On the Analysis of Independent Sets via Multilevel Splitting

Slava Vaisman

(joint work with Dirk Kroese)

School of Mathematics and Physics
The University of Queensland Australia
r.vaisman@uq.edu.au, kroese@maths.uq.edu.au

June, 2016
Overview

1. Introduction

2. Monte Carlo solution

3. Markov Inequality based Lower Bounds
Counting *independent sets* in large graphs

In graph theory, an independent set or stable set is a set of vertices in a graph, no two of which are adjacent.
The solution to this problem is given by *independence polynomial* which is equal to:

$$\mathcal{IP}(G, x) = \sum_{k=0}^{\alpha(G)} s_k x^k.$$
The solution to this problem is given by *independence polynomial* which is equal to:

\[ \mathcal{IP}(G, x) = \sum_{k=0}^{\alpha(G)} s_k x^k. \]

It is interesting from the theoretical point of view. In particular, obtaining an estimator \( \hat{s}_k \) of \( s_k \) — the number of independent set of cardinality \( k \) is in \( \#P \) (very hard even to approximate).
The solution to this problem is given by *independence polynomial* which is equal to:

\[
    \mathcal{IP}(G, x) = \sum_{k=0}^{\alpha(G)} s_k x^k.
\]

It is interesting from the theoretical point of view. In particular, obtaining an estimator \( \hat{s}_k \) of \( s_k \) — the number of independent set of cardinality \( k \) is in \#P (very hard even to approximate).

Finding a distribution of different cardinality independent sets in a graph \( G \) is equivalent to counting the number of cliques in \( G \)'s complement. The latter is of great interest to the analysis of social networks.
Monte Carlo solution

- Suppose that we have a graph $G = (V, E)$ and $1 \leq k \leq |V|$.
- Choose $Y$ uniformly at random from a set of all $k$-cardinality subsets of $\{v_1, \ldots, v_{|V|}\}$.
- Define $X = \mathbb{1}\{Y \text{ is independence set}\}$.
- Then,

$$
\mathbb{E}(X) = \left(\frac{|V|}{k}\right)^{-1} s_k.
$$
Suppose that we have a graph $G = (V, E)$ and $1 \leq k \leq |V|$. Choose $Y$ uniformly at random from a set of all $k$-cardinality subsets of $\{v_1, \ldots, v_{|V|}\}$. Define

$$X = \mathbb{1}\{Y \text{ is independence set}\}.$$ 

Then,

$$\mathbb{E}(X) = \binom{|V|}{k}^{-1} s_k.$$ 

### Rare-event problem

This algorithm will (unfortunately) fail for small values of $\mathbb{E}(X)$, because of the rare-event problem.
Our solution — Multilevel Splitting (MS)

Our solution — Multilevel Splitting (MS)

- Our method handles several types of counting problems. Let $T$ be the number of splitting levels and $N$ is a level sample size.

Problem Complexity

- Count all independent sets in a graph $O(TN|E|)$
- Count all independent sets of cardinality $k$ $O(kTN|E|)$
- Find graph’s independence number $O(\log_2(|V|TN|V||E|)$
- Find independent sets distribution $O(|V|^2TN|E|)$

Can we say something about the algorithm efficiency?
Our solution — Multilevel Splitting (MS)

- Our method handles several types of counting problems. Let $T$ be the number of splitting levels and $N$ is a level sample size.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count all independent sets in a graph</td>
<td>$\mathcal{O}(TN</td>
</tr>
<tr>
<td>Count all independent sets of cardinality $k$</td>
<td>$\mathcal{O}(kTN</td>
</tr>
<tr>
<td>Find graph’s independence number</td>
<td>$\mathcal{O}(\log_2(</td>
</tr>
<tr>
<td>Find independent sets distribution</td>
<td>$\mathcal{O}(</td>
</tr>
</tbody>
</table>
Our solution — Multilevel Splitting (MS)

- Our method handles several types of counting problems. Let $T$ be the number of splitting levels and $N$ is a level sample size.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count all independent sets in a graph</td>
<td>$\mathcal{O}(TN</td>
</tr>
<tr>
<td>Count all independent sets of cardinality $k$</td>
<td>$\mathcal{O}(kTN</td>
</tr>
<tr>
<td>Find graph’s independence number</td>
<td>$\mathcal{O}(\log_2(</td>
</tr>
<tr>
<td>Find independent sets distribution</td>
<td>$\mathcal{O}(</td>
</tr>
</tbody>
</table>

Can we say something about the algorithm efficiency?
We consider a very general problem of estimating:

\[ \ell = \mathbb{E}(Z), \quad Z \sim Q. \]
We consider a very general problem of estimating:

\[ \ell = \mathbb{E}(Z), \quad Z \sim Q. \]

Let \( Z_1, \ldots, Z_R \) be independent realizations of random variables \( Z \) such that \( \mathbb{E}(Z) = \ell \). Then,

\[ \hat{\ell} = \frac{1}{R} \sum_{i=1}^{R} Z_i \]

is a Monte Carlo estimator of \( \ell \).
Unbiased estimator

We consider a very general problem of estimating:

\[ \ell = \mathbb{E}(Z), \quad Z \sim Q. \]

Let \( Z_1, \ldots, Z_R \) be independent realizations of random variables \( Z \) such that \( \mathbb{E}(Z) = \ell \). Then,

\[ \hat{\ell} = \frac{1}{R} \sum_{i=1}^{R} Z_i \]

is a Monte Carlo estimator of \( \ell \).

We need to answer the following questions

1. Is it hard to develop an unbiased estimator \( \hat{\ell} \) for a problem in hand?
We consider a very general problem of estimating:

\[ \ell = \mathbb{E}(Z), \quad Z \sim Q. \]

Let \( Z_1, \ldots, Z_R \) be independent realizations of random variables \( Z \) such that \( \mathbb{E}(Z) = \ell \). Then,

\[ \hat{\ell} = \frac{1}{R} \sum_{i=1}^{R} Z_i \]

is a Monte Carlo estimator of \( \ell \).

We need to answer the following questions

1. Is it hard to develop an unbiased estimator \( \hat{\ell} \) for a problem in hand? — No, (at least for a bad estimator).
2. Given \( N \), how close the estimator \( \hat{\ell} \) to the real value \( \ell \)?
We consider a very general problem of estimating:

\[ \ell = \mathbb{E}(Z) , \quad Z \sim Q. \]

Let \( Z_1, \ldots, Z_R \) be independent realizations of random variables \( Z \) such that \( \mathbb{E}(Z) = \ell \). Then, 

\[ \hat{\ell} = \frac{1}{R} \sum_{i=1}^{R} Z_i \]

is a Monte Carlo estimator of \( \ell \).

We need to answer the following questions

1. Is it hard to develop an **unbiased** estimator \( \hat{\ell} \) for a problem in hand? — No, (at least for a bad estimator).

2. Given \( N \), how close the estimator \( \hat{\ell} \) to the real value \( \ell \)? — Yes, since it depends on the variance of \( Z \), that is not always available.
I don’t know the variance, did it happen to you?

Gogate and Dechter (2011)
Sampling-based Lower Bounds for Counting Queries

They also do not know the variance...
Let $Z_1, \ldots, Z_R$ be independent realizations of random variables $Z$ such that $\mathbb{E}(Z) = \ell$. Then, for a constant $0 \leq \omega < 1$, the following probabilistic lower bounds exist.

**Minimum scheme bound (MSB):**

$$
\mathbb{P}\left( \min_{1 \leq r \leq R} \left[ \frac{Z_r}{\alpha} \right] \leq \ell \right) \geq \omega,
$$

where $\alpha = \left( \frac{1}{1 - \omega} \right)^{\frac{1}{R}}$.

**Average scheme bound (ASB):**

$$
\mathbb{P}\left( \frac{1}{R} \sum_{r=1}^{R} Z_r \leq \ell \right) \geq \omega,
$$

where $\alpha = \frac{1}{1 - \omega}$.

**Maximum scheme bound (MASB):**

$$
\mathbb{P}\left( \max_{1 \leq r \leq R} \left[ \frac{Z_r}{\alpha} \right] \leq \ell \right),
$$

where $\alpha = \frac{1}{1 - \omega}$.

**Permutation scheme bound (PSB):**

$$
\mathbb{P}\left( \max_{1 \leq r \leq R} \left[ \frac{1}{\alpha} \prod_{j=1}^{r} Z_j \right]^{1/r} \leq \ell \right) \geq \omega,
$$

where $\alpha = 1/(1 - \omega)$.

**Order Statistics bound (OSB):**

$$
\mathbb{P}\left( \max_{1 \leq r \leq R} \left[ \left( \frac{1}{\alpha} \prod_{j=1}^{r} O_{(R-j+1)} \right)^{1/r} \right] \leq \ell \right) \geq \omega,
$$

where $\alpha = 1/(1 - \omega)$, and $(O_{(1)}, \ldots, O_{(R)})$ is an order statistics over the sample set $(Z_1, \ldots, Z_R)$, such that for $1 \leq r_1 < r_2 \leq R$, it holds that $O_{(r_1)} \leq O_{(r_2)}$.,
From Markov’s inequality, for arbitrary $Z_r$ ($1 \leq r \leq R$), we have:

$$
\mathbb{P} \left( \frac{Z_r}{\alpha} \geq \ell \right) \leq \frac{1}{\alpha}.
$$

Rearranging (1), and substituting $\alpha = \left(\frac{1}{1 - \omega}\right)\frac{1}{R}$, we arrive at:

$$
\mathbb{P} \left( \min_{1 \leq r \leq R} \{Z_r\} \leq \ell \right) \geq 1 - \frac{1}{\alpha} = 1 - \frac{1}{\left(\frac{1}{1 - \omega}\right)\frac{1}{R}} = 1 - \frac{1}{1 - \omega} = \omega.
$$

How tight are these bounds?
From Markov’s inequality, for arbitrary $Z_r$ ($1 \leq r \leq R$), we have:

$$
P \left( \frac{Z_r}{\alpha} \geq \ell \right) \leq \frac{1}{\alpha}.
$$

Since, the generated $R$ samples are independent, the probability that the minimum over them is also an upper bound is given by:

$$
P \left( \min_{1 \leq r \leq R} \left\{ \frac{Z_r}{\alpha} \right\} \geq \ell \right) \leq \frac{1}{\alpha^R}.
$$

(1)
From Markov’s inequality, for arbitrary $Z_r$ ($1 \leq r \leq R$), we have:

$$\mathbb{P} \left( \frac{Z_r}{\alpha} \geq \ell \right) \leq \frac{1}{\alpha}.$$ 

Since, the generated $R$ samples are independent, the probability that the minimum over them is also an **upper bound** is given by:

$$\mathbb{P} \left( \min_{1 \leq r \leq R} \left\{ \frac{Z_r}{\alpha} \right\} \geq \ell \right) \leq \frac{1}{\alpha^R}. \quad (1)$$

Rearranging (1), and substituting $\alpha = \left( \frac{1}{1-\omega} \right)^{\frac{1}{R}}$, we arrive at:

$$\mathbb{P} \left( \min_{1 \leq r \leq R} \left\{ \frac{Z_r}{\alpha} \right\} \leq \ell \right) \geq 1 - \frac{1}{\alpha^R} = 1 - \frac{1}{\left( \frac{1}{1-\omega} \right)^{R}} = 1 - \left( \frac{1-\omega}{1-\omega} \right) = \omega.$$
Proof of Minimum scheme bound

1. From Markov’s inequality, for arbitrary $Z_r$ ($1 \leq r \leq R$), we have:

$$\mathbb{P}\left(\frac{Z_r}{\alpha} \geq \ell\right) \leq \frac{1}{\alpha}.$$ 

2. Since, the generated $R$ samples are independent, the probability that the minimum over them is also an upper bound is given by:

$$\mathbb{P}\left(\min_{1 \leq r \leq R} \left\{\frac{Z_r}{\alpha}\right\} \geq \ell\right) \leq \frac{1}{\alpha^R}. \quad (1)$$

3. Rearranging (1), and substituting $\alpha = \left(\frac{1}{1-\omega}\right)^{\frac{1}{R}}$, we arrive at:

$$\mathbb{P}\left(\min_{1 \leq r \leq R} \left\{\frac{Z_r}{\alpha}\right\} \leq \ell\right) \geq 1 - \frac{1}{\alpha^R} = 1 - \frac{1}{\left(\frac{1}{1-\omega}\right)^{\frac{1}{R}}} = 1 - (1 - \omega) = \omega.$$ 

How tight are these bounds?
Numerical evaluation


(a) The $\mathcal{H}_4$ network.  
(b) The $\mathcal{W}_{16}$ network.

Figure: An example of Hypercube ($\mathcal{H}_4$) and Waxman ($\mathcal{W}_{16}$) networks.
Small and medium size Hypercube and Waxman graphs

Figure: 3% Relative Error of MS

Figure: $\omega = 0.95$ Lower bound (LB) with 10 independent runs of MS
So, why do we need these bounds?

**Figure:** Summary of CPU times in seconds that are needed to acquire the full independent set distribution via the MS algorithm using for 3% RE threshold and the LB estimator, which is based on 10 independent runs. For the $\mathcal{H}_8, \mathcal{H}_9, \mathcal{W}_{256}$ and $\mathcal{W}_{512}$ networks.
Figure: Distribution of independent sets in Hypercubes and Waxman graphs using the MS algorithm. The PMLB (for $\omega = 0.95$), was constructed using 10 independent runs of the MS algorithm.
**Solving the max-IS problem**

**Table: Performance of Splitting Algorithm on the Hypercube and Waxman networks using \( N = 1000 \).**

<table>
<thead>
<tr>
<th>( G )</th>
<th>( \alpha(G) )</th>
<th>( \hat{\alpha}(G) )</th>
<th># binary iterations.</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_4 )</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>0.22</td>
</tr>
<tr>
<td>( W_{16} )</td>
<td>–</td>
<td>7</td>
<td>4</td>
<td>0.26</td>
</tr>
<tr>
<td>( H_5 )</td>
<td>16</td>
<td>16</td>
<td>5</td>
<td>1.38</td>
</tr>
<tr>
<td>( W_{32} )</td>
<td>–</td>
<td>13</td>
<td>5</td>
<td>1.61</td>
</tr>
<tr>
<td>( H_6 )</td>
<td>32</td>
<td>32</td>
<td>6</td>
<td>13.5</td>
</tr>
<tr>
<td>( W_{64} )</td>
<td>–</td>
<td>25</td>
<td>6</td>
<td>11.5</td>
</tr>
<tr>
<td>( H_7 )</td>
<td>64</td>
<td>64</td>
<td>7</td>
<td>104</td>
</tr>
<tr>
<td>( W_{128} )</td>
<td>–</td>
<td>47</td>
<td>7</td>
<td>79</td>
</tr>
<tr>
<td>( H_8 )</td>
<td>128</td>
<td>128</td>
<td>8</td>
<td>957</td>
</tr>
<tr>
<td>( W_{256} )</td>
<td>–</td>
<td>92</td>
<td>8</td>
<td>753</td>
</tr>
<tr>
<td>( H_9 )</td>
<td>256</td>
<td>256</td>
<td>9</td>
<td>9199</td>
</tr>
<tr>
<td>( W_{512} )</td>
<td>–</td>
<td>172</td>
<td>9</td>
<td>6998</td>
</tr>
</tbody>
</table>
Table: Performance of MS on the benchmark problems from https://turing.cs.hbg.psu.edu/txn131/clique.html with $N = 1000$.

| $G$         | $|V|$ | $|E|$ | $\alpha(G)$ | $\hat{\alpha}(G)$ | $\#$ bin. | CPU  |
|------------|------|------|-------------|-------------------|----------|------|
| keller4    | 171  | 9435 | 11          | 11                | 7        | 327  |
| gen200p0944 | 200  | 17910 | 44         | 44                | 7        | 396  |
| gen200p0955 | 200  | 17910 | 55          | 55                | 8        | 822  |
| C125.9     | 125  | 6363 | $\geq 34$   | 34                | 7        | 96.1 |
Run the MS algorithm to count independent sets of cardinality $k = 34$ and obtain a counting estimate within some small RE.

During the execution of MS, accumulate unique independent sets of size 34.

Finally, try to extend these accumulated cardinality 34 independent sets with an additional vertex to obtain a 35-IS.
The C125.9 problem and our conjecture

- Run the MS algorithm to count independent sets of cardinality \( k = 34 \) and obtain a counting estimate within some small RE.
- During the execution of MS, accumulate unique independent sets of size 34.
- Finally, try to extend these accumulated cardinality 34 independent sets with an additional vertex to obtain a 35-IS.
What next?

- Theoretical work. Similar to Gogate and Dechter, develop different bounds based on Markov’s inequality.

Our experiments indicate that when applied to MS estimators, these bounds are tight. They can be used for delivering probabilistic confidence bounds for large data problems.

Explore the uniformity of MS.

An interesting practical direction is to develop a MS software with multiple CPUs / GPU.

Thank you!

Questions please.
What next?

- Theoretical work. Similar to Gogate and Dechter, develop different bounds based on Markov’s inequality.

- Our experiments indicate that when applied to MS estimators, these bounds are tight. They can be used for delivering probabilistic confidence bounds for large data problems.
What next?

- Theoretical work. Similar to Gogate and Dechter, develop different bounds based on Markov’s inequality.

- Our experiments indicate that when applied to MS estimators, these bounds are tight. They can be used for delivering probabilistic confidence bounds for large data problems.

- Explore the uniformity of MS.
What next?

- Theoretical work. Similar to Gogate and Dechter, develop different bounds based on Markov’s inequality.

- Our experiments indicate that when applied to MS estimators, these bounds are tight. They can be used for delivering probabilistic confidence bounds for large data problems.

- Explore the uniformity of MS.

- An interesting practical direction is to develop a MS software with multiple CPUs / GPU.
What next?

- Theoretical work. Similar to Gogate and Dechter, develop different bounds based on Markov’s inequality.

- Our experiments indicate that when applied to MS estimators, these bounds are tight. They can be used for delivering probabilistic confidence bounds for large data problems.

- Explore the uniformity of MS.

- An interesting practical direction is to develop a MS software with multiple CPUs / GPU.

Thank you!

Questions please.