On the Weakly Superincreasing Distributions and the Fibonacci-Hessenberg Matrices

Morteza Esmaeili[a,b]
[a]Dept. of Mathematical Sciences
Isfahan University of Technology, Isfahan, IRAN
[b] Institute for Studies in Theoretical Physics
and Mathematics (IPM), Tehran, IRAN
Email: emorteza@cc.iut.ac.ir

Abstract

This paper deals with the interconnections between finite weakly superincreasing distributions, the Fibonacci sequence and Hessenberg matrices. A frequency distribution, to be called *the Fibonacci distribution*, is introduced that expresses the core of the connections among these three concepts. Using a Hessenberg representation of finite weakly superincreasing distributions, it is shown that, among all such n-string frequency distributions, the Fibonacci distribution achieves the maximum expected codeword length.

Keywords: Weakly superincreasing distributions, Hessenberg matrices, Fibonacci sequence, Expected codeword length.

1 Introduction and Background

The Fibonacci sequence is defined by $f_0 = 0$, $f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$, $n \geq 2$. An $n \times n$ matrix $\mathcal{A} = (a_{i,j})$ is called a (lower) Hessenberg matrix if all entries above the superdiagonal are zero, that is if $j > i+1$ then $a_{i,j} = 0$. We refer to any Hessenberg matrix whose determinant is expressed in terms of Fibonacci numbers as a *Fibonacci-Hessenberg* matrix. As example let $\mathcal{R}_{1,t}$ be the one-by-one matrix with entry $t+1$ and recursively define the $n \times n$ matrix $\mathcal{R}_{n,t}$ given by (1).

Note that in the matrix $\mathcal{R}_{n,t}$ every entry below the diagonal is 1. The determinant of $\mathcal{R}_{n,t}$ is denoted by $r_{n,t}$. Using induction and cofactor expansion along the first row, one can easily show that $r_{n,t} = tf_{n+1} + f_n$, $n \geq 1$ and hence $\mathcal{R}_{n,t}$ is a Fibonacci-Hessenberg matrix. It is worth mentioning that $r_{n,t} = tf_{n+1} + f_n$ implies $r_{n,t} = r_{n-1,t} + r_{n-2,t}$. The sequence $r_{n,3}$

[1]This research was in part supported by a grant from IPM (No. 84940013).

generates the Lucas numbers, while $r_{n,t}$, $0 \le t \le 2$, generates the generalised Fibonacci sequence starting at $t+1$. Interested readers to a literature on Fibonacci-Hessenberg matrices are referred to [1][2][5][8][9].

$$\mathcal{R}_{n,t} := \begin{pmatrix} 2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 2 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 2 & 1 \\ 1 & \cdots & \cdots & \cdots & 1 & t+1 \end{pmatrix}_{n\times n} \tag{1}$$

A sequence $(a_i)_{i\ge1}$ of numbers is called *superincreasing* if it satisfies $\sum_{i=1}^{n} a_i \le a_{n+1}$, $n \ge 1$. We define $(a_i)_{i\ge1}$ to be a *weakly superincreasing* (WS) sequence if $\sum_{i=1}^{n} a_i \le a_{n+2}$, $n \ge 1$. A frequency distribution (FD) is a list of non-negative numbers. Probability distributions (PDs) are special cases of FDs wherein the numbers add to 1. It is also obvious that any FD results in a PD when the numbers are divided by their sum. A source $\mathcal{S} = \{s_1, s_2, \cdots, s_n\}$ with FD $p_1 \ge p_2 \ge \cdots \ge p_n$ is called a *weakly superincreasing source* (WSS) if

$$p_{i+2} + p_{i+3} + \cdots + p_n \le p_i, \qquad 1 \le i \le n-3. \tag{2}$$

An infinite alphabet source $\mathcal{S}$ with FD $p_1 \ge p_2 \ge \cdots$ is also called WS if

$$p_{i+2} + p_{i+3} + \cdots = \sum_{k=i+2}^{\infty} p_k \le p_i, \qquad i \ge 1.$$

Among the well known WS distributions we may refer to the Poisson and geometric distributions[6][7]. A given distribution $p_1 \ge p_2 \ge \cdots$ introducing a WSS is referred to as a weakly superincreasing distribution (WSD). In the rest of the paper, unless otherwise stated, by WSD we mean a weakly superincreasing frequency distribution which obviously covers the class of weakly superincreasing probability distributions. Given an infinite alphabet WSD $p_1 \ge p_2 \ge \cdots$ and setting $p'_n := \sum_{k=n}^{\infty} p_k$, at least one of the two distributions $(p_1, \cdots, p_{n-1}, p'_n)$ and $(p_1, \cdots, p_{n-2}, p'_n, p_{n-1})$ is an n-symbol WSD. We refer to the so obtained finite WSD from an infinite WSD as a truncated WSD. In the rest of the paper by WSD we mean a finite WSD.

It follows from definition of WSDs, relation (2), that the frequency distribution $p_1 \ge p_2 \ge \cdots \ge p_n$ of any such distribution introduces a WS sequence $a_i := p_{n-i+1}$. On the other hand, it is easily shown by induction that, for each positive integer n, the Fibonacci sequence satisfies $\sum_{i=1}^{n} f_i = f_{n+2} - 1$ and hence the sequence $(f_1, f_1, f_2, f_3, \cdots, f_n) = (1, 1, 1, 2, \cdots, f_n)$ is a WSD. This results in the following WS probability distribution:

$$\frac{f_1}{f_{n+2}}, \ \frac{f_1}{f_{n+2}}, \ \frac{f_2}{f_{n+2}}, \ \cdots, \ \frac{f_n}{f_{n+2}}. \tag{3}$$

Due to an important property of distribution $(f_1, f_1, f_2, f_3, \cdots, f_n)$, to be stated in Section 3, we refer to this as *the Fibonacci distribution.* Another WSD related to the Fibonacci sequence is obtained from equation $\sum_{i=1}^{n} f_{2i-1} = f_{2n}$. The corresponding probability distribution is the following superincreasing distribution:

$$\frac{f_1}{f_{2n}}, \frac{f_3}{f_{2n}}, \frac{f_5}{f_{2n}}, \cdots, \frac{f_{2n-1}}{f_{2n}}. \tag{4}$$

Let $p_1 \geq p_2 \geq \cdots \geq p_n$ be a WSD and $C = \{c_1, c_2, \cdots, c_n\}$ be a Huffman code for this distribution and assume that the length of the ith codeword c_i is l_i, $1 \leq i \leq n$. As Huffman coding [4] is a bottom to top procedure and l_i is the number of times that the corresponding number p_i is amalgamated, one can easily verify that C can be constructed in a way that $l_i = i$ for $i < n$ and $l_n = n - 1$. Conversely, any optimal source code C (that is a code with minimum expected codeword length) with codeword lengths satisfying $l_n = n - 1$ and $l_i = i$, $i < n$, represents a WSD.

This paper considers connections between WSDs and the Fibonacci-Hessenberg matrices and the Fibonacci sequence. The relation between Fibonacci-Hessenberg matrices and WSDs is given in Section 2. It is shown in Section 3 that among all WSDs, the maximum expected codeword length is achieved by the Fibonacci distribution defined above.

2 Hessenberg matrices and weakly superincreasing distributions

2.1 Weakly superincreasing distributions in terms of Hessenberg matrices

In this part we characterise WSDs in terms of Fibonacci-Hessenberg matrices.

Theorem 1 An $(n-2)$-tuple $(p_1, p_2, \cdots, p_{n-2})$ satisfying $p_1 + p_2 + \cdots + p_{n-2} < p$ and $0 < p_{n-2} \leq \cdots \leq p_2 \leq p_1$ form the first $n-2$ components of an n-symbol WSD with sum p if and only if it satisfies the system of inequalities given by (5).

$$\begin{cases} 2p_1 + p_2 \geq p \\ p_1 + 2p_2 + p_3 \geq p \\ p_1 + p_2 + 2p_3 + p_4 \geq p \\ \vdots \\ p_1 + \cdots + p_{n-4} + 2p_{n-3} + p_{n-2} \geq p \\ p_1 + \cdots + p_{n-3} + 3p_{n-2} \geq p. \end{cases} \tag{5}$$

Proof Consider an n-symbol WSD $0 < p_n \le \cdots \le p_2 \le p_1$ with sum $p = \sum_{i=1}^{n} p_i$. Considering (2), constraint $p_n + p_{n-1} + \cdots + p_3 \le p_1$ is equivalent to $p - (p_1 + p_2) = p_n + p_{n-1} + \cdots + p_3 \le p_1$, and hence it holds if and only if $2p_1 + p_2 \ge p$. Similarly, condition $p - (p_1 + p_2 + p_3) = p_n + p_{n-1} + \cdots + p_4 \le p_2$ holds if and only if $p_1 + 2p_2 + p_3 \ge p$. At the end of this process, we have $p - (p_1 + p_2 + \cdots + p_{n-2}) = p_n + p_{n-1} \le p_{n-3}$ if and only if $p_1 + p_2 + \cdots + 2p_{n-3} + p_{n-2} \ge p$. Finally, condition $0 < p_n \le p_{n-1} \le p_{n-2}$ is equivalent to $p - (p_1 + \cdots + p_{n-3}) = p_{n-2} + p_{n-1} + p_n \le 3p_{n-2}$, and hence we have $0 < p_n \le p_{n-1} \le p_{n-2}$ if and only if $p_1 + \cdots + p_{n-3} + 3p_{n-2} \ge p$. ∎

The following $(n-2) \times (n-2)$ Fibonacci-Hessenberg matrix $\mathcal{Q}$, which is equal to the Fibonacci-Hessenberg matrix $\mathcal{R}_{n-2,2}$ given by (1), is the coefficient matrix of the system given above. One can easily follow the given proof process and show that the statement of this Theorem holds for each $i \le n-2$.

$$\mathcal{Q} = \begin{pmatrix} 2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 2 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 2 & 1 \\ 1 & \cdots & \cdots & \cdots & 1 & 3 \end{pmatrix}_{(n-2)\times(n-2)} \tag{6}$$

The system of equations $\mathcal{Q}X = \mathbf{1}$, where $\mathbf{1}$ is the all one column vector, has solution

$$(x_{n-2}, x_{n-3}, \cdots, x_1) = \left(\frac{f_2}{f_{n+1}}, \frac{f_3}{f_{n+1}}, \cdots, \frac{f_{n-1}}{f_{n+1}}\right) \tag{7}$$

This solution vector is the Fibonacci probability distribution with the first two terms removed. Equivalently, $(q^{(n-2)}, q^{(n-3)}, \cdots, q^{(2)}, q^{(1)}, q^{(1)}, q^{(1)})$ is the n-symbol Fibonacci distribution where $q^{(i)}$ is the determinant of $\mathcal{Q}^{(i)}$, the matrix obtained from $\mathcal{Q}$ by replacing its ith column with the all one column vector $\mathbf{1}$.

2.2 Weakly superincreasing distributions derived from Hessenberg matrices

Fibonacci-Hessenberg matrices can be viewed as the origin of many WSDs. Consider for instance the following $n \times n$ Fibonacci-Hessenberg matrix $C_{n,t}$. Let $C_{n,t}^{(i)}$ be the matrix obtained from $C_{n,t}$ by replacing its ith column with the all one column vector $\mathbf{1}$, and let $c_{n,t}$ and $c_{n,t}^{(i)}$ denote the determinants of $C_{n,t}$ and $C_{n,t}^{(i)}$, respectively.

$$
C_{n,t} := \begin{pmatrix}
2 & -1 & 0 & \cdots & \cdots & 0 \\
1 & 2 & -1 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 2 & -1 \\
1 & \cdots & \cdots & \cdots & 1 & t+1
\end{pmatrix}_{n\times n} \tag{8}
$$

Theorem 2 The equalities

$$
\begin{cases} c_{n,t}^{(i)} = f_{2(n-i)+1} + t f_{2(n-i)}, & 1 \le i \le n; \\ c_{n,t} = t + \sum_{j=1}^{n} c_{n,t}^{(j)} = f_{2n} + t f_{2n-1}; \end{cases} \tag{9}
$$

hold for any number t and any two positive integers n and $1 \le i \le n$.

Proof The proof is by induction on n and computation of determinants using cofactor expansion along the first row. ■

Different WSDs can be obtained using relations in (9). Setting $t = 0$ we obtain the WSD

$$
\left(\frac{c_{n,0}^{(n)}}{c_{n,0}}, \frac{c_{n,0}^{(n-1)}}{c_{n,0}}, \dots, \frac{c_{n,0}^{(1)}}{c_{n,0}} \right) = \left(\frac{f_1}{f_{2n}}, \frac{f_3}{f_{2n}}, \dots, \frac{f_{2n-1}}{f_{2n}} \right)
$$

which is the distribution given by (4). For $t = 1$, we obtain the following $(n+1)$-symbol WSD:

$$
\left(c_{n,1}^{(n)}, c_{n,1}^{(n)}, c_{n,1}^{(n-1)}, \dots, c_{n,1}^{(1)} \right) = (f_2, f_2, f_4, \dots, f_{2n}).
$$

Similarly, the $(n+2)$-symbol WSD

$$
(f_2, f_2, f_2, f_2 + f_4, f_4 + f_6, \dots, f_{2n-2} + f_{2n})
$$

corresponds to $t = 2$. The case $t = 3$, however, leads to the following two WSDs

$$
\begin{aligned}
&(f_2, f_4, 2f_2 + f_4, 2f_4 + f_6, \dots, 2f_{2n-2} + f_{2n}), \\
&(f_2, f_2, f_3, 2f_2 + f_4, 2f_4 + f_6, \dots, 2f_{2n-2} + f_{2n}),
\end{aligned}
$$

where the first distribution introduces an $(n+1)$-symbol WSD and the second one refers to an $(n+2)$-symbol source.

3 Expected codeword length of the Fibonacci distribution

Given a source $S = \{s_1, s_2, \cdots, s_n\}$ with distribution $p_1 \geq p_2 \geq \cdots \geq p_n$, Huffman coding algorithm [4] assigns a prefix optimal code C to S. Let c_i be a codeword of length l_i assigned to the source symbol s_i by the algorithm. The code C is optimal in the sense that it has minimum expected codeword length $L = \sum_{i=1}^{n} p'_i l_i$ among all uniquely decodable codes representing S where $p'_i := p_i / \sum_{j=1}^{n} p_j$. In this section we show that among all n-symbol WSDs a distribution has maximum expected codeword length if and only if it is a scalar multiple of the Fibonacci distribution $(f_1, f_1, f_2, \cdots, f_{n-1})$.

Theorem 3 Denote $(f_1, f_1, f_2, \cdots, f_{n-1})$, the n-symbol Fibonacci distribution, by F_n and let $P = (p_n, p_{n-1}, \cdots, p_1)$, satisfying $p_1 \geq p_2 \geq \cdots \geq p_{n-1} \geq p_n$, be an n-symbol WSD. Then among all n-symbol WSDs the given distribution P has the maximum expected codeword length if and only if $P = \alpha F_n$ for some positive number α. Moreover, the expected codeword length of F_n is $1 + \frac{f_{n+2}-3}{f_{n+1}}$ bits per symbol.

Proof Since we are concerned with the expected codeword length, without loss of generality we may assume that the given WSD $p_1 \geq p_2 \geq \cdots \geq p_{n-1} \geq p_n$ add to 1, that is it is a probability distribution. The Huffman code of this distribution has expected codeword length:

$$\begin{aligned} L &= p_1 + 2p_2 + \cdots + (n-2)p_{n-2} + (n-1)p_{n-1} + (n-1)p_n \\ &= p_1 + 2p_2 + \cdots + (n-2)p_{n-2} + (n-1)(p_{n-1} + p_n) \\ &= p_1 + 2p_2 + \cdots + (n-2)p_{n-2} + (n-1)\{1 - p_1 - p_2 - \cdots - p_{n-2}\} \\ &= (n-1) - \{(n-2)p_1 + (n-3)p_2 + \cdots + 2p_{n-3} + p_{n-2}\}. \end{aligned}$$

Thus to maximise L we precisely need to minimise $L' := (n-2)p_1 + (n-3)p_2 + \cdots + 2p_{n-3} + p_{n-2}$ under the constraints given by (5) with $p = 1$. Therefore, the problem is equivalent to the linear programming (LP) problem

$$\begin{aligned} &\text{Minimise} \quad L' = (n-2)p_1 + (n-3)p_2 + \cdots + 2p_{n-3} + p_{n-2} \\ &\text{Subject to} \\ &\begin{cases} 2p_1 + p_2 \geq 1 \\ p_1 + 2p_2 + p_3 \geq 1 \\ p_1 + p_2 + 2p_3 + p_4 \geq 1 \\ \vdots \\ p_1 + p_2 + \cdots + p_{n-4} + 2p_{n-3} + p_{n-2} \geq 1 \\ p_1 + p_2 + \cdots + p_{n-3} + 3p_{n-2} \geq 1 \\ p_1 + p_2 + \cdots + p_{n-2} < 1 \\ p_1 \geq p_2 \geq \cdots \geq p_{n-2}. \end{cases} \end{aligned} \tag{10}$$

According to (6) and (7) the string $(p_{n-2}, p_{n-3}, \cdots, p_1) = \frac{1}{f_{n+1}}(f_2, f_3, \cdots, f_{n-1})$ is the solution of $QX = \mathbf{1}$ and hence a feasible solution for this LP problem. We show that this is indeed the optimal solution.

Replacing the constraint $p_1 + p_2 + \cdots + p_{n-2} < 1$ with $p_1 + p_2 + \cdots + p_{n-2} + \beta = 1$ we see that the dual LP problem [3] of this primal problem is as follows: Maximise $y_1 + y_2 + \cdots + y_{n-2} + y_{n-1}$ subject to $\begin{pmatrix} Q^t\mathbf{1} \\ \mathbf{01} \end{pmatrix} Y^t \leq \begin{pmatrix} B^t \\ 0 \end{pmatrix}$ and $y_i \geq 0$, $1 \leq i \leq n-1$, where Q^t denotes the transpose of the matrix Q given by (6), $Q^t\mathbf{1}$ is the augmented matrix obtained by adding the all one column vector $\mathbf{1}$ to the right side of Q^t, Y^t is the transpose of $Y = (y_1, y_2, \cdots, y_{n-1})$, and B^t is the transpose of $B = (n-2, n-3, \cdots, 1)$.

It follows from $y_{n-1} \geq 0$ and the inequality $y_{n-1} = [\mathbf{01}]Y^t \leq 0$ that $y_{n-1} = 0$ and hence the dual problem may be thought as: Maximise $y_1 + y_2 + \cdots + y_{n-2}$ subject to $Q^t(y_1, y_2, \cdots, y_{n-2})^t \leq B^t$. It is easily shown by induction and using cofactor expansion along the first column that the system $Q^t(y_1, y_2, \cdots, y_{n-2})^t = B^t$ has solution

$$y_{n-2} = \frac{f'_{n-2}}{f_{n+1}}; \quad y_i = (-1)^{n-3-i}\left(f'_{n-3-i} - f_{n+1-i}y_{n-2}\right), \ 1 \leq i \leq n-3;$$

where f'_i is defined by $f'_i := f_i - 1$ if i is odd and $f'_i := f_i + 1$ if i is even. This is a feasible solution for the dual problem.

Consider a primal problem: Minimise $\sum_{i=1}^m b_i x_i$ subject to $\sum_{i=1}^m a_{ij} x_i \geq c_j$ and $x_i \geq 0$, $1 \leq j \leq n$, $1 \leq i \leq m$, and its dual problem: Maximise $\sum_{j=1}^n c_j y_j$ subject to $\sum_{j=1}^n a_{ij} y_j \leq b_i$ and $y_j \geq 0$, $1 \leq j \leq n$, $1 \leq i \leq m$. Let $X^\star = (x_1^\star, x_2^\star, \cdots, x_m^\star)$ and $Y^\star = (y_1^\star, y_2^\star, \cdots, y_n^\star)$ be feasible solutions for these primal and dual problems, respectively. Then according to Theorem 5.2 in [3], necessary and sufficient conditions for simultaneous optimality of $X^\star$ and $Y^\star$ are

$$\begin{cases} \sum_{i=1}^m a_{ij} x_i^\star = c_j \ \text{ or } \ x_j^\star = 0 \ \text{ (or both)}, & 1 \leq j \leq n; \\ \sum_{j=1}^n a_{ij} y_j^\star = b_i \ \text{ or } \ y_i^\star = 0 \ \text{ (or both)}, & 1 \leq i \leq m. \end{cases}$$

Applying this Theorem we conclude that the string

$$(p_{n-2}, p_{n-3}, \cdots, p_1) = \frac{1}{f_{n+1}}(f_2, f_3, \cdots, f_{n-1})$$

is the optimal solution of the primal LP problem given by (10). Therefore, the corresponding WSD with maximum expected codeword length is indeed the Fibonacci distribution

$$(p_n, p_{n-1}, p_{n-2}, p_{n-3}, \cdots, p_1) = \frac{1}{f_{n+1}}(f_1, f_1, f_2, \cdots, f_{n-1}) = \frac{1}{f_{n+1}}F_n.$$

Corresponding to $\left(\frac{f_2}{f_{n+1}}, \frac{f_3}{f_{n+1}}, \cdots, \frac{f_{n-1}}{f_{n+1}}\right)$ we have

$$L' = \frac{1}{f_{n+1}} \{f_2 + 2f_3 + \cdots + (n-2)f_{n-1}\}.$$

It is easily shown by induction that

$$f_2 + 2f_3 + \cdots + (n-2)f_{n-1} = (n-1)f_{n+1} - f_{n+3} + 3.$$

This implies

$$L = n - 1 - \frac{(n-1)f_{n+1} - f_{n+3} + 3}{f_{n+1}} = \frac{f_{n+3} - 3}{f_{n+1}} = 1 + \frac{f_{n+2} - 3}{f_{n+1}}. \blacksquare$$

Corollary 1 It follows from $\lim_{n\to\infty} \frac{f_{n+2}}{f_{n+1}} = \frac{1+\sqrt{5}}{2}$ that

$$L = 1 + \frac{f_{n+2}}{f_{n+1}} - \frac{3}{f_{n+1}} < 1 + \frac{1+\sqrt{5}}{2} \simeq 2.618. \blacksquare$$

References

[1] N. D. Cahill, J. R. D'Errico, D. A. Narayan and J. Y. Narayan, "Fibonacci Determinants," *The College Mathematics Journal of The Mathematical Association of America*, vol. 33, no. 3, pp. 221-225, 2002.

[2] N. D. Cahill, J. R. D'Errico, J. Spence, "Complex factorisations of the Fibonacci and Lucas numbers," *Fibonacci Quarterly*, vol. 41, no. 1, pp. 13–19, Feb. 2003

[3] V. Chvatal, *Linear Programming*, New York: W. H. Freeman and Company, 1983.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[5] M. Esmaeili, "More on the Fibonacci Sequence and Hessenberg Matrices," *Electronic Journal of Combinatorial Number Theory*, 6, #A32, 2006.

[6] R. G. Gallager and D. C. Van Voorhis, "Optimal source coding for geometrically distributed integer alphabets," *IEEE Trans. Inform. Theory*, vol. 21, no. 2, pp. 228–230, 1975.

[7] P. A. Humblet, "Optimal source coding for a class of integer alphabets," *IEEE Trans. Inform. Theory*, vol. 24, pp. 110–112, 1978.

[8] Ching Li, "The Maximum Determinant of an nxn Lower Hessenberg (0,1) matrix," *Linear Algebra and Its Applications*, 183, pp. 147–153, 1993.

[9] G. Strang and K. Borre, *Linear Algebra, Geodesy and GPS*, p. 555-557. Wellesley-Cambridge, 1997.