Asymptotic Behavior of LPP on Complete Graphs

Charles Beck, Stanley Jian, Catherine Lyu, Harrison Wang

Department of Mathematics
Columbia University in the City of New York

Law of Large Numbers

- Consider flipping a fair coin repeatedly and observing the outcome.
- Each trial is a random variable, X_i, which equals 1 if heads and 0 if tails.
- After n trials $(X_1, \ldots, X_n)$, what can we observe about the average of the outcomes?
- This result states that as n increases, the average of n outcomes approaches the average of each X_i, which, in coin-flipping, is $\mathbb{E}(X_i) = 0.5$.

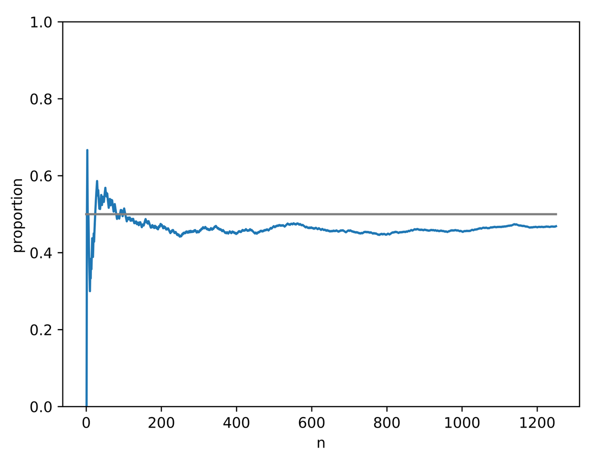

Figure 1. p = 0.5, n = 1250

$$\frac{S_n}{n} \to \mathbb{E}[X]$$

n = number of trials

$$S_n = X_1 + \ldots + X_n$$

$$\mathbb{E}[X] = 0.5$$

Extreme Value Theory

- Consider once more n repeated trials $(X_1, \ldots, X_n)$.
- Call the ith largest observation among n of these $M_n^{(i)}$ and the maximum M_n.
- M_n and $M_n^{(i)}$ contain some randomnness; if the values of our random variables change, the largest and ith largest value will likely change.
- Associate the deterministic value $b_n \approx M_n$ and $b_n^{(i)} \approx M_n^{(i)}$ (we say M_n "grows at a rate of" b_n).
- b_n is increasing in n for all distributions, so if we add more random variables, the maximum would likely be greater.
- The rate of growth of b_n, however, changes depending on distribution. Some examples are listed below:

	Exponential	Rayleigh	Power	Uniform
p.d.f.	$\lambda e^{-\lambda x}$	$2xe^{-x^2}$	$\alpha x^{-\alpha-1}$	1
c.d.f.	$1 - e^{-\lambda x}$	$1 - e^{-x^2}$	$1 - x^{-\alpha}$	x
b_n	$\frac{\log n}{\lambda}$	$\sqrt{2 \log n}$	$n^{\frac{1}{\alpha}}$	$1 - \frac{1}{n}$
$b_n^{(i)}$	$\frac{\log \frac{n}{i}}{\lambda}$	$\sqrt{2 \log \frac{n}{i}}$	$(\frac{n}{i})^{\frac{1}{\alpha}}$	$1 - \frac{i}{n}$

We denote distributions for which b_n grows quickly "heavy-tailed" distributions (e.g. Power Law) and slowly "light-tailed" distributions (e.g. exponential). The heavier the tail, the more likely one is to randomly select a greater value, and so faster the growth of the maximum.

- Maximum meteor sizes has close to heavy-tailed growth. If we find a new meteor that is larger than all previous meteors, this meteor size can easily exceed the previous maximum size significantly.
- An intuitive example of a light-tailed distribution would be the world record for hotdogs eaten in 10 minutes. It is far more difficult to eat more than 50 hot dogs in ten minutes than to eat more than 5 hotdogs, so the world record (maximum) will exceed the previous world record infrequently and by less over time.

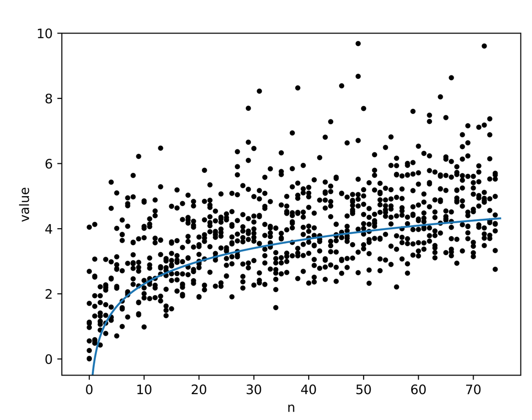

Figure 2. $\log(n)$ and simulated M_n for Exponential

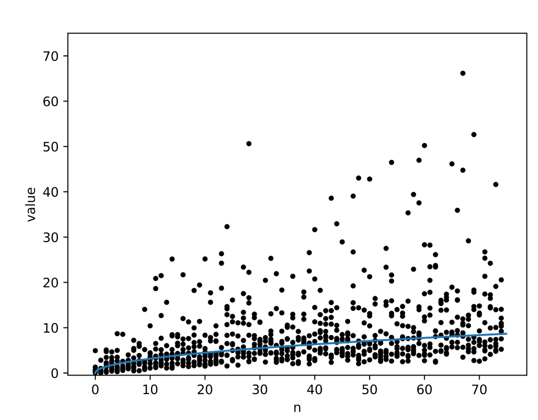

Figure 3. $\sqrt{n}$ and simulated M_n for Power (α= 2)

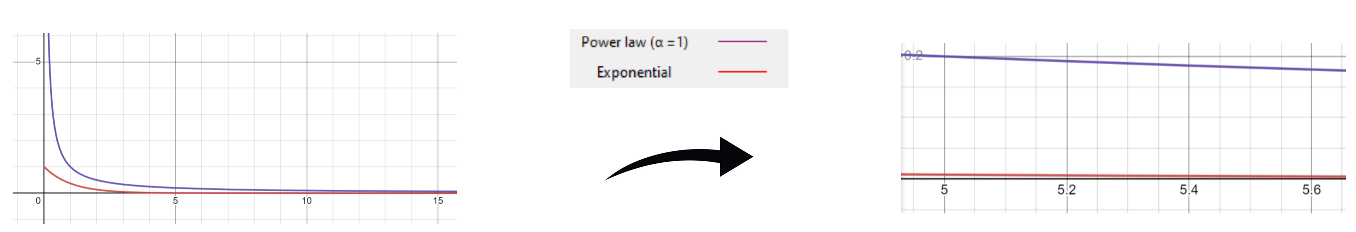

Figure 4. Comparison of heavy tail (eg. power law $\alpha = 1$) and light tail (eg. exponential) distributions

Problem Description: Last Passage Percolation (LPP)

- Consider n nodes labelled $1, \ldots, n$ where each pair of nodes is connected by an edge. Call this a complete graph with n nodes, or G_n.
- Assign to each edge a weight given by a non negative random variable $X \geq 0$.
- These edges are distributed identically - each take the same values with the same probability - and are independent from one another - each edge's value has no effect on the value of any other edge.
- Consider all paths going from 1 to n such that no vertex is repeated in the path (self-avoiding) and let the weight of any path be the sum of all edges that the path goes through.

  We wish to investigate

$$W_n = \text{largest path weight among all possible paths}$$

In particular, since the edge weights are determined by random variables, and thus the value of W_n is random as well, we cannot evaluate W_n precisely. Instead, the goal of this project is to find asymptotic (for high n) upper and lower bounds on the value of W_n for various probability distributions. This problem has previously been considered in [1].

Related Topics

Connection to Law of Large Numbers and Extreme Value Theory

- For any specific path through all vertices, each edge is independent and identically distributed. Thus, the weight of the path divided by n should approach $\mathbb{E}(X)$ according to the Law of Large Numbers. The Law of Large Numbers becomes less effective when we are not considering arbitrary edges, but rather the maximum one.
- Considering each path as a random variable, the maximum over all paths resembles the premise of the Extreme Value Theory. However, the paths themselves are not independent from each other, since two paths which share a lot of edges would likely be similar in value.
- We aim to combine these results to analyze a more convoluted mixture of the two.

The Travelling Salesman Problem

- The Travelling Salesman Problem asks for the longest path from one node to another on a weighted graph.
- The problem is NP-Hard, meaning it cannot currently be solved "quickly".
- Here, we are finding the longest path instead of the shortest one, also an NP-hard problem.
- Our problem also differs as we have random variables for edge weights instead of deterministic values. This makes it practically impossible to locate the best path, given that it changes in different iterations.
- It turns out that the non-deterministic aspect of the edges allows us to obtain a better understanding of the value of the largest path, W_n, and its behavior as n approaches infinity

Results

We wish to find a "law of large numbers" for the value of $\frac{W_n}{n}$. That is, we wish to understand how the average weight of edges on the best path changes as we increase n. In fact, we have the result that

$$\lim_{n\to\infty} \frac{W_n}{n} \approx h(n)$$

where $h(n)$ is dependent on the distribution chosen for the edge, as well as the number of nodes, n, and can be classified as follows:

Distribution	Exp e^{-x}	Rayleigh e^{-x^2}	Power $x^{-\alpha}$ $(\alpha < 1)$	Power $x^{-\alpha}$ $(\alpha = 1)$	Power $x^{-\alpha}$ $(\alpha > 1)$	Bounded $X \leq M < \infty$
$h(n)$	$\log(n)$	$\sqrt{2\log(n)}$	$n^{\frac{1}{\alpha}}$	$n\log(n)$	$n^{\frac{2}{\alpha}-1}$	M

For general random variables, M is an upper bound on the largest value the distribution can take. For exponential or power distributions, it is possible, however unlikely, to obtain an arbitrarily large value, and so they are not bounded (or $M = \infty$). For random variables like the standard uniform random variable, there is a point ($M = 1$) where any greater value is unattainable. The general result here is that

$$\frac{W_n}{n} \to M$$

as $n \to \infty$ since the more paths we have, the more likely it is for some path to have values close to M along it. Then, the average weight of edges in the path would be close to M.

For unbounded random variables, $h(n)$ grows with n and goes to ∞ in each case, since it is possible to get arbitrarily high values. You may also notice that many of the $h(n)$ values are exactly the b_n values present under Extreme Value theory. This is no coincidence, as becomes evident in the "Methods" section.

Methods

W_n behaves differently on different probability distributions, so we need to use different methods to find the probabilistic/asymptotic behavior of the upper and lower bounds. Assume, for all cases below, that we are working on G_n for a positive integer n. Also, note that these methods apply to many unbounded distributions, but bounded distributions can be understood through other means.

Upper Bound

Since the self-avoiding path from node 1 to node n can travel through no more than $n-1$ edges, it is impossible for the best path to be longer than the n largest edges in G_n (out of $\binom{n}{2}$ edges).

$$\boxed{W_n \leq \sum_{i=1}^{n} M^{(i)}_{\binom{n}{2}}}$$

Lower Bounds

$$\boxed{W_n \geq M_{\binom{n}{2}}}$$

Single Edge

We can always create a path from 1 to n passing through the largest edge. Thus, it is impossible for the maximum length path to be shorter than the weight of the largest edge, which takes the value of the random variable $M_{\binom{n}{2}}$.

Greedy Approach

1. Start at node 1.
2. Choose the largest edge from the remaining, unvisited nodes, other than node n. The weight of this edge can be represented by the random variable M_{n-i} where i is the number of unvisited nodes.
3. Repeat step 2 until all nodes besides node n are visited.
4. Connect the last node to node n.

$$\boxed{W_n \geq \sum_{i=1}^{n-2} M_i}$$

We have thus constructed a path from 1 to n, perhaps not of highest weight, but of a relatively high weight. Using our Extreme Value theory intuition, we can approximate M_i by b_i, which allows us to lower bound W_n.

k Largest Edges

We want to construct a path passing through the largest k edges in a graph. A sufficient condition is that all k edges be disjoint (no two share a vertex). With simple combinatorial tools, it is possible to show that for fixed k, the k largest edges in G_n will be completely disjoint **w.h.p.**. Further, if $\frac{k_n}{n} \to 0$, i.e. k_n grows slower than n, the k_n largest edges in G_n will be disjoint **w.h.p.** as well.

$$\boxed{W_n \geq \sum_{i=1}^{k_n} M^{(i)}_{\binom{n}{2}}}$$

$$\boxed{\begin{gathered} W_n \geq \sum_{i=k_1}^{k_2} M^{(i)}_{n^2} \\ k_1 = k_2 - nA_n + 4 \\ k_2 = Cn\log(n) \end{gathered}}$$

Graph Theoretic

1. The longest path cannot be longer than n edges, so take more than n of the largest edges in G_n, $n \log n$ of them, to be precise.
2. Find probability that there is a path from 1 to n using these edges, and show it goes to 1.
3. Lower bound W_n with smallest of these largest edges.
This result also does not hold for certain, but rather holds **w.h.p.**

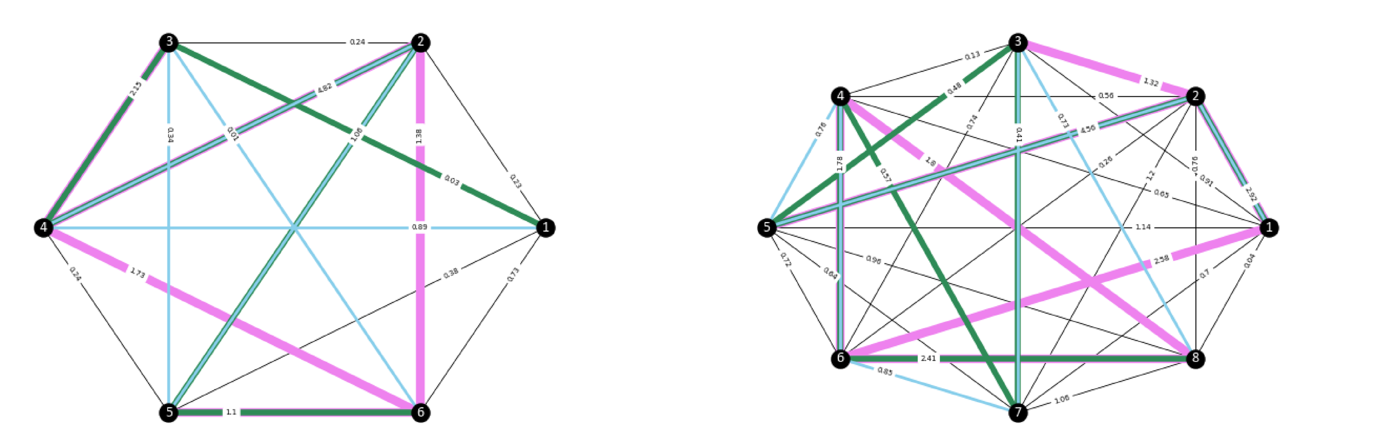

Figure 5. Above is a juxtaposition of the upper bound, the solution W_n, the single edge lower bound, and the greedy lower bound on G_6 with power law (Pareto)-distributed edges (left) and G_8 with exponential ($\lambda = 1$) distributed edges.

Acknowledgements

We would like to thank Professor Milind Hegde for his guidance throughout the proejct, Anushka Murthy for her technical assistance, Professor George Dragomir for organizing this REU, and Columbia University Math Department for the generous funding.

References

[1] Feng Wang, Xian-Yuan Wu, and Rui Zhu. Last passage percolation on the complete graph. *Statistics Probability Letters*, 164:108798, 2020.