OEIS A135342

RICHARD J. MATHAR

Abstract. A formula for the number of distinct means of nonempty subsets of $\{1\ldots,n\}$ is derived: the second difference of that counting sequence is essentially Euler's totient φ, as already stated by Siehler's formula in [1, A135342].

1. Encoding of Subsets and Means

Nonempty subsets of the set of numbers from 1 to n can be encoded as binary words $b_1b_2\ldots b_n$ of n letters where $b_i \in \{0,1\}$. The letters operate like a mask: $b_i = 1$ indicates that i is in the subset, $b_i = 0$ indicates that it is not. *Nonempty* requires that the word is $\neq 00\ldots 0$.

Definition 1. *The weight of such a binary word is*

$$w_b \equiv \sum_{i=0}^{n} b_i, \tag{1}$$

the number of the b_i which are 1.

Definition 2. *The (arithmetic) mean of such a subset selecting letters according to b_i is*

$$M_b \equiv \frac{\sum_{i=0}^{n} b_i i}{\sum_{i=0}^{n} b_i}, \tag{2}$$

In mechanics it represents the center of mass of a horizontal scale where unit masses are placed at position i for all $b_i = 1$.

Definition 3. *The torque of such a subset is*

$$t_b \equiv \sum_{i=0}^{n} i b_i, \tag{3}$$

(The descriptive name relates to the interpretation of placing unit masses in the Definition 2).

There are $\binom{n}{w}$ words with weight w. There are $2^n - 1$ words of that kind; the -1 indicates the discarded empty word at $w = 1$. This is the standard formula for the row sums of the Pascal triangle.

The means are rational numbers in the interval $1 \leq M \leq n$. The minimum and maximum are realized by the binary words $1000\ldots 0$ and $000\ldots 001$, i.e., the cases of 'loading' the scale with a unit mass at the left or right limits, i.e., selecting the subsets $\{1\}$ or $\{n\}$.

(2) is a mapping of binary words of n letters to $\mathbb{Q}$.

Date: April 28, 2023.
2020 *Mathematics Subject Classification.* Primary 05A18; Secondary 05A15.

Definition 4. *The number of distinct means in all nonempty subsets of* $\{1, 2, \dots n\}$ *is denoted by* $|\mathcal{M}_n|$.

Example 1. *For* $n = 5$ *the* $|\mathcal{M}_n| = 15$ *possible means* M *are* $\mathcal{M}_5 = \{1, 3/2, 2, 5/2, 7/3, 8/3, 11/4, 3, 7/2, 10/3, 11/3, 13/4, 4, 9/2, 5\}$.

Lemma 1. *The minimum torque for a fixed weight* w *is given by pushing all letters 1 to the beginning of the word,* $b_i = 1$ *for* $1 \le i \le w$*:*

$$t \ge \frac{w(w+1)}{2}. \tag{4}$$

The maximum torque for a fixed weight and right-most 1-letter at position n *is given by starting the word with a run of* $n - w$ *zeros followed by a run of* w *1's,* $b_i = 1$ *for* $n - w < i \le n$*:*

$$t \le \frac{w(2n+1-w)}{2}. \tag{5}$$

The minimum torque for a fixed weight and right-most 1-letter at position n *is given by starting with* $w - 1$ *1's, followed by* $n - w$ *zeros, and finishing with* $b_n = 1$*:*

$$t \ge \frac{(w-1)w}{2} + n. \tag{6}$$

The purpose of this script is to provide a proof for a formula of 2008 for $|\mathcal{M}_n|$ by Jacob Siehler. There are likley others based on properties of Sturmian words, but these are not used here.

2. Overlay of Means

Given some set b_i and its mean (2), plus another set b'_i and its mean $M_{b'}$, one can construct the mean of the *sum* of these sets by putting both sets of loads on the same scale (padding the shorter word by zeros at its end if needed), i.e., considering the mean of the word $\hat{b}$ obtained by adding the bits in the binary words, $\hat{b} \equiv b_i + b'_i$.

Remark 1. *Our problem further down applies this only for words which have no 1's at the same positions, so actually an* `xor` *operation is sufficient, but that arithmetics is more complicated.*

The mean of the overlay is

$$M_{\hat{b}} \equiv \frac{\sum_{i=0}^{n} \hat{b}_i i}{\sum_{i=0}^{n} \hat{b}_i}. \tag{7}$$

A well-knowed formula follows which says that means are added weighted by their weight:

$$M_{\hat{b}} \sum_{i=0}^{n} \hat{b}_i = \sum_{i=0}^{n} \hat{b}_i i = \sum_{i=0}^{n} (b_i + b'_i) i = M_b \sum_{i=0}^{n} b_i + M_{b'} \sum_{i=0}^{n} b'_i; \tag{8}$$

which is the same as saying that torques add arithmetically:

$$t_{\hat{b}} = t_b + t_{b'}. \tag{9}$$

3. Sets of Mean of Fixed Word-length

The set $\mathcal{M}_n$ of distinct means derived from all 2^n-1 binary words has in general an order $|\mathcal{M}_n|$ less than 2^n-1 because some means are mapped to the same rational number. All palindromic words, for example, are mapped to the same mean value $(1+n)/2$.

Proof. For palindromic words this seems to be obvious. The formal arithmetic proof starts from the property $b_i = b_{n-i+1}$ for palindromic words splitting (2) into first half-word, optional middle letter and second half-word, such that

$$M_b = \frac{\sum_{i=0}^{\lfloor n/2\rfloor} b_i i + [n\text{odd}] b_{(n+1)/2}\frac{n+1}{2} + \sum_{i=\lceil n/2+1\rceil}^{n} b_i i}{\sum_{i=0}^{n} b_i} \tag{10}$$

where $[\ldots]$ is Iverson bracket [2]. Therefore

$$\begin{aligned} M_b &= \frac{\sum_{i=0}^{\lfloor n/2\rfloor} b_i i + [n\text{odd}] b_{(n+1)/2}\frac{n+1}{2} + \sum_{i=0}^{\lfloor n/2\rfloor} b_{n-i+1}(n-i+1)}{\sum_{i=0}^{n} b_i} \\ &= \frac{\sum_{i=0}^{\lfloor n/2\rfloor} b_i i + [n\text{odd}] b_{(n+1)/2}\frac{n+1}{2} + \sum_{i=0}^{\lfloor n/2\rfloor} b_i(n-i+1)}{\sum_{i=0}^{n} b_i} \\ &= \frac{\sum_{i=0}^{\lfloor n/2\rfloor} b_i(n+1) + [n\text{odd}] b_{(n+1)/2}\frac{n+1}{2}}{\sum_{i=0}^{n} b_i} \\ &= \frac{(n+1)\sum_{i=0}^{\lfloor n/2\rfloor} b_i + \frac{n+1}{2}[n\text{odd}] b_{(n+1)/2}}{\sum_{i=0}^{n} b_i} \\ &= \frac{\frac{1}{2}(n+1)\sum_{i=0}^{\lfloor n/2\rfloor} 2b_i + \frac{n+1}{2}[n\text{odd}] b_{(n+1)/2}}{\sum_{i=0}^{n} b_i} \\ &= \frac{\frac{1}{2}(n+1)\sum_{i=0}^{n} b_i}{\sum_{i=0}^{n} b_i} = \frac{1}{2}(n+1). \end{aligned} \tag{11}$$

□

Another trivial observation is that all integers $1,\ldots,n$ are in the set $\mathcal{M}_n$ as they are realized for example by the binary word which has $b_i = 1$ at the index equal to that mean and $b_i = 0$ otherwise.

Remark 2. *Multiple realizations of a common integer mean are obtained by symmetrically flipping b_{i+k} and b_{i-k} for one or more k (generalizing the comment about palindromic words above).*

So the remaining core question is: which rational fractions may also appear in $\mathcal{M}_n$ to increase its order beyond n?

4. Incremental Deployment of Words

The superposition principle of Section 2 allows to compute the mean M of a word by starting from the word $000\ldots0$ and incrementally flipping on the bits where $b_i = 1$ one at a time in any order. The mean is not a function of any particular order in which the b_i are flipped from 0 to 1 to generate higher weights; this manuscript selects an order that is suitable to compute the first differences of the set orders $|\mathcal{M}_n|$.

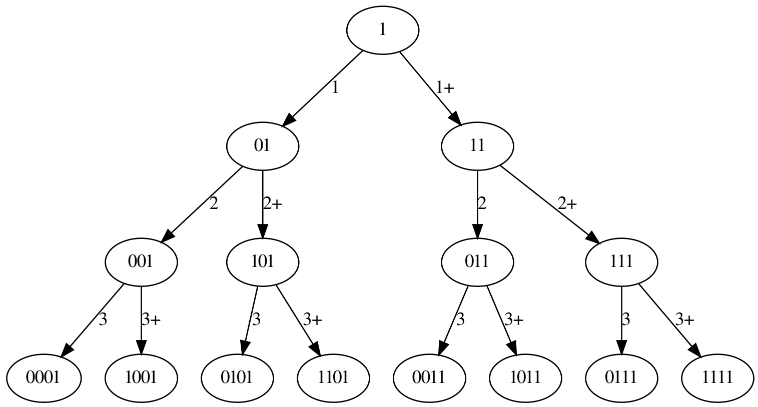

FIGURE 1. Digraph of generating binary words from the root word 1 by applying a shift-insert operation at each node. The first 3 iterations of the operation are shown. Each edge is labeled with the generation that produces the value, followed by a +-sign if the weight was increased by one.

For fixed n we start at the root word 1 with $M = w = 1$ and generate the others (children in the language of trees) by inserting the letter 0 or 1 as the new first letter in front. The digraph of Figure 1 demonstrates the recurrence: the labels of the nodes are the words, and the labels at the edges are the generations (where a + indicates that a 1 was added in front):

That particular deployment ensures that in one new generation all words of the same length with the right-most 1 at the same place appear. (It imposes a weak order on all binary words where the rank is the largest i with $b_i = 1$.) The new generation of words may have means M already produced in earlier generations (shorter words) or new means M not realized by shorter words. Our strategy is to recursive: we count the means M in generations that have now produced earlier, i.e., look at the first differences of $|M_n|$. This is essentially based on the fact that appending zeros to a word does not change its mean.

The root word has $M = 1$, weight $w = 1$. Prefixing a word with 0 or 1 may be split into a two steps: (i) a right-shift that moves all 1-bits one position to the right and insertion of 0 at the then vacant first position, (ii) an optional bit-flip $0 \to 1$ of the letter at the first position. The right-shift moves all letters b_i one place to the right, so the effect on the mean is $M \to 1 + M$; no change of w; $t \to t + w$. (8) tracks the effect of adding 0 or 1 in front of a word:

- Inserting the letter $b'_1 = 0$ ($w = 0$) does not modify M or w or t further
- Inserting letter $b'_1 = 1$ ($w' = M_{b'} = 1$) modifies M further

$$M \to \frac{Mw + 1}{w + 1} \tag{12}$$

which decreases M and increases $w \to w + 1$, $t \to t + 1$.

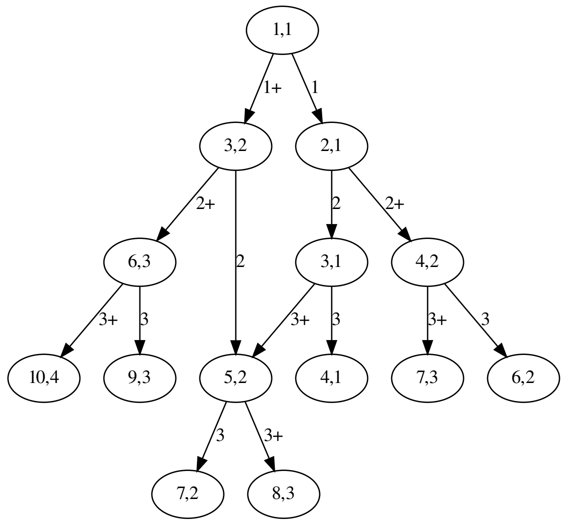

FIGURE 2. Digraph of generating binary (t, w) pairs, which are the labels at each node, from the root word 1 by applying a shift-insert operation at each node. The first 3 iterations of the operation are shown.

The net effect on mean and torque of generating a new word (child node) prefixing one letter 0 or 1 is:

$$M \to \begin{cases} 1 + M, & b_1' = 0 \\ \frac{(1+M)w+1}{w+1} = 1 + M - \frac{M}{w+1}, & b_1' = 1. \end{cases} \tag{13}$$

$$t \to \begin{cases} t + w, & b_1' = 0 \\ t + w + 1, & b_1' = 1. \end{cases} \quad w \to \begin{cases} w, & b_1' = 0 \\ w + 1, & b_1' = 1. \end{cases} \tag{14}$$

Remark 3. *This equation demonstrates that a (t, w) pair induces a $(t+w, w)$ pair and a $(t + w + 1, w + 1)$ pair. To answer the question whether some (t, w) pair appears in the deployment, a backtracking algorithm is available: (t, w) may have been a child of $(t - w, w)$ or a child of $(t - w, w - 1)$. (t, w)-values that violate (4) are not deployable, which helps to prune that backtracking algorithm.*

Figure 2 is a reduction of Figure 1 where the labels at the edges still indicate the generation and two choices of prefixing 0 or 1, but the node labels are t and w. This digraph is no longer a tree because (t, w) that are common to a subset of words have been merged into single nodes.

The further step of labelling nodes by $M = t/w$ and w leads from Figure 2 to Figure 3.

Because the main interest of this manuscript is to count distinct M without regard to w, the final reduction is to drop the w in the labels to discard that information and to merge nodes in the deployment graph with a common M. This

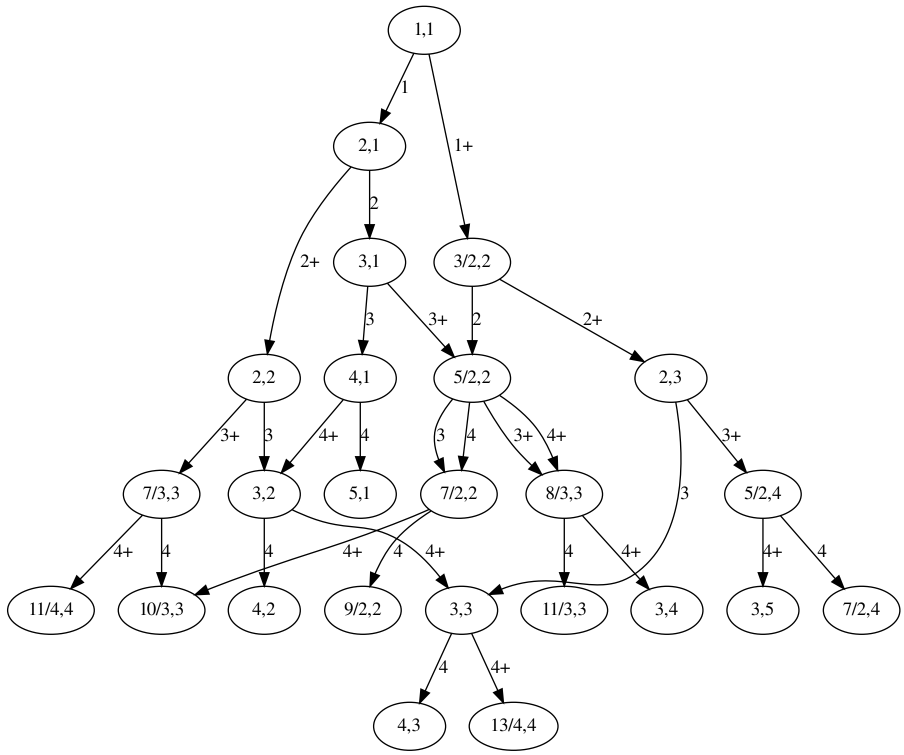

Figure 3. Digraph of distinct pairs of generated (M, w) separated by a comma in the labels, after running 4 times the shift-insert operation.

translates Figure 3 to Figure 4. It has 15 nodes after 4 shift-insertion steps, compliant with sequence [1, A135342]. At that level of representation the ratios $M = t/w$ are reduced to coprime t and w.

This graphical representation is puzzling, therefore transformed into Table 1. The table puts M-fractions with the same graph-theoretical distance to the root node (length of the shortest path between the nodes) into a single row. Row n contains the M with distance $n-1$ to the node $M = 1$ and smallest label $n-1$ of all edges incident to the node.

Example 2. *$M = 7/2$ is in row 4 because it can be reached (via $1 \to 3/2 \to 5/2$) in 3 steps from $M = 1$. It can also be reached in 4 steps from $M = 1$, but the smaller number of steps takes precedence.*

5. Statistics of Incremental Subsets of Means

Due to the particular deployment order chosen in Section 4, row n in Table 1 contains the M which appear in words with highest 1-bit at position n but not in words with highest 1-bit at position less than n.

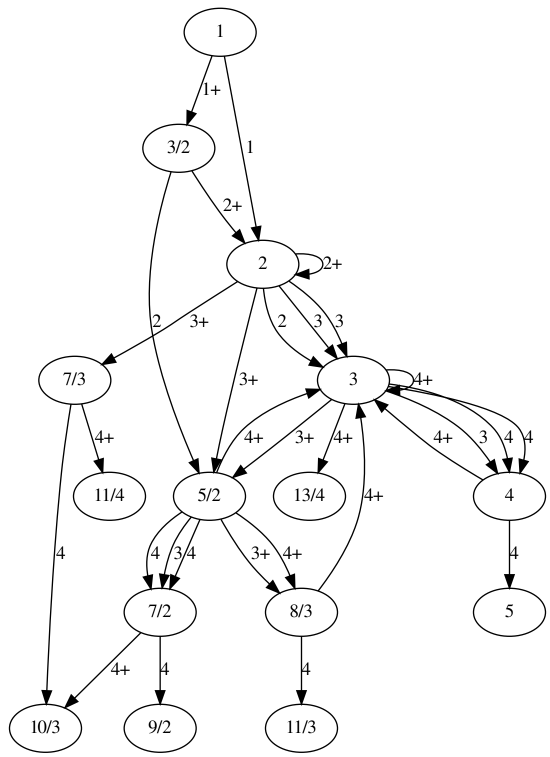

FIGURE 4. Deployment graph with nodes labeled by M, after running 4 times the shift-insert operation. The loops (edges returning to the same node) are cases where $M/(w+1) = 1$ on the right hand side in the second line of (13).

n	M											
1	1											
2	2	3/2										
3	3	5/2										
4	4	7/2	7/3	8/3								
5	5	9/2	10/3	11/3	11/4	13/4						
6	6	11/2	13/3	14/3	15/4	17/4	16/5	17/5	18/5	19/5		
7	7	13/2	16/3	17/3	19/4	21/4	21/5	22/5	23/5	24/5	23/6	25/6

TABLE 1. Fractions M generated by the deployment algorithm earliest by words with highest 1-bit at position n.

Each M is a ratio $M = t/w$ where numerator t and denominator w are coprime. For torques t and weights w which have common factors the associate entries in the table are not given by canceling common factors but by backtracking as in Remark 3.

Example 3. *The case* $(t = 12, w = 4)$ *is generated either by* $(t = 8, w = 4)$*— which is impossible because it violates* (4)*—, or by* $(t = 8, w = 3)$ *which actually is represented by* $M = 8/3$ *in the table.*

Example 4. *The nodes* $(3, 1)$*,* $(6, 2)$ *and* $(9, 3)$ *in Figure 2 are all represented by* $M = 3$ *in the table. Here backtracking gives the same value as by just canceling the common factors in* $9/3$ *and* $6/2$*.*

The structure of the table is as follows:

(1) Denominators of M in row n are at most $n - 1$. This is manifest in (2) because the weights of words in that row are $\leq n$. M-values in a row have been ordered first by increasing denominator, second by increasing numerator.

(2) 'Injection' values are the M at the heads of the columns.

Example 5. *Injection values in row 5 are* $11/4$ *and* $13/4$*. Injection values in row 6 are* $16/5$*,* $17/5$*,* $18/5$ *and* $19/5$*. Injection values in row 7 are* $23/6$ *and* $25/6$*.*

(3) The numerators t and denominators w must obey (6) and (5) to be reachable in $n - 1$ deployment steps. The actual lower limit is more restrictive: the t values up to (5) of the *previous* n are already reachable by earlier steps, which leads to the additional requirement

$$t > \frac{w[2(n-1) + 1 - w]}{2}. \tag{15}$$

The t-values that do not meet that condition are listed higher up in the same column at the same w, which mean they do not yield injection values. This leads to the arithmetic progressions of values down the columns of the table, created by the recurrence of the first line in (13).

(4) If $w = n - 1$, the requirement (15) is weaker than (4), which means the t-values have not been reached earlier — strictly equivalent to the upper limit of w described in item (1). Requirements (4) and (5) simplify to

$$\frac{n(n-1)}{2} \leq t \leq \frac{(n-1)(n+2)}{2}. \tag{16}$$

There are $n - 1$ consecutive integers in that closed interval, actually implemented by the words $0111\ldots1$, $1011\ldots1$, up to $111\ldots101$ with a single zero. Because the t must be coprime to w (to $n-1$ in this case), the number of injection points is $\varphi(n - 1)$.

6. Conclusion

In summary, the number of M values in row n of Table 1 is the same as in row $n - 1$ (created by continuing the arithmetic progressions along the columns) plus $\varphi(n - 1)$ injection points:

$$r_n = r_{n-1} + \varphi(n - 1). \tag{17}$$

The number of entries in the table up to and including row n is by definition and according to our way of deployment

$$|\mathcal{M}_n| = \sum_{i=1}^{n} r_i \quad \therefore \quad |\mathcal{M}_n| - |\mathcal{M}_{n-1}| = r_n. \tag{18}$$

Combining these two equations for the first differences yields for the second differences

$$|\mathcal{M}_n| - 2|\mathcal{M}_{n-1}| + |\mathcal{M}_{n-2}| = \varphi(n-1), \tag{19}$$

which is Siehler's equation in [1, A135342].

References

1. O. E. I. S. Foundation Inc., *The On-Line Encyclopedia Of Integer Sequences*, (2023), https://oeis.org/. MR 3822822
2. Donald E. Knuth, *Two notes on notation*, Am. Math. Monthly **99** (1992), no. 5, 403–422. MR 1163629

URL: `http://www.mpia.de/~mathar`

Hoeschstr. 7, 52372 Kreuzau, Germany