Counting without Sampling: Asymptotics of the log-Partition Function

Antar Bandyopadhyay
(Joint work with David Gamarnik)

Probability and Stochastic Processes Seminar
Technion, Haifa, Israel

Theoretical Statistics and Mathematics Unit
Indian Statistical Institute, New Delhi Centre

`http://www.isid.ac.in/∼antar`

June 19, 2007

Two Counting Problems

Definition 1 (Independent Set) *Suppose* $G := (V, E)$ *be a finite graph. We will say a subset* $I \subseteq V$ *is an* independent set *of* G, *if for any two vertices* $u, v \in I$ *there is no edge between* u *and* v.

We will denote by $\mathcal{I}_G$, *the set of all independent sets of* G.

Problem 1: Given a finite graph G, count the *number of independent sets* of G.

Definition 2 (Proper q-Coloring) *Fix* $q \geq 2$ *an integer, and suppose* $G := (V, E)$ *be a finite graph. A map* $C : V \to \{1, 2, \ldots, q\}$ *is called a* proper q-coloring *of* G, *if no two vertices of same color share an edge.*

We will denote by $\mathcal{C}_G(q)$, *the set of all proper* q*-colorings of* G.

Problem 2: Given a finite graph G, and $q \geq 2$ an integer, count the *number of proper* q*-colorings* of G.

Exact/Approximate Counting

Q: Can we do exact counting ?

A:
- Perhaps not !
- The sets are typically exponentially large.
- No polynomial time algorithm [Valiant 1979].

Q: So what do we do ?

A: We can try “approximate” counting.

Q:
- How do we approximate ?
- What kind of approximation ?

A:
- Typical approach is to use a *Markov chain Monte Carlo* (MCMC) sampling scheme.
- One need to prove *rapid mixing* for the chain.

Some Success Stories for Problems Similar to Ours (using MCMC techniques)

- Computing the permanent:
  - ▶ Jerrum and Sinclair (1989, 1997).
  - ▶ Jerrum, Sinclair and Vigoda (2004).
- Computing the volume of a convex body:
  - ▶ Dyer, Frieze and Kannan (1991).
  - ▶ Kannan, Lovasz and Simonovits (1997).
  - ▶ Lovasz and Vempala (2003).
- Counting independent set:
  - ▶ Luby and Vigoda (1997).

Remark: Such MCMC techniques typically provide a randomized ε-approximation to the counting problem, such that the running time is a polynomial in the size of the problem (e.g. the size of V), and also in the error ε.

What Do We Propose to Do ?

- We will give *deterministic* approximation schemes, which will not use sampling.

- But we will provide ε-approximation to $\log|\mathcal{I}_G|$ and $\log|\mathcal{C}_G(q)|$. (Unfortunately, this is *obviously* less efficient !)

- Moreover, we will need restrictions on our graphs ! For example, we will need *low degree* graphs, and a "*large girth*" assumption (will be more specific later).

Get Lost !!!!

We are *obviously* doing *less* than what is known !

Then why am I giving this talk ?

- Well well ... I like this work ! ☺
- But there are more reasons than just that ! ☺

Motivation and Achievements

- Our motivation comes from *statistical physics*.
- Computation of $\log|\mathcal{I}_G|$ or $\log|\mathcal{C}_G(q)|$ are interesting, because they correspond to the *free energy* for certain models in statistical physics (the models will be given later).
- We can achieve (new) explicit results for regular graphs, which are not possible to derive using the MCMC methods. To give some example:
  - ▶ We can show that for every 4-regular graph of n vertices and large girth, the number of independent sets is approximately $(1.494\ldots)^n$.
  - ▶ We can also show that if $q \geq r+1$ then for every r-regular graph with large girth, the number of proper q-coloring is approximately

$$\left[q\left(1-\frac{1}{q}\right)^{\frac{r}{2}}\right]^n.$$

- We can drop the "large girth" assumption and work with random regular graphs to get concentration results.

Two Statistical Physics Models

(1) Hard-Core Model: Given a finite graph G and a real number $\lambda > 0$, consider a (discrete) probability distribution on $\mathcal{I}_G$ given by

$$\mathbb{P}(I) \propto \lambda^{|I|} \iff \mathbb{P}(I) = \frac{\lambda^{|I|}}{Z(\lambda, G)}, \quad I \in \mathcal{I}_G\,,$$

where

$$Z(\lambda, G) := \sum_{I \in \mathcal{I}_G} \lambda^{|I|}\,.$$

Remarks:

- $\mathbb{P}$ is called the *Gibbs distribution* on $\mathcal{I}_G$.
- $Z(\lambda, G)$ is called the *partition function*.
- λ is called the *activity parameter*.
- Observe $Z(\lambda, G) = |\mathcal{I}_G|$ when $\lambda = 1$, then we are back to the original counting problem.

(2) Model on Proper q-Colorings: Given $q \geq 2$ an integer, and a finite graph G, let $\lambda_k > 0$, for $1 \leq k \leq q$. Consider a (discrete) probability distribution on $\mathcal{C}_G(q)$ given by

$$\mathbb{P}(C) = \frac{\prod_{1 \leq k \leq q} \lambda_k^{|C^{-1}(\{k\})|}}{Z(\lambda, q, G)}, \quad C \in \mathcal{C}_G(q) \quad \text{where} \quad Z(\lambda, q, G) := \sum_{C \in \mathcal{C}_G(q)} \prod_{1 \leq k \leq q} \lambda_k^{|C^{-1}(\{k\})|}.$$

Remarks:

- $\mathbb{P}$ is called the *Gibbs distribution* on $\mathcal{C}_G(q)$.
- $Z(\lambda, q, G)$ is the *partition function* and λ_k's are called the *activity parameters*.
- If all the λ_k's are equal then $Z(\lambda, q, G) = |\mathcal{C}_G(q)|$ and we are back to the original counting problem. For this case we will denote the partition function by $Z(q, G)$.

Some Families of Graphs

- **Large girth:** An infinite family of graphs $\mathcal{G}$ is defined to have *large girth*, if there exists an increasing function $f : \mathbb{N} \to \mathbb{N}$ with $\lim_{s\to\infty} f(s) = \infty$, such that for every $G \in \mathcal{G}$ with n vertices, we have

$$\text{girth}\,(G) \geq f(n)\,.$$

  Recall: girth (G) := size of the smallest cycle in G.

- **[Low degree]:** Let $\mathcal{G}\,(n, r, g)$ be the family of graphs on n vertices, such that the maximum degree of any vertex is bounded by r and each graph has girth at least g.

- **Regular:** Let $\mathcal{G}_{\texttt{reg}}\,(n, r, g)$ be the family of r-regular graphs on n vertices, such that each graph has girth at least g.

The Main Results

Algorithm Result:

Theorem 1 (Independent Sets) *For every family of graphs $\mathcal{G}$ with maximum degree at most 4 and large girth, there is an algorithm $\mathcal{A}$, such that for any $\varepsilon > 0$ and $G \in \mathcal{G}$, $\mathcal{A}$ produces a quantity $\mathfrak{Z}$ in time polynomial in $n := |V|$, such that*

$$(1-\varepsilon)\,\frac{\log|\mathcal{I}_G|}{n} \leq \mathfrak{Z} \leq (1+\varepsilon)\,\frac{\log|\mathcal{I}_G|}{n}\,.$$

Theorem 2 (Colorings) *Fix $q \geq r+1$ be two integers then*

$$\lim_{g\to\infty}\sup_{G\in\mathcal{G}(n,r,g)}\left|\frac{\log|\mathcal{C}_G(q)|}{n} - \frac{1}{n}\sum_{1\leq k\leq n}\log\left[q\left(1-\frac{1}{q}\right)^{r_{G_{k-1}}(v_k)}\right]\right| = 0.$$

where $V := \{v_1, v_2, \ldots, v_n\}$ and $G_k := G \setminus \{v_1, v_2, \ldots, v_k\}$, and by $r_G(v)$ we mean the degree of vertex v in graph G.

In particular, we can get an algorithm result for counting the number of proper q-colorings, which is similar to the previous theorem.

Results for the Regular Graphs with Large Girth:

Theorem 3 (Independent Sets) *Suppose* $\lambda < \lambda_c(r)$ *where* $\lambda_c(r) := (r-1)^{r-1}/(r-2)^r$, *then*

$$\lim_{g\to\infty} \sup_{G\in\mathcal{G}_{\mathrm{reg}}(n,r,g)} \left| \frac{\log Z(\lambda, G)}{n} - \log\left(x^{-\frac{r}{2}} (2-x)^{-\frac{r-2}{2}} \right) \right| = 0\,,$$

where x *is the unique positive solution of*

$$x = 1/(1+\lambda x^{r-1})\,.$$

In particular, if $r = 2, 3, 4, 5$ *and* $\lambda = 1$, *then the corresponding limits for* $\frac{\log|\mathcal{I}_G|}{n}$ *are respectively,* $\log 1.618\ldots$, $\log 1.545\ldots$, $\log 1.494\ldots$ *and* $\log 1.453\ldots$.

Theorem 4 (Colorings) *For every* $q \geq r+1$, *the number of* q*-colorings of graphs* $G \in \mathcal{G}_{\mathrm{reg}}(n,r,g)$ *satisfies*

$$\lim_{g\to\infty} \sup_{G\in\mathcal{G}_{\mathrm{reg}}(n,r,g)} \left| \frac{\log Z(q, G)}{n} - \log\left[q\left(1-\frac{1}{q}\right)^{\frac{r}{2}}\right] \right| = 0\,.$$

Results for the Random Regular Graphs:

Theorem 5 (Independent Sets) *For every $r \geq 2$ and every $\lambda < \lambda_c(r)$, the (random) partition function $Z(\lambda, G_r(n))$, of a random r-regular graph $G_r(n)$ satisfies*

$$\frac{\log Z(\lambda, G_r(n))}{n} \to \log\left[x^{-\frac{r}{2}}(2-x)^{-\frac{r-2}{2}}\right],$$

with high probability (w.h.p.), as $n \to \infty$, where x is the unique positive solution of $x = 1/(1 + \lambda x^{r-1})$.

Theorem 6 (Colorings) *For every $r \geq 2$ and $q \geq r+1$, the (random) partition function $Z(q, G_r(n))$ of a random r-regular graph $G_r(n)$ corresponding to the uniform distribution on proper q-colorings satisfies*

$$\frac{\log Z(q, G_r(n))}{n} \to \log\left[q\left(1 - \frac{1}{q}\right)^{\frac{r}{2}}\right].$$

w.h.p. as $n \to \infty$.

Remark: Theorem 6 was proved earlier by Achlioptas and Moore (2004) using second moment method.

Two Main Steps of the Algorithm (Illustrated only for the Independent Sets)

STEP 1 (The Cavity Equation :

- In this step we relate the computation of the partition function to the computation of the marginal probabilities.
- This is done by creating a *cavity* in the original graph.
- Let G_1 be the original graph G with one vertex, say v_1, removed.
- By definition

$$Z(\lambda, G_1) = \sum_{I \in \mathcal{I}_{G_1}} \lambda^{|I|} = \sum_{\substack{I \in \mathcal{I}_G \\ v_1 \notin I}} \lambda^{|I|} .$$

- *Cavity Equation:*

$$\frac{Z(\lambda, G_1)}{Z(\lambda, G)} = \mathbb{P}_G(v_1 \notin \mathbf{I}) .$$

Cavity Equation Continued ...

Proposition 7 *Let* $V := \{v_1, v_2, \ldots, v_n\}$*, and for* $1 \leq k \leq (n-1)$ *we define* $G_k := G \backslash \{v_1, v_2, \ldots, v_k\}$ *as the graph obtained from* G *after creating* k cavities*. Put* $G_0 = G$*. Then the following relation holds*

$$\frac{Z(\lambda, G_1)}{Z(\lambda, G_0)} = \mathbb{P}_{G_0}(v_1 \notin \mathbf{I}) \ ,$$

where $\mathbf{I}$ *is a random independent set distributed according the Gibbs measure* $\mathbb{P}$*. As a result we get*

$$Z(\lambda, G) = \prod_{k=1}^{n} \left(\mathbb{P}_{G_{k-1}}(v_k \notin \mathbf{I})\right)^{-1} \ .$$

Remark: This proposition is well known in Physics literature and also in the Markov chain based approximation algorithms for counting.

STEP 2 (Computation on Trees):

- Note our *large girth* assumption makes our graphs "*locally*" tree like !
- So in this step we only make computation for the marginal probabilities when the graph is a finite tree.
- This can be done easily by a recursive method, essentially the same cavity trick works.

Computation on Trees Continued ...

- T be a finite tree with root v_0, and let $\{v_1, v_2, \ldots, v_k\}$ be the children of v_0.
- By the cavity equation we get:

$$\begin{aligned}
\mathbb{P}_T\left(v_0 \notin \mathbf{I}\right) &= \frac{Z\left(\lambda, T \setminus \{v_0\}\right)}{Z\left(\lambda, T\right)} \\
&= \frac{1}{1 + \frac{\sum\limits_{I \in \mathcal{I}_T, v_0 \in I} \lambda^{|I|}}{Z(\lambda, T\setminus\{v_0\})}} \\
&= \frac{1}{1 + \lambda \frac{\sum\limits_{I \in \mathcal{I}_{T\setminus\{v_0\}},\, v_j \notin I \ \forall\ 1 \le j \le k} \lambda^{|I|}}{Z(\lambda, T\setminus\{v_0\})}} \\
&= \frac{1}{1 + \lambda \prod\limits_{1 \le j \le k} \mathbb{P}_{T(v_j)}\left(v_j \notin \mathbf{I}\right)}
\end{aligned}$$

where $T(v_j)$ is the tree rooted at the child v_j.

Computation on Trees Continued ...

Proposition 8 *Suppose T be a finite rooted tree with root v_0, and let $\{v_1, v_2, \ldots, v_k\}$ be $k \geq 0$ children of v_0. For each $1 \leq j \leq k$, let $T(v_j)$ denote the tree rooted at v_j consisting only the descendants of v_j (if any). Then the following recursion holds*

$$\mathbb{P}_T\left(v_0 \notin \mathbf{I}\right) = \frac{1}{1 + \lambda \prod_{1 \leq j \leq k} \mathbb{P}_{T(v_j)}\left(v_j \notin \mathbf{I}\right)}.$$

The Algorithm

INPUT: A graph G with vertex set $\{v_1, v_2, \ldots, v_n\}$, and a number $\varepsilon > 0$.

BEGIN

1. Compute the girth $g = g(G)$.

2. If $(0.9)^{\frac{g}{2}-2} \geq \varepsilon$ then find $|\mathcal{I}_G|$ by enumeration and STOP.

If not then

3. Set $Z \leftarrow 1$, $t \leftarrow \lfloor g/2 \rfloor$ and $k \leftarrow 1$.

4. Find the t-depth neighborhood $T(v_k)$ of v_k.

5. Compute the marginal probability $p = \mathbb{P}_{T(v_k)}(v_k \notin \mathbf{I})$ for the finite tree $T(v_k)$.

6. Set $Z \leftarrow Z/p$, $G \leftarrow G \setminus \{v_k\}$, $k \leftarrow k + 1$.

7. If $k \leq n$ then goto Step 4, otherwise STOP.

END

OUTPUT: Z

Why Does It Works ?

⋆ **Strong Correlation Decay:**

- We prove that under certain assumptions, for example,
  - ▶ $\lambda < \lambda_c(r) = (r-1)^{r-1}/(r-2)^r$ (for the r-regular case),
  - ▶ or $r \leq 4$ (for deriving the algorithm),

  the *influence* on the root of the boundary at a distance d decreases exponentially fast as d increases.

- A statistical physics consequence of this is the Gibbs measure on the "limiting infinite graph" (if any !) is unique, that is there is no *phase transition*.

- For the infinite r-regular trees it was shown by Kelly (1985), that there is no *phase transition* for the hard-core model if and only if $\lambda \leq \lambda_c(r)$.

- For counting independent sets we extend this result to the class of finite trees with maximum degree at most 4, which is the most crucial result for our algorithm to succeed.

Strong Correlation Decay Continued ...

- Suppose T be a finite tree with *large* depth.

- If the maximum degree of T is at most 4, then for any two boundary conditions b_1 and b_2 we show that

$$\mathbb{P}_T^{\lambda=1}\left(v_0 \notin \mathbf{I} \,\middle|\, b_1\right) \approx \mathbb{P}_T^{\lambda=1}\left(v_0 \notin \mathbf{I} \,\middle|\, b_2\right) .$$

- Moreover the error in approximation is exponentially small in the depth of the tree.

Strong Correlation Decay Continued ...

- Further, if T is a tree such that every vertex has degree r except the root, which has degree $(r-1)$ and the vertices at the last generation, which have degree 1, then for $\lambda < \lambda_c(r)$ it is know (Kelly, 1985) that

$$\mathbb{P}_T\left(v_0 \notin \mathbf{I} \,\middle|\, b\right) \approx x\,,$$

  for any boundary condition b, where x is the unique solution of the *deterministic* fixed point equation

$$x = 1/(1+\lambda x^{r-1})\,.$$

- If T is a tree with all internal vertices having degree r then under the same assumption

$$\mathbb{P}_T\left(v_0 \notin \mathbf{I} \,\middle|\, b\right) \approx \frac{1}{2-x}\,.$$

Lemma 9 *The following bounds holds for every rooted tree T with depth $t \geq 2$ and degree of any vertex at most 4*

$$\frac{1}{2} \leq \mathbb{P}_T^{\lambda=1}\left(v_0 \notin \mathbf{I} \,\middle|\, b\right) \leq \frac{8}{9},$$

and

$$\left|\mathbb{P}_T^{\lambda=1}\left(v_0 \notin \mathbf{I} \,\middle|\, b_1\right) - \mathbb{P}_T^{\lambda=1}\left(v_0 \notin \mathbf{I} \,\middle|\, b_2\right)\right| \leq (.9)^{t-2},$$

where b, b_1, b_2 are boundary conditions.

Moreover, when $\lambda < \lambda_c(r)\,(r-1)^{r-1}/(r-2)^r$, let x be the unique non-negative solution of the fixed point equation $x = 1/(1+\lambda x^{r-1})$. Suppose all the nodes of T except for leaves and the root have degree r, and suppose the root has degree $r-1$. Then for all boundary conditions b

$$\left|\mathbb{P}_T\left(v_0 \notin \mathbf{I} \,\middle|\, b\right) - x\right| \leq \alpha^t,$$

for some constant $\alpha = \alpha(\lambda) < 1$. If on the other hand, all the nodes except for leaves, have degree r (including the root), then

$$\left|\mathbb{P}_T\left(v_0 \notin \mathbf{I} \,\middle|\, b\right) - \frac{1}{2-x}\right| \leq \alpha^t,$$

with the same constant $\alpha < 1$.

Strong Correlation Decay Continued ...

Remarks:

- The proof involves only elementary math ! But at some point we had to take help of *computer* (MATLAB) [not me] !!

- The correlation decay for the counting of proper q-colorings was proved by Jonasson (2002) for finite depth r-regular tree, but his result extends to any finite tree with bounded degree, which we use for counting proper-q coloring problem.

⋆ From the Tree to the Original Graph:

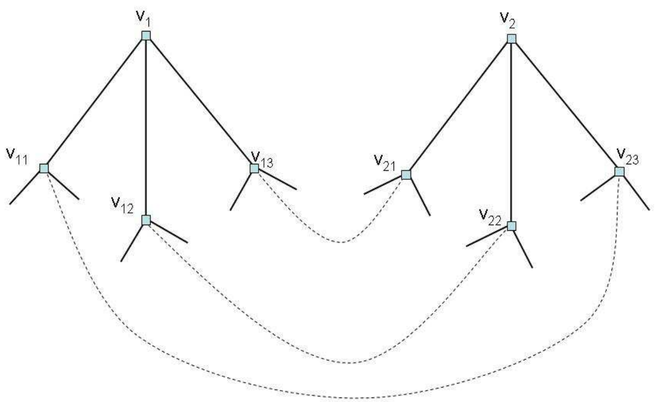
- In this step we show that the error we make by taking a *local tree* around a vertex is *small*.

  Note: The *local tree* comes from the *large girth* assumption.

- This is again done by using the strong correlation decay property and the (spacial) Markovian nature of the Gibbs distribution.

Special Tricks for the Regular Graphs

- For regular graphs creating a *cavity* destroy the regularity !
- Instead we do the following which we call the *rewiring*. Similar idea has been used in Physics literature [Mezard and Parisi, 2005].



Note: v_1 and v_2 are not neighbors and their neighbors are not neighbors !

New "Cavity" Equation for Regular Graphs

Proposition 10 *Given an r-regular graph G, and $\lambda > 0$, the graph G^o obtained from G by rewiring on nodes $v_1, v_2 \in G$, the following relation holds*

$$\frac{Z(\lambda, G^o)}{Z(\lambda, G)} = \mathbb{P}_G(v_1, v_2 \notin \mathbf{I})\mathbb{P}_{G\setminus\{v_1,v_2\}}\left(\bigcap_{1\le j\le r} [v_{1j} \notin \mathbf{I} \text{ or } v_{2j} \notin \mathbf{I}]\right)$$

where $v_{ij}, j = 1, \ldots, r$ are the neighbors of $v_i, i = 1, 2$ in G.

Strong Correlation Decay Result for Regular Graphs

Lemma 11 *Given* $r \geq 3, \lambda < (r-1)^{r-1}/(r-2)^r$ *and* $\epsilon > 0$*, there exists a sufficiently large constant* $g = g(r, \epsilon, \lambda)$ *such that for every* r*-regular graph* G *with girth* $g(G) \geq g$*, and for every pair of nodes* $v_1, v_2 \in G$ *at distance at least* $2g+1$

$$\left| \mathbb{P}_G(v_1, v_2 \notin \mathbf{I}) - \frac{1}{(2-x)^2} \right| < \epsilon,$$

and

$$\left| \mathbb{P}_{G \setminus \{v_1, v_2\}} \left(\bigcap_{1 \leq j \leq r} [v_{1j} \notin \mathbf{I} \text{ or } v_{2j} \notin \mathbf{I}] \right) - (2x - x^2)^r \right| < \epsilon,$$

where $v_{ij}, j = 1, \ldots, r$ *is the set of neighbors of* v_i *in* G*,* $i = 1, 2$*, and* x *is the unique positive solution of* $x = 1/(1 + \lambda x^{r-1})$*.*

A Technical Result needed for Regular Graphs

Lemma 12 *Given an n-node r-regular graph G, consider any integer $4 \leq g \leq g(G)$. The rewiring operation can be performed for at least $(n/2)-(2g+1)r^{2g}$ steps on pairs of nodes which are at least $2g+1$ distance apart. In every step the resulting graph is r-regular with girth at least g.*

Some Final Remarks

- A recent work of Weitz (2006) provides a *fully polynomial approximation scheme* for any finite graph with low degree (maximum degree at most 5) for the problem of counting the independent sets, but it does not give explicit limit results such as ours for the regular graphs.

- Gamarnik and Katz (2006) (personal communication) have extended the work of Weitz (2006) for other counting problems, e.g. counting colorings, and counting matchings on general finite graphs.

- It seems to me that each of this is a "success story" for making a rigorous argument for a very powerful method of statistical physics, called the **cavity method** ! But the full math picture is yet to be discovered.

Thank You