Large deviations for random graphs (30 Jan)

**Cramér’s theorem and the Gärtner-Ellis theorem.** Last semester, we covered Cramér’s theorem—for finite alphabets, \(\mathbb{R}\), and \(\mathbb{R}^d\)—along with the Gärtner-Ellis theorem. Cramér’s theorem handles the case of the empirical mean of iid rvs taking values in \(\mathbb{R}\) or \(\mathbb{R}^d\). In particular, for rvs taking values in \(\mathbb{R}\), it says that

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) \leq - \inf_{x \in F} \Lambda^*(x) \quad \text{and} \quad \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(G) \geq - \inf_{x \in G} \Lambda^*(x),
\]

where \(\mu_n\) is the law of the empirical mean of iid \(X_i \in \mathbb{R}\), \(F \subset \mathbb{R}\) is any closed set, \(G \subset \mathbb{R}\) is any open set,

\[
\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\} \quad \text{and} \quad \Lambda(\lambda) := \log \mathbb{E}[e^{\lambda X_1}].
\]

The Gärtner-Ellis theorem extends the LDP to non-iid rvs, but requires an assumption on the logarithmic moment generating function, and the notion of an “exposing hyperplane” which plays into the lower bound.

**Large deviations for dense random graphs.**

We recall that our two aims in studying large deviations are:

1. to determine the probabilities of random events; and

2. to determine the conditional probabilities of events given the occurrence of a rare one.

In this context, by II we mean to determine the structure of a random graph, given that it satisfies some atypical property. For example, consider the following, wherein we will denote by \(G(n, p)\) the Erdős-Rényi random graph on \(n\) vertices with edge-presence probability \(p\). Denoting by \(T_{n,p}\) the number of triangles in \(G(n, p)\), we ask:

What is the most likely structure of the graph if the rare event \(E := \{T_{n,p} \geq (1 + \delta)\mathbb{E}T_{n,p}\}\) occurs, where \(\delta > 0\) is a given constant?

For example, are the extra triangles likely arising from a small subset of vertices with high interconnectivity or likely arising from an excess of edges, spread uniformly throughout the graph? The answer is surprising—both can happen. If \(p\) is smaller than a particular threshold, then there exist \(0 < \delta_1 < \delta_2\) such that:

- if \(0 < \delta \leq \delta_1\) or \(\delta \geq \delta_2\) then, conditional on \(E\), the graph behaves like \(G(n, r)\) for some \(r > p\); and

- if \(\delta_1 < \delta < \delta_2\), then the conditional structure is not like some \(G(n, r)\).
In other words, there is a “middle” range of triangle excess for which the graph is heterogeneous, but it is homogeneous for both smaller and larger values of triangle excess. Chatterjee [Cha16] observed

“There is probably no way that [this] result could have been guessed from intuition; it was derived purely from a set of mathematical formulas.”

Motivated by the pursuit of interesting questions à la II, we will begin our journey to proving the LDP for dense random graphs; the following lectures should expand upon the preliminaries and discuss the related topics of the LDP on sparse random graphs and nonlinear large deviations. In this context, dense refers to the average vertex degree being comparable to the total number of vertices—think of a fixed value of $p$, which is not decreasing to 0 as $n$ increases (i.e., the sparse case). But, before we can state the LDP for dense random graphs, we must become familiar with language from the theory of graph limits.

What is a graphon? Consider a graph $G$. We will exclusively consider finite, simple graphs—those with undirected, single edges and no loops. We will denote the vertex set of $G$ by $V(G)$ and its edge set by $E(G)$. By a graph homomorphism between two graphs $H$ and $G$, we mean a map from $V(H)$ to $V(G)$ which preserves edge adjacency. In other words, if $\varphi$ is a graph homomorphism between $H$ and $G$, then for every edge $\{v, w\} \in E(H)$, $\{\varphi(v), \varphi(w)\} \in E(G)$. We denote the number of such maps by $\text{hom}(H,G)$.

For example, $\text{hom}(\bullet, G) = |V(G)|$, $\text{hom}(\rightarrow, G) = 2|E(G)|$, and $\text{hom}(\triangle, G)$ is six times the number of triangles in $G$. If we normalize by the total number of possible maps, we get the homomorphism density of $H$ into $G$,

$$t(H,G) = \frac{\text{hom}(H,G)}{|V(G)||V(H)|},$$

which may be interpreted as the probability that a randomly chosen map from $V(H)$ to $V(G)$ preserves the adjacency of edges. This number also represents the density of $H$ as a subgraph in $G$ asymptotically as $n = |V(G)| \to \infty$. For example, $t(\rightarrow, G) = \frac{2}{n^2}|E(G)|$ while the density of edges in $G$ is $\frac{2}{n(n-1)}|E(G)|$.

Let $\{G_n\}_{n \geq 1}$ be a sequence of graphs, with a number of nodes tending to infinity. Lovász and Szegedy [LS06] proved that, if $t(H,G_n)$ tends to a limit $t(H)$ for every $H$, then there is a natural “limit object” in the form of a function $f \in W$, where $W$ is the space of all graphons—measurable functions $f : [0, 1]^2 \to [0, 1]$ that satisfy $f(x, y) = f(y, x)$ for all $x, y$. Conversely, every such function arises as the limit of an appropriate graph sequence.

Any graph can be represented as a graphon. Let $G$ be a graph on $\{1, \ldots, n\}$. We can express $G$ as a graphon $f^G$ in the following way. For any $(x, y) \in [0, 1]^2$, identify the unique integers $i$ and $j$ such that

$$\frac{i-1}{n} < x \leq \frac{i}{n} \quad \text{and} \quad \frac{j-1}{n} < y \leq \frac{j}{n}.$$
Then define
\[
f^G(x, y) = \begin{cases} 
1 & \text{if } \{i, j\} \text{ is an edge in } G, \\
0 & \text{otherwise.}
\end{cases}
\] (1)

The key point is that all simple graphs, regardless of the number of vertices, are all expressed in terms of an element of $\mathcal{W}$.

Let us see some examples of (1). Suppose we have $\Delta$ and we label the edges clockwise. Then we find that $f^G(x, y) = 1$ when:

\[
0 < x \leq \frac{1}{3} \quad \text{and} \quad \frac{1}{3} < y \leq \frac{2}{3};
\]
\[
0 < x \leq \frac{1}{3} \quad \text{and} \quad \frac{2}{3} < y \leq 1; \quad \text{and when}
\]
\[
\frac{1}{3} < x \leq \frac{2}{3} \quad \text{and} \quad \frac{2}{3} < y \leq 1,
\]
as well as when these intervals are reflected about $y = x$ (since the graph is undirected). This graphon is depicted in Figure 1.

Figure 1: The graphon naturally associated with the complete graph on three vertices. The pairs $(x, y)$ for which $f^G(x, y) = 1$ are shaded gray; the pairs $(x, y)$ for which $f^G(x, y) = 0$ are unshaded.

An alternative definition of the graphon naturally associated to a graph $G$ is to match the gray tiles with the 1’s in the adjacency matrix of $G$, match unshaded tiles with the 0 entries in the adjacency matrix, and scale the resulting picture to the unit square. Indeed, as the adjacency matrix of $\Delta$ is
\[
\begin{bmatrix}
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0
\end{bmatrix},
\]
it is clear that if we rotate the adjacency matrix and perform this operation, we recover the graphon depicted in Figure 1.

Homomorphism densities extend to graphons. To see how, observe that we could alternatively express $t(\to, G)$ by assigning each vertex of $G$ a mass of $1/n$ and by integrating the edge
indicator over all pairs of vertices. For example,
\[ t(\rightarrow, G) = \frac{1}{n^2} \sum_{i,j \in V(G)} 1_{(i,j) \in E(G)} = 2|E(G)|/n^2. \]

Analogously, if \( f \in \mathcal{W} \) is a graphon, we can define
\[ t(\rightarrow, f) = \int_{[0,1]^2} f(x,y) \, dx \, dy. \]

We generalize this as follows. If \( H \) is a simple graph on \( \{1, 2, \ldots, k\} \) with edge set \( E(H) \) and \( f \in \mathcal{W} \), define
\[ t(H, f) := \int_{[0,1]^k} \prod_{(i,j) \in E(H)} f(x_i, x_j) \, dx_1 \cdots dx_k. \]

We use \( t(H, f) \) to define convergence of a sequence of graphs in the following way. We say a sequence of graphs \( \{G_n\}_{n \geq 1} \) converges to \( f \) if, for every finite simple graph \( H \),
\[ \lim_{n \to \infty} t(H,G_n) = t(H,f). \]

**Example 1.** Fix \( p \in (0,1) \) and let \( G_{n,p} \) be a random graph from the \( G(n, p) \) model. For any fixed graph \( H \), with probability 1,
\[ t(H, G_{n,p}) \to p^{|E(H)|} \quad \text{as } n \to \infty. \]

On the other hand, if \( f \) is identically equal to \( p \), then \( t(H, f) = p^{|E(H)|} \). Therefore, the sequence \( \{G_{n,p}\}_{n \geq 1} \) converges with probability 1 to the deterministic function \( f(x,y) \equiv p \) as \( n \to \infty \).

**Cut topology and quotient space.** The space of graphons is the completion of the space of finite, simple graphs. To see this and to define limits on spaces of graphons, we need to choose a topology. We might choose the topology arising from the cut distance
\[ d(\square)(f, g) := \sup_{S,T} \left| \int_{S \times T} (f(x,y) - g(x,y)) \, dx \, dy \right|, \]
where the supremum ranges over measurable subsets \( S \) and \( T \) of \([0,1] \). However, we would prefer that the distance does not depend on relabelings of the vertices of the sequences of graphs giving rise to each graphon. We therefore further define an equivalence relation on \( \mathcal{W} \):
\[ f \sim g \text{ if } f(x,y) = g_\sigma(x,y) := g(\sigma x, \sigma y), \]
for a measure-preserving bijection \( \sigma \) of \([0,1] \). We denote by \( \overline{g} \) the closure in \((\mathcal{W}, d(\square))\) of the orbit of \( \{g_\sigma\} \). Then let \( \mathcal{W} \) by the quotient space arising from \( \tau \), the quotient map \( g \mapsto \overline{g} \). By
design, \( d_\square \) is invariant under \( \sigma \), so there is an induced metric \( \delta_\square \) on \( \widetilde{W} \) given by

\[
\delta_\square(\tilde{f}, \tilde{g}) := \inf_{\sigma} d_\square(f, g_\sigma) = \inf_{\sigma_1, \sigma_2} d_\square(f_{\sigma_1}, g_{\sigma_2}).
\]

To summarize, \( \delta_\square(\tilde{f}, \tilde{g}) \) first measures the largest discrepancy between the integrals of two labeled graphons over measurable boxes (which is the reason for the \( \square \) notation) in \([0, 1]^2\), before minimizing the largest discrepancy over all possible relabelings. Therefore, \( (\widetilde{W}, \delta_\square) \) is a metric space, which may serve as the setting for graph limits.

**Some important graph limit results.** To any finite graph \( G \), take the graphon \( f^G \) and its orbit \( \tilde{G} = \tau f^G = \tilde{f}^G \in \widetilde{W} \). The first result was proved in [LS06] and the second in [BCL+08], both relying on Szemerédi’s regularity lemma; we take them for granted.

**Theorem 1.** A sequence of graphs \( \{G_n\}_{n \geq 1} \) converges to a limit \( f \in \mathcal{W} \) if and only if \( \delta_\square(\tilde{G}_n, \tilde{f}^f) \to 0 \) as \( n \to \infty \).

Recall that \( \{G_n\}_{n \geq 1} \) converges to \( f \) if, for every \( H, t(H, G_n) \to t(H, f) \) as \( n \to \infty \). Theorem 1 says that, to determine convergence, it suffices to check that the \( \delta_\square \) distance between the equivalence class of the natural graphon associated with \( G_n \) and the equivalence class of \( f \) converges to 0 with \( n \). We also have the following.

**Theorem 2.** \( \widetilde{W} \) is compact under the cut metric \( \delta_\square \).

Theorem 2 implies that \( \widetilde{W} \) is complete, so, invoking Theorem 1, we conclude that the space of graphons is indeed the completion of the space of finite graphs, with the cut metric.

**Rate function notation.** We nearly possess all the language and notation necessary to state the LDP for dense random graphs. What remains is to specify notation for the rate function. Fix \( p \in (0, 1) \). For \( u \in [0, 1] \), let

\[
I_p(u) := u \log \frac{u}{p} + (1 - u) \log \frac{1 - u}{1 - p}.
\]

For \( h \in \mathcal{W} \), let

\[
I_p(h) := \int_{[0, 1]^2} I_p(h(x, y)) \, dx \, dy.
\]

Finally, for \( \tilde{h} \in \widetilde{W} \), let \( I_p(\tilde{h}) := I_p(h) \), where \( h \) is any element of \( \tilde{h} \). It is a fact that the righthand side is the same for all elements of \( \tilde{h} \) [CV11].

We will also require \( I(u) := u \log u + (1 - u) \log(1 - u) \) and, for any \( \tilde{h} \in \widetilde{W} \),

\[
I(\tilde{h}) := \int_{[0, 1]^2} I(h(x, y)) \, dx \, dy.
\]
Large deviation principle. The $G(n, p)$ model induces a probability measure $\tilde{\mathbb{P}}_{n,p}$ on $\tilde{\mathcal{W}}$. That is, for a realization $G$ of $G(n, p)$, associate to it the graphon $f^G$ and then apply the quotient map $\tau$ to obtain its orbit $\tilde{G}$. For $\tilde{\mathbb{P}}_{n,p}$, we have the following LDP.

**Theorem 3.** For any closed set $\tilde{F} \subseteq \tilde{\mathcal{W}}$ and open set $\tilde{U} \subseteq \tilde{\mathcal{W}}$,

$$\limsup_{n \to \infty} \frac{2}{n^2} \log \tilde{\mathbb{P}}_{n,p}(\tilde{F}) \leq - \inf_{\tilde{h} \in \tilde{F}} I_p(\tilde{h}) \quad \text{and} \quad \liminf_{n \to \infty} \frac{2}{n^2} \log \tilde{\mathbb{P}}_{n,p}(\tilde{U}) \geq - \inf_{\tilde{h} \in \tilde{U}} I_p(\tilde{h}).$$

Theorem 3 accomplishes I in the setting of dense random graphs. However, to accomplish II, we require an understanding of conditional probabilities; this is the role of Theorem 4.

**Theorem 4.** Denote by $\tilde{F}^*$ the set of minimizers of $I_p$ in $\tilde{F}$. Then, for each $n \geq 1$ and $\varepsilon > 0$,

$$\mathbb{P} \left( \delta_{\tilde{F}^*}(\tilde{G}_{n,p}, \tilde{F}^*) \geq \varepsilon \mid \tilde{G}_{n,p} \in \tilde{F} \right) \leq e^{-C(\varepsilon, \tilde{F}) n^2},$$

where $C(\varepsilon, \tilde{F})$ is a positive constant depending only on $\varepsilon$ and $\tilde{F}$.

Consider the case when $\tilde{F}^*$ contains only a single element–say, $\tilde{h}^*$. Theorem 4 then states that the conditional distribution of $\tilde{G}_{n,p}$ given $\tilde{G}_{n,p} \in \tilde{F}$ converges to the point mass at $\tilde{h}^*$ as $n \to \infty$; this is a conditional law of large numbers.

A remark about Theorem 3. As Chatterjee [Cha16] observes, Theorem 3 appears to be a result specific to the $G(n, p)$ model. In fact, Theorem 3 enables us to approximate the number of graphs on $n$ vertices that have any given property, so long as the property “behaves well” with respect to the cut metric. We can make this heuristic precise as follows.

Consider any Borel set $\tilde{A} \subseteq \tilde{\mathcal{W}}$ and let

$$\tilde{A}_n := \left\{ \tilde{h} \in \tilde{A} : \tilde{h} = \tilde{G} \text{ for some } G \text{ on } n \text{ vertices} \right\}.$$

In other words, $\tilde{A}_n$ contains graphons in $\tilde{A}$ which correspond to graphons arising from graphs on $n$ vertices. Corollary 1 then follows from Theorem 3 with $p = 1/2$.

**Corollary 1.** Denote by $\text{cl}(\tilde{A})$ and $\text{int}(\tilde{A})$ the closure and interior of $\tilde{A}$, respectively. Then, for any measurable $\tilde{A} \subseteq \tilde{\mathcal{W}}$,

$$- \inf_{\tilde{h} \in \text{int}(\tilde{A})} I(\tilde{h}) \leq \liminf_{n \to \infty} \frac{2 \log |\tilde{A}_n|}{n^2} \leq \limsup_{n \to \infty} \frac{2 \log |\tilde{A}_n|}{n^2} \leq - \inf_{\tilde{h} \in \text{cl}(\tilde{A})} I(\tilde{h}).$$

A remark about Theorem 4. The heuristic behind Theorem 4 is that, if $\tilde{f}^G \in \tilde{F}$ for some closed $\tilde{F} \subseteq \tilde{\mathcal{W}}$ satisfying

$$\inf_{\tilde{h} \in \text{int}(\tilde{F})} I_p(\tilde{h}) = \inf_{\tilde{h} \in \tilde{F}} I_p(\tilde{h}) > 0, \quad (2)$$

then $\tilde{f}^G$ should resemble one of the minimizers of $I_p$ in $\tilde{F}$. That is, given that $\tilde{f}^G \in \tilde{F}$, we might expect that $\delta_{\square}(\tilde{f}^G, \tilde{F}^*) \approx 0$, where we recall that $\tilde{F}^*$ is the set of minimizers of $I_p$ in $\tilde{F}$ and define

$$\delta_{\square}(\tilde{f}^G, \tilde{F}^*) := \inf_{\tilde{h} \in \tilde{F}^*} \delta_{\square}(\tilde{f}^G, \tilde{h}).$$

One barrier to this heuristic is that a minimizer may not exist in $\tilde{F}$, but the compactness of $\tilde{W}$ resolves this. Specifically, if we assume that $I_p$ is lower semi-continuous on $\tilde{F}$—this is true, we just have not yet shown it—and because $\tilde{F}^*$ is closed, a minimizer must exist. In fact, the proof of Theorem 4 requires little else than this observation and Theorem 3. We prove it now.

**Proof of Theorem 4.** $\tilde{W}$ is compact, so $\tilde{F}$ must also be compact, as it is a closed subset. $I_p$ is lower semi-continuous on $\tilde{F}$, so the compactness of $\tilde{F}$ implies that $I_p$ attains its minimum on $\tilde{F}$. This establishes that $\tilde{F}^*$ is nonempty. The lower semi-continuity of $I_p$ implies that $\tilde{F}^*$ is closed, therefore compact.

We fix $\varepsilon > 0$ and define $\tilde{F}_\varepsilon := \left\{ \tilde{h} \in \tilde{F} : \delta_{\square}(\tilde{h}, \tilde{F}^*) \geq \varepsilon \right\}$, which is closed. We observe that

$$\tilde{p}_{n,p} \left( \delta_{\square}(\tilde{G}_{n,p}, \tilde{F}^*) \geq \varepsilon \mid \tilde{G}_{n,p} \in \tilde{F} \right) = \frac{\tilde{p}_{n,p} \left( \tilde{G}_{n,p} \in \tilde{F}_\varepsilon \right)}{\tilde{p}_{n,p} \left( \tilde{G}_{n,p} \in \tilde{F} \right)}.$$

In order to use (2), we define

$$I_1 := \inf_{\tilde{h} \in \tilde{F}} I_p(\tilde{h}) \quad \text{and} \quad I_2 := \inf_{\tilde{h} \in \tilde{F}_\varepsilon} I_p(\tilde{h}).$$

With Theorem 4, we see that

$$\limsup_{n \to \infty} \frac{1}{n^2} \log \tilde{p}_{n,p} \left( \delta_{\square}(\tilde{G}_{n,p}, \tilde{F}^*) \geq \varepsilon \mid \tilde{G}_{n,p} \in \tilde{F} \right) \leq I_1 - I_2.$$

If we can conclude that $I_1 < I_2$, then we will be done. It is obvious that $I_1 \leq I_2$. However, if $I_1 = I_2$, the compactness of $\tilde{F}_\varepsilon$ implies the existence of an $\tilde{h} \in \tilde{F}_\varepsilon$ such that $I_p(\tilde{h}) = I_2$. This is impossible as it would imply that $\tilde{h} \in \tilde{F}^*$ and hence $\tilde{F}_\varepsilon \cap \tilde{F}^* \neq \emptyset$. \hfill $\Box$

**Reflecting on our progress.** We have introduced the concepts and notation from the theory of graph limits which are necessary for phrasing LDP answers to I and II. While we have seen, excepting some supporting results, how to obtain Theorem 4 from Theorem 3, we have not begun the arduous process of proving Theorem 3. It should therefore be the aim of further sessions to prove Theorem 3, before proceeding to discuss applications of the dense-case LDP, as well as the nonlinear LDP and applications to sparse-case LDP.
References


