Advances in Mathematics 299 (2016) 396–450

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim

Nonlinear large deviations

Sourav Chatterjee [*,1], Amir Dembo [2]

Department of Statistics, Stanford University, Sequoia Hall, 390 Serra Mall, Stanford, CA 94305, United States

A R T I C L E   I N F O

Article history:
Received 24 September 2015
Received in revised form 29 April 2016
Accepted 23 May 2016
Available online 2 June 2016
Communicated by Charles Fefferman

MSC:
60F10
05C80
60C05
05A20

Keywords:
Large deviations
Sparse random graphs
Regularity lemma
Arithmetic progressions
Exponential random graph models
Concentration of measure

A B S T R A C T

We present a general technique for computing large deviations of nonlinear functions of independent Bernoulli random variables. The method is applied to compute the large deviation rate functions for subgraph counts in sparse random graphs. Previous technology, based on Szemerédi's regularity lemma, works only for dense graphs. Applications are also made to exponential random graphs and three-term arithmetic progressions in random sets of integers.

* Corresponding author.
E-mail addresses: souravc@stanford.edu (S. Chatterjee), adembo@stanford.edu (A. Dembo).

[1] Sourav Chatterjee's research partially supported by NSF grant DMS-1441513.
[2] Amir Dembo's research partially supported by NSF grant DMS-1106627.

http://dx.doi.org/10.1016/j.aim.2016.05.017

1. Introduction

1.1. A motivating example

Let $G(N,p)$ be the Erdős–Rényi random graph on N vertices with edge probability p, that is, the classical model where any two vertices are connected by an edge with probability p, independent of all else. Let T denote the number of triangles in this graph. It has been an open question in the random graph literature for a long time [23] to determine the behavior of the upper tail of T, that is, probabilities of the type $\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))$. The main difficulty with this problem, and the reason why it may be appealing to a probabilist, is that the standard tools from concentration of measure and other probability inequalities do not seem to work so well in this setting, in spite of the fact that the number of triangles in an Erdős–Rényi graph is simply a degree three polynomial of independent Bernoulli random variables.

After a series of successively improving suboptimal results by many authors over many years, a big advance was made by Kim and Vu [29] and simultaneously by Janson et al. [22] in 2004 who showed that if $p \geq N^{-1}\log N$, then

$$\exp(-c_1(\delta)N^2p^2\log(1/p)) \leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \leq \exp(-c_2(\delta)N^2p^2)\,,$$

where $c_1(\delta)$ and $c_2(\delta)$ are constants depending on δ only.

Several years later, the logarithmic discrepancy between the exponents on the two sides was removed by Chatterjee [12] and independently by DeMarco and Kahn [18,19], where it was shown that when $p \geq N^{-1}\log N$,

$$\begin{aligned}\exp(-c_1(\delta)N^2p^2\log(1/p)) &\leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \\ &\leq \exp(-c_2(\delta)N^2p^2\log(1/p))\,.\end{aligned}$$

This still left open the question of determining the dependence of the exponent on δ. When p is fixed and N tends to infinity, the problem was solved in 2011 by Chatterjee and Varadhan [16], confirming a conjecture from an unpublished manuscript of Bolthausen, Comets and Dembo [4]. In [16], it was shown that for fixed $p \in (0,1)$ and $\delta > 0$,

$$\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) = \exp(-c(\delta,p)N^2(1-o(1))) \tag{1.1}$$

as $N \to \infty$, where

$$c(\delta,p) = \frac{1}{2}\inf_f\{I_p(f) : T(f) \geq (1+\delta)p^3\}\,, \tag{1.2}$$

where $f : [0,1]^2 \to [0,1]$ is any Lebesgue measurable function that satisfies $f(x,y) = f(y,x)$ for all x and y,