Shapes of Binary Trees

STEVEN FINCH

June 24, 2004

This is a sequel to our treatment of various attributes of trees [1], expressed in
the language of probability. Let \( \{ \Phi_\tau : 0 \leq \tau \leq 1 \} \) be standard Brownian excursion.

Define the \( \Phi_\tau \)-norm

\[
\| Y \|_p = \begin{cases} 
\left( \int_0^1 |Y_\tau|^p \, d\tau \right)^{1/p} & \text{if } 0 < p < \infty, \\
\max_{0 \leq \tau \leq 1} |Y_\tau| & \text{if } p = \infty
\end{cases}
\]

and a (new) seminorm

\[
\langle Y \rangle_p = \begin{cases} 
\left( \int_0^1 \int_0^v \left| Y_u + Y_v - 2 \min_{u \leq \tau \leq v} Y_\tau \right|^p \, du \, dv \right)^{1/p} & \text{if } 0 < p < \infty, \\
\max_{0 \leq u < v \leq 1} \left| Y_u + Y_v - 2 \min_{u \leq \tau \leq v} Y_\tau \right| & \text{if } p = \infty.
\end{cases}
\]

We examined \( \| Y \|_p \) earlier [2]; \( \langle Y \rangle_p \) is a less familiar random variable but nevertheless important in the study of trees. Note that \( \langle Y \rangle_p \) is not a norm since, for any constant \( c \), \( \langle c \rangle_p = 0 \) even if \( c \neq 0 \).

Let \( T \) be an ordered (strongly) binary tree with \( N = 2n+1 \) vertices. The distance between two vertices of \( T \) is the number of edges in the shortest path connecting them. The height of a vertex is the number of edges in the shortest path connecting the vertex and the root.

The Wiener index \( d_1(T) \) is the sum of all \( \binom{N}{2} \) distances between pairs of distinct vertices of \( T \), and the diameter \( d_\infty(T) \) is the maximum such distance. If \( \delta(v, w) \) denotes the distance between vertices \( v \) and \( w \), then

\[
d_\lambda(T) = \left( \frac{1}{2} \sum_{v, w} \delta(v, w)^\lambda \right)^{1/\lambda}, \quad \lambda > 0,
\]

includes both the Wiener index and diameter as special cases.

\(^0\)Copyright © 2004 by Steven R. Finch. All rights reserved.
The **internal path length** $h_1(T)$ of a tree is the sum of all $N$ heights of vertices of $T$, and the **height** $h_\infty(T)$ is the maximum such height. Let $o$ denote the root of $T$. The generalization

$$h_\lambda(T) = \left( \sum_v \delta(v, o)^\lambda \right)^{1/\lambda}, \quad \lambda > 0,$$

includes both the internal path length and height as special cases. If we restrict attention to only those $n + 1$ vertices $v_k$ that are leaves (terminal nodes) of $T$, listed from left to right, then a sequence $\delta(\hat{v}_1, o), \delta(\hat{v}_2, o), \ldots, \delta(\hat{v}_{n+1}, o)$ emerges. This is called the **contour** of $T$.

The **width** $w_\infty(T)$ of a tree is the maximum of $\sum \delta(\hat{v}, o)$ over all $\lambda \geq 0$, where $\delta(\hat{v}, o)$ is the number of vertices of height $l$ in $T$. Note that

$$w_\lambda(T) = \left( \sum_{l=0}^{h_\infty(T)} \zeta_l(T)^\lambda \right)^{1/\lambda}, \quad \lambda > 0,$$

includes the trivial case $w_1(T) = N$. The sequence $\zeta_0(T), \zeta_1(T), \ldots, \zeta_{h_\infty}(T)$ is known as the **profile** of $T$.

### 0.1. Uniform Combinatorial Model.

In this model, the $(2^n)/(n+1)$ ordered binary trees are weighted with equal probability, where $N = 2n + 1$ is fixed.

Janson [3] determined the joint distribution of internal path length and Wiener index:

$$\left( \frac{h_1(T)}{2N^{3/2}}, \frac{d_1(T)}{2N^{5/2}} \right) \rightarrow (\|Y\|_1, \langle Y \rangle_1)$$

as $N \to \infty$. The marginal distribution of $h_1(T)$ was obtained earlier by Takács [4, 5, 6]; the result for $d_1(T)$ is apparently new. No explicit formula for $P(\langle Y \rangle_1 \leq x)$ is known; see [2] for the corresponding result for $\|Y\|_1$. We have expected values

$$E(\|Y\|_1) = \frac{1}{2}\sqrt{\frac{\pi}{2}}, \quad E(\langle Y \rangle_1) = \frac{1}{4}\sqrt{\frac{\pi}{2}}$$

and correlation coefficient

$$\frac{\text{Cov}(\|Y\|_1, \langle Y \rangle_1)}{\sqrt{\text{Var}(\|Y\|_1)} \sqrt{\text{Var}(\langle Y \rangle_1)}} = \sqrt{\frac{48 - 15\pi}{50 - 15\pi}} = 0.5519206030\ldots$$

As an aside, we mention that $\|Y\|_1 - \langle Y \rangle_1 \geq 0$ always. Underlying the joint moment [3]

$$E(\|Y\|_1^k (\|Y\|_1 - \langle Y \rangle_1)^l) = \frac{k!! \sqrt{\pi}}{2^{(7k+9l-4)/2} \Gamma((3k + 5l - 1)/2)} a_{k,l}$$
is the following interesting quadratic recursion [7, 8, 9, 10, 11, 12]:

\[ a_{k,l} = 2(3k + 5l - 4)a_{k-1,l} + 2(3k + 5l - 6)(3k + 5l - 4)a_{k,l-1} + \sum_{0<i+j<k+l} a_{i,j}a_{k-i,l-j} \]

with \( a_{0,0} = -1/2, a_{1,0} = 1 = a_{0,1} \) and \( a_{k,l} = 0 \) when \( k < 0 \) or \( l < 0 \). All \( a_{k,l} \) but \( a_{0,0} \) are positive integers when \( k \geq 0 \) and \( l \geq 0 \). Applications include the enumeration of connected graphs with \( n \) vertices and \( n + m \) edges. We have asymptotics [3, 13]

\[ a_{k,0} \sim \frac{1}{2\pi} 6^k (k-1)!, \quad a_{0,l} \sim C \cdot 50^l ((l-1)!)^2, \]

where the precise identity of the constant

\[ C = 1/50 \cdot (0.981038...) = 0.01962... = \frac{1}{50.9664...} \]

remains an unsolved problem.

Chassaing, Marckert & Yor [14] determined the joint distribution of height and width:

\[ \left( \frac{h_\infty(T)}{N^{1/2}}, \frac{w_\infty(T)}{N^{1/2}} \right) \rightarrow \left( \int_0^1 \frac{dt}{Y_t} \|Y\|_\infty \right) \]

as \( N \to \infty \). The marginal distribution of height was obtained earlier by Rényi & Szekeres and Stepanov [15, 16, 17, 18, 19, 20, 21, 22, 23, 24]; earlier works on width include [25, 26, 27, 28, 29, 30]. It turns out that the marginal distributions are identical (up to a factor of 2) and that this is the first of several theta distributions [31] we will see here:

\[ P \left( \frac{1}{2} \int_0^1 \frac{dt}{Y_t} \leq x \right) = P \left( \|Y\|_\infty \leq x \right) = \frac{\sqrt{2\pi} 5^{5/2}}{\pi x^3} \sum_{k=1}^{\infty} k^2 e^{-\pi^2 k^2/(2x^2)}. \]

The expected values thus coincide:

\[ E \left( \frac{1}{2} \int_0^1 \frac{dt}{Y_t} \right) = E \left( \|Y\|_\infty \right) = \sqrt{\frac{\pi}{2}}. \]

Rényi & Szekeres also computed the location of the maximum of the probability density [15]:

\[ \text{mode} (\|Y\|_\infty) = \frac{1}{2} \left( 2.3151543618... \right) = \frac{1}{2} \sqrt{\frac{2}{0.3731385248...}}. \]
Returning to the joint distribution formula, it is clear that \( h_\infty(T) \) and \( w_\infty(T) \) are negatively correlated. A numerical estimate (let alone an exact expression) for the correlation coefficient evidently remains open [14, 32].

For the generalized height and diameter parameters, we have marginal distributions [3, 14, 33, 34, 35]:

\[
\frac{h_\lambda(T)}{2N(\lambda + 2)/(2\lambda)} \to \| Y \|_\lambda, \quad \frac{d_\lambda(T)}{2N(\lambda + 4)/(2\lambda)} \to \langle Y \rangle_\lambda
\]
as \( N \to \infty \). The latter includes the special cases of Wiener index \((\lambda = 1, \text{ as mentioned before})\) and diameter \((\lambda = \infty)\):

\[
P(\langle Y \rangle_\infty \leq x) = \frac{1024\sqrt{2\pi}^{5/2}}{3x^9} \sum_{k=1}^{\infty} k^2 \left[ (3 + \pi^2 k^2) x^4 - 36\pi^2 k^2 x^2 + 64\pi^4 k^4 \right] e^{-8k^2/x^2},
\]
which possesses expected value

\[
E(\langle Y \rangle_\infty) = \frac{4}{3} \sqrt{2\pi}
\]
and maximum location [33]

\[
\text{mode}(\langle Y \rangle_\infty) = 3.2015131492... = \sqrt{\frac{8}{0.7805116813...}}.
\]
Nothing is known for other values of \( \lambda \) (even \( \lambda = 2 \) seems to have been neglected). It would also be good to learn the value of the correlation coefficient of \( d_\infty(T) \) and \( h_\infty(T) \), or of \( d_\infty(T) \) and \( w_\infty(T) \).

Consider finally the minimum height \( \eta(T) \) of a leaf, that is,

\[
\eta(T) = \min_{1 \leq k \leq n+1} \delta(\hat{\eta}_k, 0),
\]
and the height \( \delta(\hat{\eta}_{[n/2]}, 0) \) of the central leaf. It is known that

\[
E(\eta) \to \sum_{k=1}^{\infty} 2^{k+1-2^k} = 1.5629882961...
\]
as \( N \to \infty \) [36, 37]. It is also known that [38, 39, 40, 41]

\[
\sqrt{n} P \left( \frac{\delta(\hat{\eta}_{[n/2]}, 0)}{\sqrt{n}} \leq x \right) \to \frac{1}{2\sqrt{\pi}} \int_0^x t^2 e^{-t^2/4} dt = P \left( \sqrt{X_1^2 + X_2^2 + X_3^2} \leq \frac{x}{\sqrt{2}} \right),
\]
the Maxwell distribution from thermodynamics, where \( X_1, X_2, X_3 \) are independent standard normal variables. (This can also be written in terms of the chi square distribution with 3 degrees of freedom.) Can these results be related to Brownian excursion in some way? We will report more on the properties of leaves of \( T \) later.
0.2. **Critical Galton-Watson Model.** In this model, the size $N = 2n + 1$ is free to vary: All ordered binary trees are included but with weighting $2^{-N}$. (We omit subcritical and supercritical cases for reasons of space.)

Let $T$ be a random tree. The probability that $T$ has precisely $N$ vertices is clearly

$$
\frac{1}{n+1} \left( \frac{2n}{n} \right) 2^{-N} \sim \sqrt{\frac{2}{\pi}} N^{-3/2},
$$

hence the expected number of vertices of $T$ is infinite. We examine this result in another way. If

$$
\nu_l = \sum_{k=0}^{l} \zeta_k
$$

where $\zeta_k$ is the number of vertices of height $k$ in $T$, then $E(\nu_l) = l + 1$ and $\text{Var}(\nu_l) = (2l + 1)(l + 1)l/6$, both which $\to \infty$ as $l \to \infty$. More complicated conditional distributions are due to Pakes [43, 44]:

$$
\lim_{l \to \infty} \mathbb{P}\left( \frac{\nu_l}{l^2} \leq x \mid \zeta_l > 0 \right) = \int_{0}^{x} f(t) \, dt,
$$

$$
\lim_{l \to \infty} \mathbb{P}\left( \frac{\nu_l}{l^2} \leq x \mid \zeta_m > 0 \text{ for all positive integers } m \right) = \int_{0}^{x} g(t) \, dt,
$$

where the first density function is given by

$$
f(t) = \frac{2}{\sqrt{2\pi t^{3/2}}} \sum_{k=0}^{\infty} \left( \frac{(2k + 1)^2}{t} - 1 \right) \exp\left( -\frac{(2k + 1)^2}{2t} \right)
$$

with mean $1/3$, variance $2/45$, and Laplace transform

$$
\int_{0}^{\infty} e^{-st} f(t) \, dt = \sqrt{2s} \text{csch} \left( \sqrt{2s} \right).
$$

The second density function is not explicitly known, but has mean $1/2$, variance $1/12$ and satisfies

$$
\int_{0}^{\infty} e^{-st} g(t) \, dt = \text{sech}^2 \left( \sqrt{\frac{s}{2}} \right).
$$
Consequently \( g(t) \) is the convolution of \( \hat{g}(t) \) with itself, where
\[
\hat{g}(t) = \frac{1}{\sqrt{2\pi t}^{3/2}} \sum_{k=0}^{\infty} (-1)^k (2k + 1) \exp \left( -\frac{(2k + 1)^2}{8t} \right),
\]
but this appears to be as far as we can go.

Define \( T_l \) to be the subtree of \( T \) consisting of all \( \nu_l \) vertices up to and including height \( l \). We have the parameters \( d_\lambda(T_l), h_\lambda(T_l) \) and \( w_\lambda(T_l) \) available for study, but little seems to be known. Of course, \( w_1(T_l) = \nu_l \). Athreya [45], building on [46, 47, 48], proved that \( E(\hat{\nu}_\infty(T_l)) \sim \ln(l) \) as \( l \to \infty \), which contrasts nicely with the fact that \( P(\nu_\infty = 0) \to 1 \) as \( \nu \to \infty \). See also [49, 50, 51, 52, 53, 54, 55]. Kesten, Ney & Spitzer [56, 57, 58] demonstrated that \( P(h_\infty(T_l) = j) \sim 2/j^2 \) as \( j \to \infty \); further references include [59, 60, 61]. Can exact distributional results be found? What about other values of \( \lambda \)? Is anything known about diameter for Galton-Watson trees?

Just as the limit behavior for the uniform model is related to Brownian excursion, the limit behavior for the critical GW model is related to what is known as the two-sided three-dimensional Bessel process \( \{\mathbb{B}_t: -\infty < t < \infty\} \). That is, \( \{\mathbb{B}_t: t \geq 0\} \) and \( \{\mathbb{B}_{-t}: t \geq 0\} \) are independent copies of standard 3D radial Brownian motion \( \sqrt{W_1^2 + W_2^2 + W_3^2} \), each starting from zero [35, 62]. It would be good to learn more about the concrete distributional results arising from this correspondence.

**0.3. Leaves of Maximum Height.** Our closing remarks are concerned not with binary trees, but instead with labeled rooted trees. Choose such a tree \( T \) with \( \nu \) vertices uniformly out of the \( \nu^{N-2} \) possibilities (we agree that the root is labeled 1). Out of all possible parameters (suitably generalized), we mention only the minimum height \( \eta(T) \) of a leaf. Meir & Moon [37] computed that
\[
E(\eta) \to 9 \sum_{k=1}^{\infty} \frac{1}{4^k(1 + 2 \cdot 4^{-k})^2} = 1.6229713847...
\]
as \( N \to \infty \). A more difficult problem involves counting the leaves \( \hat{\nu}_k \) at prescribed distance from the root. Kesten & Pittel [63] proved, for leaves of maximum height, that there exists a probability distribution \( q_l \) such that
\[
\lim_{N \to \infty} P(\zeta_{h_\infty}(T) = l) = q_l, \quad l \geq 1.
\]
Further, \( q_l \) is the unique nonnegative solution of the system of equations
\[
l! e^l q_l = \sum_{k=1}^{\infty} k^l q_k, \quad \sum_{k=1}^{\infty} q_k = 1
\]
and thus \( q_1 = 0.602... \), \( q_2 = 0.248... \), \( q_3 = 0.094... \), \( q_4 = 0.035... \) with mean 1.636... and standard deviation 0.995... No exact expressions for these quantities are known. What is the corresponding distribution for the uniform ordered binary tree case?
0.4. Addendum. Janson \cite{64} computed that
\[
E \left( \int_0^1 (1/Y_t) \, dt \cdot \|Y\|_\infty \right) = 1 + \sum_{m=1}^\infty \frac{\ln [m(m+1)]}{m(m+1)} \\
= 1 + 2.0462774528... \\
= \pi - 0.0953152007..., \\
\]
\[
\frac{\Cov \left( \int_0^1 (1/Y_t) \, dt, \|Y\|_\infty \right)}{\sqrt{\Var \left( \int_0^1 (1/Y_t) \, dt \right) \sqrt{\Var (\|Y\|_\infty)}}} = \frac{3(3.0462774528... - \pi)}{\pi(\pi - 3)} \\
= -0.6428251027...
\]
and the infinite series \cite{65, 66} is a Lüroth analog of Lévy’s constant $\pi^2/(6 \ln(2))$. Why is the joint distribution of height and width of trees related to the ergodic theory of numbers? (Coincidences as such do not happen without a reason.) Another recent result is
\[
C = \frac{\sqrt{15}}{20\pi^2} = 0.0196207628...
\]
in connection with the asymptotics of the sequence $a_{0,t}$, due to Kotesovec \cite{12}.

REFERENCES


[32] C. Donati-Martin, Some remarks about the identity in law for the Bessel bridge $r_0 \frac{ds}{r(s)} \overset{(law)}{=} \sup_{s \leq 1} r(s)$, Studia Sci. Math. Hungar. 37 (2001) 131–144; MR1834327 (2002j:60141).


