THE EXPECTED VARIATION OF RANDOM BOUNDED INTEGER SEQUENCES OF FINITE LENGTH

RUDOLFO ANGELES, DON RAWLINGS, LAWRENCE SZE, AND MARK TIEFENBRUCK

Received 24 November 2004 and in revised form 15 June 2005

From the enumerative generating function of an abstract adjacency statistic, we deduce the mean and variance of the variation on random permutations, rearrangements, compositions, and bounded integer sequences of finite length.

1. Introduction

When the finite sequence of integers \( w = 1, 3, 2, 4, 3 \) is sketched as below,

\[
\begin{align*}
\text{w} = & \\
1 & \quad 2 & \quad 2 & \quad 4 & \quad 3
\end{align*}
\]

its most compelling aspect is its vertical variation, that is, the sum of the vertical distances between its adjacent terms. Denoted by \( \text{var} \ w \), the vertical variation of the sequence in (1.1) is \( \text{var} \ w = 2 + 1 + 0 + 2 + 1 = 6 \). Our purpose here is to compute the mean and variance of \( \text{var} \) on four classical sets of combinatorial sequences.

To formalize matters and place our problem in the context of other work, let \([m]^n\) denote the set of sequences \( w = x_1 x_2 \cdots x_n \) of length \( n \) with each \( x_i \in \{1, 2, \ldots, m\} \). For a real-valued function \( f \) on \([m]^2\), the \( f \)-adjacency number of \( w = x_1 x_2 \cdots x_n \in [m]^n \) is defined to be

\[
\text{adf} \ w = \sum_{k=1}^{n-1} f(x_k x_{k+1}).
\]
Some specializations of the $f$-adjacency number have been considered elsewhere. For instance, if $f(xy)$ is 1 when $x < y$ and 0 otherwise, then $\text{adf} w$ is known as the rise number of $w$ [1, 3, 4]. For the selection $f(xy) = |y - x|$, $\text{adf} w = \text{var} w$. In a sorting problem of computer science, Levcopoulos and Petersson [5] introduced the related notion of oscillation ($\text{var} w - n + 1$) as a measure of the presortedness of a sequence of $n$ distinct numbers. In [6], compositions were enumerated by their ascent variation, the $f$-adjacency statistic induced by $f(xy) = y - x$ if $x < y$ and 0 otherwise. For the case $f(xy) = h(|y - x|)$ where $h$ is a linear, convex, or concave increasing real-valued function, Chao and Liang [2] described the arrangements of $n$ distinct integers for which $\text{adf}$ achieves its extreme values.

Besides considering the distribution of $\text{var}$ on the set $[m]^n$, we also consider it on the set of rearrangements $R_n(i_1,i_2,\ldots,i_m)$ consisting of sequences of length $n = i_1 + i_2 + \cdots + i_m$ which contain $l$ exactly $i_l$ times, on the set of permutations $S_n = R_n(1,1,\ldots,1)$ of $\{1,2,\ldots,n\}$, and on the set of compositions of $m$ into $n$ parts $C_n(m) = \{x_1x_2\cdots x_n \in [m]^n : x_1 + x_2 + \cdots + x_n = m\}$. For $m,n \geq 2$, Table 1.1 displays the mean and variance of $\text{var}$ on these four sets. The $k$th falling factorial of $n$ is $n^k = n(n-1)\cdots(n-k+1)$, and, for $r$ a real number, $[r]$ denotes the greatest integer less than or equal to $r$. The results in Table 1.1 are new. David and Barton [3, Chapter 10] present the distributions of several statistics (some $f$-adjacency numbers, some not) primarily on permutations. We also note that Tiefenbruck [7] derived a generating function for compositions with bounded parts by a close relative of $\text{var}$. We leave open questions concerning the asymptotic behavior of $\text{var}$.

### 2. Enumerative factorial moments for $f$-adjacencies

Before working specifically with $\text{var}$, we discuss the enumerative generating function for $\text{adf}$ on sequences as developed by Fédou and Rawlings [4]. Let $[m]^*$ denote the set of sequences of 1, 2, ..., $m$ of finite length (including the empty sequence of length 0). For $w = x_1x_2\cdots x_n \in [m]^*$, we define $Z^w = z_1^{x_1}z_2^{x_2}\cdots z_m^{x_n}$. The enumerative generating function for $\text{adf}$ over $[m]^*$ is then defined to be $G(p) = \sum_{w \in [m]^*} p^{\text{adf} w} Z^w$.

By manipulating $G(p)$, we will obtain all of the information in Table 1.1 (and more). As a brief outline of our approach, note that the coefficient of $p^k z_1^{i_1}z_2^{i_2}\cdots z_m^{i_m}$ in $G(p)$ is

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Expected value of $\text{var}$</th>
<th>Variance of $\text{var}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_n$</td>
<td>$\frac{n^2 - 1}{3}$</td>
<td>$(n-2)(n+1)(4n-7)$</td>
</tr>
<tr>
<td>$[m]^n$</td>
<td>$\frac{2}{3m} \sum_{1 \leq i &lt; j \leq m} (y-x)i_jy$</td>
<td>$\frac{90}{m^2}$</td>
</tr>
<tr>
<td>$R_n(i)$</td>
<td>$\frac{2(n-1)}{(m-1)^{i-1}} \sum_{x=1}^{\lfloor m/2 \rfloor} (m-2x)^{i-1}$</td>
<td>See (3.10)</td>
</tr>
<tr>
<td>$C_n(m)$</td>
<td>$\frac{2(n-1)}{(m-1)^{i-1}} \sum_{x=1}^{\lfloor m/2 \rfloor} (m-2x)^{i-1}$</td>
<td>See (3.18)</td>
</tr>
</tbody>
</table>
just the number of rearrangements \( w \) in \( R_n(i) \) with \( \text{adf} \ w = k \). Thus, by dividing the coefficient of \( z_1^i z_2^j \cdots z_m^n \) in \( G'(1) \) by the cardinality of \( R_n(i) \), we will obtain the mean of \( \text{adf} \). So, in general, we compute the \( d \)th enumerative factorial moment \( G^{(d)}(1) = \sum_{w \in [m]^*} (\text{adf} \ w)^d Z^w \).

From the work of Fédou and Rawlings [4], it follows that

\[
G(p) = \frac{1}{D(p)}, \quad (2.1)
\]

where

\[
D(p) = 1 - \sum_{n \geq 1} \sum_{x_1 \cdots x_n \in [m]^n} Z^{x_1 \cdots x_n} \prod_{k=1}^{n-1} \left(p f(x_k, x_{k+1}) - 1\right). \quad (2.2)
\]

Examples are presented in [4, 6] for which \( D \) has a closed form. We do not know a closed form for \( D \) when \( \text{adf} = \text{var} \) (that is, when \( f(x, y) = |y - x| \)). Nevertheless, (2.1) is still useful in computing the mean and variance of \( \text{var} \).

Although the formula for taking the \( d \)-fold derivative with respect to \( p \) of a function of the form in (2.1) is known, we provide a short derivation. To avoid the quotient and chain rules, rewrite (2.1) as \( GD = 1 \). Differentiating the latter \( d \) times, \( d \geq 1 \), and dividing by \( d! \) gives

\[
\sum_{j=0}^{d} \frac{G^{(d-j)} D^{(j)}}{(d-j)! j!} = 0. \quad (2.3)
\]

To solve for \( G^{(d)} \), consider the system

\[
\begin{align*}
\frac{G^{(d)}}{d!} \frac{D^{(0)}}{0!} &+ \frac{G^{(d-1)}}{(d-1)!} \frac{D^{(1)}}{1!} + \frac{G^{(d-2)}}{(d-2)!} \frac{D^{(2)}}{2!} + \cdots + \frac{G^{(0)}}{0!} \frac{D^{(d)}}{d!} = 0, \\
\frac{G^{(d-1)}}{(d-1)!} \frac{D^{(0)}}{0!} &+ \frac{G^{(d-2)}}{(d-2)!} \frac{D^{(1)}}{1!} + \cdots + \frac{G^{(0)}}{0!} \frac{D^{(d-1)}}{(d-1)!} = 0, \\
&\vdots \\
\frac{G^{(1)}}{1!} \frac{D^{(0)}}{0!} &+ \frac{G^{(0)}}{0!} \frac{D^{(1)}}{1!} = 0,
\end{align*}
\]

(2.4)

where the top \( d \) equations arise from repeated application of (2.3). Cramer’s rule applied
Variation of random sequences

to the above system yields

\[
\frac{G^{(d)}}{d!} = \frac{(-1)^d}{D^{d+1}} \begin{vmatrix}
D^{(1)} & D^{(2)} & D^{(3)} & \cdots & D^{(d)} \\
\frac{1!}{0!} & \frac{2!}{1!} & \frac{3!}{2!} & \cdots & \frac{d!}{d!} \\
\frac{D^{(0)}}{0!} & \frac{D^{(1)}}{1!} & \frac{D^{(2)}}{2!} & \cdots & \frac{D^{(d-1)}}{(d-1)!} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \frac{D^{(0)}}{0!} & \frac{D^{(1)}}{1!}
\end{vmatrix}
\]

(2.5)

which, when expanded, implies that

\[
G^{(d)} = \sum_{v=1}^{d} \frac{(-1)^v}{D^{v+1}} \sum_{j_1+\cdots+j_v = d \atop j_k \geq 1} \left( \frac{d}{j_1 \cdots j_v} \right) D^{(j_1)} \cdots D^{(j_v)}. 
\]

(2.6)

To determine the enumerative factorial moment \(G^{(d)}(1)\), we see from (2.2) that

\[
D^{(j)}(1) = -\sum_{r=2}^{j+1} D^{(j)}_r, 
\]

(2.7)

where

\[
D^{(j)}_r = \sum_{x_1, \ldots, x_r \in [m]^r} \sum_{l_1+\cdots+l_{r-1} = j \atop l_k \geq 1} \left( \frac{j}{l_1 \cdots l_{r-1}} \right) \prod_{k=1}^{r-1} f(x_k x_{k+1})^{l_k}. 
\]

(2.8)

For instance,

\[
D_2' = \sum_{xy \in [m]^2} f(xy)z_x z_y, \quad D_2'' = \sum_{xy \in [m]^2} f(xy)^2 z_x z_y, \\
D_3' = 2 \sum_{vx y \in [m]^3} f(vx) f(xy) z_v z_x z_y. 
\]

(2.9)

Further setting \(\vec{j} = (j_1, \ldots, j_v)\), \(s(\vec{j}) = j_1 + \cdots + j_v\),

\[
\binom{d}{\vec{j}} = \binom{d}{j_1 \cdots j_v}, \quad \text{and} \quad D^{(\vec{j})}_{\vec{\mu}} = \sum_{r_1+\cdots+r_v = \mu \atop r_k \geq 2} D^{(j_1)}_{r_1} \cdots D^{(j_v)}_{r_v}, 
\]

(2.10)

it follows from (2.6) and (2.7) that

\[
G^{(d)}(1) = \sum_{v=1}^{d} \frac{1}{D^{v+1}(1)} \sum_{s(\vec{j}) = d \atop j_k \geq 1} \binom{d}{\vec{j}} \sum_{\mu=2v}^{d+v} D^{(\vec{j})}_{\vec{\mu}}. 
\]

(2.11)
As \( D(1) = 1 - (z_1 + \cdots + z_m) \), extracting the contributions made by all \( w \in [m]^n \) from both sides of (2.11) gives the \( d \)th enumerative factorial moment of \( \text{adf} \) over \([m]^n\) as

\[
\sum_{w \in [m]^n} (\text{adf } w)^d Z^w = \sum_{\nu=1}^{d} \sum_{\mu=2\nu} (d) \left( \sum_{j=1}^{d+\nu} \left( \sum_{i=1}^{m} z_i \right)^{n-\mu} \right) D^{(j)}_{\mu} \tag{2.12}
\]

valid for \( d \geq 1 \). When \( d = 1, 2 \), (2.9) and (2.12) imply that

\[
\sum_{w \in [m]^n} \text{adf } w Z^w = (n - 1) \left( \sum_{i=1}^{m} z_i \right)^{n-2} \sum_{xy \in [m]^2} f(xy) z_x z_y \tag{2.13}
\]

and that

\[
\sum_{w \in [m]^n} (\text{adf } w)^2 Z^w = (n - 1) \left( \sum_{i=1}^{m} z_i \right)^{n-2} \sum_{xy \in [m]^2} (f(xy))^2 z_x z_y \\
+ 2(n - 2) \left( \sum_{i=1}^{m} z_i \right)^{n-3} \sum_{vxy \in [m]^3} f(vx)f(xy) z_x z_y z_y \tag{2.14}
\]

\[+ (n - 2)(n - 3) \left( \sum_{i=1}^{m} z_i \right)^{n-4} \left( \sum_{xy \in [m]^2} f(xy) z_x z_y \right)^2. \]

3. Discussion of Table 1.1

The entries in Table 1.1 are consequences of (2.13) and (2.14) with \( f(xy) = |y - x| \) and with appropriate substitutions for \( Z \). For the mean and variance of \( \text{var} \) on the set of bounded sequences \([m]^n\), put \( z_i = 1 \) for \( 1 \leq i \leq m \). Noting that

\[
\sum_{xy \in [m]^2} |y - x| = \sum_{1 \leq x < y \leq m} 2(y - x) = 2 \left( m + 1 \right) \left( m \right) \left( m + 1 \right). \tag{3.1}
\]

it follows from (2.13) that the mean of \( \text{var} \) on \([m]^n\) is

\[
\frac{1}{m^n} \sum_{w \in [m]^n} \text{var } w = 2(n - 1) m^{n-2} \left( m + 1 \right) \left( m \right) \left( m + 1 \right) = (n - 1) \left( m^2 - 1 \right) \left( m \right) \left( m + 1 \right). \tag{3.2}
\]

As

\[
\sum_{xy \in [m]^2} |y - x|^2 = 4 \left( m + 1 \right) \left( m \right). \tag{3.3}
\]
Variation of random sequences

and as

$$\sum_{x, y \in [m]^n} |x - v||y - x| = \sum_{1 \leq x < y \leq m} 2(x - v)(y - x) + \sum_{1 \leq x < y \leq m} 4(v - x)(y - x) - \sum_{1 \leq x < y \leq m} 2(y - x)^2$$

$$= \frac{7m^2 - 8}{10} \left( \frac{m + 1}{3} \right),$$

(3.4)

implies that

$$\frac{1}{m^n} \sum_{w \in [m]^n} (\text{var } w)^2 = \frac{4(n - 1)}{m^2} \left( \frac{m + 1}{4} \right) + \frac{(n - 2)(7m^2 - 8)}{5m^3} \left( \frac{m + 1}{3} \right)$$

$$+ \frac{4(n - 2)(n - 3)}{m^3} \left( \frac{m + 1}{3} \right)^2.$$

(3.5)

Then, subbing the last result into

$$\frac{1}{m^n} \sum_{w \in [m]^n} (\text{var } w)^2 + \frac{(n - 1)(m^2 - 1)}{3m} - \left( \frac{(n - 1)(m^2 - 1)}{3m} \right)^2$$

(3.6)

and simplifying gives the variance of var as recorded in Table 1.1.

For $R_n(i)$, extracting the coefficient of $z_1^{i_1}z_2^{i_2}\cdots z_m^{i_m}$ from (2.13) leads to

$$\sum_{w \in R_n(i)} \text{var } w = 2(n - 1) \sum_{1 \leq x < y \leq m} (y - x) \left( i_1 \cdots i_x - 1 \cdots i_y - 1 \cdots i_m \right).$$

(3.7)

As the cardinality of $R_n(i)$ is

$$\binom{n}{i_1i_2\cdots i_m} = \binom{n}{\vec{i}},$$

(3.8)

it follows that the mean of var on $R_n(i)$ is

$$\left( \binom{n}{\vec{i}} \right)^{-1} \sum_{w \in R_n(\vec{i})} \text{var } w = \frac{2}{n} \sum_{1 \leq x < y \leq m} (y - x)i_xi_y.$$

(3.9)

Let $\vec{i}_r = (i_1, \ldots, i_r - 1, \ldots, i_n)$. For example, $(3,2,1,4)_{\langle 2 \rangle 3} = (3,1,-1,4)$. The variance on $R_n(\vec{i})$ is then

$$\left( \binom{n}{\vec{i}} \right)^{-1} \sum_{w \in R_n(\vec{i})} \text{var } w^2 + \frac{2}{n} \sum_{1 \leq x < y \leq m} (y - x)i_xi_y - \left( \frac{2}{n} \sum_{1 \leq x < y \leq m} (y - x)i_xi_y \right)^2,$$

(3.10)
where, upon extraction of the coefficient of \( z_1^{i_1} z_2^{i_2} \ldots z_m^{i_m} \) from (2.14), we have

\[
\sum_{w \in R_i(i)} (\text{var } w)^2 = (n-1) \sum_{1 \leq x, y \leq m} |y-x|^2 \left( \frac{n-2}{i(x\setminus y)} \right) \\
+ 2(n-2) \sum_{1 \leq v, x, y \leq m} |x-v||y-x| \left( \frac{n-3}{i(v\setminus x)} \right) \\
+ (n-2)(n-3) \sum_{1 \leq u, v, x, y \leq m} |v-u||y-x| \left( \frac{n-4}{i(u\setminus v\setminus x\setminus y)} \right).
\]  

(3.11)

The permutation entries in Table 1.1 follow from (3.9) and (3.10). Selecting \( m = n \) and \( i_k = 1 \) for \( 1 \leq k \leq n \) in (3.9) reveals the mean of var on \( S_n \) as

\[
\frac{1}{n!} \sum_{w \in S_n} \text{var } w = \frac{2}{n} \sum_{1 \leq x < y \leq n} (y-x) = \frac{2}{n} \binom{n+1}{3} = \frac{n^2 - 1}{3}.
\]

(3.12)

From (3.11), with \( m = n \) and \( i_k = 1 \) for \( 1 \leq k \leq n \),

\[
\sum_{w \in S_n} (\text{var } w)^2 = (n-1)! \sum_{1 \leq x, y \leq n} |y-x|^2 \\
+ 2(n-2)! \sum_{1 \leq v, x, y \leq n} |x-v||y-x| \\
+ (n-2)! \sum_{1 \leq u, v, x, y \leq n} |v-u||y-x| \\
= \left( \frac{4}{15} \right) (n-2)! (10n^2 + 14n - 27) \binom{n+1}{4}.
\]

(3.13)

So the variance of var on \( S_n \) is

\[
\frac{1}{n!} \sum_{w \in S_n} \text{var } w^2 + \frac{n^2 - 1}{3} \left( \frac{n^2 - 1}{3} \right)^2 = \frac{(n-2)(n+1)(4n-7)}{90}.
\]

(3.14)

For \( w = x_1 \cdots x_n \in [m]^n \), let \( \|w\| = x_1 + \cdots + x_n \). For the composition results in Table 1.1, set \( z_k = q^k \) for \( 1 \leq k \leq m \). Then (2.13) implies that

\[
\sum_{w \in [m]^n} \text{var } w q^{\|w\|} = (n-1)q^{n-2} \left( \frac{1 - q^m}{1-q} \right) \sum_{1 \leq x, y \leq m} |y-x|q^{x+y}
\]

(3.15)
and (2.14) leads to
\[
\sum_{w \in [m]^n} (\text{var } w)^2 q^{\|w\|} = (n-1) q^{n-2} \left( \frac{1 - q^m}{1 - q} \right)^{n-2} \sum_{1 \leq x, y \leq m} |y-x|^2 q^{x+y} \\
+ 2(n-2) q^{n-3} \left( \frac{1 - q^m}{1 - q} \right)^{n-3} \sum_{1 \leq v, x, y \leq m} |x-v||y-x| q^{v+x+y} \\
+ (n-2)(n-3) q^{n-4} \left( \frac{1 - q^m}{1 - q} \right)^{n-4} \sum_{1 \leq u, v, x, y \leq m} |v-u||y-x| q^{u+v+x+y}.
\]

Extracting the coefficient of \(q^m\) from (3.16) to obtain
\[
\sum_{w \in C_n(m)} \text{var } w = 2(n-1) \sum_{1 \leq x < y \leq m} (y-x) \left( \frac{m-1-x-y}{n-3} \right)
\]
\[
= 2(n-1) \sum_{1 \leq x \leq \lfloor m/2 \rfloor} \left( \frac{m-2x}{n-1} \right)
\]

and then dividing by the cardinality \(\binom{m-1}{n-1}\) of \(C_n(m)\) gives the mean of var as stated in Table 1.1. The variance is
\[
\left( \frac{m-1}{n-1} \right)^{-1} \sum_{w \in C_n(m)} \text{var } w^2 + \frac{2(n-1)}{(m-1)^{n-1}} \sum_{1 \leq x \leq \lfloor m/2 \rfloor} (m-2x)^{m-1}
\]
\[
- \left( \frac{2(n-1)}{(m-1)^{n-1}} \sum_{1 \leq x \leq \lfloor m/2 \rfloor} (m-2x)^{m-1} \right)^2,
\]

where, pulling the coefficient of \(q^m\) from (3.16), we have
\[
\sum_{w \in C_n(m)} (\text{var } w)^2 = (n-1) \sum_{1 \leq x, y \leq m} |y-x|^2 \left( \frac{m-1-x-y}{n-3} \right)
\]
\[
+ 2(n-2) \sum_{1 \leq v, x, y \leq m} |x-v||y-x| \left( \frac{m-1-v-x-y}{n-4} \right)
\]
\[
+ (n-2)(n-3) \sum_{1 \leq u, v, x, y \leq m} |v-u||y-x| \left( \frac{m-1-u-v-x-y}{n-5} \right).
\]

The sums in (3.19) are marginally simplified. For instance,
\[
\sum_{1 \leq x, y \leq m} |y-x|^2 \left( \frac{m-1-x-y}{n-3} \right) = 4 \sum_{1 \leq x \leq \lfloor m/2 \rfloor} \left( \frac{m-2x}{n} \right).
\]
As a part of the second sum on the right-hand side of (3.19), we note that

\[
\sum_{1 \leq v < x < y \leq m} (x - v)(y - x) \binom{m - 1 - v - x - y}{n - 4}
= \sum_{2 \leq x \leq \lfloor (m+1)/2 \rfloor} \left( \binom{m - 3x + 1}{n} - \binom{m - 2x + 1}{n} + x \binom{m - 2x}{n - 1} \right).
\]

(3.21)

The four-fold sums arising in the last sum in (3.19) reduce to double sums.

Acknowledgment

This work is based on work supported by the National Science Foundation under Grant no. 0097392.

References


Rudolfo Angeles: Department of Mathematics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA

Current address: Department of Statistics, Stanford University, Palo Alto, CA 94305-4065, USA

E-mail address: rangeles@stanford.edu

Don Rawlings: Department of Mathematics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA

E-mail address: drawling@calpoly.edu

Lawrence Sze: Department of Mathematics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA

E-mail address: lsze@calpoly.edu

Mark Tiefenbruck: Department of Mathematics, College of Science and Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407, USA

Current address: 8989 Jasmine Lane Street, Cottage Grove, MN 55016-3436, USA

E-mail address: mdash@alumni.northwestern.edu