The large deviation principle for the Erdos-Renyi random graph

Luo Lu

March.8, 2011
**Erdös-Rényi random graph**

Definition: Let $G(n, p)$ be the random graph on $n$ vertices where each edge is added independently with probability $p$.

Problem: Given a property $P$ and an integer $n$, roughly how many of these graphs have property $P$? What is the conditional distribution of it?

For example, $P$ maybe $\{\#\text{triangles} \geq tn^3\}$, where $t$ is a given constant. Want to define a topology together with the metric on it.

Useful notation: For every fixed graph $H$, let $N_H(G)$ denote the number of homomorphisms of $H$ into $G$ (i.e. edge-preserving maps $V(H) \rightarrow V(G)$, where $V(H)$ and $V(G)$ are the vertex sets). For example, $N_H(G) = 8$. 
Topology

- \( t(H, G) = \frac{N_H(G)}{|V(H)|^{V(G)}} \) gives the probability that a random mapping \( V(H) \to V(G) \) is a homomorphism.

- If \( t(H, G_n) \) tends to a limit \( t(H) \) for every \( H \), then Lovasz&Szegedy proved that there is a natural “limit object” in
  \[
  \mathcal{W} = \{ f : [0, 1] \times [0, 1] \to [0, 1], f(x, y) = f(y, x) \text{ for all } x, y \}
  \]

- Define \( t(H, f) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} f(x_i, x_j) \, dx_1 \cdots dx_k \).

- A sequence of graphs \( \{ G_n \}_{n \leq 1} \) is said to converge to \( f \) if for every finite simple graph \( H \),
  \[
  \lim_{n \to \infty} t(H, G_n) = t(H, f)
  \]

- A finite simple graph \( G \) on \( \{1, \cdots, n\} \) can also be represented as a graphon \( f^G \in \mathcal{W} \) in a natural way, by
  \[
  f^G(x, y) = \begin{cases} 
  1, & \text{if } ([nx], [ny]) \text{ is an edge in } G, \\
  0, & \text{otherwise.}
  \end{cases}
  \]
Topology

- A finite simple graph $G$ on $\{1, \cdots, n\}$ can also be represented as a graphon $f^G \in \mathcal{W}$ in a natural way, by

\[
f^G(x, y) = \begin{cases} 1, & \text{if } ([nx], [ny]) \text{ is an edge in } G, \\ 0, & \text{otherwise}. \end{cases}
\]

- Intuitively, the interval $[0, 1]$ represents a “continuum” of vertices, and $f(x, y)$ denotes the probability of putting an edge between $x$ and $y$.

- If $t(H, G_n)$ tends to a limit $t(H)$ for every $H$, then there is a natural “limit object” $f \in \mathcal{W}$, s.t.

\[
\lim_{n \to \infty} t(H, G_n) = t(H, f)
\]

Conversely, for each $f \in \mathcal{W}$, there exist $\{G_n\}$, whose limit is $f$. The elements of $\mathcal{W}$ are called “graphons”.

The large deviation principle for the Erdos-Renyi random graph
An example of the “convergence” under the topology

For any fixed graph $H$,

$$t(H, G(n, p)) = \frac{N_H(G)}{|V(G)||V|^H} \rightarrow p^{|E(H)|}$$

almost surely as $n \rightarrow \infty$.

On the other hand, if $f \equiv p$,

$$t(H, f) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} f(x_i, x_j) dx_1 \cdots, dx_k = p^{|E(H)|}.$$

Thus, $\lim_{n \rightarrow \infty} t(H, G_n) = t(H, f)$, the sequence of random graphs $G(n, p)$ converges a.s. to $f$ as $n \rightarrow \infty$. 

The large deviation principle for the Erdos-Renyi random graph
The “cut” metric

The notion of convergence in terms of subgraph densities can be captured by the so-called “cut distance”:

\[ d\Box(f, g) := \sup_{S, T \subseteq [0,1]} \left| \int_S \int_T [f(x, y) - g(x, y)] dx dy \right| \]

Say that \( f(x, y) \sim g(x, y) \) if \( f(x, y) = g_{\sigma}(x, y) := g(\sigma x, \sigma y) \) for some measure preserving bijection \( \sigma \) of \([0,1]\).

Since \( d\Box \) is invariant under \( \sigma \) one can define on \( \tilde{\mathcal{W}} \), which is the quotient space, the natural distance \( \delta\Box \) by

\[ \delta\Box(\tilde{f}, \tilde{g}) := \inf_{\sigma} d\Box(f, g_{\sigma}) = \inf_{\sigma} d\Box(f_{\sigma}, g) = \inf_{\sigma_1, \sigma_2} d\Box(f_{\sigma_1}, g_{\sigma_2}) \]

making \((\tilde{\mathcal{W}}, \delta\Box)\) into a metric space.

Theorem 1: A sequence of graphs \( \{G_n\}_{n \leq 1} \) converges to a limit \( f \in \mathcal{W} \) iff \( \delta\Box(\tilde{G}_n, \tilde{f}) \to 0 \) as \( n \to \infty \).
Conditional Distributions

- Define $I_p(u) := \frac{1}{2} u \log \frac{u}{p} + \frac{1}{2} (1 - u) \log \frac{1-u}{1-p}$, which is the Kullback-Leibler divergence between Bernoulli($p$) and Bernoulli($u$).

- Extend it to $W$ as

$$I_p(h) := \int_0^1 \int_0^1 I_p(h(x, y)) \, dx \, dy$$

which is the rate function—the distance between $h$ and $f \equiv p$ (the limit of $G(n, p)$).

- Theorem 2: Take any $p \in (0, 1)$. Let $\tilde{F}$ be a closed subset of $\mathfrak{W}$ satisfying

$$\inf_{\tilde{h} \in \tilde{F}} I_p(\tilde{h}) = \inf_{\tilde{h} \in \tilde{F}} I_p(\tilde{h}) > 0$$

Let $\tilde{F}^*$ be the subset of $\tilde{F}$ where $I_p$ is minimized. Then $\tilde{F}^*$ is non-empty and compact, and for each $n$, each $\epsilon > 0$,

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

where $C(\epsilon, \tilde{F})$ is a positive constant depending only on $\epsilon$ and $\tilde{F}$.
Large Deviations for triangle counts

Recall our problem at the beginning, we want to know the distribution of \( G(n, p) \) given \( T_{n,p} \geq tn^3 \), where \( T_{n,p} \) is the number of triangles. For each \( f \in \mathcal{W} \), let

\[
T(f) := \frac{1}{6} \int_0^1 \int_0^1 \int_0^1 f(x, y)f(y, z)f(z, x) dx dy dz
\]

then

\[
\{ T_{n,p} \geq tn^3 \} = \{ T(f) \geq t \} \text{ when } n \to \infty. \tag{1}
\]

Let \( H \) be the triangle graph, then

\[
t(H, f) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} f(x_i, x_j) dx_1 \cdots dx_k = 6T(f)
\]

Recall the definition \( t(H, G) = \frac{N_H(G)}{|V(G)||V(H)|} \),

\[
N_H(G) = |V(G)||V(H)| t(H, G) \to |V(G)||V(H)| t(H, f) = 6n^3 T(f).
\]

\[
T_{n,p} = N_H(G)/6 = n^3 T(f)
\]
The “Replica Symmetric Breaking” (A.J. Bray and M.A. Moore 1978)

- We consider a lattice model with infinite-ranged random interactions, $s_i$ is the spin variable at site $i$, $s_i = \pm 1$.
- The system Hamiltonian $\mathcal{BH} = -\sum_{i \neq j} J_{ij} s_i s_j$ helps define the free energy of the system.
- For a system realization, the coupling constants $J_{ij} \sim N(J_0, J^2)$.
- Let $m$ be the mean of $s_i$. $q$ be the mean of $s_i^2$. 
Figure 1: Surface plot of $m$ for a range of $1/\tilde{J}$ and $\tilde{J}_0/\tilde{J}$ (red indicates $m = 1$ and blue indicates $m = 0$)

Figure 2: Surface plot of $\sqrt{q}$ for a range of $1/\tilde{J}$ and $\tilde{J}_0/\tilde{J}$ (red indicates $\sqrt{q} = 1$ and blue indicates $\sqrt{q} = 0$)
The surface plots for \( m \) and \( \sqrt{q} \) identify 3 distinct phases

- Paramagnetic phase: \( m = 0 \) and \( q = 0 \);
- Ferromagnetic phase: \( m \neq 0 \) and \( q \neq 0 \);
- Spin-glass phase: \( m = 0 \) and \( q \neq 0 \).

Figure 3: Phase diagram for the infinite spin glass model
The “Replica Symmetric” phase

Fix $p \in [0, 1]$ and $t \in [0, 1/6)$, define the variational problem (derived from Thm.2 and Eqn.1)

$$\phi(p, t) := \inf \{ I_p(f) : f \in \mathcal{W}, T(f) \geq t \}$$

It has been shown that for $t \in (p^3/6, 1/6)$,

$$\phi(p, t) := \inf \{ I_p(f) : f \in \mathcal{W}, T(f) = t \}.$$  

Notes that there are two “extreme” functions that satisfy $T(f) = t$:

$$c_t(x, y) \equiv (6t)^{1/3}$$

and

$$\chi^t(x, y) := \begin{cases} 
1 & \text{if } \max\{x, y\} \leq (6t)^{1/3} \\
0 & \text{otherwise.}
\end{cases}$$

Recall

$$f^G(x, y) = \begin{cases} 
1 & \text{if } ([nx], [ny]) \text{is an edge in } G, \\
0 & \text{otherwise.}
\end{cases}$$

In a limiting sense, $c_t$ represents an Erdos-Renyi random graph which edge probability $(6t)^{1/3}$, while $\chi^t$ represents the union of a clique of size $m := n(6t)^{1/3}$ and a set of isolated vertices of size $n - m = n(1 - (6t)^{1/3})$.
The “Replica Symmetric” phase

\[ c_t(x, y) \equiv (6t)^{1/3} \]

and

\[ \chi^t(x, y) := \begin{cases} 
1 & \text{if } \max\{x, y\} \leq (6t)^{1/3} \\
0 & \text{otherwise.} 
\end{cases} \]

An example: \( n = 40, t = 0.000167, p = (6t)^{0.33} = 0.1, m = n(6t)^{1/3} = 4 \). The numbers of triangles in the example are \( T_1 = 15, T_2 = 4 \). However, \( T_1 = T_2 \) when \( n \to \infty \): \( \binom{n}{3} p^3 \approx \frac{1}{6} n^3 p^3 \) and \( \binom{m}{3} \approx \frac{1}{6} m^3 \approx \frac{1}{6} n^3 p^3 \).
The “Replica Symmetric” phase

- There exist $p_0 > 0$ such that if $p \leq p_0$, then there exists $p^3/6 < t' < t'' < 1/6$ such that the variational problem is solved by constant function $c_t \equiv (6t)^{1/3}$ when $t \in (p^3/6, t') \cup (t'', 1/6)$. For such $(p, t)$, the conditional distribution of $G(n, p)$ given $T_{n,p} \geq tn^3$ is asymptotically indistinguishable from the law of $G(n, (6t)^{1/3})$.

- One may call the region where $c_t$ solves the variational problem as the “replica symmetric phase” of the problem.
“Replica Symmetry Breaking”

- A simple computation shows that for any $t \in (0, 1/6)$,
  \[
  \lim_{p \to 0} \frac{l_p(c_t)}{\log(1/p)} = \frac{(6t)^{1/3}}{2} > \frac{(6t)^{2/3}}{2} = \lim_{p \to 0} \frac{l_p(\chi_t)}{\log(1/p)}
  \]

- Let $\tilde{C}$ denote the set of constant functions in $\tilde{\mathcal{Y}}$ (representing all Erdos-Renyi graphs). For each $t$, there exists $p' > 0$ and $\epsilon > 0$ such that for all $p < p'$,
  \[
  \lim_{n \to \infty} \mathbb{P}(\delta_{\square}(G(n, p), \tilde{C}) \geq \epsilon | T_{n,p} \geq tn^3) = 1
  \]

- For each $t \in (0, 1/6)$,
  \[
  \lim_{p \to 0} \lim_{n \to \infty} \mathbb{P}(\delta_{\square}(G(n, p), \tilde{\chi}_t) \geq \epsilon | T_{n,p} \geq tn^3) = 0.
  \]
  For such $(p, t)$, the conditional distribution of $G(n, p)$ given $T_{n,p} \geq tn^3$ is indistinguishable from the corresponding graph of $\tilde{\chi}_t$ in the large $n$ limit.

- From the physical point of view, it is simply the effect of replica symmetry breaking down in the “low temperature regime”.

\[\text{The large deviation principle for the Erdos-Renyi random graph}\]

“Replica-Symmetry Breaking in Spin-Glass Theories” by A.J.Bray and M.A.Moore (1978)