ON THE DISTRIBUTION OF 4-CYCLES IN RANDOM BIPARTITE TOURNAMENTS

J. W. Moon and L. Moser

Let there be given two sets of points, $P = \{P_1, \ldots, P_m\}$ and $Q = \{Q_1, \ldots, Q_n\}$, such that joining each pair of points (P_i, Q_k), for $i = 1, \ldots, m$ and $k = 1 \ldots, n$, is a line oriented towards one, and only one, of the pair. Such a configuration will be called an $m \times n$ bipartite tournament. If the line joining P_i to Q_k is oriented towards Q_k we may indicate this by $P_i \to Q_k$, and similarly if the line is oriented in the other sense. The points P_i, P_j, Q_k, and Q_ℓ will be said to form a 4-cycle if either $P_i \to Q_k \to P_j \to Q_\ell \to P_i$ or $P_i \to Q_\ell \to P_j \to Q_k \to P_i$. $C(m, n)$, the number of 4-cycles in a given $m \times n$ bipartite tournament, provides, in some sense, a measure of the degree of transitivity of the relationship indicated by the orientation of the lines, and the complete configuration may be thought of as representing the outcome of comparing each member of one population with each member of a second population, and making a decision, upon some basis, as to which component of each pair is the preferred one.

The object of this note is to investigate the distribution of $C(m, n)$ under the hypothesis that all orientations of the lines are equally likely. This parallels the study of the number of 3-cycles in ordinary tournaments made by Kendall [1], Moran [3], and others in connection with the method of paired comparisons. We will obtain sharp upper bounds for $C(m, n)$ and prove that under mild restrictions on the relative rates of growth of m and n the distribution of $C(m, n)$ tends to normality as m and n tend to infinity.

Canad. Math. Bull. vol. 5, no. 1, January 1962.

https://doi.org/10.4153/CMB-1962-002-0 Published online by Cambridge University Press

For an arbitrary $m \times n$ bipartite tournament let V_i, for $i = 1, \ldots, m$, be the number of distinct points, Q_k, such that $P_i \to Q_k$. Then

$$(1) \qquad \sum_{i=1}^{m} V_i (n - V_i)$$

equals the number of ordered pairs of distinct points (Q_k, Q_ℓ) for which there is a point P_i such that $Q_\ell \to P_i \to Q_k$, counting multiplicities.

Number the ordered pairs of distinct points, (Q_k, Q_ℓ), from 1 to $n(n-1)$ in such a way that if the ordered pair (Q_k, Q_ℓ) corresponds to s, where $s = 1, \ldots, n(n-1)$, then the ordered pair (Q_ℓ, Q_k) corresponds to $n(n-1) - (s-1)$. Let t_s denote the number of times the ordered pair corresponding to s is counted in the sum of (1). It follows that

$$(2) \qquad C(m,n) = \sum_{s=1}^{\binom{n}{2}} t_s \cdot t_{n(n-1)-(s-1)} .$$

Assume temporarily that $m \equiv n \equiv 0 \pmod 2$. We then observe that

$$(3) \qquad \sum_{s=1}^{n(n-1)} t_s = \sum_{s=1}^{\binom{n}{2}} (t_s + t_{n(n-1)-(s-1)})$$

$$= \sum_{i=1}^{m} V_i (n - V_i) \leq m \frac{n^2}{4} .$$

https://doi.org/10.4153/CMB-1962-002-0 Published online by Cambridge University Press

It is now easily seen that an upper bound for C(m, n) in (2) results by first having

$$(4) \qquad (t_s + t_{n(n-1)-(s-1)}) = m \qquad \text{for } \frac{n^2}{4} \text{ values of } s,$$

and zero for the others, and then setting $t_s = t_{n(n-1)-(s-1)}$. This gives

THEOREM 1: $C(m,n) \le \frac{m^2}{4} \cdot \frac{n^2}{4}$ if $m \equiv n \equiv 0 \pmod 2$.

If m is odd then $\frac{m^2}{4}$ is replaced by $\frac{m^2-1}{4}$, and similarly for n, as may be seen by making the appropriate changes in (3) and (4).

That the upper bound is sharp for the case $m \equiv n \equiv 0 \pmod 2$ is demonstrated by the bipartite tournament in which $P_i \to Q_k$ if, and only if, either $i \le \frac{m}{2}$ and $k \le \frac{n}{2}$ or $i > \frac{m}{2}$ and $k > n/2$, otherwise $Q_k \to P_i$. Equally simple examples suffice for the other cases.

In an m x n bipartite tournament define a random variable, $S(i, j; k, \ell)$, $i, j = 1, \ldots, m$, $i \neq j$, and $k, \ell = 1, \ldots, n$, $k \neq \ell$, to be 1 or 0 according as the points P_i, P_j, Q_k and Q_ℓ do or do not form a 4-cycle. Since only 2 of the 16 equally likely ways of orienting the 4 lines between these points yield a 4-cycle it follows that $E[S(i,j;k,\ell)] = 1/8$ and summing over all suitable pairs (i, j) and (k, ℓ) we have:

$$(5) \qquad E[C(m,n)] = \frac{1}{8}\binom{m}{2}\binom{n}{2}.$$

Similarly $E[C^2(m,n)] = E[(\Sigma\, S(i,j;k,\ell))^2]$, where the latter sum is again over the pairs (i, j) and (k, ℓ). In the expansion of this, the various types of products involved, and the number and expected value of such products are found to be as follows:

https://doi.org/10.4153/CMB-1962-002-0 Published online by Cambridge University Press

$S(i,j;k,\ell)\cdot S(i,j;k,\ell)$	$\binom{m}{2}\binom{n}{2}$	1/8
$S(i,j;k,\ell)\cdot S(i,j;k,s)$	$2n\binom{m}{2}\binom{n-1}{2}$	1/32
$S(i,j;k,\ell)\cdot S(i,h;k,\ell)$	$2m\binom{m-1}{2}\binom{n}{2}$	1/32
$S(i,j;k,\ell)\cdot S(i,j;r,s)$	$\binom{m}{2}\binom{n}{2}\binom{n-2}{2}$	1/64
(6) $S(i,j;k,\ell)\cdot S(g,h;k,\ell)$	$\binom{m}{2}\binom{m-2}{2}\binom{n}{2}$	1/64
$S(i,j;k,\ell)\cdot S(i,h;k,s)$	$4mn\binom{m-1}{2}\binom{n-1}{2}$	1/64
$S(i,j;k,\ell)\cdot S(i,h;r,s)$	$2m\binom{m-1}{2}\binom{n}{2}\binom{n-2}{2}$	1/64
$S(i,j;k,\ell)\cdot S(g,h;k,s)$	$2n\binom{m}{2}\binom{m-2}{2}\binom{n-1}{2}$	1/64
$S(i,j;k,\ell)\cdot S(g,h;r,s)$	$\binom{m}{2}\binom{m-2}{2}\binom{n}{2}\binom{n-2}{2}$	1/64

Combining these we have, where μ_k equals the k^{th} moment of $C(m,n)$ about its mean, that

$$(7) \qquad \mu_2[C(m,n)] = E[C^2(m,n)] - E^2[C(m,n)]$$

$$= \frac{1}{2^6}\binom{m}{2}\binom{n}{2}(2m+2n-1) .$$

To show that the distribution of $C(m,n)$ tends to normality it suffices (Kendall [2]) to show that, for $h = 1, 2, \ldots,$

$$\text{(i)} \quad \frac{\mu_{2h+1}}{\mu_2^{\frac{1}{2}(2h+1)}} \to 0 \quad \text{and} \quad \text{(ii)} \quad \frac{\mu_{2h}}{\mu_2^{h}} \to \frac{(2h)!}{2^h h!}$$

as m and n tend to infinity.

We shall temporarily assume that $n = o(m)$ as $m, n \to \infty$.

If $T(i,j;k,\ell) = S(i,j;k,\ell) - \frac{1}{8}$ then

https://doi.org/10.4153/CMB-1962-002-0 Published online by Cambridge University Press

$\mu_{2h+1} = E[(\Sigma\, T(i, j; k, \ell))^{2h+1}]$, for h = 1, 2, ..., where the sum is again over the pairs (i, j) and (k, ℓ).

A typical term in the expansion is $T(i_1, j_1; k_1, \ell_1) \ldots T(i_{2h+1}, j_{2h+1}; k_{2h+1}, \ell_{2h+1})$. Put all the T's, in this term, which have any values of i, j, k or ℓ in common with those of the first factor in a class with it. Add to this class any T's, in this term, which have any values of i, j, k or ℓ in common with those of any of the T's already in this class, and continue this process as long as possible. In a similar fashion form another class starting with, say, the first factor not already included in the first class.

By repeating this process, any term may be expressed as the product of classes of products, such that no two factors in different classes have any values of i, j, k, or ℓ in common, while for any factor in a class containing more than one factor there is another factor, in the same class, which does have a value of i, j, k or ℓ in common with it.

Combining all terms which have similar combinations of (i, j) and (k, ℓ) occurring, as was done in (6) for the second moment, the number of times terms of a given type appear will be a polynomial in m and n, whose largest term is of order equal to the number of distinct values of i and j, and k and ℓ, respectively, that appear in the term.

If any class in a term contains just one factor the expected value of that term will equal the expected value of the product of the remaining factors times the expected value of the single factor, or zero. Also, any term with more than h classes has expectation zero, since it is forced to have at least one class containing a single factor.

Restricting ourselves now to terms all of whose classes contain at least two factors we may make the following assertion. If such a term is to have a non-zero expectation, then each factor in a class may include at most one value of i, j, k, or ℓ not included in any other factor in the class. For, if any factor has fewer than three values of i, j, k, and ℓ in common with those of the remaining factors in the class, then its expectation is independent of the expectations of the remaining factors, as was seen in (6). As its expectation is then zero the expectation of the whole term is consequently zero.

Therefore, for any class of factors in a term the largest number of distinct values of i, j, k, and ℓ that may occur in it is four for the first factor and one new value for each of the remaining factors. Hence, the term containing the largest number of distinct values of i, j, k, and ℓ with a non-zero expectation will have h-1 classes of two factors each and one class of three factors. From the hypothesis on the relative orders of m and n, we see that the largest term in μ_{2h+1} is contributed by products of the form

$$\left[\prod_{s=1}^{h-1} T(i_s, j_s; k_s, \ell_s) \cdot T(g_s, j_s; k_s, \ell_s)\right]$$

$$\times\, T(i_h, j_h; k_h, \ell_h) \cdot T(g_h, j_h; k_h, \ell_h) \cdot T(f_h, j_h; k_h, \ell_h)$$

where different letters represent different numbers.

Therefore, μ_{2h+1} is a polynomial whose largest term is of degree 3h+1 in m and 2h in n. Hence

$$(7) \qquad \frac{\mu_{2h+1}}{\mu_2^{\frac{1}{2}(2h+1)}} = 0\left(\frac{m^{3h+1}\, n^{2h}}{m^{\frac{3}{2}(2h+1)}\, n^{2h+1}}\right) = 0\left(\frac{1}{m^{1/2}\, n}\right) \to 0 .$$

In considering μ_{2h} the same type of argument may be used to assert that its highest ordered terms arise from products of the following type:

$$\prod_{s=1}^{h} T(i_s, j_s; k_s, \ell_s) \cdot T(g_s, j_s; k_s, \ell_s) .$$

https://doi.org/10.4153/CMB-1962-002-0 Published online by Cambridge University Press

The number of times terms of this type occur is

$$(2h)!\binom{m}{h}\binom{m-h}{2}\binom{n}{2}\cdots\binom{m-3h+2}{2}\binom{n-2h+2}{2}\sim\frac{m^{3h}n^{2h}(2h)!}{2^{2h}\,h!},$$

retaining only terms of highest order.

The expectation of such a term equals

$$\{E[T(i,j;k,\ell)\cdot T(g,j;k,\ell)]\}^{h}=1/2^{6h},$$

as may be seen by direct considerations. Thus the leading term of μ_{2h} is $\frac{m^{3h}n^{2h}(2h)!}{2^{8h}\,h!}$ and that of μ_2 is $\frac{m^3n^2}{2^7}$.

It follows that

$$\text{(8)}\qquad \frac{\mu_{2h}}{\mu_2^{\,h}}\to\frac{(2h)!}{2^h\,h!}\text{ as required.}$$

For each fixed value of h the only essential difference in the arguments when $n\to cm$, $0<c\le 1$, as $m,n\to\infty$, will be the appearance of an additional factor of $(1+c)^h$ in both μ_{2h} and $\mu_2^{\,h}$, which, obviously, leaves (8) unchanged.

This proves, under the given conditions,

THEOREM 2:

$$\Pr\left\{a<\frac{C(m,n)-\frac{1}{8}\binom{m}{2}\binom{n}{2}}{\frac{1}{8}\left[\binom{m}{2}\binom{n}{2}(2m+2n-1)\right]^{\frac{1}{2}}}<b\right\}\to\frac{1}{\sqrt{2\pi}}\int_a^b e^{-y^2/2}\,dy\,.$$

https://doi.org/10.4153/CMB-1962-002-0 Published online by Cambridge University Press

REFERENCES

1. M. G. Kendall, "Rank Correlation Methods," Charles Griffin and Co., London, (1948).

2. M. G. Kendall and A. S. Stuart, "The Advanced Theory of Statistics," Charles Griffin and Co., London, 1 (1958) 111.

3. P. A. P. Moran, "On the Method of Paired Comparisons," Biometrika, 34(1947), 363-365.