Using TPA for Monte Carlo integration

Mark Huber[1] and Sarah Schott[2]

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

[2]Department of Mathematics, Duke University

20 Aug, 2010

Supported by: SAMSI Population Monte Carlo group and NSF CAREER grant DMS-05-48153

Monte Carlo Integration

Basic problem of Monte Carlo

Estimate

$$Z = \int_{\Omega} f(\vec{x})\, d\mathbb{R}^d$$

The usual strategy

Find random variable X such that $\mathbb{E}[X] = Z$

- Draw $X_1, \ldots, X_n \overset{\text{iid}}{\sim} X$
- Let $\hat{Z} = \frac{1}{n} \sum_{i=1}^{n} X$

The problem

- $$\mathrm{SD}(\hat{Z}) = \frac{\mathrm{SD}(X)}{\sqrt{n}}$$
- Variance of X is usually unknown

Many ways to create random variable

Some Examples

- Importance Sampling
- Bridge Sampling
- Path Sampling
- Harmonic Mean Estimator
- Nested Sampling

Several possibilities

- Variance could be infinite (IS, HME)
- At best, have to estimate variance

Today

Avoiding the variance problem

- Existing methods: acceptance/rejection, product estimator
- New method: TPA

$$\text{Replaces SD}(X) \text{ with } 2(\ln Z)$$

Acceptance/rejection and the Product Estimator

Abstract problem: find the measure of a set

The setup

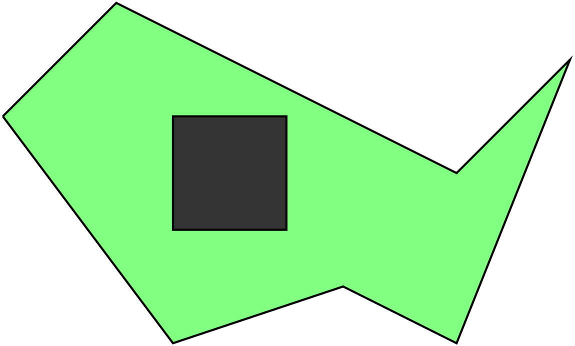
$$\mu(A) = \int_A f(\vec{x}) \, d\mathbb{R}^d$$

Find

$$Z = \mu(\Omega)$$

Abstract problem: find the measure of a set

Find easy set inside hard set

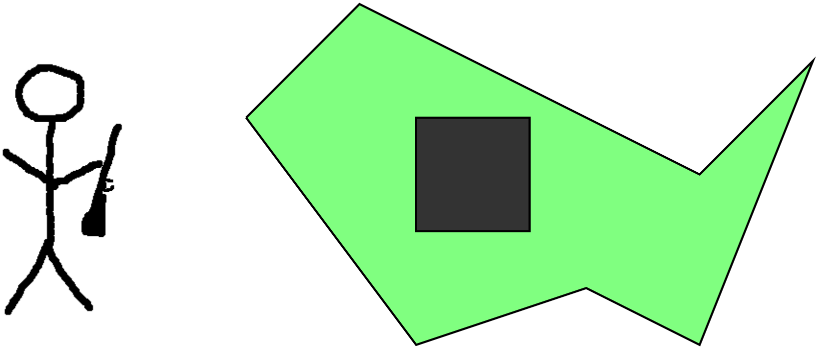


Green area = B

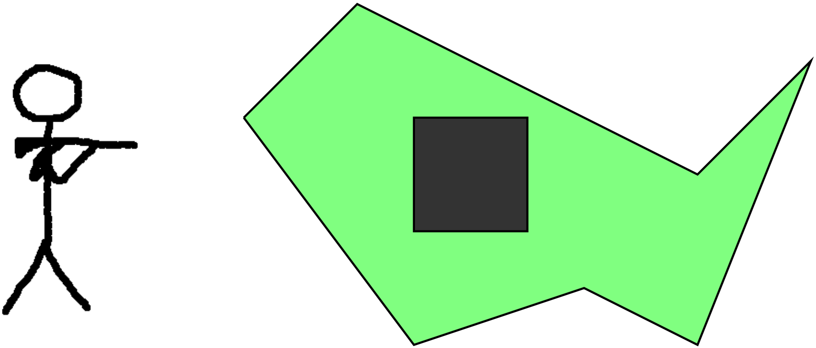
Black area = B'

$$\mu(B) = \mu(B') \frac{\mu(B)}{\mu(B')}$$

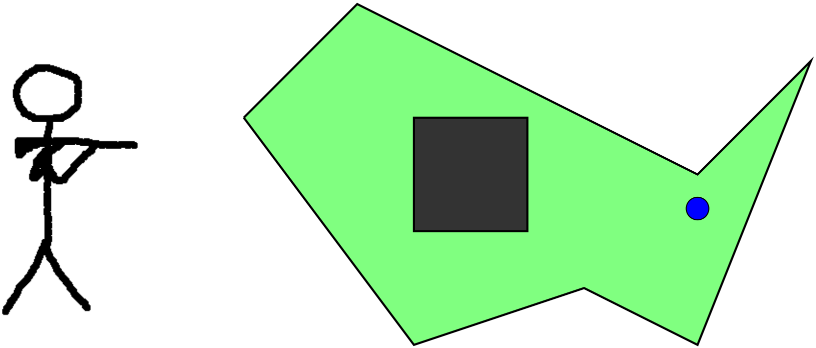
Acceptance/Rejection a.k.a "Shoot at it randomly"



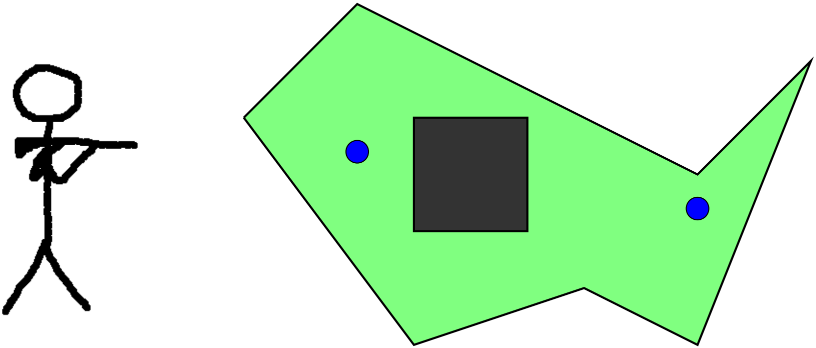
Acceptance/Rejection a.k.a "Shoot at it randomly"



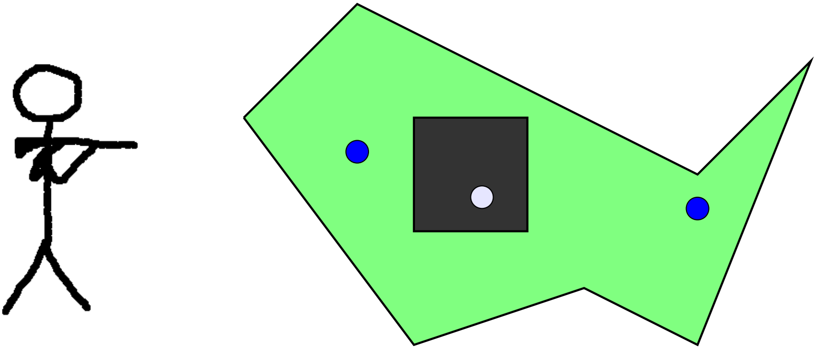
Acceptance/Rejection a.k.a "Shoot at it randomly"



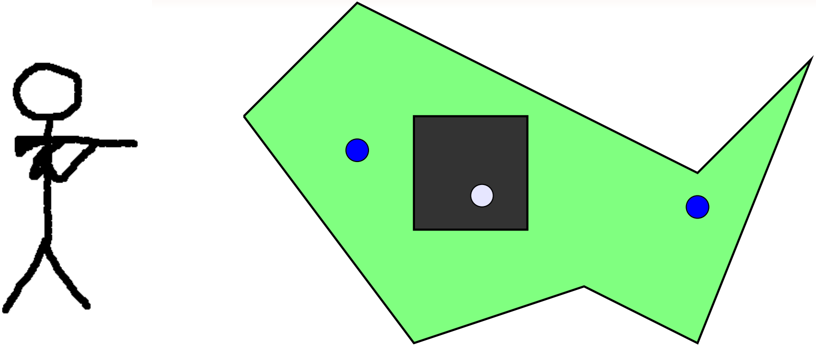
Acceptance/Rejection a.k.a “Shoot at it randomly”



Acceptance/Rejection a.k.a “Shoot at it randomly”



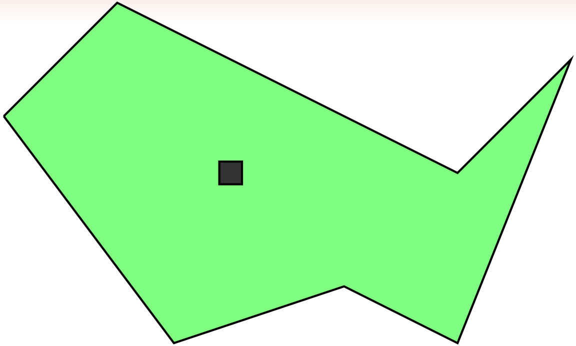
Acceptance/Rejection a.k.a "Shoot at it randomly"



Best estimate:

$$\hat{\mu}(B) = 3\mu(B'), \quad B' = \text{black rectangle inside region}$$

But nobody uses acceptance rejection

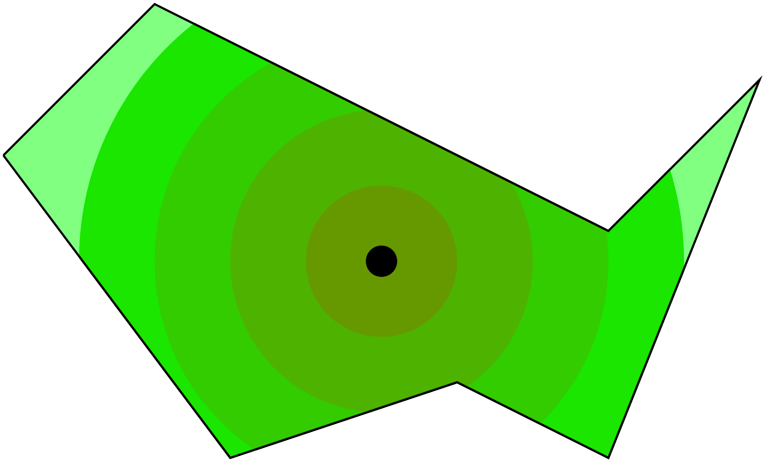


Because center square usually tiny:

$$\frac{\mu(B')}{\mu(B)} = e^{-r}$$

Need about e^r samples

Solution: Series of levels from outer edge to center



Called the product estimator

How it works (Jerrum, Valiant, Vazirani 1986)

- Fixed sequence of nested sets (cooling schedule)

$$B' = B_k \subseteq B_{k-1} \subseteq \cdots \subseteq B_0 = B$$

- For each i, estimate $p_i = \mu(B_{i+1})/\mu(B_i)$
- Final estimate product of individual estimates

$$\hat{Z} = \hat{p}_{k-1}\hat{p}_{k-2}\cdots\hat{p}_1$$

How many samples need to be taken?

Acceptance/rejection

$$O(e^r)$$

Product estimator (if for all i, $p_i \in [1/3, 2/3]$)

$$O(r^2)$$

So why does no one use the product estimator?

Need $p_i \in [1/3, 2/3]$ **for all** i

- Called a *well-balanced* schedule
- Also has applications for designing Markov chains
- In general, these schedules hard to find

The New Algorithm: TPA

Properties of TPA

- Finds optimal cooling schedule automatically
- Like product estimator, basic TPA uses $O(r^2)$ samples
- Adaptive Monte Carlo method

The Tootsie Pop Algorithm

What is a Tootsie Pop?

- Hard candy lollipops with a tootsie roll (chewy chocolate) at the center

In 1970, Mr. Owl was asked the question:

- How many licks does it take to get to the center of a Tootsie Pop?

Key Idea:

Acceptance/Rejection

- Try to get to center in one step
- If fail, start over

Tootsie Pop Algorithm

- Try to get to center
- If fail, start from current position

List of ingredients of TPA

(a) A measure space $(\Omega, \mathcal{F}, \mu)$

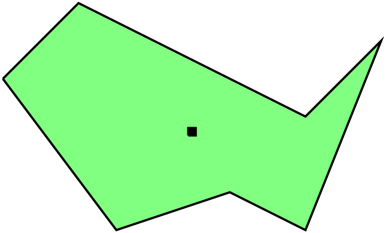
(b) Two measurable sets: the *center* B' and the *shell* B with $B' \subset B$

(c) A family of sets $\{A(\beta)\}$ where

1. $\beta' < \beta$ implies $A(\beta') \subseteq A(\beta)$,
2. $\mu(A(\beta))$ is continuous in β

(d) Two special values β_B and $\beta_{B'}$ with $A(\beta_B) = B$ and $A(\beta_{B'}) = B'$.

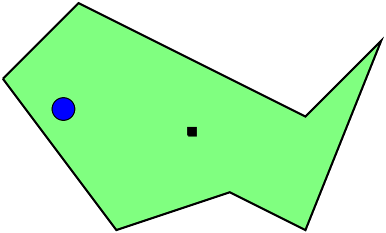
Idea behind TPA



Step	β
0	∞
1	
2	
3	

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

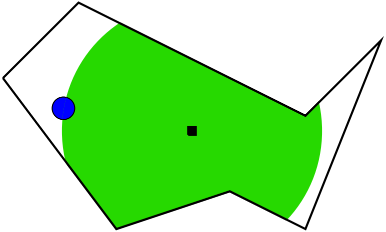
Idea behind TPA



Step	β
0	∞
1	1.72
2	
3	

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

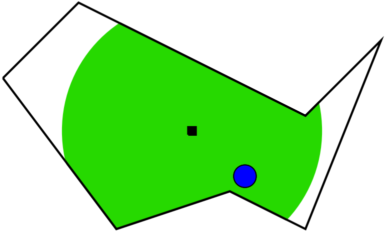
Idea behind TPA



Step	β
0	∞
1	1.72
2	
3	

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

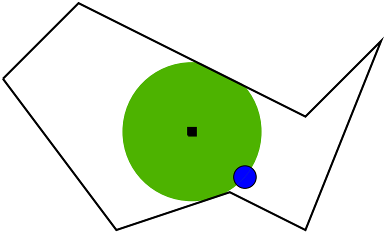
Idea behind TPA



Step	β
0	∞
1	1.72
2	0.92
3	

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

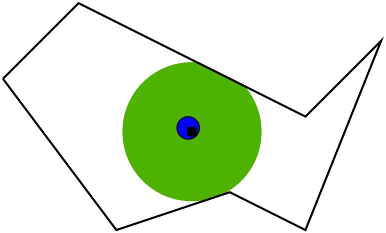
Idea behind TPA



Step	β
0	∞
1	1.72
2	0.92
3	

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

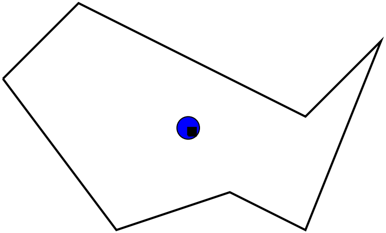
Idea behind TPA



Step	β
0	∞
1	1.72
2	0.92
3	0.09

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

Idea behind TPA



Step	β
0	∞
1	1.72
2	0.92
3	0.09

1. $\beta \leftarrow \beta_B$
2. Repeat
3. Draw $X \leftarrow \mu(A(\beta))$
4. $\beta \leftarrow \inf\{\beta' : X \in A(\beta')\}$
5. Until $\beta \leq \beta_{B'}$

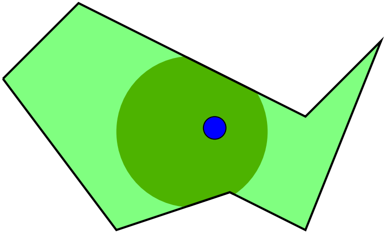
How much is shaved off at each step?

Notation: $Z(\beta) := \mu(A(\beta))$

Lemma

Say $X \sim \mu(A(\beta))$ *and* $\beta' = \min\{\beta' : X \in A(\beta')\}$. *Then*

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



Proof by picture:

Let b satisfy $Z(b)/Z(\beta) = 1/3$

Then

$$\mathbb{P}\left(\frac{Z(\beta')}{Z(\beta)} \leq 1/3\right) = \mathbb{P}(X \in A(b))$$

Product of uniforms

After k steps, measure is:

$$Z(\beta_k) = Z(\beta_B) r_1 r_2 \cdots r_k, \text{ where } r_i \stackrel{\text{iid}}{\sim} \text{Unif}([0,1])$$

Taking the logarithm:

$$\ln Z(\beta_k) = \ln Z(\beta_B) + \ln r_1 + \ln r_2 + \cdots + \ln r_k, \text{ where}$$

$$-\ln r_1, -\ln r_2, \ldots, -\ln r_k \stackrel{\text{iid}}{\sim} \text{Exp}(1)$$

So $\{\ln(Z(\beta_i))\}_{i=1}^{k}$ forms a Poisson point process!

The result

Output of TPA:

- $\ell \sim \text{Poisson}(\ln(Z(\beta_B)/Z(\beta_{B'})))$
- $\mathbb{E}[\ell] = r$, $\mathbb{V}(\ell) = r$

Repeating the Poisson point process

Suppose run the Poisson point process twice

- Result also Poisson point process rate 2 instead of rate 1

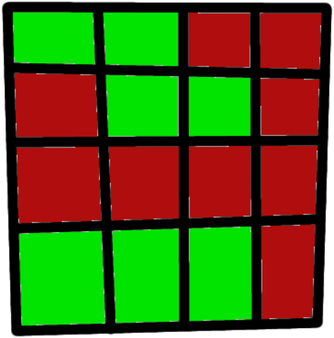
Now run k times

- Result also Poisson point process rate k instead of rate 1
- Final answer $\text{Pois}(k \ln(Z(\beta_{B'})/Z(\beta_B)))$
- Divide by k, result close to $\ln[Z(\beta_{B'})/Z(\beta_B)]$
- Exponentiate, result close to $Z(\beta_{B'})/Z(\beta_B)$
- Can use Chernoff's Bound to choose k large enough

Example

Example: The Ising model

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



$h(x) = 13$ (# adj like colored plots)

$$\pi(x) = \frac{\exp(2\beta h(x))}{Z(\beta)}$$

parameter β is inv temp

Putting Ising into framework

Add an auxiliary variable

$$[Y|X] \sim \text{Unif}([0, \exp(2\beta h(X))])$$

With Y:

- (X, Y) uniform over weird shaped space
- As β grows, more allowable values for Y
- When $\beta = 0$, $Y \sim \text{Unif}([0, 1])$ independent of X
- So $Z(0) = 2^V$ (number of configurations of X)

Method works for any exponential family

Integrated likelihood for Ising

Often called "doubly-intractable"

$$Z = \int_0^\infty p_\beta(b) \frac{\exp(2bh(x))}{Z(b)}\, db,$$

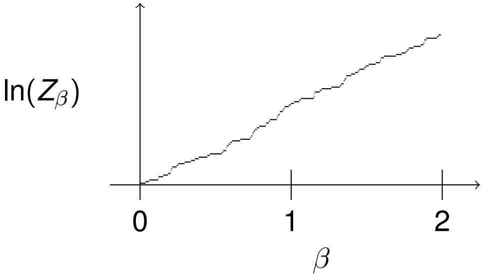
Intractable part is function $Z(b)$

- Parameter only one dimensional
- Easy to do numerically if you know $Z(\beta)$ over $(0, \infty)$

Poisson point process runs from $\ln(Z(\infty))$ **to** $\ln(Z(0))$

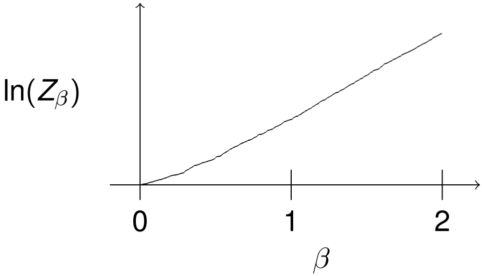
- Yields entire function $Z(b)$ for b from 0 to ∞
- Called *omnithermal* approximation

Use omnithermal approximation



One run of TPA

Use omnithermal approximation



Sixteen runs of TPA

Current work

Extending TPA

Basic TPA building block for more algorithms

- Use of particles for approximate sampling
- Creating a well-balanced cooling schedule
- Sparse versions of cooling schedule
- Embedding discrete problems in continuous space

Multiple particles

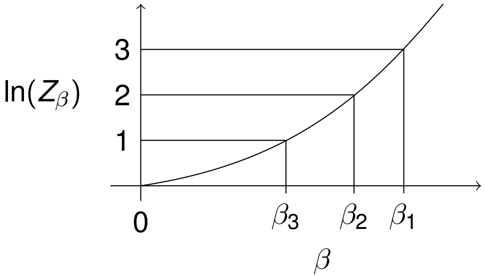
Begin with k samples from $\mu(B)$

- Idea from Nested Sampling (Skilling 2006)
- Remove outermost shell past outermost particle
- At next step add one more particle
- Approximate samples: clone randomly, run Markov chain

Change in algorithm

- Changes rate from Poisson process 1 to k

Well-balanced cooling schedule



$$\frac{Z(\beta_i)}{Z(\beta_{i-1})} \approx \frac{1}{e}$$

- Used in simulated and parallel tempering
- Reweight state space to make Markov chain mix better

Sparse cooling schedule

Often $\ln(Z(\beta))$ **has nice properties**

- For exponential families, it is convex
- Often nearly straight
- Can be used to limit number number of β's checked
- More advanced: Štefankovič, Vempala and E. Vigoda (2009)
- Right now: $r^{3/2}$. Goal: r^1

Conclusions

New algorithm: TPA

- Guaranteed performance for Monte Carlo integration
- (No variance estimate or unknown derivatives appear)
- Automatically generates cooling schedule
- Speed: r^2

Next steps

- Go from r^2 down to r

References

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

M. Huber and S. Schott
Using TPA for Bayesian inference
Bayesian Statistics 9

J. Skilling,
Nested sampling for general Bayesian computation,
Bayesian Anal., **1**(4), 833–860, 2006

D. Štefankovič,S. Vempala, and E. Vigoda,
Adaptive Simulated Annealing: A Near-Optimal Connection between Sampling and Counting,
J. of the ACM, **57**(3), 1–36, 2009

Postscript: This is not Nested Sampling

Key differences

- Nested sampling determines the nested sets for you
- TPA you can create them any way you want
- Nested sampling output has unknown fixed variance
- Variance of NS output must be estimated
- Nested sampling problems go from easy to hard
- TPA sampling problems go from hard to easy