In Case You Still Don't Know What I Did This Summer

by Mark Tiefenbruck

Let $\sigma = \sigma_1\sigma_2\cdots\sigma_n$ be a word of length n such that $\sigma_1, \ldots, \sigma_n \in \mathbb{N}$. It shall be assumed that $\sigma_l = 0$ for all l outside of this range. Then, define the variation of σ:

$$var(\sigma) = \sum |\sigma_i - \sigma_{i-1}|.$$

The focus of this discussion will be the enumeration of words based on $var(\sigma)$, with various restrictions placed on the letters of σ.

Start by defining two sequences of generating functions:

$$T_k(p,q,z) = \sum_{i,j,n\geq 0} T_{i,j,n,k} p^i q^j z^n,$$
$$U_k(p,q,z) = \sum_{i,j,n\geq 0} U_{i,j,n,k} p^i q^j z^n.$$

$T_{i,j,n,k} = \#$ of words σ of length n such that $0 \leq \sigma_1, \ldots, \sigma_n \leq k$, $var(\sigma) = i$, and $\sum \sigma_l = j$.

$U_{i,j,n,k} = \#$ of words σ of length n such that $1 \leq \sigma_1, \ldots, \sigma_n \leq k$, $var(\sigma) = i$, and $\sum \sigma_l = j$.

Now I shall present two arguments with which we can form relationships between these two series.

First, suppose you have a word σ of length $n \geq 1$ such that $0 \leq \sigma_1, \ldots, \sigma_n \leq k$. Then, define a new word σ' of length n such that $\sigma'_l = \sigma_l + 1, 1 \leq l \leq n$. It follows that $1 \leq \sigma_1, \ldots, \sigma_n \leq k+1$, $var(\sigma') = var(\sigma) + 2$ (since σ'_0 and σ'_{n+1} remain 0 and all other differences remain the same), $\sum \sigma'_l = n + \sum \sigma_l$. Now, suppose you have a word of length 0 and do the same. Clearly, nothing changes. So, after noting that every word of length n such that $1 \leq \sigma_1, \ldots, \sigma_n \leq k+1$ can be obtained this way, we conclude that

$$U_{k+1}(p,q,z) = 1 + p^2(T_k(p,q,zq) - 1), \tag{1}$$

where the 1s account for the word of length 0. Special care should be taken to note that the third argument of T_k is zq.

Instead, now, suppose you have two words σ and σ' of lengths $n \geq 0$ and $n' \geq 0$ such that $0 \leq \sigma_1, \ldots, \sigma_n \leq k$ and $1 \leq \sigma'_1, \ldots, \sigma'_{n'} \leq k$. Let σ'' be the concatenation of the two words with a 0 inserted between them. Then, $n'' = n + n' + 1$, $0 \leq \sigma''_1, \ldots, \sigma''_{n''} \leq k$, $var(\sigma'') = var(\sigma) + var(\sigma')$, and $\sum \sigma''_{l''} = \sum \sigma_l + \sum \sigma'_{l'}$. We see that every word containing at least one 0 can be obtained in this manner. Every word that does not contain a 0 is described by $U_k(p, q, z)$, so we conclude that

$$T_k(p, q, z) = T_k(p, q, z) * z * U_k(p, q, z) + U_k(p, q, z). \tag{2}$$

Equation (1) gives a relationship between U_{k+1} and T_k, and equation (2) gives a relationship between T_k and U_k. Therefore, we have a recurrence relationship for the two generating functions. Now we need some initial conditions. There are no words of length $n \geq 1$ such that $1 \leq \sigma_1, \ldots, \sigma_n \leq 0$. However, it is essential to the argument above that there is always one word of length 0. Therefore, we may conclude that

$$U_0 = 1. \tag{3}$$

This is wonderful and all, but it would be so much better to have an explicit formula for T_k, don't you think? I think so. So. Let's assume, somewhat arbitrarily but reasonably, after a little bit of work, that T_k can be expressed in the form

$$T_k(p, q, z) = \frac{A_k(p, q, z)}{B_k(p, q, z)}.$$

Then, we do some algebra, using our recurrence relationships (1) and (2).

$$\begin{aligned} U_{k+1}(p, q, z) &= \frac{(1 - p^2)B_k(p, q, zq) + p^2 A_k(p, q, zq)}{B_k(p, q, zq)} \\ T_{k+1}(p, q, z) &= \frac{(1 - p^2)B_k(p, q, zq) + p^2 A_k(p, q, zq)}{(1 - z + p^2 z)B_k(p, q, zq) - p^2 z A_k(p, q, zq)} \end{aligned}$$

Therefore,

$$A_{k+1}(p, q, z) = (1 - p^2)B_k(p, q, zq) + p^2 A_k(p, q, zq), \tag{4}$$

$$B_{k+1}(p, q, z) = B_k(p, q, zq) - z A_{k+1}(p, q, z). \tag{5}$$

Also, a quick use of equation (2) with our initial condition (3) tells us that $T_0 = \frac{1}{1-z}$. Therefore,

$$A_0 = 1, B_0 = 1 - z. \tag{6}$$

Now, define two more generating functions:

$$f(z,x) = \sum_{k=0}^{\infty} A_k x^k,$$
$$g(z,x) = \sum_{k=0}^{\infty} B_k x^k.$$

Then, using equations (4), (5), and (6),

$$f(z,x) = x(1-p^2)g(zq,x) + xp^2 f(zq,x) + 1 \tag{7}$$
$$g(z,x) = xg(zq,x) - zf(z,x) + 1. \tag{8}$$

Solving for f in (8) and substituting into (7):

$$q(xg(zq,x) - g(z,x) + 1) = zqx(1-p^2)g(zq,x) + xp^2(xg(zq^2,x) - g(zq,x) + 1) + zq$$

$$q - zq - xp^2 = qg(z,x) + (zqx(1-p^2) - qx - xp^2)g(zq,x) + x^2p^2g(zq^2,x) \tag{9}$$

Now, let

$$g(z,x) = \sum_{i=0}^{\infty} C_i(p,q,x) z^i.$$

Then, substituting into (9) and isolating the coefficients of z^i,

$$qC_i + q^i x(1-p^2)C_{i-1} - q^i(qx+xp^2)C_i + q^{2i}x^2p^2C_i = 0, i \geq 2$$

$$C_i = \frac{q^i x(p^2-1)}{q - q^i(qx+xp^2) + q^{2i}x^2p^2}C_{i-1} = \frac{q^i x(p^2-1)}{(q-xp^2q^i)(1-xq^i)}C_{i-1}, i \geq 2 \quad (10)$$

$$q - xp^2 = qC_0 - (qx+xp^2)C_0 + x^2p^2C_0$$

$$C_0 = \frac{q-xp^2}{q-(qx+xp^2)+x^2p^2} = \frac{q-xp^2}{(q-xp^2)(1-x)} = \frac{1}{1-x}. \quad (11)$$

$$-q = qC_1 + qx(1-p^2)C_0 - q(qx+xp^2)C_1 + q^2x^2p^2C_1$$

$$C_1 = \frac{x(p^2-1)C_0 - 1}{(1-xp^2)(1-qx)} \quad (12)$$

At this point, it is simple to find an explicit formula for C_i. If we first define the shifted factorial:

$$(a;q)_n = (1-a)(1-aq)...(1-aq^{n-1}), \ (1 \text{ if } n=0)$$

then it is clear that

$$C_i = \frac{q^{\binom{i}{2}}x^i(p^2-1)^i}{(xp^2;q)_i(xq;q)_i(1-x)} - \frac{q^{\binom{i}{2}}x^{i-1}(p^2-1)^{i-1}}{(xp^2;q)_i(xq;q)_i},$$

excluding the second term for $i=0$. Further, if we define the basic hypergeometric series:

$${}_r\Phi_s(a_1,...,a_r;b_1,...,b_s;q,z) = \sum_{i=0}^{\infty} \frac{(a_1;q)_i...(a_r;q)_i}{(b_1;q)_i...(b_s;q)_i(q;q)_i}(q^{\binom{i}{2}}(-1)^i)^{s-r+1}z^i,$$

then we see that

$$g(z,x) = \frac{(1-xp^2)\,{}_2\Phi_2(0,q;xp^2,xq;q,xz(1-p^2))}{x(1-x)(1-p^2)} - \frac{1}{x(1-p^2)}. \quad (13)$$

Substituting this into equation (8) easily gives an expression for $f(z,x)$ as well. Then, the numerator and denominator of T_k, respectively, are the coefficients of x^k in $f(z,x)$ and $g(z,x)$.

So this gives fairly explicit formulas for T_k and U_k. However, hypergeometric series are not so easy to manipulate. Therefore, I shall present some other interesting results.

Later on, it will be important to know formulas for $T_k(1,q,z)$ and $U_k(1,q,z)$. Substituting $p=1$ in equations (10), (11), and (12), we find that

$$g(z,x)|_{p=1} = \frac{1}{1-x} - \frac{z}{(1-x)(1-qx)} = \frac{1-qx-z}{(1-x)(1-qx)}, \tag{14}$$

$$f(z,x)|_{p=1} = \frac{1}{1-x}. \tag{15}$$

The standard methods reveal that the coefficient of x^k in equations (14) and (15), respectively, are $1 - \frac{z(1-q^{k+1})}{1-q}$ and 1. Therefore,

$$T_k(1,q,z) = \frac{1-q}{1-q-z(1-q^{k+1})}, \tag{16}$$

$$U_k(1,q,z) = \frac{1-q}{1-q-z(q-q^{k+1})}. \tag{17}$$

This is all well and good, but the purpose of this discussion is to address the topic of variations. Using the derived explicit formula for T_k, it is possible to find the exact distribution of variations among words categorized by length, sum, and bound. However, such computations are time-consuming and may contain much more information than is actually needed. So, we can do some analysis and compute the expected variation of an arbitrarily chosen word with specified length, sum, and bound. We can do this by summing all of the possible variations for words with those characteristics and dividing by the total number of words. That is,

$$E[var(\sigma)] = \frac{\text{coefficient of } q^j z^n \text{ in } \frac{\partial T_k}{\partial p}(1,q,z)}{\text{coefficient of } q^j z^n \text{ in } T_k(1,q,z)}.$$

Thus, we wish to find an explicit formula for $\frac{\partial T_k}{\partial p}(1,q,z)$. Differentiating equations (1) and (2), respectively,

$$\frac{\partial U_{k+1}}{\partial p}(1,q,z) = 2(T_k(1,q,zq)-1) + \frac{\partial T_k}{\partial p}(1,q,zq), \tag{18}$$

$$\frac{\partial T_k}{\partial p}(1,q,z) = \frac{\frac{\partial U_k}{\partial p}(1,q,z)}{(1-zU_k(1,q,z))^2}. \tag{19}$$

Thus, after substituting equations (16) and (17), we have a recurrence. Arbitrarily decide that

$$\frac{\partial T_k}{\partial p}(1,q,z) = \frac{2(A_k(q)-zqB_k(q))zq}{(1-q-z(1-q^{k+1})^2}. \tag{20}$$

Then, using equation (18) and then (19),

$$\frac{\partial U_{k+1}}{\partial p}(1,q,z) = \frac{2zq(1-q^{k+1})}{1-q-zq(1-q^{k+1})} + \frac{2(A_k-zq^2B_k)zq^2}{(1-q-zq(1-q^{k+1}))^2}$$

$$= \frac{2zq(1-q^{k+1})(1-q-zq(1-q^{k+1}))+2(A_k-zq^2B_k)zq^2}{(1-q-zq(1-q^{k+1}))^2},$$

$$\frac{\partial T_{k+1}}{\partial p}(1,q,z) = \frac{\frac{\partial U_{k+1}}{\partial p}(1,q,z)}{(\frac{1-q-z(1-q^{k+2})}{1-q-z(q-q^{k+2})})^2}$$

$$= \frac{2zq(1-q^{k+1})(1-q-zq(1-q^{k+1}))+2(A_k-zq^2B_k)zq^2}{(1-z(1-q^{k+2}))^2}.$$

Thus, we see that

$$A_k = qA_{k-1} + (1-q^k)(1-q),$$
$$B_k = q^2B_{k-1} + (1-q^k)^2.$$

Also, we know that $A_0 = B_0 = 0$, since $T_0 = \frac{1}{1-z}$. This gives us two fairly manageable recurrences. Doing the standard analysis, it turns out that

$$A_k = 1-(1+k-kq)q^k,$$
$$B_k = \frac{1-2(1+q)q^k+(1+k+2q-kq^2)q^{2k}}{1-q^2}.$$

Then, we can substitute these back into equation (20) and be generally happy.

So, I have some other results, but they're fairly easily derived from these by setting $q = 1$ or setting $z = 1$ in U_k or letting $k \to \infty$, etc. It is possible to throw in some other statistics, such as the number and sum of peaks and valleys in the words, by altering the initial condition (3) and the recurrences (1) and (2) ever so slightly, but I never got around to doing computations with these statistics because I was busy thinking about other things. We did some stuff with counting the occurrences of each number and so forth, but that more involves Dr. Rawlings's theory of adjacencies and is going to be written up in a separate article. Also, Dr. Lawrence Sze came up with a way to relate ascent variation with the variation that I have defined above, so I shall include his argument as a closing thought on this discussion. Thank you.

Given the same σ as before, define the ascent variation and descent variation:

$$avar(\sigma) = \sum_{i=1}^{n-1} \{\sigma_{i+1} - \sigma_i | \sigma_{i+1} \geq \sigma_i\} = \sum_{i=1}^{n-1} \frac{|\sigma_{i+1} - \sigma_i| + \sigma_{i+1} - \sigma_i}{2} = \frac{var(\sigma)}{2} - \sigma_1, \tag{1}$$

$$dvar(\sigma) = \sum_{i=1}^{n-1} \{\sigma_i - \sigma_{i+1} | \sigma_{i+1} \leq \sigma_i\} = \sum_{i=1}^{n-1} \frac{|\sigma_{i+1} - \sigma_i| - \sigma_{i+1} + \sigma_i}{2} = \frac{var(\sigma)}{2} - \sigma_n, \tag{2}$$

provided that we assume $\sigma_i \geq 0$ for the last equalities. Also, before I get too far, define the generating functions AT_k and AU_k to be analogous to T_k and U_k, with the exception that the exponent on p is $2avar(\sigma)$. Now, consider $T_k - T_{k-1}$ or $U_k - U_{k-1}$. Then, we have subtracted the terms for all words that do not contain at least one k. Thus, these generating functions describe the total variation for words that contain at least one k, under the previous restrictions. Now, consider one such σ. Separate it into two words at the last occurrence of k in σ, calling σ' the word before and including it and σ'' the word after and including it. We wish to be able to compute the total variation in terms of the ascent and descent variations. From above, we see that

$$var(\sigma') = 2dvar(\sigma') + 2k,$$
$$var(\sigma'') = 2avar(\sigma'') + 2k.$$

Therefore, since we have counted k two extra times,

$$var(\sigma) = var(\sigma') + var(\sigma'') - 2k = 2dvar(\sigma') + 2avar(\sigma'') + 2k.$$

Now, there are a few things to note. First, there is the fact that the selected k does not contribute to $dvar(\sigma')$ or to $avar(\sigma'')$. Thus, we can remove it from each word for the remainder of the analysis. Next, note that AT_k and AU_k would be the same if they used $2dvar(\sigma)$ for the exponent on p, since $avar(\sigma) = dvar(\sigma^*)$, where σ^* is the reverse of σ. Also note that σ' and σ'' now come from AT_k and AT_{k-1}, respectively. Finally, $n = n' + n'' + 1$ and $\sum_{l=1}^{n} \sigma_l = \sum_{l'=1}^{n'} \sigma'_{l'} + \sum_{l''=1}^{n''} \sigma''_{l''} + k$. Thus,

$$T_k - T_{k-1} = AT_k * zq^k * AT_{k-1} * p^{2k},$$
$$U_k - U_{k-1} = AU_k * zq^k * AU_{k-1} * p^{2k}.$$

Since we have formulas for T_k and U_k, we can find formulas based on these relationships for AT_k and AU_k, given the appropriate initial conditions, which I trust you to figure out.