

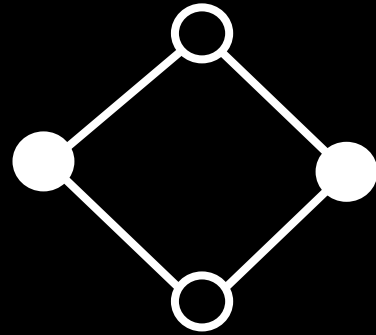
**Adaptive annealing: a near-optimal  
connection between  
sampling and counting**

**Daniel Štefankovič**  
(University of Rochester)

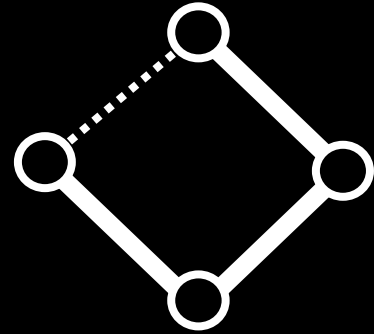
**Santosh Vempala**  
**Eric Vigoda**  
(Georgia Tech)

# Counting

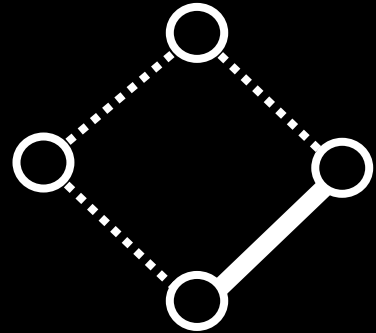
independent sets



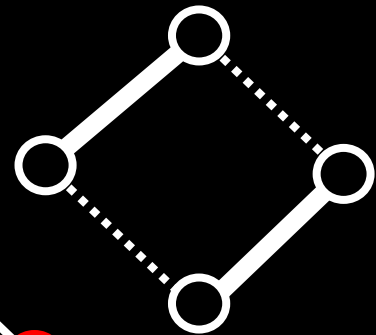
spanning trees



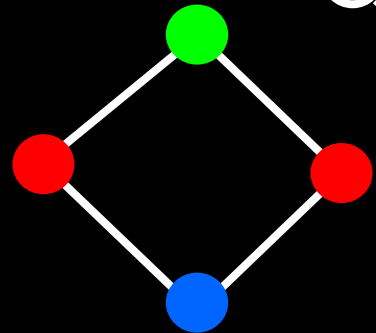
matchings



perfect matchings



k-colorings



# (approx) counting $\Leftrightarrow$ sampling

Valleau, Card'72 (physical chemistry),  
Babai'79 (for matchings and colorings),  
Jerrum, Valiant, V. Vazirani'86,

**the outcome of the JVV reduction:**

random variables:  $X_1 X_2 \dots X_t$   
such that

1)  $E[X_1 X_2 \dots X_t] = \text{"WANTED"}$

2) the  $X_i$  are easy to estimate

**squared coefficient  
of variation (SCV)**  $\frac{V[X_i]}{E[X_i]^2} = O(1)$

# (approx) counting $\Leftrightarrow$ sampling

1)  $E[X_1 X_2 \dots X_t] = \text{“WANTED”}$

2) the  $X_i$  are easy to estimate

$$\frac{V[X_i]}{E[X_i]^2} = O(1)$$

**Theorem (Dyer-Frieze'91)**

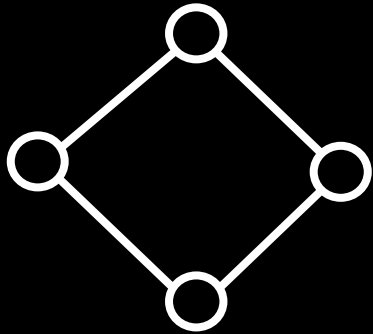
**$O(t^2/\epsilon^2)$  samples ( $O(t/\epsilon^2)$  from each  $X_i$ )**

**give**

**$1 \pm \epsilon$  estimator of “WANTED” with  $\text{prob} \geq 3/4$**

# JVV for **independent sets**

GOAL: given a graph  $G$ , estimate the number of independent sets of  $G$

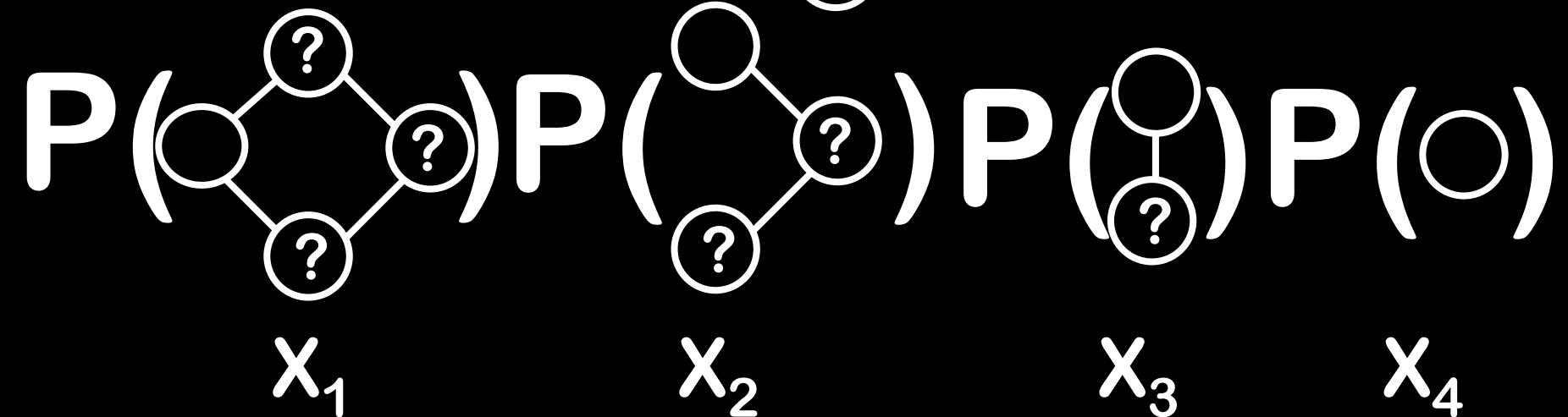
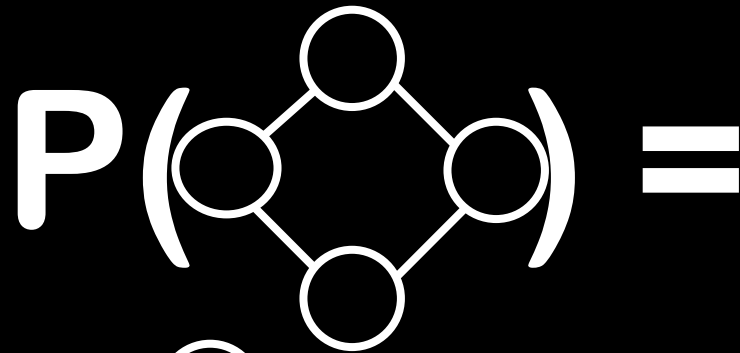


$$\# \text{ independent sets} = \frac{1}{\mathbf{P}(\text{graph})}$$

The same diamond-shaped graph as above, enclosed in large parentheses. The graph consists of 4 vertices and 4 edges forming a diamond shape.

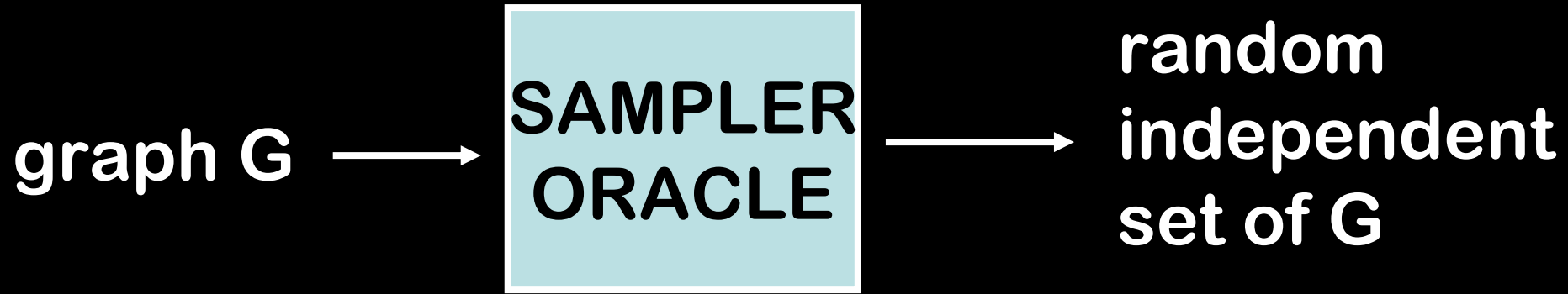
# JVV for independent sets

$$P(A \cap B) = P(A)P(B|A)$$



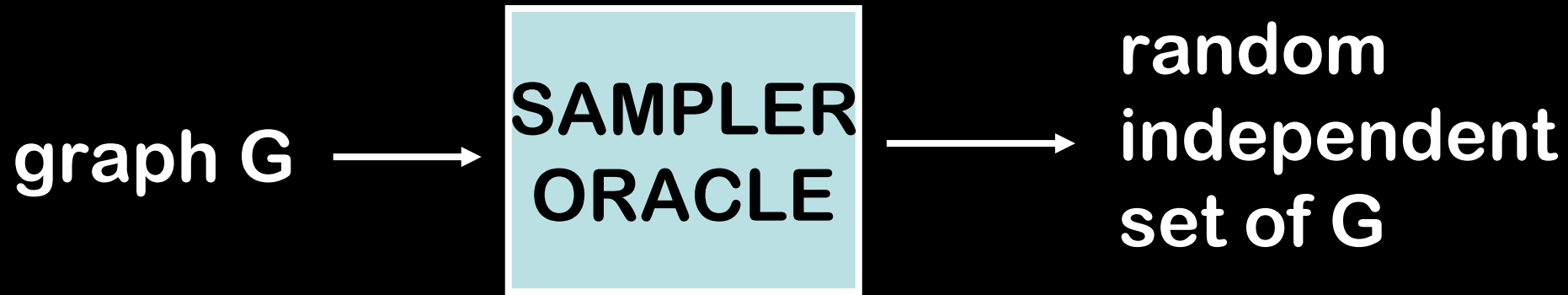
$$X_i \in [0, 1] \text{ and } E[X_i] \geq 1/2 \quad \Rightarrow \quad \frac{V[X_i]}{E[X_i]^2} = O(1)$$

**JVV: If we have a sampler oracle:**



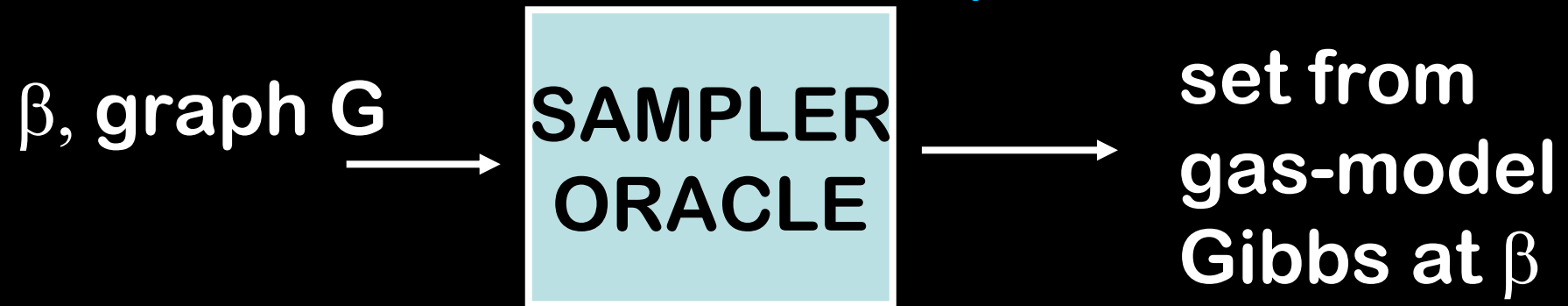
**then FPRAS using  $O(n^2)$  samples.**

**JVV:** If we have a sampler oracle:



**then FPRAS using  $O(n^2)$  samples.**

**ŠVV:** If we have a sampler oracle:



**then FPRAS using  $O^*(n)$  samples.**



# Application – independent sets

$O^*(|V|)$  samples suffice for counting

Cost per sample (Vigoda'01, Dyer-Greenhill'01)  
time =  $O^*(|V|)$  for graphs of degree  $\leq 4$ .

Total running time:

$O^*(|V|^2)$ .

# Other applications

matchings

$O^*(n^2m)$

(using Jerrum, Sinclair'89)

spin systems:

Ising model

$O^*(n^2)$  for  $\beta < \beta_c$

(using Marinelli, Olivieri'95)

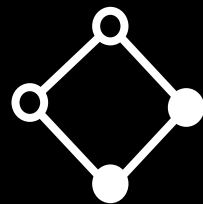
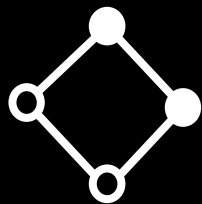
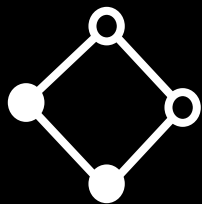
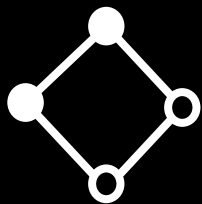
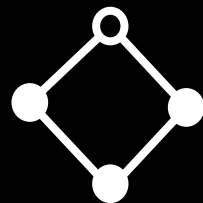
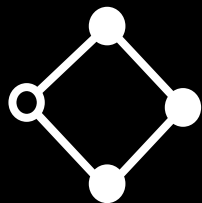
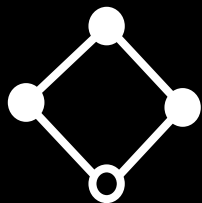
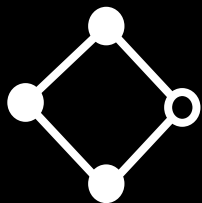
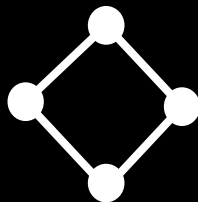
k-colorings

$O^*(n^2)$  for  $k > 2\Delta$

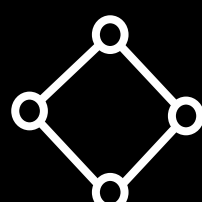
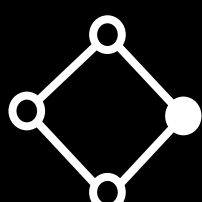
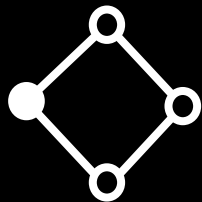
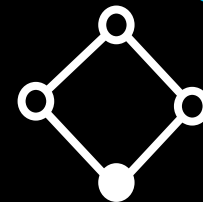
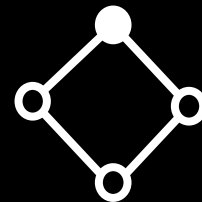
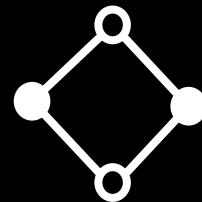
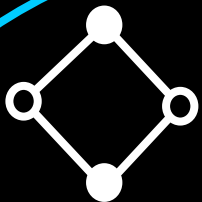
(using Jerrum'95)

total running time

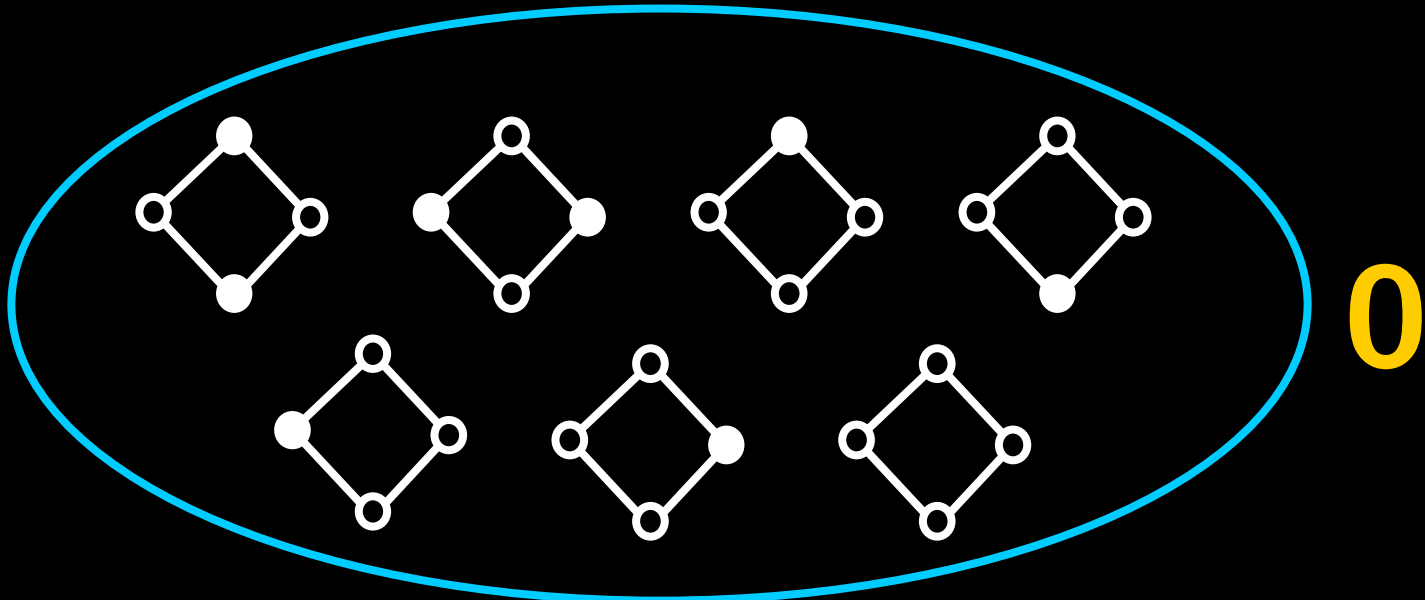
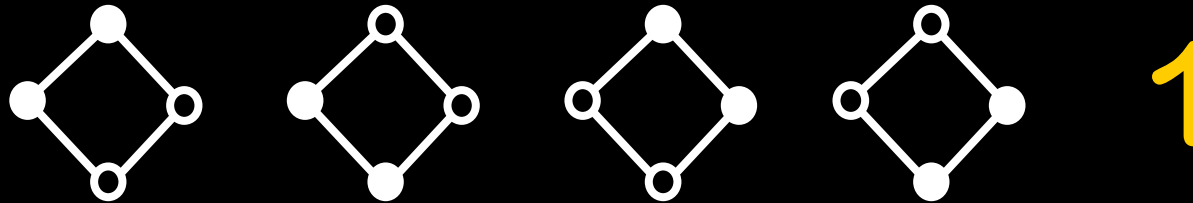
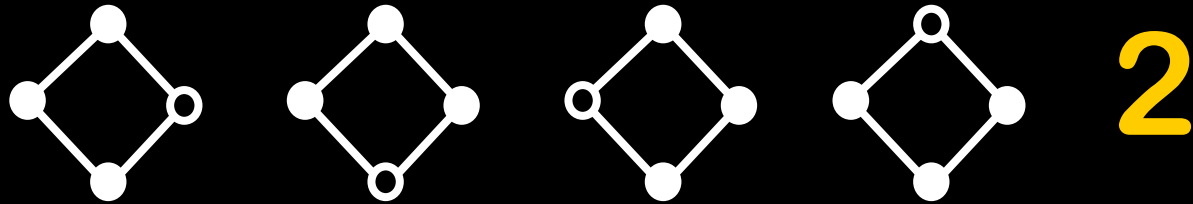
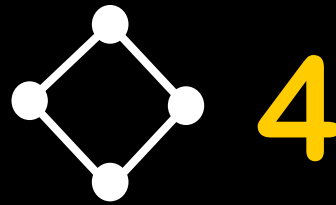
easy = hot



hard = cold



# Hamiltonian



Big set =  $\Omega$

Hamiltonian

$$H : \Omega \rightarrow \{0, \dots, n\}$$

Goal: estimate  $|H^{-1}(0)|$

$$|H^{-1}(0)| = E[X_1] \dots E[X_t]$$

# Distributions between **hot** and **cold**

$\beta$  = inverse temperature

$\beta = 0 \Rightarrow$  **hot**  $\Rightarrow$  uniform on  $\Omega$

$\beta = \infty \Rightarrow$  **cold**  $\Rightarrow$  uniform on  $H^{-1}(0)$

$$\mu_{\beta}(\mathbf{x}) \propto \exp(-H(\mathbf{x})\beta)$$

(Gibbs distributions)

# Distributions between **hot** and **cold**

$$\mu_{\beta}(\mathbf{x}) \propto \exp(-H(\mathbf{x})\beta)$$

$$\mu_{\beta}(\mathbf{x}) = \frac{\exp(-H(\mathbf{x})\beta)}{Z(\beta)}$$

**Normalizing factor = partition function**

$$Z(\beta) = \sum_{\mathbf{x} \in \Omega} \exp(-H(\mathbf{x})\beta)$$

# Partition function

$$Z(\beta) = \sum_{\mathbf{x} \in \Omega} \exp(-H(\mathbf{x})\beta)$$

have:  $Z(0) = |\Omega|$

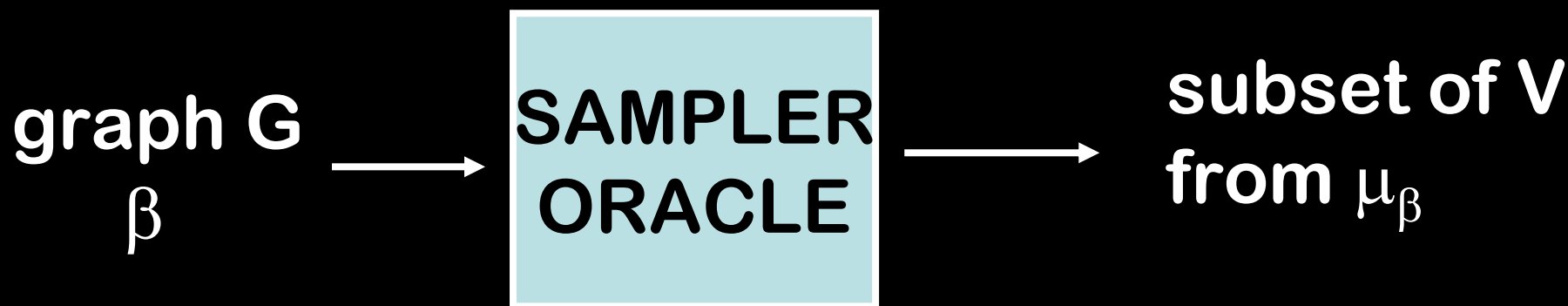
want:  $Z(\infty) = |H^{-1}(0)|$



# Assumption:

we have a sampler oracle for  $\mu_\beta$

$$\mu_\beta(\mathbf{x}) = \frac{\exp(-H(\mathbf{x})\beta)}{Z(\beta)}$$



**Assumption:**

we have a sampler oracle for  $\mu_\beta$

$$\mu_\beta(\mathbf{x}) = \frac{\exp(-H(\mathbf{x})\beta)}{Z(\beta)}$$

$\mathbf{W} \sim \mu_\beta$

# Assumption:

we have a sampler oracle for  $\mu_\beta$


$$\mu_\beta(\mathbf{x}) = \frac{\exp(-H(\mathbf{x})\beta)}{Z(\beta)}$$


$$\mathbf{W} \sim \mu_\beta \rightarrow \mathbf{X} = \exp(H(\mathbf{W})(\beta - \alpha))$$

# Assumption:

we have a sampler oracle for  $\mu_\beta$


$$\mu_\beta(\mathbf{x}) = \frac{\exp(-H(\mathbf{x})\beta)}{Z(\beta)}$$

$$\mathbf{W} \sim \mu_\beta \rightarrow \mathbf{X} = \exp(H(\mathbf{W})(\beta - \alpha))$$

can obtain the following ratio:

$$E[\mathbf{X}] = \sum_{\mathbf{s} \in \Omega} \mu_\beta(\mathbf{s}) \mathbf{X}(\mathbf{s}) = \frac{Z(\alpha)}{Z(\beta)}$$

# Our goal restated

## Partition function

$$Z(\beta) = \sum_{\mathbf{x} \in \Omega} \exp(-H(\mathbf{x})\beta)$$

Goal: estimate  $Z(\infty) = |\mathcal{H}^{-1}(0)|$

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \cdots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

$$\beta_0 = 0 < \beta_1 < \beta_2 < \dots < \beta_t = \infty$$

# Our goal restated

$$Z(\infty) = \frac{Z(\beta_1)}{Z(\beta_0)} \frac{Z(\beta_2)}{Z(\beta_1)} \dots \frac{Z(\beta_t)}{Z(\beta_{t-1})} Z(0)$$

Cooling schedule:

$$\beta_0 = 0 < \beta_1 < \beta_2 < \dots < \beta_t = \infty$$

How to choose the cooling schedule?

minimize length, while satisfying

$$\frac{V[X_i]}{E[X_i]^2} = O(1) \quad E[X_i] = \frac{Z(\beta_i)}{Z(\beta_{i-1})}$$

# Parameters: **A** and **n**

$$Z(\beta) = \sum_{\mathbf{x} \in \Omega} \exp(-H(\mathbf{x})\beta)$$

$$Z(0) = \mathbf{A}$$

$$H: \Omega \rightarrow \{0, \dots, \mathbf{n}\}$$

$$Z(\beta) = \sum_{\mathbf{k}=0}^{\mathbf{n}} \mathbf{a}_{\mathbf{k}} e^{-\beta \mathbf{k}}$$

$$\mathbf{a}_{\mathbf{k}} = |H^{-1}(\mathbf{k})|$$

# Parameters

$$Z(0) = A \quad H: \Omega \rightarrow \{0, \dots, n\}$$

	$A$	$n$
independent sets	$2^V$	$E$
matchings	$\approx V!$	$V$
perfect matchings	$V!$	$V$
k-colorings	$k^V$	$E$



# Previous cooling schedules

$$Z(0) = A \quad H: \Omega \rightarrow \{0, \dots, n\}$$

$$\beta_0 = 0 < \beta_1 < \beta_2 < \dots < \beta_t = \infty$$

“Safe steps”

$$\beta \rightarrow \beta + 1/n$$

$$\beta \rightarrow \beta (1 + 1/\ln A)$$

$$\ln A \rightarrow \infty$$

(Bezáková, Štefankovič,  
Vigoda, V. Vazirani'06)

Cooling schedules of length

$$O(n \ln A)$$

$$O((\ln n) (\ln A))$$

(Bezáková, Štefankovič,  
Vigoda, V. Vazirani'06)

No better fixed schedule possible

$$Z(0) = A \quad H: \Omega \rightarrow \{0, \dots, n\}$$

A schedule that works for all

$$Z_a(\beta) = \frac{A}{1+a} (1 + a e^{-\beta n})$$

(with  $a \in [0, A-1]$ )

has LENGTH  $\geq \Omega((\ln n)(\ln A))$

# Parameters

$$Z(0) = A \quad H: \Omega \rightarrow \{0, \dots, n\}$$

Our main result:

can get adaptive schedule  
of length  $O^* ( (\ln A)^{1/2} )$

Previously:

non-adaptive schedules  
of length  $\Omega^*( \ln A )$

# Existential part

Lemma:

for every partition function there exists  
a cooling schedule of length  $O^*((\ln A)^{1/2})$

there exists

~~can get adaptive schedule  
of length  $O^*((\ln A)^{1/2})$~~

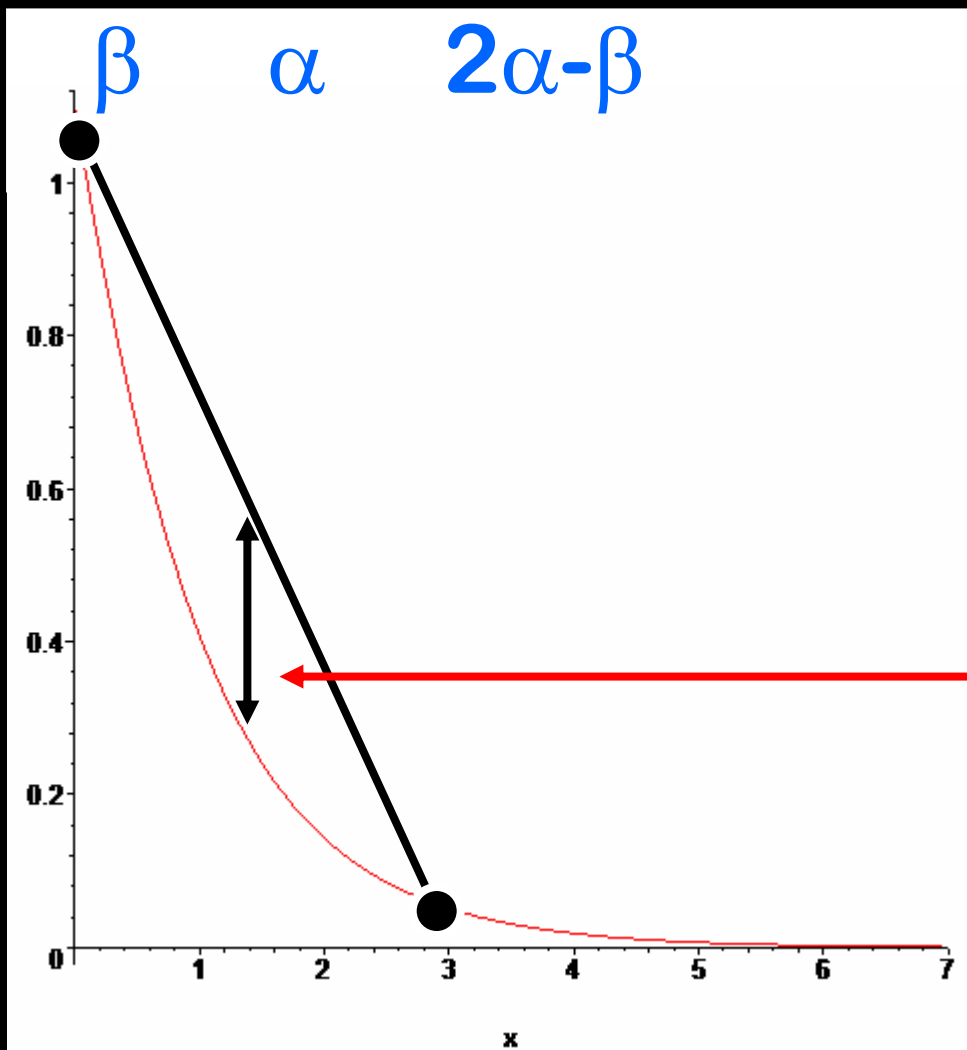
Express **SCV** using **partition function**

(going from  $\beta$  to  $\alpha$ )  $E[X] = \frac{Z(\alpha)}{Z(\beta)}$

$W \sim \mu_\beta \rightarrow X = \exp(H(W)(\beta - \alpha))$

$$\frac{E[X^2]}{E[X]^2} = \frac{Z(2\alpha - \beta) Z(\beta)}{Z(\alpha)^2} \leq C$$

$$\frac{E[X^2]}{E[X]^2} = \frac{Z(2\alpha-\beta) Z(\beta)}{Z(\alpha)^2} \leq C$$



$$f(\gamma) = \ln Z(\gamma)$$

$$\leq C' = (\ln C) / 2$$

$$f(\gamma) = \ln Z(\gamma)$$

**f is decreasing**

**f is convex**

$$f'(0) \geq -n$$

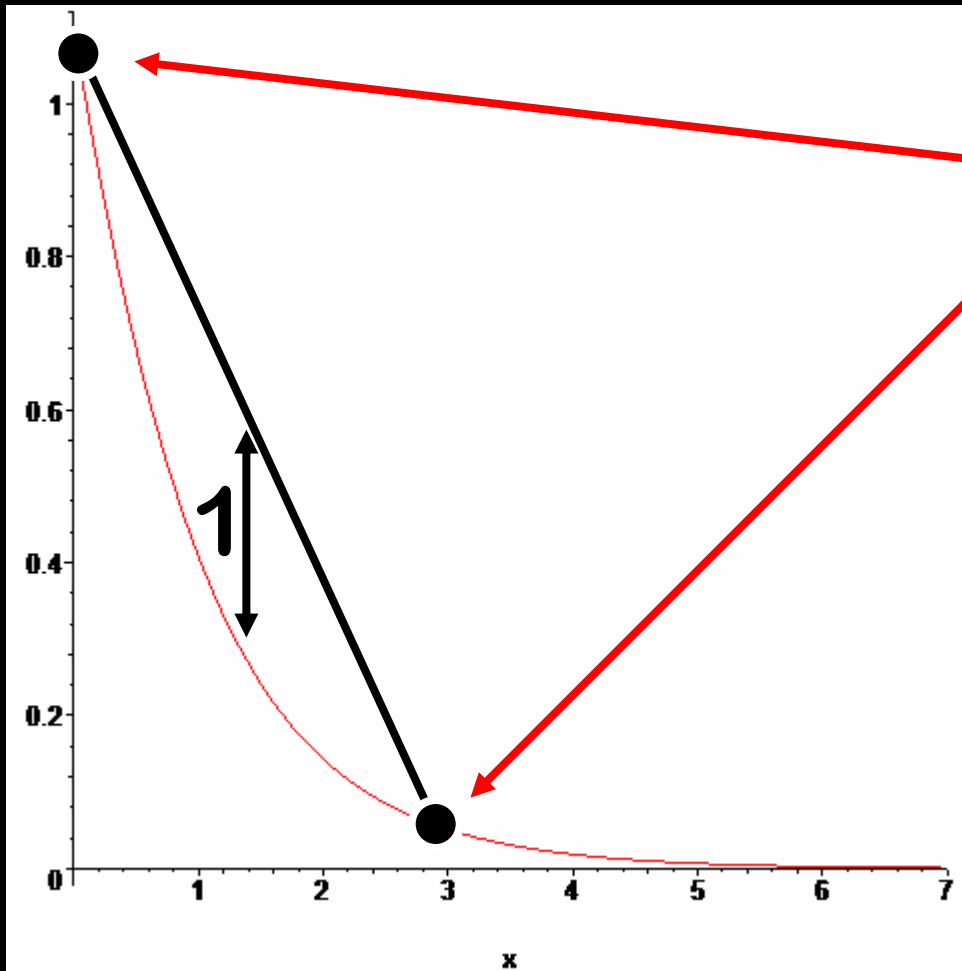
$$f(0) \leq \ln A$$



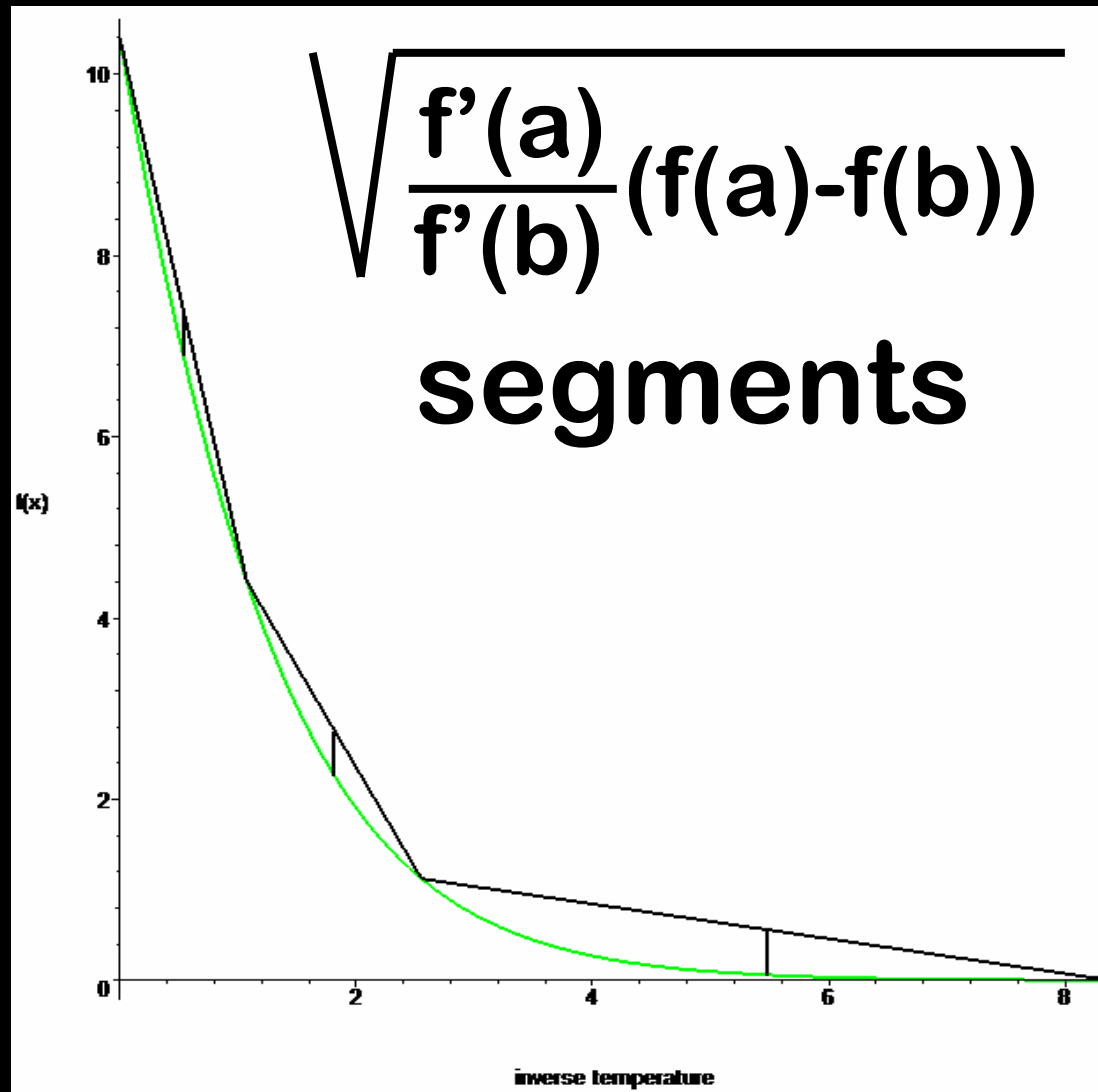
**either f or f'  
changes a lot**

**Let  $K := \Delta f$**

$$\Delta(\ln |f'|) \geq \frac{1}{K}$$

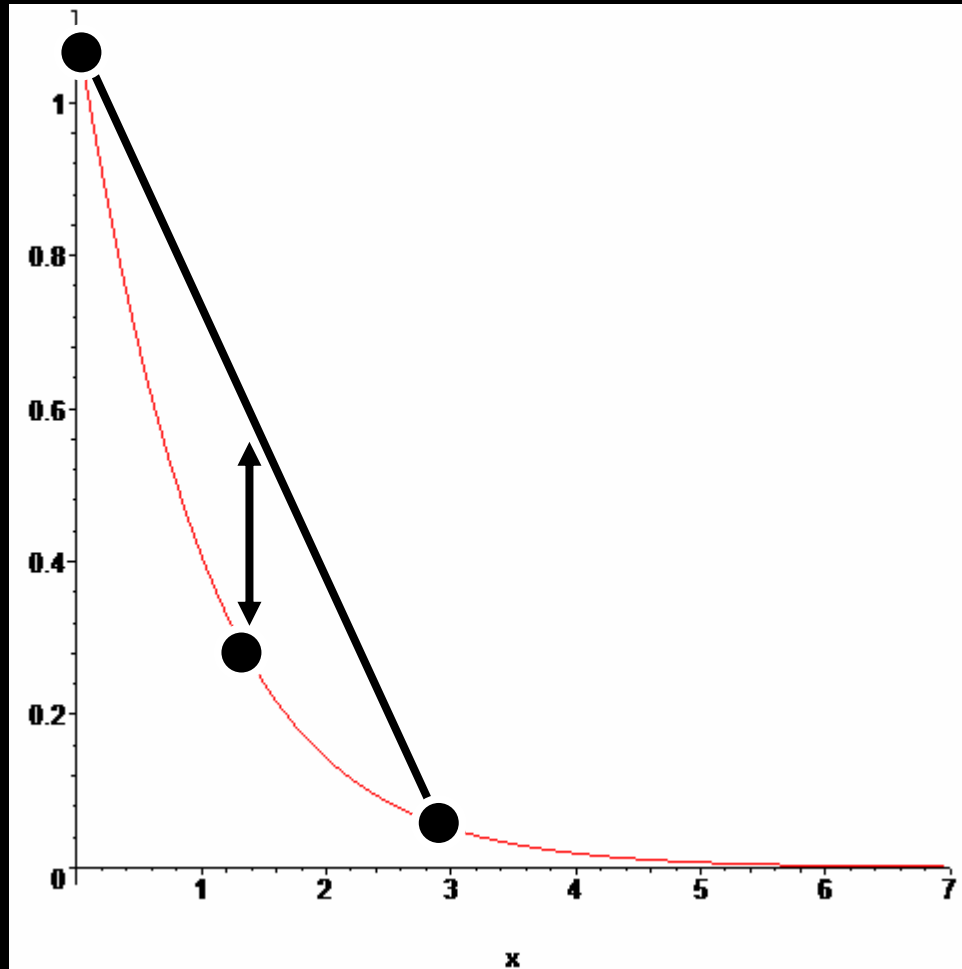


$f:[a,b] \rightarrow \mathbb{R}$ , convex, decreasing  
can be “approximated” using



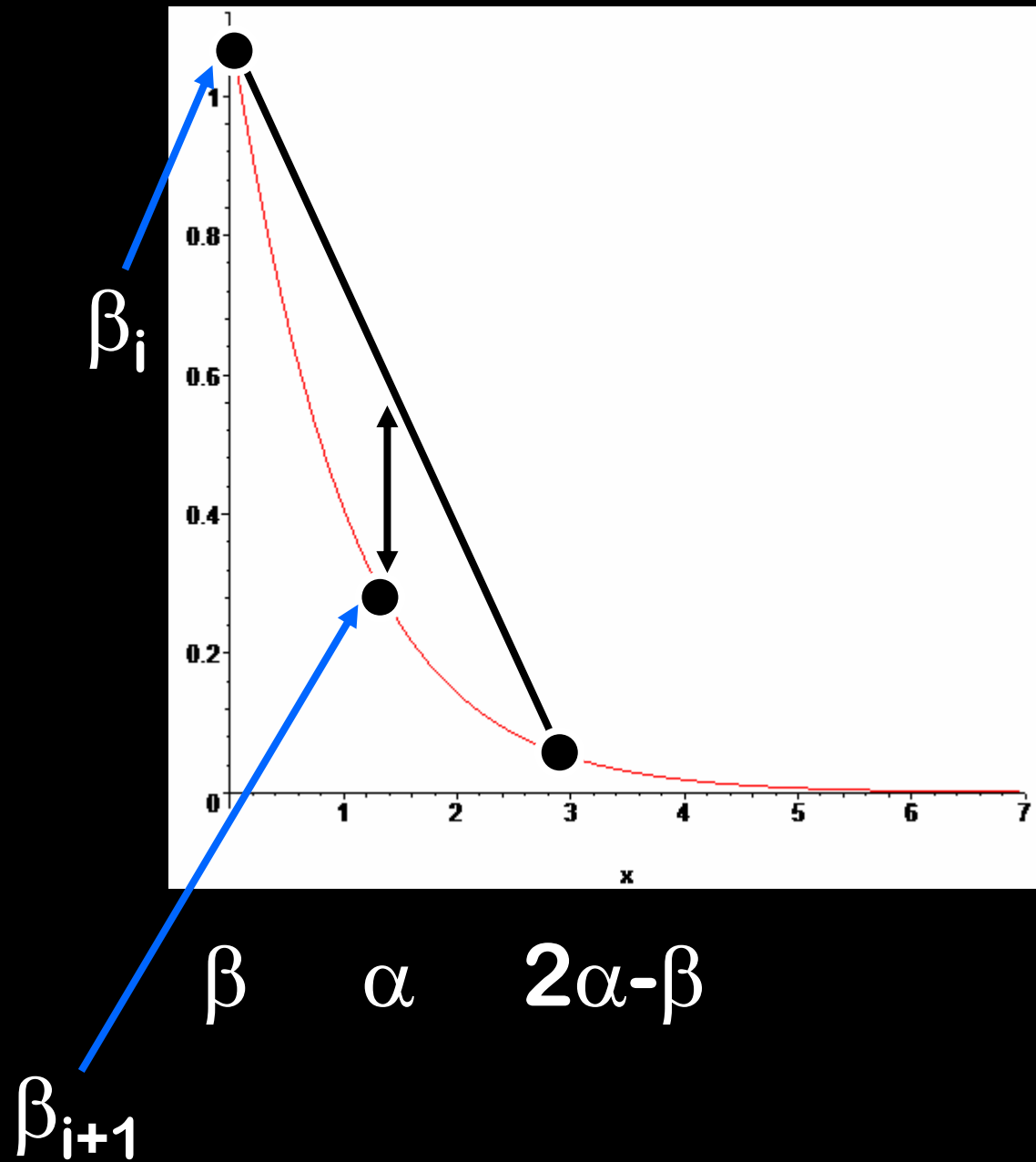


# Technicality: getting to $2\alpha - \beta$

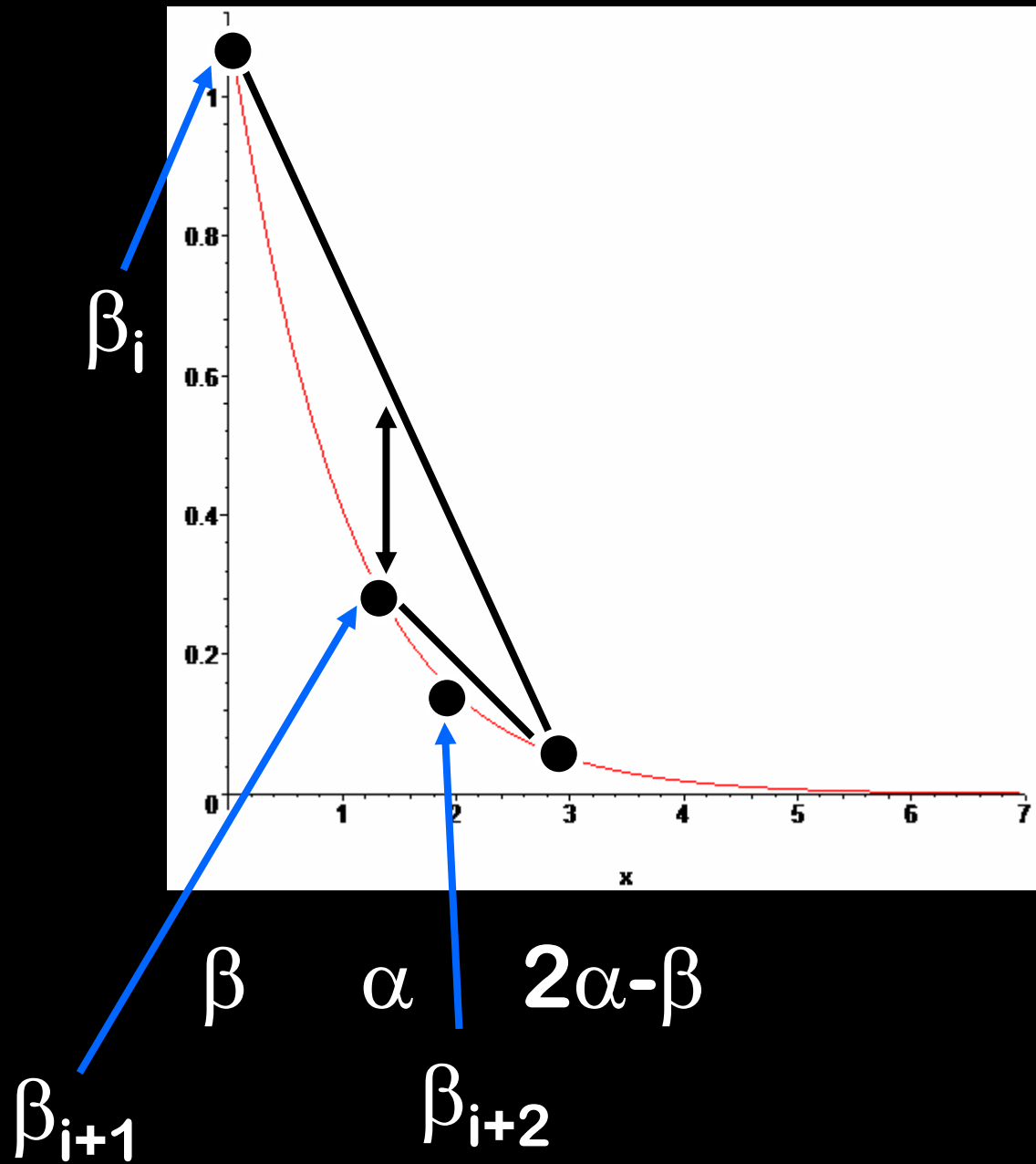


$\beta$     $\alpha$     $2\alpha - \beta$

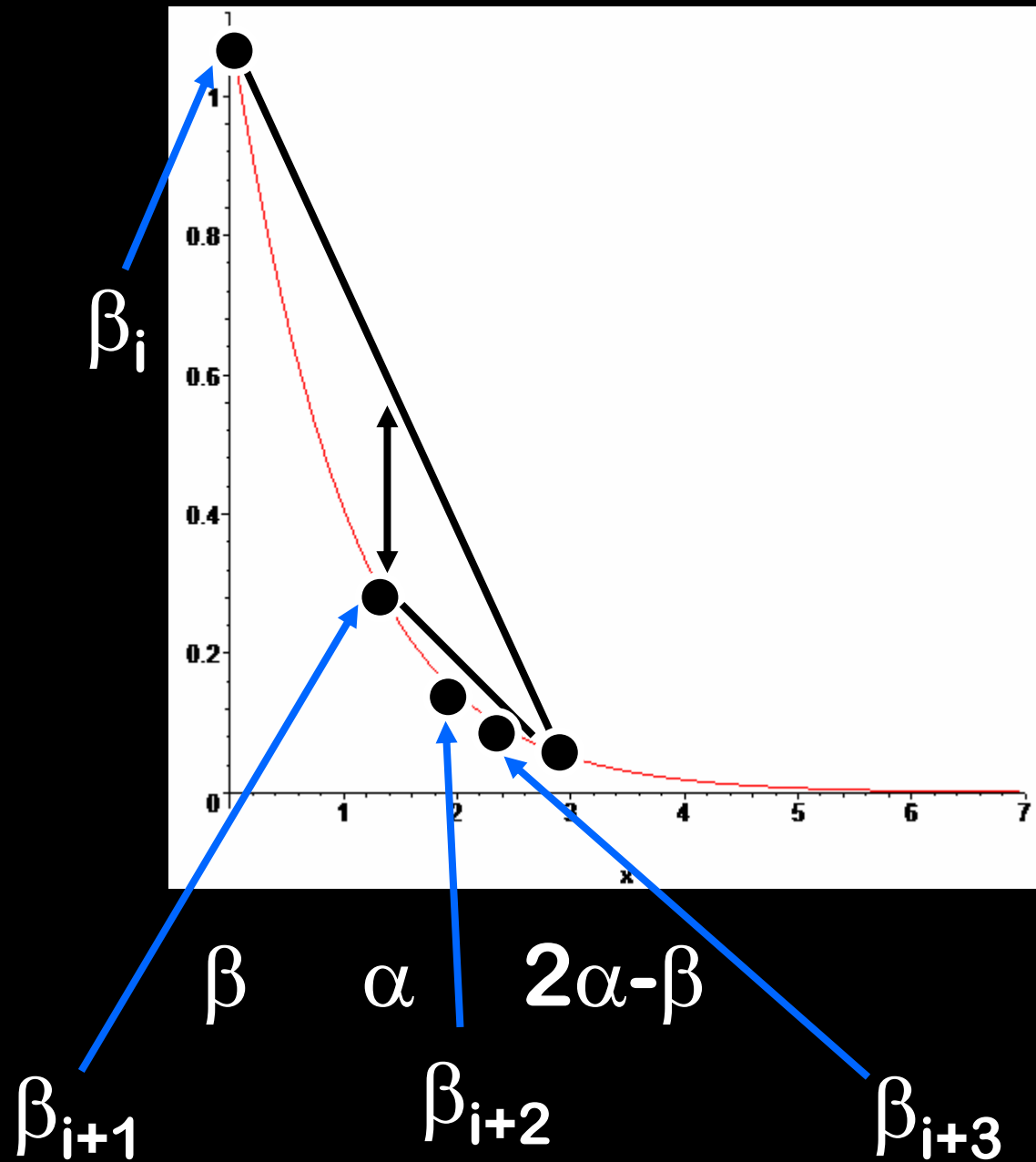
# Technicality: getting to $2\alpha-\beta$



# Technicality: getting to $2\alpha-\beta$



# Technicality: getting to $2\alpha-\beta$



In In A

extra  
steps

# Existential $\rightarrow$ Algorithmic

*there exists*

can get adaptive schedule  
of length  $O^* ( (\ln A)^{1/2} )$



can get adaptive schedule  
of length  $O^* ( (\ln A)^{1/2} )$

# Algorithmic construction

**Our main result:**

using a sampler oracle for  $\mu_\beta$

$$\mu_\beta(\mathbf{x}) = \frac{\exp(-H(\mathbf{x})\beta)}{Z(\beta)}$$

we can construct a cooling schedule of length  
 $\leq 38 (\ln A)^{1/2} (\ln \ln A) (\ln n)$

Total number of oracle calls

$$\leq 10^7 (\ln A) (\ln \ln A + \ln n)^7 \ln(1/\delta)$$

# Algorithmic construction

current inverse temperature  $\beta$

ideally move to  $\alpha$  such that

$$\mathbf{B}_1 \leq \frac{\mathbf{E}[X^2]}{\mathbf{E}[X]^2} \leq \mathbf{B}_2$$

$$\mathbf{E}[X] = \frac{Z(\alpha)}{Z(\beta)}$$

# Algorithmic construction

current inverse temperature  $\beta$

ideally move to  $\alpha$  such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2 \quad E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

$X$  is “easy to estimate”



# Algorithmic construction

current inverse temperature  $\beta$

ideally move to  $\alpha$  such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2 \quad E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

we make progress (assuming  $B_1 > 1$ )

# Algorithmic construction

current inverse temperature  $\beta$

ideally move to  $\alpha$  such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$



need to construct a “feeler” for this

# Algorithmic construction

current inverse temperature  $\beta$

ideally move to  $\alpha$  such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2 \quad E[X] = \frac{Z(\alpha)}{Z(\beta)}$$
$$\uparrow \quad = \quad \frac{Z(\beta)}{Z(\alpha)} \frac{Z(2\beta - \alpha)}{Z(\alpha)}$$

need to construct a “feeler” for this

# Algorithmic construction

current inverse temperature  $\beta$

ideally move to  $\alpha$  such that

$$B_1 \leq \frac{E[X^2]}{E[X]^2} \leq B_2$$

$$= \frac{Z(\beta)}{Z(\alpha)} \frac{Z(2\beta - \alpha)}{Z(\alpha)}$$

bad "feeler"

$$E[X] = \frac{Z(\alpha)}{Z(\beta)}$$

need to construct a "feeler" for this

Rough estimator for  $\frac{Z(\beta)}{Z(\alpha)}$

$$Z(\beta) = \sum_{k=0}^n a_k e^{-\beta k}$$

For  $W \sim \mu_\beta$  we have  $P(H(W)=k) = \frac{a_k e^{-\beta k}}{Z(\beta)}$

# Rough estimator for $\frac{Z(\beta)}{Z(\alpha)}$

If  $H(X)=k$  likely at both  $\alpha, \beta \Rightarrow$  rough estimator

$$Z(\beta) = \sum_{k=0}^n a_k e^{-\beta k}$$

$$\text{For } W \sim \mu_\beta \text{ we have } P(H(W)=k) = \frac{a_k e^{-\beta k}}{Z(\beta)}$$

$$\text{For } U \sim \mu_\alpha \text{ we have } P(H(U)=k) = \frac{a_k e^{-\alpha k}}{Z(\alpha)}$$

Rough estimator for  $\frac{Z(\beta)}{Z(\alpha)}$

For  $W \sim \mu_\beta$  we have  $P(H(W)=k) = \frac{a_k e^{-\beta k}}{Z(\beta)}$

For  $U \sim \mu_\alpha$  we have  $P(H(U)=k) = \frac{a_k e^{-\alpha k}}{Z(\alpha)}$

$$\frac{P(H(U)=k)}{P(H(W)=k)} e^{k(\alpha-\beta)} = \frac{Z(\beta)}{Z(\alpha)}$$

Rough estimator for  $\frac{Z(\beta)}{Z(\alpha)}$

$$Z(\beta) = \sum_{k=0}^n a_k e^{-\beta k}$$

For  $W \sim \mu_\beta$  we have

$$P(H(W) \in [c, d]) = \frac{\sum_{k=c}^d a_k e^{-\beta k}}{Z(\beta)}$$



Rough estimator for  $\frac{Z(\beta)}{Z(\alpha)}$

If  $|\alpha - \beta| \cdot |d - c| \leq 1$  then

$$\frac{1}{e} \frac{Z(\beta)}{Z(\alpha)} \leq \frac{P(H(U) \in [c, d])}{P(H(W) \in [c, d])} e^{c(\alpha - \beta)} \leq e \frac{Z(\beta)}{Z(\alpha)}$$

We also need  $P(H(U) \in [c, d])$   
 $P(H(W) \in [c, d])$

to be large.

Split  $\{0, 1, \dots, n\}$  into  $h \leq 4(\ln n)\sqrt{\ln A}$  intervals

$[0], [1], [2], \dots, [c, c(1+1/\sqrt{\ln A})], \dots$

for any inverse temperature  $\beta$  there exists a interval with  $P(H(W) \in I) \geq 1/8h$

We say that  $I$  is **HEAVY** for  $\beta$

# Algorithm

repeat

find an interval  $I$  which is heavy for  
the current inverse temperature  $\beta$

see how far  $I$  is heavy (until some  $\beta^*$ )

use the interval  $I$  for the feeler  $\frac{Z(\beta)}{Z(\alpha)}$   $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

either

- \* make progress, or
- \* eliminate the interval  $I$

# Algorithm

repeat

find an interval  $I$  which is heavy for  
the current inverse temperature  $\beta$

see how far  $I$  is heavy (until some  $\beta^*$ )

use the interval  $I$  for the feeler  $\frac{Z(\beta)}{Z(\alpha)}$   $\frac{Z(2\beta-\alpha)}{Z(\alpha)}$

either

- \* make progress, or
- \* eliminate the interval  $I$
- \* or make a “long move”

if we have sampler oracles for  $\mu_\beta$   
then we can get adaptive schedule  
of length  $t=O^* ( (\ln A)^{1/2} )$

independent sets  $O^*(n^2)$   
(using Vigoda'01, Dyer-Greenhill'01)

matchings  $O^*(n^2m)$   
(using Jerrum, Sinclair'89)

spin systems:

Ising model  $O^*(n^2)$  for  $\beta < \beta_c$   
(using Marinelli, Olivieri'95)

k-colorings  $O^*(n^2)$  for  $k > 2\Delta$   
(using Jerrum'95)

**input** : A black-box sampler for  $X \sim \mu_\beta$  for any  $\beta \geq 0$ , starting inverse temperature  $\beta_0$ .

**output**: A cooling schedule for  $Z$ .

Bad  $\leftarrow \emptyset$

print  $\beta_0$

**if**  $\beta_0 < \ln A$  **then**

$I \leftarrow \text{FIND-HEAVY}(\beta_0, \text{Bad})$

$w \leftarrow$  the width of  $I$

$L \leftarrow \min\{\beta_0 + 1/w, \ln A\};$  (where  $1/0 = \infty$ )

$\beta^* \leftarrow$  binary search on  $\beta^* \in [\beta_0, L]$  with precision  $1/(2n)$ , using predicate

$\text{IS-HEAVY}(\beta^*, I)$

$\beta \leftarrow$  binary search on  $\beta \in [\beta_0, (\beta^* + \beta_0)/2]$  with precision  $1/(4n)$ ,  
    using predicate  $\text{EST}(I, \beta_0, \beta) \cdot \text{EST}(I, 2\beta - \beta_0, \beta) \leq 2000$

**if**  $\beta < (\beta^* + \beta_0)/2$  **then**

$\text{PRINT-COOLING-SCHEDULE}(\beta)$  (optimal move)

**else**

**if**  $\beta = L$  **then**

$\text{PRINT-COOLING-SCHEDULE}(\beta)$  (long move)

**else**

$\gamma \leftarrow (\beta^* - \beta_0)/2$

            print  $\beta_0 + \gamma, \beta_0 + (3/2)\gamma, \beta_0 + (7/4)\gamma, \dots, \beta_0 + (2 - 2^{-\lceil \ln \ln A \rceil})\gamma$

            Bad  $\leftarrow \text{Bad} \cup I$

$\text{PRINT-COOLING-SCHEDULE}(\beta^*)$  (interval move)

**end**

**end**

**else**

    print  $\infty$

**end**