Introduction to large deviation theory

Yier Lin (University of Chicago)

Oct 31, 2022

Oregon State University, Analysis Seminar

Introduction

- Large deviation theory is one of the pillars of probability theory.
- Roughly speaking, large deviation theory concerns with the exponential decline of the probability measures of certain kinds of extreme or tail events.

What are large deviations?

- We start with an example about tossing fair coins.
- We set a Rademacher random variable $X = 1$ if we get a head, we set $X = -1$ if we get a tail.

$$P(X = -1) = \tfrac{1}{2}, \qquad P(X = 1) = \tfrac{1}{2}.$$

Coin flipping

- We toss n fair coins and record the results as $X_1, \dots, X_n$. They are i.i.d. Rademacher random variables.
- **Law of large numbers:** Set $S_n = X_1 + \cdots + X_n$, then

$$\lim_{n\to\infty} S_n/n = 0, \qquad \text{a.s.}$$

- **Central limit theorem:** Study the fluctuation around 0 in the order of $\frac{1}{\sqrt{n}}$.

Coin flipping

- The large deviation theory studies the behavior of $\mathbb{P}\big(\{\frac{S_n}{n} \approx x\}\big)$ for fixed x, as $n \to \infty$.

- As special cases,

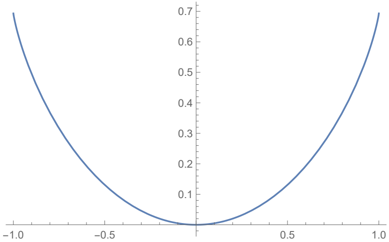
$$\mathbb{P}\Big(\frac{S_n}{n} = -1\Big) = 2^{-n} = e^{-n \log 2},$$
$$\mathbb{P}\Big(\frac{S_n}{n} = 1\Big) = 2^{-n} = e^{-n \log 2}.$$

- For general x,

$$\mathbb{P}\left(\frac{S_n}{n} \approx x\right) \approx e^{-nI_{Rad}(x)},$$

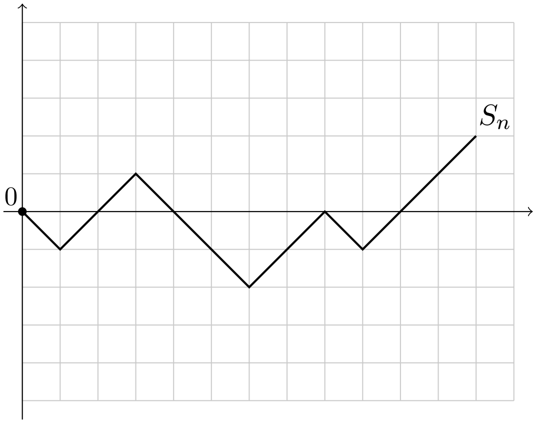
where **the rate function**

$$I_{\text{Rad}}(x) = \begin{cases} \log 2 + \frac{x+1}{2} \log \frac{x+1}{2} + \frac{1-x}{2} \log \frac{1-x}{2} & \text{if } x \in [-1, 1], \\ \infty & \text{else.} \end{cases}$$



Simple random walk

- Let $X_1, \ldots, X_n, \ldots$ be i.i.d. Rademacher random variables. Define the running sum $S_k = \sum_{i=1}^{k} X_i$. Note that we have studied the LDP of S_n in the coin flipping example.



Simple random walk

- By law of large numbers, as $n \to \infty$,

$$\{\frac{S_{nt}}{n}, t \in [0,1]\} \to 0.$$

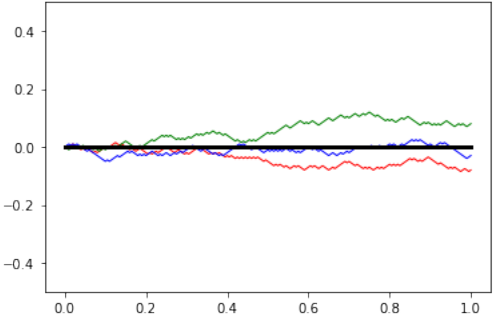


Figure 1: Samplings of simple random walk when $n = 200$.

LDP of simple random walk

- For nice $\gamma \in C([0,1])$ with $\gamma(0) = 0$, we have

$$\mathbb{P}\Big(\{\tfrac{S_{nt}}{n}, t \in [0,1]\} \approx \gamma\Big) \approx e^{-nI_{\mathrm{SRW}}(\gamma)},$$

where

$$I_{\mathrm{SRW}}(\gamma) = \int_0^1 I_{\mathrm{Rad}}\big(\dot{\gamma}(s)\big)ds.$$

Most probable shape

- Conditioned on $\frac{S_n}{n} = x$, find **most probable shape** of $\{\frac{S_{nt}}{n}, t \in [0,1]\}$, when $n \to \infty$?
- The most probable shape refers to the path γ_* where

$$\lim_{n\to\infty} \mathbb{P}\Big(\|\frac{S_{n\cdot}}{n} - \gamma_*\| \leq \varepsilon \,|\, \frac{S_n}{n} = x\Big) = 1, \qquad \forall\, \varepsilon > 0.$$

- Intuitively, the most probable shape is $\{tx, t \in [0,1]\}$.

Most probable shape

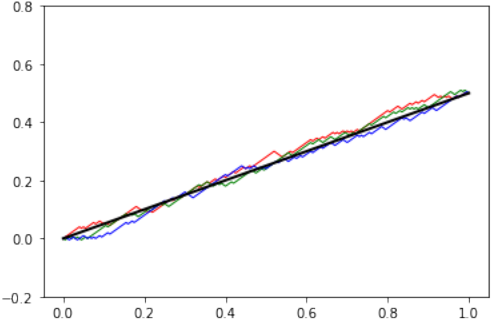


Figure 2: Samplings of simple random walk conditioned on $S_n = 0.5n$ when $n = 200$.

Most probable shape

- How to identify the most probable shape γ_* from the LDP?

$$\gamma_* = \arg\inf \Big\{ I_{\mathrm{SRW}}(\gamma) : \gamma(0) = 0, \gamma(1) = x \Big\}.$$

Note that

$$I_{\mathrm{SRW}}(\gamma) = \int_0^1 I_{\mathrm{Rad}}(\dot{\gamma}(s)) ds.$$

By convexity of I_{Rad}, the minimizer γ_* should be a straight line, $\gamma_*(t) = tx$.

- Define $B_t^k := \frac{S_{kt}}{\sqrt{k}}$. As $k \to \infty$, the process B_t^k converges to a continuous process B_t, which is a **Brownian motion**.

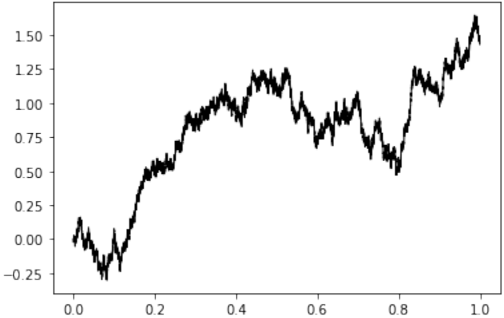


Figure 3: Simulation of B_t^k, when $k = 50000$.

The Brownian motion models the motion of the pollen particles as being moved by individual water molecules. It has the following properties:

- $B(0) = 0$.
- B has both stationary and independent increments.
- $B_t - B_s$ has a normal distribution with mean 0 and variance $|t - s|$.
- The process B_{nt} has the same law as $\sqrt{n}B_t$.

LDP of Brownian motion

- Consider the Brownian motion starting from $B_0 = 0$. We study the LDP of $\frac{B_{n\cdot}}{n}$.
- The process has the same law as $\sqrt{\varepsilon}B_\cdot$, $\varepsilon = \frac{1}{n}$. We have

$$\mathbb{P}\Big(\sqrt{\varepsilon}B_\cdot \approx \gamma\Big) \approx \exp\Big(-\varepsilon^{-1} I_{\mathrm{BM}}(\gamma)\Big),$$

where

$$I_{\mathrm{BM}}(\gamma) = \begin{cases} \int_0^1 \frac{1}{2}\dot{\gamma}(s)^2 ds, & \gamma(0) = 0 \text{ and } \gamma \in H^1([0,1]), \\ \infty & \text{else.} \end{cases}$$

Most probable shape

- Given $\sqrt{\varepsilon} B_1 = y$, what is the most probable shape?

Finding the minimizer

- The most probable shape should be the minimizer of

$$\inf \left\{ I_{\mathrm{BM}}(\gamma) : \gamma(0) = 0, \gamma(1) = y, \gamma \in H^1([0,1]) \right\}$$

- We need to minimize

$$\int_0^1 \dot{\gamma}(s)^2 ds \text{ given } \gamma(0) = 0 \text{ and } \gamma(1) = y.$$

- The minimizer has constant derivative: $\gamma(t) = ty$.

Spatial white noise

- Consider a Brownian motion B. The white noise ξ is the spatial derivative of it, i.e. $\xi(t) = \frac{dB_t}{dt}$.

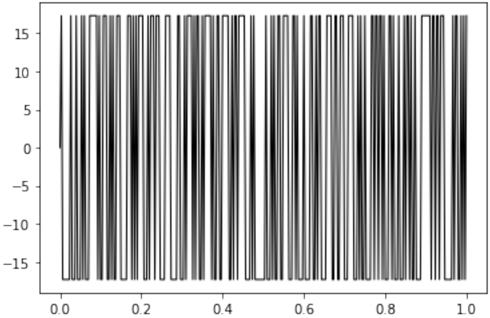


Figure 4: Simulation of spatial white noise

- Because B_t is not differentiable, ξ is a **generalized function**. For any compactly supported g,

$$\int g\xi dt = -\int g'(t)B_t dt.$$

- Informally, ξ is a Gaussian process with correlation

$$\mathbb{E}[\xi(t)\xi(s)] = \delta(t-s).$$

- For $\rho \in L^2(\mathbb{R})$, as $\varepsilon \to 0$,

$$\mathbb{P}\big(\sqrt{\varepsilon}\xi \approx \rho\big) \approx \exp\left(- \tfrac{1}{2}\varepsilon^{-1} \int_{\mathbb{R}} \rho^2 \right).$$

Space-time white noise

- Informally, the space-time white noise is a two-dimensional Gaussian process ξ with correlation function

$$\mathbb{E}[\xi(t,x)\xi(s,y)] = \delta(t-s)\delta(x-y).$$

It can also be defined as the space-time derivative of a **Brownian sheet**.

- The LDP of space-time white noise, for $\rho \in L^2(\mathbb{R}^2)$,

$$\mathbb{P}(\sqrt{\varepsilon}\xi \approx \rho) \approx \exp\Big(-\tfrac{1}{2}\varepsilon^{-1} \int_{\mathbb{R}^2} \rho^2 \Big).$$

  This is an important ingredient for us to study the LDP of stochastic PDE.

Thank you!