Boltzmann Sampling and Simulation

Michèle Soria
LIP6 – UPMC

CONFERENCE IN MEMORY OF PHILIPPE FLAJOLET

Paris–Jussieu, December 15, 2011

Random Sampling Combinatorial structures

Combinatorial modeling $\rightarrow$ Sampling of random objects

Get information on the model by working on the random objects

- visualize and compute order of growth for quantities ;
- distinguish significant properties (random vs observed)
- compare models, test software

Analytic Combinatorics $\rightarrow$ **Generic Methods**

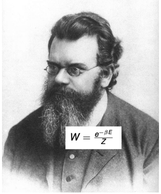
- **Recursive Method** : $n \rightarrow \gamma_n$ $\boxed{\mathbb{P}(\gamma_n) = 1/c_n}$

  compute enumeration sequences

- **Boltzmann Method** : $x \rightarrow \gamma$ $\boxed{\mathbb{P}(\gamma) \propto x^{|\gamma|}}$

  evaluate generating functions at x

Boltzmann Sampling (DuFlLoSc04)



Boltzmann model

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

uniform sampling

$$\mathcal{C} = \Phi(\mathcal{A}, \mathcal{B})$$
$$\Downarrow$$
$$\Gamma\mathcal{C} = \mathbf{F}(\Gamma\mathcal{A}, \Gamma\mathcal{B})$$

Boltzmann samplers for the random generation of combinatorial structures – DuFlLoSc04 – cited by 124

Quadratic exact-size and linear approximate-size random generation of planar graphs E Fusy - 2005 -
Boltzmann sampling of unlabelled structures P Flajolet, E Fusy, C Pivoteau- 2007 -
An unbiased pointing operator for unlabeled structures, with ... M Bodirsky, E Fusy, M Kang, S Vigerske- 2007 -
Degree distribution of random Apollonian network structures and Boltzmann sampling A Darrasse, M Soria - 2007 -
Properties of random graphs via boltzmann samplers K Panagiotou, A.Weiss - 2007 -
Random Graphs with Structural Constraints A Steger - 2007 -
Random generation of finitely generated subgroups of a free group F Bassino, C Nicaud, P Weil - 2008-
Random generation of possibly incomplete deterministic automata. F Bassino, J David, C Nicaud - 2008 -
Boltzmann oracle for combinatorial systems C Pivoteau, B Salvy, M Soria - 2008 -
Random XML sampling the Boltzmann way A Darrasse -2008 -
The degree sequence of random graphs from subcritical classes N Bernasconi, K Panagiotou, - 2009 -
Boltzmann samplers for colored combinatorial objects O Bodini, A. Jacquot- 2009 -
Boltzmann Samplers for v-balanced Colored Necklaces O Bodini, A Jaquot - 2009 -
Fast and sound random generation for automated testing and benchmarking in objective Caml B Canou, A Darrasse- 2009
Uniform random sampling of planar graphs in linear time E Fusy - 2009 -
Profiles of permutations M Lugo - 2009 -
Uniform random generation of huge metamodel instances A Mougenot, A Darrasse, X Blanc, M Soria- 2009 -
Vertices of degree k in random unlabeled trees K Panagiotou, M Sinha - 2009 -
Boltzmann sampling of ordered structures O Roussel, M. Soria- 2009 -
Autour de la génération aléatoire sous modèle de Boltzmann O Bodini - 2010
Random sampling of plane partitions O Bodini, É Fusy, C Pivoteau - 2010 -
Multi-dimensional Boltzmann sampling of languages O Bodini, Y Ponty- 2010 -
Boltzmann Samplers, Pólya Theory, and Cycle Pointing M Bodirsky, E Fusy, M Kang, S Vigerske- 2010 -
Boltzmann generation for regular languages with shuffle A Darrasse, K Panagiotou, O Roussel, M Soria- 2010 -
Quadratic and linear time random generation of planar graphs E Fusy - 2010
Parametric Random Generation of Deterministic Tree Automata PC Héam, C Nicaud, S Schmitz - 2010-
Profiles of large combinatorial structures MT Lugo - 2010 -
Average Analysis of Glushkov Automata under a BST-Like Model C Nicaud, C Pivoteau, B Razet, - 2010 -
Boys-and-girls birthdays and Hadamard products O Bodini, D Gardy, O Roussel - 2011 -
Lambda terms of bounded unary height O Bodini, D Gardy, B Gittenberger- 2011 -
Boltzmann samplers for first-order differential specifications O Bodini, O Roussel, M Soria- 2011 -
Analyse d'algorithmes et génération aléatoire en théorie des langages C Nicaud - 2011 -
Enumeration and random generation of pin permutations F Bassino, M Bouvel, A Pierrot, C Pivoteau, D Rossin, 2012
Dirichlet Random Samplers for Multiplicative Structures O Bodini, J Lumbroso- 2012 -

. . .

"If you can specify it and compute values of GF, then you can sample it with Boltzmann method"

Comb. Spec. $\mathcal{C} = \Phi(\mathcal{A}, \mathcal{B}, \ldots)$
$x < \rho$; oracle for $C(x), A(x), B(x)$

→ Random Objet $\gamma \in \mathcal{C}$; $\mathbb{P}_x(\gamma) = \frac{x^{|\gamma|}}{C(x)}$

Random Number Generator

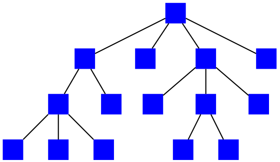
- Genericity
  Combinatorial Specifications → Equations on GF → Sampler
- Effectivity
  Sampler compiled from specifications,
  Simple probabilistic algorithms, linear time

$\mathcal{A} + \mathcal{B}$: flip a biaised coin and sample from $\mathcal{A}$ or $\mathcal{B}$
*$\mathcal{A} \times \mathcal{B}$: **independantly sample** from $\mathcal{A}$ and $\mathcal{B}$*

Binary trees :

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

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



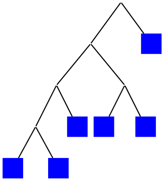
Algorithm : $\Gamma\mathcal{B}(x)$

if Bernoulli $\frac{x}{B(x)}$ **then**
 Return ■
else
 Return $\langle\, \Gamma\mathcal{B}(x)\,,\, \Gamma\mathcal{B}(x)\,\rangle$;
end if

Arbres planaires :

$$\mathcal{T} = \mathcal{Z} \times \textsc{Seq}(\mathcal{T})$$

$$T(z) = \frac{z}{1-T(z)} = \frac{1-\sqrt{1-4z}}{2}$$



Algorithm : $\Gamma\mathcal{T}(x)$

$k \leftarrow \mathrm{Geom}(T(x))$;
$\mathit{forest} \leftarrow \langle\, \underbrace{\Gamma\mathcal{T}(x), \ldots, \Gamma\mathcal{T}(x)}_{k \text{ times}} \,\rangle_{etiq}$;
Return ■ $\times$ *forest* ;

recursive process ; stops with proba 1 for x $< \rho$

SPECIFICATION OF $\mathcal{C}$ $\rightsquigarrow$ FUNCTIONAL EQUATION $\mathcal{C}(z)$

Combinatorial Class $\mathcal{C}$, labelled objects
Exponential Generating Function (EGF)
$C(z) = \sum_{\gamma \in \mathcal{C}} \frac{z^{|\gamma|}}{|\gamma|!} = \sum_{n \geq 0} \frac{c_n z^n}{n!}$

Construction	Notation	EGF
ε	1	1
atome	$\mathcal{Z}$	z
Union	$\mathcal{C} = \mathcal{A} \cup \mathcal{B}$	$C(z) = A(z) + B(z)$
Produit	$\mathcal{C} = \mathcal{A} \times \mathcal{B}$	$C(z) = A(z) \times B(z)$
Séquence	$\mathcal{C} = \text{SEQ}(\mathcal{A})$	$C(z) = (1 - A(z))^{-1}$
Ensemble	$\mathcal{C} = \text{SET}(\mathcal{A})$	$C(z) = \exp(A(z))$
Cycle	$\mathcal{C} = \text{CYC}(\mathcal{A})$	$C(x) = -\text{Log}\,(1 - A(z))$

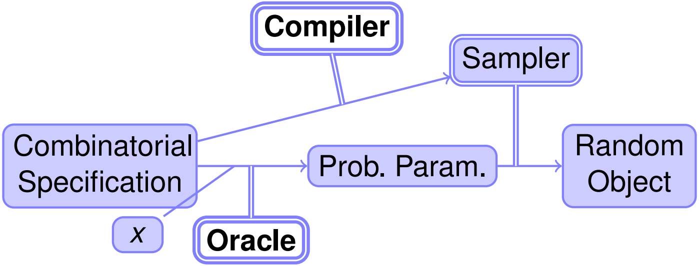
SPECIFICATION OF $\mathcal{C}$ $\rightsquigarrow$ SAMPLING ALGORITHM $\Gamma\mathcal{C}(x)$

Combinatorial Class $\mathcal{C}$, labelled objects
Exponential Generating Function (EGF)

$C(z) = \sum_{\gamma \in \mathcal{C}} \frac{z^{|\gamma|}}{|\gamma|!} = \sum_{n \geq 0} \frac{c_n z^n}{n!}$ $\qquad$ $\mathbb{P}_x(\gamma) = \frac{x^{|\gamma|}}{|\gamma|! C(x)}$

Construction	*Sampling Algorithm*
$\mathcal{C} = \emptyset$ or $\mathcal{Z}$	$\Gamma\mathcal{C}(x) := \varepsilon$ or atom
$\mathcal{C} = \mathcal{A} \cup \mathcal{B}$	$\Gamma\mathcal{C}(x) :=$ Bernoulli $\left(\frac{A(x)}{A(x)+B(x)}\right)$; return $\Gamma\mathcal{A}(x)$ or $\Gamma\mathcal{B}(x)$
$\mathcal{C} = \mathcal{A} \times \mathcal{B}$	$\Gamma\mathcal{C}(x) :=$ return $< \Gamma\mathcal{A}(x); \Gamma\mathcal{B}(x) >$ *independent calls*
$\mathcal{C} = \text{SEQ}(\mathcal{A})$	$\Gamma\mathcal{C}(x) := k \leftarrow$ Geometric$(A(x))$; return k independent $\Gamma\mathcal{A}(x)$
$\mathcal{C} = \text{SET}(\mathcal{A})$	$\Gamma\mathcal{C}(x) := k \leftarrow$ Poisson$(A(x))$; return k independent $\Gamma\mathcal{A}(x)$
$\mathcal{C} = \text{CYCLE}(\mathcal{A})$	$\Gamma\mathcal{C}(x) := k \leftarrow$ Logarithmic$(A(x))$; return k independent $\Gamma\mathcal{A}(x)$

"If you can specify it and compute values of GF, then you can sample it with Boltzmann method"



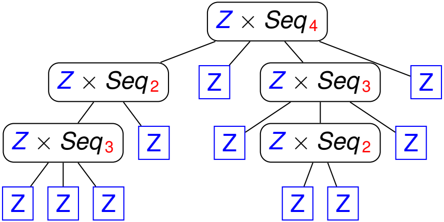
- Parameter x
- Computation of numerical values of generating functions : *oracle*
- Probabilistic Algorithms : dicrete laws with parameters $A(x)$

Complexity of free generation – Labelled Universe

Theorem (DuFlLoSc04)

The Boltzmann sampler $\Gamma C(x)$ for a class $\mathcal{C}$, of labelled objects, recursively specified on ε , Z, $\cup$, $\times$, Seq, Set, Cycle, has arithmetic complexity which is linear in the size of the result (under hypothesis oracle in $O(1)$).

(Oracle : Pivoteau-Salvy-Soria 2008, 2011)

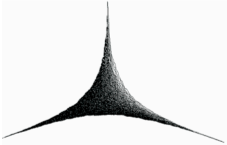


Uniform random number in [0,1] + real numbers manipulation

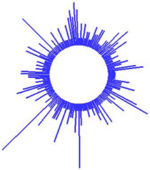
Complexity of free generation – Unlabelled Universe
Flajolet-Fusy-Pivoteau 2007

Theorem (FlFuPi07)

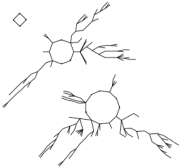
For a class $\mathcal{C}$ of non labelled objects, recursively specified on ε, Z, $\cup$, $\times$, Seq, Set, Cycle, the Boltzmann sampler has arithmetic complexity which is linear in the size of the result (under hypothesis oracle in $O(1)$).



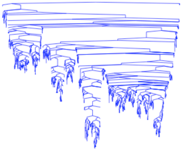
Partition plane



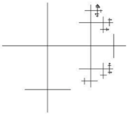
Composition circulaire



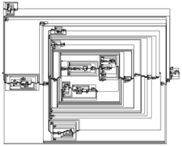
Graphe fonctionnel



Arbre non planaire



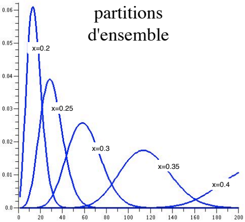
Alcool acyclique



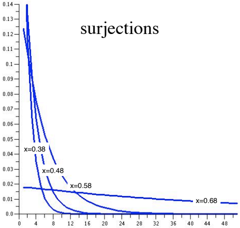
Circuit série-parallèle

Size distribution of random objects $x \to \gamma \in \mathcal{C}$

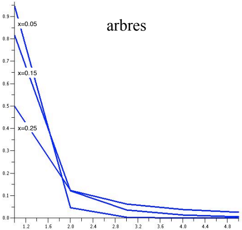
- $\mathbb{P}_x(\gamma) = \dfrac{x^{|\gamma|}}{|\gamma|!C(x)}$ uniformity for all objects of a given size
- Size N_x of the output is a random variable $\mathbb{P}(N_x = n) = \dfrac{c_n x^n}{n!C(x)}$
- Tuning x allows for aiming a certain size $\mathbb{E}(N_x) = x\dfrac{C'(x)}{C(x)}$
- Size distribution depends on $x_{(0<x<\rho)}$: continuum of distr. funct.
- Shape of distribution depends on the singularity type of $C(x)$



$C(x) = e^{(e^x - 1)}$



$C(x) = \frac{1}{2 - e^x}$



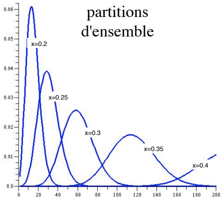
$C(x) = \frac{1 - \sqrt{1 - 4x}}{2}$

Sampling with Size Control

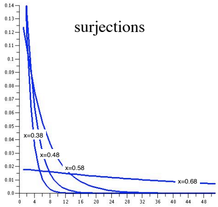
- accept sample objects of approximate size $[n(1-\varepsilon), n(1+\varepsilon)]$
- rejection : if size is not in the accepted interval, then reject and sample again

Theorem (DuFlLoSc04, FlFuPi07)

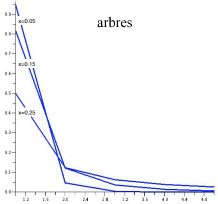
For a class $\mathcal{C}$ of (non) labelled objects, recursively specified on ε , Z, $\cup$, $\times$, Seq, Set, Cycle, the Boltzmann sampler with reject $\Gamma C(x, n, \varepsilon)$ has average arithmetic complexity in $O(n)$.



1 trial on average



$O(1)$ *trial* on average



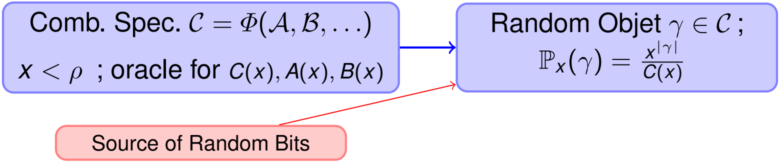
$O(n)$ *trial* on average

Current issues in Boltzmann sampling

- Extend the model
  - to other operators
  - to other universes
- Non uniform generation
  - multivariate, multicriteria sampling
  - sampling according to other distributions
  - biased sampling for realistic modeling
- Certification, implementation
  - exact simulation
  - dicrete model
  - automatization

Discrete Boltzmann Samplers

- Arithmetic Boltzmann Samplers : evaluate GFs
  $\rightarrow$ approximations ... Boltzmann distribution ?
- Boltzmann Samplers based on binary coin flipping ?



- Genericity et Effectivity
- Average boolean complexity en $O(|\gamma|)$
- Perfect Simulation of discrete random variables
  - X : Geometric, Poisson, Logarithmic
  - $\mathbb{P}(X = k)$ needs $O(k)$ bits on average

Perfect Simulation of random variables

- Simulating RV : efficient algorithms operating over reals and using real source of randomness – *Luc Devroye's treatise*
- Discrete source of randomness and finite computations ?
  - *Knuth-Yao* (1976)
    random [0,1] : $U_i = \sum_{j>0} b_j 2^{-j}, \;\; b_j \in \{0,1\}$
    Lazy evaluation for the generation of $(U_1, \ldots, U_N)$
  - *Von Neumann* (1949), $\quad e^{-x} = \sum_{n \text{ odd}} \frac{x^{n-1}}{n-1!} - \frac{x^n}{n!}$
    $x = U_1 > \ldots > U_n < U_{n+1}$
    if odd(n) **return** x **else** *do it again*
  - *Flajolet-Saheb* (1983-86) : Path length analysis of a trie

Theorem (FlSa83)

The average number of elementary coin flips for producing k bits of the exponential variable is $k + \gamma + o(1)$, and the g.f. of the distributions has a closed form expression.

Exact simulation of discrete laws for Boltzmann sampling

Algorithm : Bernoulli(p)

$i \leftarrow 1 + Geom(1/2)$;
Return $Bit_i(p)$;

$Bit_i(p)$: access with proba $\frac{1}{2^{(i+1)}}$

Algorithm : Geometric(p)

$k \leftarrow 0$;
do $i \leftarrow 1 + Geom(1/2)$;
if $Bit_i(p) = 1$ **then** Return k ;
else $k \leftarrow k + 1$;
fi ;**od** ;

O(k+1) bits for Geom(p)= k

Algorithm :Von Neuman $(x, \mathcal{P})$

do
$N \leftarrow Geom(x)$;
draw $U_1, \ldots, U_N$ indep. R.V. unif $[0, 1]$;
if $\sigma(U_1, \ldots, U_N) \in \mathcal{P}$ **then** Return N ;
od ;

$\mathbb{P}\,(VN(x, \mathcal{P}) = n) = \frac{1}{P(x)} \frac{P_n x^n}{n!}$ *Boltz*

- $U_1 < \ldots < U_n$
  $P_n = 1 \rightarrow P(x) = e^x \Longrightarrow$ Poisson
- $U_1 < \{U_2, \ldots, U_n\}$
  $P_n = (n-1)! \rightarrow P(x) = \log \frac{1}{1-x}$
  $\Longrightarrow$ Logarithmic

Buffon Machines (Flajolet-Pelletier-Soria 2011)

- Boolean Complexity Von Neuman Schema

Theorem (FlPeSo11)

Let $\mathcal{P}$ class of permutations and x parameter, von Neumann Schema exactly simulates discrete v.a. N, s.t. $\mathbb{P}(N = n) = \frac{1}{P(\lambda)} \frac{P_n \lambda^n}{n!}$, with average number of flips $O(1)$ and exponential queues.

- perfect simulation of Bernouilli laws with various parameters : π, $\exp(-1)$, $\log 2$, $\sqrt{3}$, $\cos(\frac{1}{4})$, $\zeta(5)$

$$C = \frac{7}{8}\zeta(3) - \frac{\pi^2}{12}\log 2 + \frac{1}{6}(\log 2)^3 = 0.53721\,13193\cdots$$

consumes on average less than 6 coin flips, and at most 20 coin flips in 95% of the simulations.

- Work goes on : extensions to other laws, integration in Boltzmann implementations *Jérémie Lumbroso, Maryse Pelletier, MS, ...*