Fast approximation algorithms for partition functions of Gibbs distributions

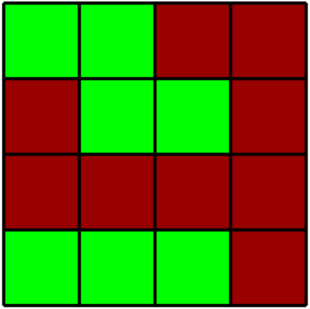
Mark Huber
Department of Mathematical Sciences
Claremont McKenna College

27 June, 2012

The Problem

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



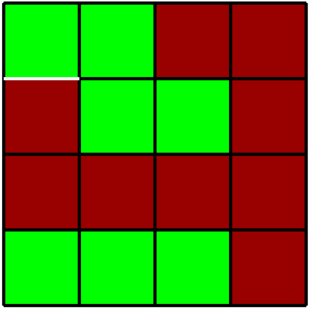
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



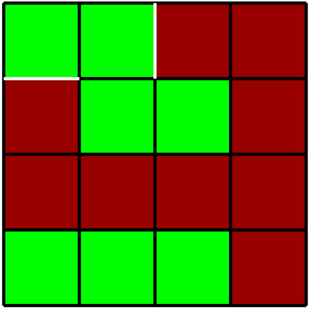
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



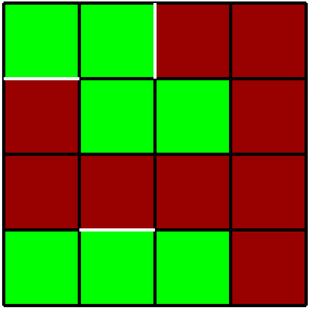
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



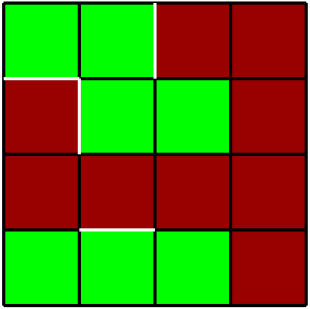
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



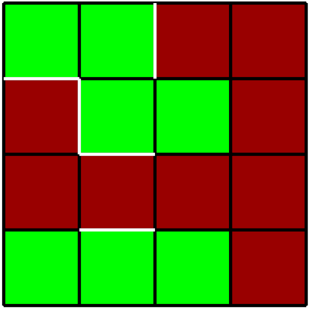
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



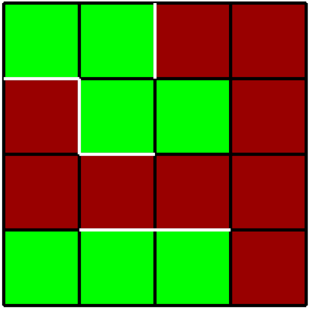
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



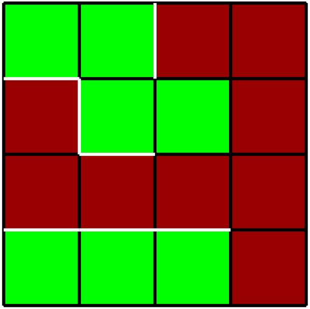
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



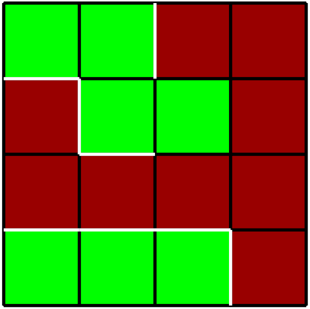
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



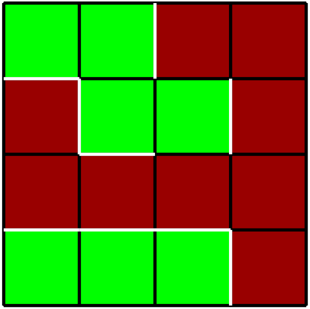
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



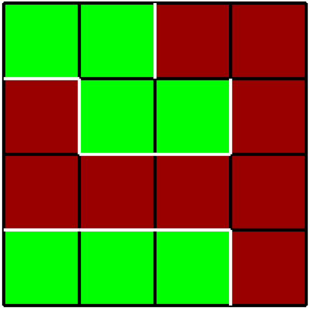
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



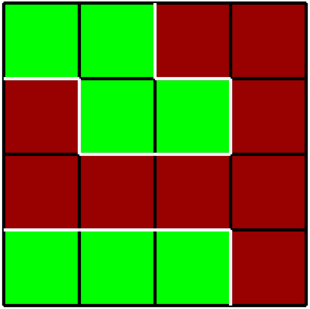
$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Autonormal (Ising) model

Besag [1974] modeled soil plots as good (green) or bad (red)



$d(x) = 11$
(number of adj nodes that disagree)

Model:

$$\mathbb{P}(X = x) = \frac{\exp(-\beta d(x))}{Z(\beta)}$$

Data

$d(x) = 11$

What is best estimate of β?

Posterior

$$f_{\text{post},\beta}(b) \propto f_{\text{prior},\beta}(b) \exp(-11b)/Z(b)$$

Finding $Z(\cdot)$

Need $Z(\cdot)$ to have posterior density

- Brute force computation long (curse of dimensionality)
- Brute force for Ising model takes $2^{\#\text{ of vertices}}$ time
- In denominator, so need relative (not additive) accuracy

Also need $Z(\cdot)$ for computing Bayes factors

- Again, for Bayes factors relative accuracy essential

$Z(\beta)$: The Partition Function

Partition function of Gibbs distribution

Ingredients:

State space $\Omega \subseteq \mathbb{R}^n$
Nonnegative function $H(x)$
Parameter β

Then the goal is to approximate:

$$Z(\beta) = \int_{x \in \Omega} \exp(-\beta H(x))\, dx$$

Discrete version

Ingredients:

State space $\Omega \subseteq \{1, 2, \ldots, c\}^n$
Nonnegative function $H(x)$
Parameter β

Then the goal is to approximate:

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

Terminology

Definition (Gibbs distribution)

X has a (discrete) *Gibbs distribution* π_β if for all $x \in \Omega$,

$$\mathbb{P}(X = x) = \frac{1}{Z(\beta)} \exp(-\beta H(x)).$$

Definition (Partition function)

The *partition function* of a Gibbs distribution is

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

Today's result

An approximation algorithm where

- Given $\epsilon > 0$
- Given the ability to sample from π'_β for $\beta' \in [0, \beta]$
- Outputs $\hat{Z}(\beta)$ so that

$$\mathbb{P}\left(\frac{1}{1+\epsilon} \leq \frac{\hat{Z}(\beta)}{Z(\beta)} \leq 1+\epsilon\right) \geq 3/4$$

- Let $q = \ln(Z(\beta))$. Method requires a number of samples

$$O(q \ln(M)(\ln(q) + \epsilon^{-2})), \quad M = \max_x H(x)$$

Is that good?

Why is problem hard?

Typically $Z(\beta)$ is exponential in n, the input size of problem

Many methods for lowering variance

- Multistage Sampling [Valleau Card 1972]
- Bridge Sampling [Meng Wong 1996]
- Nested Sampling [Skilling 2006]

The above are not approximation algorithms

No *guarantee* on quality of estimate obtained
(But they could be faster in practice)

Approximation Algorithms

Let $q = \ln(Z(0)/Z(\beta))$...

Self-reducibility [Jerrum, Valiant, Vazirani 1986]
$O[q^2\epsilon^{-2}]$ time under best conditions

TPA [H., Schott 2010]
$O[q^2\epsilon^{-2}]$, simpler, constant ~ 20

SVV [Štefankovič, Vempala, Vigoda 2009]
$O[q(\ln(q) + \ln(M))^5\epsilon^{-2}]$ (constant $\sim 10^{10}$)

Paired Product Approximation Algorithm [H. 2012]
$O[q\ln(M)(\epsilon^{-2} + \ln(q))]$ (constant < 50)

The algorithm

New algorithm

Importance sampling

Multistage sampling

TPA

Paired Estimator

Importance Sampling

Usually $Z(0)$ easy to find, so need to estimate

$$\frac{Z(\beta)}{Z(0)}$$

For $X \sim \pi_0$, set $W = \exp(-\beta H(X))$

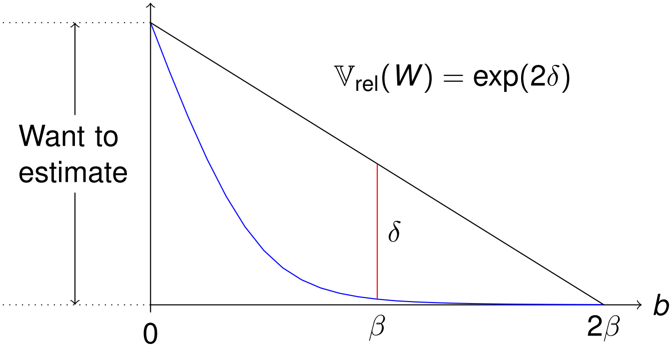
$$\mathbb{E}[W] = \frac{Z(\beta)}{Z(0)}$$

Relative variance (square of coefficient of variation)

$$\mathbb{V}_{\text{rel}}(W) = \frac{\mathbb{V}(W)}{\mathbb{E}[W]^2} = \frac{Z(2\beta)Z(0)}{Z(\beta)} - 1$$

The relative variance picture

Take the log: $z(b) = \ln(Z(b))$, $\ln\left(\frac{Z(\beta)}{Z(0)}\right) = z(\beta) - z(0)$.



Only care about 0 to β, harder to sample at 2β!

Paired Estimator

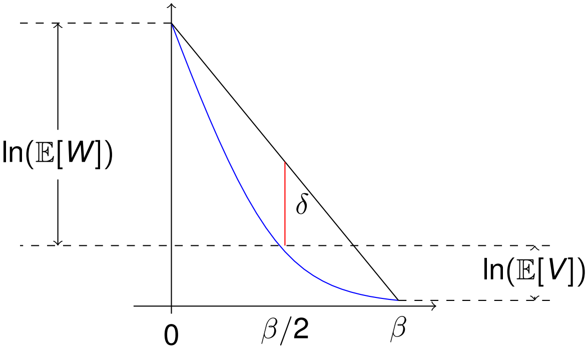
Solution to 2β problem: use two estimators

$$X \sim \pi_0, \quad Y \sim \pi_\beta$$

$$W := \exp(-(\beta/2)H(X)), \quad V := \exp((\beta/2)H(Y))$$

$$W \text{ estimates } \frac{Z(\beta/2)}{Z(0)}, \quad V \text{ estimates } \frac{Z(\beta/2)}{Z(\beta)}$$

The relative variance for paired estimator

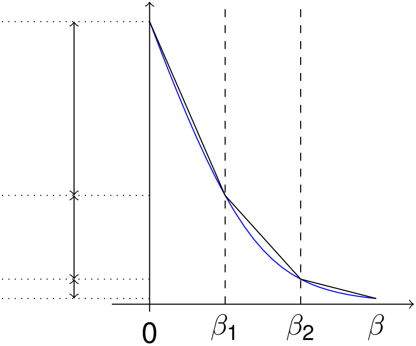


$$\mathbb{V}_{\text{rel}}(W) = \mathbb{V}_{\text{rel}}(V) = \exp(2\delta)$$

But what to do about δ being too large?

Multistage sampling

Breaking the interval into pieces [Valleau and Card 1972]



$$\frac{Z(\beta)}{Z(0)} = \frac{Z(\beta_1)}{Z(0)} \frac{Z(\beta_2)}{Z(\beta_1)} \frac{Z(\beta)}{Z(\beta_2)}$$

Relative variance of product estimator

Relative variance of $\prod W_i$ (and $\prod V_i$)

$$-1 + \exp\left(\sum_i \delta_i\right)$$

New question:

How do we make $\sum_i \delta_i$ small without using too many intervals?

Making $\sum \delta_i$ small

Fact (Lemma 3.3 of H. 2012)

$\sum \delta_i$ is small if the vertical drop in z over each interval is small.

Call such a set of intervals well balanced

- To get well balanced intervals use TPA [H. Schott 2010]
- Time to get/number of such intervals $O(q \ln(M) \ln(q))$
- Once have intervals, $O(q \ln(M) \epsilon^{-2})$ time to approximate $\prod W_i$ and $\prod V_i$

Well balanced schedules also good for Markov chains

- Necessary for fast parallel tempering (see [Woodard, Schmidler, H. 2009])

The Algorithm

1. Obtain an initial estimate of q with TPA
2. Obtain a well balanced set of intervals with TPA
3. Use the well balanced schedule to get $\bar{W}$, $\bar{V}$
4. Final estimate $\bar{W}/\bar{V}$

Conclusions

Most important thing
Previous practical methods quadratic in q (roughly the dimension), new method nearly linear

Uses some old ideas
Importance Sampling, Multistage Sampling, Product Estimator

Add some new ideas
Paired Product Estimator, Well-balanced schedule through TPA

The resulting approximation algorithm
Simple to implement, guaranteed quality of estimate

References

M. Huber
The Paired Product Estimator approach to approximating Gibbs partition functions
arXiv:1206.2689, 2012

M. Huber and S. Schott,
Using TPA for Bayesian inference
Bayesian Statistics 9, 257–282, 2010

J.P. Valleau and D.N. Card,
Monte Carlo Estimation of the Free Energy by Multistage Sampling
J. Chem. Phys. **57**, 5457–5462, 1972