A New Approach to Perfect Sampling From Nasty Distributions

Mark Huber
NSF Postdoc
Visiting--Department of Statistics
Stanford University
mhuber@orie.cornell.edu
www-stat.stanford.edu/~mhuber

Joint work with:

James Fill
Department of Mathematical Sciences
The Johns Hopkins University
jimfill@jhu.edu

Generating Perfect Samples

- The Randomness Recycler
  - New approach
  - First linear time algorithms for some problems
  - Replaces traditional Markov chains
  - Constructs dual Markov chain
- Older Perfect Samplers (complete couplers)
  - Coupling from the past (CFTP)
- Acceptance/Rejection
  - Restricted permutations

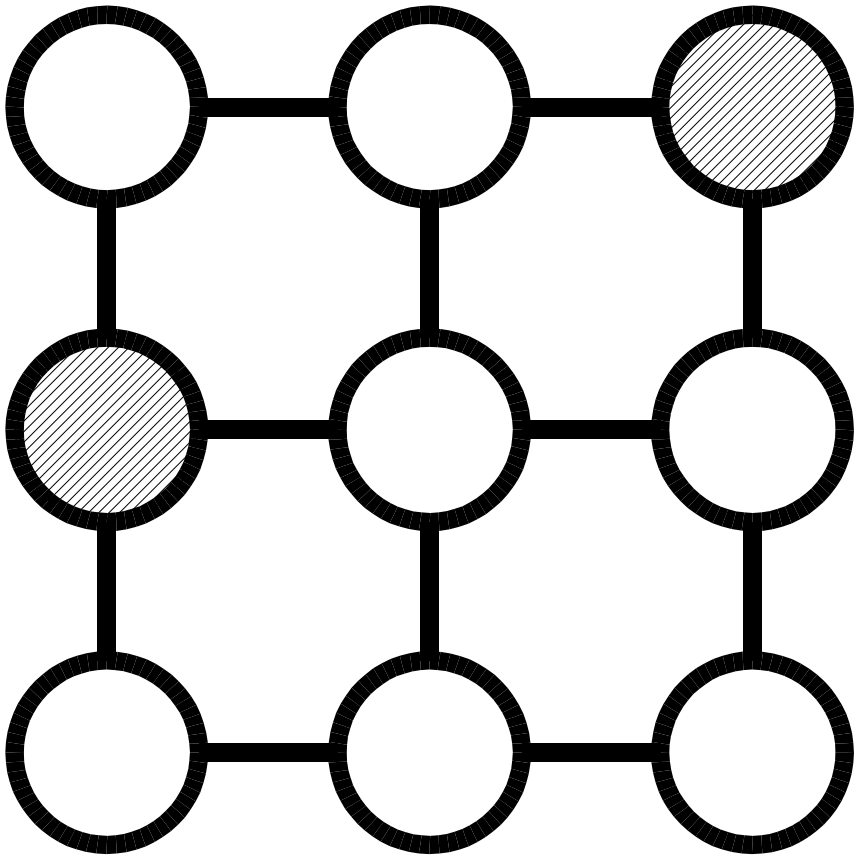
Rubik's Cube Perfect Sampler

- Uniformly position and orient corner cubes
- Uniformly position and orient edge cubes
- Accept if valid position (1/36)
- Things to note
  - Linear time
  - Partially mixed up cube

Overview

- Randomness Recycler (Fill, H. '00)
  - Slowly grows a stationary sample
  - Interruptible perfect sample
  - Different from previous perfect sampling approaches such as CFTP & FMMR
  - Each step accepts or rejects and recycles
- Ex: The Hard Core Gas Model
  - Phase Transition
  - *NP*-hard in certain cases

Hard Core Gas Model (aka Independent sets of a graph)



Distribution

$$P(\textit{choosing } I) = \pi(I) = \frac{\lambda^{|I|}}{Z_\lambda}$$

$\lambda < 1$: Small size independent sets

$\lambda = 1$: Uniform over independent sets

$\lambda > 1$: Large size independent sets

Why is this hard?

Let Δ be the maximum degree of the graph

Dyer, Frieze, & Jerrum '98

Can't sample in polynomial time when

$$\lambda > \frac{25}{\Delta}$$

unless

$$RP = NP$$

Sampling History

Single Site chain

Edge chain (Luby & Vigoda '97)

Perfect sampling with single site using CFTP (Haggstrom & Nelander '97)

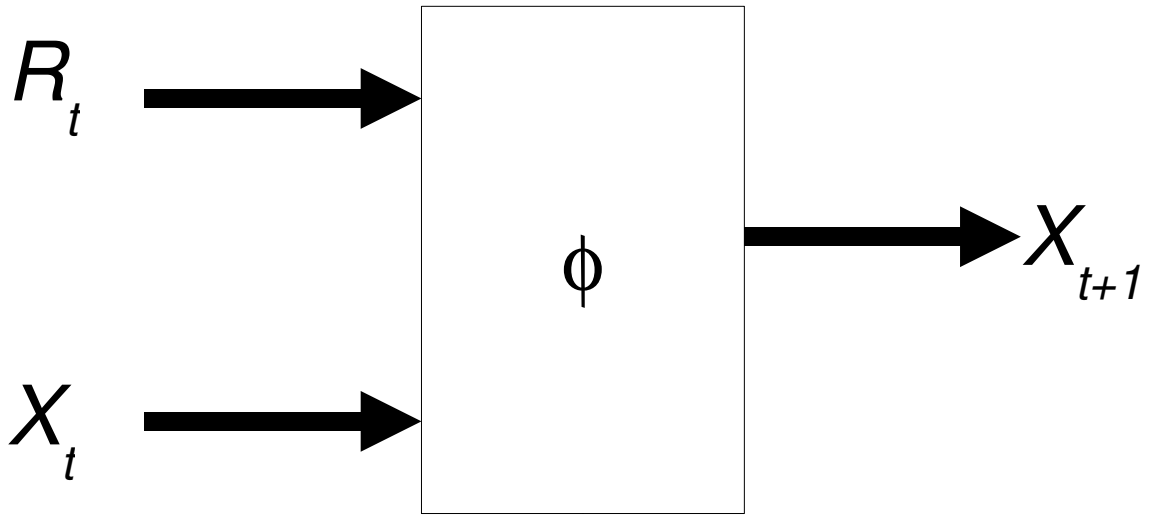
Dyer-Greenhill chain ('97)

Perfect sampling with Dyer-Greenhill (H. '98) using CFTP or FMMR

Luby Vigoda ' 99 better analysis of single site

Markov chains

Stochastic Iterative Scheme



ϕ: deterministic, R_t are random bits

complete coupling, grand coupling

Stationary Distribution

For large t,

X_t is approximately distributed as π

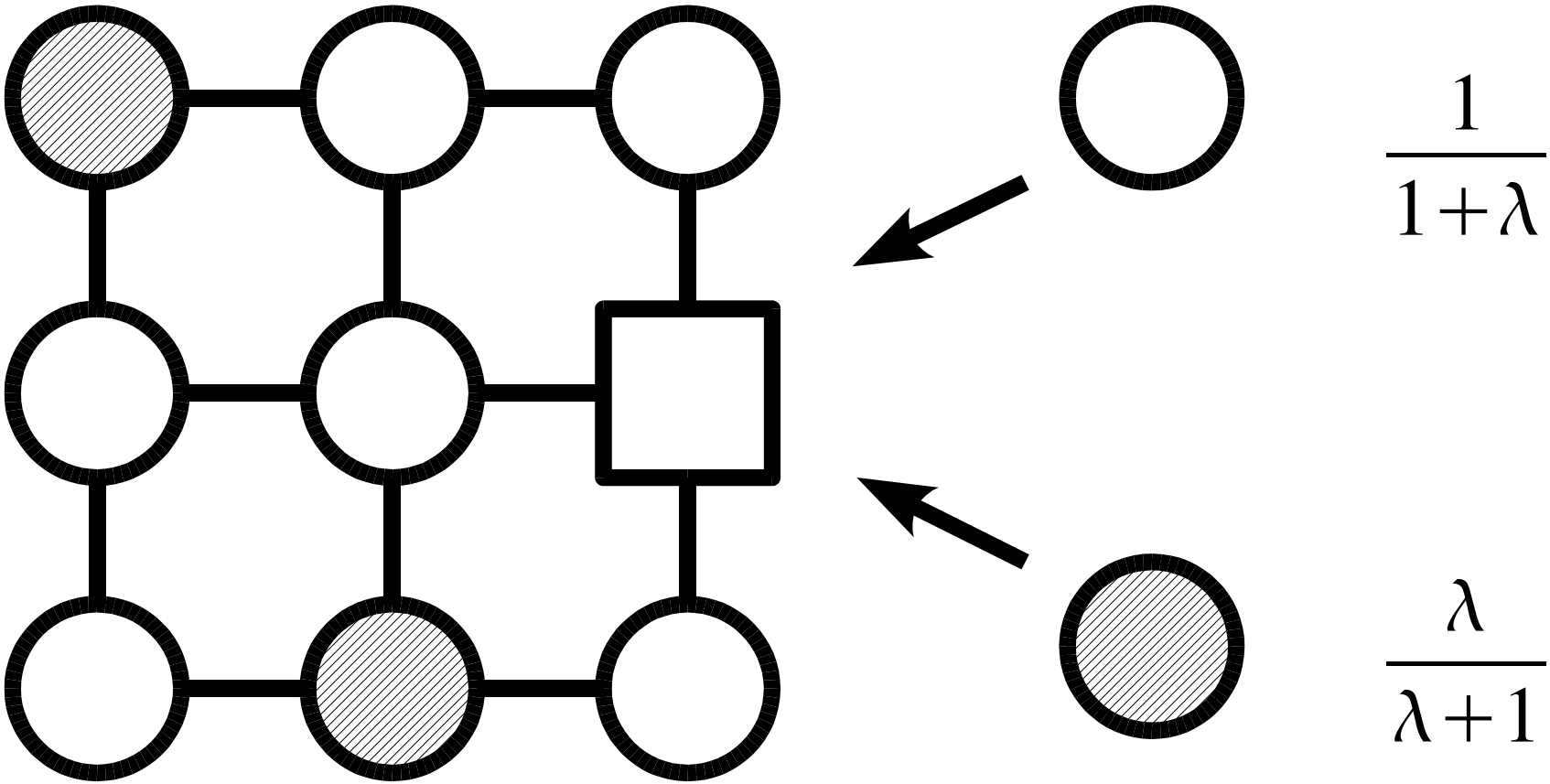
the stationary distribution

Problems:

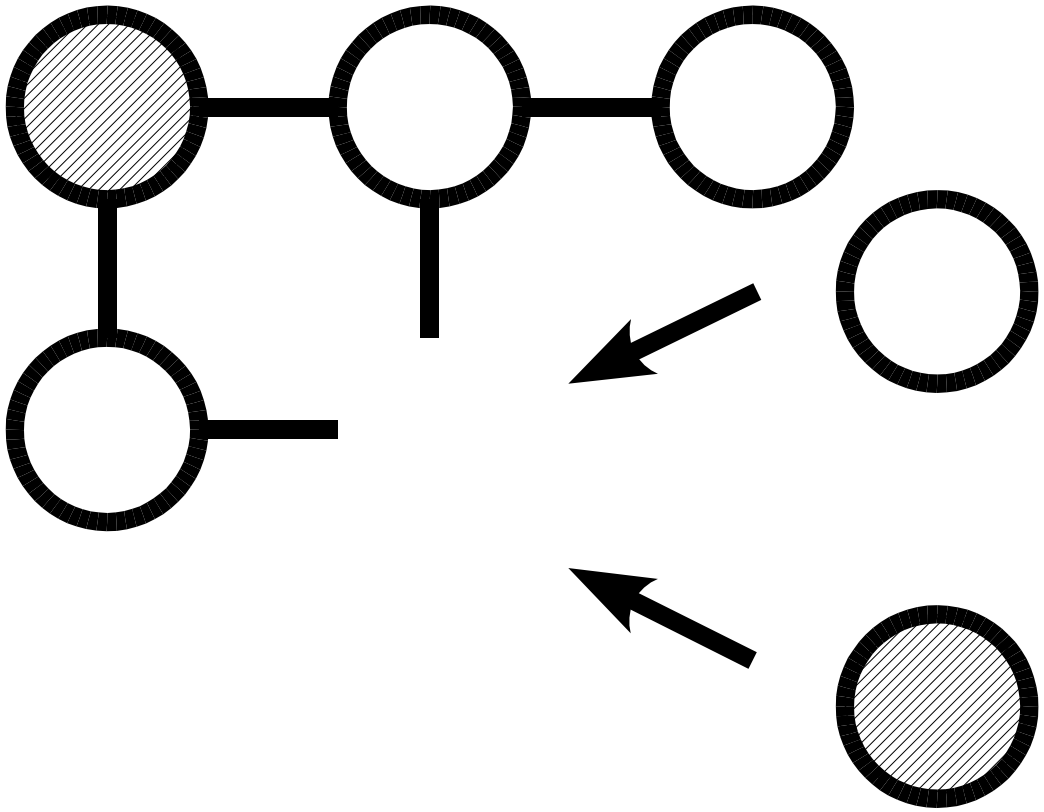
How quick is convergence?

Only approximately stationary

Heat Bath



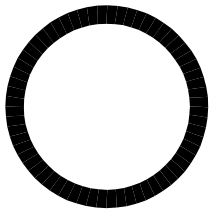
RR: Somewhere in the middle

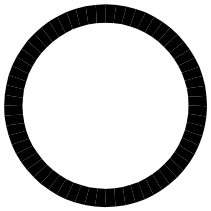


G_4

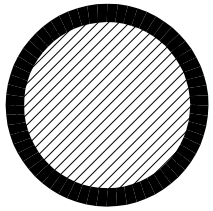
Randomness Recycler: Start

G_1 π_{G_1}



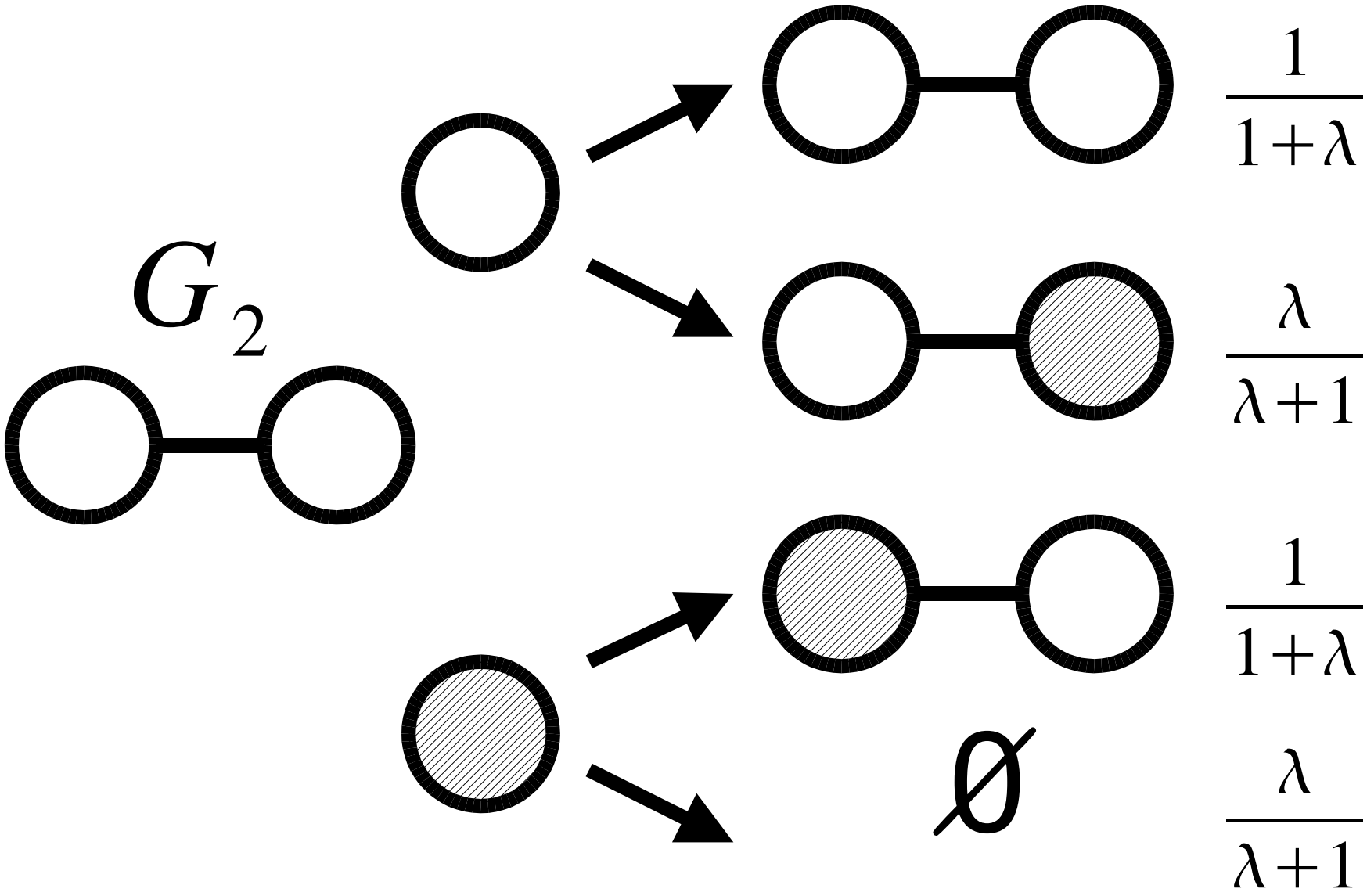


$$\frac{\lambda^0}{1+\lambda} = \frac{1}{1+\lambda}$$



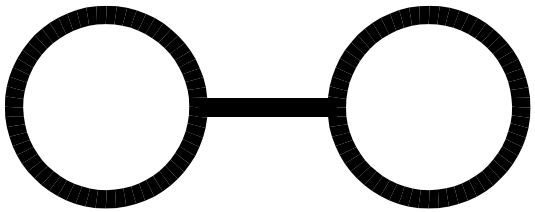
$$\frac{\lambda^1}{1+\lambda} = \frac{\lambda}{1+\lambda}$$

RR: Add a node

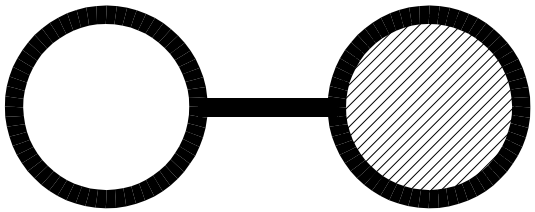

G_2
$\frac{1}{1+\lambda}$
$\frac{\lambda}{\lambda+1}$
$\frac{1}{1+\lambda}$
$\emptyset$
$\frac{\lambda}{\lambda+1}$

Reality Check

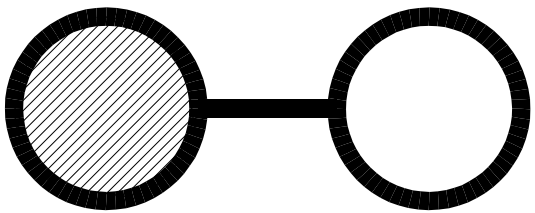
Probability of making it to G_2 and $x=$



$$\left(\frac{1}{1+\lambda}\right)\left(\frac{1}{1+\lambda}\right)$$



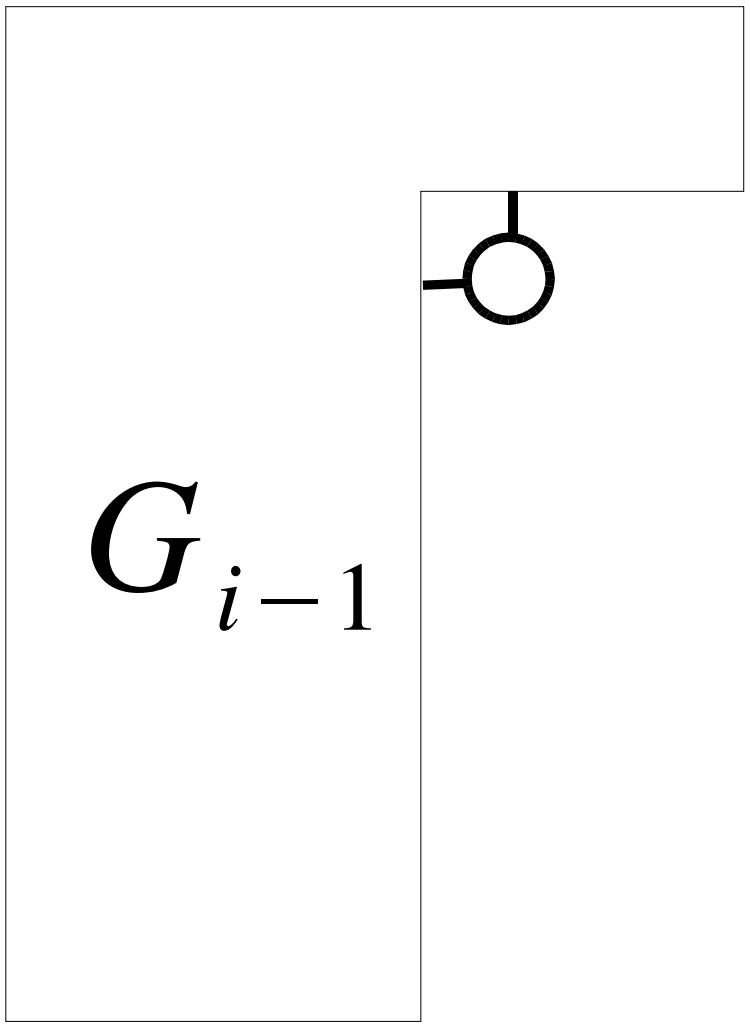
$$\left(\frac{1}{1+\lambda}\right)\left(\frac{\lambda}{1+\lambda}\right) \propto \pi_{G_2}(x)$$



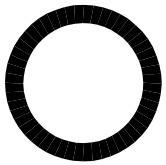
$$\left(\frac{\lambda}{1+\lambda}\right)\left(\frac{1}{1+\lambda}\right)$$

$$P(X=x|at\,G_2)=\pi_{G_2}(x)$$

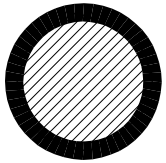
General Add Step



Neighbors all unoccupied:

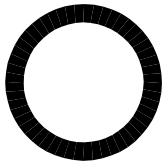


$$\frac{1}{1+\lambda} \quad G_i$$



$$\frac{\lambda}{\lambda+1} \quad G_i$$

Neighbor occupied:



$$\frac{1}{1+\lambda} \quad G_i$$

$$\frac{\lambda}{\lambda+1} \quad G_0$$

Dual Chain

X_t : state of the chain at time t

X_t^* : set of nodes that are part of graph

$$X_{t+1}^* = \begin{cases} \emptyset & \frac{\lambda}{1+\lambda} \\ X_t^* \vee \{v\} & \frac{1}{1+\lambda} \end{cases}$$

Stationarity Preserving steps

$$P(X_t = x_t | X_0^* = x_{0,}^* ..., X_t^* = x_t^*) = \pi_{X_t^*}(x_t)$$

$$\downarrow$$

$$P(X_{t+1} = x_{t+1} | X_0^* = x_{0,}^* ..., X_{t+1}^* = x_{t+1}^*) = \pi_{X_{t+1}^*}(x_{t+1})$$

After induction:

$$P(X_0 = x_0 | X_0^* = x_0^*) = \pi_{X_0^*}(x_0)$$

$$\downarrow$$

$$P(X_t = x_t | X_0^* = x_{0,}^* ..., X_t^* = x_t^*) = \pi_{X_t^*}(x_t)$$

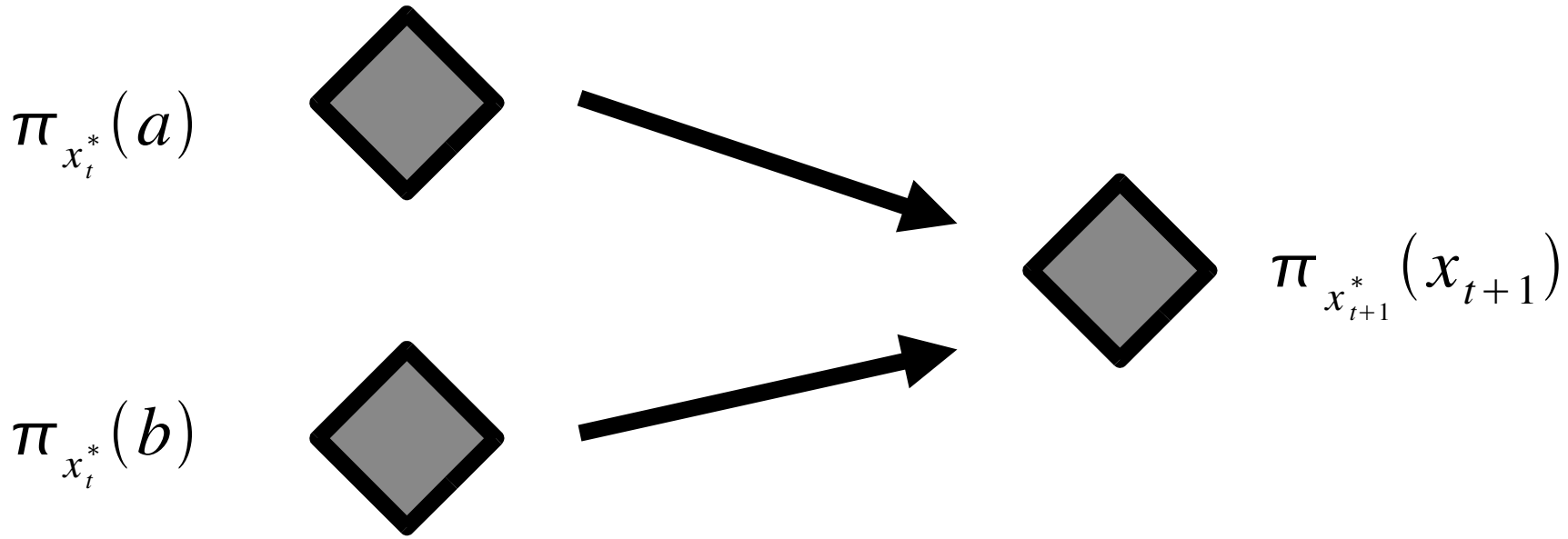
$$\downarrow$$

$$P(X_t = x_t | X_t^* = V) = \pi_V(x_t) = \pi(x_t)$$

Insuring Stationarity Preserving

Lemma: Steps preserve stationarity if

$$\sum_{x_t} \pi_{x_t^*}(x_t) \cdot P\left((x_t, x_t^*), (x_{t+1}, x_{t+1}^*)\right) \propto \pi_{x_{t+1}^*}(x_{t+1})$$



Proof...

Given: $P(X_t = x_t | X^*_0 = x^*_{0,}..., X^*_t = x^*_t) = \pi_{X^*_t}(x_t)$

$$P(X_{t+1} = x_{t+1} | A^*_{t+1})$$

$$= \frac{P(X_{t+1} = x_{t+1}, X^*_{t+1} = x^*_{t+1} | A^*_t)}{P(X^*_{t+1} = x^*_{t+1} | A^*_t)}$$

$$= C_1(A^*_{t+1}) \sum_{x_t} P(X_{t+1} = x_{t+1}, X^* = x^*_{t+1}, X_t = x_t | A^*_t)$$

$$= C_1(A^*_{t+1}) \sum_{x_t} P(X_t = x_t | A^*_t)$$

$$\cdot P(X_{t+1} = x_{t+1}, X^* = x^*_{t+1} | A^*_t, X_t = x_t)$$

$$= C_1(A^*_{t+1}) \sum_{x_t} \pi_{x^*_t}(x_t) \cdot P((x_{t,} x^*_t), (x_{t+1}, x^*_{t+1}))$$

$$= C_1(A^*_{t+1}) C_2(A^*_{t+1}) \pi_{x^*_{t+1}}(x_{t+1})$$

Why this is good

$$\sum_{x_{t+1}} P\left(X_{t+1}=x_{t+1}\middle|A^*_{t+1}\right)$$
$$=\sum_{x_{t+1}} C_1\left(A^*_{t+1}\right)C_2\left(A^*_{t+1}\right)\pi_{x^*_{t+1}}\left(x_{t+1}\right)$$
$$=C_1\left(A^*_{t+1}\right)C_2\left(A^*_{t+1}\right)\sum_{x_{t+1}} \pi_{x^*_{t+1}}\left(x_{t+1}\right)=1$$

So

$$P\left(X_0=x_0\middle|X^*_0=x^*_0\right)=\pi_{x^*}\left(x_0\right)$$
$$\downarrow$$
$$P\left(X_t=x_t\middle|X^*_t=V\right)=\pi\left(x_t\right)$$

Example: Hard Core Gas Model

Case 1: $X^*_{t+1} = \emptyset$

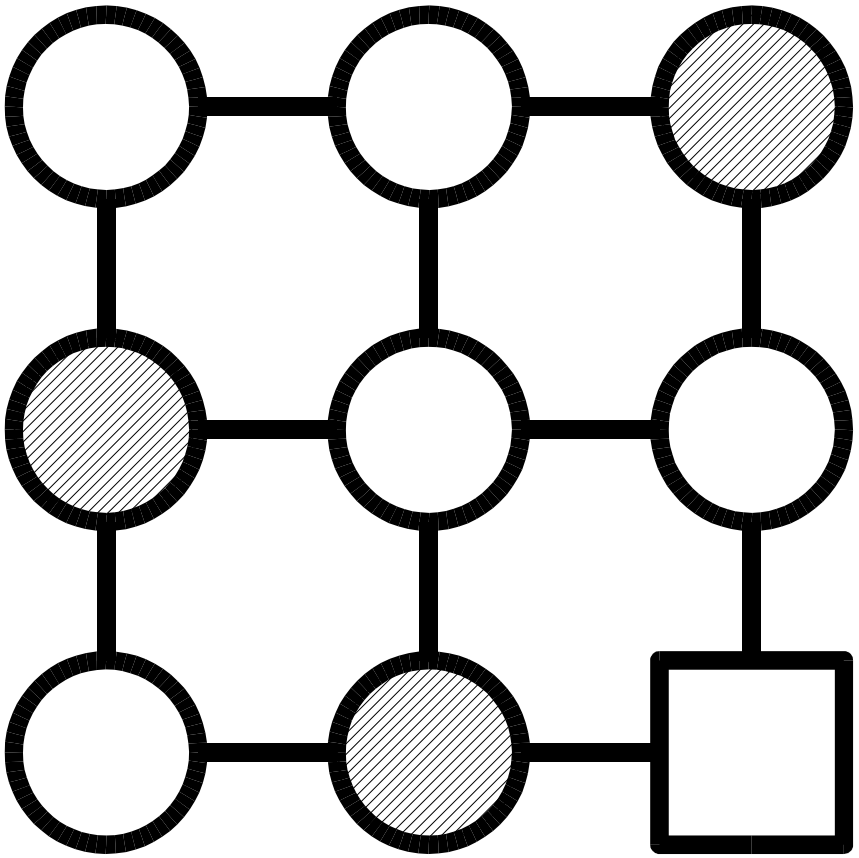
Case 2: $X^*_{t+1} = X^*_t \cup \{v\}$

If we put v in

$$\sum_{x_t} \pi_{x^*_t}(x_t) \cdot P\left((x_t, x^*_t), (x_{t+1}, x^*_{t+1})\right)$$

$$= \frac{\lambda^{|x_t|}}{z_{x^*_t}} \cdot \frac{\lambda}{1+\lambda} = \frac{\lambda^{|x_{t+1}|}}{z_{x^*_{t+1}}} \cdot \frac{z_{x^*_{t+1}}}{z_{x^*_t}(1+\lambda)}$$

Modifying rejection move



Bad news:

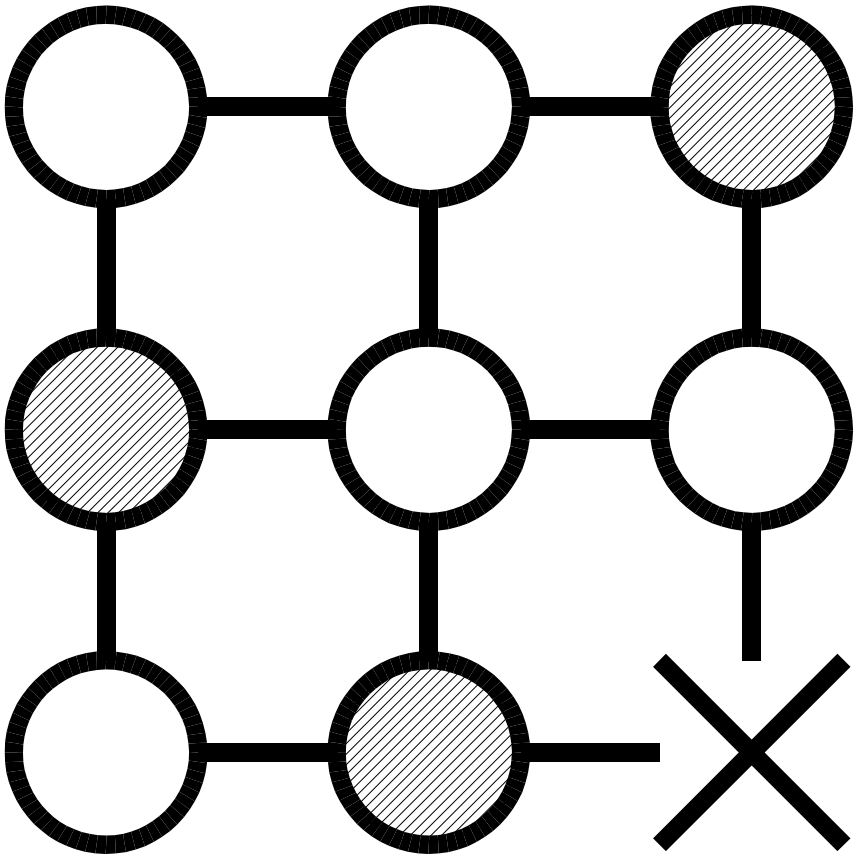
Low chance of finishing

$$\left(\frac{1}{1+\lambda}\right)^n$$

New Idea:

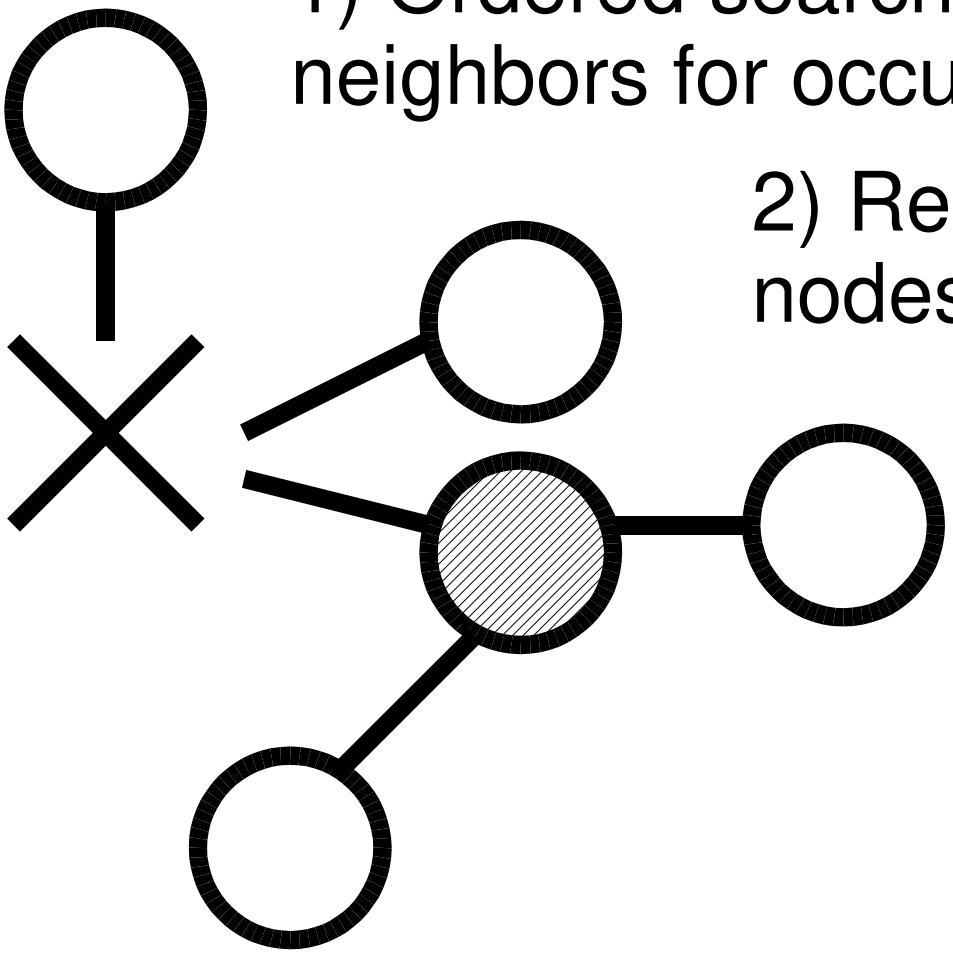
Rejection doesn't have to mean starting over

How rejection biases sample



Rejection means x_t has at least one neighbor of v occupied

How to unbias



1) Ordered search of neighbors for occupied site

2) Remove these nodes from x_t^*

Does this work?

$$|x^*_{t+1}| < |x^*_t|$$

Reconstruct: x_t from x_{t+1}

Suppose *k* neighbors of *v* removed

First *k - 1* neighbors of *v* are unoccupied

Last neighbor occupied in x_t

$$\frac{\lambda^{|x_t|}}{z_{x^*_t}} P\left(\left(x_{t,} x^*_t\right), \left(x_{t+1}, x^*_{t+1}\right)\right)$$

$$= \frac{z_{x^*_{t+1}}}{z_{x^*_t}} \frac{\lambda^{|x_{t+1}|}}{z_{x^*_{t+1}}} \lambda \frac{\lambda}{1+\lambda}$$

The algorithm

Start with $x^* = \emptyset$

Repeat

  Find $v \notin x^*$

  Case 1: v has no occupied neighbors

    Set v to occupied with prob $1/(1+\lambda)$

    Set v to unoccupied with prob $\lambda/(1+\lambda)$

    Set $x^* \leftarrow x^* \cup \{v\}$

  Case 2: has occupied neighbor

    Accept v unoccupied w/ prob $1/(1+\lambda)$

      Set $x^* \leftarrow x^* \cup \{v\}$

    Otherwise do rejection procedure

Until $x_t^* = V$

Running Time

Idea: bound run time by showing $|X_t^*|$ increases on average at each step

$$E[|X_{t+1}^*| - |X_t^*| \,||X_t^*|]$$

$$\geq \left(\frac{1}{1+\lambda}\right)(+1) - \left(\frac{\lambda}{1+\lambda}\right)(2\Delta - 1) = \alpha$$

α is positive exactly when: $\lambda < \frac{1}{2\Delta - 1}$

Expected running time bound: $\frac{|V|}{\alpha}$

More tricks

Rejection: start with random neighbor of v while searching for occupied site

Use better potential function

$$\phi_t = |X_t^*| - c|X_t|$$

Ensure that $V \setminus X_t^*$ is connected, effectively reduces Δ by 1

Altogether

$$\lambda < \frac{4}{3\Delta - 4}$$

General Procedure

- Start with simplified version of problem
- Build up to original problem one step at a time
  - At each step accept or reject as needed to make sample stationary
  - If reject, remove part of old sample to undo knowledge gained
- When at original problem, stop

Strong Stationary Stopping Times

If τ is a stopping time satisfying

$$P(X_t = x_t | T \geq \tau, X_0 = x_0) = \pi(x_0)$$

then τ is a strong stationary stopping time

Essentially undirected randomization

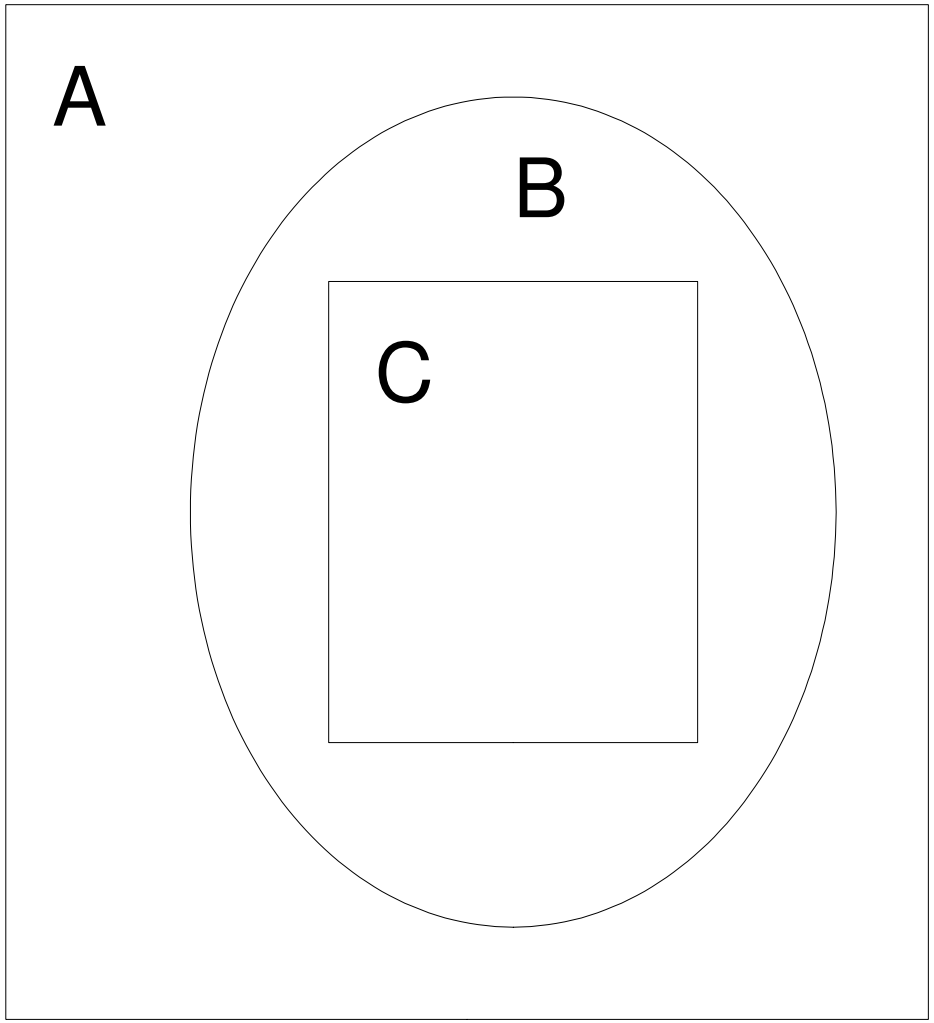
When node of interest selected, treat as algorithm

Hard Core: When node not adjacent to occupied sites in $V \setminus X_t^*$ selected

More problems

	Approximate	CFTP	DR
Ising model			
Proper colorings			
Random Cluster			
Autonormal			
Hard Core Gas			
Move Ahead 1			
Two sided interval permutations			

Acceptance/Rejection



Easy to sample from A

Prob. of acceptance

$$\frac{|B|}{|A|}$$

Running Time

$$E[T] = \frac{|A|}{|B|} \leq \frac{|A|}{|C|}$$

Restricted Permutations

Minc's Inequality

$$per(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i} \approx \prod_{i=1}^{n} \frac{r_i + 1/2 \log(2\pi r_i)}{e}$$

Van der Waerden's Inequality

$$per(A) \leq \prod_{i=1}^{n} (r_i / e)$$

1	0	1
1	1	0
0	1	1

Col. sums=row sums= γn

$$O\left(n^{3/2 + 1/(2\gamma) + 2/(\gamma\sqrt{\log n})}\right)$$

Other Perfect Samplers

- Need Markov chain, complete coupling
  - Monotonicity (Propp, Wilson '96)
  - Bounding chains (H. '98)
- Versions
  - Coupling from the past (Propp, Wilson '96)
  - Fill, Machida, Murdoch, Rosenthal ('00)
  - Read once CFTP (Wilson '00)
  - Coupling from the past (CFTP)
- Either interruptible or read-once

Conclusions

- New technique giving stationary samples
  - Different from classical Markov chains
  - Different from complete coupling perfect samplers (CFTP and FMMR)
  - Interuptible and read-once
- First linear time algorithms for several problems of interest
- Challenges
  - Handling rejection automatically
  - Making it run quickly in practice