Law of the Iterated Logarithm in $G(n, p)$

Austen James Matt Larson Andrew Salmon

Yale University

4 August 2016

$G(n, p)$

- $G(n, p)$ is a random graph model.

$G(n, p)$

- $G(n, p)$ is a random graph model.
- Connect any two vertices with probability p where $0 < p < 1$.

$G(n, p)$

- $G(n, p)$ is a random graph model.
- Connect any two vertices with probability p where $0 < p < 1$.
- $G(n, p)$ exhibits a number of well-known properties relating to connectivity, coloring, the appearance of certain subgraphs, etc.

The Law of Iterated Logarithm

- The Law of Iterated Logarithm (LIL) is a property of sequences of random variables, similar to the Central Limit Theorem (CLT).

The Law of Iterated Logarithm

- The Law of Iterated Logarithm (LIL) is a property of sequences of random variables, similar to the Central Limit Theorem (CLT).
- While the CLT involves the convergence to a normal distribution, the LIL examines the greatest deviations from 0.

The Law of Iterated Logarithm

- The Law of Iterated Logarithm (LIL) is a property of sequences of random variables, similar to the Central Limit Theorem (CLT).
- While the CLT involves the convergence to a normal distribution, the LIL examines the greatest deviations from 0.

Theorem (Khinchin (1924) and Kolmogorov (1929))

Let $S_n = \sum_{i=1}^{n} T_i$ with T_i iid random variables with 0 mean and 1 variance.

$$\mathbb{P}\left[\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right] = 1$$

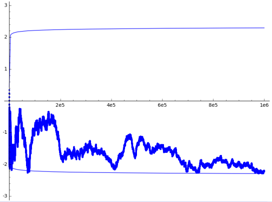
The Law of Iterated Logarithm

- The Law of Iterated Logarithm (LIL) is a property of sequences of random variables, similar to the Central Limit Theorem (CLT).
- While the CLT involves the convergence to a normal distribution, the LIL examines the greatest deviations from 0.

Theorem (Khinchin (1924) and Kolmogorov (1929))

Let $S_n = \sum_{i=1}^{n} T_i$ with T_i iid random variables with 0 mean and 1 variance.

$$\mathbb{P}\left[\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1\right] = 1$$



History of the LIL

- Not true in general without independence.

History of the LIL

- Not true in general without independence.
- 1940's: Random Walks with some conditions

History of the LIL

- Not true in general without independence.
- 1940's: Random Walks with some conditions
- 1970's: Martingales

History of the LIL

- Not true in general without independence.
- 1940's: Random Walks with some conditions
- 1970's: Martingales
- 1985: Random Walks with weaker conditions

History of the LIL

- Not true in general without independence.
- 1940's: Random Walks with some conditions
- 1970's: Martingales
- 1985: Random Walks with weaker conditions
- 2016: Random Variables coming from Random Graphs

Central Limit Theorems in Random Graph Counts

- 1988: Ruciński proves graph counts of small subgraphs are asymptotically normal.

Central Limit Theorems in Random Graph Counts

- 1988: Ruciński proves graph counts of small subgraphs are asymptotically normal.
- 1994: Janson proves that some spanning graphs, like spanning trees, Hamiltonian cycles, and perfect matchings, are asymptotically log-normal.

Janson (1994)

Let T_n be the number of trees contained in a random graph in $G(n, p)$.

$$\frac{\log T_n - \mu_n}{\sigma_n} \to N(0, 1)$$

Central Limit Theorems in Random Graph Counts

- 1988: Ruciński proves graph counts of small subgraphs are asymptotically normal.
- 1994: Janson proves that some spanning graphs, like spanning trees, Hamiltonian cycles, and perfect matchings, are asymptotically log-normal.

Janson (1994)

Let T_n be the number of trees contained in a random graph in $G(n,p)$.

$$\frac{\log T_n - \mu_n}{\sigma_n} \to N(0,1)$$

- μ_n (the "mean") is $\log \mathbb{E}[T_n] - \frac{1-p}{p}$ (recall p is the edge probability).

Central Limit Theorems in Random Graph Counts

- 1988: Ruciński proves graph counts of small subgraphs are asymptotically normal.
- 1994: Janson proves that some spanning graphs, like spanning trees, Hamiltonian cycles, and perfect matchings, are asymptotically log-normal.

Janson (1994)

Let T_n be the number of trees contained in a random graph in $G(n, p)$.

$$\frac{\log T_n - \mu_n}{\sigma_n} \to N(0, 1)$$

- μ_n (the "mean") is $\log \mathbb{E}[T_n] - \frac{1-p}{p}$ (recall p is the edge probability).
- σ_n (the "standard deviation") is $\sqrt{\frac{2(1-p)}{p}}$

A more precise statement

Main result

Let T_n be the number of trees contained in a random graph in $G(n, p)$.

- $\mathbb{P}\left[\limsup_{n\to\infty} \frac{\log T_n - \mu_n}{\sigma_n\sqrt{2\log\log n}} = 1\right] = 1$

A more precise statement

Main result

Let T_n be the number of trees contained in a random graph in $G(n, p)$.

- $$\mathbb{P}\left[\limsup_{n\to\infty} \frac{\log T_n - \mu_n}{\sigma_n\sqrt{2\log\log n}} = 1\right] = 1$$

- μ_n and σ_n are as in Janson (1994).

Proof Strategy

- Work with a different random graph model, $G(n, m)$ where graphs are drawn uniformly on n vertices under the constraint that they have exactly m edges.

Proof Strategy

- Work with a different random graph model, $G(n, m)$ where graphs are drawn uniformly on n vertices under the constraint that they have exactly m edges.
- If X_{nm} is the number of trees in a random graph from $G(n, m)$, we need to show that, for some conditions:

$$\mathbb{P}[X_{nm} \geq K\mathbb{E}X_{nm}] \leq n^{-4}$$

Proof Strategy

- Work with a different random graph model, $G(n, m)$ where graphs are drawn uniformly on n vertices under the constraint that they have exactly m edges.
- If X_{nm} is the number of trees in a random graph from $G(n, m)$, we need to show that, for some conditions:

$$\mathbb{P}[X_{nm} \geq K\mathbb{E}X_{nm}] \leq n^{-4}$$

- To prove this concentration lemma, it is enough to show, by Markov's inequality, that

$$\mathbb{E}X_{nm}^k \leq C^k(\mathbb{E}X_{nm})^k$$

Proof Strategy

- Work with a different random graph model, $G(n, m)$ where graphs are drawn uniformly on n vertices under the constraint that they have exactly m edges.
- If X_{nm} is the number of trees in a random graph from $G(n, m)$, we need to show that, for some conditions:

$$\mathbb{P}[X_{nm} \geq K\mathbb{E}X_{nm}] \leq n^{-4}$$

- To prove this concentration lemma, it is enough to show, by Markov's inequality, that

$$\mathbb{E}X_{nm}^k \leq C^k(\mathbb{E}X_{nm})^k$$

- Given this, we can prove the upper and lower bound for the limsup term using the Borel-Cantelli lemma.

About the Moment $\mathbb{E}X_{nm}^k$

- In $G(n, m)$, the appearance of a (labeled) subgraph only depends on the number of edges that subgraph contains.

About the Moment $\mathbb{E}X_{nm}^k$

- In $G(n, m)$, the appearance of a (labeled) subgraph only depends on the number of edges that subgraph contains.
- So the probability that the union of trees $\bigcup_{i=1}^{k} T_i$ is a subgraph of some graph from $G(n, m)$ depends only on the size of the edge overlap.

About the Moment $\mathbb{E}X_{nm}^k$

- In $G(n,m)$, the appearance of a (labeled) subgraph only depends on the number of edges that subgraph contains.
- So the probability that the union of trees $\bigcup_{i=1}^{k} T_i$ is a subgraph of some graph from $G(n,m)$ depends only on the size of the edge overlap.
- If X_{nm} is the number of trees in a random graph with m edges, then the kth moment $\mathbb{E}\left[X_{nm}^k\right]$ is equal to:

$$\mathbb{E}\left[X_{nm}^k\right] = \sum_{(T_1,\ldots,T_k)} \mathbb{P}\left[\bigcup_i T_i\right]$$

The Quantity $M(a)$

- But it makes more sense to try to group tuples of trees by the number of edges they contain.

$$\sum_{e=n-1}^{k(n-1)} \mathbb{P}\left[e\left(\bigcup_i T_i\right)\right] \cdot \#\left\{(T_1, \ldots, T_k) : e\left(\bigcup_i T_i\right) = e\right\}$$

The Quantity $M(a)$

- But it makes more sense to try to group tuples of trees by the number of edges they contain.

$$\sum_{e=n-1}^{k(n-1)} \mathbb{P}\left[e\left(\bigcup_i T_i\right)\right] \cdot \#\left\{(T_1, \ldots, T_k) : e\left(\bigcup_i T_i\right) = e\right\}$$

- Rewriting in terms of missing edges yields:

$$\sum_{a=0}^{(k-1)(n-1)} M(a) \frac{\binom{\binom{n}{2} - k(n-1) + a}{m - k(n-1) + a}}{\binom{\binom{n}{2}}{m}}$$

The Quantity $M(a)$

- But it makes more sense to try to group tuples of trees by the number of edges they contain.

$$\sum_{e=n-1}^{k(n-1)} \mathbb{P}\left[e\left(\bigcup_i T_i\right)\right] \cdot \#\left\{(T_1, \ldots, T_k) : e\left(\bigcup_i T_i\right) = e\right\}$$

- Rewriting in terms of missing edges yields:

$$\sum_{a=0}^{(k-1)(n-1)} M(a) \frac{\binom{\binom{n}{2} - k(n-1) + a}{m - k(n-1) + a}}{\binom{\binom{n}{2}}{m}}$$

Definition of $M(a)$

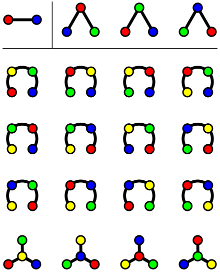
$M(a)$ is the number of k-tuples $(T_1, \ldots, T_k)$ of trees on a graph on n labeled vertices such that:

$$e\left(\bigcup_{i=1}^{k} T_i\right) = k(n-1) - a$$

$M(a)$ and the Combinatorics of Trees

- From Cayley's formula, we know that there are n^{n-2} total trees on n labeled vertices.

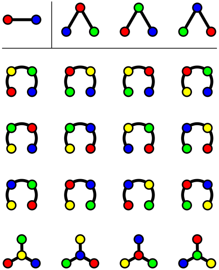
Figure: Cayley's formula for $n = 2, 3, 4$, image taken from Wikipedia



$M(a)$ and the Combinatorics of Trees

- From Cayley's formula, we know that there are n^{n-2} total trees on n labeled vertices.
- For two random trees, what is the distribution of the size of the edge intersection?

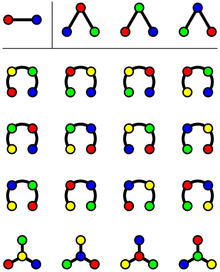
Figure: Cayley's formula for $n = 2, 3, 4$, image taken from Wikipedia



$M(a)$ and the Combinatorics of Trees

- From Cayley's formula, we know that there are n^{n-2} total trees on n labeled vertices.
- For two random trees, what is the distribution of the size of the edge intersection?
- For k (uniformly chosen) random trees, how many edges will overlap in total?

Figure: Cayley's formula for $n = 2, 3, 4$, image taken from Wikipedia



Electrical Networks and Spanning Trees

- We can convert a graph to an electrical network by setting the resistance to 1.

Electrical Networks and Spanning Trees

- We can convert a graph to an electrical network by setting the resistance to 1.

Spanning tree edge probability (Kirchoff)

Let T be a spanning tree chosen uniformly from a graph H. Let e be any edge in H.

$$\mathbb{P}[e \in E(T)] = i(e)$$

That is, the probability that this edge is in T is equal to the current traveling along this edge when we inject one amp at one of the endpoints of e and remove one amp from the other endpoint:

Electrical Networks are Graphs

- Many laws relating to electrical networks can be formalized more generally using graph theory.

Electrical Networks are Graphs

- Many laws relating to electrical networks can be formalized more generally using graph theory.

Ohm's law

Let ab be an edge in graph G. Let V_{ab} be the voltage difference across ab and let R_{ab} be the resistance of ab. Then:

$$i(ab) = \frac{V_{ab}}{R_{ab}}$$

Electrical Networks are Graphs

- Many laws relating to electrical networks can be formalized more generally using graph theory.

Ohm's law

Let ab be an edge in graph G. Let V_{ab} be the voltage difference across ab and let R_{ab} be the resistance of ab. Then:

$$i(ab) = \frac{V_{ab}}{R_{ab}}$$

Rayleigh Monotonicity Law

Let G, G' be graphs such that $G \subseteq G'$. Then for every edge $e \in E(G)$:

$$i_G(e) \geq i_{G'}(e)$$

An iterative construction

- We need to construct k-tuples random trees with total edge intersection a.

An iterative construction

- We need to construct k-tuples random trees with total edge intersection a.
- Choose some partition $(\ell_2, \ell_3, \ldots, \ell_k)$ such that $\sum_i \ell_i = a$.

An iterative construction

- We need to construct k-tuples random trees with total edge intersection a.
- Choose some partition $(\ell_2, \ell_3, \ldots, \ell_k)$ such that $\sum_i \ell_i = a$.
- Given a sequence $T_1, \ldots, T_{t-1}$ of trees, an additional tree T is ℓ_t-overlapping if T shares exactly ℓ_t edges with $\bigcup_{i=1}^{t-1} T_i$.

An iterative construction

- We need to construct k-tuples random trees with total edge intersection a.
- Choose some partition $(\ell_2, \ell_3, \ldots, \ell_k)$ such that $\sum_i \ell_i = a$.
- Given a sequence $T_1, \ldots, T_{t-1}$ of trees, an additional tree T is ℓ_t-overlapping if T shares exactly ℓ_t edges with $\bigcup_{i=1}^{t-1} T_i$.
- Counting the number of such trees gives us how often our construction can continue.

Counting ℓ_t-overlapping Trees

- Reduction (Moon, 1967): the maximum degree of random trees is almost always smaller than $\log n$.

Counting ℓ_t-overlapping Trees

- Reduction (Moon, 1967): the maximum degree of random trees is almost always smaller than $\log n$.

Overlapping Tree Lemma

Suppose we have chosen trees $T_1, \ldots, T_{t-1}$ with maximum degree less than $\log n$. Let S be a fixed set of ℓ_t edges and call a tree T_t ℓ_t-**overlapping** if it only uses precisely the edges in S as well as unused edges not in $T_1, \ldots, T_{t-1}$. Then the number of ℓ_t-overlapping trees is at most

$$2^{\ell_t} e^{2-2t+O(\log^5(n)/n)} n^{n-3-\ell_t}$$

Counting ℓ_t-overlapping Trees

- Reduction (Moon, 1967): the maximum degree of random trees is almost always smaller than $\log n$.

Overlapping Tree Lemma

Suppose we have chosen trees $T_1, \ldots, T_{t-1}$ with maximum degree less than $\log n$. Let S be a fixed set of ℓ_t edges and call a tree T_t ℓ_t-**overlapping** if it only uses precisely the edges in S as well as unused edges not in $T_1, \ldots, T_{t-1}$. Then the number of ℓ_t-overlapping trees is at most

$$2^{\ell_t} e^{2-2t+O(\log^5(n)/n)} n^{n-3-\ell_t}$$

- The results from electrical networks are well-behaved because of the high minimum degree in $K_n \setminus \bigcup_i T_i$.

A Proper Upper Bound on $M(a)$ for any k

- We can get an upper bound on $M(a)$ because any trees that satisfy the condition for $M(a)$ contain a procedure of selecting ℓ_t-overlapping trees.

A Proper Upper Bound on $M(a)$ for any k

- We can get an upper bound on $M(a)$ because any trees that satisfy the condition for $M(a)$ contain a procedure of selecting ℓ_t-overlapping trees.
- By summing over all possible partitions $\mathcal{P}(a)$, we use the multinomial theorem.

A Proper Upper Bound on $M(a)$ for any k

- We can get an upper bound on $M(a)$ because any trees that satisfy the condition for $M(a)$ contain a procedure of selecting ℓ_t-overlapping trees.
- By summing over all possible partitions $\mathcal{P}(a)$, we use the multinomial theorem.

Upper Bound

The total number of trees with a edge overlaps is at most:

$$M(a) \leq \sum_{\ell \in \mathcal{P}(a)} \prod \ell_t\text{-overlapping Trees}$$

$$= \frac{n^{k(n-2)} e^{-2\binom{k}{2} + O(\log^5(n)/n)} \left(2\binom{k}{2}\right)^a}{a!}$$

Some Remarks on the Upper Bound

- For $k = o(n)$,

$$\frac{M(a)}{N^k} \leq \frac{e^{2\binom{k}{2}2\binom{k}{2}^a}}{a!}(1 + o(1))$$

Some Remarks on the Upper Bound

- For $k = o(n)$,
$$\frac{M(a)}{N^k} \le \frac{e^{2\binom{k}{2}} 2\binom{k}{2}^a}{a!}(1 + o(1))$$
- The upper bound is a Poisson distribution with parameter $\lambda = 2\binom{k}{2}$.

Some Remarks on the Upper Bound

- For $k = o(n)$,
$$\frac{M(a)}{N^k} \leq \frac{e^{2\binom{k}{2}} 2\binom{k}{2}^a}{a!}(1 + o(1))$$
- The upper bound is a Poisson distribution with parameter $\lambda = 2\binom{k}{2}$.
- We can infer convergence to a Poisson distribution because $\sum_{i=0}^{\infty} M(a)$ must equal the total number of k-tuples of trees.

The Law of Iterated Logarithm

- This asymptotic estimate on $M(a)$ is sufficient to prove the concentration equality.

The Law of Iterated Logarithm

- This asymptotic estimate on $M(a)$ is sufficient to prove the concentration equality.
- Using similar techniques, we also proved the Law of Iterated Logarithm for the appearance of perfect matchings in random graphs in $G(n, p)$.

Acknowledgements

- Thanks to MAA Mathfest for hosting.
- Thanks to Daniel Montealegre for serving as our advisor.
- Thanks to SUMRY for giving us the opportunity to work on this problem.