FURTHER CONTRIBUTIONS TO THE THEORY OF PROBABILITY DISTRIBUTIONS OF POINTS ON A LINE—III

By P. V. Krishna Iyer

Indian Council of Agricultural Research, New Delhi

Introduction

Parts I and II of this series (1950, 1951) dealt with the moment and the probability generating functions of a number of distributions arising from m points possessing one of k characters arranged on a line. The discussions were simplified to a considerable extent by the use of certain results established by the author (1950) [*vide* Fréchet (1940) also]. These results which are of wide applicability have been used in this paper (i) for simplifying the discussions of a number of other distributions considered by Tukey (1949), Dixon (1940), Stevens (1939), Mann (1945), Kermack and McKendrick (1937), and Kendall (1945) and (ii) to generalize some of the distributions considered by these authors. Besides, a new distribution useful in testing k samples has also been discussed. The paper consists of six sections. Section 1 deals with moments of the distributions considered by Tukey, Dixon and Stevens. Sections 2, 3, 4 and 5 are devoted respectively to the distribution of the number of (i) positive or negative differences between successive observations, (ii) peaks and troughs or runs up and down, (iii) positive or negative differences between all possible pairs of observations and (iv) Kendall and Babington Smith's circular triads for a given sample. Section 6 deals with a special distribution for the number of positive or negative differences arising in k samples. It has been pointed out that this distribution can be used for testing the significance of the differences between k samples.

1. Simple Methods for Derivation of Moments of Certain Distributions

Random group size distribution. Tukey (1949) has recently considered the distribution of n_b, the number of boxes each containing b balls, when m balls are independently distributed in N boxes, the chance of a ball entering a box being p. The factorial moments of this distribution follow immediately by considering the expectation for one, two, three, etc., boxes each containing b balls.

In particular we derive the following results:

$$U \equiv E(n) = \sum_{r=0}^{\nu-1} \sum_{i} A_r (1 - S_r)^{-1},$$

where

$$S_r = p_{i_1} + p_{i_2} + \ldots + p_{i_r}$$
$$H \equiv E(\Sigma\lambda_i \bar{Y}_i) = UY$$

$$E(\Sigma\lambda_i \bar{Y}_i)^2 = \sum_{r=0}^{\nu-1} \Sigma \frac{K(1 - S_r) + 2YB_r}{(1 - S_r)^2} A_r$$

where

$$K = \sum_{i=1}^{N} p_i \; Y_i^2$$

and

$$B_r = p_{i_1} Y_{i_1} + p_{i_2} Y_{i_2} \ldots + p_{i_r} Y_{i_r}.$$

$$E(n^2) = \sum_{r=0}^{\nu-1} \Sigma \frac{1 + S_r}{(1 - S_r)^2}$$

$$E(n\Sigma\lambda_i Y_i) = \sum_{r=0}^{\nu-1} \Sigma \frac{Y + B_r}{(1 - S_r)^2} A_r$$

$$E(\Sigma\lambda_i^2 Y_i^2) = KU + 2\sum_{r=0}^{\nu-1} \Sigma C_r (1 - S_r)^{-2} A_r,$$

where

$$C_r = p_{i_1}^2 Y_{i_1}^2 + p_{i_2}^2 Y_{i_2}^2 + \ldots + p_{i_r}^2 Y_{i_r}^2$$

Using these results, the variance of the ratio estimate, correct to order $\frac{1}{\nu}$ works out to

$$\frac{H^2}{U^2}\left[\frac{1}{U^2}\sum_{r=0}^{\nu-1}\Sigma \frac{1 + S_r}{(1 - S_r)^2} A_r - \frac{2}{UH}\Sigma\Sigma \frac{Y + B_r}{(1 - S_r)^2} A_r \right.$$
$$\left. + \frac{1}{H^2}\Sigma\Sigma \frac{K(1 - S_r) + 2YB}{(1 - S_r)^2} A_r + \frac{U}{nH^2} \sigma_\omega^2 \right].$$

SUMMARY

The gain in efficiency achieved by avoiding the repetition of sub-units in a two-stage sampling plan has been explicitly worked out, the percentage reduction in the within unit component of the variance being very nearly equal to the over-all sampling fraction. Expressions have been given for estimating the variance from a sample. Optimum values of m and n have been worked out which minimize the expected cost for a given variance.

A sampling procedure has been postulated which though carried out with replacement, ensures a given number of distinct primary units and an expression for variance of the estimate has been worked out.

4

The probability for a box of b balls from m of the balls is

$$\binom{m}{b} p^b (1-p)^{m-b}. \tag{1.1}$$

When they are distributed in N boxes, the expected number of boxes each having b balls is

$$\mu'_{[1]}(b) = N\binom{m}{b} p^b (1-p)^{m-b}. \tag{1.2}$$

The probability for two boxes of b balls each is

$$\frac{m!}{(b!)^2 (m-2b)!} p^{2b} (1-p)^{m-2b}. \tag{1·3}$$

The expectation for two boxes of b balls each is

$$\mu'_{[2]}(b, b) = \frac{N^{(2)}\, m!}{(b!)^2 (m-2b)!} p^{2b} (1-p)^{m-2b}. \tag{1·4}$$

The expectation for two boxes, one of b balls, and the other of c balls is

$$\mu'_{[1,1]}(b, c) = \frac{N^{(2)}\, m!}{b!\, c!\, (m-b-c)!} p^{b+c} (1-p)^{m-b-c}. \tag{1.5}$$

The rth factorial moment about the origin for the distribution of n_b is given by

$$\mu'_{[r]}(b, b, \ldots r \text{ times}) = \frac{N^{(r)}\, m!}{(b!)^r (m-rb)!} p^{rb} (1-p)^{m-rb}. \tag{1.6}$$

Distribution of groups in a sequence of alternatives.—Stevens (1939) has dealt at length with the distribution of the number of cells containing at least one object, when m objects are distributed in N cells at random. The rth factorial moment of this distribution follows from a consideration of the expectation for r cells, each containing at least one object. Thus

$$\mu'_{[1]} = N \text{ (probability for a cell to contain at least one object)}$$

$$= N\left\{1 - \left(1 - \frac{1}{N}\right)^m\right\}. \tag{1.7}$$

$$\mu'_{[2]} = N^{(2)} \text{ (probability for two cells with at least one object)}$$

$$= N^{(2)}\left\{1 - 2\left(1 - \frac{1}{N}\right)^m + \left(1 - \frac{2}{N}\right)^m\right\}. \tag{1.8}$$

$$\mu'_{[r]} = N^{(r)} \text{ (probability for } n \text{ cells such that each contains at least one object)}$$

$$= N^{(r)}\left\{\sum_{s=0}^{r} (-)^s \binom{r}{s}\left(1 - \frac{s}{N}\right)^m\right\}. \tag{1.9}$$

Distribution of Dixon's 'm'.—In developing a non-parametric method of testing the significance of the difference between two samples 0_m and 0_n of sizes m and n respectively, Dixon (1940) has considered the distribution of m_i, the number of observations of 0_m lying between two consecutive values of 0_n when 0_m and 0_n are pooled together and arranged in ascending or descending order. Let m_i be the number of observations in the ith interval of the $(n+1)$ intervals of the sample 0_n and belonging to the sample 0_m. The rth factorial moment for the distribution of m_i is the expectation for r observations of 0_m in the ith interval. The probability of having r of the 0_m observations in the ith interval is

$$\frac{r!}{(n+r)^{(r)}} \tag{1.10}$$

This can be seen from the fact that r observations of the sample 0_m can be inserted in the $(n+1)$ intervals of 0_n in $\binom{n+r}{r}$ ways of which one alone has got all the r observations in the ith interval. Now the expectation for r of the observations to lie in a given interval is

$$\frac{\mu'_{[r]}}{r!} = \binom{m}{r} \Big/ \binom{n+r}{r} = \frac{m^{(r)}}{(n+r)^{(r)}}. \tag{1.11}$$

The factorial product moment for the joint distribution of m_i and m_j of orders l and k respectively is given by

$$\frac{\mu'_{[l,k]}\,(m_i^l m_j^k)}{l!\,k!} = \text{The expectation for } l \text{ and } k \text{ of the observations of } 0_m \text{ in the intervals } i \text{ and } j \text{ respectively.}$$

$$= \frac{m^{(k+l)}}{(n+k+l)^{(k+l)}} \cdot \frac{(k+l)!}{k!\,l!} \tag{1·12}$$

2. Distribution of Signs of Differences

A sample of n different observations can be arranged in a sequence in $n!$ ways. Let there be R positive and S negative differences between the successive values in a given arrangement of the observations. It is obvious that the sum of the quantities S and R is equal to $(n-1)$. Moore and Wallis (1943), and later Mann (1945) have considered the distribution of S in the $n!$ arrangements. The cumulants of this distribution for the following general cases have been considered in this section:

(i) The sample consists of observations belonging to l classes with fixed probabilities $p_1, p_2, \ldots p_l$.

(ii) The sample consists of $n_1, n_2, \ldots n_l$ observations having the values $\theta_1, \theta_2, \ldots \theta_l$ respectively.

For convenience some of the results given by Mann have also been obtained by the methods described in Part I.

The first four factorial moments can be obtained from the expectation for one, two, three and four *positive* signs.

$$\left.\begin{aligned}
\mu'_{[1]} &= (n-1)\tfrac{1}{2}\\
\frac{\mu'_{[2]}}{2!} &= \binom{n-2}{1}\frac{1}{3!} + \binom{n-2}{2}\left(\frac{1}{2}\right)^2\\
\frac{\mu'_{[3]}}{3!} &= \binom{n-3}{1}\frac{1}{4!} + 2\binom{n-3}{2}\left(\frac{1}{3!}\right)\left(\frac{1}{2}\right) + \binom{n-3}{3}\left(\frac{1}{2}\right)^3\\
\frac{\mu'_{[4]}}{4!} &= \binom{n-4}{1}\frac{1}{5!} + \binom{n-4}{2}\left\{\left(\frac{1}{3!}\right)^2 + 2\left(\frac{1}{4!}\right)\left(\frac{1}{2}\right)\right\}\\
&\quad + \binom{n-4}{3}\left(\frac{3!}{2!\,1!}\right)\left(\frac{1}{3!}\right)\left(\frac{1}{2}\right)^2 + \binom{n-4}{4}\left(\frac{1}{2}\right)^4
\end{aligned}\right\} \quad (2.1)$$

The factorial κ's reduce to the following:

$$\left.\begin{aligned}
\kappa_{[1]} &= (n-1)\frac{1}{2}\\
\kappa_{[2]} &= 2(n-2)\frac{1}{3!} + (5-3n)\left(\frac{1}{2}\right)^2\\
\kappa_{[3]} &= 3!\,(n-3)\frac{1}{4!} + 12(5-2n)\left(\frac{1}{3!}\right)\left(\frac{1}{2}\right)\\
&\quad + 4(5n-11)\left(\frac{1}{2}\right)^3\\
\kappa_{[4]} &= 4!\,(n-4)\frac{1}{5!} + 12(16-5n)\left(\frac{1}{3!}\right)^2\\
&\quad + 4!\,(17-5n)\left(\frac{1}{4!}\right)\left(\frac{1}{2}\right)\\
&\quad + 4!\,(15n-44)\left(\frac{1}{3!}\right)\left(\frac{1}{2}\right)^2 + 2(279-105n)\left(\frac{1}{2}\right)^4
\end{aligned}\right\} \quad (2\cdot 2)$$

The κ's can be calculated by using the relations

$$\left.\begin{aligned}
\kappa_1 &= \kappa_{[1]},\\
\frac{\kappa_2}{2!} &= \frac{\kappa_{[2]}}{2!} + \frac{\kappa_{[1]}}{2!},\\
\frac{\kappa_3}{3!} &= \frac{\kappa_{[3]}}{3!} + \frac{\kappa_{[2]}}{2!} + \frac{\kappa_{[1]}}{3!},\\
\frac{\kappa_4}{4!} &= \frac{\kappa_{[4]}}{4!} + \frac{3!}{2!\,1!}\frac{\kappa_{[3]}}{2!\,3!} + \frac{\kappa_{[2]}}{2!}\left\{\frac{1}{(2!)^2} + \frac{2!}{3!}\right\} + \frac{\kappa_{[1]}}{4!}
\end{aligned}\right\} \quad (2\cdot 3)$$

The above relations have been obtained by equating the coefficients of t, t^2, t^3, ... on either side of the identity

$$\sum \frac{\kappa_r}{r!} t^r = \sum \frac{\kappa_{[r]}}{r!} (e^t - 1)^r .$$

It can be seen that the first four cumulants are linear expressions in n. That the higher cumulants are also linear in n can be established, as in Part I, from the fact that the difference equation connecting the moment generating-functions (M.G.F.) for n, $n - 1$, ... observations is given by

$$M_n = \sum_{r=1}^{n} \frac{\theta^{r-1}}{r!} M_{n-r}, \qquad (2\cdot4)$$

where $\theta = e^t - 1$.

The probability-generating-function is obtained by putting $e^t = \xi$ and changing M to ϕ. It may also be observed that the distribution of S can be worked out by using the recurrence relation

$$P_n (S) = (S + 1) P_{n-1} (S) + (n - S) P_{n-1} (S - 1), \qquad (2.5)$$

where $P_n (S)$ stands for the number of arrangements having S negative differences.

So far we discussed the distribution of S when all the observations are different from one another. We shall now examine the distribution of S when the probability of the values being $\theta_1, \theta_2, \ldots \theta_l$ are $p_1, p_2, \ldots p_l$ respectively, *i.e.*, for free sampling. The factorial moments for the distribution of S can be shown to be as follows:

$$\left.\begin{aligned}
\mu'_{[1]} &= (n - 1) a_2, \\
\frac{\mu'_{[2]}}{2!} &= \binom{n-2}{1} a_3 + \binom{n-2}{2} a_2^2, \\
\frac{\mu'_{[3]}}{3!} &= \binom{n-3}{1} a_4 + 2 \binom{n-3}{2} a_3 a_2 + \binom{n-3}{3} a_2^3, \\
\frac{\mu'_{[4]}}{4!} &= \binom{n-4}{1} a_5 + \binom{n-4}{2} (a_3^2 + 2a_1 a_4) \\
&\quad + \binom{n-4}{3} \frac{3!}{2!\,1!} a_2^2 a_3 + \binom{n-4}{4} a_2^4,
\end{aligned}\right\} \qquad (2.6)$$

where the a's are unitary symmetric functions in p's.

The factorial cumulants reduce to the following expressions:

$$\begin{aligned}\kappa_{[1]} &= (n-1)\, a_2,\\ \kappa_{[2]} &= 2\,(n-2)\, a_3 + (5-3n)\, a_2^2,\\ \kappa_{[3]} &= 3!\,(n-2)\, a_4 + 12\,(5-2n)\, a_3 a_2 + 4\,(5n-11)\, a_2^3\\ \kappa_{[4]} &= 4!\,(n-4)\, a_5 + 12\,(16-5n)\, a_3^2 + 4!\,(17-5n)\, a_4 a_2\\ &\quad + 4!\,(15n-44)\, a_3 a_2^2 + 2\,(279-105n)\, a_2^4, \end{aligned} \tag{2.7}$$

The M.G.F. of the distribution satisfies the difference equation

$$M_n = \Sigma a_{r+1}\, \theta^r\, M_{n-r-1}, \tag{2·8}$$

where $\theta = e^t - 1$.

The results for non-free sampling can be obtained by substituting

$$\frac{n_g^{(r)} n_h^{(s)} \ldots}{n^{(r+s\ldots)}} \text{ for } p_g^{\,r}\, p_h^{\,s} \ldots .$$

in 2.6 after expanding them in terms of p's. The first and the second moments after this substitution are given below:

$$\left.\begin{aligned}\kappa_1 &= \frac{1}{n}\, \Sigma\, n_r\, n_s,\\ \kappa_2 &= \frac{\Sigma\, n_r n_s n_t}{n\,(n-1)} + \frac{\Sigma\, n_r^2 n_s^2}{n^2\,(n-1)} - \frac{2\,\Sigma\, n_r n_s n_t u_u}{n^2\,(n-1)};\end{aligned}\right\} \tag{2.9}$$

$r > s > t > u$.

It can be easily seen that both for free and non-free sampling the cumulants are either linear functions in n or can be put in the form $An + B + 0\left(\frac{1}{n}\right)$, where A and B are finite. Therefore the distribution tends to the normal form as $n \to \infty$.

3. Runs Up and Down

A run of length p may be defined as p successive observations in ascending or descending order. For n different observations the distribution of the total number of runs of length p or more has been considered in detail by Kermack and McKendrick (1937), Levene and Wolfowitz (1944) and others. The total number of peaks and troughs discussed by Kendall (1945) is equal to the total number of runs *minus* one. We shall extend these results to the case in which the observations belong to l classes either with fixed probabilities $p_1, p_2, \ldots p_l$ or are such that $n_1, n_2, \ldots n_l$ observations belong to the different classes. In the first instance, the results for the case where all the observations are different are obtained by the approach described in the earlier papers.

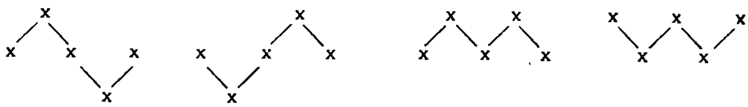
Peaks and troughs

The first four factorial moments are the expectations for one, two, three and four peaks and troughs. The probability for a peak or trough is 2/3. Two peaks and troughs can be obtained from (i) four points and (ii) five points as indicated below:

(i) *Two peaks and troughs from four points*

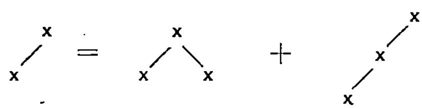
(*a*) (*b*)

(ii) *Two peaks and troughs from five points*

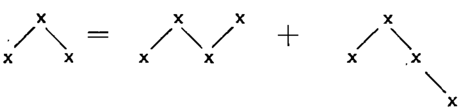


(*c*) (*d*) (*e*) (*f*)

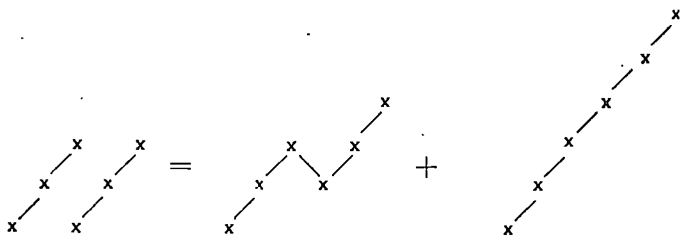
The probabilities for (i) and (ii) shown above are 5/12 and 9/20 respectively. They can be evaluated by using the relation that the product of the probabilities for two given configurations is equal to the sum of the probabilities of the configurations that can be formed from the given ones. Thus the equalities of the probabilities for the configurations shown below are obvious:



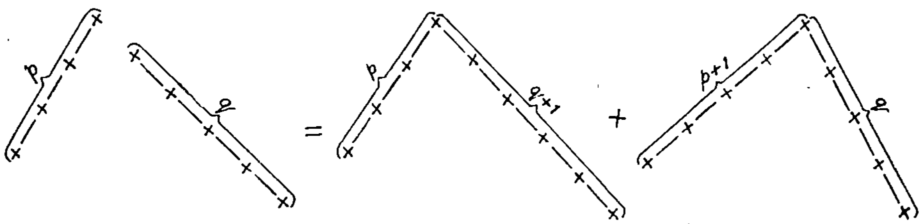
$$\frac{1}{2} = \frac{1}{3} + \frac{1}{6}$$



$$\frac{1}{3} = \left[\frac{1}{3} - \left(\frac{1}{3!} - \frac{1}{4!}\right)\right] + \left(\frac{1}{3!} - \frac{1}{4!}\right)$$

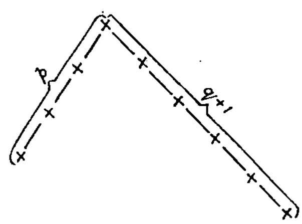


$$\left(\frac{1}{3!} \times \frac{1}{3!}\right) = \left(\frac{1}{3!} \times \frac{1}{3!} - \frac{1}{6!}\right) + \frac{1}{6!}$$



$$\frac{1}{p!} \times \frac{1}{q!} = \frac{1}{(p+q)\,(p-1)\,q!} + \frac{1}{(p+q)!\,p!\,(q-1)!}$$

The probability for the configuration



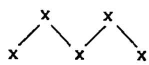
may also be obtained independently as given below. The required arrangement can be formed by keeping the highest value as the peak and dividing the remaining $(p+q-1)$ values into two groups of $p-1$ and q and arranging them on either side of the peak in ascending and descending order. This can be done in $\binom{p+q-1}{p-1}$ ways. The total number of ways of arranging the $(p+q)$ values involved in the configuration is $(p+q)!$. Hence the probability of the required configuration is

$$\frac{(p+q-1)!}{q!\,(p-1)!\,(p+q)!} = \frac{1}{q!\,(p-1)!\,(p+q)}$$

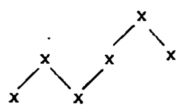
The configurations and their probabilities which are required for calculating the third and fourth factorial moments are given below:

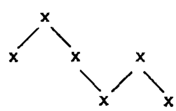
Third factorial moment (*Three peaks and troughs*)

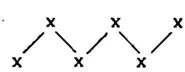
(1) *Five points*


$\frac{2}{15}$ (2)

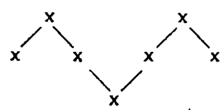
(2) *Six points*

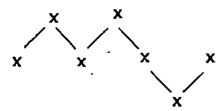

$\frac{1}{18}$ (2)

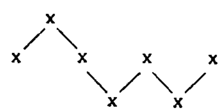

$\frac{1}{18}$ (2)

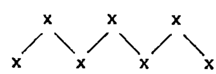

$\frac{61}{720}$ (4)

(3) *Seven points*


$\frac{132}{7!}$ (2)

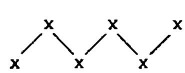

$\frac{181}{7!}$ (2)


$\frac{181}{7!}$ (2)

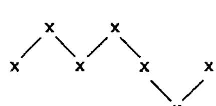

$\frac{272}{7!}$ (2)

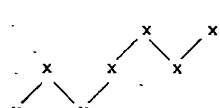
Fourth factorial moment (*Four peaks and troughs*)

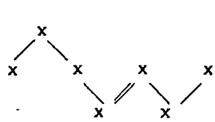
(1) *Six points*

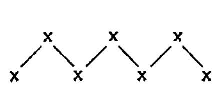

$\frac{61}{720}$ (2)

(2) *Seven points*

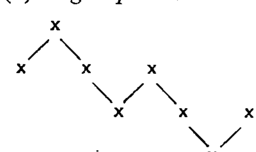

$\frac{181}{7!}$ (2)

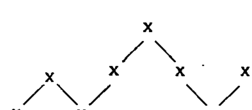

$\frac{169}{7!}$ (2)

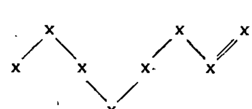

$\frac{181}{7!}$ (2)

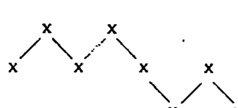

$\frac{272}{7!}$ (6)

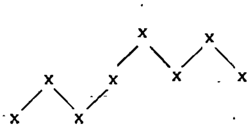
(3) *Eight points*


$\frac{589}{8!}$ (2)

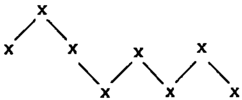

$\frac{643}{8!}$ (2)


$\frac{643}{8!}$ (2)

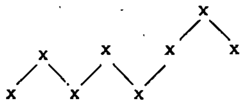

$\frac{875}{8!}$ (2)


$\frac{875}{8!}$ (2)

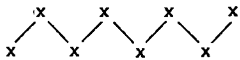
Eight points—Contd.



$\frac{917}{8!}$ (4)

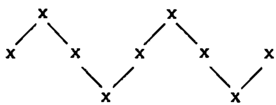


$\frac{917}{8!}$ (4)

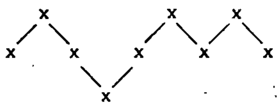


$\frac{1385}{8!}$ (6)

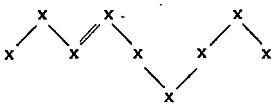
(4) *Nine points*



$\frac{2701}{9!}$ (2)

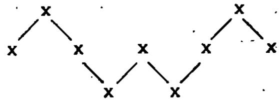


$\frac{3736}{9!}$ (2)

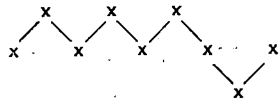


$\frac{3736}{9!}$ (2)

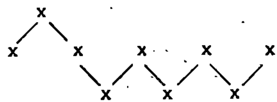
Nine points—Contd.



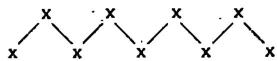
$\frac{3526}{9!}$ (2)



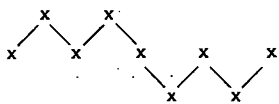
$\frac{5263}{9!}$ (2)



$\frac{5263}{9!}$ (2)



$\frac{7936}{9!}$ (2)



$\frac{5095}{9!}$ (2)

[The figures in the brackets give the number of such configurations involved in the moments.]

Using the results given above the first four factorial moments bout the origin are as follows:

$$\left.\begin{aligned}
\mu'_{[1]} &= (n-2)\, 2/3, \\
\frac{\mu'_{[2]}}{2!} &= 2\,(n-3)\, 5/24 + 2\,(n-4)\, 27/120 + \binom{n-4}{2} (2/3)^2, \\
\frac{\mu'_{[3]}}{3!} &= 2\,(n-4)\, 2/15 + 4\,(n-5)\, 101/720 + 2\,(n-6)\, \frac{766}{7!} \\
&\quad + (n-5)(n-6)(5/12)(2/3) + (n-6)(n-7)(9/20)(2/3) \\
&\quad + \binom{n-6}{3} (2/3)^3, \\
\frac{\mu'_{[4]}}{4!} &= 2\,(n-5)\, \frac{61}{6!} + 2\,(n-6)\, \frac{1347}{7!} + 2\,(n-7)\, \frac{11448}{8!} \\
&\quad + 2\,(n-8)\, \frac{37256}{9!} + (n-6)^{(2)} \frac{4}{15} \times \frac{2}{3} \\
&\quad + (n-7)^{(2)} \frac{2}{3} \times \frac{101}{720} \times 4 + (n-8)^{(2)} \frac{2}{3} \times \frac{1532}{7!} \\
&\quad + (n-6)^{(2)} \frac{5}{12} \times \frac{5}{12} \times \frac{1}{2} + (n-8)^{(2)} \frac{9}{20} \times \frac{9}{20} \times \frac{1}{2} \\
&\quad + (n-7)^{(2)} \frac{5}{12} \times \frac{9}{20} + (n-7)^{(3)} \frac{10}{24} \times \left(\frac{2}{3}\right)^2 \times \frac{1}{2} \\
&\quad + \binom{n-8}{3} (2/3)^2 \times \frac{9}{20} \times 3 + \binom{n-8}{4} (2/3)^4.
\end{aligned}\right\} \quad (3.1)$$

The moments obtained from the above expressions agree with those given by Kendall.

We shall now get the first and the second moments for the general case where the observations do not differ from one another. The first moment is the expectation for a single peak or trough. The probability for a single peak or trough from three observations is given by

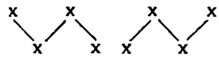
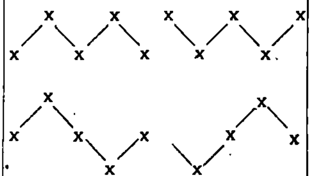
$$P_1(3) = 4\Sigma p_r\, p_s\, p_t + \Sigma p_r^2 p_s + \Sigma p_r\, p_s^2,$$

where $r > s > t$. This is obvious from the fact that a peak or a trough can be formed from three different observations in four ways and in one way if two of the observations are equal. Thus

$$\mu'_{[1]} = (n-2)\, P_1(3) \quad (3.2)$$

To obtain the second factorial moment we investigate the various ways of obtaining two peaks and troughs which cannot be treated as independent. The table below shows the number of ways of having two peaks and troughs from four and five observatiorns. Two peaks and troughs formed from six or more observations will be found to be independent.

Two peaks and troughs from four and five observations

No. of observations	Nature of observations taken in ascending order	Description of configuration	Number of configurations
4	All different.		10
4	The first or the last two observations being same and others different.	Do	4
4	The second and the third observations being same and others different.	Do	2
4	Two pairs of equal observations.	Do	2
5	All different.		54
5	First or last two observations same and others different.	Do	20
5	Second and third observations same and others different.	Do	17
5	Three observations same and others different.	Do	2
5	Three observations same, the remaining two being equal to some other value.	Do	1

Using the results of the above Table it can be easily seen that

$$\frac{\mu'_{[2]}}{2!} = (n-3)\, P_2(4) + (n-4)\, P_2(5) + \binom{n-4}{2} \{P_1(3)\}^2, \quad (3.3)$$

where

$$P_2(4) = 10\Sigma p_r p_s p_t p_u + 4\Sigma p_r^2 p_s p_t + 2\Sigma p_r p_s^2 p_t + 4\Sigma p_r p_s p_t^2 + 2\Sigma p_r^2 p_s^2,$$

$$P_2(5) = 54\Sigma p_r p_s p_t p_u p_v + 20\Sigma p_r^2 p_s p_t p_u + 17\Sigma p_r p_s^2 p_t p_u + 17\Sigma p_r p_s p_t^2 p_u + 20\Sigma p_r p_s p_t p_u^2 + 2\Sigma p_r^3 p_s p_t + 2\Sigma p_r p_s^3 p_t + 2\Sigma p_r p_s p_t^3 + \Sigma p_r^3 p_s^2 + \Sigma p_r^2 p_s^3,$$

where $r > s > t > u > v$.

The corrected second moment reduces to

$$\mu_2 = (n - 2) P_1 (3) + 2 (n - 3) P_2 (4) + 2 (n - 4) P_2 (5) + (16 - 5n) \{P_1 (3)\}^2 \quad (3.4)$$

The moments for non-free sampling can be obtained by substituting $\frac{n_r^{(l)} n_s^{(m)} n_t^{(o)} \ldots}{n^{(l+m+o \ldots)}}$ for $p_r^l p_s^m p_t^o \ldots$ in $\mu'_{[1]}$ and $\mu'_{[2]}$ about the origin. The substitution should be made only after the expansion of $\{P_1 (3)\}^2$ in terms of p's.

4. Distribution of the Total Number of Positive or Negative Signs

Let a sample consisting of n observations which are all different occur in a given order. Let x be the total number of positive differences between all the pairs of observations taken in the order left to right. Kendall's (1945) rank correlation coefficient τ is defined by

$$\tau = \frac{2 (x - y)}{n (n - 1)}, \text{ where } x + y = \frac{n (n - 1)}{2}$$

Kendall has discussed in detail the distribution of $S = (x - y)$ for the case where all the observations are different. Later he has also obtained

Frequency distribution of x *for values of* n *from* 2 *to* 5

Values of n / No. of signs	1	2	3	4	5
0	1	1	1	1	1
1		1	2	3	4
2			2	5	9
3			1	6	15
4				5	20
5				3	22
6				1	20
7					15
8					9
9					4
10					1

the variance of τ for tied rankings. The main purpose here is to show how the cumulants for the distribution of x can be evaluated by using methods similar to the one developed by the author in earlier communications (1948, 1950). This method consists in writing the first four factorial moments from the distributions for two, three, four and five different observations quoted above from Kendall (1945).

It can be established without much difficulty that

$$\left.\begin{aligned}
\mu'_{[1]}(n) &= \binom{n}{2} E_1(2), \\
\frac{\mu'_{[2]}(n)}{2!} &= \binom{n}{3} E_2(3) + \binom{n}{4} \frac{4!}{(2!)^3} \{E_1(2)\}^2, \\
\frac{\mu'_{[3]}(n)}{3!} &= \binom{n}{3} E_3(3) + \binom{n}{4} E_3(4, 3) \\
&\quad + \binom{n}{5}\left(\frac{5!}{2!\,3!}\right) 2E_1(2)\, E_2(3) \\
&\quad + \binom{n}{6}\left(\frac{6!}{(2!)^3\,3!}\right) \{E_1(2)\}^3, \\
\frac{\mu'_{[4]}(n)}{4!} &= \binom{n}{4} E_4(4) + \binom{n}{5}\left\{E_4(5, 4) + \left(\frac{5!}{2!\,3!}\right) 2E_1(2)E_3(3)\right\} \\
&\quad + \binom{n}{6}\left[\left(\frac{6!}{2!\,4!}\right) 2E_1(2)\, E_3(4, 3) + \left(\frac{6!}{(3!)^2\,2!}\right) \{E_2(3)\}^2\right] \\
&\quad + \binom{n}{7}\left(\frac{7!}{(2!)^3\,3!}\right)\left(\frac{3!}{2!\,1!}\right) \{E_1(2)\}^2 E_2(3) \\
&\quad + \binom{n}{8} \frac{8!}{(2!)^4\,4!} \{E_1(2)\}^4,
\end{aligned}\right\} \quad (4.1)$$

where

$$E_r(s) = \frac{1}{r!}\,\{\text{the } r\text{th factorial moment for the distribution of positive signs involving } s \text{ observations}\}$$

$$= \frac{\mu'_{[r]}(s)}{r!},$$

and

$$E_r(s, s-1) = \frac{1}{r!}\,\{\mu'_{[r]}(s) - s\mu'_{[r]}(s-1)\}.$$

Thus

$$E_1(2) = \mu'_{[1]}(2) = \frac{1}{2},\ E_2(3) = \frac{\mu'_{[2]}(3)}{2!} = \frac{5}{6},$$

$$E_3(3) = \frac{\mu'_{[3]}(3)}{3!} = \frac{1}{6},\ E_3(4, 3) = \frac{1}{3!}\{\mu'_{[3]}(4) - 4\mu'_{[3]}(3)\} = \frac{5}{2},$$

$$E_4(4) = \frac{\mu'_{[4]}(4)}{4!} = \frac{35}{24},\ E_4(5, 4) = \frac{1}{4!}\{\mu'_{[4]}(5) - 5\mu'_{[4]}(4)\}$$

$$= \frac{1424}{120}$$

Substituting the values of E's in the above equations and reducing we get

$$\mu'_{[1]}(n) = \frac{n(n-1)}{4},\ \mu_2(n) = \frac{n(n-1)(2n+5)}{72}.$$

$\mu_3(n)$ and $\mu_4(n)$ can be similarly calculated and agree with those that can be obtained from Kendall's (1945) values.

We may now write down the moments for the distribution of x when the observations belong to k classes with probabilities $p_1, p_2, \ldots p_k$.

$$\mu'_{[1]}(n) = \binom{n}{2} a_2,$$

$$\frac{\mu'_{[2]}(n)}{2!} = \binom{n}{3} 5a_3 + \binom{n}{4} 3a_2^2,$$

$$\frac{\mu'_{[3]}(n)}{3!} = \binom{n}{3} a_3 + \binom{n}{4} 60a_4 + \binom{n}{5} 20 \times 5a_2a_3 + \binom{n}{6} 15a_2^3,$$

$$\frac{\mu'_{[4]}(n)}{4!} = \binom{n}{4} 35a_4 + \binom{n}{5}\{1424a_4 + 20a_2a_3\} + \binom{n}{6}\{1800\, a_2a_4$$

$$+ 250a_3^2\} + \binom{n}{7} 1575a_2^2a_3 + \binom{n}{8} 105a_2^4,$$

where the a's are the unitary symmetric functions in p's such that $r > s > t \ldots$ The corresponding values for non-free sampling are obtained by substituting $\frac{n_r n_s n_t}{n(n-1)(n-2)}$ for $p_r\, p_s\, p_t \ldots$ This substitution gives

$$\mu_1(n) = \frac{\Sigma n_r n_s}{2},$$

$$\mu_2(n) = \frac{3}{4}\Sigma n_r n_s + \frac{1}{6}\Sigma n_r n_s n_t - \frac{1}{4}\Sigma n_r n_s (n_r + n_s).$$

5. Cumulants of Circular Triads 'd'

The methods outlined in the above section can be extended for obtaining the first four moments for the distribution of Kendall and Babington Smith's (1945) circular triads 'd'. For this we have simply to obtain the frequency distribution of 'd' for three, four, five, six and seven observations. These distributions have already been given by Kendall and Babington Smith. Using these distributions we get

$$E_1(3) = \frac{1}{4},\ E_2(4) = \frac{3}{8},\ E_2(5, 4) = \frac{15}{16},\ E_3(5) = \frac{25}{16},$$

$$E_3(6, 5) = \frac{15}{2},\ E_4(5) = \frac{25}{64},\ E_4(6, 5) = \frac{105}{8},\ E_4(7, 6) = \frac{5670}{64}.$$

$E_4(6, 5)$ and $E_4(7, 6)$ are obtained in succession as follows: First $E_4(5)$ is calculated from the distribution for five observations. Now we make use of the general expression for $\mu'_{[4]}(n)$ and substitute $n = 6$ and equate it to the actual fourth factorial moment for $n = 6$. This gives $E_4(6, 5)$. Similarly $E_4(7, 6)$ is determined by using the value of $E_4(5)$ and $E_4(6, 5)$ and putting $n = 7$ in $\mu'_{[4]}(n)$ and equating the expression to the actual factorial moment for $n = 7$.

The first four factorial moments can be written from the above values and are as follows:—

$$
\left.
\begin{aligned}
\mu'_{[1]}(n) &= \binom{n}{3} E_1(3), \\
\frac{\mu'_{[2]}(n)}{2!} &= \binom{n}{4} E_2(4) + \binom{n}{5} E_2(5, 4) + \binom{n}{6} \frac{6!}{(3!)^2\, 2!} \{E_1(3)\}^2, \\
\frac{\mu'_{[3]}(n)}{3!} &= \binom{n}{5} E_3(5) + \binom{n}{6} E_3(6, 5) + \binom{n}{7} \left(\frac{7!}{3!\, 4!}\right) \\
&\quad \times 2E_1(3)\, E_2(4) + \binom{n}{8} \left(\frac{8!}{3!\, 5!}\right) 2E_1(3)\, E_2(5, 4) \\
&\quad + \binom{n}{9} \left(\frac{9!}{(3!)^4}\right) \{E_1(3)\}^4, \\
\frac{\mu'_{[4]}(n)}{4!} &= \binom{n}{5} E_4\, 5) + \binom{n}{6} E_4(6, 5) + \binom{n}{7} E_4(7, 6) \\
&\quad + \binom{n}{8} \left(\frac{8!}{3!\, 5!}\right) 2E_1(3)\, E_3(5) + \binom{n}{8} \left(\frac{8!}{(4!)^2\, 2!}\right) \{E_2(4)\}^2 \\
&\quad + \binom{n}{9} \left\{ \left(\frac{9!}{3!\, 6!}\right) 2E_1(3) E_3(6, 5) + \left(\frac{9!}{4!\, 5!}\right) 2E_2(4) \right. \\
&\quad \left. \times E_2(5, 4) \right\} + \binom{n}{10} \left[\frac{10!}{(5!)^2\, 2!} \{E_2(5, 4)\}^2 \right. \\
&\quad \left. + \left(\frac{10!}{(3!)^2\, 4!\, 2!}\right) \left(\frac{3!}{2!\, 1!}\right) \{E_1(3)\}^2 E_2(4) \right] \\
&\quad + \binom{n}{11} \left(\frac{11!}{(3!)^2\, 5!\, 2!}\right) \left(\frac{3!}{2!\, 1!}\right) \{E_1(3)\}^2 E_2(5, 4) \\
&\quad + \binom{n}{12} \left(\frac{12!}{(3!)^4\, 4!}\right) \{E_1(3)\}^4.
\end{aligned}
\right\}
$$

6. A Distribution for Testing Two or More Samples

Let two samples of sizes n_1 and n_2 occur in the manner shown below:

Sample I: $a_1, a_2, a_3, \ldots a_{n_1}$,

Sample II: $b_1, b_2, b_3, \ldots b_{n_2}$.

Consider the signs of the differences

$$
\begin{aligned}
&a_1 - b_1,\ a_1 - b_2,\ a_1 - b_3,\ \ldots\ a_1 - b_{n_2}, \\
&a_2 - b_1,\ a_2 - b_2,\ a_2 - b_3,\ \ldots\ a_2 - b_{n_2}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&a_{n_1} - b_1,\ a_{n_1} - b_2,\ a_{n_2} - b_3,\ \ldots\ a_{n_1} - b_{n_2}.
\end{aligned}
$$

It will be observed that the cumulants of this distribution can also be written as follows:

$$\left.\begin{aligned}\kappa_1(n_1, n_2) &= \kappa_1(n) - \kappa_1(n_1) - \kappa_1(n_2),\\ \kappa_2(n_1, n_2) &= \kappa_2(n) - \kappa_2(n_1) - \kappa_2(n_2),\\ &\cdots\cdots\cdots\cdots\\ &\cdots\cdots\cdots\cdots\\ \kappa_r(n_1, n_2) &= \kappa_r(n) - \kappa_r(n_1) - \kappa_r(n_2),\end{aligned}\right\} \quad (6.3)$$

where $\kappa(n_1, n_2)$ refers to the cumulants of the distribution for positive or negative differences between the two samples of sizes n_1 and n_2 and $n = n_1 + n_2$. This follows from the fact that the distribution for the signs of difference between and within samples are independent of each other.

The extension of these results to k samples is obvious.

$$\kappa_r(n_1, n_2, \ldots n_k) = \kappa_r(n) - \sum_{i=1}^{k} \kappa_r(n_i) \quad (6.4)$$

The above results can be used to decide whether k given samples belong to the same population. This is done by comparing the observed number of positive or negative differences between the kC_2 pairs of samples and comparing them with the expected value on the basis of its expected variance.

It may also be observed that the method of analysis developed above can be further extended for the analysis of a randomized block experiment also. Suppose there are k blocks with l varieties in each block as shown below:

Blocks / Varieties	1	2	3	k
v_1	a_1	a_2	a_3	a_k
v_2	b_1	b_2	b_3	b_k
v_3	c_1	c_2	c_3	c_k
....	....	....	....	....
....	....	....	....	....
v_l	l_1	l_2	l_3	l_k

The number of positive differences between the rows v_1 and v_2, v_1 and v_3, ... v_1 and v_l; v_2 and v_3, v_2 and v_4, ... v_2 and v_l; v_3 and

v_4, v_3 and v_5, ... excluding those arising from the blocks will give us some idea of the significance of the differences between the varieties.

Cumulants of this distribution are as shown below:

$$\kappa_r \left[\begin{matrix} k \\ k \\ \vdots \\ l \text{ times} \end{matrix} \right] = \kappa_r (kl) - l\kappa_r (k) - k\kappa_r (l) \qquad (6 \cdot 5)$$

Making use of 6.5, a test of significance for the general difference between the varieties can be made on the basis of normalised deviates between the observed and the expected numbers of positive or negative differences provided the number of replications for each variety is not small.

The modifications that are necessary when the samples do not have all the observations different from one another are considered below: The first two factorial moments for free and non-free sampling for the case of two samples are

$$\mu_{[1]} (n_1, n_2) = n_1 n_2 \Sigma p_r p_s,$$

$$\frac{\mu_{[2]}' (n_1, n_2)}{2!} = \left[\binom{n_1}{1} \binom{n_2}{2} + \binom{n_1}{2} \binom{n_2}{1} \right] 2 \Sigma p_r p_s p_t + \binom{n_1}{1} \binom{n_2}{2} \Sigma p_r^2 p_s + \binom{n_1}{2} \binom{n_2}{1} \Sigma p_r p_s^2 + 2 \binom{n_1}{2} \binom{n_2}{2} (\Sigma p_r p_s)^2$$

and

$$\mu_{[1]}' (n_1, n_2) = \frac{n_1 n_2 \Sigma n_r n_s}{n^{(2)}}$$

$$\frac{\mu_{[2]} (n_1, n_2)}{2!} = \left\{ \binom{n_1}{1} \binom{n_2}{2} + \binom{n_1}{2} \binom{n_2}{1} \right\} \frac{2 \Sigma n_r n_s n_t}{n^{(3)}} + \frac{\binom{n_1}{1} \binom{n_2}{2} \Sigma n_r (n_r - 1) n_s}{n^{(3)}} + \frac{\binom{n_1}{2} \binom{n_2}{1} \Sigma n_r n_s (n_s - 1)}{n^{(3)}} + \frac{2}{n^{(4)}} \binom{n_1}{2} \binom{n_2}{2} \left\{ \Sigma n_r^{(2)} n_s^{(2)} + 2 \Sigma n_r^{(2)} n_s n_t + 6 n_r n_s n_t n_u \right\}, \qquad (6.6)$$

where $r > s > t > u$.