Joint distribution of rises, falls, and number of runs in random sequences

Yong Kong

School of Public Health

Yale University

300 Cedar Street, New Haven, CT 06520, USA

email: yong.kong@yale.edu

Abstract

By using the matrix formulation of the two-step approach to the distributions of runs, a recursive relation and an explicit expression are derived for the generating function of the joint distribution of rises and falls for multivariate random sequences in terms of generating functions of individual letters, from which the generating functions of the joint distribution of rises, falls, and number of runs are obtained. An explicit formula for the joint distribution of rises and falls with arbitrary specification is also obtained.

1 Introduction

For a sequence $\sigma = (a_1, a_2, \ldots, a_n)$ where $a_i \in \{1, 2, \ldots, k\}$, a pair of consecutive elements a_i and a_{i+1} is called a *rise* when $a_i < a_{i+1}$, a *fall* when $a_i > a_{i+1}$, and a *level* when $a_i = a_{i+1}$. If r, s, and l denote the number of rises, falls, and levels, they satisfy the obvious relation: $r + s + l = n - 1$.

A *run* in σ is defined as a stretch of consecutive identical integers separated by other integers or the end of the sequence (Balakrishnan and Koutras, 2002; Mood, 1940). If integer i appears n_i times in σ, we call σ of *specification* $\mathbf{n} = [n_1, n_2, \ldots, n_k]$, with $\sum_{i=1}^{k} n_i = n$. If b denotes the number of runs, it's easy to see that $b + l = n$, and $r + s + 1 = b$. If integer i has b_i runs, then $\sum_{i=1}^{k} b_i = b$. We are interested in the number $A(\mathbf{n}; r, s, \mathbf{b})$ of sequences with exactly r rises, s falls, and numbers of runs $\mathbf{b} = [b_1, \ldots, b_k]$ for a given specification $\mathbf{n}$. In Table 1 the 10 permutations of specification $\mathbf{n} = [2, 3]$ are listed.

When $\mathbf{n} = [1, 1, \ldots, 1]$ and only rises are considered, $A([1, 1, , \ldots, 1]; r)$ is the well-known Eulerian number $A(k, r)$ (Graham *et al.*, 1994). Simon Newcomb's problem is a generalization of the Eulerian number with arbitrary specification (Dillon and Roselle, 1969). The explicit formula for this problem is given by (Dillon and Roselle, 1969; MacMahon, 1915)

$$A(\mathbf{n}; r) = \sum_{j=0}^{r} (-1)^j \binom{n+1}{j} \prod_{i=1}^{k} \binom{n_i + r - j}{n_i}.$$

The joint distribution of rise and falls with arbitrary specification was studied by different researchers (Carlitz, 1972; Fu and Lou, 2000; Jackson, 1977; Reilly, 1979). Note that these researchers, except in Reilly (1979), included an additional rise at the beginning and an additional fall at the end of each sequence. For this problem, as Carlitz pointed out, no simple explicit formulas have been found (Carlitz, 1972, 1975). In Theorem 4.1 of the present paper we give an explicit formula for $A(\mathbf{n}; r, s)$, the joint distribution of rise and falls with arbitrary specification $\mathbf{n}$.

In this paper we use the matrix formulation of the two-step approach to distributions of runs (Kong, 2006, 2015, 2016, 2017a,b) to obtain the generating function $\mathcal{G}_k$ for the whole system in terms of the generating function g_i of individual integers ($i = 0, \ldots, k$). The advantage of expressing $\mathcal{G}_k$ in terms of g_i is that, by using appropriate g_i's, various joint distributions can be obtained easily. The previously obtained results are straightforward specializations, and new results of joint distribution of rises, falls, and number of runs are obtained similarly. In Theorem 2.1 we first obtain a recursive relation for $\mathcal{G}_k$, and in Theorem 3.1 the explicit formula of $\mathcal{G}_k$ is obtained in terms of g_i. The joint distributions of rises, falls, and number of runs are obtained in Corollary 3.3, Corollary 3.4, and Corollary 3.5 . In Theorem 4.1 an explicit formula is derived for the joint distribution of rises and falls.

In the following, let

$$\mathcal{G}_k = \sum_{\mathbf{n},r,s,\cdots} A(\mathbf{n}; r, s, \cdots) w^r u^s \cdots \mathbf{x}^{\mathbf{n}}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_k]$ and $\mathbf{x}^{\mathbf{n}} = x_1^{n_1} \cdots x_k^{n_k}$. We use the notation $[x^n]f(x)$ to denote the coefficient of x^n in the series expansion of $f(x)$.

2 Recursive equation for $\mathcal{G}_k$

Theorem 2.1 (Recursive equation for $\mathcal{G}_k$)**.** *The generating function $\mathcal{G}_k$ satisfies the recursive equation*

$$\mathcal{G}_k = \mathcal{G}_{k-1} + \frac{g_k(1 + w\mathcal{G}_{k-1})(1 + u\mathcal{G}_{k-1})}{1 - wug_k\mathcal{G}_{k-1}} \tag{1}$$

with $\mathcal{G}_1 = g_1$.

Proof. Based on the matrix formulation of the two-step approach to the distributions of runs (Kong, 2006, 2017a), $\mathcal{G}_k$ is given by

$$\mathcal{G}_k = \mathbf{e}_k \mathbf{M}_k^{-1} \mathbf{y}_k, \tag{2}$$

Table 1: The 10 permutations of multiset $\{1,1,2,2,2\}$ for $k=2$, $n_1=2$, and $n_2=3$.

	r	s	l	b	b_1	b_2	$w^r u^s v^l t_1^{b_1} t_2^{b_2}$
$1\ \overline{1\ 2}\ 2\ 2$	1	0	3	2	1	1	$wv^3t_1t_2$
$\overline{1\ \underline{2}}\underline{\ \overline{1}}\overline{\ 2}\ 2$	2	1	1	4	2	2	$w^2uvt_1^2t_2^2$
$\overline{1\ 2}\ \underline{2\ \overline{1}}\overline{\ 2}$	2	1	1	4	2	2	$w^2uvt_1^2t_2^2$
$\overline{1\ 2}\ 2\ \underline{2\ 1}$	1	1	2	3	2	1	$wuv^2t_1^2t_2$
$\underline{2\ 1}\ \overline{1\ 2}\ 2$	1	1	2	3	1	2	$wuv^2t_1t_2^2$
$\underline{2\ \overline{1}}\overline{\ \underline{2}}\underline{\ \overline{1}}\overline{\ 2}$	2	2	0	5	2	3	$w^2u^2t_1^2t_2^3$
$\underline{2\ \overline{1}}\overline{\ 2}\ \underline{2\ 1}$	1	2	1	4	2	2	$wu^2vt_1^2t_2^2$
$2\ \underline{2\ 1}\ \overline{1\ 2}$	1	1	2	3	1	2	$wuv^2t_1t_2^2$
$2\ \underline{2\ \overline{1}}\overline{\ \underline{2}}\underline{\ 1}$	1	2	1	4	2	2	$wu^2vt_1^2t_2^2$
$2\ 2\ \underline{2\ 1}\ 1$	0	1	3	2	1	1	$uv^3t_1t_2$

where

$$\mathbf{e}_k = [\underbrace{1, 1, \cdots, 1}_{k}],$$

$$\mathbf{y}_k = [g_1, g_2, \cdots, g_k]^T,$$

and

$$\mathbf{M}_k = \begin{bmatrix} 1 & -g_1 w & -g_1 w & \cdots & -g_1 w \\ -g_2 u & 1 & -g_2 w & \cdots & -g_2 w \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -g_{k-1} u & \cdots & -g_{k-1} u & 1 & -g_{k-1} w \\ -g_k u & \cdots & -g_k u & -g_k u & 1 \end{bmatrix}. \tag{3}$$

Here the indeterminates w and u are used to track the rises and falls, respectively, and g_i's are the generating function for the ith integer. Eq. (3) can be written in the following block form:

$$\mathbf{M}_k = \begin{bmatrix} \mathbf{M}_{k-1} & \mathbf{b} \\ \mathbf{c} & 1 \end{bmatrix},$$

where

$$\mathbf{b} = [-g_1 w, -g_2 w, \cdots, -g_{k-1} w]^T = -w\mathbf{y}_{k-1},$$

and

$$\mathbf{c} = [-g_k u, -g_k u, \cdots, -g_k u] = -u g_k \mathbf{e}_{k-1}.$$

By applying the following matrix identity to Eq. (2),

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c} & d \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \frac{1}{h}\mathbf{A}^{-1}\mathbf{b}\mathbf{c}\mathbf{A}^{-1} & -\frac{1}{h}\mathbf{A}^{-1}\mathbf{b} \\ -\frac{1}{h}\mathbf{c}\mathbf{A}^{-1} & \frac{1}{h} \end{bmatrix}$$

where

$$h = d - \mathbf{c}\mathbf{A}^{-1}\mathbf{b},$$

the recursive equation Eq. (1) is obtained:

$$\begin{aligned} \mathcal{G}_k = \mathbf{e}_k\mathbf{M}_k^{-1}\mathbf{y}_k &= [\mathbf{e}_{k-1}, 1] \begin{bmatrix} \mathbf{A}^{-1} + \frac{1}{h}\mathbf{A}^{-1}\mathbf{b}\mathbf{c}\mathbf{A}^{-1} & -\frac{1}{h}\mathbf{A}^{-1}\mathbf{b} \\ -\frac{1}{h}\mathbf{c}\mathbf{A}^{-1} & \frac{1}{h} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{k-1} \\ g_k \end{bmatrix} \\ &= \mathcal{G}_{k-1} + \frac{g_k(1 + w\mathcal{G}_{k-1})(1 + u\mathcal{G}_{k-1})}{1 - wug_k\mathcal{G}_{k-1}}. \end{aligned}$$

□

3 Explicit formula for $\mathcal{G}_k$ and generating functions of joint distributions of rises, falls, and runs

Theorem 3.1 (Explicit formula for $\mathcal{G}_k$)**.** *The explicit formula for $\mathcal{G}_k$ in terms of the generating functions g_i is given by*

$$\mathcal{G}_k = \frac{\prod_{i=1}^k (1 + g_i u) - \prod_{i=1}^k (1 + g_i w)}{u \prod_{i=1}^k (1 + g_i w) - w \prod_{i=1}^k (1 + g_i u)} = \frac{P_k - Q_k}{uQ_k - wP_k}, \tag{4}$$

where $P_k = \prod_{i=1}^k (1+g_i u)$ *and* $Q_k = \prod_{i=1}^k (1+g_i w)$.

Proof. By induction using Eq. (1) in Theorem 2.1. □

From Theorem 3.1 generating functions of various joint distributions can be obtained by using appropriate g_i's. Some examples are:

Corollary 3.2. *The explicit formula for* $\mathcal{G}_k$ *of joint distribution of rises and falls in terms of* x_i *is given by*

$$\begin{aligned} \mathcal{G}_k &= \sum_{\mathbf{n},r,s} A(\mathbf{n};r,s) w^r u^s \mathbf{x}^{\mathbf{n}} \\ &= \frac{\prod_{i=1}^k [1-x_i(1-u)] - \prod_{i=1}^k [1-x_i(1-w)]}{u\prod_{i=1}^k [1-x_i(1-w)] - w\prod_{i=1}^k [1-x_i(1-u)]}. \end{aligned} \tag{5}$$

Proof. Let

$$g_i = \sum_{j=1} x_i^j = \frac{x_i}{1-x_i}.$$

□

Eq. (5) was obtained by Carlitz (1972), Jackson (1977) and Reilly (1979) using different methods.

Corollary 3.3. *The explicit formula for* $\mathcal{G}_k$ *for the joint distribution of rises, falls, and the total number of runs is given by*

$$\begin{aligned} \mathcal{G}_k &= \sum_{\mathbf{n},r,s,b} A(\mathbf{n};r,s,b) w^r u^s t^b \mathbf{x}^{\mathbf{n}} \\ &= \frac{\prod_{i=1}^k [1-x_i(1-ut)] - \prod_{i=1}^k [1-x_i(1-wt)]}{u\prod_{i=1}^k [1-x_i(1-wt)] - w\prod_{i=1}^k [1-x_i(1-ut)]}. \end{aligned} \tag{6}$$

Proof. Let

$$g_i = t\sum_{j=1} x_i^j = \frac{x_i t}{1-x_i}.$$

□

As an example, from Eq. (6) we get for $k=2$, $n_1=2$, and $n_2=3$:

$$[x_1^2 x_2^3]\mathcal{G}_k = \sum_{r,s,b} A(\mathbf{n}; r,s,b) w^r u^s t^b = ut^2 + wt^2 + 3wut^3 + 2wu^2t^4 + 2w^2ut^4 + w^2u^2t^5,$$

which is enumerated in Table 1.

Corollary 3.4. *The explicit formula for $\mathcal{G}_k$ for the joint distribution of rises, falls, levels, and the total number of runs is given by*

$$\begin{aligned}\mathcal{G}_k &= \sum_{\mathbf{n},r,s,l,b} A(\mathbf{n}; r,s,l,b) w^r u^s v^l t^b \mathbf{x}^{\mathbf{n}} \\ &= \frac{\prod_{i=1}^k [1 - x_i(v-ut)] - \prod_{i=1}^k [1 - x_i(v-wt)]}{u\prod_{i=1}^k [1 - x_i(v-wt)] - w\prod_{i=1}^k [1 - x_i(v-ut)]}.\end{aligned} \tag{7}$$

Proof. Let

$$g_i = t\sum_{j=1} x_i^j v^{j-1} = \frac{x_i t}{1-vx_i}.$$

□

Corollary 3.5. *The explicit formula for $\mathcal{G}_k$ for the joint distribution of rises, falls, levels, and the numbers of runs for each integer is given by*

$$\begin{aligned}\mathcal{G}_k &= \sum_{\mathbf{n},r,s,l,\mathbf{b}} A(\mathbf{n}; r,s,l,\mathbf{b}) w^r u^s v^l \mathbf{t}^{\mathbf{b}} \mathbf{x}^{\mathbf{n}} \\ &= \frac{\prod_{i=1}^k [1 - x_i(v-ut_i)] - \prod_{i=1}^k [1 - x_i(v-wt_i)]}{u\prod_{i=1}^k [1 - x_i(v-wt_i)] - w\prod_{i=1}^k [1 - x_i(v-ut_i)]}.\end{aligned} \tag{8}$$

Proof. Let

$$g_i = t_i \sum_{j=1} x_i^j v^{j-1} = \frac{x_i t_i}{1 - v x_i}.$$

$\square$

For $k = 2$, $n_1 = 2$, and $n_2 = 3$, Eq. (8) gives

$$[x_1^2 x_2^3]\mathcal{G}_k = \sum_{r,s,\mathbf{b}} A(\mathbf{n}; r, s, \mathbf{b}) w^r u^s t_1^{b_1} t_2^{b_2}$$

$$= uv^3 t_1 t_2 + wv^3 t_1 t_2 + 2wuv^2 t_1 t_2^2 + wuv^2 t_1^2 t_2 + 2wu^2 v t_1^2 t_2^2 + 2w^2 uv t_1^2 t_2^2 + w^2 u^2 t_1^2 t_2^3,$$

as shown in Table 1.

4 Explicit formula for $A(\mathbf{n}; r, s)$, the number of rises and falls for a given specification

From Eq. (5) of Corollary 3.2 we obtain the explicit formula for $A(\mathbf{n}; r, s)$.

Theorem 4.1 (Explicit formula for $A(\mathbf{n}; r, s)$)**.** *The explicit formula for $A(\mathbf{n}; r, s)$ is given by*

$$A(\mathbf{n}; r, s) = {\sum_{d_i=0}^{n_i}}' {\sum_{t_i=0}^{n_i}}' \left[\prod_{i=1}^{k} f(n_i, r - t + 1, d_i, t_i) - \prod_{i=1}^{k} f(n_i, r - t, d_i, t_i) \right]$$

where the sum $\sum'$ *is taken for* $d+t=r+s+1$ *and* $t \le r$ *with* $d=\sum_i d_i$ *and* $t=\sum_i t_i$*. The function* f *is defined as*

$$f(n,m,d,t) = \begin{cases} \binom{n-1}{t-1}\binom{m+t-1}{t}(-1)^t, & d=0, \\ \frac{m}{d}\binom{n-1}{d+t-1}\binom{m+t-1}{d+t-1}\binom{d+t-1}{t}(-1)^t, & d \neq 0. \end{cases} \tag{9}$$

Proof. Let $P_k = \prod_{i=1}^{k}[1-x_i(1-u)]$ and $Q_k = \prod_{i=1}^{k}[1-x_i(1-w)]$.

$$\begin{aligned} \mathcal{G}_k &= \frac{P_k - Q_k}{uQ_k - wP_k} = \frac{P_k - Q_k}{uQ_k\left[1-\frac{wP_k}{uQ_k}\right]} = \frac{P_k - Q_k}{uQ_k}\sum_i \left[1-\frac{wP_k}{uQ_k}\right]^i \\ &= \sum_i \frac{w^i}{u^{i+1}}\left[\frac{P_k^{i+1}}{Q_k^{i+1}} - \frac{P_k^i}{Q_k^i}\right] = \sum_i \frac{w^i}{u^{i+1}}\left[\prod_{a=1}^{k}\frac{[1-x_a(1-u)]^{i+1}}{[1-x_a(1-w)]^{i+1}} - \prod_{a=1}^{k}\frac{[1-x_a(1-u)]^{i}}{[1-x_a(1-w)]^{i}}\right]. \end{aligned}$$

Each term in the last expression can be further expanded by binomial theorem, and after changing variables, we get

$$\begin{aligned} A(\mathbf{n};r,s) &= \left[w^r u^s \prod_{i=1}^{k} x_i^{n_i}\right]\mathcal{G}_k \\ &= \sum_{j_i=0}^{n_i}{\sum_{d_i=0}^{j_i}}'{\sum_{t_i=0}^{n_i-j_i}}'\left[\prod_{i=1}^{k} h(n_i,j_i,r-t+1,d_i,t_i) - \prod_{i=1}^{k} h(n_i,j_i,r-t,d_i,t_i)\right], \end{aligned}$$

where

$$h(n,j,m,d,t) = \binom{j}{d}\binom{n-j}{t}\binom{m}{j}\binom{n-j+m-1}{n-j}(-1)^{j+d+t}.$$

By using Gosper-Zeilberger method (Petkovšek *et al.*, 1996), we can simplify the above expression by getting rid of the sums of j_i. Let

$$f(n, m, d, t) = \sum_{j=0}^{n} h(n, j, m, l, t)$$

be the sum of $h(n, j, m, d, t)$ with respect to j. The Gosper-Zeilberger method constructs $G(n, j)$ as

$$G(n, j) = h(n, j, m, d, t)\frac{(d - j)(n + m - j)}{n - t + 1 - j},$$

which satisfies

$$(n - d - t + 1)h(n + 1, j, m, l, t) - nh(n, j, m, l, t) = G(n, j + 1) - G(n, j).$$

Summing with respect to j leads to the following linear recurrence equation satisfied by $f(n, m, d, t)$:

$$(n - d - t + 1)f(n + 1, m, d, t) - nf(n, m, d, t) = 0.$$

By solving the recurrence we obtain the identity in Eq. (9). $\square$

References

Balakrishnan, N. and Koutras, M. V., 2002. *Runs and Scans with Applications*. John Wiley & Sons, New York, USA.

Carlitz, L., 1972. Enumeration of sequences by rises and falls: a refinement of the Simon Newcomb problem. *Duke Math. J.* 39, 267–280.

Carlitz, L., 1975. Permutations, sequences and special functions. *SIAM Review* 17, 298–322.

Dillon, J. F. and Roselle, D. P., 1969. Simon Newcomb's problem. *SIAM J. Appl. Math.* 17, 1086–1093.

Fu, J. C. and Lou, W. W., 2000. Joint distribution of rises and falls. *Annals of the Institute of Statistical Mathematics* 52, 415–425.

Graham, R. L., Knuth, D. E., and Patashnik, O., 1994. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition.

Jackson, D. M., 1977. The unification of certain enumeration problems for sequences,. *Journal of Combinatorial Theory, Series A* 22, 92–96.

Kong, Y., 2006. Distribution of runs and longest runs: A new generating function approach. *Journal of the American Statistical Association* 101, 1253–1263.

Kong, Y., 2015. Distributions of runs revisited. *Communications in Statistics - Theory and Methods* 44, 4663–4678.

Kong, Y., 2016. Number of appearances of events in random sequences: A new approach to non-overlapping runs. *Communications in Statistics - Theory and Methods* 45, 6765–6772.

Kong, Y., 2017a. The mth longest runs of multivariate random sequences. *Annals of the Institute of Statistical Mathematics* 69, 497 – 512.

Kong, Y., 2017b. Number of appearances of events in random sequences: a new generating function approach to *Type II* and *Type III* runs. *Annals of the Institute of Statistical Mathematics* 69, 489–495.

MacMahon, P., 1915. *Combinatory Analysis.* Cambridge University press.

Mood, A. M., 1940. The distribution theory of runs. *Annals of Mathematical Statistics* 11, 367–392.

Petkovsěk, M., Wilf, H. S., and Zeilberger, D., 1996. $A = B$. A K Peters Ltd, Wellesley, MA, USA.

Reilly, J. W., 1979. A combinatorial solution of two related problems in sequence enumeration. *Journal of Combinatorial Theory, Series A* 26, 304–307.