

# Exact Random Sampling From Independent Sets

Mark Huber\*

School of Operations Research and Industrial Engineering

Cornell University

Ithaca, NY 14850

mhuber@orie.cornell.edu

October 1, 1998

## Abstract

We consider the problem of randomly sampling from the set of independent sets of a graph, where the probability of selecting a particular independent set is proportional to  $\lambda$  raised to the size of the independent set. This distribution is the hard core gas model in statistical physics. If  $\Delta$  is the maximum degree of the graph, Dyer and Greenhill gave a method for efficiently approximately sampling from this distribution when  $\lambda \leq 2/(\Delta - 2)$  by utilizing a rapidly mixing Markov chain. We show here how to use their chain to exactly sample from this distribution. This method is shown to run slightly faster than the method of Dyer and Greenhill, and as a corollary gives a better bound on the mixing time of their chain.

## 1 Introduction

An independent set of a graph is a subset of nodes such that no two nodes in the set are connected by an edge. In physics, this corresponds to the hard

---

\*This work was partially supported by an Office of Naval Research Fellowship and NSF grants CCR-9307391, CCR-9700029, and DMS-9505155, and ONR grant N00014-96-1-00500

core gas model, where molecules occupy sites (nodes of the graph) and the molecules are large enough so that no two may occupy adjacent sites at the same time.

The probability of choosing a particular independent set  $\mathcal{I}$  is  $\pi(\mathcal{I}) = \lambda^{|\mathcal{I}|} / Z_\lambda$ , where  $Z_\lambda$  is a normalization constant known as the partition function. In the hard core gas model, higher values of  $\lambda$  correspond to higher values of pressure on the gas. In optimization, a higher value of  $\lambda$  makes it more likely that the sample will be close to the maximum size independent set.

If  $\lambda$  is greater than  $2^n$ , with probability at least  $1/2$  this sample will be a maximum size independent set. Finding the maximum size independent set is an  $NP$  hard problem, and so above a certain value of  $\lambda$  it is impossible to sample efficiently from this distribution unless  $RP = NP$ . In fact, even if  $\lambda = 1$ , so that we are sampling uniformly from the set of independent sets, this problem is  $NP$  hard when the maximum degree of the graph  $\Delta$  is 25 [1].

One method for generating a random sample is to construct a Markov chain with this distribution as the stationary distribution, and then run the chain “for a long time”. A discrete time Markov chain is a stochastic process  $\{X_t\}$  with the property that  $P(X_{t+1} = j | X_t = i, X_{t-1}, \dots) = P(X_{t+1} = j | X_t = i) = p_{ij}$ . Since there are finite number of states in the chains we consider, we may form the  $p_{ij}$  into a transition matrix  $P$ . If the chain is ergodic, that is, connected and aperiodic, the chain will have a unique stationary distribution which satisfies  $\pi P = \pi$ . If the chain is run for a number of steps equal to the mixing time of the chain, then no matter where the chain was started, the final state will be roughly distributed according to the stationary distribution. The closeness of the sample distribution to the stationary distribution is usually measured by some metric such as the total variation distance.

The chain which is currently the fastest known is a chain of Dyer and Greenhill [2]. They use a technique called path coupling to analyze their chain and show that it is rapidly mixing when  $\lambda \leq 2/(\Delta - 2)$ .

A previous chain of Luby and Vigoda [5] had polynomial mixing time when  $\lambda \leq 1/(\Delta - 3)$ , which is more restrictive with respect to  $\lambda$  than the chain of Dyer and Greenhill. However, Luby and Vigoda’s chain has the advantage of being a complete coupling chain. This allowed the chain to be used in the coupling from the past (CFTP) protocol of Propp and Wilson [6]. CFTP has two important advantages over straightforward running of the Markov chain. First, CFTP generates samples which are exactly distributed according to the stationary distribution. There is no need for a metric here, because the

sample distribution and the stationary distribution are one and the same.

Secondly, CFTP contains a termination criterion which tells the user when to stop running the algorithm. When running the chain forward, if the chain has run for longer than the mixing time, there is no indication. With CFTP, once the chain has run long enough to mix, the termination criterion informs the user, and allows the algorithm to end. Hence even though our CFTP algorithm is only proven to run in polynomial time over the same range of  $\lambda$  as the Dyer and Greenhill chain, it may be effective for  $\lambda$  outside this range as well.

Another method for generating exact samples from this chain, worst case neighbor analysis, runs in linear expected time when  $\lambda \leq 0.8/(\Delta - 1)$ . Therefore it has the most restrictions on  $\lambda$ , but it does run faster than the Markov methods when it is guaranteed to run quickly.

We now show how to run the chain of Dyer and Greenhill as a complete coupling chain. This gives a polynomial expected running time over the same set of parameters as their chain,  $\lambda \leq 2/(\Delta - 2)$ , but also allows us to use exact sampling technique of coupling from the past. Dyer and Greenhill analyzed their chain using the technique of path coupling, yielding a bound of

$$\tau(\epsilon) \leq \min \left\{ \frac{2(1 + \lambda)}{2 - (\Delta - 2)\lambda} n \log_2(n), \lceil 2n^2 e(1 + \lambda)(\ln(n) + 1) \rceil \lceil \ln(\epsilon^{-1}) \rceil \right\}$$

on the mixing time, provided  $\lambda \leq 2/(\Delta - 2)$ . The following theorem summarizes our result.

**Theorem 1** *Utilizing our algorithm, exact samples may be taken from the hard core gas model in expected time*

$$4 \min \left\{ \frac{2(1 + \lambda)}{2 - (\Delta - 2)\lambda} n \log_2(2n), 2n^2(1 + \lambda) \right\}$$

if  $\lambda \leq 2/(\Delta - 2)$ .

While our proof yields a complete coupling chain that can be used to generate exact samples, it also gives a slightly stronger bound on the mixing time when  $\lambda$  is close to  $2/(\Delta - 2)$ .

**Corollary 1** *The Dyer and Greenhill chain has mixing time*

$$\tau(\epsilon) \leq \min \left\{ \frac{2(1 + \lambda)}{2 - (\Delta - 2)\lambda} n \log_2(n\epsilon^{-1}), n^2 \log_2(\epsilon^{-1}) \right\}$$

when  $\lambda \leq 2/(\Delta - 2)$ .

Throughout this paper, we will describe independent sets as elements of  $\{0, 1\}^V$ , where a node is colored 0 if it is not in the independent set, and colored 1 if it is included in the independent set.

## 2 Coupling From the Past

To utilize CFTP, the chain has to be a complete bounding chain. Normally, one step of the Markov chain is taken for a specific value of the chain. That is,  $X_{t+1} = g_t(X_t)$  where  $g_t : X_t \rightarrow \Omega$  is a function chosen at random so that  $P(g_t(X_t) = j | X_t = i) = p_{ij}$ .

In a complete coupling, we take a random function  $f_t : \Omega \rightarrow \Omega$  such that  $P(f_t(i) = j) = p_{ij}$  for all  $i, j \in \Omega$ . In other words,  $f_t$  describes how to take a step in the chain no matter what our starting state was.

Often, choosing a random function  $f$  which satisfies the transition probabilities simultaneously for all states is no more difficult than choosing a function  $g$ . Consider the chain of Dyer and Greenhill for independent sets. One step of the chain is described below.

**Chain Step** *Dyer and Greenhill*

**Set**  $X = X_t$

**Choose** a vertex  $v$  uniformly at random from  $V$

**Choose**  $U$  uniformly from  $[0, 1]$

**Case 1:**  $v$  has no neighbors colored 1 in  $X_t$ , then

**if**  $U \leq \frac{1}{1+\lambda}$

**Set**  $X(v) = 1$

**Case 2:**  $v$  has exactly 1 neighbor  $w$  colored 1

**if**  $U \leq \frac{\lambda}{4(1+\lambda)}$

**Set**  $X(v) = 1, X(w) = 0$

**Set**  $X_{t+1} = X$

There are three kinds of moves in this chain. In case 1,  $v$  has no neighbors colored 1, and so it is possible that  $v$  may be colored either 0 or 1 in the next step. In case 2,  $v$  has exactly one neighbor  $w$  colored 1 and so it is a legal move to swap the values of  $v$  and  $w$ . This chain has the desired stationary distribution because it satisfies the reversibility condition  $\pi(x)p_{xy} = \pi(y)p_{yx}$ .

The key feature to note in this chain is that once the random vertex  $v$  and random uniform  $U$  are chosen, step 3 may be applied to any state in the chain to yield a new state chosen according to the transition probabilities of the Markov chain. Hence a random  $f$  for all states may be found as quickly as a random  $g$  for a particular state

Let  $F_t = f_1 \circ f_2 \circ \dots \circ f_t$ . If  $X_0$  was the starting state of the chain at time 0, after  $t$  steps the chain will be in state  $F_t(X_0)$ . CFTP may be used to draw samples exactly from the stationary distribution when  $F_t(X_0)$  is eventually a constant with probability 1 as  $t$  goes to infinity [6]. Intuitively,  $F_t$  is a constant if no matter where we started the chain at time 0, after  $t$  time steps it is in the same state.

Now,  $F_t : \Omega \rightarrow \Omega$  so the trivial way of checking to see if it is a constant is to check the value of  $F_t(i)$  for all  $i \in \Omega$ . This would be highly inefficient, since the size of  $\Omega$  is the number of independent sets in the graph, a number which might be exponentially large in the number of nodes. A more efficient method is needed to determine when  $F_t$  is constant.

For more details on how knowing when  $F_t$  is constant may be turned into an algorithm for exact sampling using CFTP, the original paper of Propp and Wilson [6] gives the complete algorithm. For our purposes, all we need is the following theorem.

**Theorem 2** *Suppose that  $F_t$  is provably constant with probability  $1/2$  after  $T$  time steps. Then CFTP will generate exact samples from the stationary distribution of the Markov chain in expected time  $4T$ .*

## 2.1 The Bounding Chain

In order to show that  $F_t$  is constant, we introduce a second chain with state space  $\{\{0\}, \{1\}, \{0, 1\}\}^V$  which we shall refer to as a bounding chain. A chain with state space  $2^{C^V}$  is a bounding chain for a chain with state space  $C^V$ , if whenever  $X \in C^V$  and  $Y \in 2^{C^V}$ ,  $X(v) \in Y(v)$  for all vertices  $v$ . This idea was introduced in [4] and independently in [3] as a means for showing when  $F_t$  is constant.

Start with  $Y_0 = \{0, 1\}^V$ , so that trivially  $X_0(v) \in Y_0(v)$  for all  $v$ . At each step, we will ensure that  $X_t(v) \in Y_t(v) \Rightarrow X_{t+1}(v) \in Y_{t+1}(v)$  for all  $v$ . The following describes how to take steps which satisfy this requirement.

**Bounding Chain Step****Set**  $Y = Y_t$ **Choose** a vertex  $v$  uniformly at random from  $V$ **Choose**  $U$  uniformly from  $[0, 1]$ **Case 1:** All neighbors of  $v$  are colored  $\{0\}$ **if**  $U \leq \frac{1}{1+\lambda}$ **Set**  $Y(v) = \{1\}$ **else**  $Y(v) = \{0\}$ **Case 2:**  $v$  has exactly 1 neighbor  $w$  colored  $\{1\}$ , rest are  $\{0\}$ **if**  $U \leq \frac{\lambda}{4(1+\lambda)}$ **Set**  $Y(v) = \{1\}, Y(w) = \{0\}$ **Case 3:**  $v$  has more than one neighbor colored  $\{1\}$ **Set**  $Y(v) = \{0\}$ **Case 4:** One neighbor  $w$  colored  $\{0, 1\}$ , rest colored 0**if**  $U \leq \frac{\lambda}{4(1+\lambda)}$ **Set**  $Y(v) = \{1\}, Y(w) = \{0\}$ **elseif**  $\frac{\lambda}{4(1+\lambda)} \leq U \leq \frac{\lambda}{(1+\lambda)}$ **Set**  $Y(v) = \{0, 1\}$ **elseif**  $U \geq \frac{\lambda}{1+\lambda}$ **Set**  $Y(v) = \{0\}$ **Case 5:** One neighbor  $w$  colored  $\{1\}$ , at least one colored  $\{0, 1\}$ **if**  $U \leq \frac{\lambda}{4(1+\lambda)}$ **Set**  $Y(v) = \{0, 1\}, Y(w) = \{0, 1\}$ **Case 6:** More than one neighbor is unknown, rest are 0**if**  $U \leq \frac{\lambda}{1+\lambda}$ **Set**  $Y(v) = \{0, 1\}$ **elseif**  $U \geq \frac{\lambda}{1+\lambda}$ **Set**  $Y(v) = \{0\}$ 

In cases 1, 2, and 3 the vertex  $v$  and all its neighbors are completely known, so these steps are the same as the regular chain. In case 4, one neighbor is unknown, the rest of the neighbors are 0. Hence if  $U$  is small enough that we switch if the unknown is colored 1, or just turn  $v$  to 1 if the unknown is off, then we know that  $v$  will be 1 and the unknown will be 0. If  $U$  tries to turn  $v$  on but isn't low enough to switch, then  $v$  will be the opposite value of the unknown, and so will be unknown as well.

In case 5, if one neighbor is 1 and at least one neighbor is unknown, we don't have to worry about turning  $v$  to 1 directly, but if a switch move is

called for, we do not know how to proceed, and both  $v$  and its 1 neighbor must be changed to unknown. In case 6, if more than one neighbor is unknown, then trying to turn  $v$  to 1 or trying a switch move leads to  $v$  being unknown.

### 3 Analysis

We analyze this chain by looking at the expected change in the number of nodes colored  $\{0, 1\}$  at each step. Let  $D_t = \{v : X_t(v) = \{0, 1\}\}$  and  $A_t = V \setminus D_t$ . From the algorithm, it is clear that the size of  $D_t$  changes by at most two at each step.  $|D_t|$  increases in size when nodes move from  $A_t$  to  $D_{t+1}$ , and decreases when nodes move from  $D_t$  to  $A_{t+1}$ . Let  $V_t$  be the random vertex chosen at time  $t$  by the algorithm. Then  $P(v = V_t) = 1/n$  for all  $v$ , and

$$E[|D_{t+1}| \mid |D_t|] = |D_t| + \frac{1}{n} \sum_v E[|D_{t+1}| - |D_t| \mid |D_t|, v = V_t].$$

So we now compute  $E[|D_{t+1}| - |D_t| \mid |D_t|, v = V_t]$  for  $v$  satisfying cases 1 through 6. Let  $N_i$  denote the number of nodes that are in case  $i$ .

Suppose first that  $v \in A_t$ . Let  $c(v)$  denote the case that  $v$  falls into, from 1 through 5. Let  $C_i$  denote the value of  $E[|D_{t+1}| - |D_t| \mid |D_t|, v = V_t]$  given that  $c(v) = i$ , and  $v$  is in  $A_t$ . Then in cases 1 or 2, the node always stays in  $A_t$ , so our expectation is 0 and  $C_1 = C_2 = 0$ . If  $v$  is in case 3, let  $w$  be the single unknown neighbor. With probability  $\lambda/(4(1+\lambda))$ ,  $v$  attempts to swap with  $w$  and so  $w$  moves to  $A_t$ . With probability  $3\lambda/(4(1+\lambda))$ ,  $v$  attempts to turn to color 1, resulting in  $v$  becoming unknown. Hence

$$\begin{aligned} C_3 &= \frac{3\lambda}{4(1+\lambda)} - \frac{\lambda}{4(1+\lambda)} \\ &= \frac{\lambda}{2(1+\lambda)}. \end{aligned}$$

In case 4, switching  $v$  leads to both  $v$  and its neighbor colored 1 to be moved to  $D_t$ , so  $C_4 = 2\frac{\lambda}{4(1+\lambda)}$ . Finally, in case 5, attempting to turn  $v$  to color 1 results in  $v$  moving to  $D_t$ . hence  $C_5 = \frac{\lambda}{1+\lambda}$ .

Now suppose that  $v$  was in  $D_t$  to start. In taking a step, we do not consider at all the color of  $D_t$ . Hence we can treat such an occurrence as always moving  $v$  out of  $D_t$ , and then treating it as though it was in  $A_t$ . Hence for  $v \in D_t$  and case  $i$ ,

$$E[|D_{t+1}| - |D_t| \mid |D_t|, v = V_t] = C_i - 1.$$

Altogether, we have that

$$\begin{aligned}
E[|D_{t+1}| | |D_t|] &= \sum_{i=1}^5 \frac{1}{n} \left[ \sum_{v \in A_t, c(v)=i} C_i + \sum_{v \in D_t, c(v)=i} (C_i - 1) \right] \\
&= -\frac{|D_t|}{n} + \sum_{i=1}^5 \frac{1}{n} \sum_{v: c(v)=i} C_i \\
&= -\frac{|D_t|}{n} + N_3 \cdot C_3 + N_4 \cdot C_4 + N_5 \cdot C_5 \\
&\leq -\frac{|D_t|}{n} + \frac{1}{n} [N_3 + N_4 + 2N_5] \max\{C_3, C_4, \frac{1}{2}C_5\}
\end{aligned}$$

Now  $N_3$ ,  $N_4$  and  $N_5$  are not boundless. Every vertex counted in  $N_3$  and  $N_4$  is adjacent to at least one vertex in  $D_t$ . Every vertex counted by  $N_5$  is adjacent to at least two vertices in  $D_t$ . Taken as a whole, the vertices in  $D_t$  are adjacent to at most  $|D_t|\Delta$  different vertices. Hence

$$N_3 + N_4 + 2N_5 \leq \Delta |D_t|.$$

In our case,  $C_3 = C_4 = \frac{1}{2}C_5$ , and so  $C = \max\{C_3, C_4, \frac{1}{2}C_5\} = \frac{1}{2} \frac{\lambda}{1+\lambda}$ , so that

$$\begin{aligned}
E[|D_{t+1}| | |D_t|] &\leq |D_t| + \frac{|D_t|}{n} + \frac{|D_t|\Delta}{n} C \\
&= |D_t| \left( 1 - \frac{\Delta\lambda/(2(1+\lambda)) - 1}{n} \right) \\
&= |D_t|\beta
\end{aligned}$$

There are two ways to proceed in the analysis at this point, depending on the value of  $\beta$ . First, if  $\beta \leq 1$ , then  $|D_t|$  is a martingale with respect to the natural filtration. The expected time for  $|D_t|$  to reach 0 in such a case was computed in [4]. Here we present a better argument for dealing with such cases which improves upon the time bound given in [4] by a factor of 2.

**Theorem 3** *Suppose that  $X_0, X_1, \dots$  is a supermartingale on  $\{0, \dots, n\}$  such that  $P(X_{t+1} \neq i | X_t = i) > p_i > 0$ . Let  $\tau$  be the first time that  $X_\tau = 0$ . Then*

$$E[\tau] \leq \sum_i \frac{2i}{p_i}.$$



**Proof.** Let  $N_i$  be the number of times that  $X_t = i$ . Since the probability that  $X_{t+1} < X_t$  is positive for every value of  $X_t$ ,  $N_i$  are finite with probability 1. We know that  $N_i$  is a nonnegative integer, and we proceed by bounding  $P(N_i \geq x + 1 | N_i \geq x)$  for  $x \geq 1$ .

Suppose that  $N_i \geq x$ , so for some time  $t$ ,  $X_t = i$ , and  $X_{t'} = i$  for  $x - 1$  values of  $t'$  before  $t$ . Then for  $N_i \geq x + 1$ , the value of  $X_{t'} = i$  for some  $t' > t$ . We upper bound the probability that this return to  $i$  occurs by lower bounding the probability that it never occurs.

One way in which  $X_t$  will never return to  $i$  is by moving below  $i$  at the next step, and then staying there. Let  $d_j$  denote  $P(X_{t+1} = i + j)$  for all integers  $j$ . Let  $\tau_{i,0}$  be the first time the  $X$  process hits  $\{0, 1\}$ . Then

$$\begin{aligned} P(X_{\tau_{i,0}} = i | X_{t+1} = j) &= E[X_{\tau_{i,0}}] \\ &\leq E[X_{t+1}] \\ &= i + j \end{aligned}$$

Therefore  $P(X_{\tau_{i,0}} = i | X_{t+1} = j) \leq 1 + j/i$  and  $P(X_{\tau_{i,0}} = i | X_{t+1} = j) \geq -j/i$ . The probability that  $X_{t+1} = j$  is just  $d_j$ , so  $P(X_{\tau_{i,0}} = i) \geq \sum_{j \leq -1} -j d_j / i$ .

Intuitively, this is the expected change downward, and must be larger than the expected change upward. More formally, let  $p_{down} = \sum_{j \leq -1} d_j$  and  $p_{up} = \sum_{j \geq 1} d_j$ . Then

$$\sum_{j \leq -1} j d_j + \sum_{j \geq 1} j d_j = E[X_{t+1} - X_t | X_t] \geq 0$$

Now,  $\sum_{j \geq 1} j d_j \geq \sum_{j \geq 1} d_j = p_{up}$ . Hence  $\sum_{j \leq -1} -j d_j \geq p_{up}$ .

On the other hand,

$$\sum_{j \leq -1} j d_j \leq \sum_{j \leq -1} d_j = -p_{down},$$

so  $\sum_{j \leq -1} -j d_j \geq p_{down}$ . Putting the two together, we have that

$$2 \sum_{j \leq -1} -j d_j \geq p_{down} + p_{up} = p_i.$$

Therefore  $P(X_{\tau_{i,0}} = i) \geq p_i / (2i)$ .

Hence  $P(N_i \geq x + 1 | N_i \geq x) \leq 1 - p_i / (2i)$ , and put another way,  $P(N_i \geq x + 1) \leq (1 - p_i / (2i)) P(N_i \geq x)$ . Now  $E[N_i] = \sum_{x \geq 1} P(N_i \geq x)$  is just a geometric series. Trivially bounding  $P(N_i \geq 1)$  by 1 yields

$$E[N_i] \leq 2i / p_i,$$

and the theorem follows.  $\square$

One way in which the value of  $|D_t|$  changes at each step is by picking a node in  $|D_t|$  and changing its value to  $\{0\}$ . This happens with probability  $|D_t|/(n(1+\lambda))$ . In terms of our theorem,  $p_i \geq i/(n(1+\lambda))$ . Hence the expected time until  $|D_t| = 0$  is bounded above by  $n^2(1+\lambda)$ . Together with the CFTP protocol, this proves the first half of theorem 1.

If  $\beta < 1$ , then we have a sequence which grows smaller more rapidly than a general martingale. In this case,

$$E[|D_t|] = E[E[|D_t||D_{t-1}|]] \leq E[\beta|D_{t-1}|] = \beta E[|D_t|]$$

and an inductive argument can be used to show that  $E[|D_t|] \leq \beta^t E[|D_0|]$ . We know that  $|D_0| \leq n$ . Furthermore,  $|D_t|$  is a nonnegative integer, so if  $E[|D_t|] < 1$ , then  $P(|D_t| = 0) > 1/E[|D_t|]$  by Markov's inequality. Hence the time such that  $P(|D_t| = 0) > 1/2$  is at most the time that  $\beta^t < 1/(2n)$ .

We need one more well known fact. For all  $x$ ,  $1+x \leq e^x$ . In our case this implies that

$$\beta^t \leq \exp\left(\frac{2 - (\Delta - 2)\lambda}{2(1 + \lambda)}t\right).$$

Setting  $t = 2 \ln 2(1+\lambda)/(2 - (\Delta - 2)\lambda) \ln(2n)$  gives  $P(|D_t| = 0) > 1/2$ , which completes the proof of theorem 1.

## References

- [1] Martin Dyer, Alan Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. Technical Report ECS-LFCS-98-391, University of Edinburgh, 1998.
- [2] Martin Dyer and Catherine Greenhill. A genuinely polynomial-time algorithm for sampling two-rowed contingency tables. In *25th International Colloquium on Automata, Languages and Programming*, pages 339 – 350. EATCS, 1998.
- [3] Olle Häggström and Karin Nelander. Exact sampling from anti-monotone systems, 1997. Preprint.
- [4] Mark L. Huber. Exact sampling and approximate counting techniques. In *Proceedings of the 30th Annual Symposium on the Theory of Computing*, pages 31–40, 1998.

- [5] Michael Luby and Eric Vigoda. Approximately counting up to four (extended abstract). In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 682–687, 1997.
- [6] James G. Propp and David B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1 and 2):223–252, 1996.