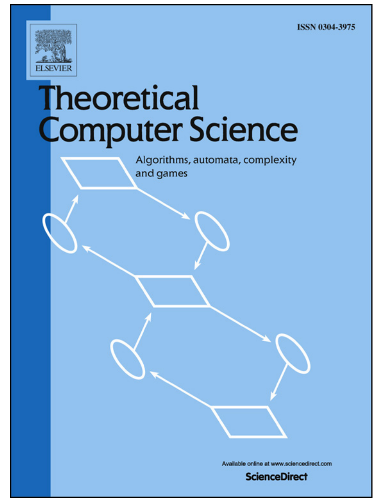
Accepted Manuscript



Efficient random sampling of binary and unary-binary trees *via* holonomic equations

Axel Bacher, Olivier Bodini, Alice Jacquot

PII: S0304-3975(17)30547-9
DOI: http://dx.doi.org/10.1016/j.tcs.2017.07.009
Reference: TCS 11249

To appear in: *Theoretical Computer Science*

Received date: 13 January 2016
Revised date: 10 July 2017
Accepted date: 14 July 2017

Please cite this article in press as: A. Bacher et al., Efficient random sampling of binary and unary-binary trees *via* holonomic equations, *Theoret. Comput. Sci.* (2017), http://dx.doi.org/10.1016/j.tcs.2017.07.009

Efficient random sampling of binary and unary-binary trees *via* holonomic equations

Axel Bacher† Olivier Bodini‡ Alice Jacquot‡

July 18, 2017

† : Research Institute for Symbolic Computation (RISC), Johannes Kepler University, Altenbergerstraße 69, A-4040 Linz, Austria. E-mail: abacher@risc.jku.at

‡ :Université Paris 13, Sorbonne Paris Cité, LIPN, CNRS(UMR 7030), F-93430, Villetaneuse, France. E-mail: bodini@lipn.fr, jacquot@lipn.fr.

Abstract

We present a new uniform random sampler for binary trees with n internal nodes consuming $2n + \Theta(\log(n)^2)$ random bits on average. This makes it quasi-optimal and out-performs the classical Remy algorithm. We also present a sampler for unary-binary trees with n nodes taking $\Theta(n)$ random bits on average. Both are the first linear-time algorithms to be optimal up to a constant.

Keywords: Random Sampling, Remy Algorithm, Holonomic Equation, Plane Trees

1 Introduction

Rooted Plane trees are central data structures in computer science, and have been widely studied in both mathematics and computer science. They are a natural way of representing hierarchy and arise in a huge number of algorithms such as depth first search, sorting algorithms and search algorithms. In this paper, we focus on uniform random sampling of Catalan and Motzkin trees (*i.e.*, rooted plane binary tees and rooted plane unary-binary tees). More precisely, we address the search of a algorithm which returns uniformly a tree having n nodes with minimal cost in terms of random bit, while keeping good time and space complexities. Contrariwise with the uniform-real-variable model where the measure of complexity is the number of calls to iid uniform real variables U on the compact $[0, 1]$, we deal with the realistic and tractable random-bit model introduced by Von Neumann and further developed by Knuth and Yao. In this model, the unit of complexity is the random bit. For obvious reasons, this notion is much more natural in computer science and coherent with Shannon information entropy [13]. This notion have already been investigated [9, 6] It is clear that the uniform random sampler of an object γ inside a set of cardinality C needs at least $\ln_2(C)$ random bits.

The efficient (in time space and random bits) sampling of trees is of central interest in many domain. Since these structures appear in numerous settings