Asymptotic results for a stochastic model of rumor propagation on finite graphs

Elcio Lebensztayn

Institute of Mathematics, Statistics and Scientific Computation

University of Campinas

ELAM – 2018

Overview of the talk

- **Subject:** The Maki–Thompson model for the propagation of a rumour within a finite homogeneous population – A Large Deviations Principle for the proportion of the population never hearing the rumour.
- **Paper:** A Large Deviations Principle for the Maki–Thompson rumour model. *Journal of Mathematical Analysis and Applications*, 2015.

1. Introduction to large deviations.
2. The Maki–Thompson rumour model.
3. Main result.
4. Ideas of the proof.

Meaning of large deviations

Idea: Asymptotic behaviour of small probabilities on an exponential scale.

$$P(Y_n \in A) \approx \exp\{-b_n I(A)\} \text{ as } n \to \infty,$$

for a sequence $\{Y_n\}$ of random variables, a sequence $\{b_n\}$ of positive numbers with $b_n \to \infty$, and a coefficient $I(A) \geq 0$.

Often one is interested in the probability of large deviations of Y_n, far away from its typical value.

Ex.: Let $\{X_i\}$ be independent and identically distributed random variables with $P(X_i = 0) = P(X_i = 1) = 1/2$. Define $S_n = \sum_{i=1}^n X_i$, $n \geq 1$.

- Weak Law of Large Numbers: For every $x > 0$,

$$P\big(|S_n - n/2| \geq nx\big) \to 0 \text{ as } n \to \infty.$$

  Therefore, the typical value of S_n is $n/2$.

Large deviations · Maki–Thompson model · Main result · Ideas of the proof

Meaning of large deviations

- Central Limit Theorem: For every $x \in \mathbb{R}$,

$$P\big(S_n - n/2 \geq x\sqrt{n}\big) \to 1 - \Phi(x/\sigma) \text{ as } n \to \infty,$$

  where Φ is the distribution function of $Z \sim N(0,1)$, and $\sigma^2 = 1/4$.

  The deviations of S_n from $n/2$ are typically of the order $\sqrt{n}$.

  Consequently, for every $x < 1/2$, we have that

$$P(S_n \leq nx) \to 0 \text{ as } n \to \infty,$$

  and for every $x > 1/2$, we have that $P(S_n \geq nx) \to 0$ as $n \to \infty$.

- Large Deviations: To quantify precisely the exponential decay rate at which these probabilities converge to 0.

  (Useful tool when an approximation of these small probabilities is needed).

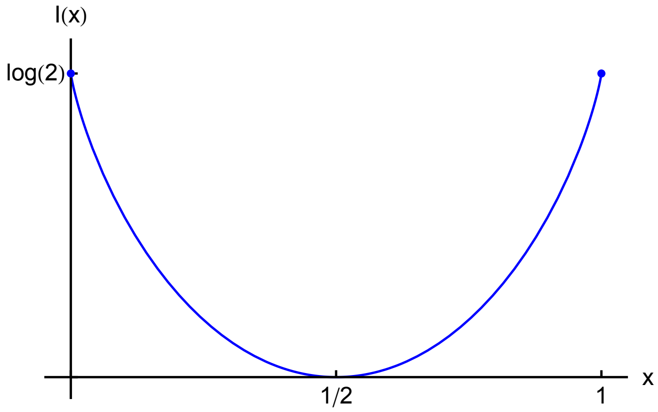
Example – Large Deviations Theorem

For every $x < 1/2$: $\lim_{n\to\infty} -\frac{1}{n} \log P(S_n \le nx) = I(x)$,

and for every $x > 1/2$: $\lim_{n\to\infty} -\frac{1}{n} \log P(S_n \ge nx) = I(x)$,

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

with the usual convention that $0 \log 0 = 0$.



References:

- Klenke (2014), Ch. 23.
- Dembo and Zeitouni (2010).

Maki–Thompson rumour model (1973)

Stochastic model for the propagation of a rumour within a population

- Closed homogeneously mixing population of $N+1$ individuals.

People subdivided into three classes:

ign Ignorants: those not aware of the rumour.

spr Spreaders: who are spreading the rumour.

sti Stiflers: who know the rumour but have ceased communicating it after meeting somebody already informed.

Notation: At time t, $X(t)$ ign, $Y(t)$ spr and $Z(t)$ sti.

Initially: $X(0) = N$, $Y(0) = 1$ and $Z(0) = 0$.

$X(t) + Y(t) + Z(t) = N + 1$ for all $t \ge 0$.

Maki–Thompson rumour model (1973)

Stochastic model for the propagation of a rumour within a population

- The rumour is spread by directed contact of the spreaders with other individuals.
- Process $\{(X(t), Y(t))\}_{t\geq 0}$ is a continuous-time Markov chain with

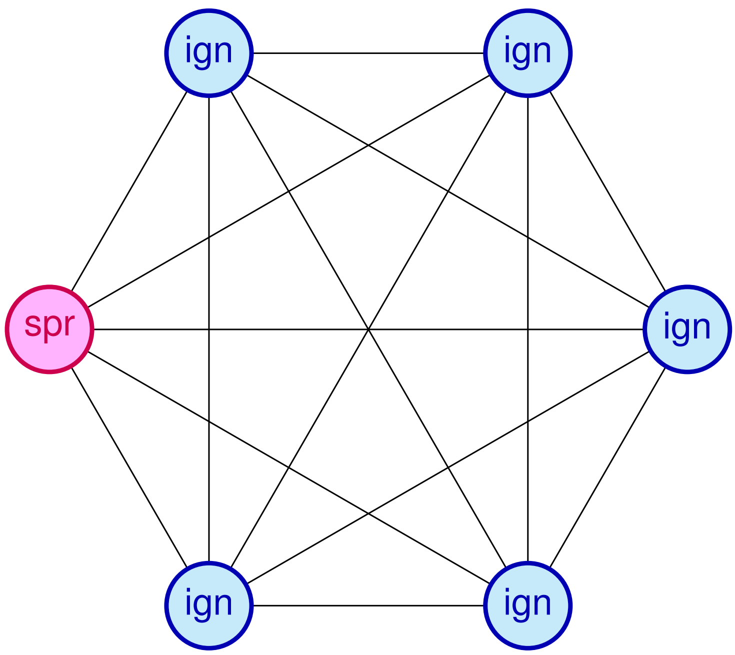
interaction	infinitesimal rate	transition	transformation
spr ign	XY	$(-1,1)$	ign $\Rightarrow$ spr,
spr spr or sti	$Y(N-X)$	$(0,-1)$	spr $\Rightarrow$ sti.

- For instance, the first case means that

$$P(X(t+dt)=i-1, Y(t+dt)=j+1 \,|\, X(t)=i, Y(t)=j) = i\,j\,dt + o(dt).$$

A realization of MT model on K_6 $(N=5)$

Time $t=0$:



References – Rumour models

- Classical models: Daley and Kendall (1965), Maki and Thompson (1973).
- General reference: Daley and Gani (1999), Ch. 5.
- Limit theorems: Sudbury (1985), Watson (1988), Pittel (1990).
- Distribution of final quantities related to the rumour process: Lefèvre and Picard (1994) (using martingales), Pearce (2000) (pgf method).
- Limit theorems for general stochastic rumour models: Lebensztayn et al. (2011a,b), Arruda et al. (2015).

Definitions

For a realization of the Maki–Thompson model on K_{N+1}, we define

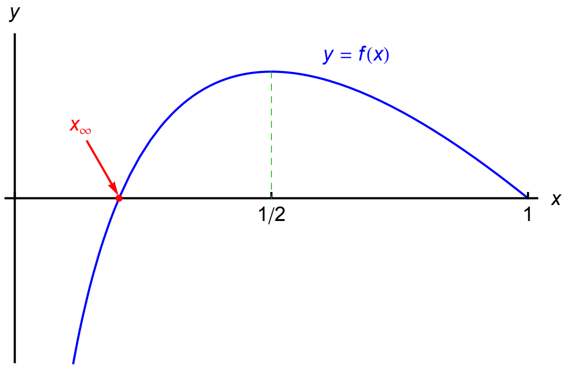
- $\tau^{(N)} = \inf\left\{t : Y^{(N)}(t) = 0\right\}$: Extinction time of the process.
- $X_F^{(N)} = X^{(N)}(\tau^{(N)})$: Final number of ignorant individuals in the population.
- $X_F^{(N)}/N$: Proportion of the originally ignorant individuals who remained ignorant at the end of the process.

Limit Theorems

Law of Large Numbers – Sudbury (1985)

$$\lim_{N\to\infty} \frac{X_F^{(N)}}{N} = x_\infty \quad \text{in probability},$$

where $x_\infty \approx 0.2032$ is the unique root of the function $f(x) = 2(1-x) + \log x$ in the interval $(0,1)$.



For large N, approximately a fifth of the people are not aware of the rumour at the moment that the spreading process stops, with high probability.

Large deviations Maki–Thompson model Main result Ideas of the proof

Limit Theorems

Central Limit Theorem – Watson (1988)

$$\sqrt{N}\left(\frac{X_F^{(N)}}{N} - x_\infty\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2) \text{ as } N\to\infty,$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $\mathcal{N}(0,\sigma^2)$ is the Gaussian distribution with mean zero and variance given by

$$\sigma^2 = \frac{x_\infty(1-x_\infty)}{1-2x_\infty} \approx 0.2727.$$

For large N, the proportion of the population never hearing the rumour is approximately normal with mean x_∞ and variance σ^2/N.

Main result

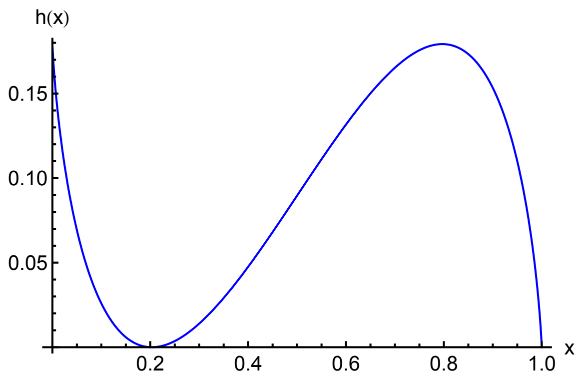
Large Deviations Principle for the ultimate proportion of ignorants

Definitions:

- $v_\infty = 1 - x_\infty \approx 0.7968$ and $\rho = 2 + \log x_\infty + \log(1 - x_\infty) \approx 0.1792$.
- $h : [0, 1) \to \mathbb{R}$ given by

$$h(x) = x \log x + (1 - x)[\rho - \log(1 - x)],$$

with the usual convention that $0 \log 0 = 0$.



1. $h(x_\infty) = 0$.
2. h is $\searrow$ on $[0, x_\infty]$ and $[v_\infty, 1)$, and is $\nearrow$ on $[x_\infty, v_\infty]$.
3. h is strictly convex on $[0, 1/2]$, and strictly concave on $[1/2, 1)$.

Main result

Large Deviations Principle for the ultimate proportion of ignorants

- $H : [0, \infty) \to [0, \infty]$ given by $H(x) = \begin{cases} h(x) & \text{if } 0 \le x < 1, \\ \infty & \text{if } x \ge 1. \end{cases}$

Theorem

Let ν_N be the probability distribution of the random variable $N^{-1} X_F^{(N)}$ on $[0, \infty)$. Then the following conclusions hold.

(a) For each closed set $F \subset [0, \infty)$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \nu_N(F) \le - \inf_{x \in F} H(x).$$

(b) For each open set $G \subset [0, \infty)$,

$$\liminf_{N\to\infty} \frac{1}{N} \log \nu_N(G) \ge - \inf_{x \in G} H(x).$$

Main ideas – Large Deviations Principle

- Formula for the probability mass function $P(X_F^{(N)} = i), i = 0, \ldots, N-1$, in terms of factorials and the enumeration $\{d_n\}$ of certain automata.
- Asymptotic estimates and bounds for $n!$ (Stirling (1730)) and for d_n (Good (1961), Korshunov (1978), Bassino and Nicaud (2007)).
- Some mathematical analysis.

$$\Downarrow$$

Auxiliary results which concern the asymptotic behaviour of normalized logarithms of probabilities of certain events.

Theorem 1

For every $x \in [0, 1)$, we have that

$$\lim_{N\to\infty} -\frac{1}{N} \log P(X_F^{(N)} = \lfloor Nx \rfloor) = \lim_{N\to\infty} -\frac{1}{N} \log P(X_F^{(N)} = \lceil Nx \rceil) = h(x).$$

Large deviations Maki–Thompson model Main result **Ideas of the proof**

Main ideas – Large Deviations Principle

Theorem 2

(a) If $0 \leq x < x_\infty$, then

$$\lim_{N\to\infty} -\frac{1}{N} \log P(X_F^{(N)} \leq Nx) = h(x).$$

(b) If $x_\infty < x < y \leq v_\infty$, then

$$\lim_{N\to\infty} -\frac{1}{N} \log P(Nx \leq X_F^{(N)} \leq Ny) = h(x).$$

(c) If $v_\infty \leq x < y < 1$, then

$$\lim_{N\to\infty} -\frac{1}{N} \log P(Nx \leq X_F^{(N)} \leq Ny) = h(y).$$

The LDP follows from standard arguments of the Large Deviations Theory.

Closed formula for the probability mass function of $X_F^{(N)}$

- For $n \geq 1$, let d_n denote the number of nonisomorphic unlabelled initially connected complete and deterministic automata with n states over a 2-letter alphabet.
- Sequence A006689 in Sloane's *On-Line Encyclopedia of Integer Sequences* – First terms: $1, 12, 216, 5248, 160675, 5931540$.

  Recursive formula (Liskovets (1969)):

$$d_1 = 1 \quad \text{and} \quad d_n = \frac{n^{2n}}{(n-1)!} - \sum_{i=1}^{n-1} \frac{n^{2(n-i)}}{(n-i)!} d_i \quad \text{for } n \geq 2.$$

Theorem 3

For each $i = 0, \ldots, N-1$, we have that $P(X_F^{(N)} = i) = \dfrac{(N-1)!}{i!} \dfrac{d_{N-i}}{N^{2(N-i)}}$.

Asymptotic estimates

$$P(X_F^{(N)} = i) = \frac{(N-1)!}{i!} \frac{d_{N-i}}{N^{2(N-i)}} \quad \text{for } i = 0, \ldots, N-1.$$

- Stirling (1730): $n! \sim \sqrt{2\pi}\, n^{n+1/2}\, e^{-n}$ as $n \to \infty$.
- Korshunov (1978), Bassino and Nicaud (2007): $d_n \sim \kappa\, n \left\{ {2n \atop n} \right\}$,

  where $\kappa = 2 - \dfrac{1}{v_\infty}$ and $\left\{ {2n \atop n} \right\}$ is a Stirling number of the second kind (number of ways of partitioning a set of $2n$ elements into n nonempty subsets).
- Good (1961): $\left\{ {2n \atop n} \right\} \sim \alpha\, \beta^n\, n^{n-1/2}$, where $\alpha = \sqrt{\dfrac{1}{2\pi\,(2\,v_\infty - 1)}}$

  and $\beta = \dfrac{1}{e\, v_\infty\,(1 - v_\infty)}$ (Saddle point method).

References I

Arruda, G. F., Lebensztayn, E., Rodrigues, F. A. and Rodríguez, P. M. (2015).
A process of rumor scotching on finite populations. *R. Soc. Open Sci.* **2** 17 p.

Bassino, F. and Nicaud, C. (2007).
Enumeration and random generation of accessible automata. *Theoret. Comput. Sci.* **381** 86–104.

Daley, D. J. and Gani, J. (1999).
Epidemic Modelling: an Introduction. Cambridge University Press, Cambridge.

Daley, D. J. and Kendall, D. G. (1965).
Stochastic rumours. *J. Inst. Math. Appl.* **1** 42–55.

Dembo, A. and Zeitouni, O. (2010).
Large Deviations Techniques and Applications. Corrected reprint of the second (1998) edition. Springer-Verlag, Berlin.

References II

Good, I. J. (1961).
An asymptotic formula for the differences of the powers at zero. *Ann. Math. Statist.* **32** 249–256.

Klenke, A. (2014).
Probability Theory. A Comprehensive Course. 2nd. ed. Springer-Verlag, London.

Korshunov, D. (1978).
Enumeration of finite automata. *Probl. Kibern.* **34** 5–82, in Russian.

Lebensztayn, E., Machado, F. P. and Rodríguez, P. M. (2011a).
Limit theorems for a general stochastic rumour model. *SIAM J. Appl. Math.* **71** 1476–1486.

Lebensztayn, E., Machado, F. P. and Rodríguez, P. M. (2011b).
On the behaviour of a rumour process with random stifling. *Environ. Modell. Softw.* **26** 517–522.

References III

Lefèvre, C. and Picard, P. (1994).
Distribution of the final extent of a rumour process. *J. Appl. Probab.* **31** 244–249.

Liskovets, V. A. (1969).
The number of initially connected automata. *Kibernetika* **3** 16–19, in Russian. Engl. transl.: *Cybernetics* **4** 259–262.

Maki, D. P. and Thompson, M. (1973).
Mathematical Models and Applications. With emphasis on the social, life, and management sciences. Prentice-Hall, Englewood Cliffs, N. J.

Pearce, C. E. M. (2000).
The exact solution of the general stochastic rumour. *Math. Comput. Modelling* **31** 289–298.

Pittel, B. (1990).
On a Daley–Kendall model of random rumours. *J. Appl. Probab.* **27** 14–27.

References IV

Sloane, N. J. A. (2010).
The On-Line Encyclopedia of Integer Sequences. Published electronically at http://oeis.org, Sequence A006689.

Stirling, J. (1730).
Methodus Differentialis: Sive Tractatus de Summatione et Interpolatione Serierum Infinitarum. Typis Gul. Bowyer, London.

Sudbury, A. (1985).
The proportion of the population never hearing a rumour. *J. Appl. Probab.* **22** 443–446.

Watson, R. (1988).
On the size of a rumour. *Stochastic Process. Appl.* **27** 141–149.