The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 242–245

A TREE COUNTING PROBLEM[1]

By J. W. Moon

University of Alberta

1. Introduction. If n, k, and m are fixed integers satisfying the inequalities $1 \leqq k \leqq n$ and $0 \leqq m \leqq n - k$, let $p = (p_1, p_2, \cdots, p_k)$ and $e = (e_1, e_2, \cdots, e_k)$ denote partitions of n and m into k integers with the property that $0 \leqq e_i \leqq p_i - 1$ for $i = 1, 2, \cdots, k$. Suppose that n labelled points are split into k classes with p_i points in the ith class and that a tree T_n is formed on these n points such that e_i edges join points in the ith class to each other. (Since the tree has $n - 1$ edges, it follows that there will be $n - 1 - m$ edges of the tree joining points in different classes. For definitions not given here, see [13].) Na and Rapoport [12], in their study of random graphs, recently derived the following formula for $T(p, e)$, the number of labelled trees T_n that can be formed in this way.

Theorem. $T(p, e) = n^{k-2} \prod_{i=1}^{k} \binom{p_i - 1}{e_i} p_i^{e_i} (n - p_i)^{p_i - 1 - e_i}$.

Their proof was based on a result due to Kirchoff [8] that has been rediscovered by several people. Kirchoff's theorem states that the number of trees contained as subgraphs of a given graph G is equal to the determinant of a certain matrix whose entries depend on G. Certain classes of graphs have a simple enough structure for the determinant to be evaluated explicitly and several formulas for counting trees with different properties have been obtained in this way (see [10]).

A different approach is taken in [11]; it is shown that the various formulas derived by evaluating these determinants may also be derived, frequently with less effort, by direct and straightforward arguments based on little more than properties of multinomial coefficients and the method of inclusion and exclusion. Our object here is to give a derivation of the formula for $T(p, e)$ based on this approach.

2. Two lemmas. A *forest* is a graph that has no cycles; its connected components are trees. Let $j_1, j_2, \cdots, j_l$ denote the number of points in the l trees making up a given forest F_n with $n = j_1 + j_2 + \cdots + j_l$ labelled points. The following result is proved in [9] and [11].

Lemma 1. *The number of trees T_n with n labelled points that contain a given forest F_n is equal to $j_1 j_2 \cdots j_l n^{l-2}$.*

Notice that if $l = n$ and $j_1 = \cdots = j_n = 1$, then Lemma 1 reduces to Cayley's formula, n^{n-2}, for the total number of trees T_n.

The following identity is essentially a multionomial extension due to Hurwitz of Abel's binomial identity; an elementary proof has been given by Riordan [15] recently and Rényi [14] has also given a proof.

Received 13 May 1967.

[1] This paper was prepared during the Summer Institute of the Canadian Mathematical Congress at Dalhousie University, Halifax.

LEMMA 2.

$$\sum (p!/j_1! \cdots j_l!) j_1^{j_1-1} \cdots j_l^{j_l-1} = l! \binom{p-1}{l-1} p^{p-l},$$

where the sum is over all solutions in positive integers to the equation $j_1 + j_2 + \cdots + j_l = p$.

3. Proof of theorem. Suppose the tree T_n is such that h_i of its edges join points belonging to the ith class of p_i points. These h_i edges determine a forest F_{p_i} consisting of $l_i = p_i - h_i$ trees. The number of ways of forming a forest F_{p_i} of l_i trees is

$$(l_i!)^{-1} \sum (p_i!/j_1! \cdots j_{l_i}!) j_1^{j_1-2} \cdots j_{l_i}^{j_{l_i}-2},$$

where the sum is over all solutions in positive integers to the equation $j_1 + j_2 + \cdots + j_{l_i} = p_i$. This may be seen by considering the number of ways of splitting the p_i points into l_i subsets and then forming a tree on the points of each subset.

If we now apply Lemma 1 to each forest $F_n = \bigcup_{i=1}^k F_{p_i}$, we conclude that the number of trees T_n containing (at least) h_i edges joining points of the ith class of points is equal to

$$n^{\Sigma l_i - 2} \prod \{(l_i!)^{-1} \sum (p_i!/j_1! \cdots j_{l_i}!) j_1^{j_1-1} \cdots j_{l_i}^{j_{l_i}-1}\},$$

where (here and elsewhere) i varies from 1 to k in the product and the first sum. (If there are t admissible forests F_n that are contained in a given tree T_n, then the tree T_n will be counted t times in this expression.) With the aid of Lemma 2 this expression simplifies to

$$n^{k-2} \prod \binom{p_i-1}{h_i} p_i^{h_i} n^{(p_i-1)-h_i}.$$

It now follows by the method of inclusion and exclusion (see, for example, [6], p. 96) that the number of trees T_n containing exactly e_i edges joining points in the ith class is given by the formula

$$T(p, e) = n^{k-2} \sum \{\prod (-1)^{h_i-e_i} \binom{h_i}{e_i} \binom{p_i-1}{h_i} p_i^{h_i} n^{(p_i-1)-h_i}\},$$

where the sum is over $h_1, h_2, \cdots, h_k$ such that $h_i = e_i, e_i + 1, \cdots, p_i - 1$ for each i. If we let $t_i = h_i - e_i$ then t_i runs from 0 to $p_i - 1 - e_i$ and the formula may be rewritten as

$$\begin{aligned} T(p, e) &= n^{k-2} \prod \binom{p_i-1}{e_i} p_i^{e_i} \{ \sum \binom{p_i-1-e_i}{t_i} (-p_i)^{t_i} n^{(p_i-1-e_i)-t_i} \} \\ &= n^{k-2} \prod \binom{p_i-1}{e_i} p_i^{e_i} (n - p_i)^{p_i-1-e_i}. \end{aligned}$$

This completes the proof of the theorem.

4. Special cases. We conclude by mentioning a few special cases and easy consequences of the theorem that have been proved separately.

Let $T(n; p_1, p_2, \cdots, p_k)$ denote the number of trees T_n that are contained in a graph G_n the complement of which consists of k disjoint complete graphs having $p_1, p_2, \cdots, p_k$ points, respectively. The formula

$$(1) \qquad T(n; p_1, p_2, \cdots, p_k) = n^{k-2} \prod (n - p_i)^{p_i - 1}$$

is obtained by setting $e_1 = \cdots = e_k = 0$ in the theorem. This formula was first proved by Austin [1].

When $k = 2$ and $p_1 = a$ and $p_2 = b$, where $a + b = n$, then formula (1) reduces to

$$(2) \qquad T(n; a, b) = a^{b-1}b^{a-1}.$$

This formula was also proved by Scoins [16], Glicksman [7], and others.

Let $W(n, j)$ denote the number of trees T_n contained in a graph G_n the complement of which consists of j disjoint edges and $n - 2j$ isolated points. Weinberg [17] proved that

$$(3) \qquad W(n, j) = n^{n-2}(1 - 2/n)^j.$$

This formula may be deduced from formula (1) by setting $k = n - j$ and $p_1 = \cdots = p_j = 2$ and $p_{j+1} = \cdots = p_{n-j} = 1$.

Let $F(n, l)$ denote the number of forests F_n of l trees such that l given points all belong to different trees of the forest. If we let $k = n - l + 1$, $p = (l, 1, \cdots, 1)$, and $e = (l - 1, 0, \cdots, 0)$, then it is not difficult to see that

$$l^{l-2}F(n, l) = T(p, e) = n^{n-l-1}l^{l-1}$$

or

$$(4) \qquad F(n, l) = ln^{n-l-1}.$$

This formula was first stated by Cayley [4]; Rényi [14] gave a proof based on the identity in Lemma 2.

Let $C(n, t)$ denote the number of trees T_n in which a given point, say the first, is joined to exactly t other points. It is easy to show (see Rényi [14]) that

$$C(n, t) = \binom{n-1}{t}F(n - 1, t);$$

hence

$$(5) \qquad C(n, t) = \binom{n-2}{t-1}(n - 1)^{n-1-t}.$$

This formula was first proved by Clarke [5] (see also de Bruijn [3] and Bedrosian [2]).

Other results can be derived in a similar fashion. We leave it as an exercise for the reader to show that Lemma 1 may also be deduced as a direct consequence of the theorem.

REFERENCES

[1] Austin, T. (1960). The enumeration of labelled chromatic graphs and trees. *Canad. J. Math.* **12** 535–545.

[2] Bedrosian, S. D. (1964). Trees of a full graph as an occupancy problem. *I.E.E.E. Trans. on Circuit Theory.* **CT-11** 290–291.

[3] de Bruijn, N. G. (1964). Problem 16. *Nieuw Archief voor Wiskunde* **12** 69.

[4] Cayley, A. (1889). A theorem on trees. *Quart. J. Math.* **23** 376–378.

[5] Clarke, L. E. (1958). On Cayley's formula for counting trees. *J. London Math. Soc.* **33** 471–475.

[6] Feller W. (1957). *An Introduction to Probability Theory and Its Applications.* Wiley, New York.

[7] Glicksman, S. (1963). On the representation and enumeration of trees. *Proc. Cambridge Phil. Soc.* **59** 509–517.

[8] Kirchoff, G. (1847). Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. *Ann. Phys. Chem.* **72** 497–508.

[9] Moon, J. W. (1964). On the second moment of the complexity of a graph. *Mathematika* **11** 95–98.

[10] Moon, J. W. (1967). Various proofs of Cayley's formula for counting trees. *A Seminar on Graph Theory.* F. Harary, ed. Holt, Boston, 70–78.

[11] Moon, J. W. (1967). Enumerating labelled trees. To appear in *Graph Theory and Theoretical Physics.* F. Harary, ed. Academic Press, London.

[12] Na, H. S. and Rapoport, A. (1967). A formula for the probability of obtaining a tree from a graph constructed randomly except for an "exogamous bias". *Ann. Math. Statist.* **38** 226–241.

[13] Ore, O. (1962). *Theory of Graphs.* A.M.S. Colloq. Publ. **38,** Providence.

[14] Rényi A. (1959). Some remarks on the theory of trees. *Publ. Math. Inst. Hung. Acad. Sci.* **4** 73–85.

[15] Riordan, J. (1967). Abel identities and inverse relations. Conference on Combinatorial Mathematics and its Applications, Chapel Hill.

[16] Scoins, H. I. (1962). The number of trees with nodes of alternate parity. *Proc. Cambridge Phil. Soc.* **58** 12–16.

[17] Weinberg, L. (1958). Number of trees in a graph. *Proc. I.R.E.* **49** 1954–1955.