LIMIT THEOREMS FOR EXPONENTIAL RANDOM GRAPHS

by Alessandra Bianchi - UNIVERSITÀ DI PADOVA

[joint work with Francesca Collet and Elena Magnanini]

SEMINÁRIO DE PROBABILIDADE - MAY 23, 2022 ISTITUTO DE MATEMÁTICA - UFRJ

1. Setting

a. Exponential random graphs

(see Park, Newman '04 for history)

We set

* $\mathcal{G}_n := \{G \text{ simple graph on } [n]\}$, $G = (V, E)$
* $V = \{1, \dots, n\} =: [n]$, $n \in \mathbb{N}$
* $E \subseteq \mathcal{E}_n := \{(i,j) : i, j \in [n], i \neq j\}$

$\longrightarrow$ ExpRG defined by a Gibbs probability measure on $\mathcal{G}_n$

<u>Edge-triangle model</u> (Strauss's model)

partition function

For $\alpha, h \in \mathbb{R}$

$$\boxed{\mathbb{P}_{n;\alpha,h}(G) = \frac{e^{H_{n,\alpha,h}(G)}}{Z_{n,\alpha,h}}}, \qquad Z_{n,\alpha,h} = \sum_{G \in \mathcal{G}_n} e^{H_{n,\alpha,h}(G)}$$

Where $H_{n;\alpha,h}(G) = \frac{\alpha}{n} T(G) + h E(G)$

triangles in G edges in G

$$\boxed{\text{Remark: If } \alpha = 0 \Longrightarrow \mathbb{P}_{n,0,h}(G) \propto e^{hE(G)} = \mathbb{P}^{ER}_{n,p}(G) \quad p = \frac{e^h}{1+e^h}}$$

Equivalently:

$$G_n \xleftrightarrow[\text{su}]{1-1} \{0,1\}^{E_n} \ni \underline{x} \longrightarrow \mathbb{P}_{n;\alpha,h}(\underline{x}) = e^{H_{n,\alpha,h}(\underline{x})} / Z_{n;\alpha,h}$$

where $$\boxed{H_{n;\alpha,h}(\underline{x}) = \frac{\alpha}{n} \sum_{\{i,j,k\} \in T_n} x_i x_j x_k + h \sum_{i \in E_n} x_i}$$

T_n → set of triangles over $[n]$

Goal: Determine the limiting behavior of the edge density

$$\frac{2E(G)}{n^2} \in [0,1]_n = \left\{0, \frac{2}{n^2}, \dots, 1-\frac{1}{n}\right\}$$

Equivalently: $\underline{X} = (X_i)_{i \in E_n}$ with law $\mathbb{P}_{n,\alpha,h}$, $S_n := \sum_{i \in E_n} X_i$

$$2\frac{S_n}{n^2} \quad \left(= \frac{S_n}{n^2/2} \right)$$

2. Relevant results on ExpRG

a. Free energy : maximizers of $\mathbb{P}_{n;\alpha,h}$ as $n \to \infty$

Let $f_n(\alpha,h) := \frac{1}{n^2} \ln Z_{n;\alpha,h}$ and $f(\alpha,h) = \lim_{n\to\infty} f_n(\alpha,h)$

(finite size) free energy (infinite size) free energy

Thm [Chatterjee, Diaconis '13]: If $\alpha > -2$: $I(u) = u \ln u + (1-u)\ln(1-u)$

* $$\boxed{f(\alpha,h) \overset{(1)}{=} \sup_{u\in[0,1]} \left(\frac{\alpha}{6} u^3 + \frac{h}{2} u - \frac{1}{2} I(u) \right) = \frac{\alpha}{6} u^{*3} + \frac{h}{2} u^* - \frac{1}{2} I(u^*)}$$

where u^* denotes the/a maximizer of (1).

asymptotic graphon

* if u^* is unique : as $n \to \infty$, $\mathbb{P}_{n;\alpha,h}$ and $\mathbb{P}^{ER}_{n,u^*}$ have the same limit

if u^* is not unique: as $n \to \infty$, $\mathbb{P}_{n;\alpha,h}$ behaves like $\mathbb{P}^{ER}_{n,u}$ where u randomly chosen from the set of maximizers.

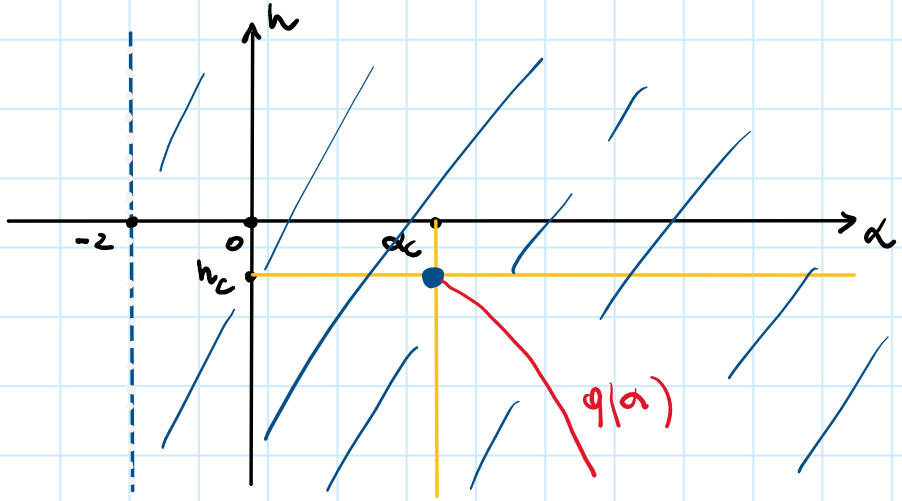
b. Phase diagram in (α, h) [Chatterjee, Dey '10 & Radin, Yin '13]

* $RS = \{(\alpha,h) : \alpha > -2\}$ Replica Symmetric regime

$\longrightarrow RS = \mathcal{U}^{RS} \sqcup \mathcal{M}^{RS}$

where $\mathcal{U}^{RS} = \{(\alpha,h) : \alpha > -2$, (1) has unique maximizer: $u^*\}$

$\mathcal{M}^{RS} = \{(\alpha,h) : \alpha > -2$, (1) has two maximizers: $u_1^* < u_2^*\}$



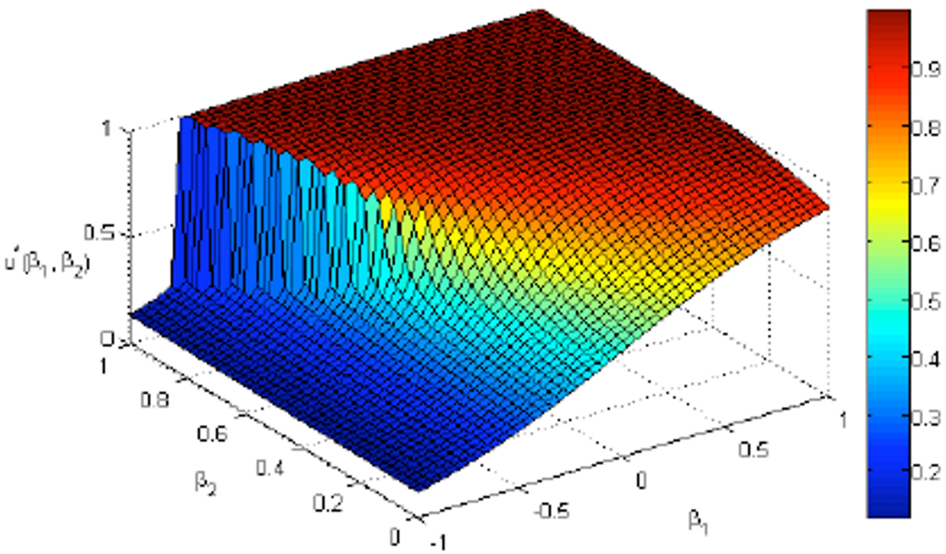
- (α_c, h_c) = critical point $\in \mathcal{U}^{RS}$

  $\alpha_c = \frac{27}{8}$, $h_c = \ln 2 - \frac{3}{2}$

- $q(\alpha) = \mathcal{M}^{RS}$

Thm [Radin, Yin '13]

* If $(\alpha,h) \in \mathcal{U}^{RS} \setminus (\alpha_c, h_c) \Longrightarrow f(\alpha,h)$ is analytic
* If $(\alpha,h) = (\alpha_c, h_c) \Longrightarrow f(\alpha_c,h_c)$ has second order discontinuity
* If $(\alpha,h) \in \mathcal{M}^{RS} \Longrightarrow f(\alpha,h)$ has first order discontinuity



Consequences:

Non-uniqueness of $\mathbb{P}_{n;\alpha,h}$ as $n \to \infty$ and $(\alpha,h) \in \mathcal{M}^{RS}$

(Picture by Sukhada Fadnavis, taken from [Chatterjee, Diaconis '13])

3. Main Results

a. Convergence of $2S_n/n^2$

Thm 1 [B., Collet, Magnanini '21]

(i.) $\forall (\alpha,h) \in \mathcal{U}^{RS}$: $\frac{2S_n}{n^2} \xrightarrow[n\to\infty]{a.s.} u^*$ w.r.t. $\mathbb{P}_{n;\alpha,h}$

(ii.) $\forall (\alpha,h) \in \mathcal{M}^{RS}$: $\mathbb{P}_{n;\alpha,h}\left(\frac{2S_n}{n^2} \in B_\varepsilon(u_1^*) \cup B_\varepsilon(u_2^*)\right) \geq 1 - e^{-cn^2}$

$\forall \varepsilon > 0$, $n \gg 1$, $c = c(\varepsilon,\alpha,h) > 0$

Proof of (i): It follows from exponential convergence in probability (Ellis's Theorem)

* cumulant generating function : $\longrightarrow$ $c_n(t) = \frac{2}{n^2} \ln \mathbb{E}_{n;\alpha,h}\left(e^{t\frac{2S_n}{n^2}}\right)$, $t \in \mathbb{R}$

(direct computation)

* free energies $f_n(\alpha,h)$, $f(\alpha,h)$ $\longrightarrow$ $= 2\underbrace{\left(f_n(\alpha,h+t) - f_n(\alpha,h)\right)}_{\approx 2\partial_h f_n(\alpha,h)\cdot t}$

Hence $c'(0) = \left(\lim_{n\to\infty} c_n(t)\right)'\Big|_{t=0} = 2\partial_h f(\alpha,h) = u^*$

Proof of (ii): The previous argument breaks as for $(\alpha, h) \in \mathcal{U}^{RS}$, $f(\alpha, h)$ has first order discontinuities.

Key Points of the proof: (1) LDP for Erdös-Rényi graphs [Chatterjee, Varadhan '11]

(2) LDP for integrals of exponential functional [Varadhan's Lemma]

$\xrightarrow{(1)+(2)}$ LDP for the edge-triangle model, with speed n^2 and explicit rate function, with minimizers provided in [Chatterjee, Diaconis '13]

$\Downarrow$

the 2 minimizers have edge density u_1^* and u_2^*, and the result follows

Formalization require to introduce "graphon space" $\mathcal{W}$:
Further details at the end of the presentation.

b. Speed of convergence in $\mathcal{U}^{RS}$

As a byproduct of the LDP for the edge-triangle model:

Prop: (a) If $\alpha \in (-2, \alpha_c)$ (so that $(\alpha, h) \in \mathcal{U}^{RS}$ $\forall h \in \mathbb{R}$), $\forall \varepsilon > 0$

$$\Longrightarrow \frac{C_1}{n} \leq \mathbb{E}_{n;\alpha,h}\left(\left|\frac{2S_n}{n^2} - u^*\right|\right) \leq \frac{C_2}{n^{1-\varepsilon}} \qquad C_1, C_2 > 0$$

(b) If $(\alpha, h) = (\alpha_c, h_c)$, $\forall \varepsilon > 0$

$$\Longrightarrow \frac{C_3}{n^{1/2}} \leq \mathbb{E}_{n;\alpha,h}\left(\left|\frac{2S_n}{n^2} - u^*\right|\right) \leq \frac{C_4}{n^{1/2-\varepsilon}} \qquad C_3, C_4 > 0$$

Proof ideas: $\forall \delta > 0$:

$$\delta \cdot \underbrace{\mathbb{P}_{n;\alpha,h}\left(\left|\frac{2S_n}{n^2} - u^*\right| \geq \delta\right)}_{\text{LDP}} \leq \mathbb{E}_{n;\alpha,h}\left(\left|\frac{2S_n}{n^2} - u^*\right|\right) \leq \underbrace{\mathbb{P}_{n;\alpha,h}\left(\left|\frac{2S_n}{n^2} - u^*\right| \geq \delta\right)}_{\text{LDP}} + \delta$$

- then optimize with the choice of $\delta = \delta(n)$.

C. Fluctuations of $2S_n/n^2$

Set: $m_n \equiv m_n(\alpha,h) := \mathbb{E}_{n,\alpha,h}\left(\frac{2S_n}{n^2}\right)$ average edge-density

$v_n \equiv v_n(\alpha,h) = \partial_h m_n(\alpha,h)$ variance edge-density

Remark: $m_n = 2\partial_h f_n(\alpha,h) = 2\partial_h\left(\frac{1}{n^2}\log Z_{n;\alpha,h}\right) = \sum_G \frac{2S_n}{n^2}\frac{e^{H_{n,\alpha,h}(G)}}{Z_{n,\alpha,h}} = \mathbb{E}_n\left(\frac{2S_n}{n^2}\right)$

$v_n = 2\partial^2_{hh} f_n(\alpha,h) = \dots = \frac{2}{n^2}\mathrm{Var}(S_n) \geq 0$

Thm 2 [B., Collet, Magnanini '21]

$\forall(\alpha,h)\in U^{RS}\setminus(\alpha_c,h_c)$, it holds the CLT

$$\frac{S_n - \frac{n^2}{2}m_n}{n/\sqrt{2}} \xrightarrow[n\to\infty]{d} \mathcal{N}\big(0, v(\alpha,h)\big) \quad \text{w.r.t. } \mathbb{P}_{n;\alpha,h}$$

where $v(\alpha,h) = \lim_{n\to\infty} v_n(\alpha,h) = 2\partial^2_{hh} f(\alpha,h) = \partial_h u^*$

Proof ideas: $f(\alpha,h)$ is analytic on $\mathcal{U}^{RS} \setminus (\alpha_c, h_c)$

1. Uniform convergence of $f_n(\alpha,h)$ and its derivatives (Lee-Yang Theorem)

   - $m_n(\alpha,h) = 2\partial_h f_n(\alpha,h) \longrightarrow 2\partial_h f(\alpha,h) = u^*$ as $n \to \infty$
   - $v_n(\alpha,h) = 2\partial^2_{hh} f_n(\alpha,h) \longrightarrow 2\partial^2_{hh} f(\alpha,h) = \partial_h u^*$

2. Convergence of the moment generating function

$$M_n(t) = \mathbb{E}_{n;\alpha,h}\left(e^{t\frac{\left(S_n - \frac{n^2}{2}m_n\right)}{n/\sqrt{2}}} \right) \quad \text{for } t \in \mathrm{Int}(0)$$

$$\rightarrow \ln M_n(t) = \frac{n^2}{2}\left(\underbrace{c_n\left(\frac{\sqrt{2}t}{n}\right) - \frac{\sqrt{2}t}{n}c_n'(0)}_{\text{second order Taylor's expansion}}\right)$$

function of cumulants of $\frac{2S_n}{n^2}$

Recall that $c_n(t) = 2\left(f_n(\alpha,h+t) - f_n(\alpha,h)\right)$

Note: Existence of the second derivatives of $f(\alpha,h)$ is crucial

d. Conjectures : [B., Collet, Magnanini '21]

Conjecture 1 : Non-standard limit theorem at (α_c, h_c)

$$\frac{S_n - \frac{n^2}{2} m_n}{n^{3/2}/2} \xrightarrow[n\to\infty]{d} Y \qquad \text{w.r.t. } \mathbb{P}_{n;\alpha_c,h_c}$$

where Y is real r.v. with density $\ell_c(y) \propto e^{-\frac{81}{64}y^4}$

Conjecture 2 : Coefficients of the mixture of Dirac measures

If $(\alpha,h) \in \mathcal{M}^{RS}$:

$$\frac{2 S_n}{n^2} \xrightarrow[n\to\infty]{d} \kappa\, \delta_{u_1^*} + (1-\kappa)\, \delta_{u_2^*}$$

where $\kappa = \dfrac{D(u_1^*)}{D(u_1^*) + D(u_2^*)}$ and $D(u_i^*) = \left(1 - 2\alpha (u_i^*)^2 (1-u_i^*)\right)^{-1/2}$

4. Mean-field approximation

a. Mean-field model

: For $\underline{x} \in \{0,1\}^{E_m}$, let

$$\overline{\mathbb{P}}_{m,\alpha,h}(\underline{x}) := e^{\overline{H}_{m,\alpha,h}(\underline{x})} / \overline{Z}_{m,\alpha,h}$$

where $\boxed{\overline{H}_{m,\alpha,h}(\underline{x}) = \frac{4}{3}\frac{\alpha}{m^4}\Big(\sum_{i\in E_m} x_i\Big)^3 + h\sum_{i\in E_m} x_i} = m^2\left(\frac{\alpha}{6}u^3 + \frac{h}{2}u\right)$

replace $\frac{\alpha}{m}\sum_{\{i,j,k\}\in T_m} x_i x_j x_k$

$u = \frac{2S_m(\underline{x})}{m^2}$, with $u \in [0,1]_m$

Hence:

$$\overline{Z}_{m,\alpha,h} = \sum_{u\in[0,1]_m} N_u\, e^{m^2\left(\frac{\alpha}{6}u^3 + \frac{h}{2}u\right)} = \sum_{u\in[0,1]_m} e^{m^2\Big(\underbrace{\frac{\alpha}{6}u^3 + \frac{h}{2}u - \frac{1}{2}I(u)}_{U_{\alpha,h}(u)}\Big) + o(m^2)}$$

$$\boxed{\text{Remark: } \sup_{u\in[0,1]} U_{\alpha,h}(u) = f(\alpha,h)}$$

free energy of $\overline{\mathbb{P}}_{m,\alpha,h}$

b. Mean-field results [B., Collet, Magnanini '21]

Let $\bar{f}_n(\alpha,h) = \frac{1}{n^2} \log \bar{Z}_{n;\alpha,h}$ and $\bar{f}(\alpha,h) = \lim_{n\to\infty} \bar{f}_n(\alpha,h)$

Thm 3 [Free energy]

$\forall (\alpha,h) \in RS$: $\quad \bar{f}(\alpha,h) = f(\alpha,h)$

Thm 4 [Convergence of the edge density]

(i) $\forall (\alpha,h) \in \mathcal{U}^{RS}$: $\quad \frac{2 S_n}{n^2} \xrightarrow[n\to\infty]{a.s.} u^*$, w.r.t. $\bar{\mathbb{P}}_{n;\alpha,h}$

(ii) $\forall (\alpha,h) \in \mathcal{M}^{RS}$: $\quad \frac{2 S_n}{n^2} \xrightarrow[n\to\infty]{d} \kappa\, \delta_{u_1^*} + (1-\kappa)\, \delta_{u_2^*}$, w.r.t. $\bar{\mathbb{P}}_{n;\alpha,h}$

where $\boxed{\kappa = \frac{D(u_1^*)}{D(u_1^*) + D(u_2^*)}}$

Prop 2 : [Speed of convergence]

(i) $\forall (\alpha, h) \in \mathcal{U}^{RS} \setminus (\alpha_c, h_c)$: $\lim\limits_{n\to\infty} n\, \overline{\mathbb{E}}_{n;\alpha,h}\left(\left| 2\frac{S_n}{n^2} - u^* \right| \right) = c_1(\alpha,h) < \infty$

(ii) At (α_c, h_c): $\lim\limits_{n\to\infty} n^{1/2} \cdot \overline{\mathbb{E}}_{n;\alpha,h}\left(\left| \frac{2S_n}{n^2} - u^* \right| \right) = c_2(\alpha_c,h_c) < \infty$

Thm 5 [Fluctuations of the edge density]

(i) $\forall (\alpha,h) \in \mathcal{U}^{RS} \setminus (\alpha_c,h_c)$ and with $\overline{m}_n = \overline{\mathbb{E}}_{n,\alpha,h}\left(\frac{2S_n}{n^2} \right)$

$$\frac{S_n - \frac{n^2}{2}\overline{m}_n}{n/\sqrt{2}} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \underline{v(\alpha,h)}), \quad \text{w.r.t. } \overline{\mathbb{P}}_{n,\alpha,h}$$

↳ $2\partial^2_{hh} f(\alpha,h)$

(ii) At (α_c, h_c): $\boxed{\frac{S_n - \frac{n^2}{2}\overline{m}_n}{n^{3/2}/2} \xrightarrow[n\to\infty]{d} Y}$, w.r.t. $\overline{\mathbb{P}}_{n,\alpha_c,h_c}$

where Y is real r.v. with density $\ell_c(y) \propto e^{-\frac{81}{64}y^4}$

5. Large deviations for edge-triangle model

a. Graph limits [Borgs, Chayes, Lóvasz, Sós, Vesztergombi '06, '08, '12]

Let $(G_n)_{n\geq 1}$ sequence of graphs $\longrightarrow$ Identify its limit as $n\to\infty$

* $g : [0,1]^2 \longrightarrow [0,1]$ measurable, symmetric is called graphon
* We say that "$G_n \xrightarrow[n\to\infty]{} g$" if, $\forall$ H fixed graph

$$t(H, G_n) \xrightarrow[n\to\infty]{} t(H,g) := \int_{[0,1]^{V(H)}} \prod_{(i,j)\in E(H)} g(x_i, x_j)\, dx_i\, dx_j$$

density of H in G_n

Idea: • $[0,1]$ is a continuum of vertices

• $g(x_i, x_j)$ is the probability that (x_i, x_j) is an edge of limiting graph

Example: $(G_n)_{n\geq 1}$ are Erdös-Rényi $RG(p) \Longrightarrow$ "$G_n \xrightarrow[n\to\infty]{} g$", $g \equiv p$, a.s.

b. Graphon space $\mathcal{W}$

Let $\mathcal{W} := \{ g \text{ graphon} \}$ and $\mathcal{G}_n = \{ G \text{ simple graph on } [n] \}$

1. $\bigcup_{n \in \mathbb{N}} \mathcal{G}_n \subset \mathcal{W}$: $\mathcal{G}_n \ni G \longrightarrow g^G$ s.t. $g^G(x,y) := \begin{cases} 1 & \text{if } (\lceil nx \rceil, \lceil ny \rceil) \in E(G) \\ 0 & \text{otherwise} \end{cases}$

2. Distance on $\mathcal{W}$ (cut distance) :

$$d_{\square}(g_1, g_2) := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} \big(g_1(x,y) - g_2(x,y) \big) \, dx \, dy \right|$$

3. $(\mathcal{W}, d_{\square}) \xrightarrow{\sim} (\widetilde{\mathcal{W}}, \delta_{\square})$ compact metric space

where $\widetilde{\mathcal{W}} = \mathcal{W}/\sim$ and $\sim$ equivalence relation s.t.

$g_1 \sim g_2$ if $g_1(x,y) = g_2(\sigma(x), \sigma(y))$ for some $\sigma : [0,1] \xrightarrow[\text{su}]{1-1} [0,1]$ Lebesgue measure preserving

c. Large deviations on $\tilde{W}$

Thm [Chatterjee, Varadhan '11]: $\forall\ p \in (0,1) \Longrightarrow$

$(\tilde{\mathbb{P}}^{ER}_{n,p})_{n\geq 1}$ satisfies a LDP on $(\tilde{W}, \delta_\square)$ with speed n^2 and rate function

$$\mathcal{J}^{ER}_p(\tilde{g}) = \frac{1}{2} \iint_{[0,1]^2} I_p\left(\tilde{g}(x,y)\right) dx\, dy$$

$\hookrightarrow I_p(u) = u \ln \frac{u}{p} + (1-u) \ln \frac{1-u}{1-p}$

* From this result + Varadhan's Lemma

[with exponential tilting $\mathcal{U}_{\alpha,h}(\tilde{g})$]

$\longrightarrow (\tilde{\mathbb{P}}_{n;\alpha,h})_{n\geq 1}$ satisfies a LDP on $(\tilde{W}, \delta_\square)$ with speed n^2 and rate function

$$\boxed{\mathcal{J}_{\alpha,h}(\tilde{g}) = \mathcal{J}^{ER}_{1/2}(\tilde{g}) - \mathcal{U}_{\alpha,h}(\tilde{g}) - \inf_{\tilde{g}\in\tilde{W}} \left\{ \mathcal{J}^{ER}_{1/2}(\tilde{g}) - \mathcal{U}_{\alpha,h}(\tilde{g}) \right\}}$$

* Minimizers of $\mathcal{J}_{\alpha,h}$ are given in [Chatterjee, Diaconis '13]:

(constant graphons) $\tilde{u}_1^*$, $\tilde{u}_2^*$: $\tilde{u}_i^*(x,y) = u_i^*$, $i=1,2$; $x,y \in [0,1]$

where u_1^* and u_2^* are solutions of the scalar problem defining $f(\alpha,h)$.

d. Heuristics on non-standard CLT at (α_c, h_c)

Setting $V_n = 2\left(S_n - \frac{n^2}{2}u^*\right)/n^{3/2}$:

$$\mathbb{P}_{n;\alpha_c,h_c}\left(V_n \in dx\right) = \mathbb{P}_{n;\alpha_c,h_c}\left(\frac{2S_n}{n^2} \in u^* + \frac{dx}{\sqrt{n}}\right)$$

based on LDP $\approx e^{-n^2 \mathcal{J}_{\alpha_c,h_c}\left(u^* + \frac{x}{\sqrt{n}}\right) + o(n^2)} dx$

Taylor's expansion of $\mathcal{J}_{\alpha,h}$ up to forth-order $\approx \boxed{e^{-\frac{81}{64}x^4}} + o(n^2) dx$

→ same density as mean-field model

Thanks for your attention !!!