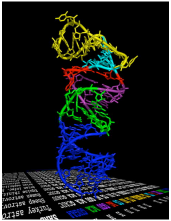
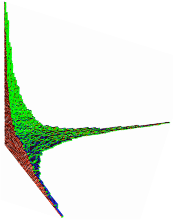
Boltzmann sampling

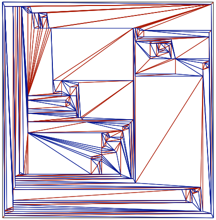
Carine Pivoteau

LIP6 – UPMC

based on work by P. Duchon, P. Flajolet, E. Fusy,
G. Louchard, C. Pivoteau and G. Schaeffer







Random generation: different approaches

Fixed size random uniform generation:

- Ad hoc methods
  - bijections, surjections, ...
    $\mathcal{A} = \phi(\mathcal{B})$ and $\Gamma\mathcal{B}(n) \;\Rightarrow\;$ random sampler $\Gamma\mathcal{A}(n)$
    $a_n = f(a_{n-1}) \;\Rightarrow\;$ incremental algorithm $\Gamma\mathcal{A}(n)$
  - rejection
    $\mathcal{A} \subset \mathcal{B}$ and $\Gamma\mathcal{B}(n) \;\Rightarrow\;$ random sampler $\Gamma\mathcal{A}(n)$
- Recursive method : counting + recursive process
  - *Nijenhuis, Wilf, 1978*
  - *Flajolet, Zimmermann, Van Cutsem, 1994*
    preprocessing time (to compute g.f. coefficients): $O(n^2)$
    random generation time : $O(n \log n)$

Approximate size random uniform generation:

- Boltzmann sampling...

Boltzmann method

Random sampling under Boltzmann model

- approximate size sampling,
- size distribution spread over the whole combinatorial class, but uniform for a sub-class of objects of the same size,
- control parameter,
- automatized sampling: the sampler is compiled from specification automatically,
- very large objects can be sampled.
  - → large scale simulations
  - → observation of random structures limit properties...

Boltzmann samplers for the random generation of combinatorial structures. P. Duchon, P. Flajolet, G. Louchard, G. Schaeffer. Combinatorics, Probability and Computing, 13(4-5):577-625, 2004. Special issue on Analysis of Algorithms.

Boltzmann sampling of unlabelled structures. Ph. Flajolet, E. Fusy, C. Pivoteau. Proceedings of ANALCO07, january 2007.

Model definition

Definition

In the unlabelled case, Boltzmann model assigns to any object $c \in \mathcal{C}$ the following probability:

$$\mathbb{P}_x(c) = \frac{x^{|c|}}{C(x)}$$

In the labelled case, this probability becomes:

$$\mathbb{P}_x(c) = \frac{1}{\hat{C}(x)} \frac{x^{|c|}}{|c|!}$$

A free Boltzmann sampler $\Gamma C(x)$ for the class $\mathcal{C}$ is a process that produces objects from $\mathcal{C}$ according to this model.

→ 2 objects of the same size will be drawn with the same probability.

Unlabelled unions, products, sequences

Suppose $\Gamma A(x)$ and $\Gamma B(x)$ are given:

Disjoint unions

Boltzmann sampler ΓC for $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$:

With probability $\frac{A(x)}{C(x)}$ do $\Gamma A(x)$ else do $\Gamma B(x)$ $\rightarrow$ Bernoulli.

Products

Boltzmann sampler ΓC for $\mathcal{C} = \mathcal{A} \times \mathcal{B}$:

Generate a pair $\langle\ \Gamma A(x)\ ,\ \Gamma B(x)\ \rangle$ $\rightarrow$ independent calls.

Sequences

Boltzmann sampler ΓC for $\mathcal{C} = \textsc{Seq}(\mathcal{A})$:

Generate k according to a geometric law of parameter $A(x)$

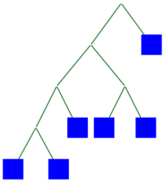
Generate a k-tuple $\langle\ \Gamma A(x)\ ,\ldots,\ \Gamma A(x)\ \rangle$ $\rightarrow$ independent calls.

Remark: $A(x)$, $B(x)$ and $C(x)$ are given by an ***oracle***.

Binary trees

$$\mathcal{B} = \mathcal{Z} + \mathcal{B} \times \mathcal{B}$$

$$B(z) = z + B(z)^2 = \frac{1 - \sqrt{1 - 4z}}{2}$$



Algorithm: $\Gamma B(x)$

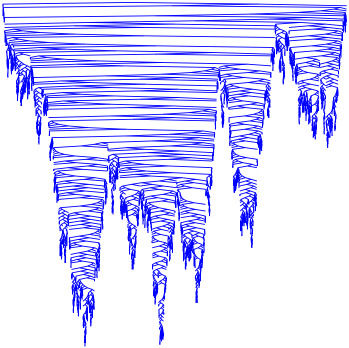
$b \leftarrow \mathrm{Bern}(x/B(x))$;
if $b = 1$ **then**
  Return ■
else
  Return $\langle\ \Gamma B(x)\ ,\ \Gamma B(x)\ \rangle$;
end if

Examples of specifications with $\{\cup, \times, \text{Seq}\}$

Regular specifications (non recursive).

- integer compositions, permutations,...
- polyominos that have rational g.f.: column-convex,

- regular languages,



Context-free specifications.

- any algebraic language,
- tree-like structures
  - k-ary, 2–3–4 trees, ...,
  - triangulations,
  - noncrossing graphs,
  - general planar rooted trees,
  - ...

Labelled classes

Same algorithms, with exponential generating functions

construction	sampler
$\mathcal{C} = \emptyset$ or $\mathcal{Z}$	$\Gamma C(x) := \varepsilon$ or atom
$\mathcal{C} = \mathcal{A} + \mathcal{B}$	$\Gamma C(x) :=$ Bern $\frac{\hat{A}(x)}{\hat{C}(x)} \longrightarrow \Gamma A(x) \mid \Gamma B(x)$
$\mathcal{C} = \mathcal{A} \times \mathcal{B}$	$\Gamma C(x) := \langle\ \Gamma A(x)\ ;\ \Gamma B(x)\ \rangle$
$\mathcal{C} = \textsc{Seq}(\mathcal{A})$	$\Gamma C(x) :=$ Geom $\hat{A}(x) \Longrightarrow \Gamma A(x)$

Put the labels at the end !

Size control – parameter tuning

- Free samplers: produce objects with randomly varying sizes!
- Approximate and exact size samplers: use rejection.
- Tuned samplers: choose x so that expected size is n.

$$\mathbb{E}_x(N) = x\frac{C'(x)}{C(x)} \quad \text{or} \quad x\frac{\hat{C}'(x)}{\hat{C}(x)}$$

- Size distribution determines the cost of rejection.