

LARGE DEVIATIONS AND EXACT ASYMPTOTICS FOR CONSTRAINED EXPONENTIAL RANDOM GRAPHS

MEI YIN

December 19, 2014

ABSTRACT. We present a technique for approximating generic normalization constants subject to constraints. The method is then applied to derive the exact asymptotics for the conditional normalization constant of constrained exponential random graphs.

1. INTRODUCTION

In a recent work [2], Chatterjee and Dembo presented a general technique for computing large deviations of nonlinear functions of independent Bernoulli random variables. In detail, let f be a function from $[0, 1]^n$ to \mathbb{R} , they considered a generic normalization constant of the form

$$F = \log \sum_{x \in \{0,1\}^n} e^{f(x)} \quad (1.1)$$

and investigated conditions on f under which the approximation

$$F = \sup_{x \in [0,1]^n} (f(x) - I(x)) + \text{lower order terms} \quad (1.2)$$

is valid, where $I(x) = \sum_{i=1}^n I(x_i)$ and

$$I(x_i) = \sum_{i=1}^n (x_i \log x_i + (1 - x_i) \log(1 - x_i)). \quad (1.3)$$

The sufficient condition they came up with consists of two parts. They first assumed that f is a twice continuously differentiable function on $[0, 1]^n$ and introduced some shorthand notation. Let $\|\cdot\|$ denote the supremum norm. For each i and j , let

$$f_i = \frac{\partial f}{\partial x_i} \text{ and } f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (1.4)$$

and define $a = \|f\|$, $b_i = \|f_i\|$, and $c_{ij} = \|f_{ij}\|$. In addition to this minor smoothness condition on the function f , they further assumed that the gradient vector $\nabla f(x) = (\partial f / \partial x_1, \dots, \partial f / \partial x_n)$ satisfies a low complexity gradient condition: For any $\epsilon > 0$, there is a finite subset of \mathbb{R}^n denoted by $\mathcal{D}(\epsilon)$ such that for all $x \in [0, 1]^n$, there exists $d = (d_1, \dots, d_n) \in \mathcal{D}(\epsilon)$ with

$$\sum_{i=1}^n (f_i(x) - d_i)^2 \leq n\epsilon^2. \quad (1.5)$$

2000 *Mathematics Subject Classification.* 60F10, 05C80, 60C05.

Key words and phrases. large deviations, normalization constants, exponential random graphs.

Mei Yin's research was partially supported by NSF grant DMS-1308333.

Theorem 1.1 (Theorem 1.5 in [2]). *Let F , a , b_i , c_{ij} , and $\mathcal{D}(\epsilon)$ be defined as above. Let I be defined as in (1.3). Then for any $\epsilon > 0$, F satisfies the upper bound*

$$F \leq \sup_{x \in [0,1]^n} (f(x) - I(x)) + \text{complexity term} + \text{smoothness term}, \quad (1.6)$$

where

$$\text{complexity term} = \frac{1}{4} \left(n \sum_{i=1}^n b_i^2 \right)^{1/2} \epsilon + 3n\epsilon + \log |\mathcal{D}(\epsilon)|, \text{ and} \quad (1.7)$$

$$\begin{aligned} \text{smoothness term} &= 4 \left(\sum_{i=1}^n (ac_{ii} + b_i^2) + \frac{1}{4} \sum_{i,j=1}^n (ac_{ij}^2 + b_i b_j c_{ij} + 4b_i c_{ij}) \right)^{1/2} \\ &+ \frac{1}{4} \left(\sum_{i=1}^n b_i^2 \right)^{1/2} \left(\sum_{i=1}^n c_{ii}^2 \right)^{1/2} + 3 \sum_{i=1}^n c_{ii} + \log 2. \end{aligned} \quad (1.8)$$

Moreover, F satisfies the lower bound

$$F \geq \sup_{x \in [0,1]^n} (f(x) - I(x)) - \frac{1}{2} \sum_{i=1}^n c_{ii}. \quad (1.9)$$

Chatterjee and Dembo [2] then applied this general result in several different settings and in particular to derive the exact asymptotics for the normalization constant of exponential random graphs. Let s be a positive integer. We recall the definition of an s -parameter family of exponential random graphs. Let H_1, \dots, H_s be fixed finite simple graphs (“simple” means undirected, with no loops or multiple edges). By convention, we take H_1 to be a single edge. Let ζ_1, \dots, ζ_s be s real parameters and let N be a positive integer. Consider the set \mathcal{G}_N of all simple graphs G_N on N vertices. Let $\text{hom}(H_i, G_N)$ denote the number of homomorphisms (edge-preserving vertex maps) from the vertex set $V(H_i)$ into the vertex set $V(G_N)$ and $t(H_i, G_N)$ denote the homomorphism density of H_i in G_N ,

$$t(H_i, G_N) = \frac{|\text{hom}(H_i, G_N)|}{|V(G_N)|^{|V(H_i)|}}. \quad (1.10)$$

By an s -parameter family of exponential random graphs we mean a family of probability measures \mathbb{P}_N^ζ on \mathcal{G}_N defined by, for $G_N \in \mathcal{G}_N$,

$$\mathbb{P}_N^\zeta(G_N) = \exp \left(N^2 \left(\zeta_1 t(H_1, G_N) + \dots + \zeta_s t(H_s, G_N) - \psi_N^\zeta \right) \right), \quad (1.11)$$

where ψ_N^ζ is the normalization constant,

$$\psi_N^\zeta = \frac{1}{N^2} \log \sum_{G_N \in \mathcal{G}_N} \exp \left(N^2 \left(\zeta_1 t(H_1, G_N) + \dots + \zeta_s t(H_s, G_N) \right) \right). \quad (1.12)$$

These exponential models are widely used to characterize the structure and behavior of real-world networks as they are able to predict the global structure of the networked system based on a set of tractable local features. For practitioners, one of the key objectives while studying this model is to evaluate the normalization constant ψ_N^ζ of the probability measure \mathbb{P}_N^ζ , since averages of various quantities of interest may be obtained by differentiating ψ_N^ζ with respect to appropriate parameters. Computation of ψ_N^ζ is also important in statistics because it is crucial

for carrying out maximum likelihood estimates and Bayesian inference of unknown parameters. Based on a large deviation principle for Erdős-Rényi graphs established in Chatterjee and Varadhan [4], Chatterjee and Diaconis [3] developed an asymptotic approximation for ψ_N^ζ and connected the occurrence of a phase transition in the exponential model with the non-analyticity of ψ_N^ζ . Further investigations quickly followed, see for example [1, 6, 7, 8, 9, 10, 11]. However, since the approximation relies on Szemerédi's lemma, the error bound on ψ_N^ζ is of the order of some negative power of $\log^* N$ and this method is also not applicable for sparse exponential random graphs.

To improve on the approximation, Chatterjee and Dembo [2] utilized Theorem 1.1. They introduced an equivalent definition of the homomorphism density so that the normalization constant for exponential random graphs (1.12) takes the same form as the generic normalization constant (1.1). This new notion of the homomorphism density is denoted by $t(H, x)$ and is constructed in the following way. Let k be a positive integer and let H be a finite simple graph on the vertex set $[k] = \{1, \dots, k\}$. Let E be the set of edges of H and let $m = |E|$. Let N be another positive integer and let $n = \binom{N}{2}$. Index the elements of $[0, 1]^n$ as $x = (x_{ij})_{1 \leq i < j \leq N}$ with the understanding that if $i < j$, then x_{ji} is the same as x_{ij} , and for all i , $x_{ii} = 0$. Let $t(H, x) = T(x)/N^2$, where $T : [0, 1]^n \rightarrow \mathbb{R}$ is defined as

$$T(x) = \frac{1}{N^{k-2}} \sum_{q \in [N]^k} \prod_{\{l, l'\} \in E} x_{ql'}. \quad (1.13)$$

For any graph G_N , if $x_{ij} = 1$ means there is an edge between the vertices i and j and $x_{ij} = 0$ means there is no edge, then $t(H, x) = t(H, G_N)$, where $t(H, G_N)$ is the homomorphism density defined by (1.10). Furthermore, if we let G_x denote the simple graph whose edges are independent, and edge (i, j) is present with probability x_{ij} and absent with probability $1 - x_{ij}$, then this newly defined homomorphism density $t(H, x)$ gives the expected value of $t(H, G_x)$. Chatterjee and Dembo checked that $T(x)$ satisfies both the smoothness condition and the low complexity gradient condition as assumed in Theorem 1.1. In detail, they showed in Lemmas 5.1 and 5.2 of [2] that

$$\|T\| \leq N^2, \quad \left\| \frac{\partial T}{\partial x_{ij}} \right\| \leq 2m, \quad (1.14)$$

$$\left\| \frac{\partial^2 T}{\partial x_{ij} \partial x_{i'j'}} \right\| \leq \begin{cases} 4m(m-1)N^{-1}, & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3; \\ 4m(m-1)N^{-2}, & \text{if } |\{i, j, i', j'\}| = 4, \end{cases} \quad (1.15)$$

and for any $\epsilon > 0$,

$$|\mathcal{D}_T(\epsilon)| \leq \exp\left(\frac{cm^4 k^4 N}{\epsilon^4} \log \frac{Cm^4 k^4}{\epsilon^4}\right), \quad (1.16)$$

where c and C are universal constants. By taking $f(x) = \zeta_1 T_1(x) + \dots + \zeta_s T_s(x)$ in Theorem 1.1, they then gave a concrete error bound for the normalization constant ψ_N^ζ , which is seen to be F/N^2 in this alternative interpretation of (1.1). This error bound is significantly better than the negative power of $\log^* N$ and allows a small degree of sparsity for ζ_i .

Theorem 1.2 (Theorem 1.6 in [2]). *Let s be a positive integer and H_1, \dots, H_s be fixed finite simple graphs. Let N be another positive integer and let $n = \binom{N}{2}$. Define T_1, \dots, T_s accordingly as in the above paragraph. Let ζ_1, \dots, ζ_s be s real parameters and define ψ_N^ζ as in (1.12). Let*

$f(x) = \zeta_1 T_1(x) + \cdots + \zeta_s T_s(x)$, $B = 1 + |\zeta_1| + \cdots + |\zeta_s|$, and I be defined as in (1.3). Then

$$\begin{aligned} -cBN^{-1} &\leq \psi_N^\zeta - \sup_{x \in [0,1]^n} \frac{f(x) - I(x)}{N^2} \\ &\leq CB^{8/5} N^{-1/5} (\log N)^{1/5} \left(1 + \frac{\log B}{\log N}\right) + CB^2 N^{-1/2}, \end{aligned} \quad (1.17)$$

where c and C are constants that may depend only on H_1, \dots, H_s .

The unconstrained exponential family of random graphs (1.11) introduced above assumes no prior knowledge of the graph before sampling, but in many situations partial information of the graph is already known beforehand. To tackle this issue, we studied the constrained exponential random graph model in [5]. For clarity, we assume that the edge density of the graph is approximately known to be e , though the argument runs through without much modification for more general constraints. Take $t > 0$. The conditional normalization constant $\psi_{N,t}^{e,\zeta}$ is defined analogously to the normalization constant for the unconstrained exponential random graph model,

$$\psi_{N,t}^{e,\zeta} = \frac{1}{N^2} \log \sum_{G_N \in \mathcal{G}_N: |e(G_N) - e| \leq t} \exp(N^2 (\zeta_1 t(H_1, G_N) + \cdots + \zeta_s t(H_s, G_N))), \quad (1.18)$$

the difference being that we are only taking into account graphs G_N whose edge density $e(G_N)$ is within a t neighborhood of e . Correspondingly, the associated conditional probability measure $\mathbb{P}_{N,t}^{e,\zeta}(G_N)$ is given by

$$\mathbb{P}_{N,t}^{e,\zeta}(G_N) = \exp\left(N^2 \left(\zeta_1 t(H_1, G_N) + \cdots + \zeta_s t(H_s, G_N) - \psi_{N,t}^{e,\zeta}\right)\right) \mathbb{1}_{|e(G_N) - e| \leq t}. \quad (1.19)$$

Using the large deviation principle established in Chatterjee and Varadhan [4] and Chatterjee and Diaconis [3], we developed an asymptotic approximation for $\psi_{N,t}^{e,\zeta}$ [5]. Nevertheless, this approximation suffers from the same problem: the error bound on $\psi_{N,t}^{e,\zeta}$ is of the order of some negative power of $\log^* N$ and is not applicable in the sparse setting.

As shown in [5], just like the unconstrained normalization constant ψ_N^ζ , the conditional normalization constant $\psi_{N,t}^{e,\zeta}$ encodes essential information about the constrained exponential model (1.19) and helps to predict the structure and behavior of a typical random graph drawn from this model. Seeing the power of nonlinear large deviations in deriving a concrete error bound for ψ_N^ζ , we naturally wonder if it is possible to likewise obtain a better estimate for $\psi_{N,t}^{e,\zeta}$. The following sections will be dedicated towards this goal. Due to the imposed constraint, instead of working with a generic normalization constant of the form (1.1) as in Chatterjee and Dembo [2], we will work with a generic conditional normalization constant in Theorem 2.1 and then apply this result to derive a concrete error bound for the conditional normalization constant of constrained exponential random graphs in Theorems 3.1 and 3.2.

2. NONLINEAR LARGE DEVIATIONS

Let f and h be two continuously differentiable functions from $[0, 1]^n$ to \mathbb{R} . Assume that f and h satisfy both the smoothness condition and the low complexity gradient condition described at the beginning of this paper. Let a, b_i, c_{ij} be the supremum norms of f and let $\alpha, \beta_i, \gamma_{ij}$ be the corresponding supremum norms of h . For any $\epsilon > 0$, let $\mathcal{D}_f(\epsilon)$ and $\mathcal{D}_h(\epsilon)$ be finite subsets

of \mathbb{R}^n associated with the gradient vectors of f and h respectively. Take $t > 0$. Consider a generic conditional normalization constant of the form

$$F^c = \log \sum_{x \in \{0,1\}^n: |h(x)| \leq tn} e^{f(x)}. \quad (2.1)$$

Theorem 2.1. *Let F^c , a , b_i , c_{ij} , α , β_i , γ_{ij} , $\mathcal{D}_f(\epsilon)$, and $\mathcal{D}_h(\epsilon)$ be defined as above. Let I be defined as in (1.3). Let $K = \log 2 + 2a/n$. Then for any $\delta > 0$ and $\epsilon > 0$, F^c satisfies the upper bound*

$$F^c \leq \sup_{x \in [0,1]^n: |h(x)| \leq (t+\delta)n} (f(x) - I(x)) + \text{complexity term} + \text{smoothness term}, \quad (2.2)$$

where

$$\text{complexity term} = \frac{1}{4} \left(n \sum_{i=1}^n m_i^2 \right)^{1/2} \epsilon + 3n\epsilon + \log \left(\frac{12K \left(\frac{1}{n} \sum_{i=1}^n \beta_i^2 \right)^{1/2}}{\delta\epsilon} \right) \quad (2.3)$$

+ $\log |\mathcal{D}_f(\epsilon/3)| + \log |\mathcal{D}_h((\delta\epsilon)/(6K))|$, and

$$\text{smoothness term} = 4 \left(\sum_{i=1}^n (\ln n_{ii} + m_i^2) + \frac{1}{4} \sum_{i,j=1}^n (\ln n_{ij}^2 + m_i m_j n_{ij} + 4m_i n_{ij}) \right)^{1/2} \quad (2.4)$$

$$+ \frac{1}{4} \left(\sum_{i=1}^n m_i^2 \right)^{1/2} \left(\sum_{i=1}^n n_{ii}^2 \right)^{1/2} + 3 \sum_{i=1}^n n_{ii} + \log 2,$$

where

$$l = a + nK, \quad (2.5)$$

$$m_i = b_i + \frac{2K\beta_i}{\delta}, \quad (2.6)$$

$$n_{ij} = c_{ij} + \frac{2K\gamma_{ij}}{\delta} + \frac{6K\beta_i\beta_j}{n\delta^2}. \quad (2.7)$$

Moreover, F^c satisfies the lower bound

$$F^c \geq \sup_{x \in [0,1]^n: |h(x)| \leq (t-\delta_0)n} (f(x) - I(x)) - \epsilon_0 n - \eta_0 n - \log 2, \quad (2.8)$$

where

$$\delta_0 = \frac{\sqrt{6}}{n} \left(\sum_{i=1}^n (\alpha\gamma_{ii} + \beta_i^2) \right)^{1/2}, \quad (2.9)$$

$$\epsilon_0 = 2\sqrt{\frac{6}{n}}, \quad (2.10)$$

$$\eta_0 = \frac{\sqrt{6}}{n} \left(\sum_{i=1}^n (ac_{ii} + b_i^2) \right)^{1/2}. \quad (2.11)$$

Proof of the upper bound. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a function that is twice continuously differentiable, non-decreasing, and satisfies $g(x) = -1$ if $x \leq -1$ and $g(x) = 0$ if $x \geq 0$. Let $L_1 = \|g'\|$ and $L_2 = \|g''\|$. Chatterjee and Dembo [2] described one such g :

$$g(x) = 10(x+1)^3 - 15(x+1)^4 + 6(x+1)^5 - 1, \quad (2.12)$$

which gives $L_1 \leq 2$ and $L_2 \leq 6$. Define

$$\psi(x) = Kg((t - |x|)/\delta). \quad (2.13)$$

Then clearly $\psi(x) = -K$ if $|x| \geq t + \delta$, $\psi(x) = 0$ if $|x| \leq t$, and $\psi(x)$ is non-decreasing for $-(t + \delta) \leq x \leq -t$ and non-increasing for $t \leq x \leq t + \delta$. We also have

$$\|\psi\| \leq K, \quad \|\psi'\| \leq \frac{2K}{\delta}, \quad \|\psi''\| \leq \frac{6K}{\delta^2}. \quad (2.14)$$

Let $e(x) = n\psi(h(x)/n)$. The plan is to apply Theorem 1.1 to the function $f + e$ instead of f only. Note that

$$\sum_{x \in \{0,1\}^n: |h(x)| \leq tn} e^{f(x)} \leq \sum_{x \in \{0,1\}^n} e^{f(x)+e(x)}. \quad (2.15)$$

We estimate $f(x) + e(x) - I(x)$ over $[0, 1]^n$. There are three cases.

- If $|h(x)| \leq tn$, then

$$f(x) + e(x) - I(x) = f(x) - I(x) \leq \sup_{x \in [0,1]^n: |h(x)| \leq (t+\delta)n} (f(x) - I(x)). \quad (2.16)$$

- If $|h(x)| \geq (t + \delta)n$, then

$$f(x) + e(x) - I(x) = f(x) - nK - I(x) \leq a + n \log 2 - nK \leq -a \leq \sup_{x \in [0,1]^n: |h(x)| \leq (t+\delta)n} (f(x) - I(x)). \quad (2.17)$$

- If $|h(x)| = (t + \delta')n$ for some $0 < \delta' < \delta$, then

$$f(x) + e(x) - I(x) \leq f(x) - I(x) \leq \sup_{x \in [0,1]^n: |h(x)| \leq (t+\delta)n} (f(x) - I(x)). \quad (2.18)$$

This shows that

$$\sup_{x \in [0,1]^n} (f(x) + e(x) - I(x)) \leq \sup_{x \in [0,1]^n: |h(x)| \leq (t+\delta)n} (f(x) - I(x)). \quad (2.19)$$

We check the smoothness condition for $f + e$ first. Note that

$$\|f + e\| \leq a + nK = l, \quad (2.20)$$

and for any i ,

$$\left\| \frac{\partial(f + e)}{\partial x_i} \right\| \leq b_i + \frac{2K\beta_i}{\delta} = m_i, \quad (2.21)$$

and for any i, j ,

$$\left\| \frac{\partial^2(f + e)}{\partial x_i \partial x_j} \right\| \leq c_{ij} + \frac{2K\gamma_{ij}}{\delta} + \frac{6K\beta_i\beta_j}{n\delta^2} = n_{ij}. \quad (2.22)$$

Next we check the low complexity gradient condition for $f + e$. Let

$$\epsilon' = \frac{\epsilon}{3\|\psi'\|} \text{ and } \tau = \frac{\epsilon}{3\left(\frac{1}{n} \sum_{i=1}^n \beta_i^2\right)^{1/2}}. \quad (2.23)$$

Define

$$\mathcal{D}(\epsilon) = \{d^f + \theta d^h : d^f \in \mathcal{D}_f(\epsilon/3), d^h \in \mathcal{D}_h(\epsilon'),$$

$$\text{and } \theta = j\tau \text{ for some integer } -\|\psi'\|/\tau < j < \|\psi'\|/\tau\}. \quad (2.24)$$

Note that

$$|\mathcal{D}(\epsilon)| \leq \frac{2\|\psi'\|}{\tau} |\mathcal{D}_f(\epsilon/3)| |\mathcal{D}_h(\epsilon')|. \quad (2.25)$$

Let $e_i = \partial e / \partial x_i$. Take any $x \in [0, 1]^n$ and choose $d^f \in \mathcal{D}_f(\epsilon/3)$ and $d^h \in \mathcal{D}_h(\epsilon')$. Choose an integer j between $-\|\psi'\|/\tau$ and $\|\psi'\|/\tau$ such that $|\psi'(h(x)/n) - j\tau| \leq \tau$. Let $d = d^f + j\tau d^h$ so that $d \in \mathcal{D}(\epsilon)$. Then

$$\begin{aligned} \sum_{i=1}^n (f_i(x) + e_i(x) - d_i)^2 &= \sum_{i=1}^n \left((f_i(x) - d_i^f) + (\psi'(h(x)/n)h_i(x) - j\tau d_i^h) \right)^2 \\ &\leq 3 \sum_{i=1}^n (f_i(x) - d_i^f)^2 + 3(\psi'(h(x)/n) - j\tau)^2 \sum_{i=1}^n h_i(x)^2 + 3\|\psi'\|^2 \sum_{i=1}^n (h_i(x) - d_i^h)^2 \\ &\leq \frac{1}{3}n\epsilon^2 + 3\tau^2 \sum_{i=1}^n \beta_i^2 + 3\|\psi'\|^2 n\epsilon'^2 = n\epsilon^2. \end{aligned} \quad (2.26)$$

Thus $\mathcal{D}(\epsilon)$ is a finite subset of \mathbb{R}^n associated with the gradient vector of $f + e$. The proof is completed by applying Theorem 1.1. \square

Proof of the lower bound. Fix any $y \in [0, 1]^n$ such that $|h(y)| \leq (t - \delta_0)n$. Let $Y = (Y_1, \dots, Y_n)$ be a random vector with independent components, where each Y_i is a Bernoulli(y_i) random variable. Let $Y^{(i)}$ be the random vector $(Y_1, \dots, Y_{i-1}, 0, Y_{i+1}, \dots, Y_n)$. Let

$$A_1 = \{x \in \{0, 1\}^n : |h(x)| \leq tn\}, \quad (2.27)$$

$$A_2 = \{x \in \{0, 1\}^n : |g(x, y) - I(y)| \leq \epsilon_0 n\}, \quad (2.28)$$

$$A_3 = \{x \in \{0, 1\}^n : |f(x) - f(y)| \leq \eta_0 n\}. \quad (2.29)$$

Let $A = A_1 \cap A_2 \cap A_3$. Then

$$\begin{aligned} \sum_{x \in \{0, 1\}^n : |h(x)| \leq tn} e^{f(x)} &= \sum_{x \in A_1} e^{f(x) - g(x, y) + g(x, y)} \\ &\geq \sum_{x \in A} e^{f(x) - g(x, y) + g(x, y)} \\ &\geq e^{f(y) - I(y) - (\epsilon_0 + \eta_0)n} \mathbb{P}(Y \in A). \end{aligned} \quad (2.30)$$

We first consider $\mathbb{P}(Y \in A_1)$. Let $U = h(Y) - h(y)$. For $t \in [0, 1]$ and $x \in [0, 1]^n$ define $u_i(t, x) = h_i(tx + (1-t)y)$. Note that

$$U = \int_0^1 \sum_{i=1}^n (Y_i - y_i) u_i(t, Y) dt, \quad (2.31)$$

which implies

$$\mathbb{E}(U^2) = \int_0^1 \sum_{i=1}^n \mathbb{E}((Y_i - y_i) u_i(t, Y) U) dt. \quad (2.32)$$

Let $U_i = h(Y^{(i)}) - h(y)$ so that $Y^{(i)}$ and U_i are functions of random variables $(Y_j)_{j \neq i}$ only. By the independence of Y_i and $(Y^{(i)}, U_i)$, we have

$$\mathbb{E}((Y_i - y_i)u_i(t, Y^{(i)})U_i) = 0. \quad (2.33)$$

Therefore

$$\begin{aligned} |\mathbb{E}((Y_i - y_i)u_i(t, Y)U)| &\leq \mathbb{E}|((u_i(t, Y) - u_i(t, Y^{(i)}))U)| + \mathbb{E}|u_i(t, Y^{(i)})(U - U_i)| \quad (2.34) \\ &\leq \left\| \frac{\partial u_i}{\partial x_i} \right\| \|U\| + \|u_i\| \|U - U_i\| \\ &\leq 2\alpha t \gamma_{ii} + \beta_i^2. \end{aligned}$$

This gives

$$\mathbb{P}(Y \in A_1^c) \leq \mathbb{P}(|U| \geq \delta_0 n) \leq \frac{\mathbb{E}(U^2)}{\delta_0^2 n^2} \leq \frac{\sum_{i=1}^n (\alpha \gamma_{ii} + \beta_i^2)}{\delta_0^2 n^2} = \frac{1}{6}. \quad (2.35)$$

Next we consider $\mathbb{P}(Y \in A_2)$. Note that

$$\mathbb{E}(g(Y, y)) = I(y) \quad (2.36)$$

and

$$\begin{aligned} \text{Var}(g(Y, y)) &= \sum_{i=1}^n \text{Var}(Y_i \log y_i + (1 - Y_i) \log(1 - y_i)) \quad (2.37) \\ &= \sum_{i=1}^n y_i(1 - y_i) \left(\log \frac{y_i}{1 - y_i} \right)^2. \end{aligned}$$

For $x \in [0, 1]$, since $|\sqrt{x} \log x| \leq 1$, we have

$$x(1 - x) \left(\log \frac{x}{1 - x} \right)^2 \leq (|\sqrt{x} \log x| + |\sqrt{1 - x} \log(1 - x)|)^2 \leq 4. \quad (2.38)$$

Therefore

$$\mathbb{P}(Y \in A_2^c) \leq \mathbb{P}(|g(Y, y) - I(y)| \geq \epsilon_0 n) \leq \frac{\text{Var}(g(Y, y))}{\epsilon_0^2 n^2} \leq \frac{4}{\epsilon_0^2 n} = \frac{1}{6}. \quad (2.39)$$

Finally we consider $\mathbb{P}(Y \in A_3)$. Let $V = f(Y) - f(y)$. For $t \in [0, 1]$ and $x \in [0, 1]^n$ define $v_i(t, x) = f_i(tx + (1 - t)y)$. Note that

$$V = \int_0^1 \sum_{i=1}^n (Y_i - y_i) v_i(t, Y) dt, \quad (2.40)$$

which implies

$$\mathbb{E}(V^2) = \int_0^1 \sum_{i=1}^n \mathbb{E}((Y_i - y_i) v_i(t, Y) V) dt. \quad (2.41)$$

Let $V_i = f(Y^{(i)}) - f(y)$ so that $Y^{(i)}$ and V_i are functions of random variables $(Y_j)_{j \neq i}$ only. By the independence of Y_i and $(Y^{(i)}, V_i)$, we have

$$\mathbb{E}((Y_i - y_i) v_i(t, Y^{(i)}) V_i) = 0. \quad (2.42)$$

Therefore

$$\begin{aligned} |\mathbb{E}((Y_i - y_i)v_i(t, Y)V)| &\leq \mathbb{E}|((v_i(t, Y) - v_i(t, Y^{(i)}))V| + \mathbb{E}|v_i(t, Y^{(i)})(V - V_i)| \quad (2.43) \\ &\leq \left\| \frac{\partial v_i}{\partial x_i} \right\| \|V\| + \|v_i\| \|V - V_i\| \\ &\leq 2atc_{ii} + b_i^2. \end{aligned}$$

This gives

$$\mathbb{P}(Y \in A_3^c) \leq \mathbb{P}(|V| \geq \eta_0 n) \leq \frac{\mathbb{E}(V^2)}{\eta_0^2 n^2} \leq \frac{\sum_{i=1}^n (ac_{ii} + b_i^2)}{\eta_0^2 n^2} = \frac{1}{6}. \quad (2.44)$$

Combining (2.35), (2.39) and (2.44), we have

$$\mathbb{P}(Y \in A) \geq 1 - \mathbb{P}(Y \in A_1^c) - \mathbb{P}(Y \in A_2^c) - \mathbb{P}(Y \in A_3^c) \geq \frac{1}{2}. \quad (2.45)$$

Plugging this into (2.30) and taking supremum over y completes the proof. \square

3. APPLICATION TO EXPONENTIAL RANDOM GRAPHS

As mentioned earlier, we would like to apply Theorem 2.1 to derive the exact asymptotics for the conditional normalization constant of constrained exponential random graphs. Recall the definition of an s -parameter family of conditional exponential random graphs introduced earlier, where we assume that the “ideal” edge density of the graph is e . Let

$$f(x) = \zeta_1 T_1(x) + \cdots + \zeta_s T_s(x) \text{ and } h(x) = T_1(x) - N^2 e, \quad (3.1)$$

where $T_i(x)/N^2$ is the equivalent notion of homomorphism density as defined in (1.13). Let $n = \binom{N}{2}$. We compare the conditional normalization constant $\psi_{N,t}^{e,\zeta}$ (1.18) for constrained exponential random graphs with the generic conditional normalization constant F^c (2.1). Note that the constraint $|e(G_N) - e| \leq t$ may be translated into $|T_1(x) - N^2 e| \leq N^2 t$, and if we further redefine t to be $(1 - 1/N)t'/2$ then we arrive at the generic constraint $|h(x)| \leq t'n$ as in (2.1). Thus $\psi_{N,t}^{e,\zeta} = F^c/N^2$. In the following we give a concrete error bound for $\psi_{N,t}^{e,\zeta}$ using the estimates in Theorem 2.1.

Theorem 3.1. *Let s be a positive integer and H_1, \dots, H_s be fixed finite simple graphs. Let N be another positive integer and let $n = \binom{N}{2}$. Define T_1, \dots, T_s accordingly as in the paragraph before Theorem 1.2. Let ζ_1, \dots, ζ_s be s real parameters and define $\psi_{N,t}^{e,\zeta}$ as in (1.18). Let $f(x) = \zeta_1 T_1(x) + \cdots + \zeta_s T_s(x)$, $B = 1 + |\zeta_1| + \cdots + |\zeta_s|$, and I be defined as in (1.3). Take $\kappa > 8$. Then*

$$\begin{aligned} &\sup_{x \in [0,1]^n: |h(x)| \leq (t' - cn^{-1/(2\kappa)})n} \frac{f(x) - I(x)}{N^2} - CBN^{-1/2} \leq \psi_{N,t}^{e,\zeta} \quad (3.2) \\ &\leq \sup_{x \in [0,1]^n: |h(x)| \leq (t' + cn^{-1/(2\kappa)})n} \frac{f(x) - I(x)}{N^2} + CB^{8/5} N^{(8-\kappa)/(5\kappa)} (\log N)^{1/5} \left(1 + \frac{\log B}{\log N} \right) \\ &\quad + CB^2 N^{(2-\kappa)/(2\kappa)}, \end{aligned}$$

where $t' = 2Nt/(N-1)$ and c and C are constants that may depend only on H_1, \dots, H_s and e .

Proof. Chatterjee and Dembo [2] checked that $T_i(x)$ satisfies both the smoothness condition and the low complexity gradient condition stated at the beginning of this paper, which readily implies that f and h satisfy the assumptions of Theorem 2.1. Recall that the indexing set for quantities like b_i and γ_{ij} , instead of being $\{1, \dots, n\}$, is now $\{(i, j) : 1 \leq i < j \leq N\}$, and for

simplicity we write (ij) instead of (i, j) . Let $a, b_{(ij)}, c_{(ij)(i'j')}$ be the supremum norms of f and let $\alpha, \beta_{(ij)}, \gamma_{(ij)(i'j')}$ be the corresponding supremum norms of h . For any $\epsilon > 0$, let $\mathcal{D}_f(\epsilon)$ and $\mathcal{D}_h(\epsilon)$ be finite subsets of \mathbb{R}^n associated with the gradient vectors of f and h respectively.

Based on the bounds for T_i (1.14) (1.15) (1.16), we derive the bounds for f and h .

$$a \leq CBN^2, \quad b_{(ij)} \leq CB, \quad (3.3)$$

$$c_{(ij)(i'j')} \leq \begin{cases} CBN^{-1}, & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3; \\ CBN^{-2}, & \text{if } |\{i, j, i', j'\}| = 4, \end{cases} \quad (3.4)$$

$$|\mathcal{D}_f(\epsilon)| \leq \prod_{i=1}^s |\mathcal{D}_i(\epsilon/(\zeta_i s))| \leq \exp\left(\frac{CB^4 N}{\epsilon^4} \log \frac{CB}{\epsilon}\right). \quad (3.5)$$

$$\alpha \leq CN^2, \quad \beta_{(ij)} \leq C, \quad (3.6)$$

$$\gamma_{(ij)(i'j')} \leq \begin{cases} CN^{-1}, & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3; \\ CN^{-2}, & \text{if } |\{i, j, i', j'\}| = 4, \end{cases} \quad (3.7)$$

$$|\mathcal{D}_h(\epsilon)| = |\mathcal{D}_1(\epsilon)| \leq \exp\left(\frac{CN}{\epsilon^4} \log \frac{C}{\epsilon}\right). \quad (3.8)$$

We then estimate the lower and upper error bounds for $\psi_{N,t}^{e,\zeta}$ using the bounds on f and h obtained above. First the lower bound:

$$\sum_{(ij)} ac_{(ij)(ij)} \leq CB^2 N^3, \quad \sum_{(ij)} b_{(ij)}^2 \leq CB^2 N^2. \quad (3.9)$$

$$\sum_{(ij)} \alpha \gamma_{(ij)(ij)} \leq CN^3, \quad \sum_{(ij)} \beta_{(ij)}^2 \leq CN^2. \quad (3.10)$$

Therefore

$$\delta_0 \leq cn^{-1/4} \leq cn^{-1/(2\kappa)}, \quad (3.11)$$

$$\frac{\epsilon_0 n + \eta_0 n + \log 2}{N^2} \leq CN^{-1} + CBN^{-1/2} + CN^{-2} \leq CBN^{-1/2}. \quad (3.12)$$

This gives

$$\psi_{N,t}^{e,\zeta} \geq \sup_{x \in [0,1]^n: |h(x)| \leq (t' - cn^{-1/(2\kappa)})n} \frac{f(x) - I(x)}{N^2} - CBN^{-1/2}, \quad (3.13)$$

Next the more involved upper bound: Assume that $n^{-1/4} \leq \delta \leq 1$ and $0 < \epsilon \leq 1$. Since $K \leq CB$, this implies that

$$l \leq CBN^2, \quad m_{(ij)} \leq CB\delta^{-1}, \quad (3.14)$$

$$n_{(ij)(i'j')} \leq \begin{cases} CBN^{-1}\delta^{-1}, & \text{if } |\{i, j, i', j'\}| = 2 \text{ or } 3; \\ CBN^{-2}\delta^{-2}, & \text{if } |\{i, j, i', j'\}| = 4. \end{cases} \quad (3.15)$$

The following estimates are direct consequences of the bounds on $l, m_{(ij)}$, and $n_{(ij)(i'j')}$.

$$\sum_{(ij)} ln_{(ij)(ij)} \leq CB^2 N^3 \delta^{-1}, \quad \sum_{(ij)} m_{(ij)}^2 \leq CB^2 N^2 \delta^{-2}, \quad (3.16)$$

$$\sum_{(ij)(i'j')} ln_{(ij)(i'j')}^2 \leq CB^3 N^3 \delta^{-2}, \quad (3.17)$$

$$\sum_{(ij)(i'j')} m_{(ij)}(m_{i'j'} + 4)n_{(ij)(i'j')} \leq CB^3 N^2 \delta^{-4}, \quad (3.18)$$

$$\sum_{(ij)} n_{(ij)}^2 \leq CB^2\delta^{-2}, \quad \sum_{(ij)} n_{(ij)} \leq CBN\delta^{-1}. \quad (3.19)$$

Therefore

$$\begin{aligned} \text{complexity term} &\leq CBN^2\delta^{-1}\epsilon + CN^2\epsilon + \log \frac{CB}{\delta\epsilon} + \frac{CB^4N}{\epsilon^4} \log \frac{CB}{\epsilon} + \frac{CB^4N}{\delta^4\epsilon^4} \log \frac{CB}{\delta\epsilon} \\ &\leq CBN^2\delta^{-1}\epsilon + \frac{CB^4N}{\delta^4\epsilon^4} \log \frac{CB}{\delta\epsilon}. \end{aligned} \quad (3.20)$$

$$\text{smoothness term} \leq CB^{3/2}N^{3/2}\delta^{-1} + CB^2N\delta^{-2} + CBN\delta^{-1} + C \leq CB^2N^{3/2}\delta^{-1}. \quad (3.21)$$

Taking $\epsilon = (B^3 \log N)/(\delta^3 N)^{1/5}$, this gives

$$\begin{aligned} \psi_{N,t}^{e,\zeta} &\leq \sup_{x \in [0,1]^n: |h(x)| \leq (t'+\delta)n} \frac{f(x) - I(x)}{N^2} + CB^{8/5}N^{-1/5}(\log N)^{1/5}\delta^{-8/5} \left(1 + \frac{\log B}{\log N}\right) \\ &\quad + CB^2N^{-1/2}\delta^{-1}. \end{aligned} \quad (3.22)$$

For n large enough, we may choose $\delta = cn^{-1/(2\kappa)}$ as in (3.11), which yields a further simplification

$$\begin{aligned} \psi_{N,t}^{e,\zeta} &\leq \sup_{x \in [0,1]^n: |h(x)| \leq (t'+cn^{-1/(2\kappa)})n} \frac{f(x) - I(x)}{N^2} + CB^{8/5}N^{(8-\kappa)/(5\kappa)}(\log N)^{1/5} \left(1 + \frac{\log B}{\log N}\right) \\ &\quad + CB^2N^{(2-\kappa)/(2\kappa)}. \end{aligned} \quad (3.23)$$

□

We can do a more refined analysis of Theorem 3.1 when ζ_i 's are non-negative for $i \geq 2$.

Theorem 3.2. *Let s be a positive integer and H_1, \dots, H_s be fixed finite simple graphs. Let N be another positive integer and let $n = \binom{N}{2}$. Let ζ_1, \dots, ζ_s be s real parameters and suppose $\zeta_i \geq 0$ for $i \geq 2$. Define $\psi_{N,t}^{e,\zeta}$ as in (1.18). Let $B = 1 + |\zeta_1| + \dots + |\zeta_s|$ and I be defined as in (1.3). Take $\kappa > 8$. Then*

$$\begin{aligned} -cBN^{-1/\kappa} &\leq \psi_{N,t}^{e,\zeta} - \sup_{|x-e| \leq t} \left\{ \zeta_1 x + \dots + \zeta_k x^{e(H_k)} - \frac{1}{2}I(x) \right\} \\ &\leq CB^{8/5}N^{(8-\kappa)/(5\kappa)}(\log N)^{1/5} \left(1 + \frac{\log B}{\log N}\right) \\ &\quad + CB^2N^{-1/\kappa}, \end{aligned} \quad (3.24)$$

where $e(H_i)$ denotes the number of edges in H_i and c and C are constants that may depend only on H_1, \dots, H_s, e , and t .

Remark. *If $H_i, i \geq 2$ are all stars, then the conclusions of Theorem 3.2 hold for any ζ_1, \dots, ζ_s .*

Remark. *As an example, consider the case where $s = 2$, H_1 is a single edge and H_2 is a triangle. Theorem 3.2 shows that the difference between $\psi_{N,t}^{e,\zeta}$ and $\sup_{|x-e| \leq t} \{\zeta_1 x + \zeta_2 x^3 - \frac{1}{2}I(x)\}$ tends to zero as long as $|\zeta_1| + |\zeta_2|$ grows slower than $N^{(\kappa-8)/(8\kappa)}(\log N)^{-1/8}$, thereby allowing a small degree of sparsity for ζ_i . When ζ_i 's are fixed, it provides an approximation error bound of order $N^{(8-\kappa)/(5\kappa)}(\log N)^{1/5}$, substantially better than the negative power of $\log^* N$ given by Szemerédi's lemma.*

Proof. Fix $t > 0$. We find upper and lower bounds for

$$L_N = \sup_{x \in [0,1]^n: |h(x)| \leq (t' + cn^{-1/(2\kappa)})_n} \frac{f(x) - I(x)}{N^2} \quad (3.25)$$

and

$$M_N = \sup_{x \in [0,1]^n: |h(x)| \leq (t' - cn^{-1/(2\kappa)})_n} \frac{f(x) - I(x)}{N^2} \quad (3.26)$$

in Theorem 3.1 when N is large.

On one hand, by considering $g(x, y) = x_{ij}$ for any $(\frac{i-1}{N}, \frac{i}{N}] \times (\frac{j-1}{N}, \frac{j}{N}]$ and $i \neq j$, we have

$$L_N \leq \sup_{\substack{g: [0,1]^2 \rightarrow [0,1], g(x,y)=g(y,x) \\ |e(g)-e| \leq t + \frac{c}{2}n^{-1/(2\kappa)}}} \left\{ \zeta_1 t(H_1, g) + \cdots + \zeta_k t(H_k, g) - \frac{1}{2} \iint_{[0,1]^2} I(g(x, y)) dx dy \right\}, \quad (3.27)$$

$$M_N \leq \sup_{\substack{g: [0,1]^2 \rightarrow [0,1], g(x,y)=g(y,x) \\ |e(g)-e| \leq t}} \left\{ \zeta_1 t(H_1, g) + \cdots + \zeta_k t(H_k, g) - \frac{1}{2} \iint_{[0,1]^2} I(g(x, y)) dx dy \right\}. \quad (3.28)$$

It was proved in Chatterjee and Diaconis [3] that when ζ_i 's are non-negative for $i \geq 2$, the above supremum may only be attained at constant functions on $[0, 1]$. Therefore

$$L_N \leq \sup_{|x-e| \leq t + \frac{c}{2}n^{-1/(2\kappa)}} \left\{ \zeta_1 x + \cdots + \zeta_k x^{e(H_k)} - \frac{1}{2} I(x) \right\}, \quad (3.29)$$

$$M_N \leq \sup_{|x-e| \leq t} \left\{ \zeta_1 x + \cdots + \zeta_k x^{e(H_k)} - \frac{1}{2} I(x) \right\}. \quad (3.30)$$

On the other hand, by considering $g'(x, y) = x_{ij} \equiv x$ for any $i \neq j$, we have

$$L_N \geq \sup_{|\frac{N-1}{N}x-e| \leq t} \left\{ \zeta_1 x + \cdots + \zeta_k x^{e(H_k)} - \frac{1}{2} I(x) \right\} + O\left(\frac{1}{N}\right), \quad (3.31)$$

$$M_N \geq \sup_{|\frac{N-1}{N}x-e| \leq t - \frac{c}{2}n^{-1/(2\kappa)}} \left\{ \zeta_1 x + \cdots + \zeta_k x^{e(H_k)} - \frac{1}{2} I(x) \right\} + O\left(\frac{1}{N}\right). \quad (3.32)$$

The $O(1/N)$ factor comes from the following consideration. The difference between $I(g')$ and $I(x)$ is easy to estimate, while the difference between $t(H_i, g')$ and $t(H_i, x) = x^{e(H_i)}$ is caused by the zero diagonal terms x_{ii} . We do a broad estimate of (1.13) and find that it is bounded by c_i/N , where c_i is a constant that only depends on H_i . Putting everything together,

$$L_N = M_N = \sup_{|x-e| \leq t} \left\{ \zeta_1 x + \cdots + \zeta_k x^{e(H_k)} - \frac{1}{2} I(x) \right\} + O\left(\frac{1}{N^{1/\kappa}}\right). \quad (3.33)$$

The rest of the proof follows. \square

REFERENCES

- [1] Aristoff, D, Radin, C.: Emergent structures in large networks. *J. Appl. Prob.* 50, 883-888 (2013)
- [2] Chatterjee, S., Dembo, A.: Nonlinear large deviations. arXiv: 1401.3495 (2014)
- [3] Chatterjee, S., Diaconis, P.: Estimating and understanding exponential random graph models. *Ann. Statist.* 41, 2428-2461 (2013)
- [4] Chatterjee, S., Varadhan, S.R.S.: The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.* 32, 1000-1017 (2011)
- [5] Kenyon, R., Yin, M.: On the asymptotics of constrained exponential random graphs. arXiv: 1406.3662 (2014)
- [6] Radin, C., Ren, K., Sadun, L.: The asymptotics of large constrained graphs. *J. Phys. A* 47, 175001 (2014)
- [7] Radin, C., Sadun, L.: Phase transitions in a complex network. *J. Phys. A* 46, 305002 (2013)
- [8] Radin, C., Sadun, L.: Singularities in the entropy of asymptotically large simple graphs. arXiv: 1302.3531 (2013)
- [9] Radin, C., Yin, M.: Phase transitions in exponential random graphs. *Ann. Appl. Probab.* 23, 2458-2471 (2013)
- [10] Yin, M., Rinaldo, A., Fadnavis, S.: Asymptotic quantization of exponential random graphs. arXiv: 1311.1738 (2013)
- [11] Zhu, L.: Asymptotic structure of constrained exponential random graph models. arXiv: 1408.1536 (2014)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF DENVER, DENVER, CO 80208, USA

E-mail address: mei.yin@du.edu