Random Graphs from Restricted Classes

Lecture Notes for the Summer School "Random Graphs, Geometry and Asymptotic Structure"

Lecture 2

Lecture notes by Konstantinos Panagiotou and Elisabetta Candellero

August 21, 2013

1 Boltzmann generation of Random Combinatorial Structures

Our aim is to describe a *general* purpose sampling algorithm for a given class of graphs.
Notation: we denote by $\mathcal{A}$ any class of graphs, and write $\mathcal{A} = \bigcup_{n\geq 0} \mathcal{A}_n$, where $\mathcal{A}_n$ consists of all elements of $\mathcal{A}$ that have n vertices. We will always assume that graphs are **well-labeled**, meaning that each graph in $\mathcal{A}_n$ has labels from the set $[n]$. If a graph G is not well-labeled, then we write $\rho(G)$ for the graph obtained by *relabeling* G canonically, hence $\rho(G)$ is well-labeled.

Remark 1.1. *A canonical relabeling is such that it preserves the order of the labels.*

For every $n \geq 0$, let us set $a_n = |\mathcal{A}_n|$. We will write

$$A(z) := \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n$$

for the **exponential generating function** of $\mathcal{A}$. Analogously to the class of trees, we can define the *Rooted Class* corresponding to $\mathcal{A}$, which we will denote by

$$\mathcal{A}_n^\bullet := \mathcal{A}_n \times [n].$$

The corresponding exponential generating function becomes

$$A^\bullet(x) := \sum_{n=0}^{\infty} \frac{n a_n}{n!} x^n = xA'(x).$$

1.1 Starting Point

The aim of this section is to look at possible *decompositions* and *recursive description of classes.* We start by defining the **Base Class** $\mathcal{X}$: this is the *single graph*, consisting of only one vertex.

Suppose we are given two classes of graphs $\mathcal{B}$ and $\mathcal{C}$ with exponential generating functions $B(x) := \sum_{n=0}^{\infty} b_n x^n/n!$ and $C(x) := \sum_{n=0}^{\infty} c_n x^n/n!$ respectively. Then, their **disjoint union** is defined as

$$\mathcal{A} = \mathcal{B} + \mathcal{C}, \quad \text{with} \quad \mathcal{B} \cap \mathcal{C} = \emptyset,$$

and the corresponding exponential generating function will be $A(x) = B(x)+C(x)$. Similarly, we can define the **labeled product** $\mathcal{A} = \mathcal{B} * \mathcal{C}$ as follows. Here $\mathcal{A}$ maps bijectively to the following set. Consider $(b, c) \in \mathcal{B} \times \mathcal{C}$ and label the resulting structure arbitrarily so that it is well-labeled. More formally, $\mathcal{A}$ is in bijection with

$$\bigcup_{b\in\mathcal{B},\, c\in\mathcal{C}} b * c,$$

where

$$b * c := \{(b', c') \ : \ (b', c') \text{ is well-labeled and } \rho(b') = b,\ \rho(c') = c\}.$$

Writing $b_k = |\mathcal{B}_k|$ and $c_k = |\mathcal{C}_k|$ we arrive at the relation

$$a_n = \sum_{k=0}^{n} b_k c_{n-k} \binom{n}{k},$$

implying $A(x) = B(x)C(x)$.

Remark 1.2. *The factor* $1/n!$ *in the definition of the exponential generating function is justified by the presence of labeling.*

Next we consider the **Set** construction. $\mathcal{A} = \mathrm{Set}(\mathcal{B})$ means that $\mathcal{A}$ is in bijection with an *unordered* collection of graphs in the class $\mathcal{B}$ (where again the labels can be distributed in any possible way). In particular, for every such collection with k graphs there are $k!$ *ordered* collections. Therefore,

$$A(x) = \sum_{k\geq 0} \frac{(B(x))^k}{k!} = e^{B(x)}.$$

The last construction that we consider is the so-called **substitution**, denoted by $\mathcal{A} = \mathcal{B} \circ \mathcal{C}$. For each $k \geq 0$, write $\mathrm{Set}_k(\mathcal{C})$ for the collection of sets in $\mathcal{C}$ with k elements, and $\mathcal{B}_k$ for the set of graphs in $\mathcal{B}$ with k vertices.

In other words, for each two classes of graphs $\mathcal{B}$ and $\mathcal{C}$, the class $\mathcal{A}$ consists of all graphs that are obtained from graphs from $\mathcal{B}$, where each node is replaced by a graph from $\mathcal{C}$.

Hence, we can write

$$\mathcal{A} = \sum_{k\geq 0} \mathcal{B}_k \times \mathrm{Set}_k\, \mathcal{C},$$

which implies $A(x) = \sum_k b_k \frac{(C(x))^k}{k!} = B(C(x))$.

1.2 Sampling

In this section we aim at answering the following question.

Question. *How do we sample randomly from a class that is specified in terms of "Base Class", "Disjoint Union", "Product", "Set", "Substitution"?*

In general, uniform sampling from $\mathcal{A}_n$ is difficult and/or computationally expensive, but there is an indirect way of approaching the problem. The idea is to proceed as we did in the case of trees, i.e., by *dropping* the requirement on the size of the sampler to be exactly n.

Assume that $A(x) < \infty$, and to each $a \in \mathcal{A}_n$ (i.e., $|a| = n$) assign the probability

$$\mathbb{P}_x(a) = \frac{x^{|a|}}{|a|!A(x)} = \frac{x^n}{n!A(x)}. \tag{1}$$

This is the so-called **Boltzmann model**, inspired from statistical physics (see [1]).

Remark 1.3. *Sampling from Base Class and finite classes is easy.*

Remark 1.4. *The probability defined in (1) only depends on the size of the element a, and not on a itself, hence (1) is* ***uniform*** *when conditioned on $|a| = n$.*

1.2.1 Boltzmann Sampler for the Disjoint Union $\mathcal{A} = \mathcal{B} + \mathcal{C}$

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

Now, suppose we have samplers $\Gamma B(x)$ and $\Gamma C(x)$, for the classes $\mathcal{B}$ and $\mathcal{C}$. The Boltzmann model associated to $A(x)$ is a mixture of those associated to $B(x)$ and $C(x)$. Given a Bernoulli generator defined as

$$\text{Bern}(p) = 1 \text{ with probability } p, \quad \text{Bern}(p) = 0 \text{ with probability } 1 - p,$$

the Boltzmann sampler $\Gamma A(x)$ is obtained by the following idea. Given the values of $B(x)$ and $C(x)$, we can consider the outcome of

$$\text{Bern}\left(\frac{B(x)}{B(x) + C(x)}\right) = \text{Bern}\left(\frac{B(x)}{A(x)}\right).$$

and choose whether we sample from $\mathcal{A}$ of from $\mathcal{B}$. We proceed as follows.

$$\begin{aligned} \Gamma A(x) : \quad & \textbf{if} \quad \text{Bern}\left(\frac{B(x)}{A(x)}\right) = 1, \quad \textbf{then} \quad \text{set } \gamma := \Gamma B(x) \\ & \textbf{else} \quad \text{set } \gamma := \Gamma C(x) \\ & \textbf{return } \gamma. \end{aligned}$$

Now let $b \in \mathcal{B}$. Since we are dealing with Bernoulli random variables, the probability of choosing a graph (in one "coin-flip") in $\mathcal{B}$ among the ones in $\mathcal{A}$ is given by $B(x)/A(x)$. Therefore, in view of (1), we have that

$$\mathbb{P}(\Gamma A(x) = b) = \frac{B(x)}{A(x)}\, \mathbb{P}(\Gamma B(x) = b) = \frac{B(x)}{A(x)}\, \frac{x^{|b|}}{|b|!B(x)} = \frac{x^{|b|}}{|b|!A(x)}.$$

The same calculation can be done to show that for every $c \in \mathcal{C}$ we have

$$\mathbb{P}(\Gamma A(x) = c) = \frac{x^{|c|}}{|c|!A(x)}.$$

1.2.2 Boltzmann Sampler for the Labeled Product $\mathcal{A} = \mathcal{B} * \mathcal{C}$

In a similar way as above, we assume that we have Boltzmann samplers for $\mathcal{B}$ and $\mathcal{C}$, and define (recursively) the Boltzmann sampler $\Gamma A(x)$.

$\Gamma A(x)$: $(\Gamma B(x), \Gamma C(x))$ –Note: the components are *independent* of each other–
$\gamma_B := \Gamma B(x)$
$\gamma_C := \Gamma C(x)$
return (γ_B, γ_C) with a randomly chosen relabeling.

Let $a \in \mathcal{A}$. The definition of the labeled product implies that there are $b \in \mathcal{B}$ and $c \in \mathcal{C}$ such that a maps bijectively to some well-labeled (b', c'), where

$$\rho(b') = b, \quad \text{and} \quad \rho(c') = c.$$

Therefore

$$\begin{aligned}
\mathbb{P}(\Gamma A(x) = a) &= \mathbb{P}(\Gamma B(x) = b)\, \mathbb{P}(\Gamma C(x) = c)\, \mathbb{P}(\text{relabeling is the correct one}) \\
&= \mathbb{P}(\Gamma B(x) = b) \mathbb{P}(\Gamma C(x) = c) \left[\binom{|b| + |c|}{|b|} \right]^{-1} \\
&= \frac{x^{|b|}}{|b|!B(x)}\, \frac{x^{|c|}}{|c|!C(x)}\, \frac{|b|!\,|c|!}{(|b| + |c|)!} \\
&= \frac{x^{|b|+|c|}}{B(x)C(x)}\, \frac{1}{(|b| + |c|)!} = \frac{x^{|a|}}{|a|!A(x)}.
\end{aligned}$$

1.2.3 Boltzmann Sampler for the set construction

In the Boltzmann model, the chance that a random graph from $\mathcal{A}$ is a collection of k elements of $\mathcal{B}$ is given by

$$\frac{B(x)^k}{k!}\, \frac{1}{A(x)} = e^{-B(x)}\, \frac{B(x)^k}{k!}.$$

Since this is a Poisson distribution, the procedure to define $\Gamma A(x)$ is as follows:

$$\begin{aligned} \Gamma A(x)\ :\ & k \ \text{ is a random variable } \sim \mathrm{Po}(B(x)) \\ & \left.\begin{aligned} \gamma_1 &:= \Gamma B(x) \\ \vdots \quad & \quad \vdots \\ \gamma_k &:= \Gamma B(x) \end{aligned}\right\} \text{independent copies!} \\ & \textbf{return } (\gamma_1, \dots, \gamma_k) \text{ with a randomly chosen labeling.} \end{aligned}$$

Boltzmann Sampler for Substitution $\mathcal{A} = \mathcal{B} \circ \mathcal{C}$: Exercise!

Exercises

Exercise 1. *Given the Substitution construction $\mathcal{A} = \mathcal{B} \circ \mathcal{C}$:*

(i) Find its generating function.

(ii) Develop a Boltzmann Sampler.

Exercise 2. *Let $\mathcal{B}$ be the class of binary trees (each vertex has either 0 or 2 children).*

(i) Compute the generating function for $\mathcal{B}$.

(ii) Develop a Boltzmann sampler for $\mathcal{B}$.

(iii) Denote by N_x the size of the Boltzmann sampler. Show that the following hold:

$$\mathbb{E}(N_x) = \frac{xB'(x)}{B(x)};$$
$$\mathrm{Var}(N_x) = \frac{x^2 B''(x) + xB'(x)}{B(x)} - \left(\frac{xB'(x)}{B(x)}\right)^2.$$

Exercise 3. *Let $\mathcal{T}^\bullet$ be the class of rooted labeled trees.*

(i) Define $\mathcal{T}^\bullet$ in terms of the product and set constructions.

(ii) Show that the sampler from Lecture 1 is actually a Boltzmann sampler for $\mathcal{T}^\bullet$.

References

[1] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combinatorics, Probability and Computing*, 13(4-5):577625, 2004.