

Large Deviations

Question: (Coming from Insurance Maths
+ Testing of Statistical Hypothesis)

$X_1, X_2, \dots \stackrel{iid}{\sim} \mu$ on $\mathbb{R}$
(Law)

• $X_1, X_2, \dots, X_m$ are ind $\forall m \in \mathbb{N}$

• $P[X_i \in B] = \mu(B) \quad \forall i=1,2,\dots$
 $\forall B \subseteq \mathbb{R}$
(Borel)

Assume: X_i 's have finite mean m

Take $a > m$ $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

$p_n = P[\bar{X}_n > a] \rightarrow 0$ as $n \rightarrow \infty$
 $\text{ss } m$ (WLLN)

Qn: How fast does $p_n \rightarrow 0$?

Insurance $\rightarrow$ Cramer ✓

Testing $\rightarrow$ Chernoff and Bahadur

Bahadur and Rao (1960)

Exc: $X_1, X_2, \dots \stackrel{iid}{\sim} N(0,1)$. $m=0$

Take $a > 0$. Show that

$$p_n = P[X_n > a] \sim \underbrace{\frac{1}{\sqrt{2\pi}} \frac{1}{a\sqrt{n}}}_{\text{polynomial correction}} \underbrace{e^{-a^2 n/2}}_{\text{Main Exponential contribution}}$$

as $n \rightarrow \infty$.

[Hint: Use Mill's ratio for $N(0,1)$]

$$\begin{aligned} \text{Exc} \Rightarrow \frac{\log p_n}{n} &= \frac{\text{const}}{n} + \frac{\log \frac{1}{a\sqrt{n}}}{n} - \frac{a^2 n/2}{n} \\ &\rightarrow -\frac{a^2}{2}. \end{aligned}$$

$$P_n \approx e^{-a^2 n / 2} = e^{-\frac{a^2}{2} n}$$

Qn: $X_1, X_2, \dots$ iid μ on $\mathbb{R}$
(Law)

$E(X_1) = m \in (-\infty, \infty)$. Take $a > m$.

Compute $\lim_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n > a]$.

LD Upper Bound

Suppose, for now, X_i is ≥ 0

$$P[\bar{X}_n > a] \leq \frac{E(\bar{X}_n)}{a} = \frac{m}{a}$$

(Markov Ineq)

$\rightarrow [0, \infty]$ -valued function

$$M_\mu(\theta) = M(\theta) = E(e^{\theta X_1}) = \int_{\mathbb{R}} e^{\theta x} d\mu(x)$$

(MGF of μ or X_1) Assume finite $\forall \theta \in \mathbb{R}$ $\theta \in \mathbb{R}$

$\downarrow$ $[0, \infty)$ -valued f'n.

$M(0) = 1$. For all $\theta > 0$,

$$p_n = P[\bar{X}_n > a]$$

$$= P\left[\theta \sum_{i=1}^n X_i > \theta n a\right]$$

$$= P\left[e^{\theta \sum_{i=1}^n X_i} > e^{n\theta a}\right]$$

Markov's
Ineq

$$\frac{E\left(e^{\theta \sum_{i=1}^n X_i}\right)}{e^{n\theta a}} \stackrel{\text{iid}}{=} \frac{(M(\theta))^n}{e^{n\theta a}}$$

Define $\Lambda_{\mu}(\theta) = \Lambda(\theta) = \log M(\theta)$, $\theta \in \mathbb{R}$

$\uparrow$
 $(-\infty, \infty)$ -valued
"
 $\mathbb{R} \cup \{\infty\}$

$\Lambda(0) = \log M(0) = 0$
 $M(\theta) = e^{\Lambda(\theta)}$
 $\forall \theta \in \mathbb{R}$

$\forall \theta > 0$, we have:

$$p_n = P[\bar{X}_n > a] \leq \frac{e^{n\Lambda(\theta)}}{e^{n\theta a}}$$

$$= e^{-n(\theta a - \Lambda(\theta))}$$

$$\forall \theta \geq 0$$

$$\Rightarrow \frac{1}{n} \leq \inf_{\theta \geq 0} e^{-n(\theta a - \Lambda(\theta))}$$

$$= \exp \left\{ -n \left[\sup_{\theta \geq 0} (\theta a - \Lambda(\theta)) \right] \right\}$$

$$\stackrel{(?)}{=} \exp \left\{ -n \left[\sup_{\theta \in \mathbb{R}} (\theta a - \Lambda(\theta)) \right] \right\}$$

$$= e^{-n \Lambda^*(a)} \quad [\because a > m]$$

Defn: $\Lambda^*(a) = \sup_{\theta \in \mathbb{R}} [\theta a - \Lambda(\theta)]$

↑
Fenchel - Legendre
transform / convex
conjugate of $\Lambda(\theta)$

$$\Lambda(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\mu(x), \quad \theta \in \mathbb{R}.$$

Note: (1) $\Lambda^*(a)$

$$= \sup \{ \theta a - \Lambda(\theta) : \Lambda(\theta) < \infty \}$$

$$(2) \quad \Lambda^*(a) = \sup_{\theta \in \mathbb{R}} [\theta a - \Lambda(\theta)]$$

$\uparrow$
 $[0, \infty]$ -valued $\geq \theta a - \Lambda(\theta) = 0$

$$\Lambda: \mathbb{R} \rightarrow (-\infty, \infty]$$

$$\Lambda^*: \mathbb{R} \rightarrow [0, \infty]$$

Properties of Λ and Λ^*

① Λ is a convex $(-\infty, \infty]$ -valued function. (Holder's inequality)

② Λ^* is a convex $[0, \infty]$ -valued function.

Remark: Λ^* is lsc, i.e.,

$$\forall u \in [0, \infty), \quad (\Lambda^* > u)$$

$$= \{a : \Lambda^*(a) > u\}$$

is open.

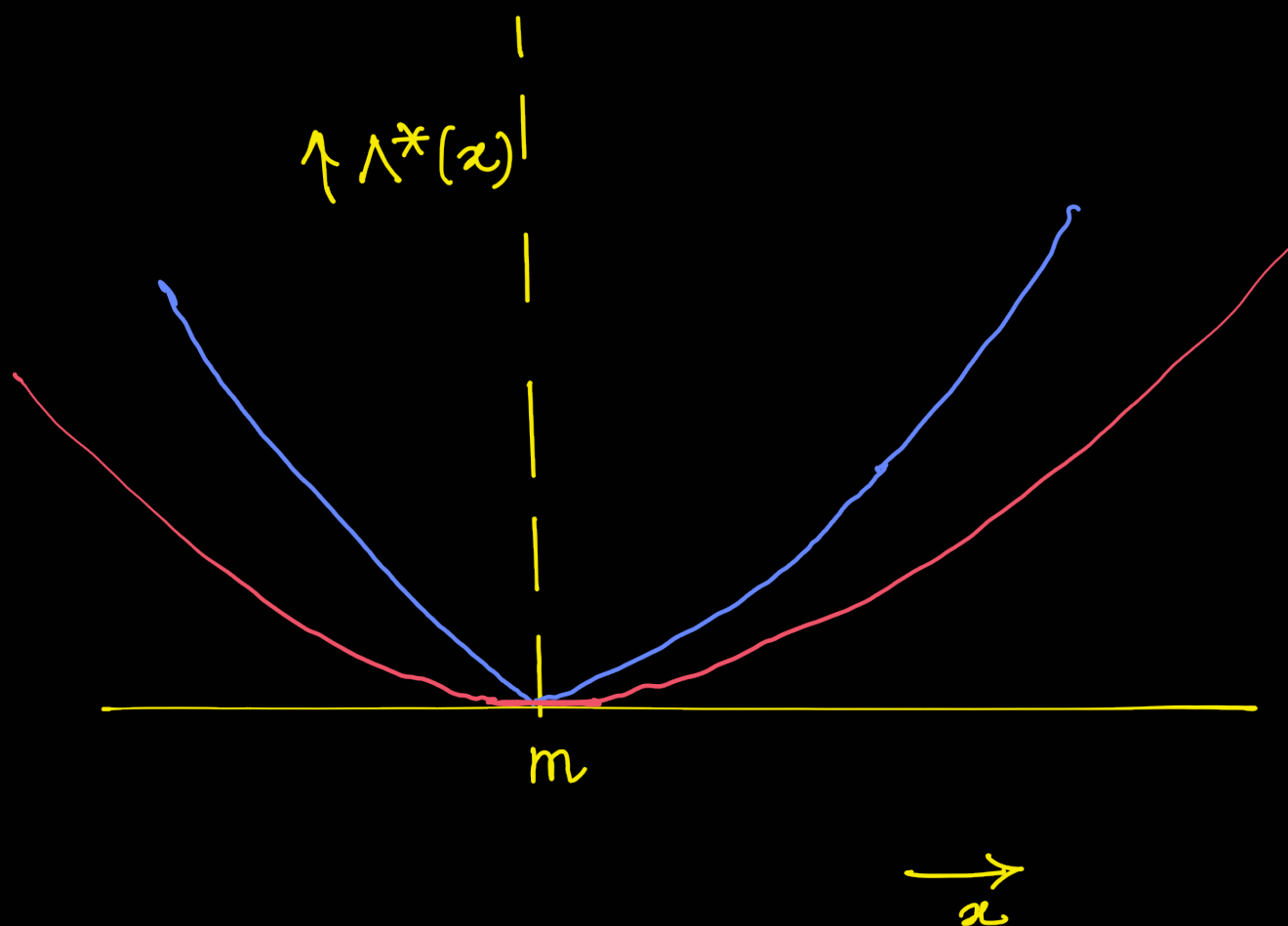
③ X_i 's have finite mean m
 $\Rightarrow \Lambda^*(m) = 0$ +

$$(i) \quad \forall \underline{x \geq m}, \quad \Lambda^*(x) = \sup_{\theta \geq 0} [\theta x - \Lambda(\theta)]$$

In particular, Λ^* is nondecreasing on $[m, \infty)$.

$$(ii) \quad \forall y \leq m, \quad \Lambda^*(y) = \sup_{\theta \leq 0} [\theta y - \Lambda(\theta)]$$

In particular, Λ^* is nonincreasing on $(-\infty, m]$.



Think of $\Lambda^*(x)$ as a measure of how close

x is to m .

Proof of (3): $\forall \theta \in \mathbb{R}$

$$\Lambda(\theta) = \log \int_{\mathbb{R}} e^{\theta x} d\mu(x)$$

$$\geq \int_{\mathbb{R}} \underbrace{\log e^{\theta x}}_{\theta x} d\mu(x) \quad \left[\because \log(\cdot) \text{ is strictly concave} \right]$$

$$= \theta \int_{\mathbb{R}} x d\mu(x) = \theta m$$

$$\Rightarrow \forall \theta \in \mathbb{R}, \quad \theta m - \Lambda(\theta) \leq 0$$

with equality iff $\theta = 0$. Therefore

$$\Lambda^*(m) = \sup_{\theta \in \mathbb{R}} [\theta m - \Lambda(\theta)] = 0.$$

(i) $\forall x \geq m$, we have

$$\left. \begin{array}{l} \theta < 0 \Rightarrow \theta x - \Lambda(\theta) \leq \theta m - \Lambda(\theta) \leq 0 \\ \theta = 0 \Rightarrow \theta x - \Lambda(\theta) = 0x - \Lambda(0) = 0 \end{array} \right\}$$

$$\Rightarrow \forall x \geq m \quad \Lambda^*(x) = \sup_{\theta \in \mathbb{R}} [\theta x - \Lambda(\theta)]$$

$$= \sup_{\theta \geq 0} [\theta x - \Lambda(\theta)]$$

Now take $y \geq x \geq m$. Then $\forall \theta \geq 0$,

$$\theta y - \Lambda(\theta) \geq \theta x - \Lambda(\theta)$$

$$\Rightarrow \sup_{\theta \geq 0} [\theta y - \Lambda(\theta)] \geq \sup_{\theta \geq 0} [\theta x - \Lambda(\theta)]$$

$$\Rightarrow \Lambda^*(y) \geq \Lambda^*(x)$$

$\Rightarrow \Lambda^*$ is nondecreasing on $[m, \infty)$.

(ii) $\rightarrow$ Exc

Go back to LDUB

We have proved that

$$p_n = P[\bar{X}_n > a] \leq e^{-n\Lambda^*(a)}$$

$$\Lambda^*(a) = \infty \Rightarrow \boxed{p_n = 0}$$

$$0 < \Lambda^*(a) < \infty \Rightarrow p_n \rightarrow 0 \text{ at least as fast as } \underline{e^{-n\Lambda^*(a)}}$$

$$\Lambda^*(a) = 0 \Rightarrow \boxed{P_n \leq 1} \quad (\text{trivial bound})$$

(won't happen in "most" cases)

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n > a] \leq -\Lambda^*(a)$$

(LDUB)

Large Deviation Lower Bound

$$\text{To show: } \liminf_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n > a] \geq -\Lambda^*(a)$$

(LDLB)

Claim: It is enough to show: $\forall \delta > 0$,

$$(*) \dots \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{\mu}(-\delta < \bar{X}_n < \delta) \geq -\Lambda_{\mu}^*(0).$$

Why does $(*) \Rightarrow \text{LDLB}$?

Firstly, using a shifting of measure argument, we can establish that $\forall x \in \mathbb{R}$ and $\forall \delta > 0$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{\mu}[x - \delta < \bar{X}_n < x + \delta] \geq -\Lambda_{\mu}^*(x).$$

(**)

$$\boxed{(*) \Rightarrow (**)} \Rightarrow (LDLB)$$

Define a prob measure τ on $\mathbb{R}$ by

$$\tau(B) = \mu(x+B) \quad , \quad B \subseteq \mathbb{R} \text{ (Borel)}$$

$$[x+B := \{x+b : b \in B\}]$$

$$X_1, X_2, \dots \stackrel{iid}{\sim} \mu \Rightarrow X_1-x, X_2-x, \dots \stackrel{iid}{\sim} \tau$$

$$\Rightarrow \boxed{P_\tau(-\delta < \bar{X}_n < \delta) = P_\mu(x-\delta < \bar{X}_n < x+\delta)}$$

Exc: Show that $\Lambda_\tau^*(0) = \Lambda_\mu^*(x)$.

[Hint: First show $\Lambda_\tau(\theta) = \Lambda_\mu(\theta) - \theta x$]

$$\text{Therefore } \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_\mu[x-\delta < \bar{X}_n < x+\delta]$$

$$= \liminf_{n \rightarrow \infty} \frac{1}{n} \log P_\tau[-\delta < \bar{X}_n < \delta]$$

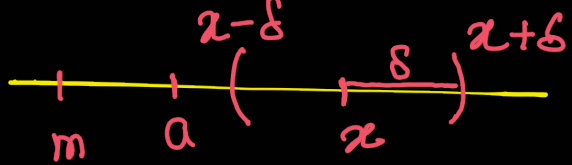
$$\stackrel{(*)}{\geq} -\Lambda_\tau^*(0) \stackrel{\text{Exc}}{=} -\Lambda_\mu^*(x).$$

Sketch of Proof of $(**) \Rightarrow LDLB$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_{\mu} [x - \delta < \bar{X}_n < x + \delta] \geq -\Lambda_{\mu}^*(x). \quad (**)$$

$$\Rightarrow \text{LDLB}$$

Take any $x > a$ and a $\delta > 0$ such that $x - \delta > a$



$$\Rightarrow (x - \delta, x + \delta) \subseteq (a, \infty)$$

$$\Rightarrow P[\bar{X}_n > a] = P[\bar{X}_n \in (a, \infty)]$$

$$\geq P[\bar{X}_n \in (x - \delta, x + \delta)]$$

$$\Rightarrow \lim_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n > a] \geq \lim_{n \rightarrow \infty} \frac{1}{n} \log P[x - \delta < \bar{X}_n < x + \delta]$$

$$\stackrel{(**)}{\geq} -\Lambda^*(x)$$

Taking sup over $x > a$, we get

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n > a] \geq \sup_{x > a} [-\Lambda^*(x)]$$

$$= - \inf_{x > a} \Lambda^*(x) \quad \checkmark$$

$$= - \underline{\Lambda^*(a)} \quad \left[\begin{array}{l} \text{Using cont} \\ \text{under} \\ \text{extra} \\ \text{assumption} \end{array} \right]$$

Sketch of proof of (*)

With loss of generality, assume first that X_i 's are nonneg r.v.s, i.e.,

$$\mu((-\infty, 0)) = 0.$$

To show: $\forall \delta > 0,$

$$(*) \dots \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(-\delta < \bar{X}_n < \delta) \geq -\Lambda^*(0).$$

$$P[-\delta < \bar{X}_n < \delta] = P[0 \leq \bar{X}_n < \delta]$$

$$\geq P[X_1 = X_2 = \dots = X_n = 0]$$

$$= (\mu(\{0\}))^n$$

$$\Rightarrow \lim_{n \rightarrow \infty} \frac{1}{n} \log P[-\delta < \bar{X}_n < \delta] \geq \log(\mu(\{0\})).$$

Therefore proof of (*) in this special case will be complete if we show

$$- \Lambda^*(0) = \log(\mu(\{0\})).$$

To show this, note that

$$\Lambda^*(0) = \sup_{\theta \in \mathbb{R}} [\theta \cdot 0 - \Lambda(\theta)]$$

$$= - \inf_{\theta \in \mathbb{R}} \Lambda(\theta)$$

$$\Rightarrow - \Lambda^*(0) = \inf_{\theta \in \mathbb{R}} \Lambda(\theta)$$

Exc: $\mu((-\infty, 0)) = 0 \Rightarrow \Lambda(\theta)$ is a
nondecreasing function
of θ

$$\text{Hence } \inf_{\theta \in \mathbb{R}} \Lambda(\theta) = \lim_{\theta \searrow -\infty} \Lambda(\theta)$$

$$= \lim_{\theta \searrow -\infty} \log \int_{[0, \infty)} e^{\theta x} \mu(dx)$$

Note:
$$e^{\theta x} \xrightarrow{\text{as } \theta \searrow -\infty} \begin{cases} 1 & \text{if } x=0 \\ 0 & \text{if } x>0 \end{cases}$$

Exc: Using DCT, show that $\Lambda(\theta) \searrow \log \mu(\{0\})$ as $\theta \searrow -\infty$

$$\Rightarrow -\Lambda^*(0) = \inf_{\theta \in \mathbb{R}} \Lambda(\theta) = \lim_{\theta \searrow -\infty} \Lambda(\theta) \stackrel{\text{Exc}}{=} \log(\mu(\{0\})).$$

This completes the proof_{of (*)} in this special case.

Similarly, you can show (*) when X_i 's are non-positive, i.e., $\mu([0, \infty)) = 0$.

In the general case, the proof of (*) needs a change of measure (also known as "exponential tilting") argument + a truncation argument.

Idea: Change the measure cleverly so that the rare event in the original measures becomes a typical event under the new measure and then use it to obtain the lower bound in (*).
(See the lecture-notes).

Exc: Following the proof of $(*) \Rightarrow \text{LDLB}$, show that (*) implies the following:

For any open $U \subseteq \mathbb{R}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n \in U] \geq - \inf_{x \in U} \Lambda^*(x).$$

Remark: Working harder on the upper bound, it is possible to show:

For any closed $F \subseteq \mathbb{R}$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P[\bar{X}_n \in F] \leq - \inf_{x \in F} \Lambda^*(x).$$

n = speed Λ^* = rate function
of the "LDP".

Defn: Suppose (S, d) is a

Polish space and $\{\mu_n\}_{n \geq 1} \subseteq \mathcal{P}(S)$

= t.s.c. prob measures on $(S, \mathcal{B}_S)$.

We say that $\{\mu_n\}_{n \geq 1}$ satisfies

LDP (Large Deviation Principle)

with speed $\{a_n\}_{n \geq 1} \subseteq (0, \infty)$

($a_n \nearrow \infty$) and an lsc rate function

$I: S \rightarrow [0, \infty]$ if

• $\forall$ open $U \subseteq S$,

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \log \mu_n(U) \geq \overbrace{-\inf_{x \in U} I(x)}^{-I(U)} \geq -J(U)$$

• $\forall$ closed $F \subseteq S$,

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{a_n} \log \mu_n(F) \leq \overbrace{-\inf_{x \in F} I(x)}^{-I(F)} \leq -J(F)$$

More generally, for any $B \subseteq S$ (Borel), define

$$I(B) := \inf_{x \in B} I(x).$$

Thm: (Cramer's Thm)

If $X_1, X_2, \dots \stackrel{\text{iid}}{\sim} \mu$ on $\mathbb{R}$,
(Law)

then the seq $\mu_n := P[\bar{X}_n \in \cdot]$,

$n \geq 1$, of prob measures on $\mathbb{R}$

satisfies LDP with speed

$a_n = n$ and rate function

$$\underline{I} = \Lambda^*.$$

Remark:

Suppose $\{\mu_n\}_{n \geq 1} \subseteq \mathcal{P}(S)$ satisfies

"LDP" with speed $a_n \nearrow \infty$ but "rate
function" $I: S \rightarrow [0, \infty]$, which is
not necessarily lsc. In this

situation, define the largest lsc minorant
 J of I as follows:

increasing seq in $n \leq I(x)$

$$J(x) := \lim_{n \rightarrow \infty} \left[\inf \{ I(y) : y \in B(x, \frac{1}{n}) \} \right],$$

↑
well-defined
ptwise

$x \in S$

Claim 1: J is lsc. $J \leq I$.
(ptwise)
(largest such)

Claim 2:
- $I(U) \geq -J(U)$
- $I(F) \leq -J(F)$

Therefore without loss of generality,
We may assume that I is lsc.

Notation: $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I, \{a_n\}_{n \geq 1})$. (LSC)

Convention: If $a_n = n \quad \forall n \in \mathbb{N}$, most imp special case

then we write $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I)$ (LSC)

Exc: Show that $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I, \{a_n\})$
if and only if $\forall$ Borel $B \subseteq S$,

$$\underline{-I(B^o)} \leq \liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mu_n(B)$$

$$\leq \limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mu_n(B)$$

$$\leq \underline{-I(\overline{B})}.$$

Defn: For a seq $\{Y_n\}_{n \geq 1}$ of S -valued r.v.s, we write $\{Y_n\} \sim \text{LDP}(I, \{a_n\})$ if $\{\mu_n\} \sim \text{LDP}(I, \{a_n\})$, where $\forall n \in \mathbb{N}$, $\mu_n = \text{law of } Y_n$, i.e., $\forall n \in \mathbb{N}$, $\mu_n(B) = P[Y_n \in B]$, $B \subseteq S$.
(Borel)

Qn: Is the rate function unique?

Ans: Yes (Exc)

(Lec 12 from the course)

Remark: for "nice" sets $B \subseteq S$
(Borel)

and for "large" n , $\mu_n(B) \approx e^{-a_n I(B)}$
whenever $\{\mu_n\} \sim \text{LDP}(I, \{a_n\})$.

In part, if $\{\mu_n\} \sim \text{LDP}(I)$,
 then $\mu_n(B) \approx e^{-n I(B)}$ for
 large n and a "nice" B .

e.g. — Cramer's Thm
 $P(\bar{X} > a) \approx e^{-n \lambda^*(a)}$

$\forall \underbrace{a > m}$

Towards the theory of LDP

Thm: (Laplace Principle): Let $h: [0,1] \rightarrow \mathbb{R}$
 be a cont f^n . Then we have

$$\lim_{n \rightarrow \infty} \underbrace{\frac{1}{n} \log \int_0^1 e^{-nh(x)} dx}_{\text{An}} = - \min_{x \in [0,1]} h(x).$$

" "

log of the L^p -norm of $e^{-h(x)}$
 ($p=n$)

$\xrightarrow{n \rightarrow \infty}$ log of the L^∞ -norm of $e^{-h(x)}$.

Proof: Let $a_n = \frac{1}{n} \log \int_0^1 e^{-nh(x)} dx$
 $\forall n \geq 1$.

Let $\min_{x \in [0,1]} h(x) = l$. Thus

$\forall x \in [0,1], \quad h(x) \geq l$

$$\Rightarrow \int_0^1 e^{-nh(x)} dx \leq e^{-nl} \quad \forall n \geq 1$$

$$\Rightarrow \limsup_{n \rightarrow \infty} a_n \leq -l = -\min_{x \in [0,1]} h(x)$$

Remains to show:

$$\liminf_{n \rightarrow \infty} a_n \geq -l.$$

To this end, fix $\varepsilon > 0$, and consider any point $x_0 \in [0, 1]$.

Using continuity of h , take an open interval J containing x_0 throughout which h is at most $\underline{h(x_0) + \varepsilon}$. Then we have

$$\begin{aligned} \int_{[0,1]} e^{-nh(x)} dx &\geq \int_J e^{-nh(x)} dx \\ &\geq \int_J e^{-n(h(x_0) + \varepsilon)} dx \end{aligned}$$

$$\Rightarrow \int_{[0,1]} e^{-nh(x)} dx \geq e^{-n(h(x_0) + \varepsilon)} \underbrace{|J|}_{\substack{\uparrow \\ \text{Leb meas of } J}}^0$$

$$\Rightarrow \liminf_{n \rightarrow \infty} a_n = \liminf_{n \rightarrow \infty} \frac{1}{n} \log \int_{[0,1]} e^{-nh(x)} dx$$

$$\geq -h(x_0) - \varepsilon$$

Since $\varepsilon > 0$ is arbitrary, we have

$$\liminf_{n \rightarrow \infty} a_n \geq -h(x_0)$$

for each $x_0 \in [0,1]$.

$\Rightarrow$ Taking max over $x_0 \in [0,1]$, we get

$$\liminf_{n \rightarrow \infty} a_n \geq \max_{x_0 \in [0,1]} [-h(x_0)]$$

$$= - \min_{x_0 \in [0,1]} h(x_0)$$

$$= -1 \quad \square$$

Go back to LDP

Suppose that $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I)$ on S .

$$\mu_n(B) \stackrel{n \rightarrow \infty}{\approx} e^{-n I(B)}$$

$$\Rightarrow \int_S d\mu_n(x) \stackrel{n \rightarrow \infty}{\approx} \int_S e^{-n I(x)} dx$$

$$\Rightarrow \int_S e^{-n h(x)} \underline{d\mu_n(x)} \stackrel{n \rightarrow \infty}{\approx} \int_S e^{-n(h(x) + I(x))} dx$$

$$\Rightarrow \frac{1}{n} \log \int_S e^{-n h(x)} d\mu_n(x) \stackrel{n \rightarrow \infty}{\approx} - \inf_{x \in S} [h(x) + I(x)]$$

Thm (Varadhan's Integral Lemma):

Let $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I)$ on a Polish space S and $h: S \rightarrow \mathbb{R}$.

is a bdd cont function. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \int_S e^{-nh(x)} d\mu_n(x) = - \inf_{x \in S} [h(x) + \underline{I}(x)]$$

→ see Lec 12 from notes.

Weak LDP - replace the closed sets in the upper bound by compact sets.

$\left\{ \mu_n \right\}_{n \geq 1} \sim \text{WLDP}(I)$ on a Polish space S if

i) $I: S \rightarrow [0, \infty]$ is lsc.

ii) $\forall$ open $U \subseteq S$, $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(U) \geq -I(U)$.

iii) $\forall$ compact $K \subseteq S$, $\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(K) \leq -I(K)$.

Defn: $\{\mu_n\}_{n \geq 1} \subseteq \mathcal{P}(S)$. We say that $\{\mu_n\}_{n \geq 1}$ is exponentially tight if $\forall \alpha \in (0, \infty)$,

$\exists$ a compact set $K = K(\alpha) \subseteq S$ such
↳ doesn't depend on n

that $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(K^c) \leq -\alpha$.

Remark: It just means that the μ_n 's are "exponentially concentrated" (with prob $\geq 1 - e^{-n\alpha}$) on ^{a very} large compact set $K \subseteq S$.

Thm: If $\{\mu_n\}_{n \geq 1} \sim \text{WLDP}(I)$
and $\{\mu_n\}_{n \geq 1}$ is exponentially tight,
then $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I)$.

Pg 123 (Lecture 14) of notes - Pg 75-76 as well.

Remark: While establishing weak conv on complicated spaces ($C[0,1]$, say), we often prove finite-dimensional conv along with tightness. Similarly, to establish LDP, we verify WLDP along with exponential tightness.

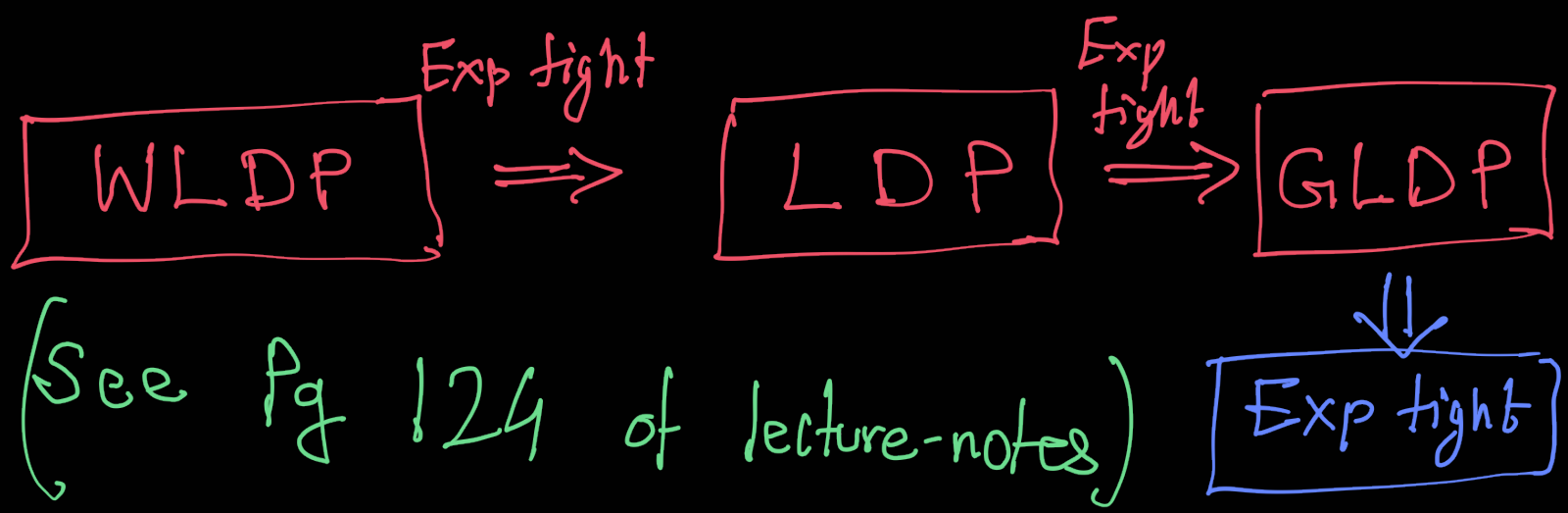
Defn: (i) A LD rate function $I: S \rightarrow [0, \infty]$ (isc by defn) is called a good rate function

if for each $\alpha \in [0, \infty)$, the set
 $(I \leq \alpha) := \{x \in S : I(x) \leq \alpha\}$
 is compact.

(2) If $\{\mu_n\}_{n \geq 1} \sim \text{LDP}(I)$ with
 a good rate function I , then we
 say that $\{\mu_n\}_{n \geq 1} \sim \text{GLDP}(I)$.

Thm: If $\{\mu_n\} \sim \text{LDP}(I)$ and $\{\mu_n\}$
 is exponentially tight, then $\{\mu_n\} \sim \text{GLDP}(I)$.

Cor: If $\{\mu_n\} \sim \text{WLDP}(I)$ and $\{\mu_n\}$
 is exponentially tight, then $\{\mu_n\} \sim \text{GLDP}(I)$.



Thm: If $\{\mu_n\} \sim \text{GLDP}$, then
 $\{\mu_n\}_{n \geq 1}$ is exponentially tight.

(See Pg 126 of lecture-notes)

Cramer's Thm on $\mathbb{R}^d$

$$\mu \in \mathcal{P}(\mathbb{R}^d)$$

$X_1, X_2, \dots \xrightarrow{\text{iid}} \mu$
($\mathbb{R}^d$ -valued r.v.s) (Law)

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad n \geq 1.$$

($\mathbb{R}^d$ -valued
r.v.)

$$\Lambda(\theta) := \log \mathbb{E}(e^{\theta \cdot X_1}), \quad \theta \in \mathbb{R}^d$$

(multivariate
log mgf)

$\theta \cdot x := \langle \theta, x \rangle = \text{the}$
usual inner product on $\mathbb{R}^d$.

Assume $\Lambda(\theta) \in (-\infty, \infty) \quad \forall \theta \in \mathbb{R}^d$

Define $\Lambda^*(x) \doteq \sup_{\theta \in \mathbb{R}^d} [\theta \cdot x - \Lambda(\theta)],$
 $(\geq 0) \quad x \in \mathbb{R}^d$

Thm: (Cramer) Under our assumptions,
 $\{\mu_n \doteq P[\bar{X}_n \in \cdot]\}_{n \geq 1} \sim \text{GLDP}(\Lambda^*).$

Sketch of Proof: (Lec 7-10 of
class-notes $\rightarrow$ details)

Step 1: $\{\mu_n\}_{n \geq 1} \sim \text{WLDP}(\Lambda^*)$

Step 2: $\{\mu_n\}_{n \geq 1}$ is exponentially tight

Step 3: Step 1 + Step 2

$\Rightarrow \{\mu_n\}_{n \geq 1} \sim \text{GLDP}(\Lambda^*).$

<u>Weak Conv</u>	<u>Large Deviations</u>
Finite-dim weak conv	WLDP
Tightness	Exp Tightness
Prokhorov's Thm	Thm (*)
Continuous Mapping Thm	Contraction Principle

Thm (*): If $\{\mu_n\}_{n \geq 1} \subseteq \mathcal{P}(S)$ is exponentially tight, then $\{\mu_n\}$ has a subseq $\{\mu_{n_k}\}$ such that

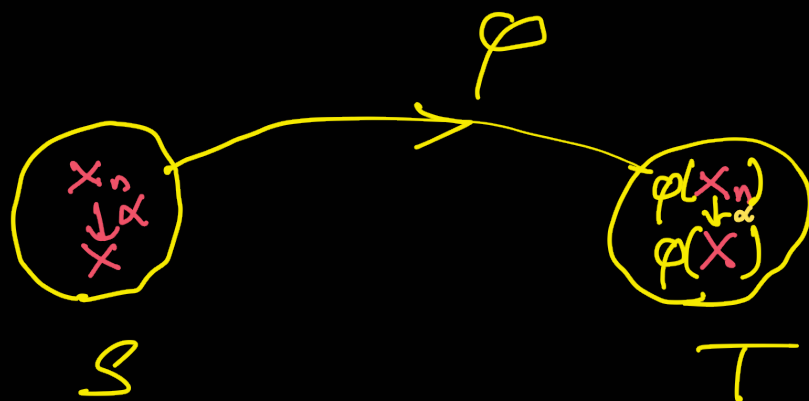
$$\{\mu_{n_k}\}_{k \geq 1} \rightsquigarrow \text{GLDP}(I, \{n_k\}_{k \geq 1})$$

for some rate function $I: S \rightarrow [0, \infty]$.

Go Back to Weak Convergence

Thm: (Cont Mapping Thm) Suppose S, T are two Polish spaces and $\varphi: S \rightarrow T$ is a cont map.

If $X_n \xrightarrow{d} X$ as S -valued r.v.s,
then $\varphi(X_n) \xrightarrow{d} \varphi(X)$ as T -valued
r.v.s.



Remark: If φ is X -almost

surely cont, i.e., if

$$P[X \in D_\varphi] = 0,$$

where $D_\varphi = \{s \in S : \varphi \text{ is discont at } s\}$,

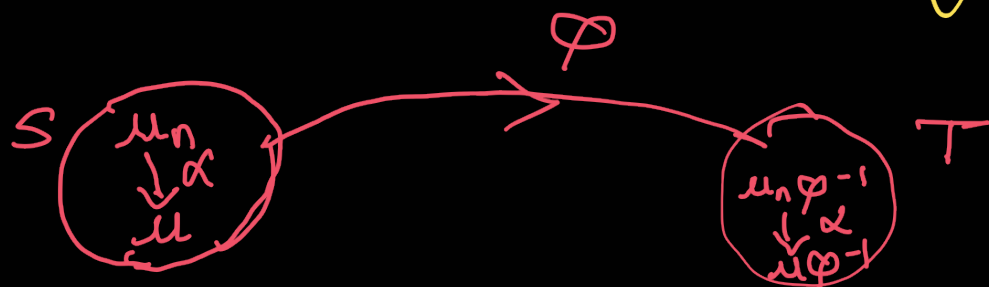
$$\text{then } \varphi(X_n) \xrightarrow{d} \varphi(X).$$

Thm (Cont Mapping Thm): Suppose

S, T are Polish spaces and $\varphi: S \rightarrow T$ is a cont map.

If $\mu_n \xrightarrow{d} \mu$ in $\mathcal{P}(S)$, then

$$\mu_n \circ \varphi^{-1} \xrightarrow{d} \mu \circ \varphi^{-1} \text{ in } \mathcal{P}(T).$$



Remark: As long as

$$\mu\left(\underbrace{\{s \in S: \varphi \text{ is discontinuous at } s\}}_{D\varphi}\right) = 0$$

then $\mu \circ \varphi^{-1} \xrightarrow{\alpha} \mu \circ \varphi^{-1}$.

Back to LDP

Thm: (Contraction Principle for Good LDP) Suppose S, T are two Polish spaces and $\varphi: S \rightarrow T$ is a cont map.

If $\{\mu_n\}_{n \geq 1} \sim \text{GLDP}(I)$

on S , then

$\{\mu_n \circ \varphi^{-1}\}_{n \geq 1} \sim \text{GLDP}(J)$

on T , where

$$J(y) \doteq \inf \{ I(x) : \varphi(x) = y \},$$

$y \in T$

(Convention $\inf \emptyset = +\infty$.)

Pg 158 of Lec 17 of

Lecture notes.

Remark: Even to show that J is lsc,
we need the goodness of I . Therefore,

this result won't hold if we simply replace the GLDP by LDP everywhere.

Thm: (Contraction Principle for a general LDP)

Suppose S, T are two Polish spaces and

$\varphi: S \rightarrow T$ is a cont map. If

$\{\mu_n\} \sim \text{LDP}(I)$ on S , then

$\{\mu_n \circ \varphi^{-1}\} \sim \text{LDP}(\tilde{I})$ on T , where

$\tilde{I}$ is the largest lsc minorant ^{lsc envelope}

of I (defined in the prev thm) defined by

$$\tilde{I}(y) \doteq \inf \{ I(x) : \varphi(x) = y \},$$

$y \in T.$

Sanov's Thm for Finitely Supported
RVs (Lec 17, 18, 20 of
lecture-notes)

$$A_0 = \{a_1, a_2, \dots, a_k\} \subseteq \mathbb{R}.$$

$$\text{WLOG } a_1 < a_2 < \dots < a_k.$$

$$M = \{ \text{soa prob vectors } \underline{\mu} = (\mu_1, \dots, \mu_k) \\ \text{on } A_0 \quad (\mu_i = \text{prob of } a_i) \}$$

$$\mathcal{P}(A_0) = M = \{ (\mu_1, \mu_2, \dots, \mu_k) \in [0, 1]^k : \\ \mu_1 + \mu_2 + \dots + \mu_k = 1 \} \\ \subseteq \mathbb{R}^k.$$

Equip M with the restriction (from $\mathbb{R}^k$) topology.

Let $\mathcal{M}$ = Borel σ -field on M .

Fix $\underline{\mu}^0 = (\mu_1^0, \mu_2^0, \dots, \mu_k^0) \in M^0$

Each $X_i: \Omega \rightarrow A_0$ ^{$P \leftarrow \text{prob meas}$} $0 < \mu_i < 1 \quad \forall i$

Suppose $X_1, X_2, \dots$ ^{iid} $\underline{\mu}^0$,
(A_0 -valued r.v.s) (Law)

i.e., $\forall i = 1, 2, \dots, k$

$$P[X_i = a_i] = \mu_i^0.$$

Defn: The (seq of) empirical measures associated with $\{X_i\}_{i \geq 1}$ is defined to be the sequence of M -valued r.v.s

$$L_n^{(B)} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(B), \quad n \geq 1, \quad B \subseteq A_0.$$

= Proportion of X_i 's in B

[Here for any $u \in \mathbb{R}$, δ_u on $\mathbb{R}$, denotes the Dirac delta measure,

i.e.,
$$\delta_u(B) = \begin{cases} 1 & \text{if } u \in B, \\ 0 & \text{if } u \notin B. \end{cases}$$

Remark: $\{L_n\}$ is a seq of random measures on A_0 and hence is an M -valued r.v.

Exc: Show that for P -almost all $\omega \in \Omega$,

$$L_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \xrightarrow{\alpha} L$$

as $n \rightarrow \infty$, where L is the prob measure on A_0 corr to the prob vector $\underline{\mu}^0 = (\mu^0_1, \dots, \mu^0_k)$.

Take Borel $B \subseteq M$ such that

$\underline{\mu}^0 \notin \overline{B}$. Then above exc

$$\Rightarrow (*) \dots P[L_n \in B] \rightarrow 0 \text{ as } n \rightarrow \infty$$

Qn: How fast does the conv hold in (*)?

Ans $\rightarrow$ Relative Entropy

Defn: (1) Entropy of $\underline{\mu}$,

$$H(\underline{\mu}) := - \sum_{i=1}^k \mu_i \log \mu_i, \quad (\mu_1, \mu_2, \dots, \mu_k) \in M.$$

(2) Entropy of $\underline{\mu}$ relative to $\underline{\mu}^0$,

$$H(\underline{\mu} | \underline{\mu}^0) := \sum_{i=1}^k \mu_i \log \frac{\mu_i}{\mu_i^0}, \quad \underline{\mu} \in M.$$

[Conventions: $0 \log 0 := 0$, $0 \log \frac{0}{0} := 0$,
 ~~$0 \log \frac{+ve}{0} := +\infty$.~~]

Exc: (1) $H(\underline{\mu} | \underline{\mu}^0) \geq 0 \quad \forall \quad \underline{\mu} \in M$

(2) $H(\underline{\mu} | \underline{\mu}^0) = 0$ if and only if $\underline{\mu} = \underline{\mu}^0$.

Remark: $H(\underline{\mu} | \underline{\mu}^0)$ measures how random / disordered $\underline{\mu}$ is relative to $\underline{\mu}^0$. In other words, we can think of this as a measure of how far $\underline{\mu}$ is from $\underline{\mu}^0$.

Thm (Sanov's Thm for finitely supported r.v.)

Suppose $X_1, X_2, \dots \stackrel{\text{iid}}{\sim} \underline{\mu}^0 \in \mathcal{M}^0$.

Let $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, $n \in \mathbb{N}$

be the associated empirical measure seq.

Then we have

$$\{L_n\}_{n \geq 1} \sim \text{GLDP}(I) \text{ on } \mathcal{M},$$

where $I: \mathcal{M} \rightarrow [0, \infty)$ is defined by

$$I(\underline{\mu}) = H(\underline{\mu} | \underline{\mu}^0), \quad \underline{\mu} \in \mathcal{M}.$$

[for a B s.t. $\underline{\mu}^0 \notin \overline{B}$,

$$P[L_n \in B] \approx e^{-n I(B)}.$$

Cor: (Cramer's Thm for finitely supported r.v.s)

Suppose $X_1, X_2, \dots \stackrel{\text{iid}}{\sim} \mu^0$ ($\in M^0$).
(Law on A_0)

$$\text{Let } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad n \geq 1.$$

Then $\{\bar{X}_n\} \sim \text{GLDP}(\bar{I})$ on $[a_1, a_k]$,

where

$$\bar{I}(x) := \inf \{H(\mu | \mu^0) : \mu \in M, \sum_{i=1}^k a_i \mu_i = x\},$$

$$(\inf \phi := +\infty) \quad x \in [a_1, a_k].$$

Proof: Exe (Sanov's Thm + contraction principle for GLDP)