GENERAL NUMERATION II. DIVISION SCHEMES

by

D. W. DUBOIS

0. This is the second in a series of two papers on numeration schemes. Whereas the first paper emphasized grouping as exemplified in the partition of a number so as to obtain its base two numeral, the present paper takes as its point of departure the method of «repeated divisions», as in the calculation of the base two numeral for a number by dividing it by two, then dividing the quotient by two, etc., and collecting the remainders. This method is a sort of classification scheme — odd or even. Now for classification schemes *par excellence* one naturally thinks of Biology with its phylla, classes, orders, etc. A first big difference between the biological divisions and the divisions associated with any based numeration scheme is that the number of subdivisions of a division varies; for example, one floral family might have over 50 genera, while another has fewer than 5 genera. There is simply no uniformity as to the numbers of subdivisions. The division schemes of this paper allow for this nonuniformity.

A non-numerical consideration has furnished impetus to the investigation of division schemes in the form of a vague conjecture about library classifications and shelving of books. As a partial explanation of the widespread discontent among library users with the arrangement of books on the shelves, I have conjectured that in any large library, however well-designed it is, there will be *dislocations* — that is, there will exist categories of books on «closely related» subjects which are shelved «far apart», as if they *weren't* closely related, not because of a clerical error but as a part of the shelving scheme. Since a whole hour of deliberation on the conjecture by Professors Erdös and Entringer failed to yield a solution, or even a reasonable combinatorial formulation of it, it would seem that the library dislocation problem is non-trivial, if it exists at all.

However, a very special sort of library dislocation problem is illuminated by the main theorem here, by means of a botanical interpretation. The idea is that a perfectly reasonable coding scheme is employed which attaches to each lowest division — i.e. to each species, a «numeral» in the form of a string of whole numbers. Then, in a natural way, to each numeral is assigned a whole number — a numeration scheme is thus defined. Now the problem arises whether the numeration scheme preserves order — if not then there is a «dislocation». The solution is simple but remarkable. If the classification scheme is symmetric (uniform) in the sense that at each level, the divisions at that level all have the same number of subdivisions at the next level (say each division contains V classes, each class contains W orders, etc.) then the numeration scheme preserves order. *But not otherwise.*

The models for the division schemes used here are rooted trees with labels and having the same length for every branch (measured from root to tip). The labels are sequences of whole numbers. Such a uniform tree suggests the inflorescence which botanists call a (compound) *umbel.*

1. Definitions and statement of the problems. A *numbered tree* is an object $\mathbf{U} = \langle \mathbf{T}, * \rangle$ consisting of set $\mathbf{T}$ (whose members are *twigs*) and a map $*$ from $\mathbf{T}$ into the set W of all nonnegative integers, satisfying the following conditions:

i) $\varnothing \in \mathbf{T}$. $\varnothing$ is the *root*.

ii) Every member of $\mathbf{T}$, except $\varnothing$, is a finite sequence (*not* a numeral) of whole numbers, written backward (e.g. $(n_3 n_2 n_1 n_0)$).

iii) For all n_0 in W, $(n_0) \in \mathbf{T}$ if and only if $n_0 < \varnothing^*$.

iv) For all n_i in W, $(n_k \dots n_0) \in \mathbf{T}$ if and only if $(n_{k-1} \dots n_0) \in \mathbf{T}$ and $n_k < (n_{k-1} \dots n_0)^*$.

The map $*$ is sometimes referred to as the *branching function*. The *length of the sequence* $(n_k \dots n_0)$ is k. Note that if $(n_j \dots n_0) \in \mathbf{T}$, if $(n_j \dots n_0)^* = 0$, if $m_i = n_i$ for all i from 0 to j and if length of the sequence m exceeds the length of n, then m does not belong to $\mathbf{T}$. The *height* of a numbered tree is the supremum of the lengths of its twigs.

An *umbel* is a numbered tree whose «branches» all have the same length, a branch being a maximal twig. Thus for an umbel, either the branching function never vanishes or else there is a k such that for all n in $\mathbf{T}$, $n^* = 0$ if and only if the length of n is k. A *symmetric* *umbel* is an umbel whose branching function depends only on the length of the sequences: If m and n have the same length then $m^* = n^*$.

Let $\mathbf{U} = \langle \mathbf{T}, * \rangle$ and $\mathbf{V} = \langle \mathbf{R}, ' \rangle$ be two umbels. Then $\mathbf{V}$ is a *subtree* of $\mathbf{U}$ provided for some $k \leqslant$ height of $\mathbf{U}$,

a) $\mathbf{R} = \{n ; n \in \mathbf{T}$ and length of n is at most $k\}$.
b) For all n in $\mathbf{R}$, if length of n is less than k then $n' = n^*$.
c) For all n in $\mathbf{R}$, if length of n equals k then $n' = 0$.

The subtree $\mathbf{V}$ described in the definition is denoted $\mathbf{U}_k$. In a real or pictorial model, $\mathbf{U}_k$ is obtained from $\mathbf{U}$ by uniformly pruning back to make an umbel of height k.

Let $i \geqslant 0$, let $n \in \mathbf{T}$, and suppose the length of n is at least $i - 2$. The value $C_i^* (n)$ is defined inductively:

$$C_i^* (n) = (n_{i-2} \ldots n_0)^* \, C_{i-1}^* (n) ; \qquad C_0^* (n) \equiv 1 , \qquad C_1^* (n) = \varnothing^* .$$

The function B^* is then defined, for n in $\mathbf{T}$:

$$B^* (n_k \ldots n_0) = \sum_{i \leq k} n_i \, C_i^* (n) \cdot$$

Recall that the twigs, i.e., members of $\mathbf{T}$, are not numerals. Moreover, as defined in [1], for finite sequences m and n, «$m \sim n$» means that m and n are the same except possibly for initial zeros — for some j, $m_i = n_i$ for all $i < j$; and for all $i > j$ each of m_i and n_i is either zero or undefined. The set $(\mathbf{T} \setminus \{\varnothing\}) / \sim$ of all equivalence classes by the relation $\sim$, is denoted by $[\mathbf{T}]$. Now the C_i^* are not defined over $[\mathbf{T}]$, but B^* is, because B^*, obviously, ignores initial zeros. Members of $[\mathbf{T}]$ are *numerals*, denoted with square brackets — thus if $n \in \mathbf{T} - \{\varnothing\}$ then the numeral containing n is written $[n]$.

The map defined on $[\mathbf{T}]$ by B^* is a *divided scheme*, or *division scheme*,

$$B^* : [\mathbf{T}] \longrightarrow W,$$

associated with the umbel $\mathbf{U} = \langle \mathbf{T}, * \rangle$. It is clearly a generalized numeration scheme. It is univalent provided B^* is injective, *complete* provided $B^* [\mathbf{T}]$ is an initial segment (possibly all) of W, *order-preserving in the strong sense* provided $m > n$ implies $B^* [m] > B^* [n]$. Note that B^* is both complete and order-preserving in the strong sense if and only if it maps the S^{th} numeral in $[\mathbf{T}]$ to the integer S, for all S.

The paper is devoted to proving the

MAIN THEOREM.—Assume that $\mathbf{U} = \langle \mathbf{T}, * \rangle$ is an umbel, with associated division scheme B^*. If $\mathbf{U}$ is symmetric then B^* is complete and order-preserving in the strong sense. If B^* is complete and order-preserving in the strong sense, if $\mathbf{U}$ is infinite, and if the branching function never assumes the value 1, then $\mathbf{U}$ is symmetric.

2. Proof of the theorem, consisting of a series of constructions and lemmas about an umbel $U = \langle \mathbf{T}, * \rangle$.

A. LEMMA.—Recursion relations. Assume n_{k+1} and m_{k-1} are nonzero members of W, and that the sequences below belong to $\mathbf{T}$.

i)

$$B^* [n_k \dots n_0] = n_k \ C_k^* (n_k \dots n_0) + B^* [n_{k-1} \dots n_0].$$

ii) Let

$$p = (n_{k+1}\, n_k\, q_{k-1} \dots q_0), \quad q = (q_{k-1} \dots q_0).$$

Then

$$B^* [p] = (n_{k+1}\, q^* + n_k)\ C_k^* + B^* [q],$$

where C_k^* may be evaluated at either p or q.

iii)

$$C_i^* \ n_k \dots n_0) = (n_{i-2} \dots n_0)^* \cdot \ C_{i-1}^* (n_k \dots n_0).$$

B. Construction of A^*. Assume $\mathbf{U} = \langle \mathbf{T}, * \rangle$ is an umbel, that the branching function never assumes the value 1, and that $\mathbf{U}$ is infinite. Thus the $*$-values are all at least 2. Let S be any nonnegative integer. Divisions with remainder are performed as follows:

$$\begin{array}{ll} \mathbf{S} = \varnothing^* Q_0 + s_0, & 0 \leq s_0 < \varnothing^* \\ Q_0 = (s_0)^* Q_1 + s_1, & 0 \leq s_1 < (s_0)^* \\ \dots & \\ Q_{k-2} = (s_{k-2} \dots s_0)^* Q_{k-1} + s_{k-1}, & 0 \leq s_{k-1} < (s_{k-2} \dots s_0)^* \\ Q_{k-1} = (s_{k-1} \dots s_0)^* Q_k + s_k, & \begin{cases} Q_k = 0 \\ Q_{k-1} < (s_{k-1} \dots s_0)^* \end{cases} \end{array}$$

Since every divisor is, by hypothesis, at least 2, the quotients decrease and finally become zero, say the k^{th} quotient is the first one to be zero, as signalled by Q_{k-1} being less than $(s_{k-1} \dots s_0)^*$. Then the sequence $(s_k \dots s_0)$ belongs to $\mathbf{T}$, and $[s_k \dots s_0] \in [\mathbf{T}]$. The function A^* is defined thus:

$$A^* : W \longrightarrow [\mathbf{T}], \quad A^* S = [s_k \dots s_0].$$

[In case U is of finite height k, A^* is still defined for those integers for which the algorithm terminates in time, that is at step $k-1$ or earlier. The condition that $*$ be never equal to one could be relaxed, so long as it is replaced by some sort of chain condition; e.g. assume that for all infinite sequences n, if for all k, $(n_k \dots n_0) \in \mathbf{T}$, then for all k there exists $k' > k$ such that $(n_{k'} \dots n_0)^* \geqslant 2$.]

C. Relations between A^* and B^*.

LEMMA C.1.—For any umbel, $A^* (B^* [n]) \equiv [n]$, (n in $\mathbf{T}$). Hence the associated division scheme is univalent.

PROOF.—Suppose n belongs to $\mathbf{T}$. The equation $A^* S = [n]$ has a solution, namely $S = B^* [n]$. To see this requires merely an examination of the formula for $B^* [n]$ and of the construction for A^*. The key inequalities follow from criterion iv) for membership in $\mathbf{T}$ (see § 1); the quotients Q_0, Q_1, ... are defined by the equations below, S being $B^* [n]$.

$$\begin{array}{ll} S = \varnothing^* Q_0 + n_0, & n_0 < \varnothing^* \\ Q_0 = (n_0)^* Q_1 + n_1, & n_1 < (n_0)^* \\ \dots & \\ Q_{k-1} = n_k = (n_{k-1} \dots n_0)^* \cdot Q_k + n_k, & n_k < (n_{k-1} \dots n_0)^* \\ Q_k = 0. & \end{array}$$

Thus $A^* (B^* [n]) = [n]$, as asserted.

LEMMA C.2.—For any umbel, if $A^*(S)$ is a twig in it ($A^*(S) \in \mathbf{T}$), then $B^*(A^* S) = S$. In case $\mathbf{U}$ is infinite, the associated division scheme is complete and univalent.

PROOF.—The first claim may be verified by simple computations. Univalence was done in Lemma C.1. As to completeness, for case of an infinite umbel $\mathbf{U}$, the first part suffices since then $A^*(S)$ belongs to $\mathbf{T}$ for all S.

D. Symmetric umbels.

LEMMA D.1. Assume that $\mathbf{U} = \langle \mathbf{T}, * \rangle$ is a symmetric umbel, finite or infinite. The associated division scheme B^* is complete, univalent, and order-preserving in the strong sense: i.e. the S^{th} numeral in $[\mathbf{T}]$ is mapped on S.

PROOF.—Let $n = (n_k \dots n_0)$ be any member of $\mathbf{T}$. Define

$$M_{-1} = \varnothing^*, \quad M^*_{i-1} = (n_{i-1} \ \dots \ n_0)^*, \quad C^*_i = C^*_i(n).$$

These values are independent of n; M^*_{i-1} and C_i^* depend only on i. The sequence, with $D_k^* = M^*_{k-1} - 1$

$$\{ [C^*_k ; \quad D^*_k] \}_{k \geq 0}$$

defines a *gauged scheme* [1], $P : \mathbf{N}^* \longrightarrow W$, where $\mathbf{N}^*$ is the set of all numerals $[n]$ satisfying $0 \leqslant n_i \leqslant D_i^*$, and $P[n] = \Sigma\, n_i C_i^*$. But this is precisely the division scheme associated with $\mathbf{U}$. Utilization of recursion relation iii) of paragraph A permits the calculation, for any $j \geqslant 1$,

$$1 + \sum_{i=0}^{j} C^*_i D_i = 1 + \Sigma\,(C^*_i M_{i-1} - C^*_i) = 1 + \Sigma\,(C^*_{i+1} - C^*_i) =$$
$$= 1 + C^*_{j+1} - C^*_0 = C^*_{j+1} \, .$$

This equality is a necessary and sufficient condition for P, hence also B^*, to be complete and univalent (Theorem 3 of [1]).

The order-preserving property is now proved by induction. That is, for every S in B* [T], the S^{th} numeral in **T** maps on S; to be proved by induction on S. First, it is obvious that $0 = B^*[0]$, and $1 = B^*[1]$, by definition of B*:

$$B^*[0] = 0 \cdot C_0^*, \quad B^*[1] = 1 \cdot C_0^* = 1 \cdot 1 = 1.$$

Assume then that $S \in B^*[\mathbf{T}]$, $S > 1$, and that for all R, if $R \leqslant S$, then R is the image of the R^{th} numeral in [**T**]. Let $B^*[n] = S$:

$$S = B^*[n] = \sum_0^k n_i C_i^*, \quad 0 \leq n_i \leq D.$$

There is an index j such that for all $i < j$, $n_i = D_i$ while $n_j < D_j$ — unless S is the maximum member of B* [**T**] in which case all is done already. By the condition above for completeness and univalence, namely $C_j^* = 1 + \Sigma_{i<j} D_i C_i^*$, the successor $1 + S$ of S is

$$1 + S = n_k C_k^* + \ldots + n_{j+1} C_{j+1}^* + n_j C_j^* + 1 + \sum_{i<j} D_i C_i^* =$$

$$= n_k C_k^* + \ldots + n_{j+1} C_{j+1}^* + (1 + n_j) C_j^* + \sum_{i<j} 0 \cdot C_i^*.$$

Hence $1 + S$ can be written in the form

$$1 + S = \sum_{i=0}^k m_i C_i^* = B^*[m],$$

$$m_i = n_i \text{ for } i \geq j + 1$$

$$m_j = 1 + n_j$$

$$m_i = 0 \text{ for } i < j.$$

This numeral $[m]$ is precisely the successor of n in [**T**]. The induction is complete; so also is the proof of the lemma. The proof contains all but a trivial part of the proof of the following lemma.

LEMMA D.2.—The class of all division schemes associated with symmetric umbels coincides with the class of all complete and univalent gauged schemes.

E. Effects of asymmetry are now studied.

LEMMA E.—Assume $\mathbf{U} = \langle \mathbf{T}, * \rangle$ is an umbel of height at least $k+1$. Assume the subtree $\mathbf{U}_{k-1}$ is symmetric and that $\mathbf{U}_k$ is not symmetric. Assume that n^*, for n of length k, is at least 2. Then in the division scheme associated with $\mathbf{U}_{k+1}$ there is an inversion, i.e. a pair (p, q) of twigs in $\mathbf{T}_{k+1}$ such that $[p] < [q]$ and $B^*[p] > B^*[q]$.

PROOF.—The hypotheses imply that $*$ and C_i^* depend only on length for sequences of length $k-2$ or less, but that among sequences of length $k-1$, $*$ is not constant. If the numerals in $[\mathbf{T}_{k-1}]$ are scanned in increasing order there must appear among the $*$-values, either an upturn or a downturn.

CASE 1.—Assume that there is an upturn but no downturn. Let z be the zero sequence of length $k-1$. For some m of length $k-1$ and larger than z, the value m^* exceeds z^*. Let $m^* = M + 1$; then $M \geqslant z^*$. Consider two new numerals, $[p] \in [\mathbf{T}_k]$ and $[e] \in [\mathbf{T}_{k+1}]$:

$$[p] = [M, m_{k-1} \ldots m_0], \quad [e] = [10 \ldots 0],$$

$[e]$ having $k+1$ zeros. It is easily checked that both $[p]$ and $[e]$ belong to $[\mathbf{T}_{k+1}]$, and that $[p] < [e]$. By application of the recursion relation ii) of § A, $B^*[p]$ and $B^*[e]$ can be expressed as follows ($B^*[z] = 0$):

$$B^*[p] = MC_k^* + B^*[m]$$

$$B^*[e] = (1 \cdot z^* + 0)\, C_k^* + B^*[z] = z^* C_k^* .$$

Since $z^* \leqslant M$, the above relations permit the conclusion:

$$B^*[e] \leq MC_k^* < B^*[p]$$

(note $B^*[m] > 0$). Thus the pair (p, e) is inverted.

CASE 2.—Assume there is a downturn. Then there exists a pair (m, n) of sequences of length $k-1$ in $\mathbf{T}_k$ such that $[n]$ is the successor of $[m]$ while $m^* > n^*$. Since $\mathbf{U}_{k-1}$ is symmetric, Lemma D shows that $B^*[n]$ is the successor of $B^*[m]$: $B^*[n] = 1 + B^*[m]$. Two sequences, p and q, of length $k+1$ in $\mathbf{T}_{k+1}$, and with $[p] < [q]$, are defined as follows.

$$[p] = [10\, m_{k-1} \ldots m_0], \quad [q] = [10\, n_{k-1} \ldots n_0].$$

That these belong to $\mathbf{T}_{k+1}$ depends on the assumption that for any sequence of length k, the $*$-value exceeds one. Hence the admissibility of the «1» in $(k+1)^{st}$ place. By the same recursion relation ii) the relations below are valid:

$$B^*[p] = m^* C_k^* + B^*[m] \geq (1 + n^*) C_k^* + B^*[n] - 1 =$$

$$= n^* C_k^* + B^*[n] + (C_k^* - 1) = B^*[q] + (C_k^* - 1) > B^*[q].$$

The last inequality uses the assumption that ${}_1*$ is not constant on sequences of length $k-1$; this implies that there are at least two such sequences and so for sequences of length $k-2$, the $*$-value is at least two. Hence $C_k^* \geqslant 1$. Thus the pair (p, q) satisfies $[p] < [q]$, $B^*[p] > B^*[q]$ (see Lemma D.1) and again there is an inversion.

Thus in every case an inversion appears and the lemma is proved.

Examination of the proof for Case 1 shows that for $\mathbf{U}_k$, there is a gap in the set $B^*[\mathbf{T}_k]$; in other words, the associated division scheme for $\mathbf{U}_k$ is incomplete.

COROLLARY.—With the hypotheses of Lemma E add the assumption that in $\mathbf{T}_{k-1}$ there exists m with $m^* > z^*$, $z = (0, \ldots, 0)$ being of length

$k-1$. Then the division scheme associated with $\mathbf{U}_k$ is incomplete.

F. Proof of the main theorem completed. Lemmas D.1, D.2 and E settle it.

References

[1] DUBOIS, D. W.: *General Numeration I. Gauged schemes.* This journal.

Department of Mathematics and Statistics
University of New Mexico
Albuquerque NM 87131
USA