# Fast approximate counting of linear extensions

Jacqueline Banks[1], Scott Garrabrant[2], Mark Huber[3], Anne Perizzolo[4]

[1]Department of Statistics, UC Riverside

[2]Department of Mathematics, Pitzer College

[3]Department of Mathematical Sciences, Claremont McKenna College

[4]Department of Mathematics, Columbia University

9 Jan, 2011

# The Problem

## Ranking objects

- Suppose $A, B, C, D$ each have a number
- Can compare any pair to find out which has the lower number
- Suppose data is $A < C$, $B < C$, and $B < D$

## Linear extensions

- A *linear extension* is a ranking consistent with data
- Five linear extensions for example above:

$$ABCD, ABDC, BACD, BADC, BDAC$$

- Estimate that $A$ is ranked first is $2/5$

# Counting linear extensions is hard

### How hard is it?

- Let $\Omega_{LE}$ denote the set of linear extensions
- Finding $\#\Omega_{LE}$ is a #P complete problem
- (Counting version of NP complete problem)

# Naive Acceptance/Rejection approach

To estimate $\#\Omega_{LE}$:

**Repeat** $N$ **times**

- Draw $X$ uniformly from $\Omega_{LE}$

**Output** $N/(\#$ **of times** $X$ **is** $ABCD)$

# Why is this not a good idea

## Suppose have $n$ objects to rank

- $\Omega_{LE}$ can be as large as $n!$ with $n$ objects to rank
- Need $N \approx n!$ to have significant chance of hitting $ABCD$ at all!

## More formally

- To get relative error $1 + \epsilon$ with probability at least $1 - \delta$, need

$$N = O\left(\#\Omega_{LE} n^3 (\ln n) \frac{1}{\epsilon^2} \ln\left(\frac{1}{\delta}\right)\right)$$

- Today, new algorithm

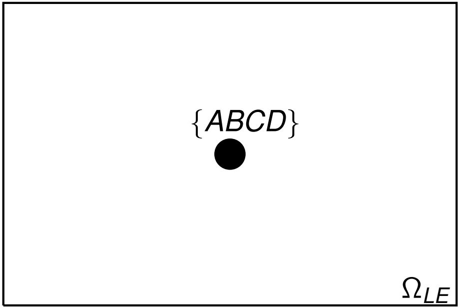
$$N = O\left((\ln \#\Omega_{LE})^2 \frac{1}{\epsilon^2} n^3 (\ln n) \ln\left(\frac{1}{\delta}\right)\right)$$

# To make this work

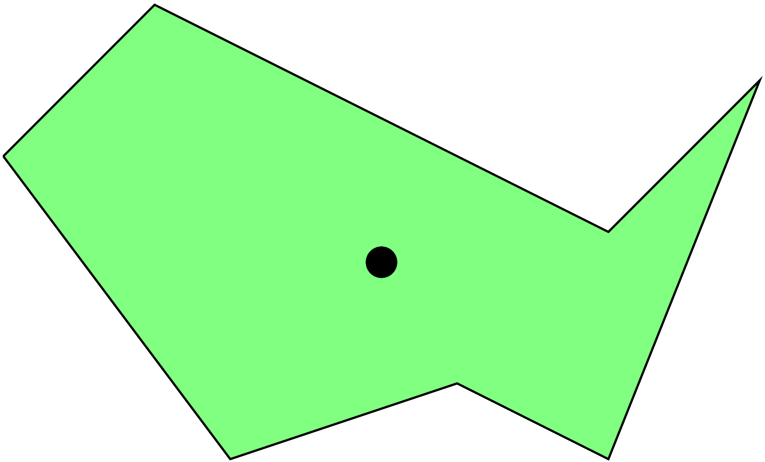
**Ingredients:**

- TPA: better than accept/reject for problems with continuous parameter
- Method for adding continuous parameter to discrete state space
- Yields new distribution on $\Omega_{LE}$ that is not uniform
- New sampling method for this nonuniform distribution

# Linear extensions, counting = discrete problem

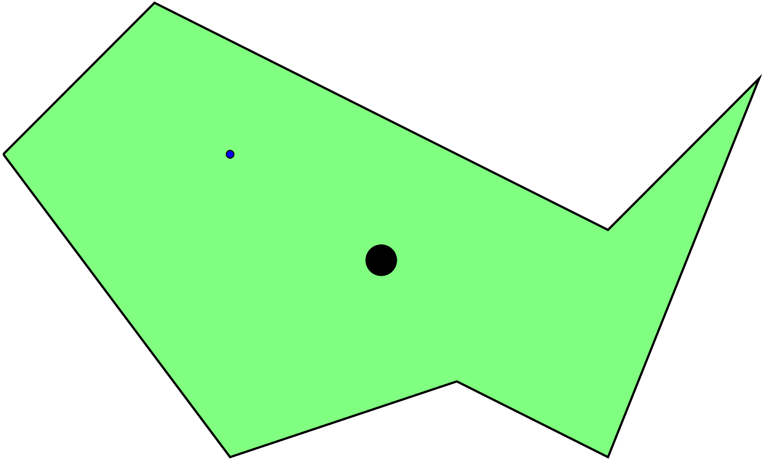


# Area = continuous problem



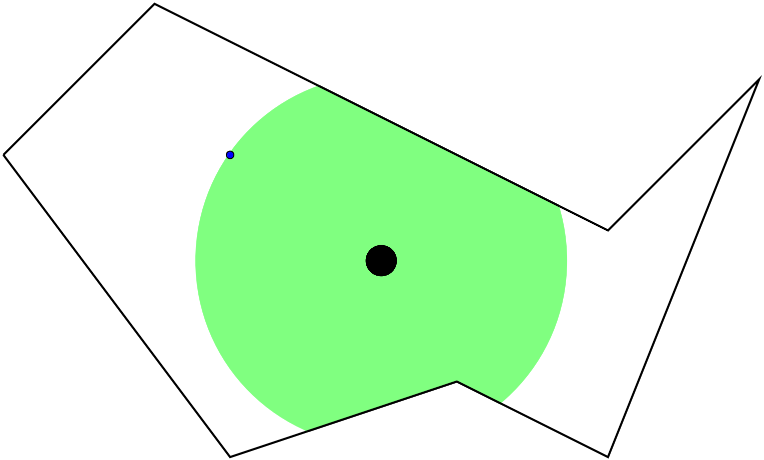
Black circle in center area 1, want area larger region

# Area = continuous problem



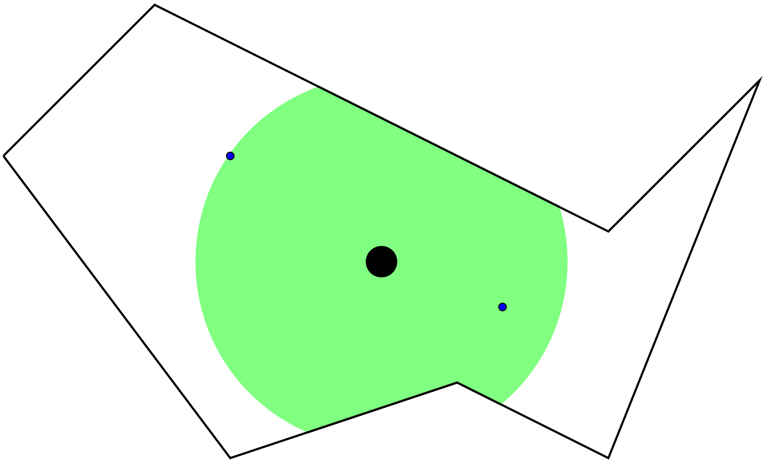
Black circle in center area 1, want area larger region

# Area = continuous problem



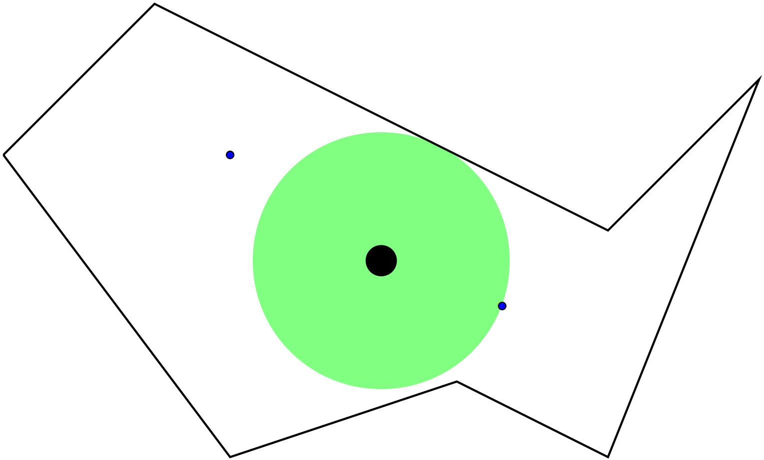
Black circle in center area 1, want area larger region

# Area = continuous problem



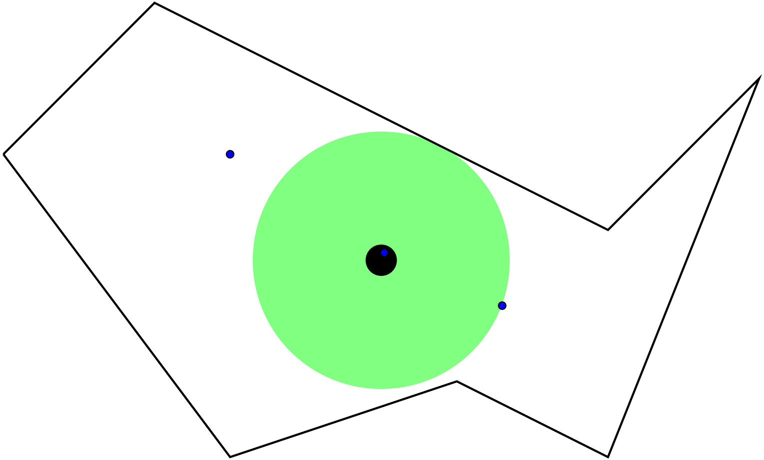
Black circle in center area 1, want area larger region

# Area = continuous problem



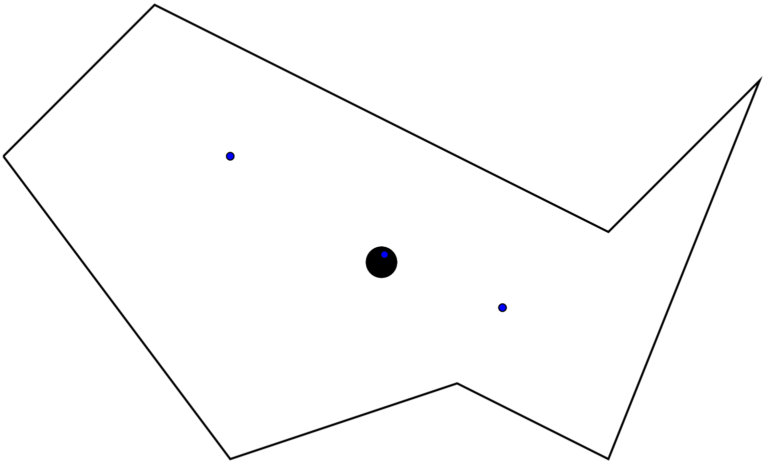
Black circle in center area 1, want area larger region

# Area = continuous problem



Black circle in center area 1, want area larger region

# Area = continuous problem



Black circle in center area 1, want area larger region

# How many steps before reaching center?

### Steps to reach center is a random variable

- Previous example, it was 2
- Distribution is Poisson(ln(size of large region))

### Proof idea

- Shaves off uniform amount each step
- Absolute value of natural log of uniform on $[0, 1]$ is exponential (mean 1)
- So in log space, we are taking exponential steps
- Gives Poisson process
- Number of steps before reaching center also Poisson

# TPA = Tootsie Pop Algorithm

**What is a Tootsie Pop?**

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

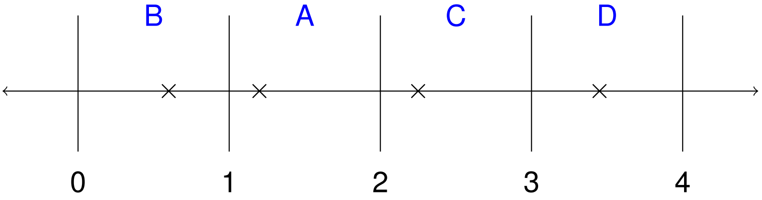
**In 1970, Mr. Owl was asked the question:**

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

# Continuizing Linear Extensions

## Linear extensions are a discrete set

- They don't have a continuous parameter!
- Solution: embed in a continuous setting
- Example:

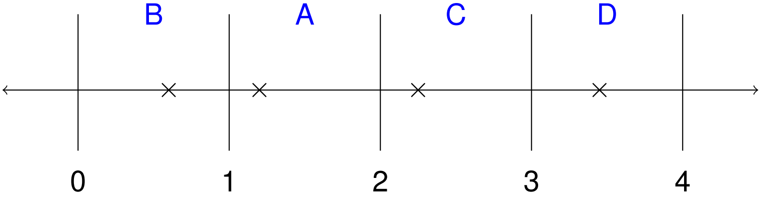



Vector: (1.21,0.62,2.25,3.45)

# Adding a continuous parameter

## One way to add parameter

- Create a home position (example: *ABCD*)
- Let $\beta$ be the farthest to the right any object is from home




$\beta = .21$ (comes from *A*)

# With parameter

$$\beta = 4, \;\; \text{set} = \Omega_{LE}$$

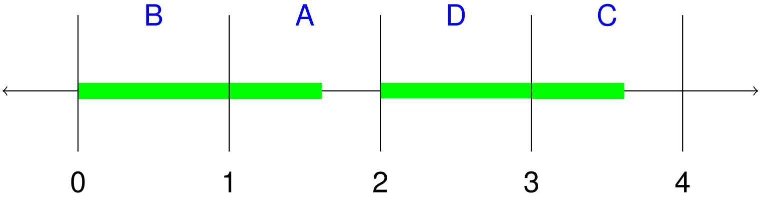
$$\beta = 0, \;\; \text{set} = \{ABCD\}$$

To use TPA, must be able to sample from any $\beta$

# The nonuniform distribution

**How does $\beta$ change the distribution on linear extensions?**

- Example: $\beta = .61$ consider *BADC*
- *A* and *C* do not have full freedom, *B* and *D* do




Volume is (1)(.61)(1)(.61)

- General weight is $\beta$ raised to the number of objects exactly $1 + \lfloor \beta \rfloor$ to the right of home

# Transposition Markov chain

**Can we sample from this distribution?**

- Use Markov chain Monte Carlo
- Pick object unfiormly, swap with one to right
- Only accept move with probability $\beta$ if moves object to $1 + \lfloor \beta \rfloor$ away from home and doesn't violate ranking data

**Does this chain mix rapidly?**

- Same technique as in [1] shows it does
- After $O(n^3 \ln n)$ steps, sample close to distribution

# Perfect sampling via bounding chains

**Even better!**

- Can draw exactly from this distribution using perfect sampling
- Coupling from the Past idea of Propp and Wilson...
- ...with bounding chains [1]
- (Can't do this with all Markov chains)
- Result: $O(n^3 \ln n)$ to get samples exactly from right distribution

# Summary: results and future work

## Use Tootsie Pop Algorithm for linear extensions

- TPA designed for continuous problems
- This is example of TPA for discrete
- Many other examples can use this technique
- Currently approx. algorithm

$$O((\ln \#\Omega_{LE})^2 n^3 \ln n)$$

## Current work

- Get algorithm down to

$$O((\ln \#\Omega_{LE}) n^3 \ln n)$$

# References


M. Huber
Fast perfect sampling from linear extensions
Discrete Mathematics, 306, pp. 420–428, 2006

M. Huber and S. Schott
Improving the product estimator by using a random cooling schedule (with comments and rejoinder)
Bayesian Statistics 9, to appear.

J. Banks, S. Garrabrant, M. Huber and A. Perizzolo
Using TPA for approximating the number of linear extensions
Submitted.