Large Deviations for Random Graphs

Honours Year Project Introductory Talk

NG Say-Yao Mark[1] A/P SUN Rongfeng[2]

[1]Student, [2]Supervisor
Department of Mathematics
National University of Singapore (NUS)

October 5, 2016

Introduction

General information about this talk.

- Introduce fundamentals and motivation for chosen topic.
- **Large deviation theory** for **random graph models**.
- Assuming some knowledge in:
  - Probability theory: *probability measure, expected value, almost sure convergence*.
  - Analysis: *limits, metric spaces, measurable functions*.
  - Graph Theory: *edge, vertex, graph homomorphisms*.
- Slides available at:

  `http://mollymr305.github.io`

Large Deviations Theory

Let's begin with *soft* introduction to large deviations theory.

Large deviations theory is a branch of probability theory which, roughly speaking, concerns the study of two things:

- **Probabilities of rare events.**
- **Conditional probabilities of events, *given* that a rare event has occurred.**

The theory of large deviations is also applied in many fields such as statistical physics, statistical hypothesis testing, etc.

A Coin Tossing Example

Let $X_1, \ldots, X_n$ be IID random variables with law

$$\mathbf{P}(X_i = 0) = \mathbf{P}(X_i = 1) = \frac{1}{2}.$$

Denote their sum by $S_n = X_1 + \cdots + X_n$.

The Strong Law of Large Numbers (LLN) tells us that

$$\frac{S_n}{n} \xrightarrow{a.s.} \frac{1}{2}, \text{ as } n \longrightarrow \infty.$$

i.e. "The *empirical mean* converges to $\frac{1}{2}$ with probability 1."

A Coin Tossing Example

We can then ask about the *asymptotic behaviour* of the probabilities rare events, for example

$$\mathbf{P}\left(\frac{S_n}{n} \geq \frac{2}{3}\right) = \mathbf{P}\left(S_n \geq \frac{2}{3}n\right), \text{ as } n \longrightarrow \infty.$$

A Coin Tossing Example

It turns out that,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(S_n \geq \frac{2}{3}n\right) = -\left(\log 2 + \frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right)$$
$$= -\log\left(\frac{2^{5/3}}{3}\right).$$

In other words: the probability of the rare event

$$\mathbf{P}\left(S_n \geq \frac{2}{3}n\right),$$

decays exponentially as $n \longrightarrow \infty$, with rate

$$\frac{2^{5/3}}{3}.$$

A Coin Tossing Example

More generally, for any $\alpha > \frac{1}{2}$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}(S_n \geq \alpha n) = -I(\alpha),$$

where $I : \mathbb{R} \to \overline{\mathbb{R}}$, called the *rate function* is defined by

$$I(z) = \begin{cases} \log 2 + z \log z + (1-z)\log(1-z) & \text{if } z \in [0,1], \\ +\infty & \text{otherwise.} \end{cases}$$

Remark 1. We set $0 \log 0 = 0$ by convention.

Remark 2. Some properties of this rate function include:

- $I(z) \geq 0$ for all $z \in \mathbb{R}$.
- $I(z) = 0$ if and only if $z = \frac{1}{2}$.

Cramer's Theorem for the Empirical Average

A further generalisation of the Coin Tossing Example.

Theorem (Cramer)

Let $\{X_i\}_{i\in\mathbb{N}}$ *be IID* $\mathbb{R}$*-valued random variables such that*

$$\varphi(t) = \mathbf{E}e^{tX_i} < \infty, \text{ for all } t \in \mathbb{R}.$$

Let $S_n = \sum_{i=1}^{n} X_i$*. Then for any* $\alpha > \mathbf{E}X_i$*,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(S_n \geq \alpha n\right) = -I(\alpha),$$

where

$$I(z) = \sup_{t\in\mathbb{R}}\{zt - \log\varphi(t)\}.$$

The Large Deviation Principle

$\star$ Starting point for the *general* large deviations theory.

Let $(\mathcal{X}, d)$ be a Polish space (i.e. separable and complete).

Definition (Good Rate Function)

A function $I : \mathcal{X} \to [0, \infty]$ is said to be a *good rate function* if

D1 $I \not\equiv \infty$.

D2 I is lower semi-continuous.
(iff $\liminf_{n\to\infty} I(x_n) \geq I(x)$ for any $x_n \to x$ in $\mathcal{X}$.)

D3 I has compact level sets.
(i.e. $I^{-1}([-\infty, c])$ is compact for every $c \in \mathbb{R}$)

For any subset $S \subseteq \mathcal{X}$ we also define $I(S) = \inf\{I(s) : s \in S\}$.

The Large Deviation Principle

Definition (LDP)

A sequence of probability measures $\{\mathbf{P}_n\}_{n\in\mathbb{N}}$ on the space $\mathcal{X}$ is said to satisfy the *Large Deviation Principle (LDP)* with rate n and good rate function I if

L1 I is a *good rate function*.

L2 For any *closed* subset $C \subseteq \mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}_n(C) \leq -I(C).$$

L3 For any *open* subset $O \subseteq \mathcal{X}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}_n(O) \geq -I(O).$$

Graph Limit Theory

We first define what is means for a graph sequences to have a limit. *For a graph G, let $V(G)$ and $E(G)$ denote the vertex and edge set respectively.*

- Let $\{G_n\}_{n\in\mathbb{N}}$ be a sequence of simple graphs whose number of vertices tend to infinity.
- For a fixed simple graph H, let $\hom(H, G_n)$ denote the number of homomorphisms from H into G_n.
- The probability that a random map $\phi : V(H) \to V(G_n)$ is a homomorphism is

$$t(H, G_n) = \frac{\hom(H, G_n)}{|V(G_n)|^{|V(H)|}}.$$

Graph Limit Theory

Lovász and Szegedy (2004) proved the following:

Theorem

Suppose $t(H, G_n)$ tends to a limit $t(H)$ for every H. Then there exists a symmetric measurable function

$$f : [0,1] \times [0,1] \to [0,1]$$

such that

$$\lim_{n\to\infty} t(H, G_n) = t(H, f) \stackrel{def.}{=} \int_{[0,1]^k} \prod_{\{i,j\}\in E(H)} f(x_i, x_j) dx_1 \cdots dx_k.$$

Conversely, every symmetric measurable function of the form $f : [0,1] \times [0,1] \to [0,1]$ corresponds to the limit of an appropriate graph sequence.

Graph Limit Theory

- The function f is the 'natural limit object' of $\{G_n\}_{n\in\mathbb{N}}$.
- This gives us a 'working definition'* for a 'graph limit'.

Definition

A sequence of simple graphs $\{G_n\}_{n\in\mathbb{N}}$ whose number of vertices tend to infinity is said to converge to a symmetric measurable function $f : [0,1]\times[0,1] \to [0,1]$ if for every finite simple graph H,

$$\lim_{n\to\infty} t(H, G_n) = t(H, f).$$

The function f is also referred to as a *graphon*.

Erdős-Rényi Random Graphs

- The Erdős-Rényi random graph $G_{n,p}$ is a graph vertex set $\{1, \ldots, n\}$, and the law $\mathbf{P}(\{i, j\}$ is an edge $) = p$ independently for all $1 \leq i < j \leq n$.
- For a given H, note that $t(H, G_{n,p})$ is a random variable.
- It can be shown that

$$t(H, G_{n,p}) \xrightarrow{a.s.} p^{|E(H)|}, \text{ as } n \longrightarrow \infty.$$

- Also, note that the constant graphon $f(x, y) = p$ gives

$$t(H, f) = \int_{[0,1]^k} \prod_{\{i,j\} \in E(H)} f(x_i, x_j) \mathrm{d}x_1 \cdots \mathrm{d}x_k = p^{|E(H)|}.$$

- In other words, the graph sequence $\{G_{n,p}\}_{n \in \mathbb{N}}$ converges to $f \equiv p$ with probability 1.

Large Deviation Principle for Random Graphs: Research

General information about this field of research:

- Large deviation theory for the Erdős-Rényi random graph was developed not too long ago by Chatterjee & Varadhan (2011).
- Other developments include large deviation theory for the Exponential Random Graph Model (ERGM) by Chatterjee & Diaconis (2013).
- However, less has been established with respect to *sparse random graphs*.

Large Deviation Principle for Random Graphs: Research

Some interesting work regarding the Erdős-Rényi random graph:

- For an Erdős-Rényi random graph $G_{n,p}$, let $T_{n,p}$ denote the number of *triangles*.
- It can be shown that $\mathbf{E}\, T_{n,p} = \binom{n}{3} p^3$.
- Consider the *rare* event $E_n = \{ T_{n,p} \geq (1+\delta)\mathbf{E}\, T_{n,p} \}$, for some $\delta > 0$.

? What is the asymptotic behaviour of $\mathbf{P}(E_n)$ as $n \to \infty$.

? Can we characterize the *conditional distribution* of $G_{n,p}$ given that the rare event E_n has occurred.

Goals for Honours Year Project

What I hope to fulfil by the end of this project:

- Better understand results concerning large deviations for:
  - Dense random graphs.
  - Sparse random graphs.
- On a side note: I hope to spend more focus on sparse random graph models.

? Try to answer some open questions in this field.

Questions & Answers

Thank you very much for your attention.

Selected References

1 Frank den Hollander. *Large deviations.* Fields Institute monographs, ISSN 1069-5273; 14.

2 László Lovász and Balázs Szegedy. *Limits of dense graph sequences.* J. Combin. Theory Ser. B, 96 no. 6, 933-957.

3 Sourav Chatterjee. *An introduction to large deviations for random graphs.* To appear in Bull. Amer. Math. Soc.