1

LARGE DEVIATION THEORY

A. DONEV

& Statistical Mechanics

Based on paper by H. Touchette & notes by E. Vanden-Eijnden

Let's start from a simple but informative example: Sum of i.i.d vars:

$$S_N = \sum_{i=1}^{N} X_i \quad , \quad X_i \text{ are i.i.d}$$

$$E[X] = 0 \quad , \quad \text{var}[X] = 1$$

for specificity

2

Law of Large numbers (LLT):

$$\lim_{N \to \infty} \frac{S_N}{N} = 0 \quad \text{(empirical mean)}$$

Central Limit Theorem (CLT):

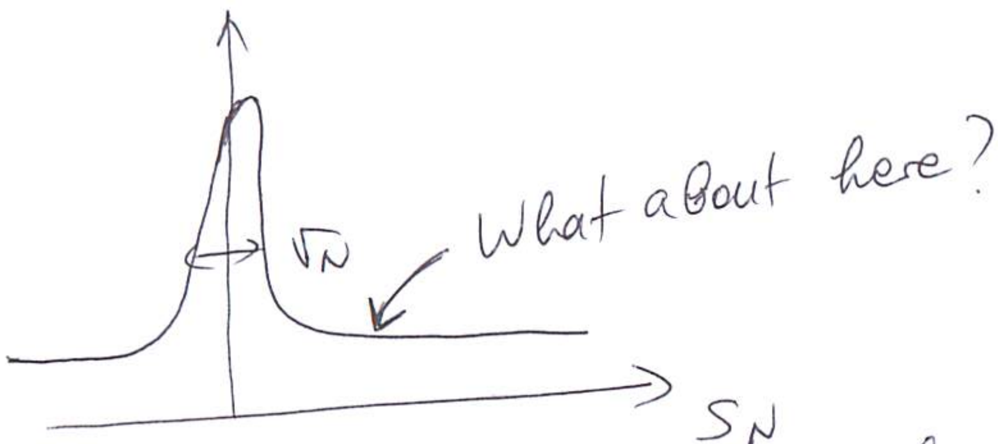
$$\frac{S_N}{N} \to \frac{1}{\sqrt{N}} \mathcal{N}(0,1) \quad \text{or}$$

normally distributed

$$\boxed{\lim_{N \to \infty} \frac{S_N}{\sqrt{N}} \to \mathcal{N}(0,1)}$$

In particular, this means that calculating $\boxed{P(S_N > N^{1/2} \cdot a)}$ can be done using Gaussian distributions

(3) The CLT <u>does not</u> tell you anything about the <u>tails</u> of the distribution of S_N, it only tells you about the behavior within a few standard deviations



How about $\boxed{P(S_N > N^{\alpha} \cdot a)}$?

$\boxed{1/2 < \alpha \leq 1}$

<u>Large Deviation Theory</u> (LDT): $\alpha = 1$

Or Moderate Deviation theory (MDT): $\alpha \in (\frac{1}{2}, 1)$

So LDT answers questions about the <u>tails</u> of the distribution of S_N (or the empirical mean), i.e. it describes <u>very (exponentially) unlikely events</u>.

LDT estimate:

$$\boxed{\lim_{N \to \infty} \frac{1}{N} \log\left[P(S_N > N a)\right] = I(a)}$$

where $I(a)$ is called the <u>rate function</u>

If $I(a)$ exists and $I(a) \neq 0$ for $\forall a$ we say S_N obeys a <u>large deviation principle</u>.

Gartner - Ellis theorem

Let A_n be a random variable parametrized by an integer $n > 0$, and define the scaled cumulant generating function :

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

where $k \in \mathbb{R}$ is a parameter

Define the Legendre - Fenchel transform

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

GT theorem:

If $\lambda(k)$ exists and is differentiable everywhere, then A_n satisfies an LD principle:

$$\boxed{\lim_{n\to\infty} \frac{1}{n} \log P\,(A_n \in da) = I(a)}$$

Means $A_n \in [a, a+da]$

Note that the Legendre-Feuchel transform is self-dual if $\lambda(k)$ is smooth (everywhere diff.)

Note: 1 $\lambda(k=0) = 0$

2
$$\lambda'(k=0) = \lim_{n\to\infty} \left. \frac{\langle A_n e^{nkA_n} \rangle}{\langle e^{nkA_n} \rangle} \right|_{k=0}$$

$$\lambda'(0) = \lim_{n\to\infty} \langle A_n \rangle$$

If $A_N = \frac{S_N}{N}$ then

$$\lambda'(0) = \lim_{n\to\infty} \frac{\langle S_n \rangle}{n} = E[X] \quad \text{(mean)}$$

3 $\boxed{\lambda \text{ is convex and positive}}$

4. There is a unique root to

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

and this realizes the maximum in

$$I(a) = \max_{k'} \left[k'a - \lambda(k') \right] = ka - \lambda(k)$$

Legendre transform

5. $I(a) > 0$ also and

6. If the GE theorem applies then $I(a)$ is strictly convex and positive

7. $$\boxed{\lambda(k) = \sup_a \{ka - I(a)\}}$$

Special case of Varadhan's theorem

Note that LDT subsumes and gives more detail about the <u>rate</u> of convergence in the LLN and CLT; (9)

<u>LLN</u>:

$$\begin{cases} \lambda'(0) = \lim_{n\to\infty} \langle A_n \rangle = a^* \\ \text{where } a^* \text{ is the (assumed unique)} \\ \text{global minimum of } I(a), \ I(a^*) = 0 \end{cases}$$

This is the weak form of the LLN

But here we get <u>more</u> – we have information about the rate of convergence

$P(A_n \in B) \to 0$ exponentially fast in n if $a^* \notin B$

<u>CLT</u>:

If $I(a)$ has a unique global minimum $I(a^*) = 0$ and is twice differentiable, then

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

and $P(A_n \in [a, a+da]) \simeq \exp\left[-\frac{n}{2} I''(a-a^*)^2\right]$

for "small" $a-a^*$, which is

$$a - a^* = O\left(\frac{1}{\sqrt{n}}\right)$$ — small deviations

So for small deviations the CLT and LDT give the same information

Take $A_n \equiv S_n / n = \dfrac{\sum x_i}{n}$ (sample mean)

$$\lambda(k) = \lim_{n \to \infty} \frac{1}{n} \ln \left\langle e^{k \sum x_i} \right\rangle = \ln \left\langle e^{kx} \right\rangle$$

The LD principle is the classical Cramér's theorem.

<u>Contraction principle</u>

$$B_n = h(A_n)$$

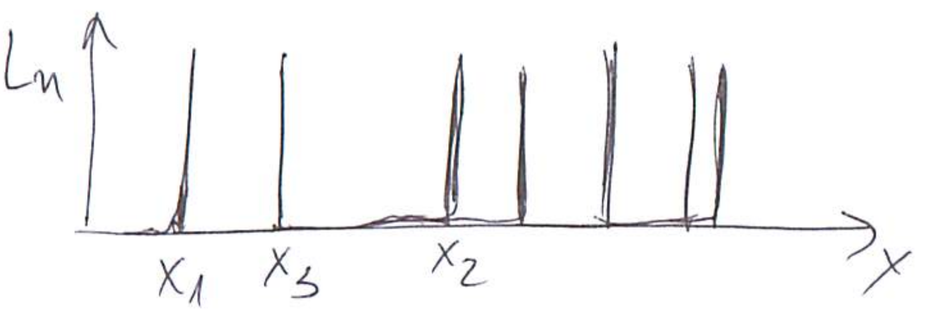
If A_n obeys an LD principle, so does B, with

$$\boxed{I_B(b) = \inf_{a : h(a) = b} I_A(a)}$$

Sanov's theorem

Let w_i be i.i.d real-valued random variables with PDF $\rho(x)$. Consider the empirical histogram or empirical density

$$L_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(w_i - x), \quad x \in \mathbb{R}$$



$L_n(x)$ is a normalized density

$L_n(x)$ is a random function, and obeys an LD principle

It can be shown (not simple) that

$$I[\mu(x)] = \int_{-\infty}^{\infty} dx\, \mu(x) \ln \frac{\mu(x)}{\rho(x)}$$

Shanon's entropy or the Kullback-Leibler divergence or relative entropy of $\mu(x)$ and $\rho(x)$

So the information-theoretic entropy is a rate function for how (un)likely it is for the empirical histogram to differ from the true distribution.

Entropy and LDT

Let $z = (z_1, \dots, z_n)$ be the microstate and $E_n = \frac{H}{n}$ be the average energy. (a random var.)

Let $I(u) = \lim_{n\to\infty} \left\{ -\frac{1}{n} \ln P[E_n \in du] \right\}$

be the rate function for the average energy per particle,

Then

$$\boxed{I(u) = \text{const.} - S(u)}$$

where $\boxed{s(u) = \lim_{n\to\infty} \frac{1}{n} \ln \Omega(u)}$

is the entropy per particle (entropy density), and $\Omega(u)$ is the number of microstates with energy u:

$$\Omega(u) = \int_{E_n(z)\in[u,u+du]} dz = \int_z \delta[E_n(z)-u]\,dz$$

So the classical (Boltzmann) <u>entropy is the rate function for energy fluctuations</u>. (here set $k_B = 1$ to avoid units)

Our generalized entropy corresponds to choosing as coarse-grained observables something other than average energy, mesoscopic $X(z)$.

$$P(X) \asymp e^{n\,S(x)}$$ (Einstein distribution)

meaning $\lim_{n\to\infty} \frac{1}{n} \log\left[P(\underline{X} \in [x, x+dx])\right] = S(x)$

where $S(x)$ is the entropy per particle associated with the coarse grained level of description $X(w)$.

The LD principle only applies if entropy is an extensive quantity.

The real content of the LD principle here is the fact that entropy is extensive, i.e., it becomes linear in n for large n. This implies the existence of a <u>thermodynamic limit</u> as $n \to \infty$ in which deviations from the <u>average</u> behavior become <u>exponentially unlikely</u> (in n).

Note that the classical free energy is the Legendre transform of $s(u)$

$$f(T) = \inf_u \{u - k_B T\, s(u)\}$$