THE COVARIANCE MATRIX OF RUNS UP AND DOWN

BY H. LEVENE AND J. WOLFOWITZ
Columbia University

1. Introduction. Let $a_1, \cdots, a_n$ be n unequal numbers and let the sequence $S = (h_1, h_2, \cdots, h_n)$ be any permutation of $a_1, \cdots, a_n$. S is to be considered a chance variable, and each of the $n!$ permutations of $a_1, \cdots, a_n$ is assigned the same probability. Consider the sequence R whose i^{th} element is the sign (+ or −) of $h_{i+1} - h_i$, $(i = 1, 2, \cdots, n - 1)$. A sequence of p consecutive + signs not immediately preceded or followed by a + sign is called a run up of length p; a sequence of p consecutive − signs not immediately preceded or followed by a − sign is called a run down of length p. The term "run" will denote both runs up and runs down. The usage of the term "length" varies; most quality control literature attributes the length $p + 1$ to the runs which we say are of length p.

As an example of our usage, the sequence

$$S = 2\ 8\ 13\ 1\ 3\ 4\ 7$$

gives the sequence

$$R = + + - + + +,$$

which has a run up of length 2, followed by a run down of length 1, followed by a run up of length 3.

Runs up and down are widely used in quality control and have been applied to economic time series. The purpose of this paper is to obtain their variances and covariances and to correct some erroneous notions prevalent in the literature about their application.

2. Notation. If the sign (+ or −) of $(h_{i+1} - h_i)$ is the initial sign of a run defined as above, we call h_i the initial turning point (i. t. p.) of the run. Then h_1 is always an i. t. p., and we adopt the convention that h_n is never an i. t. p. We define new stochastic variables as follows:

$$x_i = \begin{cases} 1 & \text{if } h_i \text{ is an i. t. p.,} \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

$$x_{pi} = \begin{cases} 1 & \text{if } h_i \text{ is the i. t. p. of a run of length } p, \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

$$w_{pi} = \begin{cases} 1 & \text{if } h_i \text{ is the i. t. p. of a run of length } p \text{ or more,} \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

for $i = 1, 2, \cdots, n$. Also

$$r = \text{the number of runs in } R, \tag{2.4}$$

$$r_p = \text{the number of runs of length } p \text{ in } R, \tag{2.5}$$

$$r'_p = \text{the number of runs of length } p \text{ or more in } R. \tag{2.6}$$

Evidently $r = \sum_{i=1}^{n} x_i$, $r_p = \sum_{i=1}^{n} x_{pi}$, and $r'_p = \sum_{i=1}^{n} w_{pi}$.

If X and Y are stochastic variables, let $E(X)$ denote the mean of X, $\sigma(XY)$ denote the covariance of X and Y, and $\sigma^2(X)$ denote the variance of X, if they exist. By the distribution function of X we shall mean a function $f(x)$ such that $P\{X < x\} \equiv f(x)$, where the symbol $P\{\ \ \}$ denotes the probability of the relation in the brackets.

3. Preliminary formulas. Let Y' be a stochastic variable with any continuous distribution function $f(y)$. Let $Y = (y_1, y_2 \cdots, y_n)$ be a sequence of n independent observations on Y'. Since $P\{y_i = y_j\} = 0$, $(i \neq j; i, j = 1, 2, \cdots, n)$, the distribution of runs up and down in Y is evidently the same as that in S. Now choose $f(y)$ to be

$$\begin{aligned} f(y) &= 0, \qquad (y \leq 0),\\ f(y) &= y, \qquad (0 \leq y \leq 1),\\ f(y) &= 1, \qquad (y \geq 1). \end{aligned}$$

Then

$$P\{y_{i-1} < y_i > y_{i+1}\} = \int_0^1 \left[\int_{y_{i+1}}^1 \left(\int_0^{y_i} dy_{i-1}\right) dy_i\right] dy_{i+1} = \frac{1}{3}.$$

By symmetry

$$E(x_i) = P\{y_{i-1} < y_i > y_{i+1}\} + P\{y_{i-1} > y_i < y_{i+1}\} = \tfrac{2}{3}, \qquad (i = 2, 3, \cdots, n-1).$$

Also $E(x_1) = 1$, and $E(x_n) = 0$.

It will be necessary hereafter to evaluate expressions of the types

$$U = \int_0^{y_{p+1}} \cdots \int_0^{y_2} \frac{(y_1)^k}{k!}\, dy_1 \cdots dy_p = \frac{(y_{p+1})^{k+p}}{(k+p)!}, \tag{3.1}$$

and

$$V = \int_{y_{p+1}}^1 \cdots \int_{y_2}^1 \frac{(y_1)^k}{k!}\, dy_1 \cdots dy_p\,. \tag{3.2}$$

From the fact that

$$\int_{y_{p+1}}^1 \cdots \int_{y_2}^1 dy_1 \cdots dy_p = \int_0^{v_{p+1}} \cdots \int_0^{v_2} dv_1 \cdots dv_p = \frac{(v_{p+1})^p}{p!}$$

where $v_j = (1 - y_j)$, $(j = 1, \cdots, p+1)$, it can easily be shown that

$$\begin{aligned} V &= \sum_{s=1}^{p} (-1)^{s+1} \frac{(v_{p+1})^{p-s}}{(k+s)!(p-s)!} + (-1)^p \frac{(y_{p+1})^{k+p}}{(k+p)!}\\ &= \sum_{r=0}^{k} (-1)^{k+r} \frac{(v_{p+1})^{p+k-r}}{(p+k-r)!\,r!}. \end{aligned} \tag{3.3}$$

We shall also need $\int_0^1 V\, dy_{p+1}$ and $\int_0^1 \int_0^{y_{p+2}} V\, dy_{p+1}\, dy_{p+2}$. Now

$$\int_0^1 V\, dy_{p+1} = \sum_{r=0}^{k} (-1)^{k+r} \frac{1}{(p+k-r+1)!\,r!}. \tag{3.4}$$

Making use of the relation,

$$\sum_{r=0}^{t} (-1)^r \frac{1}{(n-r)!\,r!} = (-1)^t \frac{1}{n(n-t-1)!\,t!}, \qquad (t < n),$$

we have

$$\int_0^1 V\, dy_{p+1} = \frac{1}{(p+k+1)p!\,k!}. \tag{3.5}$$

Similarly

$$\int_0^1 \int_0^{y_{p+2}} V\, dy_{p+1}\, dy_{p+2} = \frac{1}{(p+k+1)p!\,k!} - \frac{1}{(p+k+2)(p+1)!\,k!}. \tag{3.6}$$

4. Covariances of runs up and down. We first compute $E(r_p)$ and $E(r'_p)$. We define the symbol

$$P\{-, +^p, -\} = P\{y_{i-1} > y_i < y_{i+1} < \cdots < y_{i+p} > y_{i+p+1}\}.$$

The value of the right member is independent of i whenever it is defined (*i.e.* $i - 1 \geq 1, i + p + 1 \leq n$). Now

$$E(x_{pi}) = P\{-, +^p, -\} + P\{+, -^p, +\} = 2P\{-, +^p, -\}$$

$$= 2 \int_0^1 \int_{y_{i+p+1}}^1 \int_0^{y_{i+p}} \cdots \int_0^{y_{i+1}} \int_{y_i}^1 dy_{i-1} \cdots dy_{i+p+1} = 2\frac{p^2+3p+1}{(p+3)!},$$

$$(i = 2, 3, \cdots, n-p-1).$$

$$E(x_{p1}) = 2P\{+^p, -\} \quad \text{and} \quad E(x_{p,n-p}) = 2P\{-, +^p\}.$$

By symmetry $E(x_{p1}) = E(x_{p,n-p})$, the common value being $2\dfrac{p+1}{(p+2)!}$. Also $E(x_{pi}) = 0, (i > n-p)$.
Thus

$$E(r_p) = E\left(\sum_{i=1}^{n} x_{pi}\right) = 2E(x_{p1}) + (n-p-2)E(x_{pi})$$

$$= 2n\frac{p^2+3p+1}{(p+3)!} - 2\frac{p^3+3p^2-p-4}{(p+3)!}, \quad (p \leq n-2). \tag{4.1}$$

Besson [1], Kermack and McKendrick [5], and Wallis and Moore [6] gave the exact value, although Besson proved it only for special cases. R. A. Fisher [3] gave $\lim_{n\to\infty} \dfrac{E(r_p)}{E(r)}$.

It is clear that $E(w_{pi}) = E(x_{p,n-p})$, $(i = 2, \cdots, n-p)$, while $E(w_{p1}) = 2P\{+^p\} = 2/(p-1)!$. We then have

$$E(r'_p) = 2n\frac{p+1}{(p+2)!} - 2\frac{p^2+p-1}{(p+2)!}, \qquad (p \leq n-1). \tag{4.2}$$

Setting $p = 1$ we have

$$E(r_1') = E(r) = \tfrac{1}{3}(2n - 1). \tag{4.3}$$

Formula (4.3) was given by Bienaymé [2].

We now obtain $\sigma(r_p r_q)$. Let $(x_{pi} - E(x_{pi})) = z_{pi}$. Then

$$\begin{aligned} \sigma(r_p r_q) &= E\left\{\left[\sum_{i=1}^{n} z_{pi}\right]\left[\sum_{j=1}^{n} z_{qj}\right]\right\} \\ &= \sum_i E(z_{pi} z_{qi}) + \sum_{i<j}\sum E(z_{qi} z_{pj}) + \sum_{i<j}\sum E(z_{pi} z_{qj}). \end{aligned} \tag{4.4}$$

For $j \geq i + q + 3$, x_{qi} and x_{pj} are independent and hence $E(z_{qi}z_{pj}) = 0$. Omitting zero terms from (4.4) we have

$$\begin{aligned} \sigma(r_p r_q) = \{&\sum_i E(x_{pi} x_{qi}) + \sum_{i<j<i+q+3}\sum E(x_{qi} x_{pj}) + \sum_{i<j<i+p+3}\sum E(x_{pi} x_{qj}) \\ &- [\sum_i E(x_{pi})E(x_{qi}) + \sum_{i<j<i+q+3}\sum E(x_{qi})E(x_{pj}) \\ &+ \sum_{i<j<i+p+3}\sum E(x_{pi})E(x_{qj})]\}. \end{aligned} \tag{4.5}$$

Since $x_{pi}x_{qi} = \delta_{pq}(x_{pi})^2 = \delta_{pq}x_{pi}$, we have for the first term of the right member of (4.5)

$$\sum_{i=1}^{n} E(x_{pi} x_{qi}) = \delta_{pq} E(r_p), \tag{4.6}$$

where the Kronecker delta $\delta_{pq} = \begin{cases} 1, \text{ if } p = q, \\ 0, \text{ otherwise.} \end{cases}$

Since $x_{qi}x_{pj} = 0$ for $i < j < i + q$, the second term in the right member of (4.5) is

$$\begin{aligned} &\sum_{i=1}^{n-p-q} E(x_{qi} x_{p,i+q}) + \sum_{i=1}^{n-p-q-1} E(x_{qi} x_{p,i+q+1}) + \sum_{i=1}^{n-p-q-2} E(x_{qi} x_{p,i+q+2}) \\ &\quad = \{(n - p - q - 2)E(x_{qi} x_{p,i+q}) + (n - p - q - 3)E(x_{qi} x_{p,i+q+1}) \\ &\quad + (n - p - q - 4)E(x_{qi} x_{p,i+q+2}) \\ &\quad + E(x_{q1} x_{p,q+1}) + E(x_{q1} x_{p,q+2}) + E(x_{q1} x_{p,q+3}) \\ &\quad + E(x_{q,n-q-p} x_{p,n-p}) + E(x_{q,n-q-p-1} x_{p,n-p}) + E(x_{q,n-q-p-2} x_{p,n-p})\}. \end{aligned} \tag{4.7}$$

Now $E(x_{qi}x_{p,i+q}) = 2P\{-, +^q, -^p, +\} = 2A$, where

$$A = \int_0^1 \int_0^{y_{i+p+q+1}} \int_{y_{i+p+q}}^1 \cdots \int_{y_{i+q+1}}^1 \left[\int_0^{y_{i+q}} \cdots \int_0^{y_{i+1}} \int_{y_i}^1 dy_{i-1} \cdots dy_{i+q-1}\right] \cdot dy_{i+q} \cdots dy_{i+p+q+1}.$$

The expression within the square brackets is easily evaluated, and applying (3.6) to the result, we have

$$A = \frac{1}{(p+q+1)p!q!} - \frac{1}{(p+q+2)(p+1)!q!} - \frac{1}{(p+q+2)p!(q+1)!} + \frac{1}{(p+q+3)(p+1)!(q+1)!}.$$

Similarly, $E(x_{qi}x_{p,i+q+1}) = 2P\{-, +^q, -, +^p, -\}$, and $E(x_{qi}x_{p,i+q+2}) = 2P\{-, +^q, -, -, +^p, -\} + 2P\{-, +^q, -, +, -^p, +\}$. The other terms in the right member of (4.7) are obtained in like manner. The right member of (4.7) is symmetric in p and q; hence the second and third terms of the right member of (4.5) are equal.

We now consider the remaining terms in the right member of (4.5) for $p > q$; the result obtained also holds for $p \le q$. We write them as

$$\begin{aligned} -\Bigg\{ & \sum_{i=p+3}^{n-q} E(x_{qi})E(x_{p,i-(p+2)}) + \sum_{i=p+2}^{n-q} E(x_{qi})E(x_{p,i-(p+1)}) \\ & + \cdots + \sum_{i=p-q+1}^{n-q} E(x_{qi})E(x_{p,i-(p-q)}) + \sum_{i=p-q}^{n-q-1} E(x_{qi})E(x_{p,i-(p-q-1)}) \\ & + \cdots + \sum_{i=1}^{n-p} E(x_{qi})E(x_{pi}) + \sum_{i=1}^{n-p-1} E(x_{qi})E(x_{p,i+1}) \\ & + \cdots + \sum_{i=1}^{n-p-(q+2)} E(x_{qi})E(x_{p,i+(q+2)})\Bigg\}. \end{aligned} \tag{4.8}$$

The $(p + q + 5)$ sums in (4.8) comprise in all $\left\{(n-p)(p+q+5) - 2\sum_{k=1}^{q+2} k\right\}$ terms. Remembering that $E(x_{p,n-p}) = E(x_{p1})$, (4.8) becomes

$$\begin{aligned} -\{&[n(p+q+5) - (p^2+pq+q^2+7p+7q+16)]E(x_{qi})E(x_{pj}) \\ &+ [2p+4]E(x_{qi})E(x_{p1}) + [2q+4]E(x_{q1})E(x_{pj}) + 2E(x_{q1})E(x_{p1})\}. \end{aligned} \tag{4.9}$$

Adding the right member of (4.6), twice the right member of (4.7), and (4.9), we have

$$\begin{aligned} \sigma(r_p r_q) = {} & \\ 2n\Bigg\{ & -2\,\frac{p^2(p+q+6)(q^2+3q+1) + p(3q^3+20q^2+40q+19) + (q^3+9q^2+29q+26)}{(q+3)!(p+3)!} \\ & + 2\,\frac{-p+q+1}{(p+q+3)(q+2)!(p+1)!} - 2\,\frac{1}{(p+q+5)(q+3)!(p+1)!} \\ & - 2\,\frac{(p+q)^3 + 9(p+q)^2 + 23(p+q) + 14}{(p+q+5)!} \\ & + 2\,\frac{1}{(p+q+1)q!p!} + \delta_{pq}\,\frac{p^2+3p+1}{(p+3)!}\Bigg\} \end{aligned}$$

$$
(4.10)\qquad +2\left\{\frac{2}{(q+3)!(p+3)!}\left[\begin{array}{l} p^4(q^2+3q+1)+p^3(q^3+9q^2+19q+6) \\ \quad +p^2(q^4+9q^3+28q^2+35q+11) \\ \quad +p(3q^4+20q^3+40q^2+29q+10) \\ \quad +(q^4+9q^3+27q^2+32q+10) \end{array}\right]\right.
$$

$$
+2\,\frac{(p+q+2)(p-q-1)}{(p+q+3)(q+2)!(p+1)!}+2\,\frac{p+q+4}{(p+q+5)(q+3)!(p+1)!}
$$

$$
+2\,\frac{(p+q)^4+10(p+q)^3+29(p+q)^2+16(p+q)-19}{(p+q+5)!}
$$

$$
\left.-2\,\frac{p+q}{(p+q+1)q!p!}-\delta_{pq}\,\frac{p^3+3p^2-p-4}{(p+3)!}\right\},
$$

where δ_{pq} is the Kronecker delta. Formula (4.10) is valid for $p + q \leq n - 4$. It is symmetric in p and q. Setting $p = q$ we obtain

$$
\sigma^2(r_p) = 2n\left\{-2\,\frac{2p^5+15p^4+41p^3+55p^2+48p+26}{(p+3)!(p+3)!}\right.
$$

$$
+2\,\frac{2p^2+9p+12}{(2p+3)(2p+5)(p+3)!(p+1)!}-4\,\frac{4p^3+18p^2+23p+7}{(2p+5)!}
$$

$$
\left.+2\,\frac{1}{(2p+1)p!p!}+\frac{p^2+3p+1}{(p+3)!}\right\}
$$

$$
(4.11)\qquad +2\left\{2\,\frac{3p^6+24p^5+69p^4+90p^3+67p^2+42p+10}{(p+3)!(p+3)!}\right.
$$

$$
-4\,\frac{2p^3+11p^2+19p+9}{(2p+3)(2p+5)(p+3)!(p+1)!}
$$

$$
+2\,\frac{16p^4+80p^3+116p^2+32p-19}{(2p+5)!}
$$

$$
\left.-4\,\frac{p}{(2p+1)p!p!}-\frac{p^3+3p^2-p-4}{(p+3)!}\right\}
$$

We next evaluate $\sigma(r'_p r'_q)$. Since w_{qi} and w_{pj} are independent for $j \geq i + q + 2$, we have, corresponding to (4.5),

$$
(4.12)\qquad \begin{aligned} \sigma(r'_p r'_q) = {} & \sum_i E(w_{pi}w_{qi}) + \sum_{i<j<i+q+2}\sum E(w_{qi}w_{pj}) + \sum_{i<j<i+p+2}\sum E(w_{pi}w_{qj}) \\ & - [\sum_i E(w_{pi})E(w_{qi}) + \sum_{i<j<i+q+2}\sum E(w_{qi})E(w_{pj}) \\ & + \sum_{i<j<i+p+2}\sum E(w_{pi})E(w_{qj})]. \end{aligned}
$$

Let $G = \text{Max}(p, q)$. Then $w_{pi}w_{qi} \equiv w_{Gi}$ and we have for the first term of the right member of (4.12)

$$\sum_{i=1}^{n} E(w_{pi}w_{qi}) = E(r'_G). \tag{4.13}$$

The second term in the right member of (4.12) may be written

$$\begin{aligned}(n - p - q - 1)E(w_{qi}w_{p,i+q}) + (n - p - q - 2)E(w_{qi}w_{p,i+q+1}) \\ + E(w_{q1}w_{p,q+1}) + E(w_{q1}w_{p,q+2}).\end{aligned} \tag{4.14}$$

Now $E(w_{qi}w_{p,i+q}) = 2P\{-, +^q, -^p\}$, $E(w_{qi}w_{p,i+q+1}) = 2P\{-, +^q, -, +^p\} + 2P\{-, +^{q+1}, -^p\}$, and the other terms are obtained similarly. The third term in the right member of (4.12) will be equal to (4.14) with p and q interchanged.

The remaining terms in the right member of (4.12) become

$$\begin{aligned}-\{[n(p + q + 3) - (p^2 + pq + q^2 + 4p + 4q + 5)]E(w_{qi})E(w_{pj}) \\ + [p + 1]E(w_{qi})E(w_{p1}) + [q + 1]E(w_{q1})E(w_{pj}) + E(w_{q1})E(w_{p1})\}.\end{aligned} \tag{4.15}$$

We can now write the formula for $\sigma(r'_p r'_q)$, valid for $p + q \leq n - 2$,

$$\begin{aligned}\sigma(r'_p r'_q) = 2n \Bigg\{ &- \frac{p^2(2q + 2) + p(2q^2 + 8q + 5) + (2q + 1)(q + 2)}{(q + 2)!(p + 2)!} \\ &+ \frac{2}{(p + q + 1)q!p!} - \frac{(q + 1)(q + 2) + (p + 1)(p + 2)}{(p + q + 3)(q + 2)!(p + 2)!} \\ &- 2\,\frac{p + q + 2}{(p + q + 3)!} + \frac{(G + 1)}{(G + 2)!}\Bigg\} \\ + 2\Bigg\{&\frac{p^3(2q + 2) + p^2(2q^2 + 8q + 5) + p(2q^3 + 8q^2 + 6q - 2) + (2q^3 + 5q^2 - 2q - 6)}{(q + 2)!(p + 2)!} \\ &- 2\,\frac{p + q}{(p + q + 1)q!p!} + \frac{(p + q + 2)[(p + 1)(p + 2) + (q + 1)(q + 2)]}{(p + q + 3)(q + 2)!(p + 2)!} \\ &+ 2\,\frac{(p + q)^2 + 3(p + q) + 1}{(p + q + 3)!} - \frac{G^2 + G - 1}{(G + 2)!}\Bigg\},\end{aligned} \tag{4.16}$$

where $G = \text{Max}(p, q)$. Setting $p = q$ we obtain

$$\begin{aligned}\sigma^2(r'_p) = 2n \Bigg\{ &- 2\,\frac{(p + 1)(2p^2 + 4p + 1)}{(p + 2)!(p + 2)!} + 2\,\frac{1}{(2p + 1)p!p!} \\ &- 2\,\frac{1}{(2p + 3)(p + 2)!p!} - 4\,\frac{p + 1}{(2p + 3)!} + \frac{p + 1}{(p+2)!}\Bigg\} \\ &+ 2\Bigg\{2\,\frac{(p + 1)^2(3p^2 + 4p - 3)}{(p + 2)!(p + 2)!} - 4\,\frac{p}{(2p + 1)p!p!} \\ &+ 4\,\frac{p + 1}{(2p + 3)(p + 2)!p!} + 2\,\frac{4p^2 + 6p + 1}{(2p + 3)!} - \frac{p^2 + p - 1}{(p + 2)!}\Bigg\}.\end{aligned} \tag{4.17}$$

Setting $p = 1$, we have

$$\sigma^2(r_1') = \sigma^2(r) = \frac{16n - 29}{90}. \tag{4.18}$$

The value of $\sigma^2(r)$ was given by Bienaymé [2].

Finally, we evaluate

$$\begin{aligned}\sigma(r_p r_q') = \sum_i E(x_{pi} w_{qi}) &+ \sum_{i<j<i+q+2}\sum E(w_{qi} x_{pj}) + \sum_{i<j<i+p+3}\sum E(x_{pi} w_{qj}) \\ &- [\sum_i E(x_{pi})E(w_{qi}) + \sum_{i<j<i+q+2}\sum E(w_{qi})E(x_{pj}) \\ &+ \sum_{i<j<i+p+3}\sum E(x_{pi})E(w_{qj})].\end{aligned} \tag{4.19}$$

Let the symbol $\eta_{ij} = \begin{Bmatrix} 1, i \geq j \\ 0, i < j \end{Bmatrix}$. Then $x_{pi} w_{qi} \equiv \eta_{pq} x_{pi}$, and

$$\sum_{i=1}^{n} E(x_{pi} w_{qi}) = \eta_{pq}[E(r_p)]. \tag{4.20}$$

The remaining terms of (4.19) introduce no new difficulties, and for $p + q \leq n - 3$ we obtain

$$\begin{aligned}\sigma(r_p r_q') = 2n \Bigg\{ &- \frac{p^3(2q+2) + p^2(2q^2+13q+12) + p(6q^2+22q+23) + (2q^2+6q+15)}{(p+3)!(q+2)!} \\ &+ \frac{2}{(p+q+1)p!q!} + \frac{p-q}{(p+q+2)(p+1)!(q+1)!} \\ &+ \frac{(p-q+1)(q+2)}{(p+q+3)(p+2)!(q+2)!} + \frac{(p+2)(p+3)+(q+1)(q+2)}{(p+q+4)(p+3)!(q+2)!} \\ &- 2\frac{(p+q)^2+5(p+q)+5}{(p+q+4)!} + \eta_{pq}\frac{p^2+3p+1}{(p+3)!}\Bigg\} \\ &+ 2\Bigg\{\frac{1}{(p+3)!(q+2)!}\begin{bmatrix} p^4(2q+2) + p^3(2q^2+13q+12) \\ + p^2(2q^3+13q^2+26q+24) \\ + p(6q^3+22q^2+19q+27) \\ + (2q^3+6q^2+10q+25) \end{bmatrix} \\ &- 2\frac{p+q}{(p+q+1)p!q!} - \frac{p^2+2p-q^2+2}{(p+q+2)(p+1)!(q+1)!} \\ &- \frac{(p+2)[(p+2)(q+3)-1] - q(q+1)(q+2)}{(p+q+3)(p+2)!(q+2)!} \\ &- \frac{(p+q+3)[(p+2)(p+3)+(q+1)(q+2)]}{(p+q+4)(p+3)!(q+2)!} \\ &+ 2\frac{(p+q)^3+6(p+q)^2+8(p+q)-1}{(p+q+4)!} - \eta_{pq}\frac{p^3+3p^2-p-4}{(p+3)!}\Bigg\},\end{aligned} \tag{4.21}$$

where η_{pq} is defined as in (4.20).

5. The use of runs up and down in tests of significance. Certain misconceptions about the application of runs up and down have appeared in the literature, and it is the purpose of this section to clarify them.

Since $E(r_p)$, $\sigma^2(r_p)$, $E(r)$ and $\sigma^2(r)$ are all of the order n, it follows that r_p/r converges stochastically to

$$\lambda_p = \lim_{n\to\infty} \frac{E(r_p)}{E(r)}.$$

Let

$$\lambda'_p = \lim_{n\to\infty} \frac{E(r'_p)}{E(r)}.$$

From (4.1) and (4.2) we have

$$\lambda_1 = \frac{5}{8} = .6250$$

$$\lambda_2 = \frac{11}{40} = .2750$$

$$\lambda_3 = \frac{19}{240} = .07917$$

$$\lambda_4 = \frac{29}{1680} = .01726$$

$$\lambda'_5 = \frac{1}{280} = .00357$$

Let

$$\lambda_{pn} = \frac{E(r_p)}{E(r)}.$$

Some writers say that λ_{pn} or λ_p is "the probability of a run of length p." If the stochastic process consists in obtaining a sequence from among the $n!$ sequences S, each of which has the probability $(n!)^{-1}$, then the phrase "the probability of *a* run of length p" has no meaning. One can speak of the probability of at least one run of length p (i.e., that $r_p > 0$), of the probability of no run of length p ($r_p = 0$), of the probability that the first or fifth run (if there are five runs) in the sequence S be of length p, etc. It is possible to give *different* stochastic processes in which "the probability of a run of length p" will have meaning and be λ_{pn}, or λ_p. Consider, for example, the totality of all the *runs* in the $n!$ sequences S. There are $n!E(r)$ of them, and among these there are $n!E(r_p)$ runs of length p. Now let the stochastic process consist in drawing a *run* from the totality of all these runs, each of which is to have the same probability, which is therefore $[n!E(r)]^{-1}$. Then the probability of drawing a run of length p is λ_{pn}. It is difficult to see how this stochastic process can have relevance to most of the problems of quality control and economic time series where runs up and down are now used.

Some writers on quality control and economic time series recommend that statistical control or randomness be tested by use of $d_1, \cdots, d_{p-1}, d'_p$, where

$$d_i = r_i - E(r_i), \qquad (i = 1, 2, \cdots, (p-1)),$$
$$d'_p = r'_p - E(r'_p).$$

The availability of the covariance matrix M of $d_1, \cdots, d_{p-1}, d'_p$, which we have obtained in this paper, will assist in the construction of such tests. Also of help will be a result recently announced by one of us [7], the early publication of which is expected. This result states that in the limit with n the joint probability density function of $d_1, \cdots, d_{p-1}, d'_p$, is $Ke^{-\frac{1}{2}Q}$, where K is a constant and Q is a quadratic form in $d_1, \cdots, d_{p-1}, d'_p$ whose matrix is the inverse of the matrix M. It follows immediately that Q has in the limit the χ^2 distribution with p degrees of freedom.

We wish now to make a few remarks about the tests of significance, based on runs up and down, which are used by some contemporary writers. A description of their method can perhaps be best given by an example. With $n = 100$ and $p = 5$, say, suppose the observed values are:

Observed Values

$$\begin{aligned} r_1 &= 30 \\ r_2 &= 10 \\ r_3 &= 4 \\ r_4 &= 3 \\ r'_5 &= 3 \\ \hline \text{Total, } r &= 50 \end{aligned}$$

These writers then say that the expected values are:

Expected Values according to some writers

$$\begin{aligned} E(r_1) &= r\lambda_1 = 50\ (.6250) = 31.25 \\ E(r_2) &= r\lambda_2 = 50\ (.2750) = 13.75 \\ E(r_3) &= r\lambda_3 = 50\ (.07917) = 3.96 \\ E(r_4) &= r\lambda_4 = 50\ (.01726) = 0.86 \\ E(r'_5) &= r\lambda'_5 = 50\ (.00357) = 0.18 \\ \hline & \qquad\qquad\qquad\qquad 50.00 \end{aligned}$$

The correct expected values are given by (4.1) and (4.2) and are:

Correct Expected Values

$$\begin{aligned} E(r_1) &= 41.75 \\ E(r_2) &= 18.10 \\ E(r_3) &= 5.15 \\ E(r_4) &= 1.11 \\ E(r'_5) &= 0.22 \\ \hline & \quad 66.33 \end{aligned}$$

It should be noted that:

(a) A consequence of the incorrect method of obtaining "expected values" is that, since

$$E(r) = E(r_1) + E(r_2) + E(r_3) + E(r_4) + E(r_5'),$$

it implies that the *expected* number of runs of all lengths is equal to the *observed* number! This is obviously erroneous. In fact it follows from (4.18) and the results announced in [7] that $r - E(r)$ is in the probability sense of order $\sqrt{n}$.

(b) By using the incorrect expected values for comparison with the observed values one loses the valuable information furnished by $r - E(r)$. If this is large (in terms of its standard deviation) it is plausible to question whether statistical control or randomness exists.

6. Summary. Let $S = (h_1, \cdots, h_n)$ be a random permutation of the n unequal numbers $a_1, \cdots, a_n$, and let R be the sequence of signs ($+$ or $-$) of the differences $h_{i+1} - h_i$ ($i = 1, \cdots, n - 1$). It is assumed that each of the $n!$ sequences S is equally probable. A sequence of p successive $+$ ($-$) signs not immediately preceded or followed by a $+$ ($-$) sign is called a run up (down) of length p. Let r_p and r_p' be the number of runs up and down in R of lengths p and p or more respectively. In this paper the exact values of $\sigma(r_p r_q)$, (see formula (4.10)); $\sigma^2(r_p)$, (formula (4.11)); $\sigma(r_p' r_q')$, (formula (4.16)); $\sigma^2(r_p')$, (formula (4.17)); and $\sigma(r_p r_q')$, (formula (4.21)) are derived. A few numerical values are:

$$\sigma^2(r_1) = \frac{305n - 347}{720}, \qquad \sigma^2(r_2) = \frac{51106n - 73859}{453600},$$

$$\sigma^2(r_1') = \frac{16n - 29}{90}, \qquad \sigma^2(r_2') = \frac{57n - 43}{720}, \qquad \sigma^2(r_3') = \frac{21496n - 51269}{453600},$$

$$\sigma(r_1 r_2) = -\frac{19n + 11}{210}, \qquad \sigma(r_1' r_2') = -\frac{5n - 3}{60}, \qquad \sigma(r_1' r_3') =$$

$$-\frac{41n - 99}{630}, \qquad \sigma(r_2 r_1') = -\frac{23n + 135}{1260}, \qquad \sigma(r_1 r_2') = -\frac{117n - 79}{720},$$

$$\sigma(r_1 r_3') = -\frac{363n - 817}{5040}, \qquad \text{and} \quad \sigma(r_2 r_3') = -\frac{18346n - 49019}{453600}.$$

The values of $E(r_p)$, (formula (4.1)); and $E(r_p')$, (formula (4.2)) are also given. Certain misconceptions about the applications of runs up and down are discussed.

REFERENCES

[1] L. BESSON, (transl. and abr. by E. W. Woolard), "On the comparison of meteorological data with chance results," *Monthly Weather Review*, (U. S. Weather Bureau), Vol. 48 (1920), pp. 89–94.

[2] J. BIENAYMÉ, "Sur une question de probabilités," *Bull. Soc. Math. France*, Vol. 2 (1874), pp. 153–154.

[3] R. A. FISHER, "On the random sequence," *Quarterly Jour. Roy. Meteorological Soc.* Vol. 52 (1926), p. 250.

[4] W. O. KERMACK AND A. G. MCKENDRICK, "Tests for randomness in a series of observations," *Proc. Roy. Soc. Edinburgh*, Vol. 57 (1937), pp. 228–240.

[5] W. O. KERMACK AND A. G. MCKENDRICK, "Some distributions associated with a randomly arranged set of numbers," *Proc. Roy. Soc. Edinburgh*, Vol. 57 (1937), pp. 332–376.

[6] W. A. WALLIS AND G. H. MOORE, *A significance test for time series*, Technical Paper 1, Nat. Bureau of Econ. Res., New York, 1941.

[7] J. WOLFOWITZ, "Asymptotic distributions of ascending and descending runs," [abstract] *Bull. Amer. Math. Soc.*, Vol. 49 (1943), pp. 539–540.