Last time

$$f_n \xrightarrow[\text{in } L^2]{\text{weakly}} f \Rightarrow f_n \longrightarrow f \text{ a.s.}$$

subsequence.

Large deviations for empirical measures

$\mathcal{X}$ nice space (Polish)

X_i iid on $\mathcal{X}$ $\quad X_i \sim \mu$

empirical measures

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \quad \text{random measure}$$

LLN: $\mu_n \xrightarrow{\text{weakly}} \mu$ a.s.

Question $\nu \in \mathcal{M}_1(\mathcal{X}) = \{\text{proba. meas. on } \mathcal{X}\}$

How quickly does $\mathbb{P}(\mu_n \approx \nu)$ decay?

Remark on topology

different metrics, topologies on $\mathcal{M}_1(\mathcal{X})$

- total variation distance

$$d_{TV}(\mu, \nu) = 2 \sup_{A \subset \mathcal{X}} (\mu(A) - \nu(A))$$

if a.c. wrt the same reference then, L^1 - distance of densities.

- topology of weak convergence

$$\nu_n \xrightarrow{\text{weakly}} \nu \text{ iff } \forall C_b(\mathcal{X})$$

$$\int f \nu_n \longrightarrow \int f \nu$$

$\mathcal{M}_1(\mathcal{X})$ convex subset of $(C_b(\mathcal{X}))^*$

this is the weak-$*$-topology.
There is a metric that generates this topology. $\mathcal{M}_1(\mathcal{X})$ with this metric is polish.

2 Guessing the rate function

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\delta_{X_i}}_{Y_i}$$

Applying Cramér (without justification)

expect a LDP with rate function

$$I(\nu) = \sup_{f \in C_b} \left(\int f \, d\nu - \log \mathbb{E}\left[e^{\int f \, dY}\right] \right)$$

in $\mathbb{R}$ $\quad I(x) = \sup_{t \in \mathbb{R}} \left(xt - \log \mathbb{E}\left[e^{tX}\right] \right)$

$$= \sup_{f \in C_b} \left(\int f \, d\nu - \log \int e^{f(x)} \mu(dx)\right.$$

Lemma Fix μ, ν

$$H(\nu \| \mu) = \sup_{f \in C_b(\mathcal{X})} \left(\int f \, d\nu - \log \int e^f \, d\mu \right)$$

$$= \sup_{f \in \text{bounded meas.}} \left(\int f \, d\nu - \log \int e^f \, d\mu \right)$$

proof. $\boxed{H(\nu | \mu) \geq \text{RHS}}$

fix f. Define $\mu_f = \frac{e^f}{\int e^f d\mu} \, d\mu$

$$H(\nu|\mu) = \int \log \frac{d\nu}{d\mu} \, d\nu$$

$$= \int \log\left(\frac{d\nu}{d\mu_f} \frac{d\mu_f}{d\mu} \right) d\nu$$

$$= \underbrace{\int \log \frac{d\nu}{d\mu_f} \, d\nu}_{} + \int \log \frac{d\mu_f}{d\mu} \, d\nu$$

$$= \underbrace{H(\nu \| \mu_f)}_{\geq 0} + \underbrace{\int f \, d\nu - \int \left(\log \int e^f d\mu\right) d\nu}_{= \int f d\nu - \log \int e^f d\mu}$$

$\boxed{\sup \text{RHS} \geq H(\nu|\mu)}$

If $\frac{d\nu}{d\mu} = \varphi$ is continuous & bdd from above and below

set $f = \log \varphi$

$$H(\nu \| \mu) = \int \log \varphi \, d\nu$$

$$= \int f \, d\nu - \underbrace{\log \int e^{\overbrace{f}^{\varphi}} d\mu}_{=0}$$

In general, by Approximation!

Thm. (Sanov '61)

$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ X_i iid in Polish space $\mathcal{X}$. Then μ_n satisfy a LDP with rate function

$$I(\nu) = H(\nu \| \mu)$$

where $\mu = \mathcal{L}aw(X_i)$.

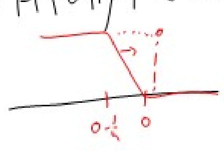
(this LDP is w.r.t. the topology of weak convergence.

Remarks:

1.) Variational formula

$$H(\nu \| \mu) = \sup_{f \in C_b} \left(\int f \, d\nu - \log \int e^f \, d\mu \right)$$

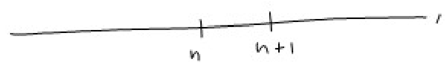
implies that $\nu \mapsto H(\nu \| \mu)$ is l.s.c.



② Def. $\mathbb{P}_n$ satisfies a <u>weak</u> LDP

- same condition for lower bound.
- upper bound only for compact sets.
- Sublevelsets of rate are closed.

Example $\mathbb{P}_n$ = uniform distribution on $[n, n+1]$



$\mathbb{P}_n$ satisfy w LDP with rate $\equiv +\infty$

Exo Weak LDP + Exponential tightness
$\implies$ LDP

Exo Proof that $\nu \longmapsto H(\nu \| \mu)$
has compact sublevelsets.

④ Typically rate fct. is almost surely infinite.

proof of Sanov
① Lower bound on open sets.
② Upper bound on compact sets.
③ exponential tightness

① Fix $\mathcal{O}$ open in $\mathcal{M}_1(\mathcal{X})$
$\nu \in \mathcal{O}$ arbitrary.

NTS

$$\varliminf \frac{1}{n} \log \mathbb{P}\left(\mu_n \in \mathcal{O}\right) \geq -\left(H(\nu|\mu) + \varepsilon\right)$$

Lets fix $\hat{X}_i$ iid distributed according to ν.

$\frac{d\nu}{d\mu} = \varphi$ Let's assume φ bdd from above and below.

$\Rightarrow \frac{d\mu}{d\nu} = \varphi^{-1}$

$$\mathbb{P}\left[\mu_n \in \mathcal{O}\right] = \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in \mathcal{O}\right]$$

$$= \mathbb{E}\left[\mathbb{1}_{\left\{\frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{X}_i} \in \mathcal{O}\right\}} \prod_{i=1}^{n} \varphi^{-1}(\hat{X}_i)\right]$$

$$\geq \mathbb{E}\left[\mathbb{1}_{\{\dots\}} \prod_{i=1}^{n} \varphi^{-1}(\hat{X}_i)\ \mathbb{1}_{\left\{\prod_{i=1}^{n} \varphi^{-1}(\hat{X}_i) \geq \exp(-n(H(\nu\|\mu)+\varepsilon))\right\}}\right]$$

$$\geq \exp\left(-n\left(H(\nu\|\mu)+\varepsilon\right)\right) \mathbb{E}\left[\mathbb{1}_{\{\dots\}} \mathbb{1}_{\{\dots\}}\right]$$

We need to show

$$\underbrace{\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{X}_i} \in \mathcal{O}\right.}_{\longrightarrow 1 \text{ by LLN}} ; \prod_{i=1}^{n} \varphi^{-1}(\hat{X}_i) \geq \exp\left(-n\left(H(\nu\|\mu)+\varepsilon\right)\right)$$

does not decay.

$$\mathbb{P}\left(\cdots \right) = \mathbb{P}\left(\prod \varphi(X_i) \leq \exp(+nH \ldots)\right.$$

$$= \mathbb{P}\left(\frac{1}{n}\sum \log \varphi(X_i) \leq H(\nu\|\mu)+\varepsilon\right]$$

by LLN $\longrightarrow 1$ $\square$

Upper bound on compact sets

$\mathcal{C} \subset \mathcal{M}_1(\mathcal{X})$ compact

We can assume that $\exists\, \alpha > 0$

$$\inf_{\nu\in\mathcal{C}} H(\nu\|\mu) > a > 0.$$

NTS: $\overline{\lim}\, \frac{1}{n}\log \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in \mathcal{C}\right] \leq -\alpha$

Recall variational formula for entropy:

$$H(\nu\|\mu) = \sup_{f\in C_b} \left(\int f\,d\nu - \log\int e^{f}\,d\mu\right)$$

Fix $\nu \in \mathcal{C}$. $H(\nu\|\mu) > \alpha$

$$\Rightarrow \exists\, f_\nu : \int f_\nu\, d\nu - \log\int e^{f_\nu}\,d\mu > \alpha.$$

Such a function f_ν exists for every $\nu\in\mathcal{C}$.

$$\mathcal{O}_\nu = \left\{ \varsigma : \int f_\nu\, d\varsigma - \log\int e^{f_\nu}\,d\mu > \alpha\right\}$$

open set containing ν.

$\Rightarrow \exists\, \nu_1, \ldots, \nu_N$ such that

$$\mathcal{C} \subseteq \bigcup_{i=1}^{N} \mathcal{O}_{\nu_i}$$

Fix ν_i.

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in \mathcal{O}_{\nu_i}\right]$$

$$= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} f_\nu(X_i) - \log\int e^{f_\nu}\,d\mu > \alpha\right]$$

$$= \mathbb{P}\left[\exp\left(\sum_{i=1}^{n} f_\nu(X_i)\right) > \exp\left\{n\left(\log\int e^{f_\nu}\,d\mu + \alpha\right)\right\}\right]$$

$$\overset{\text{Chebyshev}}{\leq} \boxed{\exp(-na)}\, \underbrace{\cancel{\exp\left(-n\log\int e^{f_\nu}\,d\mu\right)}}_{=\left(\int e^{f_\nu}\,d\mu\right)^{-n}} \underbrace{\mathbb{E}\left[\cancel{\exp\left(\sum f_\nu(X_i)\right)}\right]}_{\left(\int e^{f_\nu}\,d\mu\right)^{n}} \quad \square$$

③ Exponential tightness

NTS: $\forall r > 0 \ \exists K_r \subset \mathcal{M}_1(\mathcal{X})$ compact

s.t. $\limsup \frac{1}{n} \log \mathbb{P}\left(\mu_n \notin K_r\right) \leq -r$.

By Prokhorov $K \subset \mathcal{M}_1(\mathcal{X})$ relatively compact $\iff \forall \varepsilon > 0 \ \exists \mathcal{C} \subset \mathcal{X}$ compact s.t. $\forall \nu \in K$, $\nu(\mathcal{C}^c) < \varepsilon$

Fix $0 < b_k < \varepsilon_k$ positive sequences $\varepsilon_k \downarrow 0$

$\mathcal{C}_k$ compact in $\mathcal{X}$ s.t. $\mu(\mathcal{C}_k^c) = b_k$.

$$A_{k,\varepsilon_k} = \left\{ \nu \in \mathcal{M}_1(\mathcal{X}) : \nu(\mathcal{C}_k^c) \leq \varepsilon_k \right\}$$

$K = \bigcap_k A_{k,\varepsilon_k}$ compact.

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \notin K\right)$$

$$= \mathbb{P}\left(\frac{1}{n}\sum \delta_{X_i} \notin A_{k,\varepsilon_k} \text{ for at least one } k\right)$$

$$\leq \sum_{k \geq 1} \underbrace{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in A_{k,\varepsilon_k}\right)}_{= \mathbb{P}\left(\frac{1}{n}\sum \mathbb{1}_{A_k}(X_i) \geq \varepsilon_k\right)}$$

This is LD for coin tossing. The rate function can be made arbitrarily large by choosing the right ratio $\frac{b_k}{\varepsilon_k}$.

$\square$

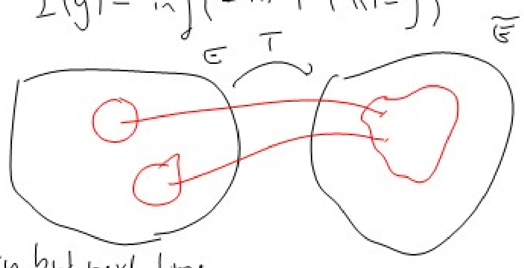
Comment: Sanov $\Rightarrow$ Cramér (at least for bdd RV)

proof is by contraction principle:

Thm: (E, d), $(\tilde{E}, \tilde{d})$ nice metric spaces. μ_n satisfy LDP on E with rate fct. I. If $T: E \to \tilde{E}$ continuous then $T_\# \mu_n$ satisfy LDP with rate

$$I'(y) = \inf\{I(x) \mid T(x) = y\}$$

idea



proof easy but next time.

Assume that X_i (bdd by R) real iid r.v.

Sanov implies that

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \quad \text{satisfy LDP}$$

with rate $H(\nu|\mu)$.

$$T : \mathcal{M}(\mathbb{R}) \longrightarrow \mathbb{R}$$

$$T\mu = \int_{-R}^{R} x \, d\mu(x)$$

Contraction gives that

$$T\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{satisfies LDP with rate}$$

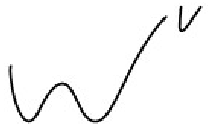
$$I(y) = \inf_{\nu : \int_{-R}^{R} x\, d\nu(x) = y} H(\nu|\mu) = \text{Cramér's rate fct.}$$

$\square$

Possible relation LD (Sanov) $\longleftrightarrow$ PDE.

Adams, Peletier, Dirr, Zimmer ('11; '13)
(Warwick) (Bath)

PDE: Gradient flow:



finite dim. $\dot{x} = -\nabla V(x)$

Riemannian Structure.

Energy

Heat eq: $\frac{du}{dt} = \Delta u$

· Felix Otto + Jordan + Kinderlehrer 98' 01'

Heat eq. gradient flow with respect to

$E = \int \rho \log \rho$; metric given by

Wasserstein.

$\sum \delta_{X_t^i} \longrightarrow$ sol. of heat eq.

interesting topic for essay!