Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

Electronic Notes in Discrete Mathematics 59 (2017) 65–77

www.elsevier.com/locate/endm

Improvements to exact Boltzmann sampling using probabilistic divide-and-conquer and the recursive method

Stephen DeSalvo [1]

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095
USA

Abstract

We demonstrate an approach for exact sampling of certain discrete combinatorial distributions, which is a hybrid of exact Boltzmann sampling and the recursive method, using probabilistic divide-and-conquer (PDC). The approach specializes to exact Boltzmann sampling in the trivial setting, and specializes to PDC deterministic second half in the first non-trivial application. A large class of examples is given for which this method broadly applies, and several examples are worked out explicitly.

Keywords: exact sampling, perfect simulation, probabilistic divide-and-conquer, Boltzmann sampler, random combinatorial structure, rejection sampling, recursive method.

[1] Email: `stephendesalvo@math.ucla.edu`

http://dx.doi.org/10.1016/j.endm.2017.05.006
1571-0653/

1 Introduction

The Boltzmann sampler [7] has transformed the way in which combinatorial structures are analyzed and randomly sampled, by taking advantage of the generating function structure. One starts with a family of combinatorial objects, $\mathcal{C}$, parameterized by various statistics like size and weight, and writes $\mathcal{C}$ as a countably infinite disjoint union of *finite* sets, for example,

$$\mathcal{C} = \bigcup_n \mathcal{C}_n = \bigcup_n \bigcup_k \mathcal{C}_{n,k}.$$

We may have n and k represent, for example, certain statistics like the weight of an integer partition and the number of parts, respectively, and $\mathcal{C}_{n,k}$ is the set of all integer partitions of size n into exactly k parts. One then samples from this set of objects via the form of the generating function.

For unlabelled structures, an object of size n is generated with probability, for some given real-valued x,

$$\Pr\left(\text{random object is of size } n\right) = \frac{c_n\, x^n}{C(x)},$$

where $C(x)$ is the generating function. For labelled structures, an object of size n is generated with probability, for some given real-valued x,

$$\Pr\left(\text{random object is of size } n\right) = \frac{c_n\, x^n}{n!\, \widehat{C}(x)},$$

where $\widehat{C}(x)$ is the exponential generating function.

A particularly general method for the random sampling of combinatorial structures is the recursive method of Nijenhuis and Wilf [9]. The method samples components one at a time in proportion to its prevalence in the overall target set, by constructing a table of values, which is equivalent to forming a conditional probability distribution. Once this table is complete, sampling may be considered to be $O(1)$ *random bits* per sample (assuming fixed floating point, otherwise $O(\log p_{k,n})$, where $p_{k,n}$ is the number of objects with weight n into exactly k components), since the majority of the work is then about applying the unranking algorithm, which is arithmetical in nature, and does not require any auxiliary randomness. The main drawback of this method is that the table size may be overwhelming, and often only a small portion of the table is utilized, even though the full table is needed in principle.