Large deviation principle

$$P(A_n \in da) \underset{n \gg 1}{\approx} \exp(-n\, I(a))\, da$$

$I(a)$ - rate function (decay rate for $a / I(a) > 0$)

- $I(a) \geq 0$
- $I(a)$ convex
- single minimum $a^* / I(a^*) = 0$

Super-exponential decay $\Rightarrow I = +\infty$

Sub-exponential decay $\Rightarrow I = 0$

$\boxed{\text{Gärtner-Ellis Theorem}}$

A_n real random variable

scaled cumulant generating function

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln \langle e^{k n A_n} \rangle$$

$$\langle e^{nkA_n} \rangle = \int_{\mathbb{R}} e^{nka}\, P(A_n \in da)$$

if $\lambda(k)$ exists and is differentiable for all $k \in \mathbb{R}$

$\Rightarrow$ large deviation principle holds for A_n

with a rate function $I(a) = \sup_{k\in\mathbb{R}} (ka - \lambda(k))$

$I(a)$ is the Legendre-Fenchel transform of $\lambda(k)$

$$I = \lambda^*$$

Plausibility argument (of part 2) (connection with saddle-point approximation of (Laplace))

$$\underbrace{P(A_n \in da) \approx \exp(-n I(a))\, da}$$

$$\langle \exp(n k A_n) \rangle \underset{n \gg 1}{\approx} \int_{\mathbb{R}} \exp\left[n [ka - I(a)] \right] da$$

saddle point

$$\approx \exp\left(n \sup_{a \in \mathbb{R}} [ka - I(a)] \right)$$

$$\lambda(k) = \lim_{n \to \infty} \frac{1}{n} \ln \langle \exp(n k A_n) \rangle$$

$$= \sup_{a \in \mathbb{R}} [ka - I(a)] \to \lambda = I^*$$

[Varadhan theorem: large deviation principle holds for A_n
$\Rightarrow \lambda = I^*$]

if $\lambda(k)$ is differentiable $\forall k \in \mathbb{R}$

$\to$ Legendre - Fenchel transform is SELF-INVERSE

$$\Rightarrow \underbrace{I = \lambda^*}$$

Examples with sample means of i.i.d. random variables

$$S_n = \frac{1}{n} \sum_i^n X_i$$

G.E.H $\to$ Cramer Theorem

$$\lambda(k) = \ln \langle \exp(kX) \rangle$$

Gaussian Sample Mean: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$$\lambda(k) = \ln\langle \exp(kX)\rangle = \mu k + \frac{1}{2}\sigma^2 k$$

$$I(s) = \sup_{k\in\mathbb{R}} \left[ks - \lambda(k)\right]$$

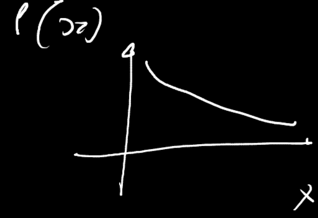
sup condition

$$\underbrace{k(s) = \frac{s-\mu}{\sigma^2}} \Leftarrow \mu + \sigma^2 k = \underbrace{\lambda'(k) = s}$$

$$I(s) = k(s)\cdot s - \lambda(k(s)) = \frac{(s-\mu)^2}{2\sigma^2}$$

<u>Exponential sample mean</u>

$$p(x) = \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) \quad 0 \le x < +\infty$$



(mean μ) $\quad S_n = \frac{1}{n}\sum_{1}^{n} X_i$

$$\lambda(k) = \ln\langle \exp(kX)\rangle = -\ln\left[1-\mu k\right] \quad \underbrace{k < 1/\mu}$$

$$I(s) = \sup_{k<1/\mu} \left[ks - \lambda(k)\right]$$

sup condition: $\lambda'(k) = s \Rightarrow k(s) = \frac{s-\mu}{\mu s}$

$\lambda'(k) = \frac{\mu}{1-\mu k}$ $\quad k < 1/\mu \Rightarrow s > 0$

$$I(s) = k(s)\, s - \lambda\left[k(s)\right] = \frac{s}{\mu} - 1 - \ln\left(\frac{s}{\mu}\right)$$

$I(\mu) = 0 \to \mu$ is the "typical value" for S_n

Properties of λ

$$\lambda(k) = \lim_{n\to\infty} \frac{1}{n} \ln \langle \exp(nkA_n)\rangle$$

- $\lambda(0) = 0$ (normalization of $P(A_n \in da)$)

$$\lambda'(0) = \lim_{n\to\infty} \left.\frac{\langle A_n \exp(nkA_n)\rangle}{\langle \exp(nkA_n)\rangle}\right|_{k=0} = \underbrace{\lim_{n\to\infty} \langle A_n\rangle}_{\text{"the typical" value}}$$

(if it exists)

$$\lambda''(0) = \lim_{n\to\infty} \left[n\left(\langle A_n^2\rangle - \langle A_n\rangle^2\right)\right]$$

$$= \lim_{n\to\infty} n \operatorname{var}(A_n) = \sigma^2 \ (= \operatorname{var}(X))$$

i.i.d. sample mean $\operatorname{var}(A_n) = \frac{\sigma^2}{n}$

- $\lambda(k)$ is convex (always)

- $\lambda(k)$ is differentiable everywhere and $\lambda(k)$ is convex $\Rightarrow$ Legendre-Fenchel

$$\left[\begin{array}{l} I(a) = k(a)\, a - \lambda(k(a)) \Leftarrow \text{"Legendre"} \\ \text{with } k(a) \text{ defined by } \lambda'(k) = a \end{array} \right]$$

$$I'(a) = \cancel{k'(a)\cdot a} + k(a) - \cancel{\lambda'(k(a))\, k'(a)}$$

if $\lambda(k)$ is strictly convex: $\overset{!}{=} k(a)$

$$\boxed{\begin{array}{l} I'(a) = k(a) \\ \lambda'(k) = a(k) \end{array}}$$

(one-to-one relations between slopes

after inversion $\lambda'(k) = a(k)$

$\searrow$ it is possible to invert $k(a) \to a(k)$ [because $\lambda'(k)$ is monotonically increasing since $k(a)$ is implicitly defined by $\lambda'(k(a)) = a$]

Properties of $I(a)$

Varadhan th.

- $I(a) \geq 0$ $\quad \lambda(0) \stackrel{d}{=} \sup_a [-I(a)] = 0$

- $I(a)$ obtained from G.E. th. are strictly convex (no linear parts)

from $I'(a) = k(a)$

$I''(a) > 0 \quad \forall a$ $\quad k'(a) = 1/a'(k)$ $\quad$ from $\lambda'(k) = a(k)$

$$\Rightarrow \underbrace{I''(a) = k'(a)}_{} \stackrel{b}{=} \underbrace{\frac{1}{\lambda''(k)} = \frac{1}{a'(k)}}_{}$$

curvature of I = inverse curvature of generating function (property of Legendre transform)

Taylor expand $I(a)$ around $a^* \,/\, I(a^*) = 0$

$$I(a) \simeq \frac{1}{2} I''(a^*)(a-a^*)^2$$

$a^* =$ "typical" value

$$\to P(A_n \in da) \underset{n \gg 1}{\simeq} \exp\left(-\frac{n}{2} I''(a-a^*)^2\right)$$

(Law of large numbers)

Gaussian approx. for a close to a^*

"typical" fluctuations around a^* are Gaussian and small (central limit theorem) and $|a - a^*| \sim \frac{1}{(n I'')^{1/2}}$ Gaussian scaling

IID sample mean $\sim \sigma / n^{1/2}$

$\boxed{\text{Large deviations} \to |a - a^*| \gg \frac{1}{n^{1/2}} \quad \text{exponential scaling}}$

Law of large numbers:

if $I(a)$ has a unique minimum a^*

$I(a^*) = k(a^*) \cdot a^* - \lambda(k(a^*)) = 0$

if $k(a^*) = 0$ (because $\lambda(0) = 0$)

(for example when $I(a)$ is computed from G-E th. and is therefore strictly convex)

then the slopes duality implies $I'(a^*) = k(a^*) = 0$

and finally $a^* = \lambda'(k(a^*)) = \lambda'(0) = \lim_{n \to \infty} \langle A_n \rangle$

a^* is the "typical" value in the limit $n \to \infty$

and since $I(a) > 0 \iff a \neq a^*$ (a^* unique minimum)

$\lim_{n \to \infty} P(A_n \in da^*) = 1$ the probability measure "concentrates" on a^* in the limit $n \to \infty$