Speeding up the product estimator using random temperatures

Mark Huber[1] and Sarah Schott[2]

[1]Department of Mathematics and Computer Science, Claremont McKenna College

[2]Department of Mathematics, Duke University

5 Aug, 2009

The basic Monte Carlo framework

Given ingredients....

- Large set B containing smaller set B'
- Ability to generate samples uniformly from B

The goal

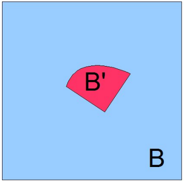
- Approximate $A = \mu(B)/\mu(B')$
- Create (ϵ, δ) randomized approximation algorithm:

$$\mathbb{P}((1+\epsilon)^{-1} \leq A/\hat{A} \leq 1+\epsilon) \geq 1-\delta.$$

Basic algorithm

Simple Monte Carlo

- Draw $X_1, \ldots, X_N$ iid from B
- Let $\hat{A}$ be $\#\{i : X_i \in B'\}/N$



Problem

- Want $\mathrm{SD}(\hat{A})/\mathbb{E}[\hat{A}] \leq \epsilon$
- Need to take $\epsilon^{-2} A = \epsilon^{-2} \mu(B)/\mu(B')$ samples to do this!

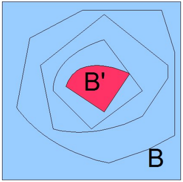
Outline

1. The basic product estimator
2. “Tootsie Pop Algorithm”
3. Accelerating the schedule

The product estimator approach

Use of telescoping products for estimate

- Algorithm: Jerrum, Valiant, Vazirani (1986)
- Name "product estimator" Fishman (1996)
- Idea: insert sequence of regions between B and B'



- Index sequence of regions by "cooling schedule" $\beta_0, \beta_1, \ldots, \beta_k$

More on the product estimator

Product Estimator

- Estimate $Z(\beta_i)/Z(\beta_{i+1})$ for all i
- Let
$$\widehat{\frac{Z(\beta_0)}{Z(\beta_k)}} = \widehat{\frac{Z(\beta_0)}{Z(\beta_1)}}\widehat{\frac{Z(\beta_1)}{Z(\beta_2)}}\cdots\widehat{\frac{Z(\beta_{k-1})}{Z(\beta_k)}}.$$

Running time

- If $Z(\beta_i)/Z(\beta_{i+1}) \leq c$ for all i...
- ...then $k \geq \ln A/\ln c$.
- Needs $28ck^2\epsilon^{-2}\ln(\delta^{-1})$ samples...
- ...at least
$$28\frac{c}{(\ln c)^2}(\ln A)^2\epsilon^{-2}\ln(\delta^{-1})$$

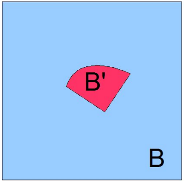
Outline

1. The basic product estimator
2. “Tootsie Pop Algorithm”
3. Accelerating the schedule

The classic question of the Tootsie Pop

Tootsie Pop

- Similar to $B' \subseteq B$
- A hard candy lollipop filled with a chocolate chewy center



- "How many licks does it take to get to the center of a Tootsie pop?"
- Tootsie Pop Algorithm = TPA

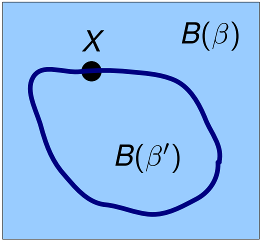
Outline of TPA

Requirement

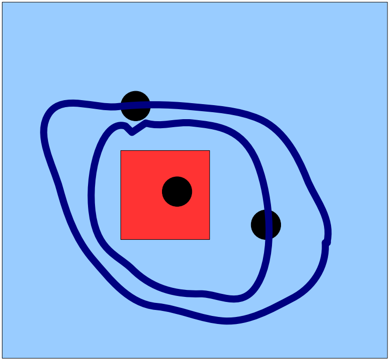
- Say temperature β indexes set $B(\beta)$
- Require: $\beta > \beta' \Rightarrow B(\beta) \subset B(\beta')$

Approach

- Draw $X \sim \text{Unif}(B(\beta))$
- Let $\beta' = \sup\{b : X \in B(b)\}$



Repeat several times



After 3 moves, went from B to B'

Chopping the partition function in half

Theorem

Draw $X \sim \text{Unif}(B(\beta))$*, let* $\beta' = \sup\{b : X \in B(b)\}$*. Then*

$$\frac{Z(\beta')}{Z(\beta)} \sim \text{Unif}([0,1]).$$

Notes

- Ex: $\mathbb{P}(Z(\beta')/Z(\beta)) \leq 0.3$ is $\mathbb{P}(X \in B(b))$ where b satisfies $Z(b)/Z(\beta) = 0.3$
- On average, cuts partition function in half

In log space

So far:

$$Z(\beta') = Z(\beta)U, \ U \sim \text{Unif}([0,1])$$

Working in log space:

$$f(\beta) := \log(Z(\beta))$$

Products change to additions:

$$f(\beta') = f(\beta) + E, \ E \sim \text{Exp}(1).$$

Poisson process

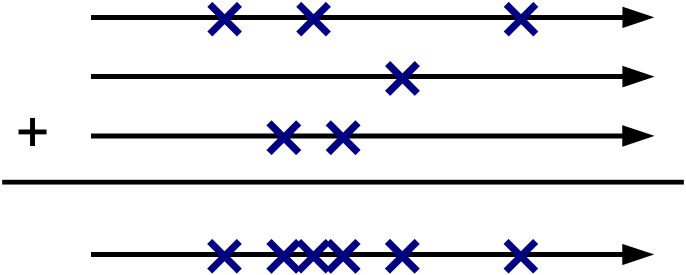
Result

- Obtain sequence of temps: $\beta_0, \beta_1, \ldots$
- $f(\beta_i)$ forms Poisson process
- For $B' \subset B$, say $B(\beta_{B'}) = B'$
- Let $N = \max\{i : f(\beta_i) > f(\beta_{B'})\}$
- Then $N \sim \text{Pois}(\ln A)$

Advantages to Poisson process

Repeat k times

- Originally has Poisson process rate 1
- Restart k times from B, run until reach B'
- Result: Poisson process rate k



Determining number of runs

How to set k?

- Let $N^k \sim \text{Pois}(k \ln A)$
- Then $\mathbb{E}[N^k/k] = \ln A$...
- ...and $\text{SD}[N^k/k] = \sqrt{(\ln A)/k}$
- Want $\text{SD}[N^k/k] \leq \epsilon$
- Set $k = \epsilon^{-2} \ln A$
- Total number of samples: $\epsilon^{-2}(\ln A)^2$

Determining number of samples

Two phases

- First phase set $k = \ln(2\delta^{-1})$
- Returns estimate of $\ln A$ (call estimate $\hat{\ell}$)
- With probability $1 - \delta/2$, $\hat{\ell} > \ln A - \sqrt{\ln A}$
- Second phase use $k = \epsilon^{-2}(\hat{\ell} + \sqrt{\hat{\ell}} + 1)$

All temperatures

Compensated Poisson process

- $A_t = N_t^k - kt$ is a martingale
- A_t close to 0 means N_t close to $f(\cdot)$
- Has right continuous sample paths
- So use approach of Doob's maximal inequality:
- Bound $\mathbb{P}(\sup\{A_t : t \in [\beta_B, \beta_{B'}]\} \leq \epsilon)$

Using Chernoff Bounds to bound entire path

Exponentiate and use Strong Markov Property

- For all $\alpha > 0$, $\exp(\alpha A_t)$ is a nonnegative submartingale

$$
\begin{aligned}
T &:= \inf\{t : A_t > \epsilon\} \\
\exp(\alpha A_t) &\geq \exp(\alpha\epsilon)\mathbb{P}(T \leq t)
\end{aligned}
$$

- Set $\alpha = \ln A/\epsilon$...
- Resulting bound:

$$\mathbb{P}(T \leq t) \leq \exp(-(1/2)k\epsilon^2(1 - 2\epsilon)/\ln A)$$

- Set $k = 2(\ln A)\epsilon^{-2} \ln \delta^{-1}$

Outline

1. The basic product estimator
2. "Tootsie Pop Algorithm"
3. Accelerating the schedule

Faster cooling schedule

Beating the $(\ln A)^2$ bound

- Suppose $Z(\beta) = \sum_{i=0}^{m} a_i \exp(-\beta i)$
- Equivalent: $w(x;\beta) = \exp(-\beta H(x))$, call $H(x)$ the *Hamiltonian*
- Then can do better than Bernoulli random variables with importance sampling
- Estimate $Z(\beta)/Z(\beta')$ by drawing $X \sim w(\cdot;\beta)$, using $\exp(-\beta H(X))/\exp(-\beta' H(X))$

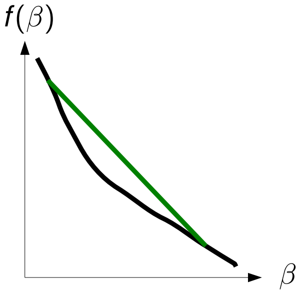
Drawback

- Not easy to find good cooling schedule for problem
- SVV uses 10^8 samples to get β_i's.

Brief description of SVV

Takes advantage of form of $f(\beta) = \ln Z(\beta)$

- Convex function
- In many places, almost concave



Where our work fits in

Hard to find those places in f

- Relatively complex procedure
- SVV uses 10^8 samples as preprocessing step

Our goal:

- Before used random changes in temp
- Are there random moves for temperature...
- ...to find these places in f automatically?

Summary: accomplishments and future goals

Fixed temp cooling schedule at least

$$28\frac{c}{(\ln c)^2}(\ln A)^2\epsilon^{-2}\ln(\delta^{-1})$$

- Returns single estimate

The TPA:

$$[(\ln A)(\ln A+\sqrt{\ln A})\epsilon^{-2}+\ln A]\ln(2\delta^{-1})$$

- Estimate good for all β

Next step:

- $\Theta(\ln A)$ for problems with Hamiltonian

References

M. L. Huber and S. Schott
A product estimator for all temperatures
preprint, 2009

M. Jerrum, L. Valiant, and V. Vazirani
Random generation of combinatorial structures from a uniform distribution.
Theoret. Comput. Sci., **43**, 169–188, 1986

D. Štefankovič, S. Vempala, and E. Vigoda.
Adaptive simulated annealing: A near-optimal connection between sampling and counting.
arXiv:cs/0612058v1[cs.DS], 2006