A TEST AGAINST TREND IN RANDOM SEQUENCES

Peter Lindqvist

Department of Mathematical Sciences
Norwegian University of Science and Technology
NO–7491 Trondheim, Norway

ABSTRACT:[1] We study a modification of Kendall's τ, replacing his permutations of n different numbers by sequences of length n. Thus repetition is allowed. In particular, binary sequences are studied.

1 Introduction

The basic tool in Kendall's τ-test is the "score" S. Suppose that ℓ digits $0, 1, ..., \ell - 1$ satisfying the transitive relations $0 < 1 < ... < \ell - 1$ are given and consider all the ℓ^n possible sequences

$$x_1, x_2, ..., x_n \tag{1}$$

of length n that can be formed by the aid of these digits. Let $S^+(x_1, ..., x_n)$ denote the number of true inequalities

$$x_i > x_j \quad \text{with} \quad i > j$$

in the sequence (1). Analogously, $S^-(x_1, ..., x_n)$ counts the number of all valid inequalities $x_i < x_j$ with $i > j$. Define

$$S(x_1, ..., x_n) = S^+(x_1, ..., x_n) - S^-(x_1, ..., x_n). \tag{2}$$

[1]AMS classification 62G60

For example, $S = 5 - 11 = -6$ for the sequence $0\,1\,1\,2\,0\,2\,1$. Originally[2], Kendall considered the distribution of S among the $\ell!$ permutations of the digits $0, 1, ..., \ell - 1$ (then $n = \ell$), and so far as I know[3] the generalizations of Kendall's τ-test rely upon the distribution of S among all

$$\frac{n!}{(2!)^{p_2}(3!)^{p_3}\cdots(r!)^{p_r}} \qquad (p_1 + 2p_2 + ... + rp_r = n)$$

possible sequences (1) consisting of p_1 digits occurring only once, p_2 pairs, p_3 triplets, and so on. Here the numbers of ties, i.e., $p_1, p_2, ..., p_r$, are regarded as fixed. See [S]. For permutations S^+ has been thoroughly investigated in [M].

We shall study a different situation, arising for example in connexion with the testing of *sequences of random digits*. In this setting the number of ties cannot be regarded as fixed *a priori*. Thus we are led to study *the distribution of S among all ℓ^n sequences (1).* As we shall learn, *this distribution approaches normality, as $n \to \infty$.*

Having in mind applications for a certain kind of sampling, we have considered the binary case $\ell = 2$ also when the probability for a 0 is p and the probability for a 1 is q, where $p + q = 1$. See Section 5.

Finally, we mention that the distribution for S in our setting of the problem is, in certain respects even simpler than the version considered by Kendall [K1], Sillitto [S], and Silverstone [S2].

2 The Basic Results

The mean value for S taken over all sequences (1) is zero by symmetry:

$$\mu(n) = 0. \tag{3}$$

When all the sequences are equiprobable, the variance is

$$\sigma^2(n) = \frac{\ell - 1}{\ell}\,\frac{n(n-1)}{2} + \frac{\ell^2 - 1}{\ell^2}\,\frac{n(n-1)(n-2)}{9} \tag{4}$$

[2]The coefficient τ was considered by Greiner (1909) and Esscher (1924). The coefficient was rediscovered by Kendall (1938).

[3]However, see [L1].

and the fourth central moment is

$$
\begin{aligned}
\mu_4(n) \;=\; & \Big(\frac{\ell^2-1}{\ell^2}\Big)^2 \frac{100n^4+328n^3-127n^2-997n-372}{2700}\, n(n-1) \\
& + \frac{\ell^2-1}{\ell^4}\,\frac{252n^3+507n^2-3623n+3652}{900}\, n(n-1) \\
& - \frac{\ell^2-1}{\ell^3}\,\frac{2n^3+3n^2-5n-15}{6}\, n(n-1) \\
& + \frac{\ell-1}{\ell^3}\,\frac{n^2+11n-25}{2}\, n(n-1).
\end{aligned} \tag{5}
$$

By symmetry

$$\mu_3(n) \;=\; 0,\; \mu_5(n) \;=\; 0,\; \mu_7(n) \;=\; 0, \ldots.$$

It is interesting to observe that for n fixed the moments approach those given by Kendall in [K1], as $\ell \to \infty$. Formula (4) is derived in Section 6, but the corresponding calculations for (5) are, to say the least, a laborious task and so $\mu_4(n)$ is given without proof, when $\ell \geq 3$.

As n grows, the distribution for S tends towards normality in the sense that the frequency between the values S_1 and S_2 tends to

$$\frac{1}{\sigma(n)\sqrt{2\pi}} \int_{S_1}^{S_2} e^{-x^2/2\sigma^2(n)}\, dx,$$

where the standard deviation is $\sigma(n) = \sqrt{\mu_2(n)}$. This follows from the Second Limit Theorem, since

$$\lim_{n\to\infty} \frac{\mu_{2k}(n)}{(\sigma(n))^{2k}} \;=\; 1\cdot 2\cdot 5\cdots(2k-1) \tag{6}$$

and the odd moments are zero. It is easy to prove (6) for small values of ℓ, but the probability function for S becomes soon too complicated, as ℓ grows. Therefore we shall prove (6) only for the binary case $\ell = 2$, see Section 4.

In the binary case the probability generating function for S is

$$f(x) \;=\; \begin{cases} \frac{1}{2^{2\nu}} \prod\limits_{k=1}^{k=\nu} \big(x^{2k-1}+2+x^{1-2k}\big), & n = 2\nu \\ \frac{2}{2^{2\nu+1}} \prod\limits_{k=1}^{k=\nu} \big(x^{k}+x^{-k}\big)^2, & n = 2\nu+1 \end{cases} \tag{7}$$

and so the characteristic function $\phi(\theta) \;=\; f(e^{i\theta})$ reduces to the simple expression (10). Our proof for (6), when $\ell = 2$ is based on $\phi(\theta)$. Furthermore, the distribution for S can be rapidly calculated via a suitable interpretation of (7).

In the general binary case, when the probability that $x_i = 0$ is p and that $x_i = 1$ is q in (1), $i = 1, 2, ..., n$, $p + q = 1$, the corresponding probability function is given by (18) and the characteristic function by (19). Now again $\mu(n) = 0$, and

$$\sigma^2(n) = \frac{n(n^2-1)}{3}\,pq \tag{8}$$

and the fourth moment is

$$\mu_4(n) = \frac{n(n^2-1)(5n^3-6n^2-5n+14)}{15}\,p^2q^2 + \frac{n(n^2-1)(3n^2-7)}{15}\,pq(p-q)^2. \tag{9}$$

All odd moments are zero and the asymptotic normality (6) holds even for $p \neq q$.

3 The Probability Function

Consider the binary case $\ell = 2$ with equiprobable sequences (1). Direct calculation of S for $n = 3$ yields

0 0 0	**0**	0 1 0	**0**	1 0 0	**2**	1 1 0	**2**
0 0 1	**−2**	0 1 1	**−2**	1 0 1	**0**	1 1 1	**0**

and so we can write

−2	**−1**	**0**	**+1**	**+2**
2	0	4	0	2

and in this manner the distribution for S can be tabulated. The results are displayed below:

2
1 2 1
2 0 4 0 2
1 2 1 2 4 2 1 2 1
2 0 4 0 6 0 8 0 6 0 4 0 2
1 2 1 2 4 4 5 4 5 8 5 4 5 4 4 2 1 2 1

The fundamental observation is that the table can be constructed via the following kind of figurates. For even n we have

$$
\begin{array}{c}
\overline{1\;2\;1} \\
1\;2\;1 \\
1\;2\;1 \qquad 1\;2\;1 \\
\overline{1\;2\;1\;2\;4\;2\;1\;2\;1} \\
1\;2\;1\;2\;4\;2\;1\;2\;1 \\
1\;2\;1\;2\;4\;2\;1\;2\;1 \quad 1\;2\;1\;2\;4\;2\;1\;2\;1 \\
\overline{1\;2\;1\;2\;4\;4\;5\;4\;5\;8\;5\;4\;5\;4\;4\;2\;1\;2\;1}
\end{array}
$$

and so on. (The overlined sequences display $n = 2$, 4 and 6.) For odd n the table looks like

$$
\begin{array}{c}
\overline{0\;2\;0} \\
0\;2\;0 \\
0\;2\;0 \quad 0\;2\;0 \\
\overline{0\;2\;0\;4\;0\;2\;0} \\
0\;2\;0\;\;4\;0\;2\;0 \\
0\;2\;0\;4\;0\;2\;0 \quad 0\;2\;0\;4\;0\;2\;0 \\
\overline{0\;2\;0\;4\;0\;6\;0\;8\;0\;6\;0\;4\;0\;2\;0} \\
0\;2\;0\;4\;0\;6\;0\;8\;0\;6\;0\;4\;0\;2\;0 \\
\underline{0\;2\;0\;4\;0\;6\;0\;8\;0\;6\;0\;4\;0\;\mathbf{4}\;0\;4\;0\;6\;0\;8\;0\;6\;0\;4\;0\;2\;0}
\end{array}
$$

and the next row, corresponding to $n = 7$, becomes

$$0\;2\;0\;4\;0\;6\;0\;12\;0\;14\;0\;16\;0\;20\;0\;16\;0\;14\;0\;12\;0\;6\;0\;4\;0\;2\;0.$$

(The first and last zeros in a row are void.) In an obvious interpretation the above process reads

$$
\begin{aligned}
& x^{-1} + 2 + x, \\
& x^{-4} + 2x^{-3} + x^{-2} + 2x^{-1} + 4 + 2x + x^2 + 2x^3 + x^4 \\
& \quad = \big(x^{-1} + 2 + x\big)\big(x^{-3} + 2 + x^3\big), \ldots..
\end{aligned}
$$

for odd n and

$$
\begin{aligned}
&2,\ 2x^{-2}+4+2x^2,\\
&2x^{-6}+4x^{-4}+6x^{-2}+8+6x^2+4x^4+2x^6\\
&\quad = 2\big(x^{-2}+2+x^2\big)\big(x^{-4}+2+x^4\big), ...
\end{aligned}
$$

for even n. This leads to the probability generating function (7).

A simple proof for the probability generating function $f(x)$ comes from considering the binary sequence

$$j_1, j_2, ..., j_n \qquad \text{where} \quad j_k = 0 \quad \text{or} \quad 1.$$

Then we have

$$S \ = \ \sum_{k=2}^{n}(j_1 - j_k) \ + \sum_{k=3}^{n}(j_2 - j_k) \ + \ ... \ + \sum_{k=n}^{n}(j_{n-1} - j_k)$$

and the index j_k appears exactly $(n-k)-(k-1)$ times and so its contribution to the score S is

$$[(n-k)-(k-1)]j_k.$$

Therefore

$$S \ = \ (n-1)j_1 + (n-3)j_2 + \cdots + (n-2k+1)j_k + \cdots + (3-n)j_{n-1} + (1-n)j_n$$

for this sequence. Now j_k is 0 or 1 so that the generating function becomes

$$\big(1+x^{n-1}\big)\big(1+x^{n-3}\big)\cdots\big(1+x^{-(n-3)}\big)\big(1+x^{-(n-1)}\big).$$

Upon multiplication, the coefficient of x^t indicates how many times $S = t$ among all possible sequences $j_1, j_2, ..., j_n$. Dividing by the total number of sequences we arrive at the probability generating function (7).

The characteristic function for S is $\phi(\theta) \ = \ f(e^{i\theta})$, $i^2 \ = \ -1$. Euler's formula yields

$$\phi(\theta) \ = \begin{cases} \prod_{k=1}^{\nu} \cos^2(k\theta), & n = 2\nu + 1 \\ \prod_{k=1}^{\nu} \cos^2\big(\frac{2k-1}{2}\,\theta\big), & n = 2\nu. \end{cases} \tag{10}$$

By definition $\phi(0) = 1$ and

$$\phi^{(k)}(0) \;=\; i^{-k}\mu_k(n), \qquad k = 1, 2, ..., n.$$

By symmetry $\phi'(0) = 0$, $\phi^3(0) = 0$,..., so that all odd moments are zero and

$$\phi^{(2k)}(0) \;=\; (-1)^k \mu_{2k}(n).$$

Direct calculations yield

$$\begin{aligned}
\mu_1(n) &= 0,\\
\mu_2(n) &= \frac{n(n^2-1)}{12},\\
\mu_3(n) &= 0,\\
\mu_4(n) &= \frac{n(n^2-1)(5n^3-6n^2-5n+14)}{240}\\
\mu_5(n) &= 0,\\
\mu_6(n) &= \frac{n(n^2-1)(35n^6-126n^5+74n^4+420n^3-829n^2-294n+1488)}{4032},\\
\mu_7(n) &= 0,\\
&.......
\end{aligned}$$

The arrangements in Section 4 will shorten such calculations.

4 Approach to Normality

In order to show that the distribution for S approaches normality in the binary case, we shall prove (6). The dichotomy in formulae (10) forces us to separate the cases

$$\lim_{\nu\to\infty} \frac{\mu_{2k}(2\nu)}{\big(\sigma(2\nu)\big)^{2k}} \;=\; \frac{(2k)!}{2^k \cdot k!}, \qquad \lim_{\nu\to\infty} \frac{\mu_{2k}(2\nu+1)}{\big(\sigma(2\nu+1)\big)^{2k}} \;=\; \frac{(2k)!}{2^k \cdot k!}.$$

However, both cases are so similar that we shall write down only the odd case $n = 2\nu + 1$. Then the characteristic function is

$$\phi(\theta) \;=\; cos^2(\theta)\, cos^2(2\theta) \cdots cos^2(\nu\theta)$$

and by logarithmic differentiation

$$\phi'(\theta) \;=\; -2\,\phi(\theta)\sum_{j=1}^{\nu} j\;\tan(j\theta). \tag{11}$$

Denoting

$$A(\theta) \;=\; -2\sum_{j=1}^{\nu} j\;\tan(j\theta), \tag{12}$$

we obviously have

$$A(0) \;=\; 0,\; A''(0) \;=\; 0,\; A^{(4)}(0) \;=\; 0, ...$$

In order to calculate the odd derivatives $A'(0),\; A'''(0), ...$ we use the expansion

$$\tan(z) \;=\; \sum_{k=1}^{\infty} \frac{2^{2k}\big(2^{2k}-1\big)}{(2k)!}\,(-1)^{k+1}\mathrm{B}_{2k} z^{2k-1} \qquad \left(|z|^2 < \frac{\pi^2}{4}\right)$$

where $\mathrm{B}_0 = 1, \mathrm{B}_2 = 1/6, \mathrm{B}_4 = -1/30, ...$ are the Bernoulli numbers. Thus

$$A(\theta) \;=\; \sum_{k=1}^{\infty} \frac{2^{2k}\big(2^{2k}-1\big)}{(2k)!}\,(-1)^{k}\mathrm{B}_{2k}\big(1^{2k}+2^{2k}+\cdots+\nu^{2k}\big)\theta^{2k-1} \tag{13}$$

for $|\theta| < \pi/2\nu$. We deduce that

$$A^{(2k-1)}(0) \;=\; \frac{(-1)^k}{k}\,2^{2k}\big(2^{2k}-1\big)\mathrm{B}_{2k}\,s_\nu(2k) \tag{14}$$

where

$$\begin{aligned} s_\nu(2k) \;&=\; 1^{2k}+2^{2k}+\cdots+\nu^{2k} \\ &=\; \frac{\nu^{2k+1}}{2k+1}+\frac{\nu^{2k}}{2}+\frac{k\nu^{2k-1}}{6}+\langle\text{lower terms}\rangle \end{aligned}$$

are well-known polynomials of degree $2k+1$.

According to (11) and (12) we obtain[4] by Leibniz rule

$$\begin{aligned}
\phi'(\theta) &= \phi(\theta)A(\theta) \\
\phi''(\theta) &= \phi'(\theta)A(\theta) + \phi(\theta)A'(\theta) \\
\phi'''(\theta) &= \phi''(\theta)A(\theta) + 2\phi'(\theta)A'(\theta) + \phi(\theta)A''(\theta) \\
&\vdots \\
\phi^{k+1}(\theta) &= \sum_{j=0}^{k} \binom{k}{j} \phi^{(j)}(\theta)A^{(k-j)}(\theta) \\
&\vdots
\end{aligned}$$

In passing, we calculate

$$\phi''(0) = A'(0) = -\frac{n\big(n^2-1\big)}{12}$$

$$\phi^{(4)} = 3\phi''(0)^2 + A'''(0) = \frac{n\big(n^2-1\big)\big(5n^3-6n^2-5n+14\big)}{240}$$

for odd n. For even n we shall arrive at the same formulae. We have obtained that

$$\sigma^2(n) = \frac{n\big(n^2-1\big)}{12} \tag{15}$$

$$\mu_4(n) = \frac{n\big(n^2-1\big)\big(5n^3-6n^2-5n+14\big)}{240}. \tag{16}$$

This shows that (6) holds at least for $k=1$ and $k=2$. For general k we use induction.

To this end, notice that at the point $\theta = 0$ we have

$$\phi^{(2k)} = (2k-1)\phi''\phi^{(2k-2)} + \sum_{j=2}^{k} \binom{2k-1}{2j-1} A^{(2j-1)}\phi^{(2k-2j)}$$

or, more conveniently,

$$\frac{\phi^{(2k)}}{(\phi'')^k} = (2k-1)\frac{\phi^{(2k-2)}}{(\phi'')^{k-1}} + \sum_{j=2}^{k} \binom{2k-1}{2j-1} \frac{\phi^{(2k-2j)}}{(\phi'')^{k-j}} \frac{A^{(2j-1)}}{(\phi'')^j}. \tag{17}$$

[4]The connexion with cumulants is obvious, since $\frac{d}{d\theta}\log|\phi(\theta)| = A(\theta)$.

According to (14) and (15), where $n = 2\nu + 1$ is odd, we have

$$\lim_{\nu\to\infty} \frac{A^{(2j-1)}}{(\phi'')^j} \;=\; 0, \qquad j = 2, 3, ..., k,$$

since $A^{(2j-1)} \approx \nu^{2j+1}$ and $(\phi'')^j \approx \nu^{3j}$. But now (17) shows that, if

$$\lim_{\nu\to\infty} \frac{\phi^{(2m)}}{(\phi'')^m} \;=\; 1 \cdot 2 \cdot \dots \cdot (2m-1)$$

holds for $m = 1, 2, ..., k-1$, then also

$$\lim_{\nu\to\infty} \frac{\phi^{(2k)}}{(\phi'')^k} \;=\; 1 \cdot 2 \cdot \dots \cdot (2k-1).$$

In other words, the desired conclusion (6) follows by induction with respect to k. This concludes our proof of the asymptotic normality.

5 Binary Sequences not Equiprobable

Consider again all 2^n sequences of length n consisting merely of 0's and 1's. But assume now that the probability for a 0 is $P(0) = p$ and the probability for an 1 is $P(1) = q$. Here $p + q = 1$. For example, the sequence 0 1 1 0 1 1 1 has probability p^2q^5. The figurates in Table I and Table II (at the end) are constructed via (6) below. The simple rule for the formation of these figurates is condensed in the formulae

$$f(x) \;=\; \begin{cases} (p+q)\prod\limits_{k=1}^{\nu} \left(pqx^{-2k} + p^2 + q^2 + pqx^{2k}\right), & n = 2\nu + 1 \\ \prod\limits_{k=1}^{\nu} \left(pqx^{1-2k} + p^2 + q^2 + pqx^{2k-1}\right), & n = 2\nu \end{cases} \tag{18}$$

for the probability generating function. (Of course, the factor $p + q$ outside the product is 1, but it is included to match Table II.)

The characteristic function $\phi(\theta) = f(e^{i\theta})$ is

$$\phi(\theta) \;=\; \begin{cases} (p+q)\prod\limits_{k=1}^{\nu} \left(p^2 + q^2 + 2pq\cos(2k\theta)\right), & n = 2\nu + 1 \\ \prod\limits_{k=1}^{\nu} \left(p^2 + q^2 + 2pq\cos((2k-1)\theta)\right), & n = 2\nu \end{cases} \tag{19}$$

For $p = q = 1/2$ we again obtain the expressions in Section 2.

Let us consider the case $n = 2\nu + 1$, the calculations for even n being similar. Now

$$\phi'(\theta) \;=\; \phi(\theta) \sum_{k=1}^{\nu} \frac{-4pqk\sin(2k\theta)}{p^2+q^2+2pq\cos(2k\theta)} \;=\; \phi(\theta)A(\theta)$$

with an obvious abbreviation. The well-known expansion

$$\frac{\varrho\sin(\psi)}{p^2+q^2-2\varrho\cos(\psi)} \;=\; \sum_{m=1}^{\infty} \varrho^m \sin(m\psi) \qquad (|\varrho|<1)$$

converges for

$$-\varrho \;=\; \min\{\frac{p}{q},\, \frac{q}{p}\} \quad \text{if} \quad p \neq q.$$

Having treated the case $p = q = 1/2$ in the previous sections, we assume that $p \neq q$ here. Then

$$\frac{4k\varrho\sin(2k\theta)}{p^2+q^2-2\varrho\cos(2k\theta)} \;=\; 4k \sum_{m=1}^{\infty} \varrho^m \sin(2km\theta)$$

and so we obtain

$$A(\theta) \;=\; \sum_{m=1}^{\infty} \left(4\varrho^m \sum_{k=1}^{\infty} k\sin(2km\theta) \right) \qquad (p \neq q).$$

Using the Maclaurin series for $\sin(2km\theta)$, we arrive at the formula

$$A(\theta) \;=\; \sum_{j=1}^{\infty} \left\{ (-1)^{j+1} \frac{2^{2j+1}\bigl(1^{2j}+\cdots+\nu^{2j}\bigr)}{(2j-1)!} \sum_{m=1}^{\infty} m^{2j-1}\varrho^m \right\} \theta^{2j-1} \qquad (20)$$

where some arrangements have been done. The corresponding convergence investigations are quite straightforward.

By (20), $A(0) = 0$, $A''(0) = 0$, $A^{(4)}(0) = 0, \ldots,$ and

$$A^{(2j-1)}(0) \;=\; (-1)^{j+1} 2^{2j+1} s_\nu(2j) \sum_{m=1}^{\infty} m^{2j-1}\varrho^m. \qquad (21)$$

(This expansion diverges for $\varrho = -1$, i. e. for $p = q$.) Here the infinite sum is easily calculated as the differentiated geometric series

$$\sum_{m=1}^{\infty} m^{2j-1} \varrho^m \;=\; \Big(\varrho \frac{d}{\varrho}\Big)^{2j-1} \frac{1}{1-\varrho} \qquad (j = 1, 2, 3, \dots).$$

A calculation yields

$$\begin{aligned} A'(0) \;&=\; \frac{4\nu(\nu+1)(2\nu+1)}{3} \frac{\varrho}{(1-\varrho)^2}, \\ A'''(0) \;&=\; -\frac{16}{15}\nu(\nu+1)(2\nu+1)(3\nu^2+3\nu-1)\frac{1+4\varrho+\varrho^2}{(1-\varrho)^4}\varrho \end{aligned}$$

and using

$$\begin{aligned} \phi''(0) \;&=\; A'(0), \\ \phi^{(4)}(0) \;&=\; 3\phi''(0)A'(0) + A'''(0) \;=\; 3A'(0)^2 + A'''(0) \end{aligned}$$

we arrive at (8) and (9). —The corresponding calculations for even n yield the same final result.

An analogous investigation as that in Section 4, but now based on (21), shows the *approach to normality* also for $p \neq q$. The difference is merely technical.

6 The Variance (with General ℓ).

Consider again all sequences $x_1, x_2, \dots, x_n$ that can be formed of the digits $0, 1, 2, \dots, \ell$. Let $P_n(t; j_0, j_1, \dots, j_{\ell-1})$ count the number of those sequences consisting of j_0 0's, j_1 1's,..., $j_{\ell-1}$ ℓ's, $j_0 + j_1 + \cdots + j_{\ell-1} = n$, for which $S = t$. For example $P_6(1; 3, 3) = 3$, $P_6(4; 2, 4) = 2$, $P_9(27; 3, 3, 3) = 1$, and $P_9(t; 3, 3, 3) = 0$ when $t \geq 28$.

Constructing the sequence $x_1, x_2, \dots, x_n, x_{n+1}$ from $x_1, x_2, \dots, x_n$ we arrive at the fundamental recursive rule

$$\begin{aligned} &P_{n+1}(t; i_0, i_1, \dots, i_{\ell-1}) \\ &\quad = P_n(t - i_1 - \cdots - i_{\ell-1}; i_0 - 1, i_1, \dots, i_{\ell-1}) \\ &\qquad +\; P_n(t + i_0 - i_2 - \cdots - i_{\ell-1}; i_0, i_1 - 1, \dots, i_{\ell-1}) + \cdots + \\ &\qquad +\; P_n(t + i_0 + \cdots + i_{\ell-2}; i_0, i_1, \dots, i_{\ell-2}, i_{\ell-1} - 1) \end{aligned} \tag{22}$$

where now $i_0 + i_1 + \cdots + i_{\ell-1} = n+1$. In passing, we notice that applying the recursive formula twice we obtain for $\ell = 2$ that

$$
\begin{gathered}
P_{n+1}(t; i_0, i_1) \;=\; P_{n-1}(t - 2i_1; i_0 - 2, i_1) \\
+P_{n-1}(t + i_0 - 1; i_0 - 1, i_1 - 1) + P_{n-1}(t + i_0 - i_1 + 1; i_0 - 1, i_1 - 1) \\
+P_{n-1}(t + 2i_0; i_0, i_1 - 2).
\end{gathered}
\tag{23}
$$

Repeated use of this identity describes exactly how the configurations in Tables I and II are built up.

Let us return to the recursive rule (22). For the calculation of the variance $\mu_2(n)$ we assume that all sequences $x_1, x_2, \ldots, x_n$ of length n are equiprobable. Then

$$
\mu_2(n) \;=\; \sum_{x_1 \cdots x_n} \frac{S^2(x_1, x_2, \ldots, x_n)}{\ell^n},
$$

where the sum is taken over all the ℓ^n possible sequences. The auxiliary quantities

$$
B_n(j_0, \ldots, j_{\ell-1}) \;=\; \sum_t t^2 P_n(t; j_0, \ldots, j_{\ell-1})
$$

satisfy according to the recursive rule (22) the formula

$$
\begin{aligned}
& B_{n+1}(i_0, \ldots, i_{\ell-1}) \\
&= \sum_{k=0}^{\ell-1} \sum_t \bigl(t + i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}\bigr)^2 \times \\
&\qquad P_n(t + i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}; i_0, \ldots, i_k - 1, \ldots, i_{\ell-1}) \\
&-2 \sum_{k=0}^{\ell-1} \sum_t \bigl(i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}\bigr) \times \\
&\qquad \bigl(t + i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}\bigr) \times \\
&\qquad P_n(t + i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}; i_0, \ldots, i_k - 1, \ldots, i_{\ell-1}) \\
&+ \sum_{k=0}^{\ell-1} \sum_t \bigl(i_0 + i_1 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}\bigr)^2 \times \\
&\qquad P_n(t + i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}; i_0, \ldots, i_k - 1, \ldots, i_{\ell-1}),
\end{aligned}
$$

where $i_0 + i_1 + \cdots + i_{\ell-1} = n+1$. Here the first inner sum over t is merely $B_n(i_0, i_1, \ldots, i_{k-1}, i_k - 1, i_{k+1}, \ldots, i_{\ell-1})$ and the second inner sum is zero by symmetry. In the last inner sum $\sum_t P_n$ is a certain number of combinations.

Therefore the above formula reduces to the simple expression

$$\begin{aligned} B_{n+1}(i_0, \ldots, i_{\ell-1}) \;=\; & \sum_{k=0}^{\ell-1} B_n(i_0, \ldots, i_{k-1}, i_k - 1, i_{k+1}, \ldots, i_{\ell-1}) \\ & + \sum_{k=0}^{\ell-1} \bigl(i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{\ell-1}\bigr)^2 \binom{n}{i_0 \cdots i_{k-1}\, i_k - 1\, i_{k+1} \cdots i_{\ell-1}}. \end{aligned}$$

Here

$$\binom{n}{i_0 \cdots i_k - 1 \cdots i_{\ell-1}} \;=\; \frac{n!}{i_0! \cdots (i_k - 1)! \cdots i_{\ell-1}!}$$

is the usual multinomial coefficient. Summing all equations with $i_0 + i_1 + \cdots + i_{\ell-1} = n+1$ and noting that

$$\ell^n \mu_2(n) \;=\; \sum_{j_0 + \cdots + j_{\ell-1} = n} B_n(j_0, \ldots, j_{\ell-1}),$$

we obtain

$$\begin{aligned} \mu_2(n+1) \;=\; & \mu_2(n) \\ & + \; \ell^{-n-1} \sum_{k=0}^{\ell-1} \sum_{j_0, \cdots, j_{\ell-1}} \bigl(i_0 + \cdots + i_{k-1} - i_{k+1} - \cdots - i_{l-1}\bigr)^2 \times \\ & \qquad \binom{n}{i_0 \cdots i_k - 1 \cdots i_{\ell-1}}. \end{aligned}$$

The parenthesis in the sum can be written as the sum of products $\pm i_\alpha i_\beta$, $\ell - 1$ of which are of the form i_α^2, $k^2 + (\ell - k - 1)^2$ of which are of the form $+i_\alpha i_\beta$ ($\alpha \neq \beta$), and $2k(\ell - k - 1)$ of which are negative. Using well-known identities like

$$\sum_{j_0 + \cdots + j_{\ell-1} = n} j_0 j_1 \binom{n}{j_0 \cdots j_{\ell-1}} \;=\; n(n-1)\ell^{n-2}$$

$$\sum_{j_0 + \cdots + j_{\ell-1} = n} j_0^2 \binom{n}{j_0 \cdots j_{\ell-1}} \;=\; n\ell^{n-1} + n(n-1)\ell^{n-2}$$

we finally obtain the equation

$$\mu_2(n+1) \;=\; \mu_2(n) + n\,\frac{\ell-1}{\ell} + \frac{n(n-3)}{3}\,\frac{\ell^2-1}{\ell^2}. \tag{24}$$

Adding the n first equations (24) and noting that $\mu_2(1) = 0$, we reach the final result (4). This concludes our proof for the variance $\sigma^2(n)$.

The fourth moment μ_4 in (5) is the result of a similar, although more tedious, calculation. However, a more effective method should be invented for higher moments. —Corresponding formulae for S^+ are given in [L1].

In passing we mention the formula

$$S(i_0, \ldots, i_{\ell-1}) \;=\; 2S^+(i_0, \ldots, i_{\ell-1}) - \frac{n^2 - \left(i_0^2 + \cdots + i_{\ell-1}^2\right)}{2},$$

$i_0 + \cdots + i_{\ell-1} = n$. Here the exceptional notation is understandable.

7 Edgeworth's Approximation

The closedness to normality of the distribution for S is good, when l^n is large. However, if l^n is not large, especially the tails of the distribution behave obstinately, so that the assumption of normality is somewhat inadequate for precisely those values of S whose significance may be in doubt. Fortunately, numerical calculations indicate that a correction based on Edgeworth's series gives an accurate approximation.

Let

$$\Phi(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

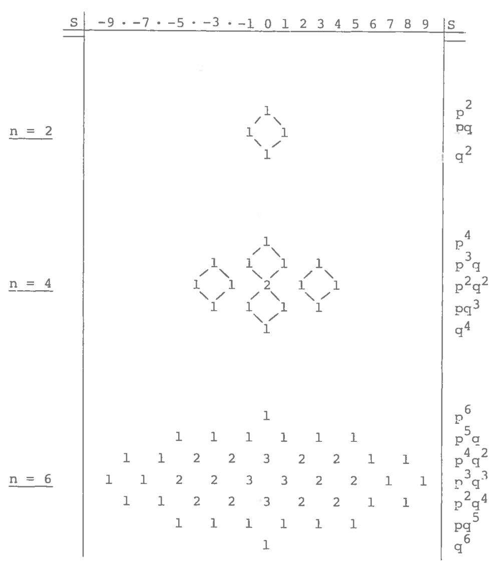
denote the normal distribution function. The approximation

$$\mathrm{P}\{S \;\leq\; \sigma(n)x\} \;\approx\; \Phi(x) + \frac{1}{4!}\Big(\frac{\mu_4(n)}{\sigma(n)^4} - 3\Big)\Phi^{(4)}(x) \tag{25}$$

is obtained from Edgeworth's series [C, page 229], terms containing $\mu_6,\ \mu_7, \ldots$ being neglected.

It stands to reason that (25) is accurate, $\mu_4(n)$ and $\sigma(n)$ being calculated from (5) and (4), provided that ℓ^n is large, say $\ell^n > 10^6$. —The dependence on ℓ is slightly puzzling. This is a point that requires further numerical investigation.

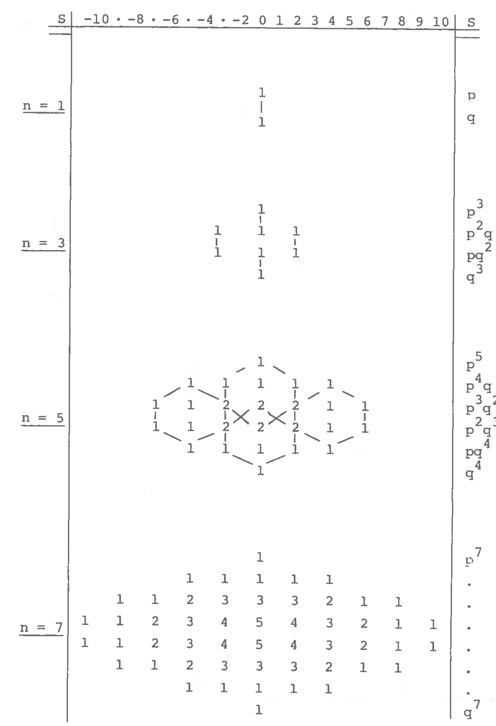
TABLE I The binary distribution for S



For example 1 / 1 1 / 1 stands for (0 1) (0 0) / (1 1) (1 0).

The p^iq^j, $i+j = n$, $p+q = 1$, denote the corresponding frequences.

TABLE II The binary distribution for S



For example 1 / 1 stands for (0) / (1) .

The p^ip^j, $i+j = n$, $p+q = 1$, denote the corresponding frequences.

References

[C] H. Cramér, *Mathematical Methods of Statistics.* Uppsala 1945.

[GR] I. Gradshteyn, I Ryzhik, *Table of Integrals, Series, and Products* (fourth edition), New York - London 1965.

[K1] M. Kendall, *A new measure of rank correlation*, Biometrika **30**, 1938, pp. 81–93.

[K2] M. Kendall, *Rank Correlation Methods* (fourth edition), Bristol 1970.

[L1] P. Lindqvist, *On a statistical "trend-test"*, Report - Mat - **A131**, Helsinki University of Technology 1978, pp. 1–13.

[L2] P. Lindqvist, *A test against trend in random sequences*, Report - Mat - **A234**, Helsinki University of Technology 1986, pp. 1–23.

[M] H. Mann, *Nonparametric tests against trend*, Econometrica **13**, 1945, pp. 245–259..

[S] G. Silitto, *The distribution of Kendall's τ coefficient of rank correlation in rankings containing ties*, Biometrika **34** , 1947, pp. 36–40.

[S2] H. Silverstone, *A note of the cumulants of Kendall's S-distribution*, Biometrika **37**, 1950, pp. 231–235.