CHAPTER 3

INVERSION FORMULAS AND THEIR APPLICATIONS

1. Differential Operator Associated with a Family of Polynomials

Consider a family

$$P_0(x),\ P_1(x),\ P_2(x),\ \ldots;$$

of polynomials in a real variable x, where $P_0(x) = 1$ and $P_n(0) = 0$, $n \geqslant 1$. Such a family is called a *normal family* of polynomials.

A *differential operator* D, *associated with the polynomials* $P_n(x)$, maps each polynomial $\varphi(x)$ onto a polynomial denoted by $\mathrm{D}\varphi(x)$, and in addition satisfies

(1) $\mathrm{D}P_n(x) = \begin{cases} nP_{n-1}(x) & \text{if } n \neq 0, \\ 0 & \text{if } n = 0. \end{cases}$

(2) $\mathrm{D}[\lambda\varphi(x)] = \lambda\, \mathrm{D}\varphi(x), \qquad \lambda = \text{constant}.$

(3) $\mathrm{D}[\varphi(x) + \varphi'(x)] = \mathrm{D}\varphi(x) + \mathrm{D}\varphi'(x).$

Property 1. *For each normal family P_n of polynomials, there exists one and only one differential operator.*

Proof: We show that every polynomial $\varphi_n(x)$ of degree n can be expressed uniquely in the form

$$\varphi_n(x) = \alpha_n P_n(x) + \alpha_{n-1}P_{n-1}(x) + \cdots + \alpha_0 P_0(x),$$

where α_n, α_{n-1}, $\ldots$, α_0 are constants. To prove this take for α_n the coefficient of x^n in $\varphi_n(x)$ divided by the coefficient of x^n in $P_n(x)$; $\varphi_{n-1}(x) = \varphi_n(x) - \alpha_n P_n(x)$ is then, at most, of degree $n-1$, and for α_{n-1}

take the coefficient of x^{n-1} in $\varphi_{n-1}(x)$, divided by the coefficient of x^{n-1} in $P_{n-1}(x)$. Next consider

$$\varphi_{n-2}(x) = \varphi_{n-1}(x) - \alpha_{n-1}P_{n-1}(x),$$

which determines α_{n-2}, etc.

From (1), (2), and (3),

$$\mathrm{D}\varphi_n(x) = n\alpha_n P_{n-1}(x) + (n-1)\,\alpha_{n-1}P_{n-2}(x) + \cdots + 0.$$

This proves the uniqueness of the differential operator associated with the polynomials $P_n(x)$.

Moreover, the operator D, defined by this equality, satisfies (1), (2), and (3).

THEOREM (Taylor). *Let $P_n(x)$ be a normal family of polynomials. If φ is a polynomial in x of degree n, then*

$$\varphi(x) = \varphi(0) + \frac{\mathrm{D}\varphi(0)}{1!}P_1(x) + \frac{\mathrm{D}^2\varphi(0)}{2!}P_2(x) + \cdots + \frac{\mathrm{D}^n\varphi(0)}{n!}P_n(x).$$

Moreover, this series expansion of $\varphi(x)$ is unique.

PROOF: We have

$$\varphi(x) = \alpha_0 P_0(x) + \alpha_1 P_1(x) + \cdots + \alpha_n P_n(x).$$

Putting $x = 0$, this becomes

$$\varphi(0) = \alpha_0.$$

Differentiating,

$$\mathrm{D}\varphi(x) = 0 + \alpha_1 P_0(x) + 2\,\alpha_2 P_1(x) + \cdots + n\alpha_n P_{n-1}(x).$$

Putting $x = 0$, we obtain

$$\mathrm{D}\varphi(0) = \alpha_1.$$

Differentiating again,

$$\mathrm{D}^2\varphi(x) = 2\,\alpha_2 P_0(x) + 2\cdot 3\,\alpha_3 P_4(x) + \cdots + n(n-1)\alpha_n P_{n-2}(x).$$

Hence,

$$\mathrm{D}^2\varphi(0) = 2\,\alpha_2.$$

More generally, let

$$1 \times 2 \times 3 \times \cdots \times k = k!;$$

then

$$D^k\varphi(0) = k!\,\alpha_k.$$

This completes the proof.

Notice that the differential operator, associated with the normal family $P_n(x) = x^n$ of polynomials, is the ordinary differential d/dx, and the formula in the above theorem becomes the standard Taylor–Maclaurin expansion.

Binomial Formula. Let y be a constant, and consider the polynomial

$$\varphi(x) = (x + y)^n.$$

By substituting the polynomials $P_n(x) = x^n$ ($P_n(0) = 0$, $n \geqslant 1$) in Taylor's formula, $\varphi(x)$ may be expanded as a Taylor series.

The differential operator, associated with the polynomials P_n, is the ordinary differential

$$D\varphi(x) = d\varphi/dx.$$

Then

$$\begin{aligned} D\varphi(x) &= n(x + y)^{n-1}, \\ D^2\varphi(x) &= n(n-1)(x + y)^{n-2}, \\ D^k\varphi(x) &= n(n-1)\cdots(n-k+1)(x+y)^{n-k}. \end{aligned}$$

Taylor's formula becomes

$$(x + y)^n = y^n + \binom{n}{1}xy^{n-1} + \binom{n}{2}x^2y^{n-2} + \cdots + \binom{n}{n}x^n.$$

This is the "binomial formula" which may be written

$$\boxed{(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}}$$

Δ Binomial Formula. Let us consider Taylor's formula relative to the polynomials

$$P_n(x) = [x]_n = x(x-1)(x-2)\cdots(x-n+1) \quad (P_n(0)=0,\ n \geqslant 1).$$

The operator Δ, defined by

$$\Delta\varphi(x) = \varphi(x+1) - \varphi(x),$$

is the differential operator associated with the polynomials P_n, since

$$\begin{aligned}\Delta[x]_n &= [x+1]_n - [x]_n \\ &= (x+1)[x]_{n-1} - (x-n+1)[x]_{n-1} = n[x]_{n-1}.\end{aligned}$$

Using Taylor's formula, relative to the polynomials $[x]_n$, to expand the polynomial

$$\varphi(x) = [x+y]_n,$$

and observing that

$$\Delta^k\varphi(x) = n(n-1)\cdots(n-k+1)[x+y]_{n-k},$$

we have

$$\boxed{[x+y]_n = \sum_{k=0}^{n} \binom{n}{k} [x]_k [y]_{n-k}} \qquad \text{(Vandermonde's formula)}$$

∇ Binomial Formula. Let us consider Taylor's formula relative to the polynomials

$$P_n(x) = [x]^n = x(x+1)(x+2)\cdots(x+n-1) \quad (P_n(0)=0,\ n \geqslant 1).$$

The operator ∇, defined by

$$\nabla\varphi(x) = \varphi(x) - \varphi(x-1),$$

is the differential operator associated with the polynomials P_n, since

$$\begin{aligned}\nabla[x]^n &= [x]^n - [x-1]^n \\ &= (x+n-1)[x]^{n-1} - (x-1)[x]^{n-1} = n[x]^{n-1}.\end{aligned}$$

Expanding the polynomial

$$\varphi(x) = [x+y]^n,$$

and observing that

$$\nabla^k \varphi(x) = n(n-1)\cdots(n-k+1)[x+y]^{n-k},$$

we obtain

$$\boxed{[x+y]^n = \sum_{k=0}^{n} \binom{n}{k} [x]^k [y]^{n-k}} \qquad \text{(Nörlund's formula)}$$

First Inversion Theorem. *Let $\varphi_n(x)$ and $\psi_n(x)$ be families of polynomials of degree n satisfying*

$$\varphi_n(x) = \sum_{k=0}^{n} \alpha_n{}^k \psi_k(x) \qquad (n = 0, 1, 2, \ldots, n_0),$$

$$\psi_n(x) = \sum_{k=0}^{n} \beta_n{}^k \varphi_k(x) \qquad (n = 0, 1, 2, \ldots, n_0).$$

If $a_0, a_1, a_2, \ldots, a_{n_0}, b_0, b_1, \ldots, b_{n_0}$ are numbers satisfying

$$a_n = \sum_{k=0}^{n} \alpha_n{}^k b_k \qquad (n = 0, 1, 2, \ldots, n_0),$$

then

$$b_n = \sum_{k=0}^{n} \beta_n{}^k a_k \qquad (n = 0, 1, 2, \ldots, n_0).$$

Proof: Clearly,

$$\varphi_n(x) = \sum_{k=0}^{n} \alpha_n{}^k \sum_{m=0}^{n} \beta_k{}^m \varphi_m(x) = \sum_{m=0}^{n} \left(\sum_{k=0}^{n} \alpha_n{}^k \beta_k{}^m \right) \varphi_m(x).$$

Writing $\alpha_n{}^m = 0$, if $m > n$; $\beta_n{}^m = 0$, if $m > n$; $\delta_n{}^m = 0$, if $m \neq n$, $= 1$ if $m = n$; and comparing the coefficients of $\varphi_m(x)$ in the above equation,

$$\delta_n{}^m = \sum_{k=0}^{n_0} \alpha_n{}^k \beta_k{}^m.$$

In other words, the matrices $((\alpha_j{}^i))$ and $((\beta_j{}^i))$ are inverses of one another; therefore, the vector equation

$$\mathbf{a} = ((\alpha_j{}^i))\mathbf{b}$$

is equivalent to

$$\mathbf{b} = ((\beta_j{}^i))\mathbf{a}.$$

REMARK: Let $P_n(x)$ and $Q_n(x)$ be families of polynomials that vanish at $x = 0$, $n \neq 0$, and let Δ and D be, respectively, their associated differential operators. What we have just shown is that the matrices

$$\left(\begin{pmatrix} Q_0(0) & 0 & 0 & \cdots \\ Q_1(0) & \dfrac{\Delta Q_1(0)}{1!} & 0 & \cdots \\ Q_2(0) & \dfrac{\Delta Q_2(0)}{1!} & \dfrac{\Delta^2 Q_2(0)}{2!} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}\right)$$

and

$$\left(\begin{pmatrix} P_0(0) & 0 & 0 & \cdots \\ P_1(0) & \dfrac{\mathrm{D}P_1(0)}{1!} & 0 & \cdots \\ P_2(0) & \dfrac{\mathrm{D}P_2(0)}{1!} & \dfrac{\mathrm{D}^2 P_2(0)}{2!} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}\right)$$

are inverses of one another.

INVERSE BINOMIAL FORMULAS. Putting $x = y + 1$, and using the binomial formula,

$$x^n = (y+1)^n = \sum_{k=0}^{n} \binom{n}{k} (x-1)^k.$$

Again, using the binomial formula,

$$(x-1)^n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} x^k.$$

By the first inversion theorem, if the numbers $a_0, a_1, a_2, \ldots, b_0, b_1, b_2, \ldots$ satisfy

$$a_n = \sum_{k=0}^{n} \binom{n}{k} b_k,$$

then

$$b_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} a_k.$$

In particular (Formula 1 of Chapter 1, Section 10)

$$n^p = \sum_{k=0}^{n} \binom{n}{k} (k! S_p^{\,k}) \qquad \text{(assuming } S_p^{\,0} = 0\text{).}$$

Therefore,

$$(n! S_p^{\,n}) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} k^p$$

or

$$\boxed{S_n^{\,m} = \frac{1}{m!} \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} k^n} \qquad \text{(Stirling's formula)}$$

STIRLING'S INVERSE FORMULAS. We recall (Chapter 1, Section 5) that the Stirling numbers $s_n^{\,k}$ of the first kind are defined by

$$[x]_n = \sum_{k=1}^{n} s_n^{\,k} x^k,$$

and (Chapter 1, Section 10)

$$x^n = \sum_{k=1}^{n} S_n^{\,k} [x]_k.$$

By the first inversion theorem, if numbers $a_1, a_2, \ldots, b_1, b_2, \ldots$ satisfy

$$a_n = \sum_{k=1}^{n} s_n^{\,k} b_k,$$

then

$$b_n = \sum_{k=1}^{n} S_n^{\,k} a_k.$$

LAH'S INVERSE FORMULAS. Since $[-x]_n$ is a polynomial of degree n, it has a series expansion

$$[-x]_n = L_n{}^1[x]_1 + \cdots + L_n{}^n[x]_n.$$

The $L_n{}^k$'s are called *Lah numbers*; changing x into $-x$,

$$[x]_n = L_n{}^1[-x]_1 + \cdots + L_n{}^n[-x]_n.$$

Then, applying the first inversion theorem,

$$a_n = \sum_{k=1}^{n} L_n{}^k b_k \Leftrightarrow b_n = \sum_{k=1}^{n} L_n{}^k a_k.$$

2. THE MÖBIUS FUNCTION

We will now generalize the first inversion theorem. The following description of the Möbius function is due essentially to Rota [15], who is convinced that the Möbius function is a fundamental unifying principle of enumeration.

Let X be a set on which is defined an order relation $\leqslant$; then, by definition

$$x \leqslant x,$$
$$x \leqslant y,\ y \leqslant x \Rightarrow x = y,$$
$$x \leqslant y,\ y \leqslant z \Rightarrow x \leqslant z.$$

X is assumed to have a unique minimal element, denoted by 0; i.e.,

$$0 \leqslant x \qquad (x \in X)$$

(if X does not contain such an element, then it can always be adjoined to X).

Finally, suppose that for all $x, y \in X$, the *segment*

$$[x, y] = \{u/u \in X, u \geqslant x, u \leqslant y\}$$

is finite. Such a set X, together with its order relation $\leqslant$, is called a *locally finite ordered set*.

For example, X might be the set of nonnegative integers with the order relation

$$x \leqslant y \Leftrightarrow x \text{ is less than or equal to } y,$$

or X might be the set of integers $\geqslant 1$ with the order relation

$$x \leqslant y \Leftrightarrow x \text{ is a divisor of } y.$$

Later, we shall come across many other examples.

As usual, write $x < y$, if $x \leqslant y$ and $x \neq y$; write $x \geqslant y$, if $y \leqslant x$. Let X be a locally finite ordered set. Let A be the set of all real-valued functions $f(x, y)$ of two variables, defined for x and y ranging over X, such that

$$\begin{aligned} f(x, y) &\neq 0 \qquad \text{if} \quad x = y, \\ f(x, y) &= 0 \qquad \text{if} \quad x \nleqslant y. \end{aligned}$$

A product $*$ is defined on A by

$$f * g(x, y) = \sum_{x \leqslant u \leqslant y} f(x, u) g(u, y).$$

The summation is over all u in the finite segment $[x, y]$.

The set A, together with the product $*$, is called the *group of arithmetic functions*. We now prove that it is, indeed, a "group."

PROPOSITION 1. *The product $*$ is associative, i.e.,*

$$(f * g) * h = f * (g * h).$$

PROOF: By definition

$$\begin{aligned} [(f * g) * h](x, z) &= \sum_{\substack{y \\ x \leqslant y \leqslant z}} h(y, z) \sum_{\substack{u \\ x \leqslant u \leqslant y}} f(x, u) g(u, y) \\ &= \sum_{\substack{u, y \\ x \leqslant u \leqslant y \leqslant z}} f(x, u) g(u, y) h(y, z) \\ &= [f * (g * h)](x, z). \end{aligned}$$

PROPOSITION 2. *The Kronecker function*

$$\delta(x, y) = \begin{cases} 1 & \text{if} \quad x = y, \\ 0 & \text{if} \quad x \neq y, \end{cases}$$

is the identity element for $*$.

PROOF: By definition

$$[f * \delta](x, y) = \sum_{x \leq u \leq y} f(x, u)\, \delta(u, y) = f(x, y).$$

PROPOSITION 3. *For each $f \in A$, there exists a left-inverse $f^{-1} \in A$ (that is, $f^{-1}f = \delta$); for a given x, $f^{-1}(x, y)$ is defined by induction on y as follows:*

(1) *if* $y = x$: $\quad f^{-1}(x, y) = \dfrac{1}{f(x, x)}$;

(2) *if* $y > x$: $\quad f^{-1}(x, y) = \dfrac{-1}{f(y, y)} \sum_{x \leq u < y} f^{-1}(x, u) f(u, y)$.

PROOF: From (1),

$$f^{-1} * f(x, x) = f^{-1}(x, x) f(x, x) = 1.$$

If $x < y$, from (2),

$$f^{-1} * f(x, y) = \sum_{x \leq u < y} f^{-1}(x, u) f(u, y) + f^{-1}(x, y) f(y, y) = 0.$$

Therefore, $f^{-1} * f = \delta$.

Now if each element α of a monoid has a left inverse α^{-1}, then α^{-1} is also the right inverse of α, since

$$\alpha\alpha^{-1} = f = \delta f = f^{-1}ff = f^{-1}\alpha\alpha^{-1}\alpha\alpha^{-1} = f^{-1}\alpha\, \delta\alpha^{-1} = f^{-1}f = \delta.$$

Therefore, Propositions 1, 2, and 3 prove that A is a group.

Furthermore, A is a ring (see Smith [18]; Carlitz [1]) possessing many remarkable properties.

The most important of these properties that concern us here is the following:

$$f = g * \alpha \Rightarrow g = f * \alpha^{-1}.$$

That is, to every function $\alpha(x, y)$ satisfying $\alpha(x, x) \neq 0$, $\alpha(x, y) = 0$ if $x \nleq y$; there corresponds a function $\beta(x, y)$ with the same properties, and such that

$$f(0, x) = \sum_{0 \leq u \leq x} \alpha(u, x) g(0, u)$$

implies

$$g(0, x) = \sum_{0 \leqslant u \leqslant x} \beta(u, x) f(0, u).$$

X being a locally finite set, the function

$$\xi(x, y) = \begin{cases} 1 & \text{if} \quad x \leqslant y, \\ 0 & \text{otherwise,} \end{cases}$$

is called the *Riemann function.*

The *Möbius function* μ is defined inductively (for all $y \geqslant x$) by

$$\mu(x, x) = 1,$$
$$\mu(x, y) = -\sum_{x \leqslant t < y} \mu(x, t).$$

The functions ξ and μ are inverses of each other and, therefore, the following is a consequence of the first inversion theorem.

THEOREM (Möbius inversion [8]). *Let X be a locally finite ordered set and let $f(x)$, $g(x)$ be functions defined on X, with*

$$f(x) = \sum_{0 \leqslant u \leqslant x} g(u) \qquad (x \in X).$$

Then

$$g(x) = \sum_{0 \leqslant u \leqslant x} \mu(u, x) f(u) \qquad (x \in X).$$

EXAMPLE 1. Let X be the set of positive integers with the order relation $\leqslant$ defined by $k \leqslant n$ if "k is less than or equal to n."

Let us invert

$$f(n) = \sum_{k=1}^{n} g(k).$$

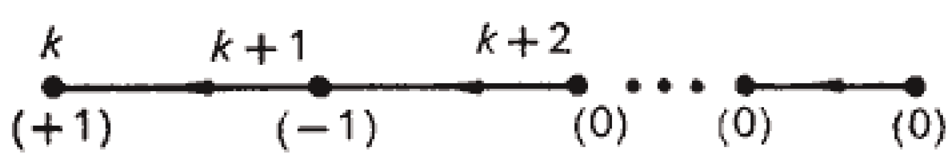

FIG. 1

For all $n \geqslant k$, the Möbius function is defined by

$$\mu(k, n) = \begin{cases} 1 & \text{if} \quad n = k \\ -1 & \text{if} \quad n = k+1 \\ 0 & \text{if} \quad n = k+2, k+3, \ldots. \end{cases}$$

Therefore,

$$g(n) = f(n) - f(n-1).$$

EXAMPLE 2. Let X be the set of positive integers, with the order relation $\leqslant$ defined by

$$y \leqslant x \qquad \text{if} \quad y \text{ divides } x \qquad (y/x).$$

Let us find the inverse of

$$f(n) = \sum_{d/n} g(d).$$

The graph of Fig. 2 illustrates the method of finding the Möbius function $\mu(d, n)$.

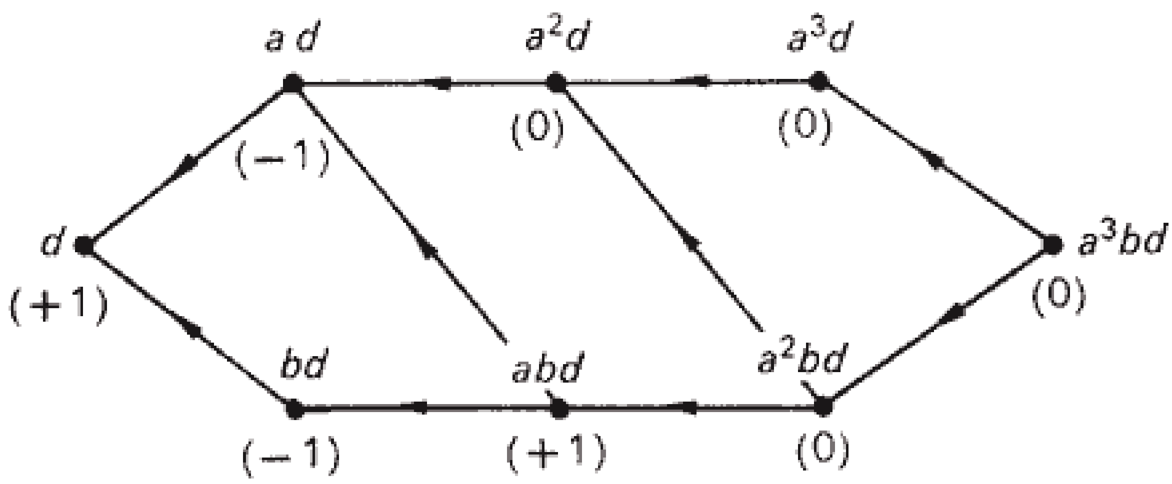

FIG. 2

It follows immediately that

$$\mu(d, n) = \begin{cases} 1 & \text{if} \quad n = d, \\ (-1)^k & \text{if} \quad n = p_1 p_2 \cdots p_k d, \\ 0 & \text{otherwise,} \end{cases}$$

where the p_i's are distinct prime numbers $\neq 1$. Therefore,

$$g(n) = \sum_{d/n} \mu(d, n) f(d).$$

This function $\mu(d, n)$—usually written as $\mu(d/n)$—is the classical function first derived by Möbius in 1832 in order to study the distribution of the prime numbers.

EXAMPLE 3. Let A be a finite set. Let us find the inverse of

$$f(A) = \sum_{S \subset A} g(S).$$

In this example, X is the lattice of subsets of A, with the order relation $\subset$ (inclusion).

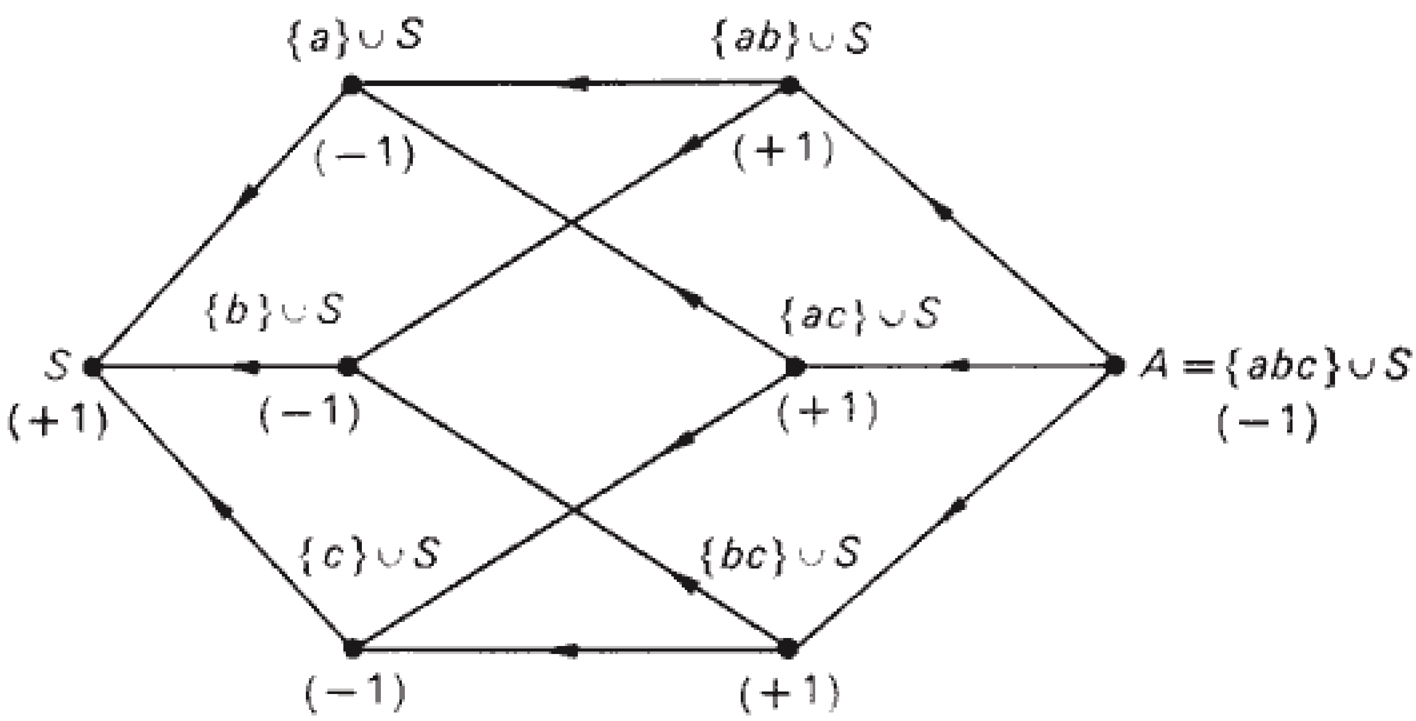

FIG. 3

Clearly,

$$\mu(S, A) = (-1)^{|A| - |S|},$$

whence,

$$g(A) = \sum_{S \subset A} (-1)^{|A| - |S|} f(S).$$

EXAMPLE 4. Let A be a set, and let

$$\mathscr{A} = (A_1, A_2, \ldots, A_k)$$

be a partition of A; then, by definition,

$$A_i \neq \varnothing,$$

$$i \neq j \Rightarrow A_i \cap A_j = \varnothing,$$

$$\bigcup A_i = A.$$

Write $\mathscr{B} < \mathscr{A}$ ($\mathscr{B}$ is a "subpartition" of $\mathscr{A}$) if

$$\left.\begin{array}{l} B_j \cap A_i \neq \varnothing \\ B_j \in \mathscr{B} \\ A_i \in \mathscr{A} \end{array}\right\} \Rightarrow B_j \subset A_i .$$

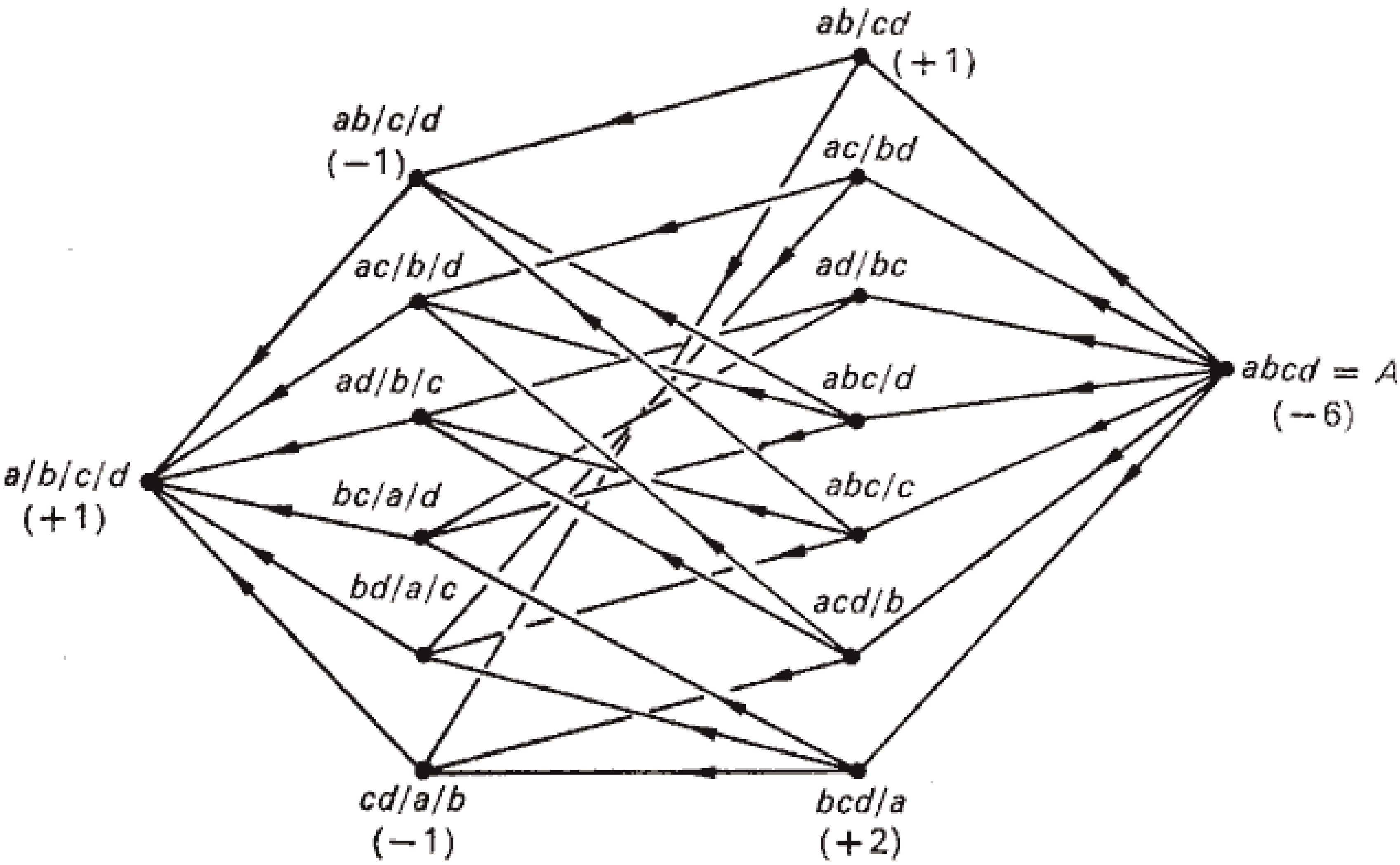

FIG. 4

If $\mathscr{A}$ is a partition of A into p classes $A_1, A_2, \ldots, A_p$, and if $\mathscr{B}$ is a subpartition of $\mathscr{A}$, such that each A_i contains n_i classes of $\mathscr{B}$, then

$$\mu(\mathscr{B}, \mathscr{A}) = (-1)^{p+n_1+n_2+\cdots+n_p}(n_1-1)!(n_2-1)!\cdots(n_p-1)!.$$

This formula was first derived by Schützenberger [16] and, with a view to different applications, has since been reformulated by Frucht [23] and Rota [15].

APPLICATION: *The circular word problem* (Moreau [11]). An *alphabet* is a set of m distinct symbols $a_1, a_2, \ldots, a_m$ called *letters*; a *word of length n* is a mapping φ of $X = \{1, 2, \ldots, n\}$ into

$$A = \{a_1, a_2, \ldots, a_m\}.$$

Two words φ and φ' are said to be *equivalent*, or the same "circular word," if

$$\varphi'(i) \equiv \varphi(i+p)(\operatorname{mod} n) \qquad (i = 1, 2, \ldots, n).$$

Let us now try to count the number of circular words of n letters.

p is said to be a *period* of a word φ if $\varphi(i+p) \equiv \varphi(i) \pmod n$ for each i; the *primitive period* of φ is defined to be the smallest period p ($1 \leqslant p \leqslant n$). For example, the primitive period of *abcabcabc* is $p = 3$.

Obviously, a period p must be a divisor of n. Let $M(p)$ be the number of circular words of primitive period p. Then, since to each such word there corresponds p different words, the total number of words is

$$m^n = \sum_{p/n} pM(p).$$

Inverting this formula, using the Möbius function $\mu(d, n)$ of Example 2,

$$\mu(d, n) = \begin{cases} (-1)^k & \text{if} \quad n = p_1 p_2 \ldots p_k d, \\ 0 & \text{otherwise}, \end{cases}$$

where the p_i's are distinct prime numbers $\neq 1$, we obtain

$$pM(p) = \sum_{q/p} \mu(q, p) m^q.$$

Therefore, the total number of circular words of length n is*

$$\sum_{p/n} M(p) = \sum_{p/n} \frac{1}{p} \sum_{q/p} \mu(q, p) m^q.$$

The problem in information theory of the "comma-free" dictionaries is a generalization of this well-known problem. A *comma-free dictionary* is a set of words, each of n letters, satisfying the property that, for any two words of the dictionary, no integer k ($1 \leqslant k < n$) exists such that the last $n - k$ letters of the first word, followed by the first k letters of the second word, form a word in the dictionary. The problem is to find a comma-free dictionary with the maximum number of words.

There are several theorems that facilitate the calculation of the Möbius function in special cases, such as:

* Pólya's theorem, which is proved directly in Chapter 5, can be proved by an extension of this same formula; see Rota, "Enumeration under Group Action," which appeared in the *Journal of Combinatorial Theory*, 1969.

THEOREM (P. Hall). *Let the ordered set* $(X, \leqslant)$ *be a semilattice—that is, there exists for all* $a, b \in X$ *a least upper bound* c *(called the union of* a *and* b *and denoted by* $a \vee b$*) such that*

$$c \geqslant a, b,$$

$$x \geqslant a, b \Rightarrow x \geqslant c.$$

If $x > y$*, and* x *is not the union of immediate predecessors of* y*, then*

$$\mu(y, x) = 0.$$

PROOF: Assume that x is not the union of immediate predecessors of y, that is, of elements of the set

$$\{z/z \in X, z \geqslant y, \quad \text{there exists no } t \text{ such that } z > t > y\}.$$

Let $a_1, a_2, \ldots, a_k$ be the immediate predecessors of y, which are $\leqslant x$, and let $b = a_1 \vee a_2 \vee \cdots \vee a_k$, then

$y \leqslant b \leqslant x$ (since $a, a' \leqslant x$ implies $a \vee a' \leqslant x$),

$b \neq x$ (otherwise, x would be the union of predecessors of a),

$b \neq y$ (since the graph contains no circuits).

Now assume that the theorem is true for $\mu(y, z)$ with $z < x$, then it is true for $\mu(y, x)$, since

$$-\mu(y, x) = \sum_{z \in [y, x)} \mu(y, z) = \sum_{y \leqslant z < b} \mu(y, z) + \mu(y, b) + 0 = 0. \quad \textbf{Q.E.D.}$$

(This result may be verified in the above examples.)

3. SIEVE FORMULAS

Let X be a finite set on which is defined, for all $x \in X$, a function $m(x) \geqslant 0$ called the *measure* (or the weight) of x. If $A \subset X$, write

$$m(A) = \sum_{x \in A} m(x) \qquad \text{if} \quad A \neq \varnothing,$$

$$m(\varnothing) = 0 \qquad \text{if} \quad A = \varnothing,$$

$m(A)$ is called the *measure of the set* A.

For example, if $m(x) = 1$ for all $x \in A$, then $m(A) = |A|$ is the *cardinality* of the set A; if $m(x)$ is a probability distribution and A is the set of events having a certain property, then $m(A)$ is the probability of such an event.

FORMULA 1. *If $A, B \subset X$ and $\bar{A}$ denotes the complement $X - A$, then*

$$\boxed{\begin{aligned} m(\bar{A}) &= m(X) - m(A) \\ m(\overline{A \cup B}) &= m(\bar{A} \cap \bar{B}) \\ m(\overline{A \cap B}) &= m(\bar{A} \cup \bar{B}) \end{aligned}}$$

The proof is immediate.

FORMULA 2. *Let A_i $(i \in \{1, 2, \ldots, q\} = Q)$ be subsets of X, and let*

$$\overline{m}(K) = m\left(\bigcup_{i \in K} A_i\right) \quad \text{if} \quad K \neq \varnothing, \qquad = 0 \quad \text{if} \quad K = \varnothing,$$

$$\underline{m}(K) = m\left(\bigcap_{i \in K} A_i\right) \quad \text{if} \quad K \neq \varnothing, \qquad = 0 \quad \text{if} \quad K = \varnothing.$$

Then

$$\boxed{\overline{m}(Q) = \sum_{K \subset Q} (-1)^{|K|+1} \underline{m}(K)}$$

PROOF: If $q = 2$, then obviously

$$m(A_1 \cup A_2) = m(A_1) + m(A_2) - m(A_1 \cap A_2).$$

Now, proceed by induction on $q = |Q|$; assume the theorem is true for all sets of $q - 1$ subsets $A_1, A_2, \ldots, A_{q-1}$ then, for q subsets $A_1, A_2, \ldots, A_q$,

$$\begin{aligned} &m(A_1 \cup A_2 \cup \cdots \cup A_{q-1} \cup A_q) \\ &\qquad = m(A_1 \cup \cdots \cup A_{q-1}) + m(A_q) - m((A_1 \cup \cdots \cup A_{q-1}) \cap A_q) \\ &\qquad = m(A_1 \cup \cdots \cup A_{q-1}) + m(A_q) - m\left(\bigcup_{i<q} A_i \cap A_q\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i<q} m(A_i) - \sum_{i<j<q} m(A_i \cap A_j) \\
&\quad + \sum_{i<j<k<q} m(A_i \cap A_j \cap A_k) - \cdots \\
&\quad + m(A_q) - \sum_{i<q} m(A_i \cap A_q) \\
&\quad + \sum_{i<j<q} m(A_i \cap A_j \cap A_q) - \cdots \\
&= \sum_{i\leqslant q} m(A_i) - \sum_{i<j\leqslant q} m(A_i \cap A_j) \\
&\quad + \sum_{i<j<k\leqslant q} m(A_i \cap A_j \cap A_k) - \cdots,
\end{aligned}
$$

which completes the induction.

FORMULA 3. *Let A_i ($i \in \{1, 2, \ldots, q\} = Q$) be subsets of X. Then*

$$\boxed{\underline{m}(Q) = \sum_{K\subset Q} (-1)^{|K|+1}\overline{m}(K)}$$

PROOF: The Möbius function required here (see Example 3) is given by

$$\mu(I, J) = (-1)^{|J|-|I|}.$$

Using the Möbius inversion theorem, Formula 2 becomes

$$(-1)^{|Q|+1}\underline{m}(Q) = \sum_{K\subset Q} (-1)^{|K|-|Q|}\overline{m}(K).$$

Hence,

$$\underline{m}(Q) = \sum_{K\subset Q} (-1)^{|K|+1}\overline{m}(K) = \sum_{i\leqslant q} m(A_i) - \sum_{i<j\leqslant q} m(A_i \cup A_j) + \cdots.$$

Formula 3 can also be obtained from Formula 2 by taking complements.

SYLVESTER'S FORMULA. *Let $A_1, A_2, \ldots, A_q \subset X$. Let $\underline{\underline{m}}(K) = \underline{m}(K)$, $K \neq \varnothing$, and $\underline{\underline{m}}(\varnothing) = m(X)$; the measure of the set of elements of X that do not belong to any of the sets A_i is*

$$\boxed{T_q^{\,0} = \sum_{k=0}^{q} (-1)^k \sum_{\substack{K\subset Q\\ |K|=k}} \underline{\underline{m}}(K)}$$

PROOF: Sylvester's formula can be deduced immediately from Formula 2, since

$$T_q^{\,0} = m\Big(\bigcap_{i\in Q}\bar{A}_i\Big) = m\Big(\overline{\bigcup_{i\in Q}A_i}\Big) = m(X) - m\Big(\bigcup_{i\in Q}A_i\Big)$$
$$= \underline{\underline{m}}(\varnothing) - \sum_{\substack{K\subset Q\\ K\neq\varnothing}}(-1)^{|K|+1}\underline{\underline{m}}(K).$$

SIEVE FORMULA. *Let $A_1, A_2, \ldots, A_q \subset X$; the measure of the set of elements of X that belong to exactly p of the sets A_i is*

$$\boxed{T_q^{\,p} = \sum_{k=p}^{q}(-1)^{k-p}\binom{k}{p}\sum_{\substack{K\subset Q\\ |K|=k}}\underline{\underline{m}}(K)}$$

PROOF: Let $P \subset Q$ be an arbitrary subset of p elements. Replacing X by $\bigcap_{i\in P} A_i$, and A_j by $A_j \cap \bigcap_{i\in P} A_i$, in Sylvester's formula gives

$$m\Big(\bigcap_{i\in P}A_i \cap \bigcap_{j\in Q-P}\bar{A}_j\Big) = m\Big(\bigcap_{i\in P}A_i\Big) - \sum_{\substack{K\supset P\\ |K|=p+1}} m\Big(\bigcap_{i\in K}A_i\Big) + \cdots$$
$$= \sum_{K\supset P}(-1)^{|K|-|P|}\underline{\underline{m}}(K).$$

Hence,

$$T_q^{\,p} = \sum_{|P|=p} m\Big(\bigcap_{i\in P}A_i \cap \bigcap_{j\in Q-P}\bar{A}_i\Big)$$
$$= \sum_{|P|=p}\ \sum_{K\supset P}(-1)^{|K|-|P|}\underline{\underline{m}}(K)$$
$$= \sum_{\substack{K\subset Q\\ |K|\geqslant p}}\ \sum_{\substack{P\subset K\\ |P|=p}}(-1)^{|K|-|P|}\underline{\underline{m}}(K)$$
$$= \sum_{k=p}^{q}(-1)^{k-p}\sum_{\substack{K\subset Q\\ |K|=k}}\underline{\underline{m}}(K)\binom{k}{p}.$$

Notice that Sylvester's formula is a special case ($p = 0$) of the sieve formula.

The sieve formulas lead to a very general method of enumeration (named the sieve method after the "Sieve of Eratosthenes," which provides a construction for recursively cataloging the prime numbers). To enumerate the elements that do not belong to any of the sets $A_1, A_2, \ldots, A_q$, start with the set X and eliminate those elements belonging to A_1, then those belonging to A_2, and so on until $\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_q$ is obtained.

APPLICATION 1 "*Problème des Rencontres*" (Montmort [24]). Let

$$\varphi = \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

be a permutation, i.e., a bijection from $\{1, 2, \ldots, n\}$ into itself. φ is said to admit a *coincidence* (or a "rencontre") *at* i, if $\varphi(i) = i$.

For example, the permutation

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 3 & 7 & 5 & 6 & 4 \end{pmatrix}$$

admits coincidences at 3, 5, and 6.

The problem is to obtain the number of permutations of order n which admit exactly p coincidences.

Let $P(n)$ be the number of permutations admitting no coincidences, and let A_i be the set of $(n-1)!$ permutations that admit a coincidence at i.

From Sylvester's formula

$$P(n) = |\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_n| = \underline{\underline{m}}(\varnothing) - \sum_{|K|=1} \underline{\underline{m}}(K) + \sum_{|K|=2} \underline{\underline{m}}(K) - \cdots,$$

where, if $|K| = k$,

$$\underline{\underline{m}}(K) = \left|\bigcap_{i \in K} A_i\right| = (n-k)!.$$

Therefore,

$$P(n) = n! - \binom{n}{1}(n-1)! + \cdots + (-1)^k \binom{n}{k}(n-k)! + \cdots + (-1)^n \binom{n}{n},$$

i.e.,

$$\boxed{P(n) = n!\left[1 - \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{(-1)^k}{k!} + \cdots + (-1)^n \frac{1}{n!}\right]}$$

Obviously, the number of permutations admitting exactly p coincidences is

$$P^p(n) = \binom{n}{p} P(n-p).$$

Thus, the problem reduces to a calculation of the $P(n)$'s, which can be computed easily, using the following formulas:

$$P(1) = 0$$
$$P(n) = nP(n-1) + (-1)^n,$$

from which we obtain

$P^p(n)$	$p = 0$	$p = 1$	$p = 2$	$p = 3$	$\cdots$
$n = 1$	0	1			
$n = 2$	1	0	1		
$n = 3$	2	3	0	1	
$n = 4$	9	8	6	0	
$n = 5$	44	45	20	10	
$\vdots$					

APPLICATION 2 "*Problème des Ménages*" (Touchard [20]). The "problème des ménages" asks for the number of ways $T(n)$ of seating n husbands (labeled $1, 2, \ldots, n$) and their respective wives (labeled $\underline{1}, \underline{2}, \ldots, \underline{n}$) at a circular table, in such a way that each man has a woman seated on either side of him, neither being his wife. Such a seating arrangement can be described by a bijection φ: Seat man 1; seat to his right woman $\varphi(1)$; seat to her right man 2; seat to his right woman $\varphi(2)$; etc. Let A_{2i-1} ($i \in \{1, 2, \ldots, n\}$) be the set of bijections φ with $\varphi(i) = \underline{i}$.

If $i \neq n$, let A_{2i} be the set of bijections φ with $\varphi(i) = \underline{i+1}$ (for $i = n$, A_{2n} denotes the set of bijections φ with $\varphi(n) = \underline{1}$). From Sylvester's formula, the solution to the "problème des ménages" is

$$T(n) = \left| \bigcap_{i \in Q} \bar{A}_i \right| = \sum_{K \subset Q} (-1)^{|K|} \underline{\underline{m}}(K),$$

where $Q = \{1, 2, \ldots, 2n\}$.

If $|K| = k$

$$\underline{\underline{m}}(K) = \left|\bigcap_{i \in K} A_i\right| = (n-k)! \qquad \text{if } K \text{ does not contain two consecutive integers from the sequence } (1, 2, \ldots, 2n, 1),$$

$$= 0 \qquad \text{otherwise.}$$

From Chapter 1 (Section 8, Fibonacci numbers), however, the number of sets K of cardinality k, not containing two consecutive integers from the sequence $(1, 2, \ldots, 2n, 1)$ is

$$f^*(2n, k) = \frac{2n}{2n-k}\binom{2n-k}{k}.$$

Hence,

$$T(n) = n! - \frac{2n}{2n-1}\binom{2n-1}{1}(n-1)!$$

$$+ \frac{2n}{2n-2}\binom{2n-2}{2}(n-2)! + \cdots + (-1)^n \frac{2n}{n}\binom{n}{n}0!.$$

Similarly, from the sieve formula, the number of ways of seating n husbands and their wives alternately around a circular table, in such a way that exactly p husbands sit next to their own wives, is

$$T^p(n) = \sum_{k=p}^{2n} (-1)^{k-p}\binom{k}{p}\frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!.$$

Application 3 *Counting prime numbers* (Euler). The problem here is to calculate the number $\varphi(n)$ of integers less than or equal to n ($n > 0$) which are coprime to n.

Decomposing n into its prime factors suppose

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_q^{\alpha_q},$$

where $p_1, p_2, \ldots, p_q$ are distinct primes not equal to 1.

Let A_i be the set of integers $\leq n$ which are multiples of p_i; then

$$|A_i| = \frac{n}{p_i},$$

$$|A_i \cap A_j| = \frac{n}{p_i p_j}, \qquad i \neq j, \quad \text{etc.}$$

Take X to be the set of positive integers $\leq n$ and $m(A)$ equal to the cardinality of A; then, from Sylvester's formula,

$$\begin{aligned}\varphi(n) &= |X| - \sum |A_i| + \sum |A_i \cap A_j| - \sum |A_i \cap A_j \cap A_k| + \cdots \\ &= n - \sum_i \frac{n}{p_i} + \sum_{i<j} \frac{n}{p_i p_j} - \sum_{i<j<k} \frac{n}{p_i p_j p_k} + \cdots.\end{aligned}$$

Finally,

$$\boxed{\varphi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_q}\right)} \qquad \text{(Euler's function)}$$

4. Distributions

Let X be a set of n objects, some being indistinguishable from others; two objects x and y are said to be in the *same class* (or of the *same kind*) if they are indistinguishable; if the classes determine a partition of X of type $1^{\lambda_1}2^{\lambda_2} \cdots n^{\lambda_n}$, then X is said to be a *collection of objects of type* $1^{\lambda_1}2^{\lambda_2} \cdots n^{\lambda_n}$.

The objects are to be put into boxes $a_1, a_2, \ldots, a_m$ which, some of them being indistinguishable from others, form a collection of type $1^{\mu_1}2^{\mu_2} \cdots m^{\mu_m}$.

A *distribution* is a mapping φ of X into A; two distributions are said to be *equivalent* if they are indistinguishable, i.e., if one can be obtained from the other by a permutation of objects of the same kind, or of boxes of the same kind; the classes of this equivalence relation are called *schemata*.

Let

$$R'(1^{\lambda_1}2^{\lambda_2} \cdots n^{\lambda_n}; 1^{\mu_1}2^{\mu_2} \cdots m^{\mu_m}).$$

be the number of schemata relative to the distributions of a collection of objects of type $1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n}$ into a collection of boxes of type $1^{\mu_1}2^{\mu_2}\cdots m^{\mu_m}$.

Let

$$R(1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n};\ 1^{\mu_1}2^{\mu_2}\cdots m^{\mu_m})$$

be the number of these schemata in which no box is left empty.

Then it is clear that, if $m_1\ m_2\ \cdots\ m_p$ denotes a partition of the integer m into parts $m_1, m_2, \ldots, m_p$:

$$R'(1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n};\ m_1m_2\cdots m_p) = \sum_{\substack{0\leqslant k_1\leqslant m_1\\ 0\leqslant k_2\leqslant m_2\\ \vdots}} R(1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n};\ k_1k_2\cdots k_p).$$

Moreover, if $\mu(k_1k_2\cdots k_p, m_1m_2\cdots m_p)$ denotes the Möbius function (Chapter 3, Section 2) on the lattice of p-tuples (Chapter 1, Section 9), then

$$\begin{aligned} &R(1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n};\ m_1m_2\cdots m_p)\\ &\quad= \sum_{\substack{0\leqslant k_1\leqslant m_1\\ 0\leqslant k_2\leqslant m_2\\ \vdots}} \mu(k_1k_2\cdots k_p, m_1m_2\cdots m_p)\times R'(1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n};\ k_1k_2\cdots k_p).\end{aligned}$$

In Chapter 5, a general method for calculating R and R' will be given; however, in a great many cases, these numbers can be calculated directly, using very simple formulas.

PROPOSITION 1

$$R(1^n; m) = S_n^{\,m} \qquad (\textit{Stirling number}),$$

$$R'(1^n; m) = S_n^{\,1} + S_n^{\,2} + \cdots + S_n^{\,m}.$$

PROOF: $R(1^n; m)$ is the number of partitions of a set of n distinct objects into m classes.

PROPOSITION 2

$$R(1^n;\ 1^m) = m!S_n^{\,m}$$

$$R'(1^n;\ 1^m) = m^n.$$

PROOF: $R(1^n; 1^m)$ is the number of surjections of X into A, and $R'(1^n; 1^m)$ is the total number of mappings of X into A.

PROPOSITION 3

$$R(n; m) = P_n^{\,m} \qquad \textit{(the number of partitions of } n \textit{ into } m \textit{ parts)},$$

$$R'(n; m) = P_n^{\,1} + P_n^{\,2} + \cdots + P_n^{\,m}.$$

PROOF: The proof is obvious.

PROPOSITION 4

$$R(n; 1^m) = \binom{n-1}{m-1},$$

$$R'(n; 1^m) = \binom{n+m-1}{n}.$$

PROOF: $R'(n; 1^m)$ is the number of solutions $u_1, u_2, \ldots, u_m$ of $u_1 + u_2 + \cdots + u_m = n$, where the u_i's are nonnegative integers. Hence (Chapter 1, Section 7)

$$R'(n; 1^m) = \frac{[n+1]^{m-1}}{(m-1)!} = \binom{n+m-1}{n}.$$

$R(n; 1^m)$ is the number of solutions $u_1, u_2, \ldots, u_m$ of $u_1 + u_2 + \cdots + u_m = n$, where the u_i's are integers greater than 0. Two solutions $(u_1, u_2, \ldots, u_m)$ and $(u_1', u_2', \ldots, u_m')$ are considered equal if and only if $u_i = u_i'$ for each i; if $s_k = u_1 + u_2 + \cdots + u_k$, each solution is completely determined by the numbers $s_1, s_2, \ldots, s_{m-1}$, satisfying

$$1 \leqslant s_1 < s_2 < \cdots < s_{m-1} \leqslant n-1,$$

and conversely.

The first formula is an obvious consequence of this remark.

THEOREM

$$R(1^{\lambda_1} 2^{\lambda_2} \cdots n^{\lambda_n}; 1^m) = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} \binom{k}{1}^{\lambda_1} \binom{k+1}{2}^{\lambda_2} \cdots \binom{k+n-1}{n}^{\lambda_n},$$

$$R'(1^{\lambda_1}2^{\lambda_2}\cdots n^{\lambda_n};1^m)=\binom{m}{1}^{\lambda_1}\binom{m+1}{2}^{\lambda_2}\cdots\binom{m+n-1}{n}^{\lambda_n}.$$

PROOF: Assume that some of the boxes may be left empty. First distribute the n_1 objects of the first kind, then the n_2 objects of the second kind, etc., giving as the number of possibilities

$$\begin{aligned}R'(n_1n_2\cdots;1^m)&=R'(n_1;1^m)R'(n_2;1^m)\cdots,\\&=\binom{n_1+m-1}{n_1}\binom{n_2+m-1}{n_2}\cdots.\end{aligned}$$

This proves the second formula.

We now prove the first formula. Let R_k or $R(K)$ denote the number of distributions into a subset K ($|K|=k$) of the set A of boxes such that no box is left empty, then

$$R_m'=\sum_{K\subset A}R(K)=\sum_{k=0}^{m}\binom{m}{k}R_k.$$

Therefore, by the inverse binomial formula (Chapter 3, Section 1),

$$R_m=\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}R_k',$$

whence the first formula.

APPLICATION: *The number of perfect partitions of an integer s.* Consider a box of weights containing

t_1 weights of one kilogram,

t_2 weights of two kilograms, etc.,

where $t_1+t_2+\cdots=s$.

The partition $1^{t_1}2^{t_2}\cdots s^{t_s}$ is said to be *perfect*, if the box of weights will weigh any object of integral weight from 1 to s kilograms in one and only one way.

The problem is to find the number of perfect partitions of the integer s containing exactly m types of weights.

RESULT. *Let the prime decomposition of* $s+1$ *be given by*

$$s + 1 = \prod_{i=0}^{\lambda_1} p_i^{\,1} \left(\prod_{i=0}^{\lambda_2} p_i^{\,2} \right)^2 \left(\prod_{i=0}^{\lambda_3} p_i^{\,3} \right)^3 \cdots,$$

where for each k, $p_0^{\,k} = 1$, *and where for each* $i \neq 0$, *the* $p_i^{\,k}$ *are distinct prime numbers different from* 1.

The number of perfect partitions of s, *which consist of exactly* m *types of weights, is equal to* $R(1^{\lambda_1} 2^{\lambda_2} \cdots n^{\lambda_n}; 1^m)$.

PROOF: Suppose

$$s + 1 = p_1' p_2' \cdots p_n',$$

where the p_i's are prime numbers (not necessarily distinct) not equal to 1.

Consider a perfect partition of s consisting of exactly m types of weights. At least one of the parts of this partition must be equal to 1; if the number of 1's in this partition is $a_1 - 1$, then the succeeding part of the partition must be equal to a_1; if the number of parts equal to a_1 is $a_2 - 1$, then the succeeding part of the partition must be equal to $a_1 a_2$, and suppose there are $a_3 - 1$ of them; etc.

Therefore,

$$s = a_1 - 1 + (a_2 - 1)a_1 + (a_3 - 1)a_1 a_2 + (a_4 - 1)a_1 a_2 a_3 + \cdots$$
$$+ (a_m - 1)a_1 a_2 \cdots a_{m-1},$$

i.e.,

$$s + 1 = a_1 a_2 \cdots a_m = p_1' p_2' \cdots p_n'.$$

Such a sequence $(a_1, a_2, \ldots, a_m)$ completely determines a perfect partition of s consisting of exactly m types of weights; as the p''s are prime numbers not equal to 1, the number of such sequences is

$$R(1^{\lambda_1} 2^{\lambda_2} \cdots n^{\lambda_n}; 1^m). \quad \textbf{Q.E.D.}$$

5. COUNTING TREES

We recall some standard graph-theoretic definitions: a *graph* is defined by a set X of *vertices* and a set of *arcs*, or pairs (x, y) of vertices;

an *edge* is a set $\{x, y\}$ of two vertices joined by an arc. Let U denote the set of arcs.

A *path* is a sequence of arcs such that the terminal vertex of each arc is the initial vertex of the succeeding arc; a path that starts and finishes at the same vertex is a *circuit*. Vertices a and b belong to the same *strongly connected component* if there exists a path from a to b, and a path from b to a. A *chain* is a sequence of distinct edges, each edge having one vertex in common with the preceding edge, and the other vertex in common with the succeeding edge.

A *cycle* is a chain that begins and ends at the same vertex. Vertices a and b belong to the same *connected component* if there exists a chain from a to b.

A *partial graph* of (X, U) is a graph (X, V) with $V \subset U$; a *subgraph* of (X, U) is a graph (S, V) with $S \subset X$, and V the subset of U, consisting of all the edges that join two elements of S. We recall [22] that, if G is a graph with n vertices, m arcs, and p connected components, then the *number of independent cycles* is

$$k(G) = m - n + p.$$

For a proof of this see, for instance, Berge [22].

A *tree* is a connected graph containing no cycles; alternatively:

THEOREM 1. *Let $H = (X, U)$ be a graph of order $|X| = n \geqslant 2$; the following properties are equivalent and each characterizes a tree*:

(1) *H is connected and contains no cycles;*
(2) *H contains no cycles and has exactly $n - 1$ edges;*
(3) *H is connected and has exactly $n - 1$ edges;*
(4) *G contains no cycle and if an edge is adjoined to H one and only one cycle is created;*
(5) *H is connected and if any edge is deleted from H, then H becomes disconnected;*
(6) *every pair of vertices of H is connected by one and only one chain.*

PROOF:

(1) $\Rightarrow$ (2) since, if p is the number of connected components, and m the number of edges,

$$p = 1, \qquad k(H) = m - n + p = 0.$$

Therefore, $m = n - p = n - 1$.

(2) $\Rightarrow$ (3) since $k(H) = 0$, and $m = n - 1$ imply

$$p = k(H) - m + n = 1,$$

and H is connected.

(3) $\Rightarrow$ (4) since $p = 1$, and $m = n - 1$ imply

$$k(H) = m - n + p = 0.$$

Therefore, H contains no cycles; moreover, adding one edge to H increases $k(H)$ by 1, i.e., there exists in the new graph, one and only one cycle.

(4) $\Rightarrow$ (5) since, if H was not connected, there must exist two vertices a and b that are not connected, and, therefore, a new cycle could not possibly be created by adjoining the edge $\{a, b\}$; therefore, $p = 1$, $k(H) = 0$, which imply $m = n - 1$. On the other hand, deleting an edge

$$m' = n' - 2, \qquad k(H') = 0,$$

and

$$p' = k(H') - m' + n' = 2,$$

i.e., H' is disconnected.

(5) $\Rightarrow$ (6). For any two vertices a and b, there exists a chain from a to b (H is connected). If there existed a second chain from a to b, then the deletion of any edge that belonged only to the second chain would not disconnect the graph.

(6) $\Rightarrow$ (1) for if H contains a cycle, at least one pair of vertices must be connected by two distinct chains.

COROLLARY. *A graph $G = (X, U)$ possesses a partial subgraph which is a tree, if and only if, it is connected.*

PROOF: If G is not connected, no partial graph of G is connected, and, therefore, G possesses no partial trees.

If G is connected, suppose there exists an edge, whose deletion does not disconnect G. If no such edge exists then G is a tree, by (5); if such an edge exists, delete it. Now repeat this process, deleting if possible, another edge that does not disconnect G, etc.

Eventually, it will not be possible to delete any more edges without disconnecting the graph, and, hence, we shall have a tree whose set of vertices is X.

This corollary gives a simple algorithm for constructing a partial tree H of a connected graph G.

Given a graph (X, U) the *degree* of a vertex $a \in X$ is, by definition, the number $d(a)$ of edges which have a as end vertex. A vertex x, with $d(x) = 1$, is called *pendant*.

THEOREM 2 (Moon [9]). *Let $T(n; d_1, d_2, \ldots, d_n)$ be the number of trees with vertices $x_1, x_2, \ldots, x_n$ and degrees $d(x_1) = d_1$, $d(x_2) = d_2, \ldots, d(x_n) = d_n$; then*

$$T(n; d_1, d_2, \ldots, d_n) = \binom{n-2}{d_1 - 1, d_2 - 1, \ldots, d_n - 1}.$$

PROOF:

(1) Clearly, from the definitions, the sum of the degrees of any graph is twice the number of edges. Therefore, by Theorem 1, for a tree, the sum of the degrees is $2(n-1)$. Hence, $T \neq 0$ only if

$$\sum_{i=1}^{n} (d_i - 1) = 2(n-1) - n = n - 2.$$

We may suppose that $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_n$, and otherwise relabel; since the above equality implies that $d_n = 1$, x_n is a "pendant" vertex of the tree.

(2) In order to prove that

$$T(n; d_1, d_2, \ldots, d_n) = \sum_{\substack{i \\ d_i \geqslant 2}} T(n-1; d_1, d_2, \ldots, d_i - 1, \ldots, d_{n-1}),$$

make a list $\mathscr{C}_i$ of the trees with vertices $x_1, x_2, \ldots, x_n$ and degrees $d(x_k) = d_k$, such that the pendant vertex x_n is joined to x_i. If $d_i \geqslant 2$,

$$|\mathscr{C}_i| = T(n-1; d_1, d_2, \ldots, d_i - 1, \ldots, d_{n-1}).$$

Since the list of the required trees is the union of these lists $\mathscr{C}_i$ $(d_i \geqslant 2)$, the equality (2) follows.

(3) The theorem is trivially true for $n = 3$; therefore, assume $n \geqslant 3$, and suppose the theorem is true for $n-1$; then

$$\begin{aligned} T(n; d_1, d_2, \ldots, d_n) &= \sum_{\substack{i \\ d_i \geqslant 2}} T(n-1; d_1, d_2, \ldots, d_i - 1, \ldots, d_{n-1}) \\ &= \sum_{\substack{i \\ d_i \geqslant 2}} \binom{n-3}{d_1 - 1, d_2 - 1, \ldots, d_i - 2, \ldots, d_{n-1} - 1} \\ &= \binom{n-2}{d_1 - 1, d_2 - 1, \ldots, d_{n-1} - 1} \\ &= \binom{n-2}{d_1 - 1, d_2 - 1, \ldots, d_n - 1} \qquad (\text{since } d_n = 1) \end{aligned}$$

(by Consequence 1 of Section 9, Chapter 1).

COROLLARY 1 (Menon [7]). *The numbers $d_1, d_2, \ldots, d_n \geqslant 1$ are degrees of a tree, if and only if*

$$\sum_{i=1}^{n} d_i = 2(n-1).$$

PROOF: This condition is equivalent to $T(n; d_1, d_2, \ldots, d_n) \neq 0$.

COROLLARY 2 (Cayley's formula [2]). *The number of trees with vertices $x_1, x_2, \ldots, x_n$ is n^{n-2}.*

PROOF: By Theorem 2 and Corollary 1, the number of trees with vertices $x_1, x_2, \ldots, x_n$ is

$$\sum_{\substack{d_1, \ldots, d_n \geqslant 1 \\ d_1 + d_2 + \cdots + d_n = 2(n-1)}} \binom{n-2}{d_1 - 1, d_2 - 1, \ldots, d_n - 1},$$

which, from Section 9, Chapter 1, equals

$$(1 + 1 + \cdots + 1)^{n-2} = n^{n-2}.$$

COROLLARY 3 (Clarke [3]). *The number of trees with vertices x_1, $x_2, \ldots, x_n$ and $d(x_1) = k$ is*

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

PROOF: The required number is

$$\sum_{d_2, d_3, \ldots, d_n} \binom{n-2}{k-1, d_2-1, d_3-1, \ldots, d_n-1}$$

$$= \frac{(n-2)!}{(k-1)!(n-k-1)!} \sum_{d_2, d_3, \ldots, d_n \geqslant 1} \binom{n-k-1}{d_2-1, d_3-1, \ldots, d_n-1}$$

$$= \binom{n-2}{k-1}(n-1)^{n-k-1}$$

(on putting the variables equal to one in the multinomial formula).

COROLLARY 4 (Moon [9]). *Let*

$$H_1 = (X_1, U_1), H_2 = (X_2, U_2), \ldots, H_p = (X_p, U_p)$$

be disjoint trees of orders $|X_i| = n_i$; *the number of trees of order* n, *having the union of the* X_i*'s as their set of vertices and containing the* H_i*'s as subgraphs, is*

$$T(H_1, H_2, \ldots, H_p) = n_1 n_2 \cdots n_p n^{p-2}.$$

PROOF: For the time being assume that each set X_i is "contracted" to a unique vertex $\bar{x}_i$; the number of trees $\bar{H}$ with vertices $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p$ and $d(\bar{x}_i) = d_i$ is, by Theorem 2,

$$\binom{p-2}{d_1-1, d_2-1, \ldots, d_p-1}.$$

Clearly, each of these trees $\bar{H}$ corresponds to $(n_1)^{d_1}(n_2)^{d_2} \cdots (n_p)^{d_p}$ trees H containing the H_i's as subgraphs. Therefore,

$$T(H_1, H_2, \ldots, H_p)$$

$$= \sum_{d_1, d_2, \ldots, d_p \geqslant 1} \binom{p-2}{d_1-1, d_2-1, \ldots, d_p-1}(n_1)^{d_1}(n_2)^{d_2} \cdots (n_p)^{d_p}$$

$$= n_1 n_2 \cdots n_p(n_1 + n_2 + \cdots + n_p)^{p-2}.$$

The corollary follows, since $n = n_1 + n_2 + \cdots + n_p$.

COROLLARY 5 (Cayley [2]). *The number of graphs with vertices* $x_1, x_2, \ldots, x_n$, *consisting of p disjoint trees, and with* $x_1, x_2, \ldots, x_p$ *belonging to p different trees, is*

$$T'(n; p) = pn^{n-p-1}.$$

PROOF: Make a list $\mathscr{C}$ of the trees with vertices $x_0, x_1, x_2, \ldots, x_n$, such that $d(x_0) = p$; by Corollary 3,

$$|\mathscr{C}| = \binom{n-1}{p-1} n^{n-p}.$$

If $P \subset \{1, 2, \ldots, n\}$ and $|P| = p$, let $\mathscr{C}_p$ be the list of trees belonging to $\mathscr{C}$ such that, for all $i \in P$, x_i is joined to x_0. Then

$$|\mathscr{C}| = \sum_P |\mathscr{C}_P| = \sum_P T'(n; p).$$

Hence,

$$\binom{n-1}{p-1} n^{n-p} = \binom{n}{p} T'(n; p).$$

Therefore,

$$T'(n; p) = \frac{(n-1)!}{(p-1)!(n-p)!} \cdot \frac{p!(n-p)!}{n!} n^{n-p} = pn^{n-p-1}. \quad \textbf{Q.E.D.}$$

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of vertices, and let $U = \{u_1, u_2, \ldots, u_q\}$ be a set of edges joining pairs of vertices in X. Let us now try to find the number $T(X, U)$ of trees, with vertices $x_1, x_2, \ldots, x_n$, none of the edges of which belong to U. Using the theory of determinants, there exists a general formula (see Berge [22, Chapter 16]). However, for the particular cases that we wish to consider here, it does not prove very useful. Let (X, V) be a graph with n vertices, q edges and p connected components with, respectively, $n_1, n_2, \ldots, n_p$ vertices.

Write

$$v(V) = \begin{cases} 0 & \text{if the graph } (X, V) \text{ contains a cycle,} \\ n_1 n_2 \cdots n_p & \text{otherwise.} \end{cases}$$

THEOREM 3 (Temperley [19]). *The number of trees with vertices x_1, $x_2, \ldots, x_n$ and no edge belonging to U is*

$$T(X, U) = n^{n-2} \sum_{V \subset U} v(V)\left(\frac{-1}{n}\right)^{|V|}.$$

PROOF: If $v \in U$, let A_v be the set of trees which contain v. Suppose $V \subset U$. If (X, V) is acyclic and has p connected components, then, by Corollary 4, the number of trees that contain all the edges of V is

$$\left|\bigcap_{v \in V} A_v\right| = v(V)n^{p-2} = v(V)n^{n-|V|-2}.$$

If (X, V) contains a cycle, then this equality is still valid, since both sides vanish. Therefore, by Sylvester's formula (Chapter 3, Section 3),

$$T(X, U) = n^{n-2} + \sum_{\substack{V \subset U \\ V \neq \varnothing}} (-1)^{|V|} v(V) n^{n-2-|V|} = n^{n-2} \sum_{V \subset U} v(V)\left(\frac{-1}{n}\right)^{|V|}.$$

COROLLARY 1 (Weinberg [21]). *If U is a set of q pairwise disjoint edges,*

$$T(X, U) = n^{n-2}\left(1 - \frac{2}{n}\right)^q.$$

PROOF: In this case, when $V \subset U$,

$$v(V) = 2^{|V|},$$

and

$$T(X, U) = n^{n-2} \sum_{k=0}^{q} 2^k \binom{q}{k}\left(\frac{-1}{n}\right)^k = n^{n-2}\left(1 - \frac{2}{n}\right)^q.$$

COROLLARY 2 (O'Neil [12]). *If U consists of q edges having a common end vertex x_1,*

$$T(X, U) = n^{n-2}\left(1 - \frac{1}{n}\right)^{q-1}\left(1 - \frac{q+1}{n}\right).$$

PROOF:

$$\sum_{V \subset U} v(V)\left(\frac{-1}{n}\right)^{|V|} = \sum_{k=0}^{q} (k+1)\binom{q}{k}\left(\frac{-1}{n}\right)^k$$

$$= \sum_{k=0}^{q} \binom{q}{k} \left(\frac{-1}{n}\right)^k + \sum_{k-1=0}^{q-1} \left(\frac{-q}{n}\right) \binom{q-1}{k-1} \left(\frac{-1}{n}\right)^{k-1}$$

$$= \left(1 - \frac{1}{n}\right)^q - \frac{q}{n}\left(1 - \frac{1}{n}\right)^{q-1}$$

$$= \left(1 - \frac{1}{n}\right)^{q-1}\left(1 - \frac{1}{n} - \frac{q}{n}\right).$$

COROLLARY 3 (Austin). *If U is the set of edges joining all the possible pairs of distinct vertices in $S \subset X$ ($|S| = s$) [i.e., (S, U) is a "complete graph"], then*

$$T(X, U) = n^{n-2}\left(1 - \frac{s}{n}\right)^{s-1}.$$

PROOF: Let $\mathscr{V}_p$ be the family of $V \subset U$, such that (S, V) is acyclic with p connected components, then

$$\sum_{V \subset U} v(V)\left(\frac{-1}{n}\right)^{|V|} = \sum_{p=1}^{s} \left(\frac{-1}{n}\right)^{s-p} \sum_{V \in \mathscr{V}_p} v(V).$$

Assume $P \subset S$, $|P| = p$ and $V \in \mathscr{V}_p$, and suppose there exists an acyclic graph with vertex set S, edge set equal to V, and a vertex of P in each connected component of (S, V), then we denote this graph by (S, V, P). By Corollary 5 of Theorem 2

$$|\{(S, V, P)/V \in \mathscr{V}_p \text{ and } (S, V, P) \text{ exists}\}| = ps^{s-p-1}.$$

Therefore,

$$\sum_{V \in \mathscr{V}_p} v(V) = \sum_{V \in \mathscr{V}_p} |\{(S, V, P)/P \subset S, |P| = p, \text{ and } (S, V, P) \text{ exists}\}|$$

$$= |\{(S, V, P)/P \subset S, |P| = p, V \in \mathscr{V}_p, \text{ and } (S, V, P) \text{ exists}\}|$$

$$= \sum_{\substack{P \subset S \\ |P| = p}} ps^{s-p-1} = \binom{s}{p} ps^{s-p-1}.$$

Therefore,

$$\sum \nu(V)\left(\frac{-1}{n}\right)^{|V|} = \sum_{p=1}^{s}\left(\frac{-1}{n}\right)^{s-p}\binom{s}{p} p s^{s-p-1}$$

$$= \sum_{p-1=0}^{s-1}\left(\frac{-s}{n}\right)^{s-p}\binom{s-1}{p-1} = \left(1-\frac{s}{n}\right)^{s-1}.$$

COROLLARY 4 (Scoins [17]; Glickman [4]). *If the graph (X, U) is the union of two disjoint complete graphs (S, V) and (T, W) with $|S| = s$ and $|T| = t$, then*

$$T(X, U) = s^{t-1}t^{s-1}.$$

PROOF: From Theorem 3,

$$\frac{T(X, V \cup W)}{n^{n-2}} = \frac{T(X, V)}{n^{n-2}} \cdot \frac{T(X, W)}{n^{n-2}}.$$

Therefore, by Corollary 3 of Theorem 3,

$$T(X, U) = n^{n-2}\left(1-\frac{s}{n}\right)^{s-1}\left(1-\frac{t}{n}\right)^{t-1}$$

$$= (s+t)^{s+t-2}\left(\frac{s+t-s}{s+t}\right)^{s-1}\left(\frac{s+t-t}{s+t}\right)^{t-1} = s^{t-1}t^{s-1}.$$

COROLLARY 5 (Moon [9, 10]). *Let U be a set of $m-1$ edges forming an open chain on a set $Y \subset X$ of m vertices, then*

$$T(X, U) = n^{n-2}\sum_{p=1}^{m}\binom{m+p-1}{m-p}\left(\frac{-1}{n}\right)^{m-p}.$$

PROOF: If $V \subset U$ determines a graph (Y, V) with p connected components,

$$|V| = |Y| - p = m - p.$$

If $m_1, m_2, \ldots, m_p$ are the numbers of vertices of these components, respectively, $m_1 + m_2 + \cdots + m_p = m$.

Therefore, for $|V| = m - p$, there are as many graphs (Y, V) as there are integer solutions $m_1, m_2, \ldots, m_p > 0$ of this equation; hence,

$$\sum_{V \subset U} v(V)\left(\frac{-1}{n}\right)^{|V|} = \sum_{p=1}^{m} \left(\frac{-1}{n}\right)^{m-p} \sum_{\substack{|V| = m-p \\ V \subset U}} v(V)$$

$$= \sum_{p=1}^{m} \left(\frac{-1}{n}\right)^{m-p} \sum_{\substack{m_1, m_2, \ldots > 0 \\ m_1 + m_2 + \cdots + m_p = m}} m_1 m_2 \cdots m_p.$$

This last summation is equal to the coefficient of x^m in the expansion of

$$(x + 2x^2 + 3x^3 + \cdots)^p = x^p(1 - x)^{-2p}.$$

By the binomial formula this coefficient is equal to

$$(-1)^{m-p} \frac{-2p(-2p-1)(-2p-2)\cdots(-2p-(m-p-1))}{(m-p)!}$$

$$= \frac{(m+p-1)(m+p-2)\cdots(2p+1)2p}{(m-p)!}$$

$$= \binom{m+p-1}{m-p}.$$

The corollary follows.

References

1. L. Carlitz, Rings of arithmetic functions, *Pacific J. Math.* **14**, 1165–1171 (1964).
2. A. Cayley, A theorem on trees, *Quart. J. Math.* **23**, 376–378 (1889).
3. L. E. Clarke, On Cayley's formula for counting trees, *J. London Math. Soc.* **33**, 471–475 (1958).
4. S. Glickman, On the representation and enumeration of trees, *Proc. Cambridge Philos. Soc.* **59**, 509–517 (1963).
5. I. Kaplansky and J. Riordan, the problème des menages, *Scripta Math.* **12**, 113–124 (1946).
6. H. Meier-Wunderli, Note on a basis of P. Hall for higher commutation in tree groups, *Comment. Helv. Math.* **26**, 1–5 (1952).
7. V. V. Menon, On the existence of trees with given degrees, *Sankhya* **26**, 63–68 (1964).

8. A. F. Möbius, Uber eine besondere Art von Umkehruhg der Reihen, *J. Reine Angew. Math.* **9**, 105–123 (1832).
9. J. W. Moon, Enumerating labeled trees, *in* "Graph Theory and Theoretical Physics" (F. Harary, ed.), pp. 261–271. Academic Press, New York, 1967.
10. J. W. Moon, On the second moment of the complexity of a graph, *Mathematika* **11**, 95–98 (1964).
11. C. Moreau, cited by E. Lucas, "Théorie des Nombres," pp. 396–397. Gauthiers-Villars, Paris, 1891.
12. P. V. O'Neill, The number of trees in a certain network, *Notices Amer. Math. Soc.* **10**, 569 (1963).
13. A. Rényi, Some remarks on the theory of trees, *Magy. Tudom. Akad. Mat. Kut. Intéz. Kölz.* **4**, 73–85 (1959a).
14. G.-C. Rota, Private communication.
15. G.-C. Rota, On the foundations of combinatorial theory I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie* **2**, 340–368 (1964).
16. M. P. Schützenberger, Contribution aux applications statistiques de la théorie de l'information, *Publ. Inst. Stat. Univ. Paris* **3**, 5–117 (1954).
17. H. I. Scoins, The number of trees with nodes of alternate parity, *Proc. Cambridge Philos. Soc.* **58**, 12–16 (1962).
18. D. A. Smith, Incidence functions as generalized arithmetic functions, *Duke Math.* **34**, 617–633 (1967).
19. H. N. V. Temperley, On the mutual cancellation of cluster integrals in Mayer's fugacity series, *Proc. Phys. Soc.* (*London*) **83**, 3–16 (1964).
20. J. Touchard, Sur un problème de permutations, *C. R. Acad. Sci.* **198**, 631–633 (1943).
21. L. Weinberg, Number of trees in a graph, *Proc. IRE* **46**, 1954–1955 (1958).
22. C. Berge, "Théorie des Graphes et ses Applications." Dunod, Paris, 1958.
23. W. R. Frucht, and G.-C. Rota, The Möbius function for partitions of a set. *Scientia (Valparaiso)* **122**, 111–115 (1963).
24. J. Touchard, Remarques sur les probabilités totales et sur le problème des rencontres, *Ann. Soc. Sci. Bruxelles* **A53**, 126–134 (1933).