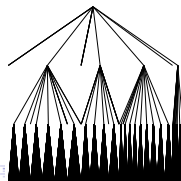
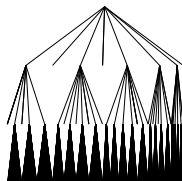
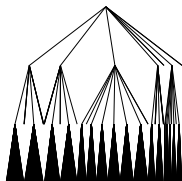


Infinite Boltzmann Samplers and Applications to Branching Processes

Hanane Tafat Bouzid

& Olivier Bodini (Laboratoire d'Informatique de Paris Nord),
& Guillaume Moroz (INRIA Nancy).

June 27 2012
Gascom, Bordeaux



Plan

- 1 Introduction
- 2 Infinite Boltzmann Sampler
- 3 Applications to Galton–Watson process and prefix of trees
- 4 Complexity analysis
- 5 Conclusion

Motivations

- Boltzmann samplers: rejection algorithms finely tuned
[Duchon, Flajolet, Louchard and Schaeffer 2001]
gives automatic random generators for **finite** objects described by a combinatorial specification.
- Works for all combinatorial objects given by the symbolic method
[Flajolet, Sedgewick 2009]
- Provides more **efficient sampling** than the recursive method
[Nijenhuis–Wilf 1978, Flajolet–Van Cutsem–Zimmermann 1994,
Denise–Zimmermann 1999]

Motivations

- Boltzmann samplers: rejection algorithms finely tuned [Duchon, Flajolet, Louchard and Schaeffer 2001] gives automatic random generators for **finite** objects described by a combinatorial specification.
- Works for all combinatorial objects given by the symbolic method [Flajolet, Sedgewick 2009]
- Provides more **efficient sampling** than the recursive method [Nijenhuis–Wilf 1978, Flajolet–Van Cutsem–Zimmermann 1994, Denise–Zimmermann 1999]
- Boltzmann samplers: optimal (linear) complexity for approximate size uniform random generation.
- Technical condition: requires a generating function **analytic at 0**.

Our aim: Extend Boltzmann samplers so that there are no longer restricted to the generation of infinite-size objects.

Boltzmann Sampler

A Boltzmann sampler $\Gamma_{\mathcal{C}}$ of the combinatorial class \mathcal{C} is an algorithm (parametrized by $x \in \mathbb{R}$) that produces an object γ of size n with probability

$$\Pr(\gamma) = \frac{x^n}{C(x)}$$

where $C(x)$ the generating function of \mathcal{C} .

Therefore

- The probability to draw an object γ of size n is uniform.
- The size S of γ is a random variable: $\Pr(S = n) = \frac{c_n x^n}{C(x)}$.
- It is easy to finely tune the Boltzmann sampler in order to generate an object of size n : by choosing x such that $E(S) = n = x \frac{C'(x)}{C(x)}$.

Boltzmann Sampler

A Boltzmann sampler $\Gamma_{\mathcal{C}}$ of the combinatorial class \mathcal{C} is an algorithm (parametrized by $x \in \mathbb{R}$) that produces an object γ of size n with probability

$$\Pr(\gamma) = \frac{x^n}{n!C(x)}$$

where $C(x)$ the generating function of \mathcal{C} .

Therefore

- The probability to draw an object γ of size n is uniform.
- The size S of γ is a random variable: $\Pr(S = n) = \frac{c_n x^n}{n!C(x)}$.
- It is easy to finely tune the Boltzmann sampler in order to generate an object of size n : by choosing x such that $E(S) = n = x \frac{C'(x)}{C(x)}$.

Boltzmann Sampler

Basic combinatorial specification

Sampler $\Gamma\mathcal{C}(x)$

ϵ
■

return ϵ
return ■

$\mathcal{C} = \mathcal{A} + \mathcal{B}$

if $\text{Bern}\left(\frac{A(x)}{C(x)}\right) = 1$ then $\Gamma\mathcal{A}(x)$ else $\Gamma\mathcal{B}(x)$

$\mathcal{C} = \mathcal{A}.\mathcal{B}$

return $(\Gamma\mathcal{A}(x), \Gamma\mathcal{B}(x))$

$\mathcal{C} = \text{Seq}(\mathcal{A})$

return k -tuple of $\Gamma\mathcal{A}(x)$ (with $k = \text{Geom}(A(x))$)

magic: independant drawings for the product!

Boltzmann Sampler

Basic combinatorial specification

Sampler

ϵ

$\Gamma\mathcal{C}(x) := \epsilon$

■

$\Gamma\mathcal{C}(x) := \blacksquare$

$\mathcal{C} = \mathcal{A} + \mathcal{B}$

$\Gamma\mathcal{C}(x) := \left(\text{Bern} \left(\frac{A(x)}{C(x)} \right) \rightarrow \Gamma\mathcal{A}(x) | \Gamma\mathcal{B}(x) \right)$

$\mathcal{C} = \mathcal{A}.\mathcal{B}$

$\Gamma\mathcal{C}(x) := (\Gamma\mathcal{A}(x), \Gamma\mathcal{B}(x))$

$\mathcal{C} = \text{Seq}(\mathcal{A})$

$\Gamma\mathcal{C}(x) := (\text{Geom}(A(x))) \Rightarrow \Gamma\mathcal{A}(x)$

magic: independant drawings for the product!

Boltzmann Sampler

Basic combinatorial specification

Sampler

$$\mathcal{C} = \text{Set}(\mathcal{A})$$

$$\Gamma\mathcal{C}(x) := (\text{Pois}(A(x)) \Rightarrow \Gamma\mathcal{A}(x))$$

$$\mathcal{C} = \text{Cycle}(\mathcal{A})$$

$$\Gamma\mathcal{C}(x) := (\text{Loga}(A(x)) \Rightarrow \Gamma\mathcal{A}(x))$$

Complexity

The Boltzmann sampler $\Gamma C(x)$ has a linear cost in the size of the object generated.

Interesting fact: nice link with computer algebra (Newton iteration scheme) for an efficient computation of $C(x)$ [Pivoteau, Salvy, Soria 2011]

Plan

- 1 Introduction
- 2 Infinite Boltzmann Sampler**
- 3 Applications to Galton–Watson process and prefix of trees
- 4 Complexity analysis
- 5 Conclusion

Infinite Boltzmann Sampler

Given a recursive combinatorial class \mathcal{C} , we introduce:

- $E_{\mathcal{C}}(z, C(z)) = 0$ the functional equation derived from the recursive specification of \mathcal{C} .
- Boltzmann-valid points $V := \{(x, y) \in \mathbb{R}_+^2 \mid E_{\mathcal{C}}(x, y) = 0\}$.

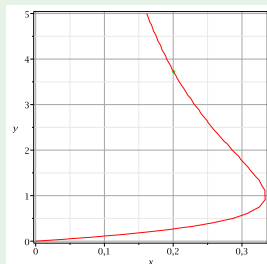
Example: Unary-binary trees

Recursive specification:

$$\mathcal{Y} = \mathcal{Z} + \mathcal{Z}\mathcal{Y} + \mathcal{Z}\mathcal{Y}^2$$

Boltzmann-valid points:

$$V := \{(x, y) \in \mathbb{R}_+^2 \mid y = x + xy + xy^2\}$$

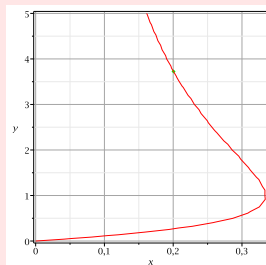


Infinite Boltzmann Sampler

The infinite Boltzmann sampler $\Gamma_{\mathcal{C}}$ of the combinatorial class \mathcal{C} is an algorithm with two parameters $(x, y) \in V$ that produced an object of size n with probability

$$\Pr(S = n) = \frac{c_n x^n}{y} \text{ and } \Pr(S = \infty) = 1 - \frac{C(x)}{y}$$

- $C(z)$ is the unique solution of $E_{\mathcal{C}}(z, C(z))$ analytic in 0 with non negative coefficients;
- if $(x, y) \in V$ such that $y = C(x)$ then it gives the **classical Boltzmann distribution**, $\Pr(S = \infty) = 0$.

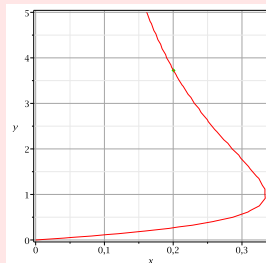


Infinite Boltzmann Sampler

The infinite Boltzmann sampler $\Gamma\mathcal{C}$ of the combinatorial class \mathcal{C} is an algorithm with two parameters $(x, y) \in V$ that produced an object of size n with probability

$$\Pr(S = n) = \frac{c_n x^n}{n! y} \text{ and } \Pr(S = \infty) = 1 - \frac{C(x)}{y} \text{ with } C(x) = \sum_{n=0}^{\infty} \frac{c_n x^n}{n!}.$$

- $C(z)$ is the unique solution of $E_C(z, C(z))$ analytic in 0 with non negative coefficients;
- if $(x, y) \in V$ such that $y = C(x)$ then it gives the **classical Boltzmann distribution**, $\Pr(S = \infty) = 0$.



Infinite Boltzmann Sampler

Combinatorial specification

Sampler

ϵ

$$\Gamma_{\mathcal{C}}(x, y) := \epsilon$$

■

$$\Gamma_{\mathcal{C}}(x, y) := \blacksquare$$

$$\mathcal{C} = \mathcal{A} + \mathcal{B} \quad \Gamma_{(x, y_A + y_B)} \mathcal{C} := \left(\text{Bern}\left(\frac{y_A}{y_A + y_B}\right) \rightarrow \Gamma_{(x, y_A)} \mathcal{A} \mid \Gamma_{(x, y_B)} \mathcal{B} \right)$$

$$\mathcal{C} = \mathcal{A} \cdot \mathcal{B} \quad \Gamma_{(x, y_A y_B)} \mathcal{C} := (\Gamma_{(x, y_A)} \mathcal{A}, \Gamma_{(x, y_B)} \mathcal{B})$$

$$\mathcal{C} = \text{Seq}(\mathcal{A}) \quad \Gamma_{(x, \frac{1}{1-y_A})} \mathcal{C} := (\text{Geom}(y_A) \implies \Gamma_{(x, y_A)} \mathcal{A})$$

Infinite Boltzmann Sampler

Combinatorial specification

Sampler

$$\mathcal{C} = \text{Set}(\mathcal{A})$$

$$\Gamma_{(x, e^{y_A})} \mathcal{C} := (\text{Pois}(y_A) \implies \Gamma_{(x, y_A)} \mathcal{A})$$

$$\mathcal{C} = \text{Cycle}(\mathcal{A})$$

$$\Gamma_{(x, \ln(\frac{1}{1-y_A}))} \mathcal{C} := (\text{Loga}(y_A) \implies \Gamma_{(x, y_A)} \mathcal{A})$$

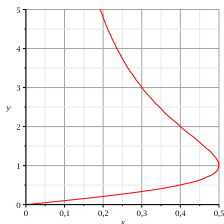
Plan

- 1 Introduction
- 2 Infinite Boltzmann Sampler
- 3 Applications to Galton–Watson process and prefix of trees**
- 4 Complexity analysis
- 5 Conclusion

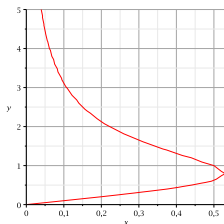
Galton–Watson process (Definition)

- Each node has a probability p_k to have k children.
- p_0 is the probability of extinction of a node.
- If $p_0 < 1/2$, nonzero probability to have an **infinite process** \Rightarrow **impossible to use the "classical Boltzmann sampler"**.

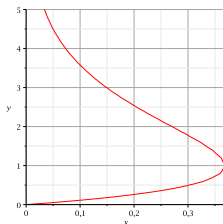
Solution: use our "infinite Boltzmann sampler"!



Binary trees



Ternary trees



Cayley trees

Works similarly for all families of trees studied by [\[Meir–Moon 1978\]](#)

Galton–Watson process (Example)

- Binary trees: $\mathcal{T} = \mathcal{Z} + \mathcal{Z}\mathcal{T}^2$.

$$\Gamma_{(x,y)}\mathcal{T} := \text{Bern}(x/y) \rightarrow (\blacksquare | [\blacksquare, \Gamma_{(x,y)}\mathcal{T}, \Gamma_{(x,y)}\mathcal{T}])$$

- Boltzmann-valid points: $\{(x, y), 0 \leq x \leq 1/2, x + xy^2 - y = 0\}$.

\equiv

- $\{(x, \frac{1-\sqrt{1-4x^2}}{2x}), 0 \leq x \leq 1/2\} \cup \{(x, \frac{1+\sqrt{1-4x^2}}{2x}), 0 \leq x < 1/2\}$

|||

\cup

|||

$$p_0 \geq 1/2$$

(classical Boltzmann sampler)

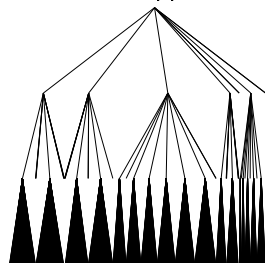
$$p_0 < 1/2$$

(infinite Boltzmann sampler)

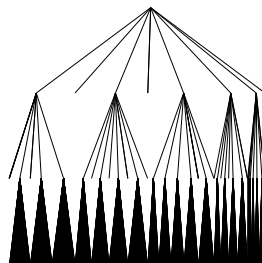
Prefix of infinite Cayley trees

- Cayley trees: $\mathcal{T} = \mathcal{Z} \times \text{Set}(\mathcal{T})$.
- Infinite-Boltzmann sampler: $\Gamma_{(x, x \cdot e^y)} \mathcal{T} = (\mathcal{Z}, [\text{Pois } y \implies \Gamma_{(x,y)} \mathcal{T}])$.

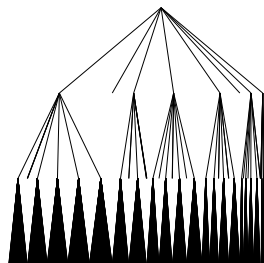
Simulation stopped at height=4:



$$(x, y) = (0.5, 0.3)$$



$$(x, y) = (0.24, 0.34)$$



$$(x, y) = (0.16, 0.20)$$

Plan

- 1 Introduction
- 2 Infinite Boltzmann Sampler
- 3 Applications to Galton–Watson process and prefix of trees
- 4 Complexity analysis**
- 5 Conclusion

Complexity

At each step of the computation, the complexity of our infinite-Boltzmann sampler (measured in the number of real-arithmetic operations) is linearly proportional to the number of atoms which have been output.

Complexity analysis with rejection

- The probability generating function of the interrupted infinite Boltzmann sampler targeted at $n_1 \leq n \leq n_2 = (1 - \epsilon)n \leq n \leq (1 + \epsilon)n$ is $F(u) = \sum_k \Pr(T_n = k)u^k$
- $$F(u) = \frac{1}{1 - \left(\sum_{k < n_1} \frac{a_k x^k u^k}{y} + u^{n_2} \sum_{k > n_2} \frac{a_k x^k}{y} \right)} \sum_{n_1 < k < n_2} \frac{a_k x^k u^k}{y}$$
$$= \frac{C^{[n_1, n_2]}(x, u)/y}{1 - 1/y(C^{<n_1}(x, u) + C^{>n_2}(x, u))}.$$
- The average cost:

$$E(T_n) = \frac{\partial}{\partial u} F(u) \Big|_{u=1} = \frac{x C'^{[n_1, n_2]}(x, 1) + x C'^{<n_1}(x, 1) + n_2 C'^{>n_2}(x, 1)}{C^{[n_1, n_2]}(x, 1)}.$$

- Assuming $C(x) = \tau - c(1 - \frac{x}{\rho})^{1/2} + O(1 - \frac{x}{\rho})$ and $x = \rho(1 - \frac{1}{2n})$:

$$E(T_n) = f(\epsilon)n$$

$$E(T_n) = f(\epsilon)n \sim \frac{K}{\epsilon}n$$

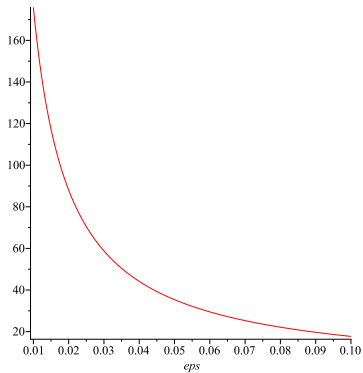


Figure: curve of $f(\epsilon)$.

Plan

- 1 Introduction
- 2 Infinite Boltzmann Sampler
- 3 Applications to Galton–Watson process and prefix of trees
- 4 Complexity analysis
- 5 Conclusion

Conclusion

Our "infinite Boltzmann sampler"

- generalizes the classical Boltzmann method for the generation of objects of infinite size.
- works for all the combinatorial specifications of the symbolic method.

In progress:

- allows automatization of random generation of finite and infinite combinatorial objects [[Maple/Sagemath packages](#), [ANR Magnum project](#)]