Sequential Importance Sampling for Counting Hamilton Cycles on Random Graphs

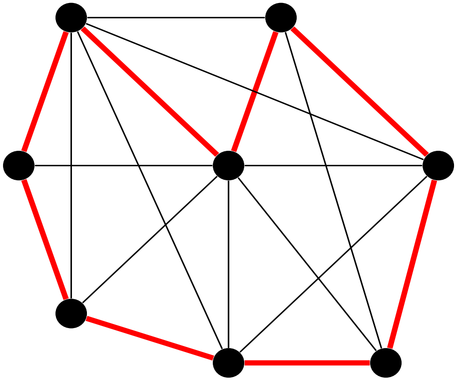
Ad Ridder (*Vrije University, Amsterdam*)
Radislav Vaisman (*University of Queensland, Brisbane*)

RESIM 2016, Eindhoven

Introduction

- Given a directed connected graph $G = (V, \mathcal{A})$ with $|V| = n$ nodes.
- Problem is to compute the number of Hamilton cycles $\mathrm{HC}(G)$.
- This is a #P-complete counting problem (Valiant 1979).
- That means the counting equivalent of the NP-complete complexity of the decision problem.

Recall Hamilton Cycles



Relation to Rare Events

- Let $\mathscr{P}(G)$ the set of all paths in G.
- Let $\mathscr{H}(G)$ the set of all Hamilton cycles in G.
- Note that $\mathscr{H}(G) \subset \mathscr{P}(G)$.
- Consider the uniform probability model on $\mathscr{P}(G)$.
- Then

$$\mathrm{HC}(G) = |\mathscr{H}(G)| = \frac{|\mathscr{H}(G)|}{|\mathscr{P}(G)|} \times |\mathscr{P}(G)| = \mathbb{P}(U \in \mathscr{H}(G)) \times |\mathscr{P}(G)|.$$

- Suffices to compute the probability $\mathbb{P}(U \in \mathscr{H}(G))$ assuming the size of $\mathscr{P}(G)$ is known.

In This Talk

- Direct and undirected graphs.
- Randomized algorithms.
- Random graphs.
- Importance sampling based on myopic rules.
- Importance sampling based on oracle knowledge.
- Complexity issues.

(Randomized) Approximate Counting

- A randomized algorithm produces a random output X_G.
- E.g., a Monte Carlo simulation.
- Unbiased $\mathbb{E}[X_G] = \mathrm{HC}(G)$.
- (ϵ, δ)-approximation if

$$\mathbb{P}\big((1-\epsilon)\mathrm{HC}(G) < X_G < (1+\epsilon)\mathrm{HC}(G)\big) > 1-\delta.$$

- Objective (Karp&Luby 1983): algorithm is FPRAS *fully polynomial randomized approximation scheme*.
- Meaning that (ϵ, δ)-approximation is obtained in a polynomial running time in n, ϵ^{-1}, and $\log \delta^{-1}$.

Independent Samples

- Consider indeed a Monte Carlo algorithm.
- Execute N i.i.d. replications of the algorithm.
- Compute the sample average estimator.
- Apply Chebyshev's inequality.
- Required sample size for (ϵ, δ)-approximation ($\delta = 1/4$) is

$$N = O\left(\frac{\mathbb{E}[X_G^2]}{\epsilon^2(\mathbb{E}[X_G])^2}\right);$$

- Let $n \to \infty$ (size of vertex set).

Observation

A Monte Carlo algorithm would be FPRAS if its relative error $\mathbb{E}[X_G^2]/(\mathbb{E}[X_G])^2$ is bounded by a polynomial function in n.

Algorithm: OSLA

- Randomized algorithm, called one-step-look-ahead.
- Construct a path from node to node.
- Start at node 1.
- Say current path of length t of distinct nodes: $(1, v_2, \ldots, v_t)$.
- Remaining nodes $R_t = V \setminus \{1, v_2, \ldots, v_t\}$.
- Let $N(v_t)$ be the 'neighbours' of node v_t (in the original graph).
- Choose v_{t+1} randomly from $N(v_t) \cap R_t$.
- Continue until either $t = n$ or $N(v_t) \cap R_t = \emptyset$.
- If $t = n$, set $R_t = \{1\}$ for completing a cycle.
- Return $X_G = \prod_t |N(v_t) \cap R_t|$.

OSLA is Unbiased

Recognize OSLA as an importance sampling simulation.

Observation

Conclude unbiasedness: $\mathbb{E}[X_G] = \mathrm{HC}(G)$.

OSLA is not FPRAS

- Consider the graph G with n nodes and $n \times n$ adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

- G has a single Hamilton cycle $\pi = (1, 2, \ldots, n, 1)$.
- In OSLA: $\mathbb{E}[X_G] = 1$; $\mathbb{E}[X_G^2] = (n-1)!$.

Observation

Relative error of OSLA estimator is not polynomially bounded.

Are There FPRAS Algorithms?

- No, not generally, unless RP=NP.
- Special cases:
  - some random digraphs (Frieze et al 1992);
  - dense undirected graphs (Dyer et al 1994);
  - random directed graphs (Rasmussen 1994);
  - random regular graphs (Frieze et al 1997);
  - dense directed graphs (Zhang et al 2011);

Randomization

- Consider counting Hamilton cycles in *random* graphs.
- $G(n, p)$ model introduced by Erdos & Renyi (1959).
  - vertex set V of fixed size n;
  - arc set $\mathcal{A}$;
  - each of $n(n-1)$ possible arcs is included with probability p independently.
- Denote $\mathcal{G}_n$ the set of all directed graphs of n vertices.
- Probability measure $\mathbb{P}_{(\mathcal{G}_n, p)}$ on $\mathcal{G}_n$.

First Analysis

- In this randomization model the expected value of the OSLA estimator $\mathbb{E}[X_G]$ becomes a (random) conditional expectation.
- Notation: $\mathbb{E}[X|\mathcal{G}_n]$.
- Calculus to show

$$\mathbb{E}_{(\mathcal{G}_n,p)}\big[\mathbb{E}[X|\mathcal{G}_n]\big] = \mathbb{E}_{(\mathcal{G}_n,p)}[X] = p\prod_{r=1}^{n-1}(rp) = p^n(n-1)!$$

$$\mathbb{E}_{(\mathcal{G}_n,p)}\big[\mathbb{E}[X^2|\mathcal{G}_n]\big] = \mathbb{E}_{(\mathcal{G}_n,p)}[X^2] = p^n(n-1)!\prod_{r=1}^{n-1}\big(1+(r-1)p\big).$$

Corollary

OSLA is logarithmically efficient for random graphs; i.e.,

$$\liminf_{n\to\infty}\frac{\log \mathbb{E}_{(\mathcal{G}_n,p)}[X^2]}{2\log \mathbb{E}_{(\mathcal{G}_n,p)}[X]} \geq 1.$$

What About FPRAS?

Denote the (random) relative error

$$R_n = \frac{\mathbb{E}[X^2|\mathcal{G}_n]}{(\mathbb{E}[X|\mathcal{G}_n])^2}.$$

Definition

FPRAS with high probability (whp) for random graphs means

$$\lim_{n\to\infty} \mathbb{P}_{(\mathcal{G}_n,p)}(R_n \text{ is bounded by polynomial in } n) = 1.$$

A Weaker Condition

Definition

We say that an algorithm is a subexponential randomized approximation scheme (SRAS) whp if for any $\eta > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{(\mathcal{G}_n,p)}\left(\frac{1}{n}\log R_n > \eta\right) = 0.$$

Eguivalently $\frac{1}{n}\log R_n \xrightarrow{\mathbb{P}} 0$.

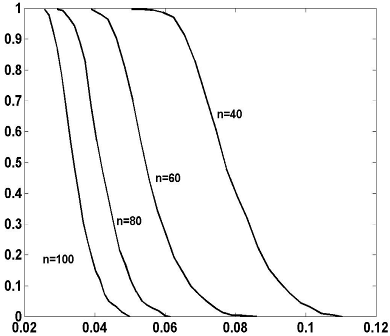
Theorem

OSLA satisfies SRAS whp for random graphs.

Proof

- Denote
$$\kappa_n = \frac{\mathbb{E}_{(\mathcal{G}_n,p)}[X^2]}{\left(\mathbb{E}_{(\mathcal{G}_n,p)}[X]\right)^2}.$$
Then $\lim_{n\to\infty} \frac{1}{n} \log \kappa_n = 0$.
- Markov inequality:
$$\mathbb{P}_{(\mathcal{G}_n,p)}\left(\frac{1}{n} \log R_n > \eta\right) \leq \frac{\mathbb{E}_{(\mathcal{G}_n,p)}\left[\frac{1}{n} \log R_n\right]}{\eta}.$$
- Jensen's inequality and Delta method:
$$\mathbb{E}_{(\mathcal{G}_n,p)}\left[\frac{1}{n} \log R_n\right] \leq \frac{1}{n} \log \mathbb{E}_{(\mathcal{G}_n,p)}[R_n] \leq \frac{1}{n} \log \kappa_n.$$

Illustration



This is a plot of

$$\mathbb{P}\left(\frac{1}{n}\log R_n > x\right)$$

for $n = 40, 60, \ldots, 100$ and $0 < x < 0.12$. For each n, 500 random graphs from the DRG model. Each instance was simulated 200 times by OSLA. The graphs in the figure are the empirical cdf's of the 500 estimates of $(1/n)\log R_n$.

What about Undirected Graphs?

- Adapt the randomization model and OSLA algorithm (straigtforwardly).
- We now get

$$\mathbb{E}_{(\mathcal{G}_n,p)}\left[\mathbb{E}[X|\mathcal{G}_n]\right] = \mathbb{E}_{(\mathcal{G}_n,p)}[X] \sim p^n(n-1)!$$

$$\mathbb{E}_{(\mathcal{G}_n,p)}\left[\mathbb{E}[X^2|\mathcal{G}_n]\right] = \mathbb{E}_{(\mathcal{G}_n,p)}[X^2] \sim p^n(n-1)!\prod_{r=1}^{n-1}\big(1+(r-1)p\big),$$

for $n \to \infty$.

Corollary

OSLA is logarithmically efficient for random undirected graphs.

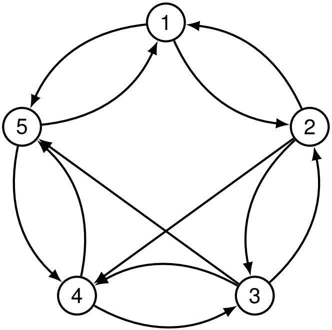
Theorem

OSLA satisfies SRAS whp for random undirected graphs.

Using an Oracle

- Randomized algorithm, also called n-step-look-ahead.
- Consider the directed graph case.
- Construct a path from node to node.
- Start at node 1.
- Say current path of length t of distinct nodes: $(1, v_2, \ldots, v_t)$.
- Let R_t be the remaining nodes and $N(v_t)$ are the neighbours of v_t.
- Ask the oracle for each node $w \in N(v_t) \cap R_t$ whether the path $(1, v_2, \ldots, v_t, w)$ can be completed to an Hamilton cycle.
- Denote W_t for the 'yes' nodes.
- Choose randomly one of the 'yes' nodes.
- Continue until a completed cycle has been constructed.
- Return $Y = \prod_t |W_t|$.

Illustration



$\mathscr{H}(G) = \{\pi_1, \pi_2, \pi_3\}$ with

$$\pi_1 = (1,2,3,4,5,1);\; X = 8;\; Y = 4$$
$$\pi_2 = (1,2,4,3,5,1);\; X = 8;\; Y = 4$$
$$\pi_3 = (1,5,4,3,2,1);\; X = 2;\; Y = 2$$

Gives

$$\mathbb{E}[X] = \mathbb{E}[Y] = \mathrm{HC}(G) = 3,$$

and

$$\mathbb{V}ar[X] = 9;\; \mathbb{V}ar[Y] = 1.$$

Analysis

Definition

A Hamilton path from node s to node t is any path from s to t that visits all nodes in the graph once. Denote by $\mu_r(p)$ the probability that there is at least one Hamilton path in a random graph with r nodes in the random graph model.

We now get

$$\mathbb{E}_{(\mathcal{G}_n,p)}\big[\mathbb{E}[Y|\mathcal{G}_n]\big] = \mathbb{E}_{(\mathcal{G}_n,p)}[Y] = p^n(n-1)!$$

$$\mathbb{E}_{(\mathcal{G}_n,p)}\big[\mathbb{E}[Y^2|\mathcal{G}_n]\big] = \mathbb{E}_{(\mathcal{G}_n,p)}[Y^2] \sim p^n(n-1)!\prod_{r=1}^{n-1}\big(1+(r-1)p\mu_{r+1}(p)\big),$$

for $n \to \infty$.

Complexities

Corollary

nSLA is logarithmically efficient for random graphs.

Theorem

nSLA satisfies SRAS whp for random graphs.

References

- M. Dyer, A. Frieze, M. Jerrum. ”Approximately counting Hamilton cycles in dense graphs”. In*Proceedings Symposium on Discrete Algorithms* 1994.
- P. Erdős and A. Rényi. ”On random graphs I”. *Publicationes Mathematicae Debrecen* 1959.
- A. Frieze and S. Suen. ”counting Hamilton cycles in random directed graphs”. *Random Structures and Algorithms* 1992.
- A. Frieze, M. Jerrum, M. Molloy, R. Robinson, N. Wormald. ”Generating and counting Hamilton cycles in random regular graphs”. *Journal of Algorithms* 1996.
- R.M. Karp and M. Luby. ”Monte-Carlo algorithms for enumeration and reliability problems”. In *Proceedings of the 24th Annual Symposium on Foundations of Computer Science* 1983.
- L.E. Rasmussen. ”Approximating the permanent: a simple approach”. *Random Structures and Algorithms*, Vol. 5, 1994, pp. 349-361.
- Rubinstein R.Y, ”Stochastic Enumeration Method for Counting NP-hard Problems”, *Methodology and Computing in Applied Probability*, 2011.
- R.Y. Rubinstein, A. Ridder and R. Vaisman. *Fast Sequential Monte Carlo Methods for Counting and Optimization*, Wiley, 2014.
- L.G. Valiant. ”The Complexity of Enumeration and Reliability Problems”. *SIAM Journal on Computing* 1979.
- J. Zhang and F. Bai. ”An improved FPRAS for counting the number of Hamiltonian cycles in dense digraphs”. *Theoretical Computer Science* 2011.