Branching processes, random trees and Brownian excursion

Vladimir Vatutin
Steklov Mathematical Institute (Moscow)

September 17, 1992

[summary by Guy Louchard]

Abstract

Using bijections between genealogical trees arising from branching processes, plane trees and classical random walks, one can derive asymptotic distributions of random variables such as tree height, distances related to the nearest mutual ancestor and labelled trees properties.

1. A few definitions

– $\mathcal{K}_N$:= the set of all plane trees K with $N+1$ vertices

$$|\mathcal{K}_N| = \frac{1}{N+1}\binom{2N}{N} \qquad \text{(Catalan number)}$$

– Let $A \subset \{0, 1, \ldots\}$,

$$\mathcal{K}_N^{(A)} := \{K \in \mathcal{K}_N\colon \text{the degree of any non-rooted vertex of K belongs to } A\}.$$

– $\mathcal{F}_N$:= the set of all rooted labelled trees with N non-rooted vertices

$$|\mathcal{F}_N| = (N+1)^{N-1}.$$

To any plane tree $K \in \mathcal{K}_{N+1}$ having m_r vertices of degree r, there corresponds $\frac{N!}{\prod_r (r!)^{m_r}}$ rooted labelled trees, with

$$\begin{cases} m_1 + 2m_2 + \cdots + Nm_N = N, \\ m_0 + m_1 + \cdots + m_N = N+1. \end{cases}$$

2. Branching processes

Let $f(s)$ be the generating function of the number ξ of direct descendants of an individual in a Galton-Watson branching process: $f(s) := \sum_k P(\xi = k)s^k$ and let $Z(n)$ be the number of individuals at time n.

There is an obvious bijection δ between any genealogical tree G and the corresponding plane tree K.

If Ω_G is the set of all genealogical trees, $f(s)$ generates a probability measure on the set of subsets of Ω_G and also a corresponding measure on the set of subsets of Ω_K (the set of all plane trees). If $\nu(G)$ is the total number of individuals in G then

$$P[G = G_N|\nu(G) = N] = \frac{p_0^{m_0}\cdots p_{N-1}^{m_{N-1}}}{P[\nu(G) = N]} \tag{1}$$

and by the bijection δ, this is equivalent to $P_\delta[K = K_N|K \in \mathcal{K}_N]$.

A few particular cases of $f(s)$ are given by:

– $f_1(s) = \frac{1}{2-s}$. Then (1) gives $\frac{1}{|\mathcal{K}_N|}$.

- $f_2(s) = \sum_{k\in A} c\alpha^k s^k$ with $f(1) = 1$, $f'(1) = 1$. Then (1) leads to

$$P_\delta[K = K_n | K \in \mathcal{K}_N^{(A)}] = \frac{1}{|\mathcal{K}_N^{(A)}|}.$$

- $f_3(s) = e^{s-1}$. Then (1) leads to

$$\frac{N!}{\prod_r (r!)^{m_r}} \frac{1}{(N+1)^{N-1}}. \tag{2}$$

3. Labelled trees

If $T_N \in \mathcal{F}_N$, set $\gamma :=$ the operation of removing labels and adding a root, so $K_{N+1} = \gamma(T_N) \in \mathcal{K}_{N+1}$. One can prove that $P[T \in \mathcal{F}_N : \gamma(T) = K_{N+1}] \equiv (2)$. So one can investigate the properties of T_N which are invariant with respect to relabelling by investigating the branching process with $f(s) = f_3(s)$. The same analysis holds with $f_4(s) = \sum_{k\in A} \frac{c}{k!} s^k$ and for $\mathcal{F}_N^{(A)}$.

4. Trees height

Let $\tau(G) = \min[n : Z(n) = 0]$ (extinction time). Then the height of K, $H(K)$ is given by $H(K) = \tau[\delta^{-1}(K)]$. One can prove that, if $f'(1) = 1, 0 \le f''(1) = B$, then

$$\lim P\left[\frac{\tau}{\sqrt{N}} \ge x | \nu = N\right] = P\left[\max_{0\le t\le 1} W_0^+(t) \ge \frac{x}{2}\sqrt{B}\right] \tag{3}$$

where $W_0^+(t)$ is the standard Brownian Excursion (B.E.) on $[0,1]$.

For example $f_1(s)$ leads to $B = 2$. If we set $n = x\sqrt{N}$, (3) gives $P[H(K) \ge n]$. An equivalent of the density can also be derived.

If one analyzes the standard random walk (R.W.) until the next return to 0, the local time at height j is equivalent in probability to the branching process $Z(j)$ with $f_1(s)$. Also f_1 leads to the Catalan statistic for plane trees and to the classical relation between asymptotic height and maximum of the B.E.

However, if $\frac{n}{\sqrt{N}} \to \infty$, with $\frac{n}{N} \le a < 1$, (3) is no longer true. It is however possible to find a function $h(n, N)$ such that

$$P[H(K) \ge n | K \in \mathcal{H}_N] \sim 4\frac{n^2}{N} \frac{\exp[-h(n,N)]}{1 - n^2/N^2}.$$

5. Nearest mutual ancestor

Let $\lambda(n, G)$ be the distance to the nearest mutual ancestor of all individuals living at time $n-1$ (given that $Z(n-1) > 0$). Let us condition on $\nu(G) = N$ and use $f_1(s)$. Call $\zeta(n, K)$ the corresponding quantity in the tree K.

One can derive $P[\zeta(n,K)/\sqrt{N} \le a | K \in \mathcal{K}_{n,N}]$ if $\frac{n}{\sqrt{N}} \to \beta$. Also $\zeta(n, K) \le na$ iff the number of upcrossings of the strip $[(n-1)a, n]$ in the corresponding R.W. is given by (1). The same arguments can be generalized to the total number of subtrees of K having their roots on level $(1-a)n$ and containing all vertices of given height n of K.

6. Generalizations

One can consider generation dependent branching processes, processes with some life length distribution, etc.