^m = \frac{1}{n-m} \sum_{\ell=m+1}^{n} X_\ell \ ; \quad \hat{S}_n^0 = \hat{S}_n .$$

Note that $\mathcal{B}_{E^n} = \mathcal{B}_E^{\otimes n}$ once the space (E, d) is separable, and we will assume this in the following. Let μ_n denote the law of $\hat{S}_n$ on E.
μ_n is a Borel measure as soon as the convex hull of the support of μ is separable. The following technical assumption

formalises the conditions required.

Assumption (*):

(a) E locally convex, Hausdorff, topological real vector space. $\mathcal{E} \subset E$ is a closed, convex subset of E such that $\mu(\mathcal{E}) = 1$ and $\mathcal{E}$ can be made into a Polish space.

(b) The closed convex hull of each compact $K \subset \mathcal{E}$ is compact.

Theorem 19: Let Assumption (*) hold. Then $(\mu_n)_{n\geq 1}$ satisfies in E (and in $\mathcal{E}$) a weak LDP with rate function Λ^* (Fenchel-Legendre transform of Λ; $\Lambda(\lambda) = \log \int_E e^{\langle \lambda, x\rangle} \mu(dx)$).

Moreover, for every open, convex subset $A \subset \mathcal{E}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x\in A} \Lambda^*(x) \quad .$$

Before we turn to the extended proof we put $E = \mathcal{E} = \mathbb{R}^d$, since it dispenses with $\mathcal{D}_\Lambda = \mathbb{R}^d$ or Λ to be steep.

Corollary 20: The sequence $(\mu_n)_{n\geq 1}$ of the laws of empirical means of $\mathbb{R}^d$-valued i.i.d. random variables satisfies a weak LDP with the convex rate function Λ^*. Moreover, if $0 \in \mathcal{D}_\Lambda^\circ$, then $(\mu_n)_{n\geq 1}$ satisfies the full LDP with the good, convex rate function Λ^*.

Proof: The weak LDP is a specialisation of Theorem 19.

$0 \in \mathcal{D}_\Lambda^\circ \implies \mu_n \subset \mathcal{M}_1(\mathbb{R}^d)$ exponentially tight. $\square$

The proof of the Theorem is quite long and it is based on the following two key lemmas.

Lemma 21: (key lemma 1)
Let Assumption (*)(a) hold true. Then, $(\mu_n)_{n \geq 1}$ satisfies the weak LDP in E with a convex rate function I.

Lemma 22: (key lemma 2)
Let Assumption (*) hold true. Then, for every open, convex subset $A \subset E$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mu_n(A) = -\inf_{x \in A} \{I(x)\},$$

where I is the convex rate function of Lemma 21.

We first complete the proof of Theorem 19 assuming the two key lemmata.

Proof of Theorem 19: In view of both Lemma 21 and Lemma 22 we shall show that $I = \Lambda^*$. By the independence of X_i, it follows that for each $\lambda \in E^*$ and every $t \in \mathbb{R}, n \in \mathbb{Z}_+$,

$$\frac{1}{n}\log \mathbb{E}\left[e^{n\langle t\lambda, \hat{S}_n\rangle}\right] = \log \mathbb{E}\left[e^{t\langle \lambda, X_1\rangle}\right] = \Lambda_\lambda(t)$$

For every $\lambda \in E^*$, the function Λ_λ is the logarithmic moment generating function of the real valued random variable $\langle \lambda, X_1\rangle$. Hence, by Fatou's lemma, it is lower semicontinuous.

We will show that

$$(+) \quad \inf_{x:\langle \lambda, x\rangle - a > 0} \{ I(x) \} \leq \inf_{z > a} \Lambda_\lambda^*(z), \quad \forall a \in \mathbb{R}, \ \forall \lambda \in E^*.$$

If $(+)$ holds we get that

$$\sup_{x\in E} \{\langle \lambda, x\rangle - I(x)\} = \sup_{a\in\mathbb{R}} \sup_{x:\langle\lambda,x\rangle - a>0} \{\langle \lambda, x\rangle - I(x)\}$$

$$\geq \sup_{a\in\mathbb{R}} \{ a - \inf_{x:\langle\lambda,x\rangle - a > 0} I(x)\} \geq \sup_{a\in\mathbb{R}} \{ a - \inf_{z>a}\{\Lambda_\lambda^*(z)\}\}$$

$= \sup_{z\in\mathbb{R}} \{ z - \Lambda_\lambda^*(z)\}$. Applying the duality lemma to $\Lambda_\lambda : \mathbb{R} \to (-\infty, \infty]$ we get that $\Lambda_\lambda(1) = \sup_{z\in\mathbb{R}}\{z - \Lambda_\lambda^*(z)\}$.

Combining this with our previous estimate yields

$$\sup_{x\in E}\{\langle\lambda, x\rangle - I(x)\} \geq \Lambda_\lambda(1) = \Lambda(\lambda).$$

The opposite inequality follows by applying ~~to~~ the following version of Varadhan's lemma: $\varliminf_{n\to\infty} \frac{1}{n}\log \mathbb{E}\left[e^{n\phi(X_n)}\right] \geq \sup_{x\in E}\{\phi(x) - I(x)\}$

where $\phi : E \to \mathbb{R}$ is lower semicontinuous. We apply it now with $\phi = \lambda \in E^*$. Thus

$$\Lambda(\lambda) = \sup_{x \in E} \{ \langle \lambda, x \rangle - I(x) \}$$, and by the duality lemma again, we conclude $I = \Lambda^*$.

We are left to show (+):

Pick $\lambda \in E^*$. If $\lambda = 0$, then $\Lambda_\lambda^*(z) = 0$ for $z = 0$, and $\Lambda_\lambda^*(z) = +\infty$ otherwise. As E is open, convex set with $\mu_n(E) = 1$ for all n we have $\inf_{x \in E} I(x) = 0$ (from Lemma 22). Hence, we now fix $\lambda \in E^*$, $\lambda \neq 0$, and $a \in \mathbb{R}$. The half-space $H_a := \{ x \in E : \langle \lambda, x \rangle - a > 0 \}$ is convex, and therefore by Lemma 22, $-\inf_{x : \langle \lambda, x \rangle - a > 0} \{ I(x) \} =$

$$\lim_{n \to \infty} \frac{1}{n} \mu_n(H_n) \geq \sup_{\delta > 0} \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\overline{H}_{a+\delta})$$

as $\overline{H_{a+\delta}} \subset H_a$ for all $\delta > 0$.

Let $Y_\ell := \langle \lambda, X_\ell \rangle$, and $\hat{Z}_n := \frac{1}{n} \sum_{\ell=1}^{n} Y_\ell = \langle \lambda, \hat{S}_n \rangle$.

$$\hat{S}_n \in \overline{H_y} \iff \hat{Z}_n \in [y, \infty).$$

By our old version of Cramér's theorem (real valued case), $\hat{Z}_n$ satisfies the LDP with rate function Λ_λ^*, and thus

$$\sup_{\delta > 0} \limsup_{n \to \infty} 1/n \log \mu_n(\overline{H}_{a+\delta}) = \sup_{\delta > 0} \{ -\inf_{z \geq a+\delta} \Lambda_\lambda^*(z) \} = -\inf_{z > a} \Lambda_\lambda^*(z),$$

giving (+). ∎

Proof of key lemma 1 (Lemma 21):

The proof is important and a bit long - we split it in several steps. The proof is based on sub-additivity, which is defined as follows.

Definition A function $f : \mathbb{Z}_+ \to [0, \infty]$ is called sub-additive if $f(n+m) \leq f(n) + f(m)$ for all $n, m \in \mathbb{Z}_+$.

Lemma 23: If $f : \mathbb{Z}_+ \to [0,\infty]$ is a sub-additive function such that $f(n) < \infty$ for all $n \geq N$ and some $N < \infty$, then

$$\lim_{n\to\infty} \frac{f(n)}{n} = \inf_{n \geq N} \frac{f(n)}{n} < \infty .$$

Proof: Pick $m \geq N$ and $M_m := \max \{ f(r) : m \leq r \leq 2m \}$

$M_m < \infty$. $\forall\, n \geq m \geq N$, $s := \lfloor n/m \rfloor \geq 1$ and

$r = n - m(s-1) \in \{m, \dots, 2m\}$.

$$\frac{f(n)}{n} \leq \frac{(s-1) f(m)}{n} + \frac{f(r)}{n} \leq \frac{(s-1) f(m)}{n} + \frac{M_m}{n} .$$

$(s-1)/n \to 1/m$ as $n \to \infty$, and therefore

$$\limsup_{n\to\infty} \frac{f(n)}{n} \leq \frac{f(m)}{m} .$$ This holds for all $m \geq N$,

and we conclude the proof by considering the infimum over $m \geq N$.

The key fact to the application of sub-additivity is the following lemma.

__Lemma 24__ : Assume (*) part (a), that is, $\mathcal{E} \subset E$ closed, convex such that $\mu(\mathcal{E}) = 1$, whereas E is locally convex, top. Hausdorff real vector space.

For every convex $A \subset \mathcal{B}_E$, the function

$$f(n) := -\log \mu_n(A) \quad \text{is sub-additive.}$$

Proof: W.l.o.g. $A \subset \mathcal{E}$.

$$\hat{S}_{m+n} = \frac{m}{m+n} \hat{S}_m + \frac{n}{n+m} \underbrace{\hat{S}^m_{m+n}}_{\frac{1}{n}\sum_{l=m+1}^{m+n}} .$$

Therefore, $\hat{S}_{m+n}$ is a convex combination of the independent random variables $\hat{S}_m$ and $\hat{S}^m_{m+n}$. We have (A convex),

$$\{\omega : \hat{S}^m_{m+n}(\omega) \in A\} \cap \{\omega : \hat{S}_m(\omega) \in A\} \subset \{\omega : \hat{S}_{m+n}(\omega) \in A\}.$$

$$\mu^{n+m}(\{\omega : \hat{S}_{m+n}(\omega) \in A\}) = \mu^n(\{\omega : \hat{S}_n(\omega) \in A\}),$$

and thus $\mu_n(A)\mu_m(A) \leq \mu_{n+m}(A)$, and henceforth

$f(n)$ is sub-additive. ∎

We now turn to the proof of key lemma 1:

$A \subset E$ open, convex, and w.l.o.g. $A \subset \mathcal{E}$.

$\mu_n(A) = \mu_n(A \cap \mathcal{E})$.

Either $\mu_n(A) = 0$ for all $n \geq 1$, in which case $\mathcal{L}_A := -\lim_{n\to\infty} 1/n \log \mu_n(A) = +\infty$, or else the limit

$\mathcal{L}_A = -\lim_{n\to\infty} 1/n \log \mu_n(A)$ exists.

We shall now show this existence. We know that $\mu_m(A) > 0$ for some m. Then there exists an $N < \infty$ such that $\mu_n(A) > 0$ for all $n \geq N$.

Since $\mathcal{E}$ is Polish, by Prohorov's theorem, any finite family of probability measures is tight.

$\exists$ compact $K \subset \mathcal{E}$ such that $\mu_r(K) > 0$ $\forall r = 1, \dots, m$.

One can easily show that there exists a point $p_0 \in A$ such that $\mu_m(B_{p_0}) > 0$ for some neighbourhood B_{p_0} of p_0 (for that, assume the contrary and get a contradiction to $\mu_m(A) > 0$).

Define the function $f : [0,1] \times \mathcal{E} \times \mathcal{E}$ by $f(a,p,q) = (1-a)p + aq$, and note that f is continuous, and $f(0, p_0, q) = p_0 \in A$. Therefore there exists $\varepsilon_q > 0$, and two neighbourhoods, W_q of q, and U_q of p_0, such that

$$(1-\varepsilon)U_q + \varepsilon W_q \subset A \qquad 0 \leq \varepsilon \leq \varepsilon_q.$$

We turn this into the following. First cover K by a finite number of these W_q, and let ε^* the minimum of the ε_q values, and let $\mathcal{U}$ denote the finite intersection of the corresponding $\mathcal{U}_q$. $\mathcal{U}$ is neighborhood of p_0, and

$$(1-\varepsilon)\mathcal{U} + \varepsilon K \subset A \qquad \forall\, 0 \leq \varepsilon \leq \varepsilon^* .$$

$\mathcal{U}$ contains a convex neighborhood V of p_0 (by the fact that $\mathcal{E}$ is convex resp. E locally convex);

$$\mu_n(A) \geq \mu_n((1-\varepsilon)V + \varepsilon K) \quad \forall n \geq 1, \forall 0 \leq \varepsilon \leq \varepsilon^*$$

$N = m\lceil 1/\varepsilon^* \rceil + 1 < \infty$; and we write each $n \geq N$ as $n = ms + r$ with $1 \leq r \leq m$. Since

$\hat{S}_n = (1-\frac{r}{n})\hat{S}_{ms} + \frac{r}{n}\hat{S}_n^{ms}$; the last inequality holds for $\varepsilon = r/n \leq \varepsilon^*$; and it follows that

$$\mu_n(A) \geq \mu^n(\{\hat{S}_{ms} \in V, \hat{S}_n^{ms} \in K\}) = \mu^n(\{\hat{S}_{ms} \in V\})\mu^n(\{\hat{S}_n^{ms} \in K\})$$

$$\geq \mu_{ms}(V)\ \mu_r(K) .$$

From our construction, $\mu_m(V) > 0$, and $\mu_{ms}(V) \overset{\text{by Lemma 24}}{\geq} \mu_m(V)^s > 0$.

Thus $\mu_n(A) > 0$ for all $n \geq N$.

Using sub-additivity, Lemma 23 and Lemma 24 we get that the limit $\mathcal{L}_A := -\lim_{n\to\infty} \frac{1}{n}\log \mu_n(A)$ exists.

Define $I(x) := \sup\{\mathcal{L}_A : x \in A, A \in \mathcal{C}^0\}$, where $\mathcal{C}^0$ is a base of the topology of E. With that it follows that μ_n satisfies the weak LDP with rate function I. This is Theorem 4.1.11 in [DZ98]. The lower bound follows immediately as for any open set G and $x \in G$ there exists an $A \in \mathcal{C}^0$ such that $x \in A \subset G$, and therefore,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\mathcal{L}_A \geq -I(x).$$

I is non-negative, and if $I(x) > \alpha$, then $\mathcal{L}_A > \alpha$ for some $A \in \mathcal{C}^0$ such that $x \in A$. Hence, $I(y) \geq \mathcal{L}_A > \alpha$ for all $y \in A$, and thus $\{x : I(x) > \alpha\}$ are open.
The upper bound (weak LDP = upper bound for compact sets) follows with finite cover of a compact $F \subset E$.
We leave the proof of the convexity of the rate function as an exercise. $\square$

__Proof of Key lemma 2 (Lemma 22):__
We shall show $\lim_{n\to\infty} 1/n \log \mu_n(A) = -\inf_{x\in A}\{I(x)\}$.

Key lemma 1 gives that $\mathcal{L}_A = -\lim 1/n \log \mu_n(A)$ exists for any open, convex $A \in \mathcal{C}^0$, and the LDP lower bound $\mathcal{L}_A \leq \inf_{x\in A} I(x)$.

It suffices to show that $\forall A \in \mathcal{C}^0 \; \forall \delta > 0$:

$$\inf_{x \in A} I(x) \leq \mathcal{L}_A + 2\delta .$$

Pick $A \in \mathcal{C}^0$ and $\delta > 0$. W.l.o.g. $\mathcal{L}_A < \infty$. There exists $N \in \mathbb{Z}_+$ such that

$$-\frac{1}{N} \log \mu_N(A \cap \mathcal{E}) = -\frac{1}{N} \log \mu_N(A) \leq \mathcal{L}_A + \delta < \infty.$$

Turn the relatively open set $A \cap \mathcal{E}$ into a Polish space. Hence, (prob. ~~m.~~ tight) there exists compact set $C \subset A \cap \mathcal{E}$ such that

$$-\frac{1}{N} \log \mu_N(C) \leq -\frac{1}{N} \log \mu_N(A) + \delta \leq \mathcal{L}_A + 2\delta .$$

Cover C by a finite number of neighbourhoods B_x ($\overline{B_x} \subset A$).
By assumption $\overline{co}(C)$ is compact.

$$\overline{co}(C) \cap \left(\bigcup_{i=1}^{k} \overline{B_i} \right) =: \tilde{C} \text{ . Hence,}$$

$K := co(\tilde{C}) \supset C$ and is a closed subset of both A and $\overline{co}(C)$. K is convex and compact,

$$-\frac{1}{N} \log \mu_N(K) \leq -1/N \log \mu_N(C) \leq \mathcal{L}_A + 2\delta$$

K convex $\Rightarrow$ $g(n) := -\log \mu_n(K)$ sub-additive.

$$-\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K) \leq \liminf_{n\to\infty} \left[-\frac{1}{nN} \log \mu_{nN}(K) \right]$$

$$\leq -\frac{1}{N} \log \mu_N(K) \leq \mathcal{L}_A + 2\delta .$$

The weak LDP upper bound for $K \subset A$ yields

$$\inf_{x \in A} I(x) \leq \inf_{x \in K} I(x) \leq -\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(K) .$$

■

4.2 Sanov's theorem

In this section we generalise Sanov's theorem which we have proven for some finite alphabet E.

Here, we take $E = \mathcal{M}_1(\Sigma)$, which is a Polish space once equipped with the Lévy metric. Note that $E \subset M(\Sigma)$ is a closed convex subset of $M(\Sigma)$ – the vector space of all finite signed measures on Σ.

We shall need the following general fact about the space $M(\Sigma)$ which it will be useful to have at our disposal.

Lemma 25: The duality relation

$$(\phi, \alpha) \in C_b(\Sigma) \times M(\Sigma) \longmapsto \int_\Sigma \phi(x)\, \alpha(dx)$$

determines a representation of $M(\Sigma)^*$ as $C_b(\Sigma)$.

Proof: Clearly the mapping determines a unique element of $M(\Sigma)^*$. Conversely, let ~~[illegible]~~ $\lambda \in M(\Sigma)^*$ be given and define $\phi(\sigma) = \lambda(\delta_\sigma)$, $\sigma \in \Sigma$. Clearly, ϕ is continuous. From the way we have defined the topology on $M(\Sigma)$ it is clear that we find a finite set $\{\psi_m\}_{m=1,\dots,M}$ of $\psi_m \in C_b(\Sigma)$ such that $|\lambda(\alpha)| \le \sum_{m=1}^{M} |\int_\Sigma \psi_m \, d\alpha|$, $\alpha \in M(\Sigma)$.

Thus ϕ is bounded.

$\lambda(\alpha) = \int_\Sigma \phi \, d\alpha$ if α is a linear combination of point masses. Such α's are dense in $M(\Sigma)$, it follows that this equation holds for all $\alpha \in M(\Sigma)$. ■

Large deviations: Let $Q \in \mathcal{M}_1(\mathcal{M}_1(\Sigma))$ be given and define $Q_n \in \mathcal{M}_1(\mathcal{M}_1(\Sigma))$ to be the distribution of

$$\nu^{(n)} = (\nu_1, \dots, \nu_n) \in \mathcal{M}_1(\Sigma)^n \longmapsto \frac{1}{n}\sum_{k=1}^{n} \nu_k \in \mathcal{M}_1(\Sigma)$$

under $Q^n \in \mathcal{M}_1(\mathcal{M}_1(\Sigma)^n)$. That is, with $L_n = \frac{1}{n}\sum_{k=1}^{n} \nu_k$,

$Q_n = Q^n \circ L_n^{-1}$. By WLLN combined with the second countability of the weak topology on $\mathcal{M}_1(\Sigma)$, one can easily check that $Q_n \Rightarrow \delta_{\mu_Q}$, where $\mu_Q \in \mathcal{M}_1(\Sigma)$ is defined by

$$\mu_Q(A) = \int_{M_1(\Sigma)} \nu(A) \, Q(d\nu) \quad , A \subset \mathcal{B}(\Sigma).$$

Thus, it is reasonable to study large deviations of $\{Q_n : n \ge 1\}$.

The rate function will be

$$I_Q(\nu) = \Lambda_Q^*(\nu) = \sup_{\phi \in C_b(\Sigma)} \left\{ \int \phi \, d\nu - \Lambda_Q(\phi) \right\} \quad \text{for}$$

$\nu \in \mathcal{M}_1(\Sigma)$, where $\Lambda_Q(\phi) = \log\left(\int_{\mathcal{M}_1(\Sigma)} \exp\left(\langle \phi, \nu \rangle\right) Q(d\nu)\right)$.

We have to check exponential tightness late. Before we discuss a particular case for Q and provide a useful representation for the rate function.

Let Q be the distribution $\tilde{\mu}_\epsilon^{\mathcal{M}_1(\Sigma)}$ of $\sigma \in \Sigma \mapsto \delta_\sigma$ under some $\mu \in \mathcal{M}_1(\Sigma)$. In this case, Q_n is the distribution $\tilde{\mu}_n$ of the empirical distribution (measure) $L_n(\sigma) = \frac{1}{n}\sum_{k=1}^n \delta_{\sigma_k}$; $\sigma \in \Sigma^n$, under μ^n and the measure μ_Q coincides with μ.

$\Lambda_{\tilde{\mu}}(\phi) = \log \int_\Sigma e^\phi d\mu$. Rate function for $\nu \in \mathcal{M}_1(\Sigma)$

$$I_{\tilde{\mu}}(\nu) = \Lambda^*_{\tilde{\mu}}(\nu) = \sup\left\{ \int_\Sigma \phi \, d\nu - \log\left(\int e^\phi d\mu\right) : \phi \in C_b(\Sigma)\right\}$$

The following lemma provide a more tractable expression for $I_{\tilde{\mu}}$.

Lemma 26: For $\nu \in \mathcal{M}_1(\Sigma)$, define

$$H(\nu|\mu) = \begin{cases} \int_\Sigma f \log f \, d\mu & \text{if } \nu \ll \mu \text{ and } f = \frac{d\nu}{d\mu}, \\ \infty & \text{otherwise.} \end{cases}$$

Then $I_{\tilde{\mu}} = H(\cdot|\mu)$.

Proof: Step 1. $\nu \ll \mu$ and define $\nu_\theta = \theta\mu + (1-\theta)\nu$ for $\theta \in [0,1]$, then $H(\nu|\mu) = \lim_{\theta \downarrow 0} H(\nu_\theta|\mu)$.

To see this, set $f = \frac{d\nu}{d\mu}$ and $f_\theta = \theta + (1-\theta) f$.

Since $x \in [0,\infty) \longmapsto x \log x$ is convex, Jensen's inequality gives that

$$H(\nu_\theta | \mu) = \int_\Sigma f_\theta \log f_\theta \, d\mu \leq (1-\theta) \int_\Sigma f \log f \, d\mu = (1-\theta) H(\nu|\mu)$$

At the same time, $x \in [0,\infty) \longmapsto \log x$ being non-decreasing and concave, $\log f_\theta \geq (\log \theta) \vee ((1-\theta)\log f)$, and thus

$$H(\nu_\theta | \mu) = \theta \int_\Sigma \log f_\theta \, d\mu + (1-\theta) \int_\Sigma f \log f_\theta \, d\mu \geq \theta \log \theta + (1-\theta)^2 H(\nu|\mu)$$

Both inequalities give a proof of step 1.

<u>Step 2:</u> If $\nu \ll \mu$, then $I_{\tilde{\mu}}(\nu) \leq H(\nu|\mu)$.

Using the lower semicontinuity of $\nu \longmapsto \Lambda^*_{\tilde{\mu}}(\nu)$ and step 1, we may assume that $f = \frac{d\nu}{d\mu} \geq \theta$ for some $\theta \in (0,1)$.

By Jensen's inequality, we then have

$$\exp\left(\int_\Sigma \phi(x) \nu(dx) - H(\nu|\mu) \right) = \exp\left(\int_\Sigma (\phi - \log f) \, d\nu \right)$$

$$\leq \int_\Sigma \frac{\exp(\phi)}{f} \, d\nu = \int_\Sigma \exp(\phi) \, d\mu .$$

Hence, $I_{\tilde{\mu}}(\nu) \leq H(\nu|\mu)$ whenever $\nu \ll \mu$.

<u>Step 3:</u>

We are left to show that if $I_{\tilde{\mu}}(\nu) < \infty$, then $d\nu = f d\mu$ and $I_{\tilde{\mu}}(\nu) \geq \int f \log f \, d\mu$. (*)

$\nu \in \mathcal{M}_1(\Sigma)$ with $I_{\tilde{\mu}}(\nu) < \infty$ implies that

$$(**) \int_\Sigma \phi \, d\nu - \log \int_\Sigma e^\phi d\mu \leq I_{\tilde{\mu}}(\nu) < \infty$$ for every

bounded continuous ϕ.

The class of ϕ's under which $(**)$ holds is closed under bounded pointwise convergence, $(**)$ holds for every bounded $\mathcal{B}(\Sigma)$-measurable function ϕ. We are now in the position to show $\nu \ll \mu$. Suppose ~~thus~~ that $A \in \mathcal{B}(\Sigma)$ exists with $\mu(A) = 0$.

By $(**)$, with $\phi = r \mathbb{1}_A$; $r\nu(A) \leq I_{\tilde{\mu}}(\nu)$, $r > 0$; and therefore $\nu(A) = 0$. Henceforth $\nu \ll \mu$, and we set $f = \frac{d\nu}{d\mu}$.

If f is uniformly positive and uniformly bounded, then $(*)$ is consequence of $(**)$ with $\phi = \log f$.

If f is uniformly positive but not necessarily bounded, set $f_n = f \wedge n$ and use $(**)$ with Fatou's Lemma to justify

$$\int_\Sigma f \log f \, d\mu = \int \log f \, d\nu \leq \varliminf_{n\to\infty} \int \log f_n \, d\nu \leq I_{\tilde{\mu}}(\nu)$$

$$+ \varlimsup_{n\to\infty} \log \left(\int_\Sigma f \wedge n \, d\mu \right) = I_{\tilde{\mu}}(\nu).$$

For the general case, proceed as in step 1: $\int f_\theta \log f_\theta \, d\mu \leq I_{\tilde{\mu}}(\nu_\theta)$ as long as $\theta \in (0,1)$. Now as $\theta \mapsto I_{\tilde{\mu}}(\nu_\theta)$ is bounded, lower semicontinuous, and convex on $[0,1]$, it is continuous on $[0,1]$.

This now completes the proof ∎

$H(\nu|\mu)$ is called the relative entropy of ν with respect to μ. We have now studied ~~st~~ all parts of the proof of the following theorem.

Theorem 27: Sanov's theorem

$\mu \in \mathcal{M}_1(\Sigma)$ and let $\tilde{\mu}_n \in \mathcal{M}_1(\mathcal{M}_1(\Sigma))$ be the distribution under μ^n of L_n. Then $H(\cdot|\mu)$ is a good, convex rate function on $\mathcal{M}_1(\Sigma)$ and $(\tilde{\mu}_n)_{n\geq 1}$ satisfies the large deviation principle on $\mathcal{M}_1(\Sigma)$ with rate function ~~E~~ $H(\cdot|\mu)$.

Remark: (1) There exist a version of Sanov's theorem where the topology on $\mathcal{M}_1(\Sigma)$ is not the weak one but instead the strong or τ-topology is introduced and used. (test functions are bounded + measurable) See [DZ98] for details.

(2) Let $H(\nu|\mu) < \infty$ and $G \in \mathcal{B}(\mathcal{M}_1(\Sigma))$ be a neighbourhood (strong topology) of ν. Set $f = \frac{d\nu}{d\mu}$, define

$$F_n(\sigma) = \prod_{m=1}^{n} f(\sigma_m) \quad , \sigma \in \Sigma^n, \text{ and let}$$

$$A_n = \{\sigma \in \Sigma^n : L_n(\sigma) \in G \text{ and } F_n(\sigma) > 0\}.$$

Using SLLN, check that $\nu^n(A_n) \to 1$ as $n \to \infty$.

Use Jensen's inequality and $x\log x \geq -e^{-1}$, $x \in [0,\infty)$, to verify the following steps

$$\log \tilde{\mu}_n(G) \geq \log \int_{A_n} \frac{1}{F_n(\sigma)} \nu^n(d\sigma)$$

Jensen

$$\geq \log(\nu^n(A_n)) - \frac{1}{\nu^n(A_n)} \int_{A_n} \log(F_n(\sigma))\, \nu^n(d\sigma)$$

$$\geq \log(\nu^n(A_n)) - \frac{1}{e\nu^n(A_n)} - \frac{1}{\nu^n(A_n)} \int_{\Sigma^n} \log F_n(\sigma)\, \nu^n(d\sigma)$$

$$= \log(\nu^n(A_n)) - \frac{1}{e\nu^n(A_n)} - \frac{n\, H(\nu|\mu)}{\nu^n(A_n)}$$

as long as $\nu^n(A_n) > 0$. Combining this with $\nu^n(A_n) \to 1$ as $n \to \infty$, we arrive at the lower bound.