



HAL
open science

Boltzmann Samplers for v -balanced Colored Necklaces

Olivier Bodini, Alice Jacquot

► **To cite this version:**

Olivier Bodini, Alice Jacquot. Boltzmann Samplers for v -balanced Colored Necklaces. 2009. hal-00432215

HAL Id: hal-00432215

<https://hal.science/hal-00432215v1>

Preprint submitted on 14 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Boltzmann Samplers for v -balanced Colored Necklaces

O. Bodini and A. Jacquot

LIP6, UMR 7606, Département CALSCI, Université Paris 6 - UPMC,
104, avenue du président Kennedy,
F-75252 Paris cedex 05, France

Abstract. This paper is devoted to the random generation of particular colored necklaces for which the number of beads of a given color is constrained (these necklaces are called v -balanced). We propose an efficient sampler (its expected time complexity is linear) which satisfies the Boltzmann model principle introduced by Duchon, Flajolet, Louchard and Schaeffer (7). Our main motivation is to show that the absence of a decomposable specification can be circumvented by mixing the Boltzmann samplers with other types of samplers.

Introduction

Necklaces are classical objects in combinatorics (10; 6; 13; 14). They naturally occur in the study of Lyndon words or in many other enumeration problems (5). For $v = (v_1, \dots, v_k)$ a k -tuple of positive integers, our interest lies in uniformly drawing v -balanced necklaces of n beads : the beads can take k distinct colors and the number n_i of beads of color i verifies the v -balance (which we define as meaning that (n_1, \dots, n_k) is collinear to v). An additional reason to focus on v -balanced structures derives from the intrinsic difficulty to enumerate and describe such objects in terms of analytic combinatorics. In particular, the generating function of v -balanced cycles is neither holonomic nor closed-formed. We attempt to draw very large necklaces in order to conjecture some limit properties on them. To that purpose, we adopt the framework of Boltzmann samplers. This approach, like the recursive method (15), allows to “automatically” build a sampler from a decomposable specification of a combinatorial class. A Boltzmann sampler does not guarantee the size of the generated object, only that the drawn object has the same probability to be drawn as any other of the same size. More details on this point can be found in the preliminary section. But we can already state that for a large number of combinatorial classes this relaxation allows to generate an object of size n in expected linear time $O(n)$ without preprocessing.

The main problem is that v -balanced necklaces do not admit decomposable s using traditional builders ($+$, \times , Seq , ...), except for some very special cases. The aim of this paper is to adapt an ad-hoc sampler for v -balanced sequences, in such a way that the sampling follows a Boltzmann model. After that, we will use this

Boltzmann sampler (which cannot be built from a decomposable specification) to obtain a Boltzmann sampler for v -balanced cycles. To this end, we need to know how to obtain a Boltzmann sampler for a class \mathcal{A} from a Boltzmann sampler for its pointing class $\Theta\mathcal{A}$.

This paper is organized in four sections. The first section defines the notations and concepts used throughout the article. After reflecting upon basic notions related to combinatorial classes and Boltzmann samplers, we define our classes of interest, namely, the v -balanced sequences and the v -balanced cycles.

The second section addresses the sampler for $(1, 1)$ -balanced cycles. We show two ways of building such a sampler. The first one uses a natural isomorphism with a decomposable combinatorial class involving Dyck paths. Unfortunately, this approach cannot be generalized, therefore we propose a second one based on an isomorphism between pointed $(1, 1)$ -balanced cycles and the weighted sum of $(1, 1)$ -balanced sequences. These sequences can be drawn with an ad-hoc sampler. In the third section, a generalization of the previous approach to v -balanced cycles is given, yielding an efficient Boltzmann sampler which, for our object of study, bypasses the need for a specification of the class.

The fourth section concludes with some perspective works.

1 Preliminaries

An *(unlabelled) combinatorial class* is a couple $\langle \mathcal{C}, s \rangle$ (generally abbreviated by \mathcal{C}) where \mathcal{C} is a set of *objects* and s is a function on \mathcal{C} called *size function* which satisfies the conditions :

- (i) $\forall a \in \mathcal{C}$, $s(a)$ is a non-negative integer;
- (ii) the number of objects of any given size is finite.

We naturally associate to a combinatorial class \mathcal{C} the ordinary generating function $C(x) = \sum c_n x^n$ where c_n is the number of objects of size n in \mathcal{C} . A *Boltzmann sampler* for an unlabelled combinatorial class \mathcal{C} is a random generator such that the probability to draw an object a of size n is $P_x(a) = \frac{x^n}{C(x)}$. These

samplers were first introduced in (7) and extended, (in particular the sampler for unlabelled cycles) in (9). Let us notice that a Boltzmann sampler depends on a parameter x which can be tuned to focus on an expected output size. More precisely, let N be the random variable of the size of the output, we can solve the equation $\mathbb{E}(N) = x \cdot \frac{C'(x)}{C(x)}$ to center the output distribution on an expected value.

Let us recall in the following table the samplers for the unlabelled operators that we need.

Boltzmann samplers are very powerful tools to efficiently generate combinatorial structures (2; 8). In particular, It can be possible to automatically build a sampler according to the specification of a combinatorial class. Our aim is to

Sampler	Description
$\Gamma_x(\mathcal{Z})$	Return \mathcal{Z} . \mathcal{Z} denotes an atomic class.
$\Gamma_x(\mathcal{A}\mathcal{B})$	Return $(\Gamma_x\mathcal{A}, \Gamma_x\mathcal{B})$.
$\Gamma_x(\mathcal{A} + \mathcal{B})$	If $Bernoulli(\frac{A(x)}{A(x)+B(x)}) = 1$ then Return $\Gamma_x\mathcal{A}$ else Return $\Gamma_x\mathcal{B}$
$\Gamma_xSeq(\mathcal{C})$	Draw l according to a geometric law of parameter $C(x)$. Return the concatenation of l calls to $\Gamma_x\mathcal{A}$.
$\Gamma_xRep_n(\mathcal{C})$	Return $\underbrace{aa\dots a}_{n \text{ times}}$ with a generated by $\Gamma_{x^n}\mathcal{A}$.
$\Gamma_xCyc(\mathcal{C})$	Let K be a random variable in \mathbb{N}^* verifying : $\mathbb{P}(K = k) = \frac{-\varphi(k)}{kCycA(x)} \log(1 - A(x^k))$ with $CycA(x)$ the generating function of $Cyc(\mathcal{A})$ Draw k according to the law of K . Draw j according to a logarithmic law of parameter $A(x^k)$. Let M by the concatenation of j calls to $\Gamma_{x^k}\mathcal{A}$ Return $\underbrace{MM\dots M}_{k \text{ times}}$.

Fig. 1. Some classical samplers with \mathcal{A} , \mathcal{B} and \mathcal{C} combinatorial classes, $A(x)$, $B(x)$ and $C(x)$ their respective generating functions. \mathcal{C} must not have neutral objects.

consider Cyc_v and Seq_v defined below, as additional basic classes, to be added to the collection of classical constructions thus increasing the expressivity of the Boltzmann model.

Definition 1. We denote by Seq_v (resp. Cyc_v) a sequence of atoms (objects of size 1), such that each atom can be colored by one of the color in $\{1, \dots, k\}$ and the numbers n_i of beads of color i verifies the v -balanced condition. That is to say (n_1, \dots, n_k) is collinear to v .

From an easy observation, it can be seen that the generating function of $Seq_{(v_1, \dots, v_k)}$ is $\sum \left(\frac{(n|v|)!}{\prod ((v_i n)!) } \right) x^{n|v|}$ where $|v| = \sum v_i$. The following proposition is a trivial consequence.

Proposition 1 (see e.g (1)). The class Seq_v is holonomic for every v .

Proof. Indeed, a quick calculation shows that this can be written as the hypergeometric function :

$${}_kF_{k-1} \left(\frac{1}{|v|}, \frac{2}{|v|}, \dots, \frac{|v|}{|v|}, \frac{1}{v_1}, \frac{2}{v_1}, \dots, \frac{v_1}{v_1}, \frac{1}{v_2}, \frac{2}{v_2}, \dots, \frac{v_2}{v_2}, \dots, \frac{1}{v_k}, \frac{2}{v_k}, \dots, \frac{v_k-1}{v_k}; \frac{|v|^{|v|}}{\prod v_i^{v_i}} \times x^{|v|} \right).$$

The hypergeometric functions are holonomic by definition.

Remark: Nevertheless, the class Seq_v is algebraic in only very few cases. For instance, $Seq_{(1,1,1)}$ is not algebraic. See (1) for a complete classification of the algebraic cases.

2 Generating (1,1)-balanced cycles

This section is dedicated to the random generation of (1, 1)-balanced cycles. This is a good example to illustrate our approach. A (1, 1)-balanced sequence (resp. cycle) is by previous definition a sequence (resp. cycle) of black atoms Z_b and white atoms Z_w such that the number of black atoms is equal to the number of white atoms. The generating function for $Seq_{(1,1)}$ is $S_{(1,1)} = {}_1F_0(1/2, 4x^2) = (1 - 4x^2)^{-\frac{1}{2}}$. In particular, $Seq_{(1,1)}$ is algebraic. We are going to use this important property in the following first approach.

2.1 First approach: through a decomposable specification

Both $Seq_{(1,1)}$ and $Cyc_{(1,1)}$ are specifiable. We can thus apply the unlabelled samplers described in the previous table. More precisely, $Seq_{(1,1)}$ is a classical combinatorial notion, names *bridges*. To generate $Cyc_{(1,1)}$, we use an isomorphism between the (1, 1)-balanced cycles and the cycles of indecomposable Dyck paths.

Let \mathcal{D} be the class of Dyck paths of specification $\mathcal{D} = Seq(\nearrow \mathcal{D} \searrow)$. Dyck paths are excursions from $(0, 0)$ to $(0, 2n)$ over the discrete lattice $\mathbb{Z}^+ \times \mathbb{Z}^+$, with displacements of $(1, 1)$ and $(1, -1)$. This can also be viewed as the class of well-formed parentheses strings. Bridges are defined by $Seq((\nearrow \mathcal{D} \searrow) + (\searrow \overline{\mathcal{D}} \nearrow))$, where $\overline{\mathcal{D}}$ is like \mathcal{D} , but with the roles of \searrow and \nearrow interchanged. So, we can generate $Seq_{(1,1)}$ by classical Boltzmann sampler principles.

The class of *indecomposable* Dyck paths is the class of specification $\nearrow \mathcal{D} \searrow$.

Proposition 2 (Raney's lemma). *The balanced cycle can also be specified as $Cyc_{(1,1)} \simeq Cyc(\nearrow \mathcal{D} \searrow)$, where Cyc is the classical constructor for cycles.*

Proof. (sketch) We can represent $Cyc_{(1,1)}$ as an excursion from $(0, 0)$ to $(0, 2n)$, with steps of $(1, 1)$ and $(1, -1)$, up to circular permutations. Such an excursion has a non-empty set S of points of minimal abscissa. As we deal with excursions up to circular permutations, we can consider that the excursion begin at any c in S (see Fig.2). In this case, we have Dyck paths and we remark that they are exactly the same up to circular permutations of their indecomposable Dyck paths.

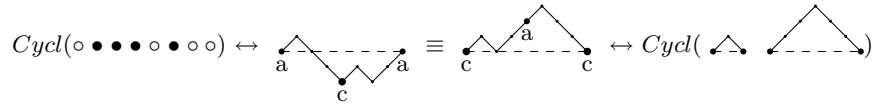


Fig. 2. Isomorphism between $Cyc_{(1,1)}$ and $Cyc(\nearrow \mathcal{D} \searrow)$.

Now, with this specification, we can use usual Boltzmann samplers for unlabelled structures to draw $(1, 1)$ -balanced cycles. We obtain a Boltzmann sampler for $Cyc(\nearrow \mathcal{D} \searrow)$ by a combination of the previous mentioned samplers according to the specification. In the following algorithms $D(x)$ and $C_{(1,1)}(x)$ respectively represent the generating functions of \mathcal{D} and $Cyc(\circ D \bullet)$.

Algorithm 1: $\Gamma_x \mathcal{D}$

Input: the parameter x

Output: an object of \mathcal{D}

- 1 Draw l according to a geometric law of parameter $x^2 D(x)$
 - 2 $M := \varepsilon$
 - 3 **for** i **from** 1 **to** l **do**
 - 4 $\lfloor M := \text{concat}(M, \circ, \Gamma_x \mathcal{D}, \bullet)$
 - 5 **return** (M) .
-

Algorithm 2: $\Gamma_x Cyc_{(1,1)}$

Input: the parameter x

Output: a $(1, 1)$ -balanced cycle

- 1 Let K be a random variable in \mathbb{N}^* verifying :
 - 2 $\mathbb{P}(K = k) = \frac{-\varphi(k)}{k C_{(1,1)}(x)} \log(1 - x^{2k} D(x^k))$
 - 3 Draw k according to the law of K .
 - 4 Draw j according to a logarithmic law of parameter $x^{2k} D(x^k)$.
 - 5 $M := \varepsilon$
 - 6 **for** i **from** 1 **to** j **do**
 - 7 $\lfloor M := \text{concat}(M, \circ, \Gamma_x \mathcal{D}, \bullet)$
 - 8 **return** $\underbrace{[MM\dots M]}_{k \text{ times}}$.
-

This method is an efficient way to draw $(1, 1)$ -balanced cycles. In particular, the basic rejection sampler $\mu Cyc_{(1,1)}(x; n, \varepsilon)$ (see (7) for details) has an $O(n)$ overall cost in average. But it relies on a very singular property of the class : it can be decomposed with usual constructors. We are now interested in another way to generate these objects. This new approach will be extended to all v -balanced objects in the last section.

2.2 Second approach: mixed samplers

We are still focused on the generation of $(1, 1)$ -balanced cycles, but now the use of an algebraic specification is avoided. The idea of our sampler can be summarized as follows : first, we adapt an ad-hoc sampler for the $(1, 1)$ -balanced sequences in such a way that this sampling follows a Boltzmann model; second, we show an isomorphism (see proposition 4) between the class $\Theta Cyc_{(1,1)}$ of pointed balanced cycles and a sum involving duplications of $Seq_{(1,1)}$. The notion of pointing classes is recalled in this part; finally, to obtain a Boltzmann sampler for $Cyc_{(1,1)}$, we

explain how to obtain a Boltzmann sampler for a class A from a Boltzmann sampler for its pointing class ΘA .

Sampler for $Seq_{(1,1)}$. Let us start by introducing our Boltzmann sampler for $Seq_{(1,1)}$ and proving its correctness:

Algorithm 3: $\Gamma_x Seq_{(1,1)}$

Input: the parameter x

Output: a balanced sequence of Z_b and Z_w .

- 1 Let L be a random variable in \mathbb{N}^* verifying $\mathbb{P}(L = l) = \frac{(2l)!}{(l!)^2} \frac{x^{2l}}{S_{(1,1)}(x)}$
 - 2 Draw l according to the law of L .
 - 3 Let M be a $(2l)$ -uple, select uniformly l positions belongs the $2l$ entries in M .
 - 4 This positions are the Z_b entries of M , the other ones are the Z_w entries.
 - 5 **return** (M) .
-

Lemma 1. *Algorithm 3 is a valid Boltzmann sampler for $Seq_{(1,1)}$.*

Proof. Let α be an output of this algorithm. The probability to draw α is the probability to draw the right length and then to draw the right positions for Z_b . So,

$$\mathbb{P}(\alpha) = \frac{|\alpha|!}{\left(\frac{|\alpha|}{2}\right)!^2} \frac{x^{|\alpha|}}{S_{(1,1)}(x)} \cdot \frac{\left(\frac{|\alpha|}{2}\right)!^2}{|\alpha|!} = \frac{x^{|\alpha|}}{S_{(1,1)}(x)}$$

An isomorphism for $\Theta Cyc_{(1,1)}$. Another classical operator, in structural combinatorics, is the pointing operator which can be defined as follows :

Definition 2. Let \mathcal{C} be a combinatorial class, the combinatorial class $\Theta \mathcal{C}$ is formally defined as $\sum_{n>0} \mathcal{C}_n \times \{\epsilon_1, \dots, \epsilon_n\}$ where the ϵ_i are distinct neutral objects (0 -sized objects) and \mathcal{C}_n is the sub-class of \mathcal{C} of the elements of size n .

The generating function of the pointed class $\Theta \mathcal{C}$ is $C^\bullet(x) = x \cdot \frac{dC}{dx}$. We can interpret this operator by saying that each object in $\Theta \mathcal{C}$ is an object in \mathcal{C} with a tagged atom on it.

Theorem 1. $\Theta Cyc_{(1,1)}$ is isomorphic to $\sum_{n>0} \varphi(n) Rep_n(Seq_{(1,1)})$, where $Rep_n(\mathcal{A})$ is the class $\{\underbrace{aa\dots a}_n; a \in \mathcal{A}\}$.

Proof. Let C_{bic} be the generating function for $Cyc(Z_w + Z_b)$, without any constraint on the number of beads (Z_w and Z_b) of each color. We can write the generating function C_{bic} as follows :

$$C_{bic} = \sum_{n>0} \frac{-\varphi(n)}{n} \log(1 - (Z_w^n + Z_b^n)) = \sum_{n>0} \frac{\varphi(n)}{n} \sum_{k>0} \frac{1}{k} (Z_b^n + Z_w^n)^k$$

$$C_{bic} = \sum_{n>0} \frac{\varphi(n)}{n} \sum_{k>0} \frac{1}{k} \sum_{p_1+p_2=k} \frac{k!}{p_1!p_2!} Z_b^{p_1n} Z_w^{p_2n}$$

The notion of diagonal for a bivariate generating function is needed for what follows. Let $f(X, Y) = \sum_i a_{i,j} X^i Y^j$ be a bivariate generating function, the function $g(Z) = \sum a_{n,n} Z^{2n}$ is called the *(1,1)-diagonal* of $f(X, Y)$ and it is denoted by Δf .

So, by definition $C_{(1,1)} = \Delta C_{bic}$ and we can use the previous formula to obtain :

$$C_{(1,1)} = \sum_{l>0} ([Z_b^l Z_w^l] C_{bic}) X^{2l} = \sum_{n>0} \frac{\varphi(n)}{n} \sum_{p>0} \frac{1}{2p} \frac{(2p)!}{(p!)^2} X^{2np}$$

Now, pointing the $Cyc(1,1)$ class yields :

$$C_{(1,1)}^\bullet = \sum_{n>0} \varphi(n) \sum_{p>0} \frac{(2p)!}{(p!)^2} X^{2np}$$

$$C_{(1,1)}^\bullet = \sum_{n>0} \varphi(n) S_{(1,1)}(X^n)$$

At this stage, it is possible to draw an object of size n in $Cyc_{(1,1)}$ using classical recursive method (12). But here we pitch on Boltzmann point of view which avoids costly preprocessing calculus.

This isomorphism allows us to describe a Boltzmann sampler for $\Theta Cyc_{(1,1)}$:

Algorithm 4: $\Gamma_x \Theta Cyc_{(1,1)}$

Input: the parameter x

Output: a $(1, 1)$ -balanced cycle of Z_b and Z_w .

```

1  $n := 1$ 
2  $S := \frac{\varphi(n)S_{(1,1)}(x^n)}{C_{(1,1)}^\bullet(x)}$ 
3 Draw a real number  $u$  uniformly in  $[0, 1]$ 
4 while  $u > S$  do
5    $n := n + 1$ 
6    $S := S + \frac{\varphi(n)S_{(1,1)}(x^n)}{C_{(1,1)}^\bullet(x)}$ 
7 return  $[\Gamma_x(Rep_n(Seq_{(1,1)}))]$ 

```

Corollary 1. *Algorithm 4 is a valid Boltzmann sampler for $\Theta Cyc_{(1,1)}$.*

Proof. This is a corollary of the correctness of our general sampler, given in section four.

A Boltzmann Sampler for $Cyc_{(1,1)}$. We have obtained a Boltzmann sampler for $\Theta Cyc_{(1,1)}$. This is enough to uniformly generate $(1,1)$ -balanced cycles. But, the sampler does not have a Boltzmann distribution for $Cyc_{(1,1)}$, so it can not be called by another constructor. For instance, $Cyc(\Theta Cyc_{(1,1)})$ is not equal to

$Cyc(Cyc_{(1,1)})$. Indeed, small objects are drawn with a smaller probability in $\Theta Cyc_{(1,1)}$. To unbiased the sampler, we are going to change its parameter according to a well-chosen density law $f_x(u)$. A similar idea occurs in (16).

Lemma 2. *Let $C(x) = \sum_{n>0} c_n x^n$ be a generating function (with $C(0) = 0$). For any fixed x in the convergence disc of C , the function $f_x(u) = \frac{C^\bullet(xu)}{uC(x)}$ is a density of probability on $[0, 1]$.*

Proof. Clearly f is non negative. Now, it remains only to prove that $\int_{u=0}^1 \frac{C^\bullet(xu)}{uC(x)} du = 1$. We can expand the serie to $C^\bullet(xu)$, $\int_{u=0}^1 \frac{\sum_{n>0} n c_n (ux)^n}{uC(x)} du$. We now swap the sum and integral, we have :

$$\sum_{n>0} \frac{n c_n x^n}{C(x)} \int_{u=0}^1 u^{n-1} du = \frac{1}{C(x)} \sum_{n>0} n c_n x^n \left[\frac{u^n}{n} \right]_{u=0}^1 = 1.$$

Theorem 2. *The following sampler (Algorithm 5) gives a valid Boltzmann sampler for \mathcal{C} with parameter x from a Boltzmann sampler for $\Theta \mathcal{C}$.*

Algorithm 5: $\Gamma_x \mathcal{C}$

Input: the parameter x

Output: an object in \mathcal{C} .

- 1 Draw a real number u according to the density law $\frac{C^\bullet(xu)}{u(C(x) - c_0)}$.
 - 2 **if** $(Bernoulli(\frac{c_0}{C(x)})) = 1$ **then**
 - 3 | **return** an object in \mathcal{C}_0 drawn uniformly.
 - 4 **else**
 - 5 | **return** $\Gamma_{ux}(\Theta \mathcal{C})$ and forget the point.
-

Proof. It is sufficient to evaluate the probability that the output be of size n . If $n = 0$, we have drawn an object in \mathcal{C}_0 . This occurs with probability $\frac{c_0}{C(x)}$. If

$n > 0$, the probability is $(1 - \frac{c_0}{C(x)}) \cdot \int_{u=0}^1 \frac{n c_n (ux)^n}{uC(x)} du = \frac{c_n x^n}{C(x)}$. In every cases, this is a Boltzmann probability.

This sampler allows us to generate extremely large $(1, 1)$ -cycles (more than 1000000 beads). Indeed, this sampler is clearly linear in the size of the output. The following figures (Fig.3) show a $(1, 1)$ -cycle of size 100 (Fig. 3(a)) (only 100 beads for the legibility.) It also shows that we can compose our sampler with the classical builder $(*, +, Seq, Cyc, MSet, PSet, \dots)$. Figure 3(b) shows a random generated a necklace of $(1, 1)$ -cycles. We can see that the necklaces do not contain a lot of $(1, 1)$ -cycles. Moreover only one of these $(1, 1)$ -cycles contains a lot of beads.

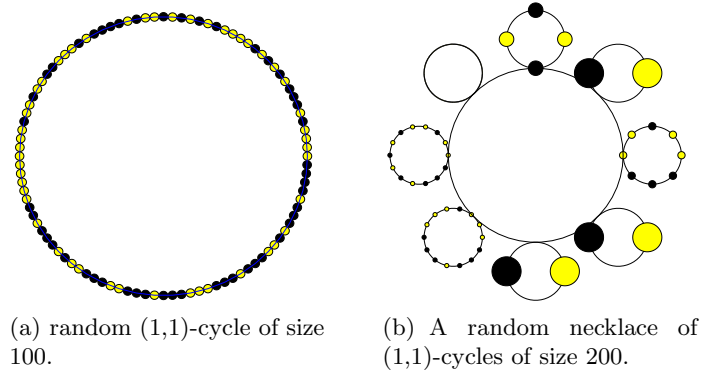


Fig. 3. Examples of Boltzmann sampling.

3 The general vectorial case

In this part, we extend the previous method to all cases, algebraic or not. Let $v = (v_1, v_2, \dots, v_m)$, and $|v| = \sum_{i=1}^m v_i$. Our goal here is to generate Cyc_v : cycles of m colors, such that the number of occurrences of each atom \mathcal{Z}_i verify the v -balance condition. We follow the same principles than in section 3.2 but the proofs are slightly more technical.

3.1 Sampler for Seq_v

Let us recall S_v , the generating function of v -balanced sequences:

$$S_v = \sum_{p>0} \frac{(|v|p)!}{m \prod_{i=1}^m ((v_i p)!)} \prod_{i=1}^m Z_i^{v_i p}.$$

Algorithm 6: $\Gamma_x Seq_v$

Input: the parameter t

Output: a v -balanced sequence of Z_i .

Let L be a random variable in \mathbb{N}^* verifying $\mathbb{P}(L = l) = \frac{(|v|l)!}{\prod_{i=1}^m (v_i l!)} t^{\sum_{i=1}^m v_i l} S_v$

- 1
 - 2 Draw l according to the law of L .
 - 3 Let M be a $(|v|l)$ -uple,
 - 4 **for** i **from** 1 **to** m **do**
 - 5 Select uniformly lv_i positions belongs the $(|v|l)$ entries not yet affected in M .
 - 6 This positions are the Z_i entries of M .
 - 7 **return** (M) .
-

Lemma 3. *Algorithm 6 is a valid Boltzmann sampler for Seq_v . Its arithmetic complexity is linear in the size of its output object.*

Proof. The proof can be easily transposed from $(1,1)$ -balanced one. The complexity result is trivial.

Now, as with the example of $(1,1)$ -balanced cycles, we are going to use this sampler for v -balanced sequences to generate v -balanced cycles.

3.2 An isomorphism for ΘCyc_v

Theorem 3. ΘCyc_v is isomorphic to $\sum_{n>0} \varphi(n) Rep_n(Seq_v)$.

Proof. Let C_{mcol} be the generating function of cycles of m atoms Z_1, \dots, Z_m .

$$C_{mcol} = \sum_{n>0} \frac{-\varphi(n)}{n} \log(1 - (\sum_{i=1}^m Z_i^n)) = \sum_{n>0} \frac{\varphi(n)}{n} \sum_{k>0} \frac{1}{k} (\sum_{i=1}^m Z_i^n)^k$$

$$C_{mcol} = \sum_{n>0} \frac{\varphi(n)}{n} \sum_{k>0} \frac{1}{k} \sum_{\sum_{i=1}^m p_i = k} \frac{k!}{\prod_{i=1}^m (p_i!)} \prod_{i=1}^m Z_i^{np_i}$$

Let C_v be the generating function of Cyc_v . This is the extraction of terms with the exponents verifying the v -balanced condition in C_{mcol} .

$$C_v = \sum_{l>0} \left(\prod_{i=1}^m Z_i^{lv_i} C_{mcol} \right) X^{v|l} = \sum_{n>0} \frac{\varphi(n)}{n} \sum_{p>0} \frac{1}{|v|p} \frac{(|v|p)!}{\prod_{i=1}^m ((v_i p)!)} X^{v|np}$$

We will now apply the same idea that we described for the (1,1)-balanced case to ΘC_v (the generating function of which is C_v^\bullet).

$$C_v^\bullet = \sum_{n>0} \varphi(n) \sum_{p>0} \frac{(|v|p)!}{\prod_{i=1}^m ((v_i p)!)} X^{v|np} = \sum_{n>0} \varphi(n) S_v(X^n)$$

This isomorphism can be used to obtain the following sampler:

Algorithm 7: $\Gamma_x \Theta Cyc_v$

Input: the parameter x

Output: a v -balanced cycle of Z_i .

```

1  $n := 1$ 
2  $S := \frac{\varphi(n) S_v(x^n)}{C_v^\bullet(x)}$ 
3 Draw a real number  $u$  uniformly in  $[0, 1]$ 
4 while  $u > S$  do
5    $n := n + 1$ 
6    $S := S + \frac{\varphi(n) S_v(x^n)}{C_v^\bullet(x)}$ 
7 return  $[\Gamma_x(Rep_n(Seq_{(1,1)}))]$ 

```

Proposition 3. *Algorithm 7 is a valid Boltzmann sampler for ΘCyc_v . Its arithmetic complexity is linear in the size of its output object.*

Proof. Let us consider the generation of a pointing cycle c .

It can be written as $c = u^p$, where u a primitive sequence (*i.e.* without replication) and p is the primitive repetition order of c ($|c| = p|u|$). There are $s = |u|$ shifts of u which produce equivalent cycles.

In Algorithm 7, the generating sequence is not necessarily primitive. So, c can be drawn as any \tilde{u}^d , with $d|p$ and \tilde{u} a shift of $\frac{p}{d}$ repetitions of u .

So, the probability to draw u is the sum for all $d|p$ of the probability to draw d as repetition order and to draw one of the s shift of \tilde{u} as motif :

$$\mathbb{P}(c) = \sum_{d|p} \frac{\varphi(d) Rep_d(S_v)}{C_v^\bullet} \cdot \frac{s \cdot (x^d)^{\frac{p}{d}}}{Rep_d(S_v)} = s \frac{x^{|c|}}{C_v^\bullet} \sum_{d|p} \varphi(d) = |c| \frac{x^{|c|}}{C_v^\bullet}$$

$\mathbb{P}(c, i) = \frac{1}{|c|} \mathbb{P}(c) = \frac{x^{|c|}}{C_v^\bullet}$ with $i \in [1, |c|]$, the choice of i corresponding to the pointed atom. The complexity ensues from the results on Boltzmann sampling.

To obtain a v -balanced cycle from an object of ΘCyc_v , we can now apply the general algorithm 5 to ΘCyc_v . As proven previously, this provides us a Boltzmann sampler for Cyc_v .

4 Conclusion

The random generation of constrained-colored structures is in general very difficult. In a previous paper (3), we already investigated the generation of the k -colored structures and size-colored structures. In this short paper, we have presented a way to efficiently generate v -balanced cycles. It is possible with our samplers to generate v -balanced cycles of sizes reaching up to one million. Nevertheless our methods can not be directly generalized to other balanced structures. For instance, we do not know how to generate $(1, 1)$ -balanced general non planar (unlabelled) trees where general non planar trees can be specified as $\mathcal{T} = \mathcal{Z.MSet}(\mathcal{T})$. This problem is a work in progress and should be solved by a method involving multivariate Boltzmann sampler. Another perspective is the generation of semi-labelled structures. In these structures each atom can take a color in $\{1, \dots, k\}$ but, if we have an atom of color $k > 1$, we need to also have at least one atom of color $k - 1$. Semi-labelling is a new interesting labelling which is in a sense between unlabelling and labelling. But very little is as of yet known about it.

Finally, we want to thank, B. Salvy for his precious knowledge on hypergeometric functions and Joachim Dehais, Jérémie Lumbroso and Yann Ponty for their careful reading of a preliminary version of the manuscript.

Bibliography

- [1] F. Beukers, G. Heckman, “ Monodromy for the hypergeometric function ${}_nF_{n-1}$ ”, in *Inventiones mathematicae*, vol 95; pp. 325-354.
- [2] O. Bodini, E. Fusy and C. Pivoteau, “Random sampling of plane partitions.” *Proceedings of Gascom’06*.
- [3] O. Bodini, A. Jacquot, “Boltzmann Samplers for Colored Combinatorial Objects”, in *Proceedings of Gascom’07*.
- [4] R. Brualdi and M. Newman, “An enumeration problem for a congruence equation”, *J. Res. Nat. Bureau Standards*, B74 (1970), 37-40.
- [5] K. Cattell, F. Ruskey, J. Sawada, C.R. Miers, and M. Serra, “Fast Algorithms to Generate Necklaces, Unlabelled Necklaces and Irreducible Polynomials over GF(2)”, *J. Algorithms*, 37 (2000) 267-282.
- [6] WYC. Chen, JD Louck, “Necklaces, MSS Sequences, and DNA Sequences”, *Adv. Appl. Math.* 18 (1997) pp.18-32
- [7] P. Duchon, P. Flajolet, G. Louchard and G. Schaeffer. “Boltzmann Samplers for the Random Generation of Combinatorial Structures”, In *Combinatorics, Probability, and Computing*, Special issue on Analysis of Algorithms, 2004, Vol. 13, No 4–5, pp. 577-625.
- [8] A. Darrasse, M. Soria, “Degree distribution of random Apollonian network structures and Boltzmann sampling”, *Proceedings of International Conference on Analysis of Algorithms DMTCS*, 2007.
- [9] P. Flajolet, E. Fusy and C. Pivoteau, “Boltzmann Sampling of Unlabelled Structures”, *Proceedings of ANALCO’07 (Analytic Combinatorics and Algorithms) Conference*, New Orleans, January 2007. SIAM Press, in print.
- [10] P. Flajolet and R. Sedgewick, “Analytic Combinatorics” published soon by Cambridge University Press.
- [11] P. Flajolet and M. Soria, private communication, 2008.
- [12] P. Flajolet, P. Zimmerman, B. Van Cutsem, “A calculus for the random generation of labelled combinatorial structures”, *Theor. Comput. Sci.* 132, 1-2 (Sep. 1994), 1-35.
- [13] H. Fredricksen, I.J. Kessler, “An algorithm for generating necklaces on beads in two colors”, *Discrete mathematics*, 1986, vol. 61, no2-3, pp. 181-188.
- [14] H. Fredricksen and J. Maiorana, “Necklaces of beads in k colors and k -ary de Bruijn sequences”, *Discrete Mathematics* 23 (1978), 207-210.
- [15] A. Nijenhuis and H. Wilf, “Combinatorial Algorithms”, Academic Press, ISBN 0125192509 (1975).
- [16] O. Roussel, ”A Boltzmann sampler for the box operator“, master thesis, 2008.