Better Numerical Integration through Randomness

Mark Huber[1] and Sarah Schott[2]

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

[2]Department of Mathematics, Duke University

18 Nov, 2009

Randomness can be used for good or evil

Saturday Morning Breakfast Cereal 17 February, 2007

Numerical Integration

Numerical integration

Not just for those who can't master integration by parts

- Many integrands have no elementary antiderivatives
- Example:

$$\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

- Some antiderivatives exponentially large in input
- Example:

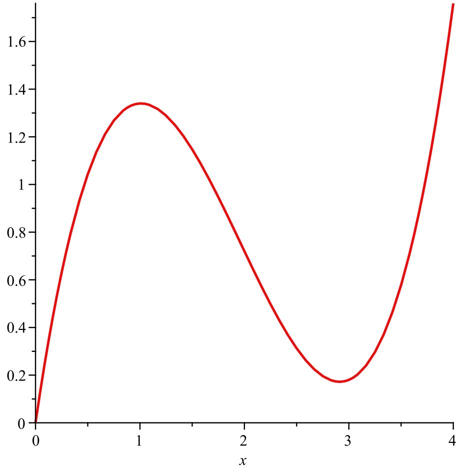
$$\int_0^a x^{1000000} \exp(-x)\, dx$$

- Antiderivative has a million (and one) terms!

1-D: easy

Fortunately, in 1-D numerical integration easy

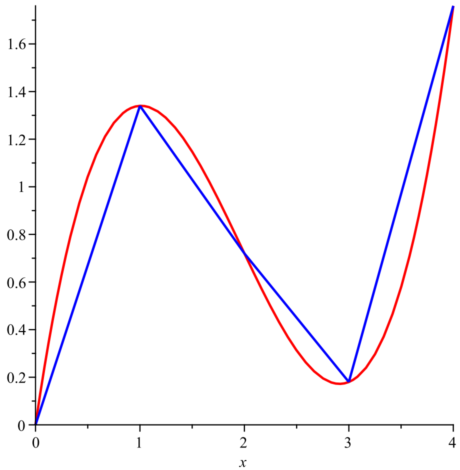
- Use rectangles, or trapezoidal rule, or Simpson's rule



1-D: easy

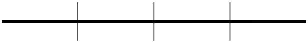
Fortunately, in 1-D numerical integration easy

- Use rectangles, or trapezoidal rule, or Simpson's rule

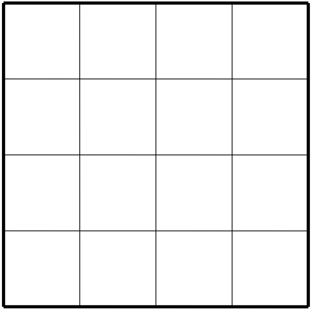


Higher dimensions, not so easy

In 1-D, k intervals



In 2-D, k^2 squares



Growth in d dimensions

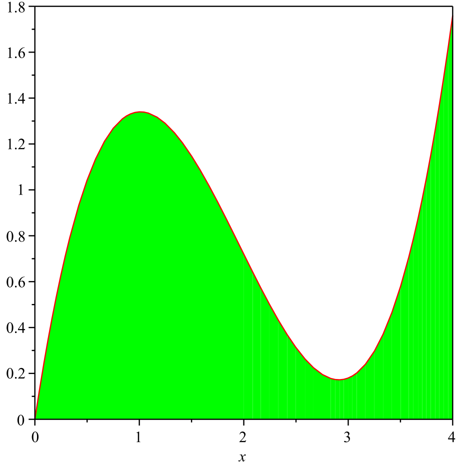
In d dimensions, need k^d squares

- Exponential growth in d
- Called "The curse of dimensionality"

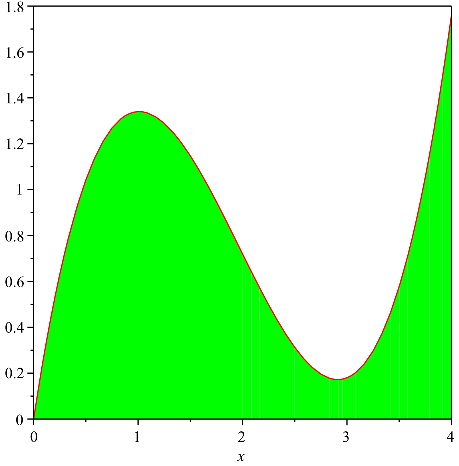
High dimensional integration arises often

- In statistics, $X_1, X_2, \ldots, X_n$ are data values
- For $A \subseteq \mathbb{R}^n$, need to find $\mathbb{P}((X_1, \ldots, X_n) \in A)$
- In combinatorics/CS, $\#P$ complete problems
- Graph with n nodes leads to n dimensional problem

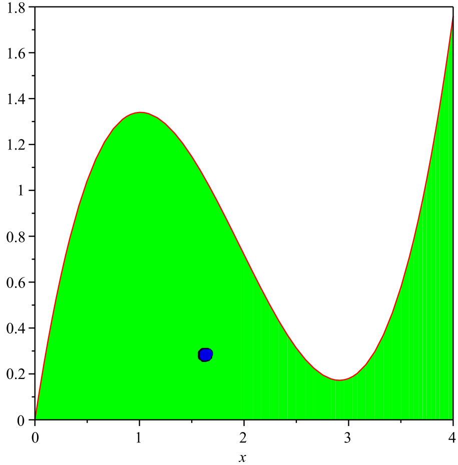
Solution: Shoot at it randomly



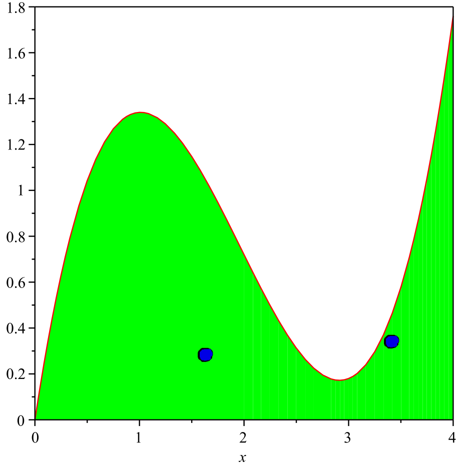
Solution: Shoot at it randomly



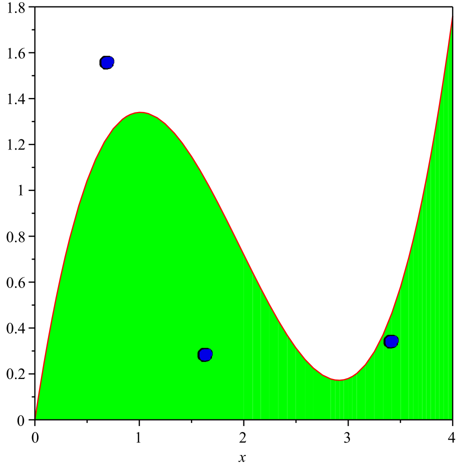
Solution: Shoot at it randomly



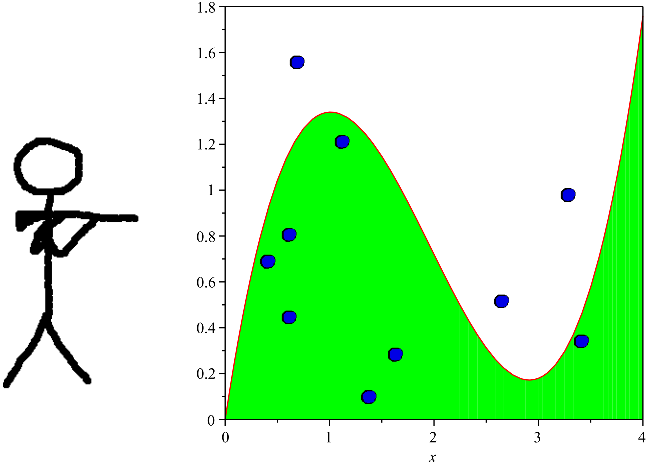
Solution: Shoot at it randomly



Solution: Shoot at it randomly



Solution: Shoot at it randomly



Analyzing the algorithm

Algorithm

- Fire n times at box
- Say you hit H times

Analysis

- Let p be chance hit green region
- Then

$$p = \frac{\text{volume(region)}}{\text{volume(box)}}$$

- Estimate

$$\hat{p} = \frac{H}{n}, \quad \widehat{\text{volume(region)}} = \frac{H}{n} \cdot \text{volume(box)}.$$

How well does it work?

True answer: 3.0933...

- After 10 iterations 5.04
- After 10^3 iterations 2.8656
- After 10^5 iterations 3.0870
- After 10^7 iterations 3.0935

About a factor of 100 per extra digit

How well does it work?

True answer: 3.0933...

- After 10 iterations 5.04
- After 10^3 iterations 2.8656
- After 10^5 iterations 3.0870
- After 10^7 iterations 3.0935

About a factor of 100 per extra digit

How well does it work?

True answer: 3.0933...

- After 10 iterations 5.04
- After 10^3 iterations 2.8656
- After 10^5 iterations 3.0870
- After 10^7 iterations 3.0935

About a factor of 100 per extra digit

How well does it work?

True answer: 3.0933...

- After 10 iterations 5.04
- After 10^3 iterations 2.8656
- After 10^5 iterations 3.0870
- After 10^7 iterations 3.0935

About a factor of 100 per extra digit

How well does it work?

True answer: 3.0933...

- After 10 iterations 5.04
- After 10^3 iterations 2.8656
- After 10^5 iterations 3.0870
- After 10^7 iterations 3.0935

About a factor of 100 per extra digit

Probability explanation

Mean value = average = expected value

$$\mathbb{E}[\hat{p}] = p$$

Standard deviation

$$\mathrm{SD}[\hat{p}] = \sqrt{\frac{p(1-p)}{n}}$$

Relative error

$$\frac{SD(\hat{p})}{p} = \sqrt{\frac{1-p}{np}} \leq \epsilon$$

Bounding error

Get relative error below ϵ

$$n \approx \frac{1-p}{p} \cdot \frac{1}{\epsilon^2}$$

- The $1/\epsilon^2$ effect called “Monte Carlo” error
- Best you can do in general
- Independent of the dimension!

Solution: Use “interior” rather than “exterior” sampling

- Need ability to sample from green region rather than box
- The algorithm is called “TPA”
- Combines “product estimator” [1] with “nested sampling” [3]

The Algorithm

The Tootsie Pop Algorithm

What is a Tootsie Pop?

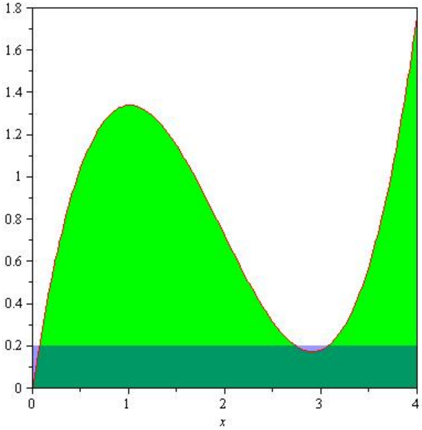
- Hard candy lollipops with a tootsie roll (chewy chocolate) at the center

In 1970, Mr. Owl was asked the question:

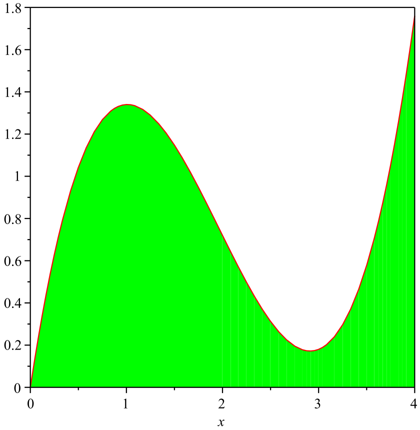
- How many licks does it take to get to the center of a Tootsie Pop?

The chocolate chewy center

- Area under .2 line almost all green
- That's our chewy center
- Hard candy is original green region
- Next question: how do we lick?

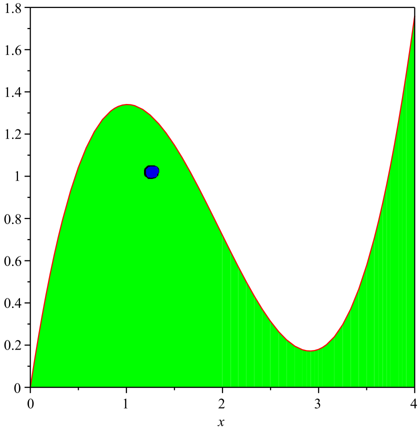


Only shoot at the green part



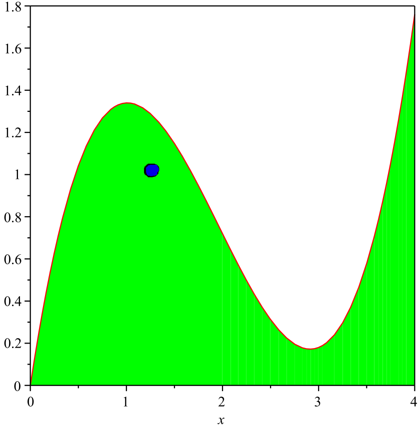
- Don't shoot bounding box
- Only shoot at green region
- Easier: don't need to maximize function!
- Usually done in high dimensions using Markov chains

Only shoot at the green part



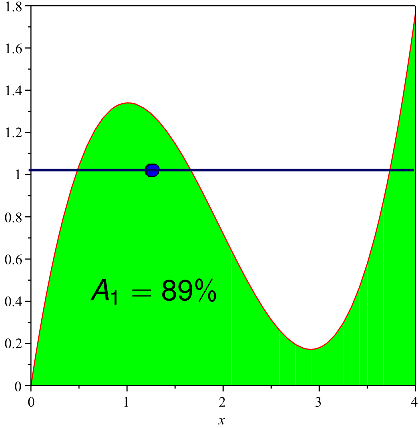
- Don't shoot bounding box
- Only shoot at green region
- Easier: don't need to maximize function!
- Usually done in high dimensions using Markov chains

Only shoot at the green part



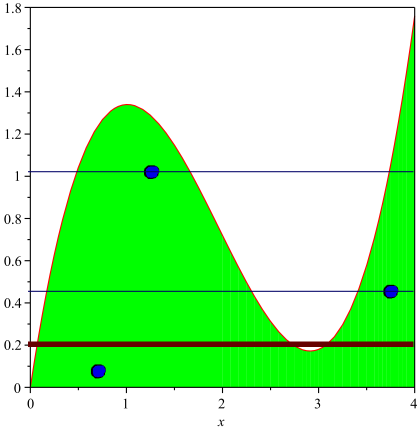
- Don't shoot bounding box
- Only shoot at green region
- Easier: don't need to maximize function!
- Usually done in high dimensions using Markov chains

Only consider region below point



Let A_1 denote the percentage of green region below the point
Note $\mathbb{P}(A_1 \leq 1/2) = 1/2$
For all $a \in [0, 1]$, $\mathbb{P}(A_i \leq a) = a$, call A_1 *uniform* on $[0, 1]$

Keep on licking!



- After choosing first point
- Draw next point from region below first point
- Continue “licking” until reach “‘center”
- In example, three licks

The number of licks

Let

$$N = \text{the number of licks}$$

Set

$$r = \frac{\text{volume(chewy center)}}{\text{volume(green region)}}, \quad A_1, A_2, \ldots \overset{\text{iid}}{\sim} \text{Unif}([0,1])$$

In terms of A_i:

$$N = \text{smallest value of } n \text{ such that } A_1 A_2 A_3 \cdots A_n \leq r$$

Multiplication to addition

Multiplication is hard:

$N =$ smallest value of n such that $\ln A_1 + \ln A_2 + \cdots + \ln A_n \leq \ln r$

Note $\ln x < 0$ **for** $x \in (0, 1)$

$N =$ smallest n such that $(-\ln A_1)+(-\ln A_2)+\cdots+(-\ln A_n) \leq -\ln r$

A day at the races

Roll A_1, move horse forward $(-\ln A_i)$
Count how many steps to reach $-\ln r$

steps	A_i	$-\ln A_i$

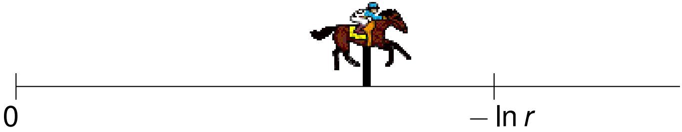
A day at the races

Roll A_1, move horse forward $(-\ln A_i)$

Count how many steps to reach $-\ln r$

steps	A_i	$-\ln A_i$
1	.23	1.469

A day at the races



Roll A_1, move horse forward ($-\ln A_i$)
Count how many steps to reach $-\ln r$

steps	A_i	$-\ln A_i$
1	.23	1.469
2	.48	.7339

A day at the races

Roll A_1, move horse forward $(-\ln A_i)$

Count how many steps to reach $-\ln r$

steps	A_i	$-\ln A_i$
1	.23	1.469
2	.48	.7339
3	.33	1.1086

Poisson process

This particular "race" is called a Poisson Process

- Very well studied process
- Again let N be number of steps until reach center
- $$\mathbb{P}(N = 1) = r, \ \ \mathbb{P}(N = i) = r\frac{(-\ln r)^{i-1}}{(i-1)!}$$
- Average value of N is $(-\ln r) + 1$
- Standard deviation of N is $\sqrt{-\ln r}$

Want to estimate r, not $-\ln r$

Repeat experiment k times

$$N_1, N_2, \ldots, N_k$$

Estimate r:

$$S = \frac{N_1 + N_2 + \cdots + N_k - k}{k}, \quad \hat{r} = e^{-S}$$

For this estimate:

$$\mathbb{E}[\hat{r}] = r, \quad \text{relative error} = \frac{-\ln r}{\sqrt{k}} = \frac{\ln(1/r)}{\sqrt{k}}.$$

The result

Basic Monte Carlo

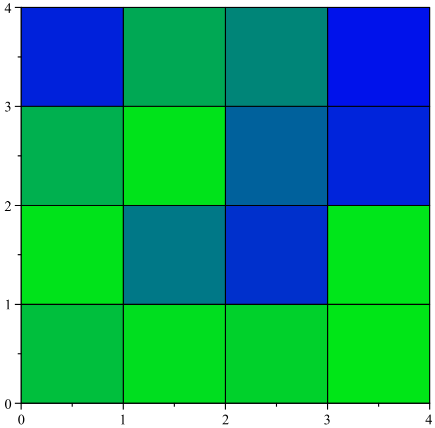
$$\#\text{ of samples} = \frac{1}{\epsilon^2} \cdot \frac{1 - r}{r}$$

TPA

$$\#\text{ of samples} = \frac{1}{\epsilon^2} \cdot \left(\ln \frac{1}{r}\right)^2$$

Applications

The Autonormal model



- Spatial model
- Soil quality for each square
- Quality between 0 (blue) and 1 (green)
- Want more likely close squares close quality

Autonormal model assigns probabilities

For vector $\vec{x} \in [0,1]^{16}$:

$$\mathbb{P}(\{d\vec{x}\}) = \frac{w_\beta(x)}{Z(\beta)}\,d\vec{x}, \quad w_\beta(x) = \prod_{i,j \text{ adjacent}} \exp(-\beta(x(i) - x(j))^2/2)$$

where

$$Z(\beta) = \int_0^1 \int_0^1 \cdots \int_0^1 w(x)\;dx(1)\;dx(2)\cdots\;dx(16)$$

Goal: given data x, find β that maximizes $w_\beta(x)/Z(\beta)$
(Called the maximum likelihood estimator.)

TPA for Autonormal

Ingredients

- $Z(0) = 1$ is hard candy shell (let $\beta_0 = 0$
- $Z(\beta)$ is chewy center
- Note $Z(\beta')$ shrinks as β' grows
- Gives nested volumes
- At β_i, sample X uniformly from volume
- Let β_{i+1} be smallest β' so X still in volume

Bayesian inference

Theorem (Bayes' Rule)

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Basic outline

- Parameters are themselves random variables with probabilities
- Take data
- Use data + Bayes' rule to update probabilities

Bayesian Example:Autonormal

Initial probabilities (called *prior*)

- Example prior: $\mathbb{P}(\beta = 1) = 1/2$, $\mathbb{P}(\beta = 0) = 1/2$
- Data $X = x$, suppose $w(x) = 4.523$
- Bayes' Rule:

$$\begin{aligned} \mathbb{P}(\beta = 1 | X \in d\vec{x}) &= \frac{\mathbb{P}(X \in d\vec{x} | \beta = 1)\mathbb{P}(\beta = 1)}{\mathbb{P}(X \in d\vec{x})} \\ &= \frac{(w_1(\vec{x})/Z(1))(1/2)}{(w_1(\vec{x})/Z(1))(1/2) + (w_0(\vec{x})/Z(0))(1/2)} \end{aligned}$$

- Need to know $Z(0)$ and $Z(1)$

$\#P$ complete problems

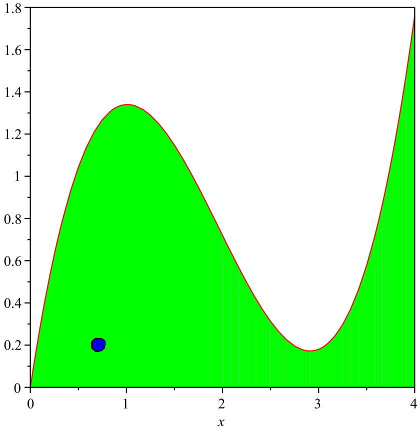
Computational Complexity

- Example: Traveling Salesman Problem
- Problem in NP: Is there a TSP path of length ≤ 60?
- Problem in $\#P$: How many TSP paths of length ≤ 60?

Some $\#P$ problems where TPA is applicable:

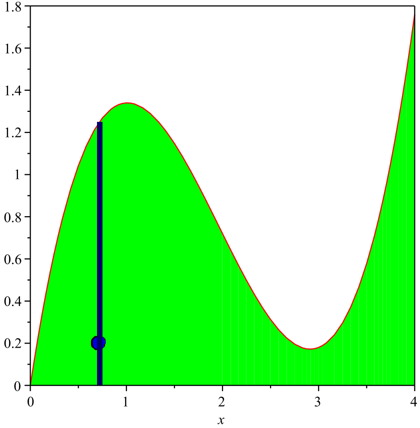
- Volume of a convex body
- Counting linear extensions of a poset
- Partition function for the Ising model

How to generate samples from green region?



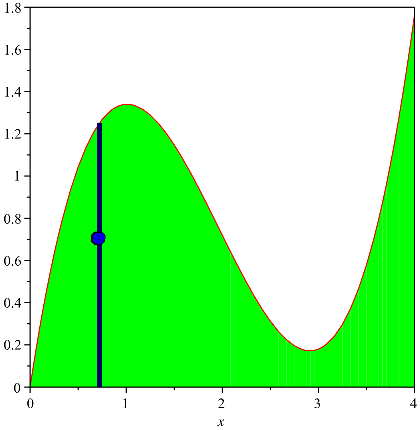
- Markov chains
- Take random steps from starting state
- Ex: Choose random vertical value
- Then choose random horizontal value

How to generate samples from green region?



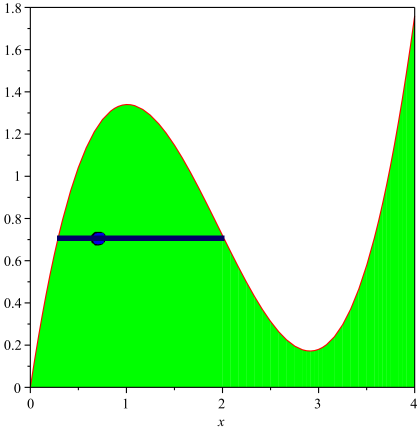
- Markov chains
- Take random steps from starting state
- Ex: Choose random vertical value
- Then choose random horizontal value

How to generate samples from green region?



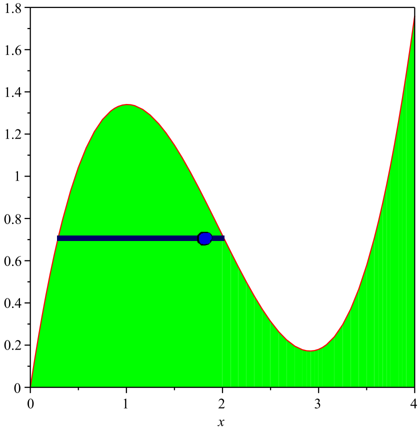
- Markov chains
- Take random steps from starting state
- Ex: Choose random vertical value
- Then choose random horizontal value

How to generate samples from green region?



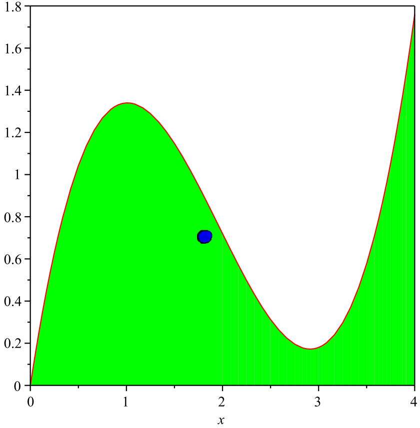
- Markov chains
- Take random steps from starting state
- Ex: Choose random vertical value
- Then choose random horizontal value

How to generate samples from green region?



- Markov chains
- Take random steps from starting state
- Ex: Choose random vertical value
- Then choose random horizontal value

How to generate samples from green region?



- Markov chains
- Take random steps from starting state
- Ex: Choose random vertical value
- Then choose random horizontal value

Raises further questions

For a Markov chain

- How many steps until state close to random?
- (How many times must a deck of cards be riffle shuffled?)
- How best to design Markov chain?
- Markov chain Monte Carlo (MCMC) most popular method, there are others

But a more important question...

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

- Purdue created a licking machine, 364 licks
- Purdue also ran human experiments: 252 licks
- U. Michagan created a licking machine, 411 licks
- Swarthmore Junior HS ran human experiment: 144 licks

Summary: results and future work

The Tootsie Pop algorithm

- Interior rather than exterior sampling
- No need for bounding box
- Do need a chewy center (usually easy to find)
- Many applications!

Open questions

- Associated sampling problems
- Exponential families need # of samples equal to

$$\frac{1}{\epsilon^2}\left(\ln \frac{1}{r}\right)$$

Can TPA match this?

References

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

M. Huber and S. Schott
Improving the product estimator by using a random cooling schedule
preprint

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