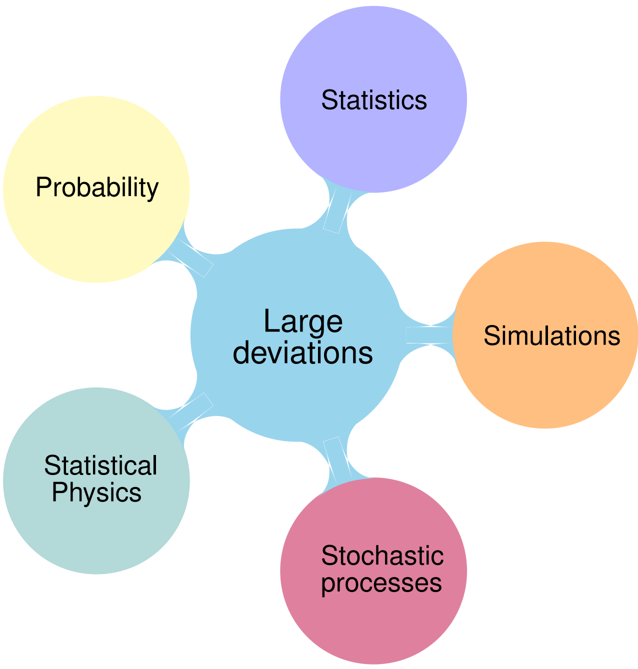
Applications of large deviation theory

Hugo Touchette

National Institute for Theoretical Physics (NITheP)
Stellenbosch, South Africa

Maties Machine Learning
20 April 2018

Overview



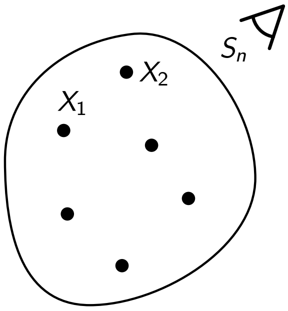
Goal

Explain how deterministic behavior arises from randomness

- Stochastic processes
- Many interacting components
- Rare events
- Typical or generic events
- Emergent determinism

Large deviation theory in one slide

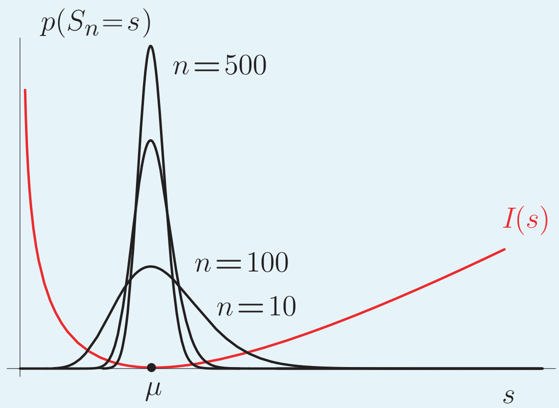
$S_n(\vec{X})$	Macro
$\vec{X} = (X_1, X_2, \ldots, X_n)$	Micro



Large deviation principle (LDP)

$$P(S_n = s) \approx e^{-nI(s)}$$

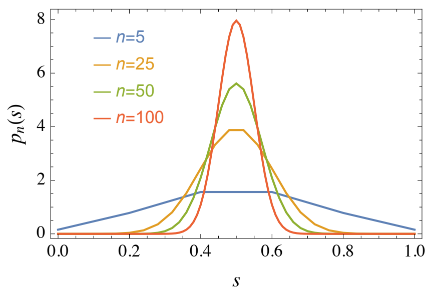
- $I(s)$ = rate function
- Exponentially rare events
- Typical value: $I(s^*) = 0$
- Concentration of probability



Coin tossing

H	T	T	H	T	$\cdots$
1	0	0	1	0	$\cdots$
X_1	X_2	X_3	X_4	X_5	$\cdots$

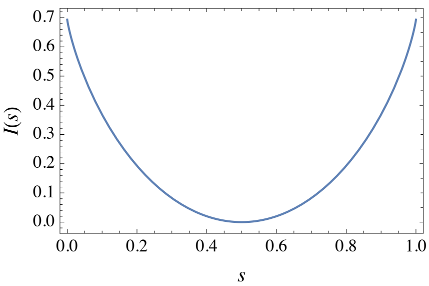
$$S_n = \frac{\#\text{ heads}}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$



LDP

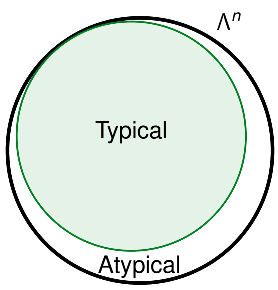
$$P(S_n = s) \approx e^{-nI(s)}$$

$$I(s) = s\log s + (1-s)\log(1-s) + \log 2$$



Typical sequences

$$\#\{\vec{X} : S_n = 0.5\} \approx 2^n$$



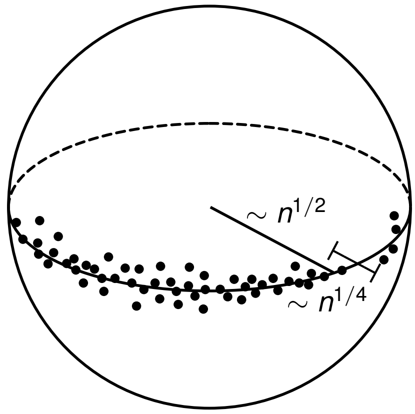
Gaussian vectors and hyperspheres

- Gaussian vector in n-dim:

$$\vec{X} = (X_1, X_2, \ldots, X_n)$$

- $X_i \sim \mathcal{N}(0,1)$ iid
- Rescaled "radius":

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$



LDP

$$P(S_n = s) \approx e^{-nI(s)}$$

- Typical value: $S_n \to 1$
- Exponential concentration

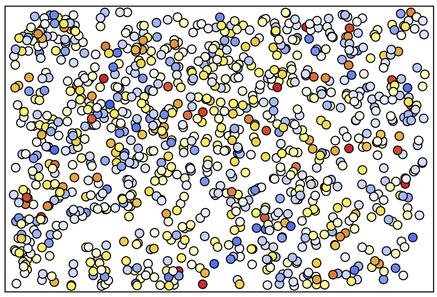
Typicality

- Most points lie on surface
- Asymptotically uniform
- Volume concentrated near surface as $n \to \infty$

Statistical physics

- Total energy:

$$U_N = \sum_{i=1}^{N} \frac{v_i^2}{2m}, \qquad N \sim 10^{23}$$



- Velocity distribution:

$$L_N(v) = \frac{\#\text{ particles } v_i \in [v, v+\Delta v]}{N\Delta v}$$

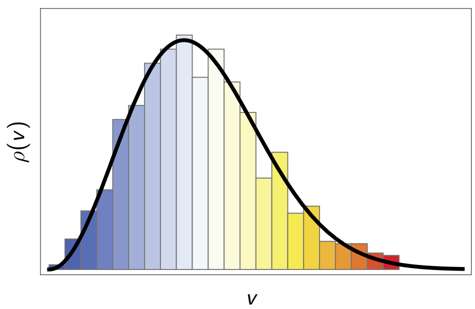
LDP

$$P(L_N = \rho) \approx e^{-NI(\rho)}$$

- Equilibrium distribution:

$$\rho^*(v) = c\, v^2 e^{-\frac{mv^2}{2k_B T}}$$

- Maxwell's distribution

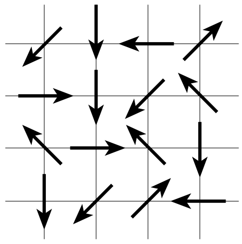


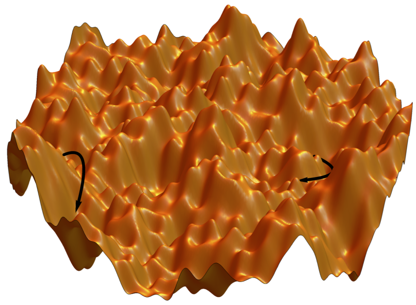
Spin glasses

- Energy:

$$H = -\sum_{ij} \underbrace{J_{ij}}_{\text{disorder}} \sigma_i \sigma_j + h \sum_i \sigma_i$$

- Find min energy (ground state)
- Count number $\Omega_N(u)$ local min with given energy





LDPs

$$\Omega_N(u) \approx e^{N\Sigma(u)}$$

- $\Sigma(u) =$ entropy

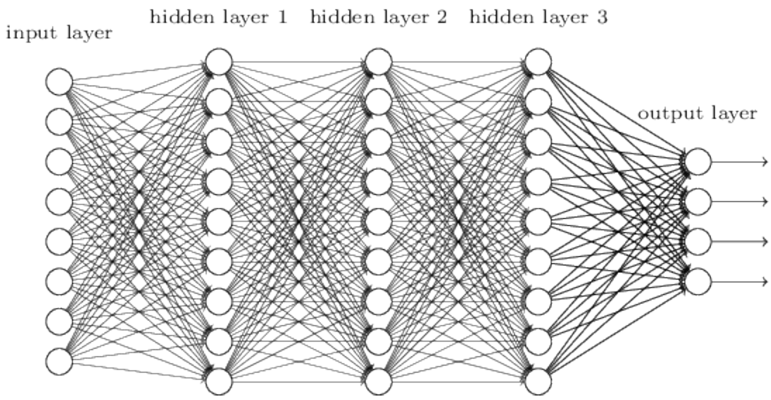
Typicality

- Exponentially many metastable states
- Most critical points are saddles in high dim

Problems from machine learning

- Neurons: $\{\sigma_i\}$
- Weights: $\{w_{ij}\}$
- Cost function:

$$\mathcal{C}(\text{input}, \text{output}, \{w_{ij}\})$$



Practical

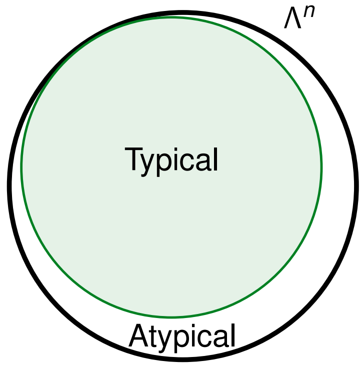
- Number of metastable states
- Typicality of data
- Overfitting fluctuations

Fundamental

- Why ML works at all?
- Generic attractors
- Represent typical features

Other applications

- Information theory
- Random graphs
- Markov process: X_t, $S_T[x]$
- Signal analysis
- Statistical physics
- Phase transitions
- Quantum systems
- ...



- Full space is huge
- Most states "look" the same

F. den Hollander, *Large Deviations*, AMS, 2000

H. Touchette, The large deviation approach to statistical mechanics, *Physics Reports* **478**, 2009

www.physics.sun.ac.za/~htouchette

Current research: LDs of Markov processes

- Process: $\{X_t\}_{t=0}^T$
- Observable: $A_T[x]$

LDP

$$P(A_T = a) \approx e^{-TI(a)}$$

Problems

- Predict how rare fluctuations happen
- Effective process describing fluctuations
- Related to non-Hermitian spectral problem