Counting snail venom

Geoff Smith, 1998

Introduction

Let $\Sigma = \{B, W\}$ be a two letter alphabet. Let Σ^* denote the set of *words* on this alphabet – that is to say finite sequences each entry of which is drawn from the alphabet. The number of terms in a sequence is its *length*. Notice that the number of words of length n is 2^n since each of the n terms may be chosen independently. The number of words involving exactly r copies of B and s copies of W is $(r+s)!/r!s!$. This quantity is known as the *binomial coefficient* $\binom{r+s}{s}$ and arises in algebra via

$$(x+y)^n = \sum_{\substack{r,s\geq 0 \\ r+s=n}} \binom{r+s}{s} x^r y^s. \qquad (1)$$

where x and y are unknowns. These quantities $\binom{r+s}{s}$ also form the entries of *Pascal's triangle* where $r+s$ is the row number and r indexes how far the entry is from the left (and s from the right). Note that $0! = 1$ in order to make everything work smoothly. Putting $x = y = 1$ in (1) yields that there are 2^n words of length n, a fact which we have already noted. The numbers from Pascal's triangle crop up frequently when performing enumeration arguments.

We will need a fact about these binomial coefficients. Notice that

$$-1 + \binom{m}{1} - \binom{m}{2} + \binom{m}{3} \ldots (-1)^m \binom{m}{m} = 0 \text{ when } m \geq 1$$

because then $-(1-1)^m = 0$ and the binomial theorem applies. This really says that if you add up the entries in any row of Pascal's triangle (except the top one), alternating the signs as you go, the answer is 0.

How big is a union of finitely many finite sets?

Suppose that you have two finite sets A and B. You can find the size of their union using

$$|A \cup B| = |A| + |B| - |A \cap B|$$

because when you work out $|A| + |B|$ the elements of $|A \cap B|$ are being 'counted twice'. You compensate for this by subtracting $|A \cap B|$.

Now suppose you have three finite sets. A very careful analysis of counting will show you that

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \quad (2)$$

What is going on here is that you first try $|A|+|B|+|C|$ but this is wrong because elements in $A\cap B$, $A\cap C$, and $B\cap C$ have been counted too much. You therefore try to eliminate this over-counting by subtracting $|A \cap B| + |A \cap C| + |B \cap C|$, but then notice that elements of $A \cap B \cap C$ have been 'over removed'. You compensate for this by adding $|A \cap B \cap C|$ and all is well. We are edging toward the inclusion-exclusion enumeration principle. Let is look at a concrete example. Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ and $C = \{2, 3, 4, 5, 6, 7\}$. Now (2) says $7 = 4 + 4 + 6 - 2 - 3 - 4 + 2$, which happily is true.

We prove validity of the Inclusion-Exclusion counting principle.

Theorem Suppose $n \in \mathbb{N}$ and A_i is a finite set for $1 \le i \le n$. It follows that

$$\begin{aligned}\left|\bigcup_{1\le i\le n} A_i\right| &= \sum_{1\le i_1\le n} |A_{i_1}| - \sum_{1\le i_1\le i_2\le n} |A_{i_1} \cap A_{i_2}| \\ &\quad + \sum_{1\le i_1\le i_2\le i_3\le n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \ldots + (-1)^{n+1}\left|\bigcap_{i=1}^{n} A_i\right|.\end{aligned}$$

Proof

Induct on $| \cup_i A_i|$. If all A_i are empty, then the theorem holds, and the induction starts without difficulty. Now focus on the case where $| \cup_i A_i| > 0$. Pick any $x \in \cup_i A_i$ and form new sets by $B_i = A_i \setminus \{x\}$ (i.e. remove x from all the sets A_i which contain it). Now $|\cup_i B_i| = |\cup_i A_i| - 1$ so the theorem holds for the sets B_i by induction. Now pop x back in to all sets from which you deleted it. The re-insertion of x makes a contribution of $+1$ on the left hand side. On the right, suppose x occurs in exactly $m(\ge 1)$ of the A_i. The re-insertion of x has the effect of adding exactly $m - \binom{m}{2} + \binom{m}{3} \ldots + (-1)^m \binom{m}{m}$ to the right hand side. However, we have prepared the ground in the previous section, so we know that this quantity is 1. The induction step is complete, and thus the inclusion-exclusion principle is valid.

Example Application 1

The Euler φ-function is $\varphi : \mathbb{N} \to \mathbb{N}$ defined by

$$\varphi(n) = |\{k \mid 1 \le k \le n, \quad \text{g.c.d.}(k, n) = 1\}|.$$

We have not made this function up. It is an important function in number theory, combinatorics and algebra, and it has sweet properties. For example, if $a, b \in \mathbb{N}$ and a, b are coprime (i.e. have 1 as their greatest common divisor), then $\varphi(ab) = \varphi(a)\varphi(b)$. In fact this follows immediately from the next proposition.

Proposition Let the prime divisors of n be $p_1, p_2, \ldots, p_k$ (without repetition). It follows that

$$\varphi(n) = n \prod_{i=1}^{k} (1 - p_i^{-1}).$$

Proof The theorem is trivially true if $n = 1$, since the empty product is 1. Thus we may assume $n > 1$, and thus $k \geq 1$. For each i in the range $1 \leq i \leq k$ we put

$$A_i = \{t \mid 1 \leq t \leq n,\ n/p_i \in \mathbb{N}\}.$$

Here we have eschewed the vertical line notation $p_i \mid n$ deliberately in order to avoid notational collision.

Notice that $|A_i| = n/p_i$. Moreover, if $i \neq j$ then $|A_i \cap A_j| = n/p_i p_j$ and so on. The inclusion-exclusion principle enables us to count the natural numbers between 1 and n (inclusive) which are *not* coprime to n in two ways.

$$n - \varphi(n) = n \sum_i 1/p_i - n \sum_{i_1 < i_2} n/p_{i_1} p_{i_2} + \ldots.$$

Rearrange, and a use a little algebra to obtain the result.

Example Application 2

Deal two packs of shuffled cards simultaneously. What is the probability that no pair of identical cards will be exposed simultaneously?

Fix the first pack, and consider all possible rearrangements of the second pack. For each i in the range $1 \leq i \leq 52$ let A_i denote the set of all arrangements of the second pack which happen to have the property that the card in position i matches the card in position i of the first pack. Obviously $|A_i| = 51!$ for every i. Moreover, if $i \neq j$, then $|A_i \cap A_j| = 50!$ and so on. Let $X = \cup_i A_i$, so the probability of at least one match is $|X|/52!$. We calculate this using the inclusion-exclusion principle.

$$\begin{aligned} |X|/52! &= (52!)^{-1} \left(\tbinom{52}{1} 51! - \tbinom{52}{2} 50! + \tbinom{52}{3} 49! - \ldots - \tbinom{52}{52} 0! \right) \\ &= 1 - 1/2! + 1/3! - \ldots - 1/52! \approx 1 - (\sum_{i=0}^{\infty} (-1)^i / i!) = 1 - 1/e. \end{aligned}$$

Thus the probability of no coincidences is (to an excellent approximation) $1/e$.

Here we have use the fact that

$$e^x = 1 + x + x^2/2! + \ldots = \sum_{i=0}^{\infty} x^i / i!$$

and put $x = -1$.

Enumerations related to snail venom.

We define a *language* L to be an arbitrary subset of Σ^*. Associated to a language we have a power series p_L in two variables b and w where

$$p_L = p_L(b, w) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} c_{mn} b^m w^n$$

and c_{mn} is the number of words in L which involve exactly m copies of the letter B and n copies of the letter W.

(1) If the language L_1 consists of all words which do not involve W then $c_{mn} = 0$ whenever $n > 0$, and $c_{m0} = 1$ for every m since there is exactly one word in the language of length m involving no W. Thus the power series is

$$p_{L_1} = \sum_{m=0}^{\infty} b^m = (1 - b)^{-1}.$$

(2) If the language L_2 consists of all words in Σ^* of length t where t is fixed, then

$$p_{L_2} = \sum_{\substack{m,n \geq 0 \\ n+m=t}} \frac{t!}{n!m!} b^m c^n.$$

If we further specify that $t = 3$ to give the language $\hat{L}_2$ we have

$$p_{\hat{L}_2} = b^3 w^0 + 3b^2 w^1 + 3b^1 w^2 + b^0 w^3 = b^3 + 3b^2 w + 3bw^2 + w^3.$$

In this case the power series collapse to being a polynomial expression, since all save a finite number of coefficients vanish. Notice that the coefficients here are drawn from a row of *Pascal's triangle*, and another way of describing p_{L_2} is $(b + w)^3$. To understand this, think of $+$ as being "or" and multiplication (juxtaposition) as being "and then". In this sense $(b + w)^3 = (b + w)(b + w)(b + w)$ exactly describes the $\hat{L}_2$. Each word in $\hat{L}_2$ is of the form: B or W, then B or W, then B or W.

(3) If the language L_3 consists of all words where there are no juxtaposed Ws and no juxtaposed Bs then there is one word in L_3 of length 0, the empty word, and exactly two words of every other length. Thus

$$p_{L_3} = 1 + \sum_{n=0}^{\infty} b^n w^{n+1} + \sum_{n=0}^{\infty} b^{n+1} w^n + \sum_{n=1}^{\infty} 2b^n w^n.$$

The particular languages of interest in the investigation of snail venom are the ones satisfying the following five conditions.

(i) The words involve an even number of occurrences of B.

(ii) The words begin and end with a letter B.

(iii) Words must not contain the subword $WWWWWWW$.

(iv) Words must not contain the subword BBB (i.e. you must not have three consecutive occurrences of B).

(v) Suppose a word involves exactly $2k$ occurrences of the letter B. Label them from the left the $1st, 2nd, \ldots 2k$-th. Between at least one $(2i-1)$-th and $2i$-th occurrence of B, there must at least two consecutive occurrences of W.

The language S specified by the first three conditions are all possible words associated with snail venom. The language defined by all five these conditions are the so-called inaccessible words, and are exactly those which do not admit of synthesis by a specific manufacturing process.

The language S is defined by conditions (i), (ii) and (iii). Let $f = (1 + w + w^2 + w^3 + w^4 + w^5 + w^6)$. The definition of p_S yields that

$$p_S = f + bfb + bfbfbfb + bfbfbfb + \ldots$$

or more formally

$$p_S = \sum_{i=0}^{\infty} b^{2i} f^{2i-1}$$

If you wish to see you many words there are in this language of length 30 and involving 8 occurrences of B (and so 22 occurrences of W) you look at the coefficient of $b^8 w^{22}$ and this is precisely the coefficient of w^{22} in f^7. This can be determined in a trice by almost any computer algebra system (AXIOM, MAPLE, REDUCE etc).

Now we proceed to the main calculation, where we also impose conditions (iv) and (v), and are particularly interested in the cases where there are 4, 6 and 8 occurrences of B, and when the word lengths does not happen to exceed 30. This final constraint can be imposed at the very end of the calculations by truncating the appropriate polynomial if necessary.

The Main Calculation

All these calculations are the same in spirit, but the details become a little more complicated as we proceed.

We introduce a formal and compact way of describing languages. The letters B and W stand for themselves. The letter G stands for *either* W *or no letter*, the letter F stands for *consecutive* W*s, somewhere between 0 and 6 in number.* We interpret $\cup$ to mean the union of languages.

Four occurrences of B

Consider the three languages $C_1 = BFBFBFB$, $C_2 = BBBFB \cup BFBBB$ and $C_3 = BGBFBGB$. The language (of inaccessible words) we are interested in is

$$C = (C_2 \cup C_3)\,.$$

The polynomial for C_1 is $b^4 f^3$. C_2 is described by $b^4(2f-1)$ and C_3 is described by $b^4 g^2 f$. $C_2 \cap C_3$ is described by $b^4(2g-1)$. Thus the inaccessible language $C_2 \cup C_3$ is described by

$$b^4(2f - 1 + g^2 f - (2g-1)) = b^4(2f + g^2 f - 2g).$$

Of course, since every single polynomial in this analysis involves the factor b^4, we could have suppressed it. We didn't for expository reasons.

Six occurrences of B

In this section we will suppress the ever present factor b^6. Consider the three languages $D_1 = BFBFBFBFBFB$, D_2 the language consisting of words containing three or more consecutive Bs, and $D_3 = BGBFBGBFBGB$. The polynomial describing D_1 is f^5 and that describing D_3 is $g^3 f^2$.

The language D_2

$D_2 = R \cup S \cup T \cup U$ where

		Language	Polynomial
R	$=$	$BBBFBFBFB$	f^3
S	$=$	$BFBBBFBFB$	f^3
T	$=$	$BFBFBBBFB$	f^3
U	$=$	$BFBFBFBBB$	f^3

The sum of the polynomials of these 4 languages is $4f^3$. We use the inclusion-exclusion principle to determine the polynomial describing D_2

		Language	Polynomial
$R \cap S$	$=$	$BBBBFBFB$	f^2
$R \cap T$	$=$	$BBBBBFB$	f
$R \cap U$	$=$	$BBBFBBB$	f
$S \cap T$	$=$	$BFBBBBFB$	f^2
$S \cap U$	$=$	$BFBBBBB$	f
$T \cap U$	$=$	$BFBFBBBB$	f^2

The sum of the polynomials of these 6 languages is $3f + 3f^2$.

		Language	Polynomial
$R \cap S \cap T$	$=$	$BBBBBFB$	f
$S \cap T \cap U$	$=$	$BFBBBBB$	f
$R \cap T \cap U$	$=$	$BBBBBB$	1
$R \cap S \cap U$	$=$	$BBBBBB$	1

The sum of the polynomials of these 4 languages is $2f+2$.

		Language	Polynomial
$R\cap S\cap T\cap U$	$=$	$BBBBBB$	1

The polynomial of this language is 1.

The language D_2 of words involving at least 3 consecutive Bs is described by the polynomial $4f^3-(3f+3f^2)+(2f+2)-1$, using the inclusion-exclusion principle.

The language $D_2\cup D_3$

$D_2\cap D_3$ is the union of the following four languages.

		Language	Polynomial
L	$=$	$BBBGBFBGB$	g^2f
M	$=$	$BGBBBFBGB$	g^2f
N	$=$	$BGBFBBBGB$	g^2f
0	$=$	$BGBFBGBBB$	g^2f

.

The sum of the polynomials of these 4 languages is $4g^2f$. We must calculate the polynomials associated with various intersections.

		Language	Polynomial
$L\cap M$	$=$	$BBBBFBGB$	gf
$L\cap N$	$=$	$BBBBBGB$	g
$L\cap O$	$=$	$BBBGBBB$	g
$M\cap N$	$=$	$BGBBBBGB$	g^2
$M\cap O$	$=$	$BGBBBBB$	g
$N\cap O$	$=$	$BGBFBBBB$	gf

The sum of the polynomials of these 6 languages is $3g+2gf+g^2$.

		Language	Polynomial
$L\cap M\cap N$	$=$	$BBBBBGB$	g
$L\cap M\cap O$	$=$	$BBBBBBB$	1
$L\cap N\cap O$	$=$	$BBBBBBB$	1
$M\cap N\cap O$	$=$	$BGBBBBBB$	g

The sum of the polynomials of these 4 languages is $2g+2$.

		Language	Polynomial
$L\cap M\cap N\cap O$	$=$	$BBBBBB$	1.

The polynomial of this languages is 1.

Thus $D_2\cap D_3$ is described by $4g^2f-(g^2+2gf+3g)+(2g+2)-1$.

We conclude that the inaccessible language $D_2\cup D_3$ is described by the polynomial

$$4f^3-(3f+3f^2)+(2f+2)-1+g^3f^2-(4g^2f-(g^2+2gf+3g)+(2g+2)-1)$$

which simplifies a little to yield

$$4f^3-f-3f^2+g^4f^3-4g^2f+g^2+2gf+g.$$

Eight occurrences of B

In this section we will suppress the ever present factor b^8. Consider the three languages $E_1=BFBFBFBFBFBFBFB$, E_2 the language consisting of words containing three or more consecutive Bs, and $E_3=BGBFBGBFBGBFBGB$. The polynomial describing E_1 is f^7 and that describing E_3 is g^4f^3.

The language E_2

$E_2=R\cup S\cup T\cup U\cup V\cup W$ where

		Language	Polynomial
R	$=$	$BBBFBFBFBFBFB$	f^5
S	$=$	$BFBBBFBFBFBFB$	f^5
T	$=$	$BFBFBBBFBFBFB$	f^5
U	$=$	$BFBFBFBBBFBFB$	f^5
V	$=$	$BFBFBFBFBBBFB$	f^5
W	$=$	$BFBFBFBFBFBBB$	f^5

The sum of the polynomials of these 6 languages is $6f^5$.

We must walk the inclusion-exclusion road. From now on, we may sometimes omit the full language description when it is clear.

		Language	Polynomial
$R\cap S$	$=$	$BBBBFBFBFBFB$	f^4
$R\cap T$	$=$	$BBBBBFBFBFB$	f^3
$R\cap U$	$=$	$BBBFBBBFBFB$	f^3
$R\cap V$	$=$	$BBBFBFBBBFB$	f^3
$R\cap W$	$=$	$BBBFBFBFBBB$	f^3
$S\cap T$	$=$	$BFBBBBFBFBFB$	f^4
$S\cap U$	$=$	$BFBBBBBFBFB$	f^3
$S\cap V$	$=$	$BFBBBFBBBFB$	f^3
$S\cap W$	$=$	$BFBBBFBFBBB$	f^3
$T\cap U$	$=$	$BFBFBBBBFBFB$	f^4
$T\cap V$	$=$	$BFBFBBBBBFB$	f^3
$T\cap W$	$=$	$BFBFBBBFBBB$	f^3
$U\cap V$	$=$	$BFBFBFBBBBFB$	f^4
$U\cap W$	$=$	$BFBFBFBBBBB$	f^3
$V\cap W$	$=$	$BFBFBFBFBBBB$	f^4

The sum of the polynomials of these 15 languages is $5f^4 + 10f^3$.

From now on we may group together similar languages under the heading *Siblings:* often languages which are the reversals of one another.

		Language	Polynomial	Siblings
$R \cap S \cap T$	$=$	$BBBBBFBFBFB$	f^3	4
$R \cap S \cap U$	$=$	$BBBBBBFBFB$	f^2	6
$R \cap S \cap V$	$=$	$BBBBFBBBFB$	f^2	4
$R \cap T \cap V$	$=$	$BBBBBBBFB$	f	2
$R \cap S \cap W$	$=$	$BBBBFBFBBB$	f^2	2
$R \cap T \cap W$	$=$	$BBBBBFBBB$	f	2

The sum of the polynomials of these 20 languages is $4f^3 + 12f^2 + 4f$.

		Language	Polynomial	Siblings
$R \cap S \cap T \cap U$	$=$	$BBBBBBFBFB$	f^2	3
$R \cap S \cap T \cap V$	$=$	$BBBBBFBBBFB$	f^2	4
$R \cap S \cap T \cap W$	$=$	$BBBBBFBFBBB$	f^2	2
$R \cap S \cap U \cap V$	$=$	$BBBBBBBFB$	f	2
$R \cap S \cap U \cap W$	$=$	$BBBBBBBB$	1	2
$R \cap S \cap V \cap W$	$=$	$BBBBFBBBB$	f	1
$R \cap T \cap U \cap W$	$=$	$BBBBBBBB$	1	1

The sum of the polynomials of these 15 languages is $9f^2 + 3f + 3$.

		Language	Polynomial	Siblings
$R \cap S \cap T \cap U \cap V$	$=$	$BBBBBBBFB$	f	2
4 others	$=$	$BBBBBBBB$	1	4

The sum of the polynomials of these 6 languages is $2f + 4$.

		Language	Polynomial	Siblings
All six	$=$	$BBBBBBBB$	1	1

The polynomial of this language is 1.

The polynomial describing this language is

$$6f^5 - (5f^4 + 10f^3) + (4f^3 + 12f^2 + 4f) - (9f^2 + 3f + 3) + (2f + 4) - 1$$

The language $E_2 \cup E_3$

We use an inclusion-exclusion argument, and express $E_2 \cap E_3$ as the union of the following six languages.

		Language	Polynomial
L	$=$	$BBBGBFBGBFBGB$	$g^3 f^2$
M	$=$	$BGBBBFBGBFBGB$	$g^3 f^2$
N	$=$	$BGBFBBBGBFBGB$	$g^3 f^2$
O	$=$	$BGBFBGBBBFBGB$	$g^3 f^2$
P	$=$	$BGBFBGBFBBBGB$	$g^3 f^2$
Q	$=$	$BGBFBGBFBGBBB$	$g^3 f^2$

The sum of the polynomials of these 6 languages is $6g^3f^2$.

		Language	Polynomial	
$L \cap M$	=	$BBBBFBGBFBGB$	g^2f^2	2
$M \cap N$	=	$BGBBBBGBFBGB$	g^3f	2
$N \cap 0$	=	$BGBFBBBBFBGB$	g^2f^2	1
others	=	various	g^2f	10

The sum of the polynomials describing these 15 languages is $10g^2f+3g^2f^2+2g^3f$

		Language	Polynomial
$L \cap M \cap N$	$BBBBBGBFBGB$	g^2f	2
$M \cap N \cap 0$	$BGBBBBBFBGB$	g^2f	2
$L \cap M \cap 0$	$BBBBBBFBGB$	gf	2
$L \cap N \cap 0$	$BBBBBBFBGB$	gf	2
$M \cap N \cap P$	$BGBBBBBBGB$	g^2	2
$L \cap M \cap P$	$BBBBFBBBGB$	gf	2
$L \cap M \cap Q$	$BBBBFBGBBB$	gf	2
$L \cap O \cap P$	$BBBGBBBBGB$	g^2	2
$L \cap N \cap P$	$BBBBBBBGB$	g	2
$L \cap N \cap Q$	$BBBBBGBBB$	g	2

The sum of the polynomials describing these 20 languages is $4g^2f+8gf+4g^2+4g$.

		Language	Polynomial	
$L \cap M \cap N \cap 0$	=	$BBBBBBFBGB$	gf	2
$M \cap N \cap 0 \cap P$	=	$BGBBBBBBGB$	g^2	1
$L \cap M \cap N \cap P$	=	$BBBBBBBGB$	g	8
$L \cap M \cap P \cap Q$	=	$BBBBFBBBB$	f	1
$L \cap N \cap O \cap Q$	=	$BBBBBBBB$	1	3

The sum of the polynomials describing these 15 languages is $2gf+g^2+8g+f+3$.

$L \cap M \cap N \cap O \cap P$	=	$BBBBBBBGB$	g	2
4 others	=	$BBBBBBBB$	1	4

The sum of the polynomials describing these 6 languages is $2g + 4$.

$$L \cap M \cap N \cap O \cap P \cap Q \quad = \quad BBBBBBBB \quad 1 \quad 1$$

The polynomial of this language is 1.

The polynomial describing the language $E_2 \cap E_3$ is therefore

$$6g^3f^2 - (3g^2f^2 + 2g^3f + 10g^2f) + (4g^2f + 4g^2 + 4g + 8gf)$$
$$-(2gf + f + g^2 + 8g + 3) + (2g + 4) - 1$$

The polynomial describing the (inaccessible) language $E_2 \cup E_3$ is therefore

$$6f^5 - (5f^4 + 10f^3) + (4f^3 + 12f^2 + 4f) - (9f^2 + 3f + 3) + (2f + 4) - 1 + g^4f^3$$
$$- \big[(6g^3f^2 - (3g^2f^2 + 2g^3f + 10g^2f + (4g^2f + 4g^2 + 4g + 8gf)$$
$$-(2gf + f + g^2 + 8g + 3) + (2g + 4) - 1\big]$$

The escargot polynomials and associated numbers

By truncating the polynomials we can focus on the case that the words in question have length no more than 30. By evaluating w to be 1, we count the words (we call these words *strings*). However, in the chemical application, we can allow for the 19 possible amino acids which may occur when a W is present by putting $w = 19$ (we call these words *sequences*). We let e_m be the polynomial which enumerates all words satisfying conditions (i), (ii) and (iii) and involve m copies of B, and are truncated to allow for the total length ≤ 30 condition (the escargot polynomials). Let f_m be the polynomials of inaccessibility, which are defined in the same way, save that conditions (iv) and (v) are also imposed.

Four occurrences of B.

$$f_4 = 1 + 3\,w + 6\,w^2 + 6\,w^3 + 6\,w^4 + 6\,w^5 + 6\,w^6 + 3\,w^7 + w^8$$

$f_4(1) = 38$ inaccessible strings of length at most 30.

$f_4(19) = 19963135442$ inaccessible sequences of length at most 30.

$$e_4(w) = 1 + 3\,w + 6\,w^2 + 10\,w^3 + 15\,w^4 + 21\,w^5 + 28\,w^6$$
$$+33\,w^7 + 36\,w^8 + 37\,w^9 + 36\,w^{10} + 33\,w^{11} + 28\,w^{12} + 21\,w^{13}$$
$$+15\,w^{14} + 10\,w^{15} + 6\,w^{16} + 3\,w^{17} + w^{18}$$

$e_4(1) = 343$ strings of length at most 30.

$e_4(19) = 1224639048862059586774 21$ sequences of length at most 30.

Six occurrences of B.

$$f_6 = 1 + 5\,w + 15\,w^2 + 35\,w^3 + 60\,w^4 + 89\,w^5$$
$$+122\,w^6 + 154\,w^7 + 175\,w^8 + 180\,w^9 + 171\,w^{10}$$
$$+154\,w^{11} + 129\,w^{12} + 96\,w^{13} + 65\,w^{14} + 41\,w^{15}$$
$$+24\,w^{16} + 12\,w^{17} + 4\,w^{18}$$

$f_6(1) = 1532$ inaccessible strings of length at most 30.

$$f_6(19) = 489875340706634924622560$$

inaccessible sequences of length at most 30.

$$e_6 = 1 + 5\,w + 15\,w^2 + 35\,w^3 + 70\,w^4 + 126\,w^5 + 210\,w^6$$
$$+325\,w^7 + 470\,w^8 + 640\,w^9 + 826\,w^{10} + 1015\,w^{11}$$
$$+1190\,w^{12} + 1330\,w^{13} + 1420\,w^{14} + 1451\,w^{15}$$
$$+1420\,w^{16} + 1330\,w^{17} + 1190\,w^{18} + 1015\,w^{19}$$
$$+826\,w^{20} + 640\,w^{21} + 470\,w^{22} + 325\,w^{23} + 210\,w^{24}$$

$$e_6(1) = 16555 \text{ strings of length at most 30.}$$

$$e_6(19) = 1119403018505084128111935530758381$$

sequences of length at most 30.

Eight occurrences of B

$$f_8 = 1 + 7\,w + 28\,w^2 + 84\,w^3 + 210\,w^4 + 442\,w^5 + 817\,w^6 + 1371\,w^7$$
$$+2125\,w^8 + 3064\,w^9 + 4129\,w^{10} + 5234\,w^{11} + 6290\,w^{12} + 7186\,w^{13}$$
$$+7808\,w^{14} + 8081\,w^{15} + 7987\,w^{16} + 7550\,w^{17} + 6818\,w^{18}$$
$$+5866\,w^{19} + 4805\,w^{20} + 3747\,w^{21} + 2771\,w^{22}$$

$$f_8(1) = 86421 \text{ inaccessible strings of length at most 30.}$$

Thus there are

$$f_8(19) = 404715377014978469067711788207 97$$

inaccessible sequences of length at most 30.

$$e_8 = 1 + 7\,w + 28\,w^2 + 84\,w^3 + 210\,w^4 + 462\,w^5 + 924\,w^6 + 1709\,w^7$$
$$+2954\,w^8 + 4809\,w^9 + 7420\,w^{10} + 10906\,w^{11} + 15330\,w^{12}$$
$$+20664\,w^{13} + 26769\,w^{14} + 33390\,w^{15} + 40166\,w^{16} + 46655\,w^{17}$$
$$+52374\,w^{18} + 56854\,w^{19} + 59710\,w^{20} + 60691\,w^{21} + 59710\,w^{22}$$

$$e_8(1) = 501827 \text{ strings of length at most 30.}$$

$$e_8(19) = 855972319026576304079305969491905$$

sequences of length at most 30. Thus about 5% of the sequences are unaccessible.