Wiener Index and Dependencies in Random Digital Trees

(joint with Hsien-Kuei Hwang and Chung-Kuei Lee)

Michael Fuchs

Department of Applied Mathematics
National Chiao Tung University

June 21st, 2016

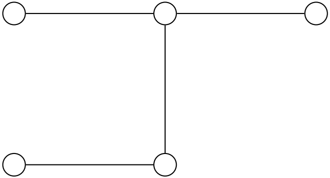
The Wiener Index

Introduced by H. Wiener in 1947 to investigate the boiling point of alkanes.

The Wiener Index

Introduced by H. Wiener in 1947 to investigate the boiling point of alkanes.

Given a graph:

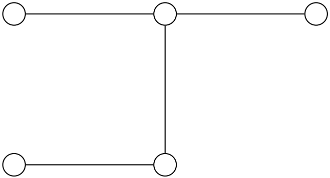


The Wiener index is the sum of distances between all unordered pairs of nodes.

The Wiener Index

Introduced by H. Wiener in 1947 to investigate the boiling point of alkanes.

Given a graph:



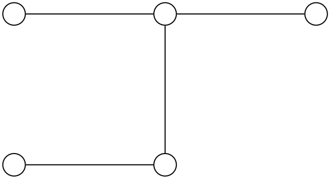
Wiener index=18

The Wiener index is the sum of distances between all unordered pairs of nodes.

The Wiener Index

Introduced by H. Wiener in 1947 to investigate the boiling point of alkanes.

Given a graph:



Wiener index=18

The Wiener index is the sum of distances between all unordered pairs of nodes.

In this talk, we will consider the Wiener index of rooted trees (trees arise as molecular graphs of acyclic organic molecules).

Families of Random Trees

There are many families of random trees:

Families of Random Trees

There are many families of random trees:

- Random plane trees;
- Random non-plane trees;
- Random binary trees;
- Random binary search trees;
- Random median-of-$(2k+1)$ search trees;
- Random quadtrees;
- Random digital search trees;
- Etc.

Families of Random Trees

There are many families of random trees:

- Random plane trees;
- Random non-plane trees;
- Random binary trees;
- Random binary search trees;
- Random median-of-$(2k+1)$ search trees;
- Random quadtrees;
- Random digital search trees;
- Etc.

Question: How does the Wiener index behave for such random trees?

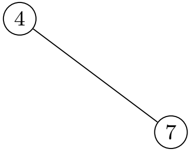
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$

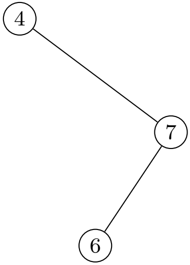
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



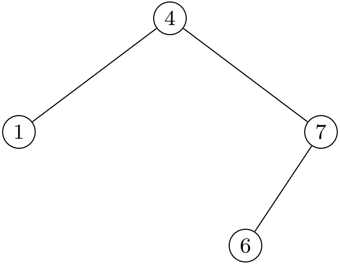
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



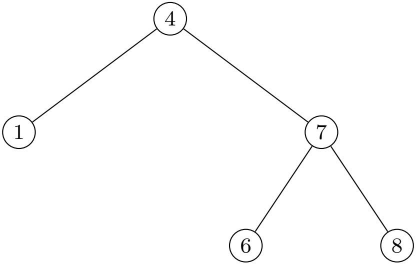
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



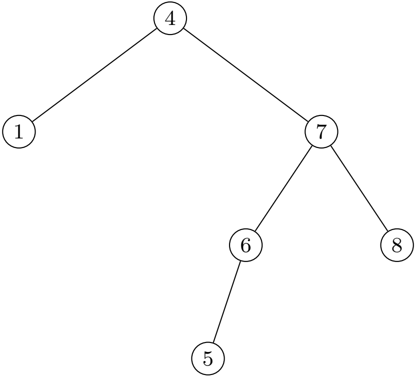
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



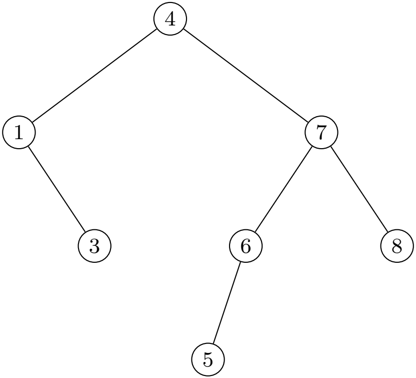
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



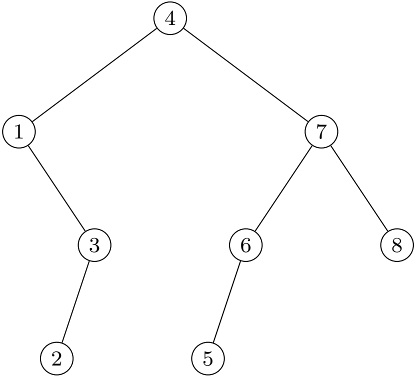
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



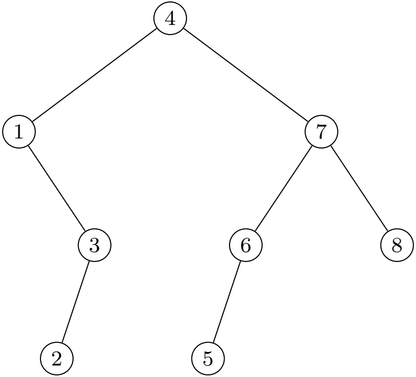
Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



Binary Search Trees

Example: Input: $4, 7, 6, 1, 8, 5, 3, 2$



Random model: Input is a random permutation of size n.

Moments of Wiener Index

T_n ... total path length.

W_n ... Wiener index.

Moments of Wiener Index

T_n ... total path length.

W_n ... Wiener index.

Theorem (Neininger 2002)

We have,

$$\mathbb{E}(W_n) \sim 2n^2 \log n$$

and

$$\mathrm{Var}(T_n) \sim \frac{21 - 2\pi^2}{3} n^2,$$
$$\mathrm{Cov}(T_n, W_n) \sim \frac{20 - 2\pi^2}{3} n^3,$$
$$\mathrm{Var}(W_n) \sim \frac{20 - 2\pi^2}{3} n^4.$$

Limit Law of Wiener Index

Theorem (Neininger 2002)

We have,

$$\left(\frac{T_n - \mathbb{E}(T_n)}{n}, \frac{W_n - \mathbb{E}(W_n)}{n^2}\right) \xrightarrow{d} (T, W),$$

where (T, W) *is a solution of*

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \stackrel{d}{=} A \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + B \begin{pmatrix} X_1^* \\ X_2^* \end{pmatrix} + \begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix}$$

with

$$A = \begin{pmatrix} 0 & U \\ U^2 & U(1-U) \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1-U \\ (1-U)^2 & U(1-U) \end{pmatrix}$$

and b_1^*, b_2^* *are functions of* U.

Random Split Trees (i)

Consider $b \geq 2$, $s > 0$ and s_0, s_1 with

$$0 \leq s_0 \leq s, \qquad 0 \leq bs_1 \leq s + 1 - s_0.$$

Random Split Trees (i)

Consider $b \geq 2$, $s > 0$ and s_0, s_1 with

$$0 \leq s_0 \leq s, \qquad 0 \leq bs_1 \leq s + 1 - s_0.$$

Moreover, consider a random vector

$$\mathbf{V} = (V_1, \dots, V_b) \in [0, 1]^b$$

with

$$\sum_{i=1}^{b} V_i = 1.$$

Random Split Trees (i)

Consider $b \geq 2$, $s > 0$ and s_0, s_1 with

$$0 \leq s_0 \leq s, \qquad 0 \leq bs_1 \leq s + 1 - s_0.$$

Moreover, consider a random vector

$$\mathbf{V} = (V_1, \ldots, V_b) \in [0,1]^b$$

with

$$\sum_{i=1}^{b} V_i = 1.$$

Assume that

$$V_i \stackrel{d}{=} V_1 := V \qquad 2 \leq i \leq b.$$

V is called *splitter*.

Random Split Trees (ii)

n balls are distributed to the infinite b-ary tree.

Random Split Trees (ii)

n balls are distributed to the infinite b-ary tree.

- If a ball is distributed to an internal node, choose the i-th subtree with probability V_i and move the ball to the chosen subtree, continue with the subtree;

Random Split Trees (ii)

n balls are distributed to the infinite b-ary tree.

- If a ball is distributed to an internal node, choose the i-th subtree with probability V_i and move the ball to the chosen subtree, continue with the subtree;
- If a ball is distributed to a leave containing $< s$ balls, put it there;

Random Split Trees (ii)

n balls are distributed to the infinite b-ary tree.

- If a ball is distributed to an internal node, choose the i-th subtree with probability V_i and move the ball to the chosen subtree, continue with the subtree;
- If a ball is distributed to a leave containing $< s$ balls, put it there;
- If a ball is distributed to a full leave, randomly put s_0 balls in the leave, randomly put s_1 balls in the subtrees, for the remaining balls choose a subtree according to the splitter, continue with the subtrees.

Random Split Trees (ii)

n balls are distributed to the infinite b-ary tree.

- If a ball is distributed to an internal node, choose the i-th subtree with probability V_i and move the ball to the chosen subtree, continue with the subtree;
- If a ball is distributed to a leave containing $< s$ balls, put it there;
- If a ball is distributed to a full leave, randomly put s_0 balls in the leave, randomly put s_1 balls in the subtrees, for the remaining balls choose a subtree according to the splitter, continue with the subtrees.

The resulting tree is called *random split tree of size* n.

Examples of Random Split Trees

Example 1: Binary search trees: $b = 2, s = s_0 = 1, s_1 = 0$ and V uniformly distributed on $[0, 1]$.

Examples of Random Split Trees

Example 1: Binary search trees: $b = 2, s = s_0 = 1, s_1 = 0$ and V uniformly distributed on $[0, 1]$.

If $b \geq 2$, $s = s_0 = b - 1, s_1 = 0$ and $V = \min\{U_1, \ldots, U_{b-1}\}$ with U_i uniformly distributed on $[0, 1]$, then b*-ary search trees*.

Examples of Random Split Trees

Example 1: Binary search trees: $b = 2, s = s_0 = 1, s_1 = 0$ and V uniformly distributed on $[0, 1]$.

If $b \geq 2$, $s = s_0 = b - 1, s_1 = 0$ and $V = \min\{U_1, \ldots, U_{b-1}\}$ with U_i uniformly distributed on $[0, 1]$, then b*-ary search trees*.

Example 2: Trie: $b = 2, s = 1, s_0 = s_1 = 0$ and V uniformly distributed on $\{p, 1 - p\}$ with $0 < p < 1$.

Examples of Random Split Trees

Example 1: Binary search trees: $b = 2, s = s_0 = 1, s_1 = 0$ and V uniformly distributed on $[0, 1]$.

If $b \geq 2$, $s = s_0 = b - 1, s_1 = 0$ and $V = \min\{U_1, \ldots, U_{b-1}\}$ with U_i uniformly distributed on $[0, 1]$, then b-*ary search trees*.

Example 2: Trie: $b = 2, s = 1, s_0 = s_1 = 0$ and V uniformly distributed on $\{p, 1 - p\}$ with $0 < p < 1$.

Assumption: V has Lebesgue density and the distribution function satisfies $F_V(x) < 1$ for all $x < 1$.

Examples of Random Split Trees

Example 1: Binary search trees: $b = 2, s = s_0 = 1, s_1 = 0$ and V uniformly distributed on $[0, 1]$.

If $b \geq 2$, $s = s_0 = b - 1, s_1 = 0$ and $V = \min\{U_1, \dots, U_{b-1}\}$ with U_i uniformly distributed on $[0, 1]$, then *b-ary search trees*.

Example 2: Trie: $b = 2, s = 1, s_0 = s_1 = 0$ and V uniformly distributed on $\{p, 1 - p\}$ with $0 < p < 1$.

Assumption: V has Lebesgue density and the distribution function satisfies $F_V(x) < 1$ for all $x < 1$.

Satisfied by Example 1 but NOT Example 2.

Moments of Wiener Index

Theorem (Munsonius 2012)

Under the assumption,

$$\mathbb{E}(W_n) \sim \frac{1}{\mu} n^2 \log n$$

with $\mu = -b\mathbb{E}(V \log V)$ *and*

$$\begin{aligned} \mathrm{Var}(T_n) &\sim \sigma_T^2 n^2, \\ \mathrm{Cov}(T_n, W_n) &\sim \sigma_C^2 n^3, \\ \mathrm{Var}(W_n) &\sim \sigma_W^2 n^4, \end{aligned}$$

where $\sigma_T^2, \sigma_C^2, \sigma_W^2 > 0$.

Limit Law of Wiener Index

Theorem (Munsonius 2012)

We have,

$$\left(\frac{T_n - \mathbb{E}(T_n)}{n}, \frac{W_n - \mathbb{E}(W_n)}{n^2}\right) \xrightarrow{d} (T, W),$$

where (T, W) is a solution of

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \stackrel{d}{=} \sum_{i=1}^{b} A_i \begin{pmatrix} X_1^{(i)} \\ X_2^{(i)} \end{pmatrix} + \begin{pmatrix} b_1^* \\ b_2^* \end{pmatrix}$$

with

$$A = \begin{pmatrix} 0 & V_i \\ V_i^2 & V_i(1 - V_i) \end{pmatrix},$$

and b_1^*, b_2^* are functions of the splitter.

Trie

Proposed by René de la Briandais in 1959.

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Example:

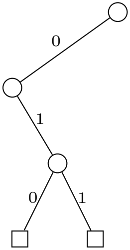
011011
010101
101110
010000
101010
001100

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Example:



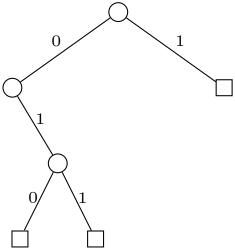
011011
010101
101110
010000
101010
001100

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Example:



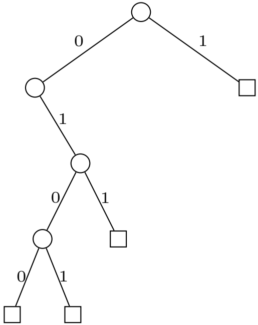
011011
010101
101110
010000
101010
001100

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Example:



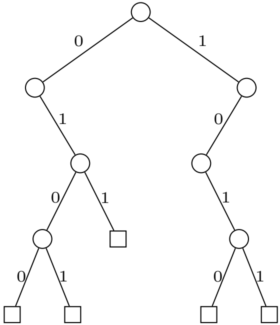
011011
010101
101110
010000
101010
001100

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Example:



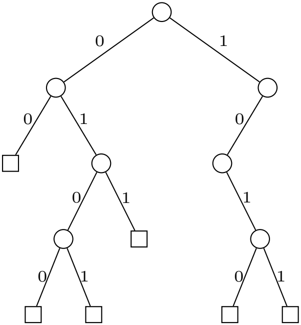
011011
010101
101110
010000
101010
001100

Trie

Proposed by René de la Briandais in 1959.

Name from the word data retrieval (suggested by Fredkin).

Example:



011011
010101
101110
010000
101010
001100

Why Studying Tries?

Why Studying Tries?

- Numerous applications.

  indexing sorted files, orthogonal range search, partial-match retrieval, pattern matching, approximate string matching, IP address or routing lookup, peer-to-peer lookup, data mining, dictionary-based syntactic pattern recognition, policy representations for network firewalls, syntactic pattern recognition, etc.

Why Studying Tries?

- Numerous applications.

  indexing sorted files, orthogonal range search, partial-match retrieval, pattern matching, approximate string matching, IP address or routing lookup, peer-to-peer lookup, data mining, dictionary-based syntactic pattern recognition, policy representations for network firewalls, syntactic pattern recognition, etc.

- Many variants and related data structures.

  Digital search trees, PATRICIA tries, radix sort, contention-resolution tree algorithms, multi-access broadcast channels, leader election algorithms, extendable hashing, polynomial factorization, etc.

Why Studying Tries?

- Numerous applications.

  indexing sorted files, orthogonal range search, partial-match retrieval, pattern matching, approximate string matching, IP address or routing lookup, peer-to-peer lookup, data mining, dictionary-based syntactic pattern recognition, policy representations for network firewalls, syntactic pattern recognition, etc.

- Many variants and related data structures.

  Digital search trees, PATRICIA tries, radix sort, contention-resolution tree algorithms, multi-access broadcast channels, leader election algorithms, extendable hashing, polynomial factorization, etc.

- Analysis of tries is interesting and challenging.

Notation and Random Model

Two types of nodes:

- Internal nodes: only used for branching;
- External nodes: nodes which hold data.

Notation and Random Model

Two types of nodes:

- Internal nodes: only used for branching;
- External nodes: nodes which hold data.

Random Model:

Bits are independent Bernoulli random variables with mean p.

Notation and Random Model

Two types of nodes:

- Internal nodes: only used for branching;
- External nodes: nodes which hold data.

Random Model:

Bits are independent Bernoulli random variables with mean p.

- $p = 1/2$: *symmetric digital search tree*;
- $p \neq 1/2$: *asymmetric digital search tree*.

Notation and Random Model

Two types of nodes:

- Internal nodes: only used for branching;
- External nodes: nodes which hold data.

Random Model:

Bits are independent Bernoulli random variables with mean p.

- $p = 1/2$: *symmetric digital search tree*;
- $p \neq 1/2$: *asymmetric digital search tree*.

Question: How does a random trie look like?

Additive Shape Parameter X_n

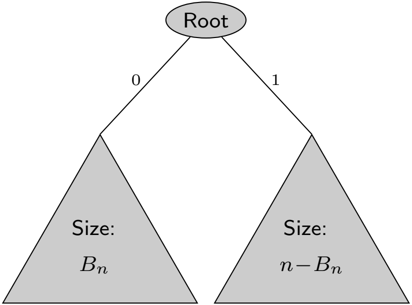
Computed recursively as follows: compute it for the two subtrees and add them up + add a toll.

Additive Shape Parameter X_n

Computed recursively as follows: compute it for the two subtrees and add them up + add a toll.

$$X_n \stackrel{d}{=} X_{B_n} + X^*_{n-B_n} + T_n$$

- $B_n \stackrel{d}{=} \text{Binomial}(n,p)$;
- $X_n \stackrel{d}{=} X^*_n$;
- X_n, X^*_n, B_n independent.
- T_n toll-function.



Size, Path Lengths and Wiener Indices

In this talk:

Size, Path Lengths and Wiener Indices

In this talk:

- **Size** S_n:

  number of internal nodes of a random trie;

Size, Path Lengths and Wiener Indices

In this talk:

- **Size** S_n:

  number of internal nodes of a random trie;

- **External Path Length** K_n:

  sum of distances of all external nodes to the root in a random trie;

Size, Path Lengths and Wiener Indices

In this talk:

- **Size** S_n:

  number of internal nodes of a random trie;

- **External Path Length** K_n:

  sum of distances of all external nodes to the root in a random trie;

- **Internal Path Length** N_n:

  sum of distances of all internal nodes to the root in a random trie;

Size, Path Lengths and Wiener Indices

In this talk:

- **Size** S_n:

  number of internal nodes of a random trie;

- **External Path Length** K_n:

  sum of distances of all external nodes to the root in a random trie;

- **Internal Path Length** N_n:

  sum of distances of all internal nodes to the root in a random trie;

- **External Wiener Index** KW_n:

  Wiener index of the external nodes in a random trie;

Size, Path Lengths and Wiener Indices

In this talk:

- **Size** S_n:

  number of internal nodes of a random trie;

- **External Path Length** K_n:

  sum of distances of all external nodes to the root in a random trie;

- **Internal Path Length** N_n:

  sum of distances of all internal nodes to the root in a random trie;

- **External Wiener Index** KW_n:

  Wiener index of the external nodes in a random trie;

- **Internal Wiener Index** NW_n:

  Wiener index of the internal nodes in a random trie.

Distributional Recurrences

Size:

$$S_n \stackrel{d}{=} S_{B_n} + S^*_{n-B_n} + 1.$$

Distributional Recurrences

Size:

$$S_n \stackrel{d}{=} S_{B_n} + S^*_{n-B_n} + 1.$$

Path Lengths:

$$K_n \stackrel{d}{=} K_{B_n} + K^*_{n-B_n} + n;$$
$$N_n \stackrel{d}{=} N_{B_n} + N^*_{n-B_n} + S_{B_n} + S^*_{n-B_n}.$$

Distributional Recurrences

Size:

$$S_n \stackrel{d}{=} S_{B_n} + S^*_{n-B_n} + 1.$$

Path Lengths:

$$K_n \stackrel{d}{=} K_{B_n} + K^*_{n-B_n} + n;$$

$$N_n \stackrel{d}{=} N_{B_n} + N^*_{n-B_n} + S_{B_n} + S^*_{n-B_n}.$$

Wiener Indices:

$$KW_n \stackrel{d}{=} KW_{B_n} + KW^*_{n-B_n} + B_n(K^*_{n-B_n} + n - B_n) + (n - B_n)(K_{B_n} + B_n);$$

$$NW_n \stackrel{d}{=} NW_{B_n} + NW^*_{n-B_n} + (S_{B_n} + 1)(N^*_{n-B_n} + S^*_{n-B_n}) + (S_{n-B_n} + 1)(N_{B_n} + S_{B_n}).$$

Mean and Variance - An Overview

Shape parameter	Mean	Variance
Size S_n	n	n
EPL K_n	$n\log n$	$\begin{cases} p\neq q: & n\log n \\ p=q: & n \end{cases}$
IPL N_n	$n\log n$	$n\log^2 n$
External Wiener Index KW_n	$n^2\log n$	$\begin{cases} p\neq q: & n^3\log n \\ p=q: & n^3 \end{cases}$
Internal Wiener Index NW_n	$n^2\log n$	$n^3\log^2 n$

Some Notation

We use the following notation:

Some Notation

We use the following notation:

- Entropy: $h = -p \log p - q \log q$;

Some Notation

We use the following notation:

- Entropy: $h = -p\log p - q\log q$;
- If $\log p/\log q \in \mathbb{Q}$, then

$$\frac{\log p}{\log q} = \frac{r}{\ell}, \qquad \gcd(r,\ell) = 1.$$

Some Notation

We use the following notation:

- Entropy: $h = -p\log p - q\log q$;
- If $\log p/\log q \in \mathbb{Q}$, then

$$\frac{\log p}{\log q} = \frac{r}{\ell}, \qquad \gcd(r,\ell) = 1.$$

and

$$\chi_k = \frac{2rk\pi i}{\log p}, \qquad (k \in \mathbb{Z}).$$

Some Notation

We use the following notation:

- Entropy: $h = -p\log p - q\log q$;
- If $\log p/\log q \in \mathbb{Q}$, then

$$\frac{\log p}{\log q} = \frac{r}{\ell}, \qquad \gcd(r,\ell) = 1.$$

and

$$\chi_k = \frac{2rk\pi i}{\log p}, \qquad (k \in \mathbb{Z}).$$

- For a function G:

$$\mathcal{F}[G](x) = \begin{cases} h^{-1}\sum_{k\in\mathbb{Z}} G(-1+\chi_k)e^{2k\pi i x}, & \text{if } \log p/\log q \in \mathbb{Q}; \\ h^{-1}G(-1), & \text{if } \log p/\log q \notin \mathbb{Q}. \end{cases}$$

Variance of Size S_n

Theorem (Régnier & Jacquet 1989; Kirschenhofer & Prodinger 1991; F., Hwang, Zacharovas 2014)

We have,

$$\mathrm{Var}(S_n) \sim \mathcal{F}[G_S](r \log_{1/p} n)n,$$

where

$$\begin{aligned} G_S(-1+\chi_k) =& \chi_k \Gamma(-1+\chi_k)\left(1 - \frac{\chi_k + 3}{2^{1+\chi_k}}\right) \\ & - \frac{1}{h}\sum_{j\in\mathbb{Z}} \Gamma(\chi_j + 1)\Gamma(\chi_{k-j}+1) \\ & - 2\sum_{j\geq 1} \frac{(-1)^j (j+1+\chi_k)\Gamma(j+\chi_k)\left(p^{j+1}+q^{j+1}\right)}{(j-1)!(j+1)(1-p^{j+1}-q^{j+1})}. \end{aligned}$$

Variance of EPL K_n

Theorem (Jacquet & Régnier 1986; Kirschenhofer, Prodinger, Szpankowski 1989; F., Hwang, Zacharovas 2014)

- $p \neq q$:

$$\mathrm{Var}(K_n) \sim h^{-3} pq \log^2(p/q) n \log n;$$

- $p = q$:

$$\mathrm{Var}(K_n) \sim \mathcal{F}[G_K](r \log_{1/p} n) n$$

Variance of EPL K_n

Theorem (Jacquet & Régnier 1986; Kirschenhofer, Prodinger, Szpankowski 1989; F., Hwang, Zacharovas 2014)

- $p \neq q$:

$$\mathrm{Var}(K_n) \sim h^{-3} pq \log^2(p/q) n \log n;$$

- $p = q$:

$$\mathrm{Var}(K_n) \sim \mathcal{F}[G_K](r \log_{1/p} n) n$$

where

$$G_K(-1+\chi_k) = \Gamma(\chi_k)\left(1 - \frac{\chi_k^2 - \chi_k + 4}{2^{\chi_k+2}}\right) + 2\sum_{\ell \geq 1} \frac{(-1)^\ell \Gamma(\chi_k+\ell)(\ell(\chi_k+\ell-1)-1)}{\ell!(2^\ell - 1)}.$$

Variance and Limit Law of IPL N_n

Theorem (F., Hwang, Zacharovas 2014)

We have,

$$\mathrm{Cov}(S_n, N_n) \sim h^{-1}\mathcal{F}[G_S](r\log_{1/p} n)n\log n$$

and

$$\mathrm{Var}(N_n) \sim h^{-2}\mathcal{F}[G_S](r\log_{1/p} n)n\log^2 n.$$

Variance and Limit Law of IPL N_n

Theorem (F., Hwang, Zacharovas 2014)

We have,

$$\mathrm{Cov}(S_n, N_n) \sim h^{-1}\mathcal{F}[G_S](r\log_{1/p} n)n\log n$$

and

$$\mathrm{Var}(N_n) \sim h^{-2}\mathcal{F}[G_S](r\log_{1/p} n)n\log^2 n.$$

Theorem (F. & Lee 2015)

We have,

$$\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}, \frac{N_n - \mathbb{E}(N_n)}{\sqrt{\mathrm{Var}(N_n)}}\right)^{\mathsf{T}} \xrightarrow{d} \mathcal{N}(0, E_2),$$

where E_2 is the 2×2 unit matrix.

External Wiener Index KW_n

Theorem (F. & Lee 2015)

- $p \neq q$:

$$\mathrm{Cov}(K_n, KW_n) \sim h^{-3} pq \log^2(p/q) n^2 \log n;$$
$$\mathrm{Var}(KW_n) \sim h^{-3} pq \log^2(p/q) n^3 \log n.$$

- $p = q$:

$$\mathrm{Cov}(K_n, KW_n) \sim \mathcal{F}[G_K](r \log_{1/p} n) n^2;$$
$$\mathrm{Var}(KW_n) \sim \mathcal{F}[G_K](r \log_{1/p} n) n^3$$

External Wiener Index KW_n

Theorem (F. & Lee 2015)

- $p \neq q$:

$$\mathrm{Cov}(K_n, KW_n) \sim h^{-3}pq\log^2(p/q)n^2\log n;$$
$$\mathrm{Var}(KW_n) \sim h^{-3}pq\log^2(p/q)n^3\log n.$$

- $p = q$:

$$\mathrm{Cov}(K_n, KW_n) \sim \mathcal{F}[G_K](r\log_{1/p} n)n^2;$$
$$\mathrm{Var}(KW_n) \sim \mathcal{F}[G_K](r\log_{1/p} n)n^3$$

and

$$\left(\frac{K_n - \mathbb{E}(K_n)}{\sqrt{\mathrm{Var}(K_n)}}, \frac{KW_n - \mathbb{E}(KW_n)}{\sqrt{\mathrm{Var}(KW_n)}}\right)^{\mathsf{T}} \xrightarrow{d} \mathcal{N}(0, E_2).$$

Internal Wiener Index NW_n

Theorem (F. & Lee 2015)

We have,

$$\mathrm{Cov}(S_n, NW_n) \sim 2h^{-1}\mathcal{F}[G_{\hat{S}}](r\log_{1/p} n)\mathcal{F}[G_S](r\log_{1/p} n)n^2\log n;$$

$$\mathrm{Cov}(N_n, NW_n) \sim 2h^{-2}\mathcal{F}[G_{\hat{S}}](r\log_{1/p} n)\mathcal{F}[G_S](r\log_{1/p} n)n^2\log^2 n;$$

$$\mathrm{Var}(NW_n) \sim 4h^{-2}(\mathcal{F}[G_{\hat{S}}](r\log_{1/p} n))^2\mathcal{F}[G_S](r\log_{1/p} n)n^3\log^2 n.$$

Internal Wiener Index NW_n

Theorem (F. & Lee 2015)

We have,

$$\mathrm{Cov}(S_n, NW_n) \sim 2h^{-1}\mathcal{F}[G_{\hat{S}}](r\log_{1/p} n)\mathcal{F}[G_S](r\log_{1/p} n)n^2 \log n;$$
$$\mathrm{Cov}(N_n, NW_n) \sim 2h^{-2}\mathcal{F}[G_{\hat{S}}](r\log_{1/p} n)\mathcal{F}[G_S](r\log_{1/p} n)n^2 \log^2 n;$$
$$\mathrm{Var}(NW_n) \sim 4h^{-2}(\mathcal{F}[G_{\hat{S}}](r\log_{1/p} n))^2\mathcal{F}[G_S](r\log_{1/p} n)n^3 \log^2 n.$$

Moreover,

$$\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}, \frac{N_n - \mathbb{E}(N_n)}{\sqrt{\mathrm{Var}(N_n)}}, \frac{NW_n - \mathbb{E}(NW_n)}{\sqrt{\mathrm{Var}(NW_n)}}\right)^{\mathsf{T}} \xrightarrow{d} \mathcal{N}(0, E_3),$$

where E_3 is the 3×3 unit matrix.

Size S_n and EPL K_n

Remark

We have,

$$\rho(K_n, KW_n) \sim 1$$

and

$$\begin{aligned} \rho(S_n, N_n) &\sim 1, \\ \rho(S_n, NW_n) &\sim 1, \\ \rho(N_n, NW_n) &\sim 1, \end{aligned}$$

where $\rho(\cdot,\cdot)$ denotes the correlation coefficient.

Size S_n and EPL K_n

Remark

We have,

$$\rho(K_n, KW_n) \sim 1$$

and

$$\begin{aligned} \rho(S_n, N_n) &\sim 1, \\ \rho(S_n, NW_n) &\sim 1, \\ \rho(N_n, NW_n) &\sim 1, \end{aligned}$$

where $\rho(\cdot, \cdot)$ *denotes the correlation coefficient.*

Question: how about the correlation between S_n and K_n?

Size S_n and EPL K_n

Remark

We have,

$$\rho(K_n, KW_n) \sim 1$$

and

$$\begin{aligned} \rho(S_n, N_n) &\sim 1, \\ \rho(S_n, NW_n) &\sim 1, \\ \rho(N_n, NW_n) &\sim 1, \end{aligned}$$

where $\rho(\cdot,\cdot)$ denotes the correlation coefficient.

Question: how about the correlation between S_n and K_n?

$\longrightarrow$ one expects strong positive correlation!

Covariance between S_n and K_n

Theorem (F. & Hwang 201?)

We have,

$$\mathrm{Cov}(S_n, K_n) \sim \mathcal{F}[G_{SK}](r \log_{1/p} n) n,$$

where

$$\begin{aligned} G_{SK}(-1+\chi_k) &= \Gamma(\chi_k)\left(1 - \frac{\chi_k + 2}{2^{\chi_k+1}}\right) \\ &\quad - \frac{1}{h} \sum_{j \in \mathbb{Z}\setminus\{0\}} \Gamma(\chi_{k-j}+1)(\chi_j - 1)\Gamma(\chi_j) \\ &\quad - \frac{\Gamma(\chi_k+1)}{h}\left(\gamma + 1 + \psi(\chi_k+1) - \frac{p\log^2 p + q \log^2 q}{2h}\right) \\ &\quad + \sum_{j \geq 2} \frac{(-1)^j (2j^2 - 2j + 1 + (2j-1)\chi_k)\Gamma(j-1\chi_k)(p^j + q^j)}{j!(1-p^j-q^j)}. \end{aligned}$$

Correlation Coefficient $\rho(S_n, K_n)$

Theorem (F. & Hwang 201?)

We have,

$$\rho(S_n, K_n) \sim \begin{cases} 0, & \text{if } p \neq q; \\ F(n), & \text{if } p = q, \end{cases}$$

where

$$F(n) = \frac{\mathcal{F}[G_{SK}](r \log_{1/p} n)}{\sqrt{\mathcal{F}[G_S](r \log_{1/p} n)\mathcal{F}[G_K](r \log_{1/p} n)}}$$

is a periodic function with

$$\textit{average value} = 0.927\cdots \ \textit{and amplitude} \leq 1.5 \times 10^{-5}.$$

Correlation Coefficient $\rho(S_n, K_n)$

Theorem (F. & Hwang 201?)

We have,

$$\rho(S_n, K_n) \sim \begin{cases} 0, & \text{if } p \neq q; \\ F(n), & \text{if } p = q, \end{cases}$$

where

$$F(n) = \frac{\mathcal{F}[G_{SK}](r \log_{1/p} n)}{\sqrt{\mathcal{F}[G_S](r \log_{1/p} n)\mathcal{F}[G_K](r \log_{1/p} n)}}$$

is a periodic function with

$$\text{average value} = 0.927\cdots \text{ and amplitude} \leq 1.5 \times 10^{-5}.$$

Question: can this behavior be ascribed to the weakness of Pearson's correlation coefficient?

Limit Laws

Theorem (F. & Hwang 201?)

- $p \neq q$:

$$\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}, \frac{K_n - \mathbb{E}(K_n)}{\sqrt{\mathrm{Var}(K_n)}}\right)^{\mathsf{T}} \xrightarrow{d} \mathcal{N}(0, I_2),$$

where I_2 is the 2×2 identity matrix.

Limit Laws

Theorem (F. & Hwang 201?)

- $p \neq q$:

$$\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}, \frac{K_n - \mathbb{E}(K_n)}{\sqrt{\mathrm{Var}(K_n)}}\right)^{\mathsf{T}} \xrightarrow{d} \mathcal{N}(0, I_2),$$

where I_2 is the 2×2 identity matrix.

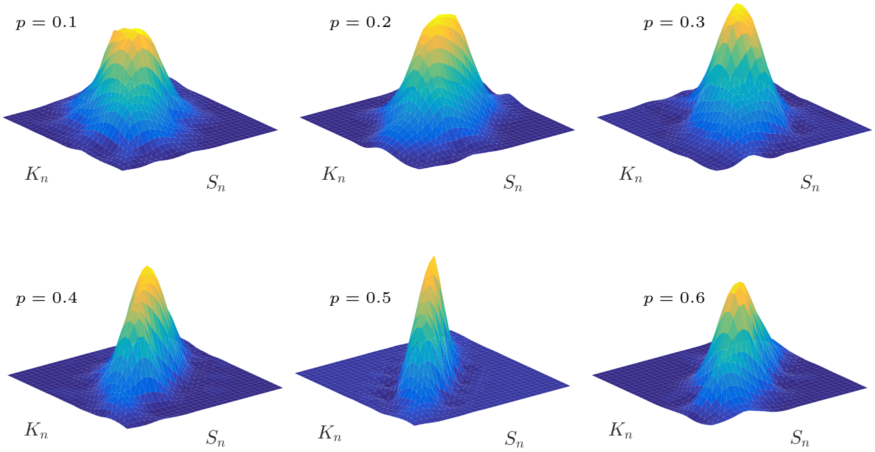
- $p = q$:

$$\Sigma_n^{-1/2} \begin{pmatrix} S_n - \mathbb{E}(S_n) \\ K_n - \mathbb{E}(K_n) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(0, I_2),$$

where Σ_n is the (asymptotic) covariance matrix:

$$\Sigma_n := n \begin{pmatrix} \mathscr{F}[G_S](r \log_{1/p} n) & \mathscr{F}[G_{SK}](r \log_{1/p} n) \\ \mathscr{F}[G_{SK}](r \log_{1/p} n) & \mathscr{F}[G_K](r \log_{1/p} n) \end{pmatrix}.$$

Joint Distribution of S_n and K_n



Summary

Summary

- Full analysis of Wiener index of tries, thereby completing the study of Wiener index of grid trees.

Summary

- Full analysis of Wiener index of tries, thereby completing the study of Wiener index of grid trees.

- Similar results for other digital trees:

  M. Fuchs and C.-K. Lee (2015). The Wiener index of random digital trees, *SIAM J. Discrete Math.*, **29:1**, 586–614.

Summary

- Full analysis of Wiener index of tries, thereby completing the study of Wiener index of grid trees.

- Similar results for other digital trees:

  M. Fuchs and C.-K. Lee (2015). The Wiener index of random digital trees, *SIAM J. Discrete Math.*, **29:1**, 586–614.

- Surprising correlation result for size and external path length in tries

Summary

- Full analysis of Wiener index of tries, thereby completing the study of Wiener index of grid trees.
- Similar results for other digital trees:

  M. Fuchs and C.-K. Lee (2015). The Wiener index of random digital trees, *SIAM J. Discrete Math.*, **29:1**, 586–614.
- Surprising correlation result for size and external path length in tries $\longrightarrow$ better probabilistic explanation?

Summary

- Full analysis of Wiener index of tries, thereby completing the study of Wiener index of grid trees.

- Similar results for other digital trees:

  M. Fuchs and C.-K. Lee (2015). The Wiener index of random digital trees, *SIAM J. Discrete Math.*, **29:1**, 586–614.

- Surprising correlation result for size and external path length in tries $\longrightarrow$ better probabilistic explanation?

- Similar surprising results for other shape parameters and other digital trees:

  M. Fuchs and H.-K. Hwang. Dependence between path length and size in random digital trees, preprint.