Migration Letters

Volume: 21, No: S3 (2024), pp. 348-363

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Depth Of A Node And Distance Between Two Nodes In A Skip List

Munera Saad Altamimi*

Abstract

The skip list was introduced as a data structure for dictionary operations. Using a binary tree representation of skip lists, we obtain the limit law for the path lengths of the leaves in the skip list, and Distance between two nodes. In our study we will study some parameters about the skip lists as a random graph. We are essentially interested in the depth of a node, the distance between two no[1]des. The techniques used are very wide: moment generating functions, techniques used in random trees, central limit theorem...etc. A precise average case analysis is performed for the parameter number of left- to-right maxima. Some additional results are also given. Due to their poor cache behavior when compares with e.g. B-trees, but fear not, a good implementation of skip lists can easily outperform B-trees while being implementable in only a couple of hundred lines. There are many problems related to skip list that are studied. We plan to continue this study in order to better characterize the very useful objects in computer Science.

Keywords: *Skip list; Dictionary operation; Distance Nodes.*

1. Introduction

A simple undirected graph G = (V, E) is a nonempty set of vertices V and a set of edges E each edge $e \in E$ E is an unordered pair of distinct vertices x, y. The two vertices x, y are then called the endpoints of e, and e is said to join x and y. The edge e is also said to be incident with x and y. To simplify our notation, when an edge joining x and y is not labeled by a name like e, we refer to it by xy, so as to avoid the cumbersome parentheses of the set notation. A graph is empty if it has no edges. Because directions are not part of the definition, the edges do not have any orientation; both xy and yx refer to the same edge and we may use them interchangeably. Figure 1 illustrates a simple graph with V = t, u, v, w, x, y, z, E = a, b, c, d, e, f [1].

The degree of a vertex v, V is the number of edges incident with v. We denote the degree of v by d (v). The total sum of degrees in a simple undirected graph, each edge contributes 2 because every edge is incident with two vertices. Thus,

$$\sum v \epsilon V d(V) = 2E$$

Example 1. Let v, u, t, w are the vertices and a, b, c, d are the edges

*Department of Mathematics, Faculty of Science and Human studies, Al- Kharj, Prince Sattam Bin Abdul Aziz University.

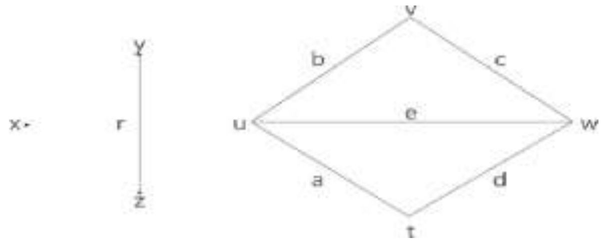

Figure 1.1: A simple Graph

Parity consideration necessitates the number of vertices with odd degree be even, as the total sum of degrees is even. A graph is defined analogously except that each edge is an ordered pair of distinct vertices [2]. The two edges xy and yx are different; the former is directed from x to y, the latter is directed from y to x. we say that xy starts at x and ends at y. The in degree $d_{in}(u)$ of a vertex u is the number of edges that start at u, and $d_{out}(u)$, the out degree of u, is the number of edges ending at u.

Example 2. Figure 1.2 shows a digraph in which $d_{in}(u) = 2$ and $d_{out}(u) = 1$.

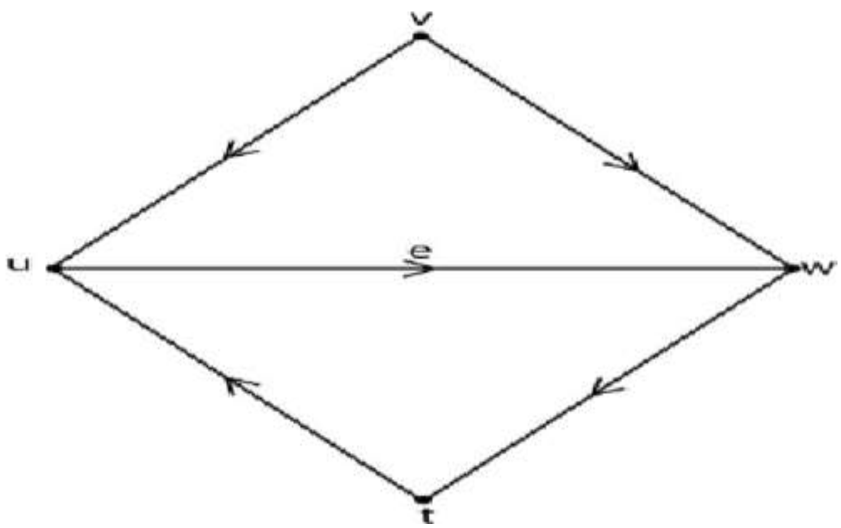

Figure 1.2: A digraph

A graph $H = (V, E)$ is a subgraph of $G = (V', E')$ if $V' \subseteq V$ and $E' \subseteq E$. We use the notation

$H \subset G$ to indicate that the graph H is a subgraph of the graph G. If $H \subset G$ and H is not G, then H is a proper subgraph of G, when a vertex is removed, the only way to maintain the integrity of the remaining part of the graph and have a consistent subgraph definition is to destroy all edges incident with it.

A path graph is a sequence $x_0, x_1, \ldots, x_k$, of distinct vertices such that x_i, x_{i+1} is an edge for $i = 0, \ldots\ldots, k-1$.

A simple graph is connected if there is a path in it from x to y, for every pair vertices x and y. A subgraph H of the graph G is a component if H is connected and is not a proper subgraph of another connected subgraph of G. In the following example (look to the example 1.1), we have three components and the vertex x is isolated because it is in a component by itself [3].

1.1. Tree

In graph theory, a tree is an undirected graph in which any two vertices are connected by exactly one path, or equivalently a connected acyclic undirected graph [1]. A forest is an undirected graph in which any two vertices are connected by at most one path, or equivalently an acyclic undirected graph, or equivalently a disjoint union of trees. The following are some of the basic theorems concerning undirected trees.

Theorem 1 [1]**.** In any tree there is exactly one path connecting any two of its vertices.

Proof: Let T = (V, E) be a tree and let u and v be two vertices of T . Because T is connected, there must be at least one path from u to v.

Assume the statement of the theorem is not true, toward a contradiction. That is, assume that there is more than one path u to v. Let P and Q be two such paths; the two paths start at u. Let w be the first common vertex between P and Q after which the two paths separate: that is, the next vertex after w on P is different from the next vertex after w on Q (w is possibly u). Because P and Q both terminate at v, they must meet again. Let x be the first common vertex between P and Q after w (x is possibly v). The part of the path P connecting w to x is a path P ′, and the part of the path Q connecting w to x is also a path Q′, and the only common vertices between P ′ and Q′ are w and x, where they both begin and end. Now, P ′ together with the reverse of Q′ forms a circuit in T . But T is circuit-free, a contra diction.

Theorem 2 [1]**.** The number of edges in a tree with n vertices is n − 1.

Proof: Let T = (V, E) be a tree with n vertices. We first show that if e = xy is an edge of T, then T e is disconnected. The two ends x and y of the edge e are connected by a unique path, according to Theorem 1.1. Because the edge xy connects x to y, there is no path from x to y other than the sequence x, y. It follows that there is no x y path in T e; that is, T e is disconnected. Moreover, the paths from x to all other vertices of T can be classified into two types: paths using e and paths not using e. Any vertex connected to x via a path not using e will remain connected to x in T e, and every vertex connected to x with a path that uses e must be connected to y first and will not be connected to x in T e, but will remain connected to y in T e. Thus T e is a graph consisting of the trees, has the same number of vertices as T , and has one less edge than T . We next proceed by induction on n. The only tree with one vertex is the empty graph on one vertex, and the theorem is clearly true for n = 1. Assume the statement is true for every tree on less than n vertices. Let the two trees of T − e be $T_1 = (V_1, E_1)$ and $T_2 = (V_2, T_2)$. It is clear that

$|V_1| < n$ and $|V_2| < n$. Thus, by induction, $|E_1| = |V_1| - 1$ and $|E_2| = |V_2| - 1$. Adding these two equations, we get $|E_1|+|E_2|=|\upsilon|-2$. Because $|E|=|E_1|+|E_2|+1$, it follows that $|E|=n-1$, and this completes the induction.

Theorem 3 [1]**.** If G is a graph with n vertices, the following statements are equivalent:

- **(a)** G is a tree.
- **(b)** G is a circuit-free and has n − 1 edges.
- **(c)** G is a connected and has n − 1 edges.

Proof: (a) implies both (b) and (c) by definition of trees and theorem 1.2. If G satisfies (b), it cannot be disconnected, for if it were, then every connected component of it is circuit-free, and if there are k > 1 such components, then each would be a tree and hence would satisfy theorem

1.2. Namely, if $T_i = (E_i, V_i)$ is one of these trees, then $E_i = V_i$ 1. But summing over all k components would yield

$$
\begin{aligned}
|E| &= \sum_{i=1}^{k} |E_i| \\
&= \sum_{i=1}^{k} \Big(|V_i| - 1\Big) \qquad (1.1) \\
&= n - k < n - 1.
\end{aligned}
$$

Using (1.1) we have a contradiction. Thus G is connected and (b) implies both (a) and (c).

Finally, assume G satisfies (c). Suppose G has a circuit and e is an edge of G contained in the circuit. Then $G - e$ is still connected. If $G - e$ has a circuit, then remove another edge contained in that circuit and the resulting graph is still connected. Continue this process until a connected graph with no circuit is obtained. The final graph is a tree on n vertices, and thus has $n - 1$ edges, by theorem 1.2. By the removal of some of the edges of G contained in circuits, the final graph has less than $n - 1$ edges, and this is impossible. This contradiction shows that G contains no circuits. Thus G is a tree and (c) implies both (a) and (b).

In computer applications, it is more common to use directed trees. These are simple digraphs without circuits, and the in degree of every vertex is 1, except for a distinguished vertex called root, which has in degree 0.

In a directed tree, if we drop edge directions we obtain the underlying undirected tree. Thus, in a directed tree, every vertex is reachable via a directed path from the root. If v is a vertex, then it is children are the set of vertices$\{x|vx \in E\}$. The leaves are the set of vertices with no children. The leaves are the terminal nodes of the tree.

1.2. Binary Trees

Definition. A binary tree consists of a finite set of nodes that is either empty, or consists of one specially designated node called the root of the binary tree, and the elements of two disjoint binary trees called the left subtree and right subtree of the root [5].

Example 3. The following is an example of binary tree.

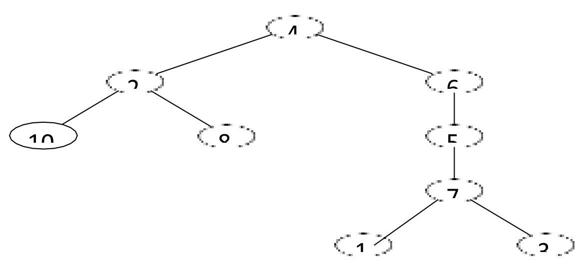

Figure 1.3: Binary tree constructed from 4, 2, 10, 8, 6, 5, 7, 1, 3. The only condition on the tree that each node must have at most two children and the distribution of numbers is at random without any order.

1.2.1. Binary Tree Terminology

Tree terminology is generally derived from the terminology of family trees.

Each root is said to be the parent of the roots of its subtrees.

Two nodes with the same parent are said to be siblings; they are the children of their parent.

The root node has no parent. [5].

A great deal of tree processing takes advantage of the relationship between a parent and its children, and we commonly say a directed edge (or simply an edge) extends from a parent to its children. Thus edges connect a root with the roots of each subtree. An undirected edge extends in both directions between a parent and a child.

Grandparent and grandchild relations can be defined in a similar manner; we could also extend this terminology further if we wished (designating nodes as cousins, as an uncle or aunt, etc.).

Definition 3. An important special kind of binary tree is the binary search tree **(BST)**. In a **BST**, each node stores some information including a unique key value, and perhaps some associated data. A binary tree is a **BST** if and only if, for every node x in the tree:

-All keys in x′s left subtree are less than the key in x, and

-all keys in x′s right subtree are greater than the key in x.

Remark 1. The reason binary-search trees are important is that the following operations can be implemented efficiently using a **BST**:

-insert a key value

-determine whether a key value is in the tree

-remove a key value from the tree

-print all of the key values in sorted order

Example 4. A binary search tree

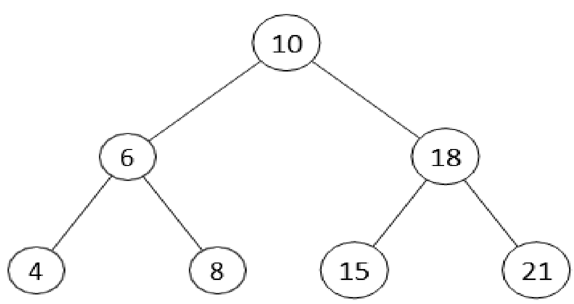

Figure 1.4: Binary search tree constructed from 10, 6, 4, 8, 15, 21, 18. Distribution of numbers under a given condition. For example 6 is on the left of 10 because it is less than 10 and 4 on the left of 6 because it is less than 6, 8 less than 10 will be on the left but it is greater than 6 then it must goes to the the right of 6, and so on

1.2.2. Depth of a node

Definition 4. The depth of a node is the number of edges from the node to the tree’s root node. The root node will have a depth of 0.

1.2.3. Height of a tree

Definition 5. The height of a tree is the number of edges on the longest path from the root to a leaf. The height of a tree T will be denoted by H (T).

Example 5. The figure below gives the depths and height in a tree:

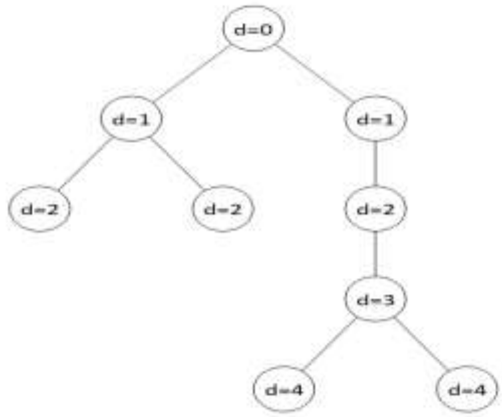

Figure 1.5: Depths of nodes where, for x node of the tree, d(x) denotes the depth of x. The height of a tree is the maximum of depths of its nodes then H = 4. The diameter of a tree is the number of nodes on the longest path between any two leaf nodes. The tree below has a diameter of 7 nodes.

1.2.4. Distance between two nodes in a Binary Tree

Definition 6. Distance between two nodes is the minimum number of edges to be traversed to reach one node from other

d(X, Y) = d(root, X) + d(root, Y)?2 ∗ (d(root to LCA(X, Y)))

where LCA(X, Y) = Lowest Common Ancestor of X, Y and d(u, v) is the distance the distance between nodes u and v.

Example 6. Let the following graph.

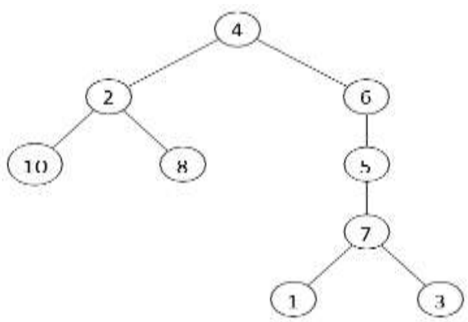

Figure 1.6: Distance (4,2)=1, Distance (4, 10)=2, Distance(6, 3)=3, Distance (2, 1)=5, Distance (3, 1)=2, Distance (10, 3)=6.

1.3. General formulation

Let T_n (We refer the reader to the monograph of Mahmoud 2003 for the details of the object) be a binary search tree with n internal nodes. And let Δ_n be the distance between two nodes chosen randomly in $\tau_{n.}$ Let also D_n be the depth of a randomly chosen node in the same tree. In a random binary search tree generated from a random permutation of 1, . . . , n, select two nodes uniformly at random, all $\binom{2}{n}$ choices being equally likely. Let the distance between the two nodes be Δ_n. This distance has a recursive structure that relates to D_n. The two nodes may come from the left subtree, form the right subtree, may involve the root and one node from the left subtree, may involve the root and node form the right subtree or may be in different subtrees with a path connecting them passing through the root. Conditioned on R being the root label, we have a recurrence:

$$\Delta_n = \begin{cases} \Delta_{R-1}, & \text{with probability } \frac{\binom{R-1}{2}}{\binom{n}{2}}; \\ \tilde{\Delta}_{n-R}, & \text{with probability } \frac{\binom{n-R}{2}}{\binom{n}{2}}; \\ (D_{R-1}+1)+(\tilde{D}_{n-R}+1), & \text{with probability } \frac{(R-1)(n-R)}{\binom{n}{2}}; \\ D_{R-1}+1, & \text{with probability } \frac{(R-1)}{\binom{n}{2}}; \\ \tilde{D}_{n-R}+1, & \text{with probability } \frac{(n-R)}{\binom{n}{2}}; \end{cases}$$

Here,(D_1,…………, D_n) and ($\tilde{D}1$,……………., $\tilde{D}n$) denote independent copies of random depths; likewise (Δ_1,…………, Δ_n) and ($\tilde{\Delta}1$,………………, $\tilde{\Delta}n$) are independent. It should be noted, however,that (D_1,…………….., D_n) and (Δ_1,……………………, Δ_n)are dependent.

Theorem 4. The distance Δ_n between two randomly selected nodes in a random binary search tree of size n has mean value

$$E[\Delta_n] = \frac{4(n+1)}{n(n-1)}[(n+3)H_{n+1} - 3n - 3] = 4\ln n + O(1)$$

Proof: By right and left symmetries, we have the recurrence

$$E[\Delta_n] = \frac{2}{\binom{n}{2}}\left[\Delta_{R-1}\binom{R-1}{2}\right] + \frac{2}{\binom{n}{2}}E[(D_{R-1}+1)(R-1)(n-R) + \frac{2}{\binom{n}{2}}E[(D_{R-1}+1)(R-1)].$$

Define $x_n = \binom{2}{n} E[\Delta n]$ to obtain the recurrence

$$x_n = \frac{2}{n}\sum_{r=1}^{n} x_{r-1} + \frac{2}{n}\sum_{r=1}^{n}(E[D_{r-1}+1)(r-1)(n-r+1).$$

The moments and limit law for D_n were obtained by classic recurrence and generating function techniques [1], as well as various probabilistic techniques [6]. These results include

$$E[D_n] = 2(n+1)H_n - 4 = 2lnn + O(1),$$

Difference nx_n and $(n-1)x_{n-1}$ and rearrange in the form

$$nx_n = (n+1)x_{n-1} + 2\sum_{r=1}^{n}(2rH_{r-1} - 3r + 3).$$

The transformation $y_n = x_{n-1}/(n+1)$ linearizes the recurrence:

$$y_n = y_{n-1} + \frac{q(n)}{n(n+1)},$$

where $q(n) = \sum_{r=1}^{n}(2rh_{r-1} - 3r + 3)$. We unwind the recurrence under the boundary condition $y_1 = 0$, get the answer and clean up the sums in it and finally obtain E $[\Delta_n]$.

2. Basic tools

2.1 Convergence of random variables

2.1.1 Almost surely convergence

Definition 7. A sequence of random variables $X_1, X_2, \ldots\ldots, X_n$ defined on the same samplespace **Ω**, converges almost surely to a random variable X, and we write

$$X_n \xrightarrow[n\to+\infty]{a.s} X$$

$$\mathbf{P}\Big(\omega \in \Omega, : lim_n X_n(\omega) = X(\omega)\Big) = 1.$$

Example 7. with a probability measure that is uniform on this space, **P** [a, b] = b − a for all

$0 \leq a \leq b \leq 1$ Define the sequence $\{X_n, n = 1, 2, 3, \cdots\}$ as follows:

$$X(s) = \begin{cases} 1, & if\, 0 \leq s \leq \frac{1}{2}; \\ 0, & otherwise. \end{cases}$$

Also, define the random variable X on this sample space as follows:

$$X(s) = \begin{cases} 1, & \text{if } 0 \leq s \leq \frac{1}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

Let $S = [0, 1]$, *we need to show that* $X_n \xrightarrow[n\to+\infty]{a.s} X$.

Define the set A as follows:

$$A = \left\{ s \in S : \lim_{n\to\infty} X_n(s) = X(s) \right\}.$$

Our aim is to prove that $\mathbf{P}(A) = 1$. Let note that $\frac{n+1}{2n} > \frac{1}{2}$, so for any $s \in [0, \frac{1}{2})$, we have

$$X_n(s) = X(s) = 1$$

Therefore, we conclude that $[0, \frac{1}{2}) \subset A$. Now if $s > \frac{1}{2}$,then

$$X(s) = 0.$$

Also, since $2s - 1 > 0$, we can write

$$X_n(s) = 0, \ \forall n > \frac{1}{2s-1},$$

this implies that,

$$\lim_{n\to+\infty} X_n(s) = 0 = X(s), \forall s > \frac{1}{2}.$$

We conclude that $(\frac{1}{2}, 1] \subset s$, we can check that $s = \frac{1}{2} \notin A$, since

$$X_n(\frac{1}{2}) = 1, \text{for all } n.$$

while $X(\frac{1}{2}) = 0$, we conclude

$$A = [0, \frac{1}{2}) \cup (\frac{1}{2}, 1] = s - \{\frac{1}{2}\}$$

Since $P(A) = 1$, we conclude $X_n \xrightarrow[n\to+\infty]{a.s} X$.

2.1.1 Convergence in probability

Definition 8. A sequence of random variables $X_1, X_2, \ldots\ldots\ldots, X_n$ defined on the same sample space Ω, converges almost surely to a random variable X, [7] [8] and we write

$$X_n \xrightarrow[n\to+\infty]{\mathbf{P}} X$$

if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\Big(\|X_n - X\| > \varepsilon\Big) = 1.$$

Example 8. *Let* X *be a random variable, and* $X_n = X + Y_n$, *where*

$$E[Y_n] = \frac{1}{n}, var(Y_n) = \frac{\sigma^2}{n}$$

Where $\sigma > 0$ *is a constant. Show that* $X_n \xrightarrow[n\to+\infty]{\mathbf{P}} X$ *in probability.*

First note that by the triangle inequality, for all a, b ∈ R we have $|a + b| \leq |a| + |b|$. Choosing $a = Y_n - E(Y_n)$ and $b = \mathbf{E}(Y_n)$,we obtain

$$|Y_n| \leq |Y_n - \mathbf{E}(Y_n)| + frac1n$$

Now, for any $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbf{P}\Big(|Y_n| \geq \varepsilon\Big) &\leq \mathbf{P}\Big(|Y_n - \mathbf{E}(Y_n)| + \frac{1}{n} \geq \varepsilon\Big) \\
&= \mathbf{P}\Big(|Y_n - \mathbf{E}(Y_n)| \geq \varepsilon - \frac{1}{n}\Big) \\
&\leq \frac{\mathbf{Var}(Y_n)}{(\varepsilon - \frac{1}{n})^2} \text{ by chebychevs in equality} \\
&= \frac{\sigma^2}{n(\varepsilon - \frac{1}{n})} \to 0 \text{ when } n \to \infty
\end{aligned}$$

Therefore, we conclude $X_n \xrightarrow[n\to+\infty]{\mathbf{P}} X$ in probability.

2.1.1 Convergence in distribution

Definition 9. Suppose that $X_{n,\ n\in N}$ and X are real-valued random variables with distribution functions $F_{n,\ n\in N}$ and F, respectively. We say that X_n converges in distribution to X as n goes to infinity [9] [10] [11] if

$\lim F_n(x) = F(x)$

For all real number x at which F is continuous.

Example 9. Let (X_n) be a sequence of independent identically distributed random variables with uniform distribution on the interval [0, 1]. Define

$Y_n = n(1 - \max_{1\leq I \leq n} X_i)$

The distribution function of Y_n is

$$\begin{aligned}
F_{Y_n} &= \mathbf{P}(Y_n \leq y) \\
&= \mathbf{P}\Big(n(1 - max_{1\leq i\leq n} X_i)\Big) \\
&= \mathbf{P}\Big((max_{1\leq i\leq n} X_i \geq 1 - \frac{y}{n}\Big) \\
&= 1 - \mathbf{P}\Big(x_1 < 1 - \frac{y}{n}, x_2 < 1 - \frac{y}{n}, \cdots, x_n < 1 - \frac{y}{n}\Big) \\
&= 1 - \mathbf{P}(x_1 < 1 - \frac{y}{n}).\mathbf{P}(x_2 < 1 - \frac{y}{n}). \cdots .\mathbf{P}(x_n < 1 - \frac{y}{n}) \\
&= 1 - F_{x_1}(1 - \frac{y}{n}).F_{x_2}(1 - \frac{y}{n}). \cdots .F_{x_n}(1 - \frac{y}{n}) \\
&= 1 - \Big[F_{x_n}\Big(1 - \frac{y}{n}\Big)\Big]^n
\end{aligned}$$

Thus,

$$F_{Y_n}(y) = \begin{cases} 0, & if\ y < 0; \\ 1 - (1 - \frac{y}{n})^n, & if\ 0 \leq y < n; \\ 1, & if\ y \geq n \end{cases}$$

Sinsce

$$\lim_{n\to\infty} (1 - \frac{y}{n})^n = exp(-y)$$

We have

$$F_{Y_n}(y) = F_y(y) \begin{cases} 0, & if\ y < 0; \\ 1 - exp(-y), & if\ y \geq n \end{cases}$$

Where F_Y (y) is the distribution function of an exponential random variable. Therefore, the sequence Y_n converges in law to an exponential distribution.

2.1.1 Generating function

Definition 10 [12]: The generating function for a sequence me a_0, a_1, a_2, is defined as ...

$$F(z) = \sum_{n=0}^{\infty} a_n z^n$$

Where z is a complex variable. Sometimes the generating function is called the z transform. The coefficient a_n of the nth power of z in this this power series is denoted by $[z^n]F(z)$. The function F (z) is formal power series, in the sense that we do not care very much about domains of convergence. We just require that it converges for some radius $|z| > 0$.

Example 10. Define unlabeled binary trees as rooted unlabeled directed trees and each vertex may have no children, one left child, one right child, or two children (one left and one right), where the left and right children are distinguished.

$$B(z) = \sum_{n=0}^{\infty} b_n z^n$$

$$B(z) = \sum_{n=0}^{\infty} b_n z^n$$

Be the generating function of the sequence $(b_n)^{\infty}_{n=0}$. The number of trees of order n with i vertices in the left subtree is $b_i b_{n-i-1}$, because we can form all possible subtrees of order i to the left of the root vertex, and independently form n − i − 1 subtrees to right of the root vertex. But i can be 0, 1, · · · or n − 1. Thus

$b_n = b_0 b_{n-1} + b_1 b_{n-2} + \cdots + b_{n-1} b_0$

Multiply both sides by z^n and sum over n from 1 to ∞to obtain

$$B(z) - b_0 = zB^2(z)$$

Although counting binary trees on n vertices is not meaningful for n = 0, the value $b_0 = 1$ isconsistent with the recurrence. Hence

$$B(z) - 1 = zB^2(z)$$

This quadratic equation in B(z) has the two solutions

$$B(z) = \frac{1 \pm \sqrt{1-4z}}{2z}$$

The initial condition $B(0) = b_0 = 1$ is satisfied only by the minus sign; the plus sign must berejected.

We use the notation $(\alpha)_k$, known as Pochhammer's symbol, to the product α(α 1).(α-k+ 1),for any

complex number α.

$$\begin{aligned} b_n &= [z^n]B(z) \\ &= [z^n]\left(\frac{1}{2z}\left[1 - \sum_{n=0}^{\infty} \frac{(\frac{1}{2})_n}{n!}(-4z)^n\right]\right) \\ &= -\frac{1}{2}(-4)^{n+1}\frac{(\frac{1}{2})_{n+1}}{(n+1)!} \\ &= 4^n \frac{\frac{1}{2} \times \frac{3}{2} \times \frac{5}{2} \times \cdots \times \frac{(2n-1)}{2}}{(n+1)!} \\ &= \frac{1}{n+1}\binom{2n}{n} \end{aligned}$$

Theorem 5. *Suppose that $A(z) = \Sigma_{n\geq 0} a_n z^n$ and $B(z) = \sum_{n\geq 0} b_n z^n$ are two generating functions, and the radius of convergence of $A(z)$ is larger than that of $B(z)$. Let $C(z) = \sum_{n\geq 0} c_n z^n$ be the product $A(z)B(z)$. Suppose further that b_{n-1}/b_n approaches a limit b as $n \to \infty$. If $A(b) \neq 0$, then $c_n \sim A(b) b_n$.*

Proof: It suffices to show $c_n/b_n \sim A(b)$. First write

$$\begin{aligned} \left|A(b) - \frac{c_n}{b_n}\right| &= \left|A(b) - \sum_{k=0}^{n} a_k b^k + \Sigma_{k=0}^{n} a_k b^k - \frac{c_n}{b_n}\right| \\ &\leq \left|A(b) - \Sigma_{k=0}^{n} a_k b^k\right| + \left|\Sigma_{k=0}^{n} a_k b^k - \frac{c_n}{b_n}\right| \text{ using triangular inequality} \end{aligned}$$

On the other hand the product $C(z) = A(z)B(z)$, gives that the coefficient $c_n = \sum_{k=0}^{n} a_n b_{n-k}$. But $A(b) = \sum_{k\geq 0} a_k b^k$ then

$$\lim_{n\to+\infty} \left(A(b) - \sum_{k\geq 0} a_k b^k\right) = 0$$

then when n is large enough the quantity $\left|A(b) - \sum_{k\geq 0} a_k b^k\right|$ becomes very small.
The term $sum_{k=0}^{n} a_k b^k - c_n/b_n$ can be written as

$$\begin{aligned} &\left|\sum_{k=0}^{n} a_k b^k - \frac{c_n}{b_n}\right| \\ &= \left|\sum_{k=0}^{n} a_k b^k - \sum_{k=0}^{n} \frac{a_n b_{n-k}}{b_n}\right| \\ &= \left|\sum_{k=0}^{n} a_k \left(b^k - \frac{b_{n-k}}{b_n}\right|\right. \\ &\leq \sum_{k=0}^{n} |a_k||b^k - \frac{b_{n-k}}{b_n}|, M \leq n, n \text{ fixed} \\ &\leq \sum_{k=0}^{M} |a_k||b^k - \frac{b_{n-k}}{b_n}| + \sum_{k=n+1}^{n} |a_k||b^k - \frac{b_{n-k}}{b_n}|. \end{aligned}$$

We have, $\forall\, k \geq 0$,

$$\lim_{m\to+\infty} \frac{b_{n-k}}{b_n} = \lim_{m\to+\infty} \left(\frac{b_{n-k}}{b_{n-k+1}} \times \frac{b_{n-k+1}}{b_{n-k+2}} \times \cdots \times \frac{b_{n-1}}{b_n}\right) \sim b^k,$$

which implies that

$$\mathop{lim}_{n\to+\infty} \left(\sum_{k=0}^{n} |a_k|||b^k - \frac{b_{n-k}}{b_n}||\right) < +\infty \text{ then } \mathop{lim}_{n\to+\infty} \left(\sum_{k=n+1}^{n} |a_k||b^k - \frac{b_{n-k}}{b_n}|\right) = 0.$$

The series $\sum_{k=0}^{n} a_k b^k$ converges, then $|A(b) - c_n/b_n|$ approaches 0 as $n \to \infty$, and the theorem follows

Example 11. *Suppose $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$ are two sequences with generating functions $A(z)$ and $B(z)$, respectively:*

$$A(z) = \sum_{n\geq 0} a_n z^n, \; B(z) = \sum_{n\geq 0} b_n z^n.$$

We have

- $A(z) + B(z) = \sum_{n\geq 0}(a_n + b_n)z^n$.
- $A(z) + B(z) \rightarrow$ *is the G.F of* $(a_n + b_n)_{n\geq 0}$.
- $A(z) - B(z) \rightarrow$ *is the G.F of* $(a_n - b_n)_{n\geq 0}$.

$$\begin{aligned} A(z).B(z) &= \Big(\sum_{n\geq 0} a_n z^n\Big)\Big(\sum_{m\geq 0} b_m z^m\Big) \\ &= \sum_{n\geq 0}\sum_{m\geq 0} a_n b_m z^{n+m} \\ &= \sum_{k=0}^{\infty} z^k \sum_{L=0}^{k} a_{k-L} b_L \\ &= \sum_{k=0}^{\infty}\Big(\sum_{L=0}^{k} a_{k-L} b_L\Big) z^k \\ &= \sum_{k=0}^{\infty} C_k z^k \end{aligned}$$

$A(z).B(z)$ *is the G.F of* $\left(\sum_{L=0}^{k} a_{k-L} b_L\right)$.

Example 12. *What is the generating function for the binomial coefficients* $\binom{n}{k}$, *for a fixed integer* n. *Let for all* $k \leq n$, $a_k = \binom{n}{k}$ *and* $A(z)$ *be the moment generating function of the sequence* $(a_k)_k$. *Then*

$$\begin{aligned} A(z) &= \sum_{k=0}^{n} C_k^n z^k \\ &= \sum_{k=0}^{\infty} \binom{n}{k} z^k 1^{n-k} = (z+1)^n \\ (a+b)^n &= \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \end{aligned}$$

Definition 11. *Let* X *be an integer random variable. The generating function (GF) of* X *denoted by* $G_X(t)$ *defined for all* $t \in]0,1[$ *by*

$$G_X(t) = \mathbf{E}[t^X] = \sum_{k=0}^{\infty} t^k \mathbf{P}(X = k).$$

Theorem 6. *If* X *is an integer random variable with generating function*G_X *then, for all integer number* m *such that* X *have moment of order* m *we have,*

$$\mathbf{E}[X(X-1)(X-2)\cdots(X-(m-1))] = G_X^{(m)}(0)$$

Where we define $G_X^{(m)}(t)$ *as the* m^{th} *derivative of* G_X

Example 13. *If* X *is a geometric random variable:* $\forall k \geq 0, \mathbf{P}(X = k) = p^k(1-p)$ *Then the generation of* X *is given by*

$$G_X(t) = \frac{q}{1-pt} \forall t \in]0,1[.$$

Remark 2. *The combinatorics and probabilistic counting* $[4, 8-10]$. *The geometric distribution plays an important role in our study.*

Theorem 7. *The generating function characterizes the distribution and if* X *and* Y *are two integer random variables. If* X *and* Y *are independent, we have*

$$G_{X+Y} = G_X . G_Y$$

Conclusion:

The skip list was introduced as a data structure for dictionary operations. Using a binary tree representation of skip lists, we obtain the limit law for the path lengths of the leaves in the skip list, and Distance between two nodes. Due to the rise in popularity of peer-to-peer systems dynamic overlay networks have recently received a lot of attention. An overlay network is a logical network formed by its participants over a wired or wireless routing network. The number of computers and users in such a network may reach millions. Therefore, research in this area has focused on improving scalability and efficiency of overlay networks. Two usual optimization properties are the speed of searching for items in the network and the speed of topology updates. Another popular parameter is expansion. Due to their poor cache behavior when compares with e.g. B-trees, but fear not, a good implementation of skip lists can easily outperform B-trees while being implementable in only a couple of hundred lines. There are many problems related to skip list that are studied. We plan to continue this study in order to better characterize the very useful objects in computer Science.

References

1. Mahmoud, H.: Evolution of random search trees, Wiley-Inter science Series in Discrete Mathematics and Optimization. John Wiley and Sons, Inc., New York (1992).
2. Aleksander Kelenc, Niko Tratnik, Ismael G. Yero, Uniquely identifying the edges of a graph: The edge metric dimension. Discrete Applied Mathematics, Volume 251, 31 December 2018, Pages 204-220.
3. Gaurav Kucheriya. Gábor Tardos, A Characterization of Edge-Ordered Graphs with Almost Linear Extremal Functions. Combinatorica (2023) 43:1111–1123 https://doi.org/10.1007/s00493-023-00052-5.
4. Ramesh Hariharan, Telikepalli Kavitha, and Kurt Mehlhorn, Faster Algorithms for Minimum Cycle Basis in Directed Graphs. 2008, Society for industrial and applied mathematics, Vol. 38, No.4, pp.1430-1447.
5. Lara Pudwell, Connor Scholten, Tyler Schrock, and Alexa Serrato, Noncontiguous Pattern Containment in Binary Trees. Hindawi Publishing Corporation ISRN Combinatorics Volume 2014, Article ID 316535, 8 pages http://dx.doi.org/10.1155/2014/316535.
6. Devroye, L.: A limit theory for random skip lists, Ann. Appl. Probab. 2 (1992) 597-609.
7. Borovkov, A.A. (2013). On Convergence of Random Variables and Distributions. In: Probability Theory. Universitext. Springer, London. https://doi.org/10.1007/978-1-4471-5201-9_6
8. George G. Roussas, An Introduction to Probability and Statistical Inference (Second Edition), Academic Press, 2015, Pages 245-271, ISBN 9780128001141, https://doi.org/10.1016/B978-0-12-800114-1.00007-X.
9. Esquerdo, P.J.R. (2011). Convergence of Random Variables. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_188
10. Bernardo Lafuerza-Guillén a and Carlo Sempi, Probabilistic norms and convergence of random variables. J. Math. Anal. Appl. 280 (2003) 9–16
11. Mahmoud, H; Boris, P.: On the most probable shape of a search tree grown from a random permutation, SIAM J. Algebraic Discrete Methods, 5(1), 69-81 (1984).

Acknowledgement: The author is grateful to the Research and Innovation Agency of the Ministry of Education of the Kingdom of Saudi Arabia. Also, he expresses his appreciation to Prince Sattam bin Abdulaziz University. Finally, the author thanks the Editor in Chief and anonymous referee for the insightful and constructive comments that improved this paper