Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 75–76. Available online at the URL
`http://algo.inria.fr/seminars/`.

Profile of Random Recursive Trees and Random Binary Search Trees

Hsien-Kuei Hwang
Institute of Statistical Science, Academia Sinica (Taiwan)

April 26, 2004

Summary by Brigitte Chauvin and Jean-Maxime Labarbe

1. Introduction

Random recursive trees model is a simple probability model useful for many applications as system generation, spread of contamination of organisms, internet interface map, stochastic growth of networks or statistical physics. A random recursive tree is constructed as follows: one starts from a root node with the label 1; at each step $n, n \geq 2$, a new node with the label n is attached uniformly at random to one of the previous nodes labelled by $1, 2, \ldots, n-1$. In this model, the labels are increasing along any path from the root to a node.

Denote by $X_{n,k}$ the number of nodes at distance k from the root in a random recursive tree with n nodes. We are interested in the *profile* of random recursive trees, i.e., the collection $\{ X_{n,k} : k \in \mathbb{N} \}$. More precisely, there exist some results about the mean and the variance of the $X_{n,k}$ but the limit distribution of $X_{n,k}$ scaled by its mean, in the range $k/\log n \sim$ Constant is a source of intriguing phenomena. Notice that the profile provides a fine and informative shape characteristic, it is related to path length, depth, height, width, … and also to generation of random trees or other algorithmic problems.

A well-known connection between random recursive trees and binary search trees is the following: by rotation, a planar tree gives a binary tree and then it appears that the profile of random recursive trees is exactly the "left" profile of binary search trees (meaning that the distance of a node is only counted for left branches). Consequently, it is of the same flavour to study the profile of binary search trees; results and phenomena for $Y_{n,k}$, the number of external nodes at level k and $Z_{n,k}$, the number of internal nodes at level k in a random binary search tree, appear as a simple transposition.

2. Main Results

Let k depend on n, let $\alpha_{n,k} = k/\log n$ and suppose that:

$$\lim_{n\to\infty} \alpha_{n,k} = \alpha$$

The main phenomena are summarized in the following proposition:

Proposition 1 (Main phenomena)**.**

- $\mathbf{E}(X_{n,k})$ *is unimodal and* $\mathbf{Var}(X_{n,k})$ *is bimodal,*
- $\forall \alpha \in [0, e),\ X_{n,k}/\,\mathbf{E}(X_{n,k}) \xrightarrow{d} X_\alpha$ *(convergence in distribution)*
- $\forall \alpha \in [0, 1],\ X_{n,k}/\,\mathbf{E}(X_{n,k}) \xrightarrow{m} X_\alpha$ *(convergence of all moments)*
- *for* $k = o(\log n)$*, (case* $\alpha = 0$*),* $(X_{n,k} - \mathbf{E}(X_{n,k}))/\sqrt{\mathbf{Var}(X_{n,k})} \xrightarrow{m} \mathcal{N}(0,1)$

– *for* $k = \log n + o(\log n)$ *(case* $\alpha = 1$*) and* $|k - \log n| \to \infty$,
$(X_{n,k} - \mathbf{E}(X_{n,k}))/\sqrt{\mathbf{Var}(X_{n,k})} \xrightarrow{m} X_1'$
– *for* $k = \log n + O(1)$, $(X_{n,k} - \mathbf{E}(X_{n,k}))/\sqrt{\mathbf{Var}(X_{n,k})}$ *does not converge in distribution.*

where X_α *and* X_1' *are limit distributions we describe further.*

The proof is based both on the contraction method and the moment method.

Let $\mu_{n,k}$ be the first moment of $X_{n,k}$. A fine study of the asymptotics of $\mu_{n,k}$ leads to

$$\text{for } 0 \le \alpha < e, \qquad \frac{\log \mu_{n,k}}{\log n} \longrightarrow \alpha - \alpha \log \alpha.$$

The second moment and the variance of $X_{n,k}$ can also be asymptotically described: if $0 \le \alpha < 2$,

$$\mathbf{Var}(X_{n,k}) \sim \Big(\frac{\Gamma(\alpha+1)^2}{(1-\alpha/2)\Gamma(2\alpha+1)} - 1\Big)\mu_{n,k}^2$$

and the variance exhibits a *bimodal behavior* when $\alpha = 1$.

For the second point in the proposition, i.e., the limit distribution, the starting point is the recurrence formula satisfied by the $X_{n,k}$ (a branching-type property):

$$X_{n,k} \stackrel{d}{=} X_{U_n,k-1} + X^*_{n-U_n,k} \tag{1}$$

where U_n is uniform over $\{0, \ldots, n-1\}$ and $X_{j,k}$ and $X^*_{j',k'}$ are independent of each other and independent of U_n. As usually, thanks to the asymptotics of the first moment $\mu_{n,k}$ of $X_{n,k}$ and because U_n/n converges in distribution to a uniform distribution U on the interval $[0,1]$, it is possible to deduce a limit equation from (1):

$$X_\alpha \stackrel{d}{=} \alpha U^\alpha X_\alpha + (1-U)^\alpha X^*_\alpha.$$

Moreover, the convergence of the first m moments is obtained for $0 \le \alpha < m^{1/(m-1)}$ and the moments of the limit distribution $\nu_m := \mathbf{E}(X_\alpha^m)$ are given by a recurrence relation:

$$\nu_m = \frac{1}{m - \alpha^{m-1}} \sum_{j=1}^{m} \binom{m}{j} \nu_j \nu_{m-j} \alpha^{j-1} \frac{\Gamma(j\alpha+1)\Gamma((m-j)\alpha+1)}{\Gamma(m\alpha+1)}.$$

In the particular case when $\alpha = 1$, the convergence of all moments in the fifth point of the Proposition is obtained by the method of moments: all moments satisfy the same type of recurrence:

$$a_{n,k} = b_{n,k} + \frac{1}{n-1}\sum_{j=1}^{n-1}(a_{j,k-1} + a_{j,k})$$

so that generating functions techniques and transfer theorem allow one to get asymptotics for higher moments. Notice that the limit distribution X_1' is nothing but $(dX_\alpha/d\alpha)|_{\alpha=1}$ and is a solution of a 'quicksort'-type equation:

$$X_1' \stackrel{d}{=} UX_1' + (1-U)X'^*_1 + U + U\log U + (1-U)\log(1-U).$$

Among open questions:

– what happens at the boundary of the interval (α_-, α_+) of convergence of binary search trees (analog of the interval $(0, e)$ for recursive trees)?
– how to prove a.s. convergence in general?
– how to plot or simulate limit laws like X_α?