THE SYMBOLIC COMPUTATION OF FUNCTIONS OF SEQUENCES OVER FINITE ALPHABETS WITH GIVEN TRANSITION PROBABILITIES BY SEQUENCE LENGTH INDEPENDENT ALGORITHMS

by

D. M. Jackson
E. Horowitz

Technical Report
No. 71-100
June 1971

Department of Computer Science
Cornell University
Ithaca, New York 14850

THE SYMBOLIC COMPUTATION OF FUNCTIONS OF SEQUENCES OVER FINITE ALPHABETS WITH GIVEN TRANSITION PROBABILITIES BY SEQUENCE LENGTH INDEPENDENT METHODS

Section

Abstract

A special case of the problem discussed in this paper occurs in connection with non-statistical classification and is introduced from this point of view. The special case concerns the computation of expectations of statistical functions of the "distance" between pairs of fixed length sequences over a binary alphabet with given a priori state transition probabilities. The general problem involves an extension to alphabets of arbitrary order and the comparison of an arbitrary number of fixed length sequences. Given a set of sequences, it is shown that for a large class of functions exact computation may be carried out by an algorithm whose computation time is independent of the length of the sequences. It is further shown that results for all functions of this class may be derived from a small number of basis functions. Two methods for computing basis functions are given. Basis functions for the commonly encountered special case involving pairs of binary sequences are given explicitly.

1. Introduction

The work presented here was motivated by a problem occurring in non-statistical classification. Accordingly, a brief description of the problem is now given.

Non-statistical classification deals with the construction of classifications of populations of objects when an adequate statistical model for the population is not known (Good [1], Nagy [2], Friedman & Rubin [3], Zadeh [4], Cover & Hart [5], Cover [6], inter al.). The objective of classification in this case is to obtain a decomposition of the population which will reveal its structure, and to suggest further models for a more sophisticated treatment of the material. The role of such techniques in applications to taxonomy, for example, is described in detail by Sokal and Sneath [7].

In a typical situation the objects of a population are characterized by a set of vectors whose components specify the value or state of a continuous or discrete attribute. The vectors are of fixed length, the latter being the number of attributes which are measured or observed. A classification algorithm develops classes of objects which are similar to each other and yet well-differentiated from members of other classes. Similarity may be regarded as a distance function on the attribute space.

An acceptable classification must be stable with respect to small changes in the characterizations of the objects since instability of a classification indicates, in the first instance, that the objects of the population have been inadequately characterized. Thus, we necessarily require that the classification not be grossly affected by occassional errors in the measuring or recording of the states of the attributes in the objects.

The effect of errors in the object characterizations is to introduce an "uncertainty" into the exact computation of the distance between pairs of objects. The estimation of these uncertainties for all pairs of objects of the population is of value in determining the stability of classifications generated by the class of algorithms which utilize a matrix of inter-object distances.

The purpose of this paper is twofold. First, to determine exactly in an analytic form the expectations of statistical functions of the distance between pairs of objects when observations on discrete attributes are susceptible to error independently. This includes the computation of the uncertainties in distance as a special case. Second, to determine, in a form convenient for subsequent numerical computation, explicit expressions for the expectations of statistical functions for the important special case concerning pairs of aequences over the binary alphabet (0,1) . Expressions will be derived for arbitrary state transition probabilities so that specialization to a particular case, for example equiprobability of error, is immediate.

The computation of these quantities is generally avoided because of the extreme combinatorial nature of the problem. The problem is particularly acute when the number of attributes is too small to derive sufficiently accurate approximations using the central limit theorem. In this paper it is shown that, for a reasonably broad class of distance functions which includes the Hamming distance function, the computing time for determining expectations of statistical functions of distance can be carried out very fast. More precisely, if the distance function is representable as a multivariate polynomial then the dependence of computation time on the number of attributes is reduced from polynomial to constant. Moreover, each expectation is expressible in terms of a small number of basis functions. Since the latter are multivariate polynomials in the a priori transition probability matrix and the arguments of the distance functions, the determination for different error models or for different pairs of objects is reduced to direct polynomial evaluation. For reference purposes, the algebraic forms of these basis functions are given in Appendix II for the commonly encountered special case involving pairs of fixed length sequences over the binary alphabet. The reductions obtained here are sufficiently substantive to enable these functions to be computed manually with relative ease, whereas previously it was unfeasible to calculate them at all.

The ability to perform by computer complex algebraic manipulations has been important in the development of the work. The SAC-1 system (Horowitz [8]) is designed specifically for performing a broad class of operations on multivariate polynomials and rational functions. Although explicit mention of the use of SAC-1 is not given in the remainder of the paper, it must be stated that the empirical results obtained with its use have led to a number of the theorems in Sections 3 and 4.

Section 2 contains a definition of the general problem, a special case of which is the comparison of pairs of binary sequences. The computational order of an algorithm for solving the problem is determined and is found to be polynomial in the length of the sequences. Section 3 shows that the computational order of the process is independent of the length of the sequences for a class of distance functions which includes Hamming distance. In addition it is shown that a solution of the problem may be given in terms of the moments and covariances of the error distribution and that these may be reduced to multivariate polynomials in the arguments of the distance function and the a priori state transition probability matrix. Section 4 gives a number of recurrence relations for reducing the computation time further. A characterization of the moments and covariance of arbitrary order is given in terms of the basis functions. Examples of the application of these recurrence relations are given in Section 5, together with an explanation of the basis functions of Appendix II.

2. The General Problem

The general problem may be characterized as follows. Suppose that $\mathcal{U}$ is a given set of sequences of length N composed of symbols from an alphabet $\mathcal{A}$ of order k, and that $\mathcal{W}$ is an arbitrary subset of $\mathcal{U}$ of given cardinality w. $\Lambda = \{u_i\}$ is a given partition, into λ parts, of the set $\mathcal{X} \equiv \{x_i\}$ of all ordered sequences of length w composed of symbols from $\mathcal{A}$. $\underline{n} = \underline{u}(\mathcal{W})$ is a vector whose i-th component gives the number of symbol positions, j, in which the ordered sequence of j-th symbols, taken one from each of the sequences of $\mathcal{W}$ in order, belong to u_i. $f(\underline{n})$ is a given real valued function (distance function) of $\underline{n}$ and ϕ is a statistical function of f. $\underset{\sim}{\Phi}$ is a λ dimensional array, giving the expectations of ϕ for all subsets of $\mathcal{U}$ of cardinality w, when the symbols of the sequences are susceptible to error independently with given a priori probabilities.

For brevity we shall refer to the construction of $\underset{\sim}{\Phi}$ as the (w,k,N) - problem, since these variables characterize the problem. The (w,k,N)-problem has λ variables, so when the number of variables is of consequence we shall refer to the (w,k,N)-problem alternatively as the λ-variable problem. In practice, the $(2,2,N)$-problem, involving the comparison of pairs of binary sequences is the most frequently encountered. The problem arises in a variety of circumstances. Message transmission in discrete noisy channels, matching problem

(Barton [9]) in which the matches may be incorrectly recognized and applications of nonstatistical classification (Jackson [10]) are examples.

To simplify the subsequent analysis, a convention is adopted for abbreviating power products, products of multinomial coefficients and products of summation operators. The convention is summarized in Appendix I.

To examine the combinatorial aspects of the general problem, we begin by considering the matrix $\underset{\sim}{H}$, where $H_{ij} = \text{Prob}(a_i|a_j)$, denoting the given a priori transition probabilities between the symbols a_i and a_j of the alphabet $\mathcal{A}$. If $x^{(p)}$ and $x^{(q)}$ are two arbitrary w-tuples of $\mathcal{X}$ whose i-th symbols are $x_i^{(p)}$ and $x_i^{(q)}$ respectively, and if $\underset{\sim}{H}^{(0)}$ is the transition probability matrix for pairs of w-tuples in $\mathcal{X}$ then:

$$H_{pq}^{(0)} = \text{Prob}(x^{(p)}|x^{(q)}) = \prod_{i=1}^{k} \text{Prob}(x_i^{(p)}|x_i^{(q)}) \tag{2.1}$$

Let $\underset{\sim}{H}^{(\Lambda)}$ be the $\lambda \times \lambda$ transition probability matrix for pairs of parts u_i, u_j of the partition Λ. Then:

$$H_{ij}^{(\Lambda)} = \text{Prob}(u_i|u_j) = \frac{1}{|u_j|} \cdot \sum_{\begin{cases}(x^{(p)}, x^{(q)}) \\ (x^{(p)}, x^{(q)}) \varepsilon u_i \times u_j\end{cases}} H_{pq}^{(0)} . \tag{2.2}$$

For brevity, the matrix $\underset{\sim}{H}^{(\Lambda)}$ will be denoted by $\underset{\sim}{P}$. $\underset{\sim}{P}$ is column stochastic,

$$\text{i.e.} \quad \underline{\rho}(\underset{\sim}{P}) = \underline{1} \tag{2.3}$$

where $\underline{1}$ is the vector with λ components, each of which is equal to one. Since the errors occur independently, it follows that the expectation of $\phi(f)$ is given by

$$\phi^*(\underline{n}) = \sum_{\left\{\substack{\underset{\sim}{I} \\ \underline{\rho}(\underset{\sim}{I})=\underline{n}}\right.} \left[\frac{\underline{n}}{\underset{\sim}{I}} \right]_\rho \cdot \underset{\sim}{P}^{\underset{\sim}{I}} \cdot \psi(\underline{m},\underline{n}) \tag{2.4}$$

where i) $\underline{m} = \underline{\sigma}(\underset{\sim}{I})$ ii) $\psi(\underline{m},\underline{n}) = \phi(f(\underline{m}),f(\underline{n}))$, and iii) $\underset{\sim}{I}$ is a $\lambda \times \lambda$ matrix with non-negative integer elements giving the number of transitions from u_j to u_i . The error matrix is accordingly given by $\{\phi^*(\underline{n}) \mid \sigma(\underline{n}) = N\}$ and is denoted by $\underset{\sim}{\Phi}$. Since the (w,k,N)-problem is a λ variable problem and the matrices $\underset{\sim}{I}$ are $\lambda \times \lambda$, then the number of matrices $\underset{\sim}{I}$ whose elements sum to N is $\binom{N+\lambda^2-1}{\lambda^2-1}$. The computation of $\underset{\sim}{\Phi}$ by (2.4) is therefore $O(N^{\lambda^2-1})$.

To illustrate the use of a partition Λ we shall consider the $(2,2,N)$-problem with the additional conditions that errors occur equiprobably with probability r , and that the distance function is independent of the order in which the sequences A and B are compared. Accordingly, for equiprobability of error:

$$\underset{\sim}{H} = \begin{bmatrix} s & r \\ r & s \end{bmatrix} \text{ where } s = 1 - r$$

Since pairs of binary vectors are involved, the set $\mathscr{X}$ is given by

$$\mathscr{X} \equiv (x_1,x_2,x_3,x_4) \equiv ((0,0),(0,1),(1,0),(1,1)) \text{ .}$$

The symmetry requirement for the distance functions induces the partition Λ on $\mathscr{X}$ given by:

$$\Lambda \equiv (u_1,u_2,u_3) \equiv ((x_1),(x_2,x_3),(x_4))$$

where $\lambda = 3$. It follows from (2.2) that

$$\underset{\sim}{H}^{(\Lambda)} = \begin{bmatrix} s^2 & rs & r^2 \\ 2rs & r^2+s^2 & 2rs \\ r^2 & rs & s^2 \end{bmatrix} \text{ .}$$

The time required for computing $\underset{\sim}{\Phi}$ for the (2,2,N)-problem is therefore $O(N^8)$. For the (2,k,N)-problem with a symmetrical distance function, computation time is $O(N^{\frac{1}{4}k^2(k+1)^2-1})$. This increases exponentially with the order k of the alphabet. It has already been stated in Section 1 that sufficient reduction for a certain class of distance functions has been achieved by the techniques to be presented here. Indeed the (2,2,N)-problem will be reduced to a condition in which it may be solved manually with relative ease.

It has been shown (Jackson & White [11]) for all distance functions that the order of computation may be reduced provided that each element of $\underset{\sim}{\Phi}$ is required only to within a prescribed accuracy ε. An upper estimate on the order of computation in this case is $N^2 k_\varepsilon^6$, where $k_\varepsilon << 2N$ and k_ε is determined from ε. Proofs of the convergence of the numerical method for the (2,2,N)-problem are given by Jackson & White [12]. Complete recomputation of $\underset{\sim}{\Phi}$ is, however, necessary if the length N, of the sequences, the statistical function ϕ, the distance function f or the a priori transition probability matrix $\underset{\sim}{H}$ are changed. In practice it is commonly necessary to compute $\underset{\sim}{\Phi}$ for a variety of N, ϕ, f and $\underset{\sim}{H}$. In Section 3 it is shown that repetition of the computation for changes in these quantities can be eradicated for a particular class of distance functions. In addition it is shown that the computation time for a single element of $\underset{\sim}{\Phi}$ is independent of N in this case.

3. Analysis for a Class of Distance Functions

Suppose that ϕ and f have the property that $\phi(f)$ belong to the class of multivariate polynomials of degree Q in λ variables. Accordingly, $\psi(\underline{m},\underline{n})$ may be expressed in the form:

$$\psi(\underline{m},\underline{n}) = \sum_{\underline{i}} a_{\underline{i}}(\underline{n}) \cdot \underline{m}^{\underline{i}} \tag{3.1}$$

The values of $a_{\underline{i}}(\underline{n})$ and the admissible values of $\underline{i}$ are determinable from the given functions ϕ and f by equating coefficients of $\underline{m}^{\underline{i}}$. Each of the vectors $\underline{i}$, $\underline{m}$, $\underline{n}$ has λ components, and $\sigma(\underline{i}) \leq Q$.

For the (w,k,N)-problem, the population $\mathcal{U}$ is embedded in the attribute space of dimension N. f is a generalized distance function on this space and is defined only at the lattice points since the attributes are discrete. For each ordered set of w points of $\mathcal{U}$ there corresponds a single point $\underline{n}$ in the event space whose dimension is $\lambda - 1$. The mapping of subsets of cardinality w in the attribute space to the lattice points of the event space is many-one. With each lattice point in the event space is associated a conditional probability distribution, $P(\underline{m}|\underline{n})$, giving the probability of the event $\underline{m}$ conditional on the event $\underline{n}$.

Since $\phi^*(\underline{n}) = \mathcal{E}(\psi(\underline{m},\underline{n})|\underline{n})$ and $\psi(\underline{m},\underline{n})$ is a linear combination of the random variables $\underline{m}^{\underline{i}}$ then:

$$\phi^*(\underline{n}) = \sum_{\underline{i}} a_{\underline{i}}(\underline{n}) \cdot B_{\underline{i}}(\underline{n},\underset{\sim}{P}) \tag{3.2}$$

where $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ is identified as the $\underline{i}$-th covariance, $\mathcal{E}(\underline{m}^{\underline{i}}|\underline{n})$, of the distribution $P(\underline{m}|\underline{n})$. The remainder of this section is concerned with determining the quantities $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$, for the λ variable case, as multivariate polynomials in $\underline{n}$ and $\underset{\sim}{P}$ and to provide practicable methods for their construction.

Definition 3.1: $G_{ij} \equiv P_{ij} \frac{\partial}{\partial P_{ij}}$ where $\frac{\partial}{\partial P_{ij}}$ denotes formal partial differentiation with respect to P_{ij} with the provision that $\{P_{ij}\}$ is treated as a set of independent variables.

The following theorem gives a simple form for $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ in terms of the operator $\underset{\sim}{G}$. Since the elements of $\underset{\sim}{P}$ are treated as independent variables when they appear as operands of $\underset{\sim}{G}$, further reduction using the column stochastic property of $\underset{\sim}{P}$, namely $\underline{\rho}(\underset{\sim}{P}) = \underline{1}$, is not permissible.

Theorem 3.1: If $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ is the $\underline{i}$-th covariance of the distribution $P(\underline{m}|\underline{n})$, where $\sigma(\underline{m}) = \sigma(\underline{n}) = N$, then

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = (\underline{\sigma}(\underset{\sim}{G}))^{\underline{i}}(\underline{\rho}(\underset{\sim}{P}))^{\underline{n}}$$

Proof: Since $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ is the $\underline{i}$-th covariance then:

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum_{\substack{\underline{m} \\ \{\sigma(\underline{m})=N}} \underline{m}^{\underline{i}} \cdot P(\underline{m}|\underline{n})$$

Now from (2.4) the distribution is given by:

$$P(\underline{m}|\underline{n}) = \begin{bmatrix} \underline{n} \\ \underset{\sim}{I} \end{bmatrix}_{\rho} \cdot \underset{\sim}{P}^{\underline{I}} \quad \text{where} \quad \underline{\sigma}(\underset{\sim}{I}) = \underline{m} \quad \text{and} \quad \underline{\rho}(\underset{\sim}{I}) = \underline{n} \; .$$

Thus, since $\underline{m}^{\underline{i}} = (\underline{\sigma}(\underset{\sim}{I}))^{\underline{i}}$, then substitution into the expression for $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ yields:

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum_{\substack{\underset{\sim}{I} \\ \{\underline{\rho}(\underset{\sim}{I}) = \underline{n}}} (\underline{\sigma}(\underset{\sim}{I}))^{\underline{i}} P(\underline{m}|\underline{n})$$

But from Definition 3.1 we have:

$$\underset{\sim}{G} \cdot P(\underline{m}|\underline{n}) = \underset{\sim}{I} \cdot P(\underline{m}|\underline{n})$$

whence

$$\begin{aligned} B_{\underline{i}}(\underline{n},\underset{\sim}{P}) &= (\underline{\sigma}(\underset{\sim}{G}))^{\underline{i}} \sum_{\substack{\underset{\sim}{I} \\ \{\underline{\rho}(\underset{\sim}{I}) = \underline{n}}} P(\underline{m}|\underline{n}) \\ &= (\underline{\sigma}(\underset{\sim}{G}))^{\underline{i}} \cdot (\underline{\rho}(\underset{\sim}{P}))^{\underline{n}} \end{aligned}$$

QED

An explicit multivariate polynomial expression for $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ is now derived by utilizing the result of Theorem 3.1 and by applying each term of $(\underline{\sigma}(\underset{\sim}{G}))^{\underline{i}}$ to $(\underline{\rho}(\underset{\sim}{P}))^{\underline{n}}$. The following lemma is used:

<u>Lemma 3.1</u>: If $S_i^{(n)}$ is a Stirling number of the second kind then the following operator identity holds:

$$\left(x\frac{d}{dx}\right)^n \equiv \sum_{i=1}^{n} s_i^{(n)} x^i \frac{d^i}{dx^i}\ ; \quad n \neq 0\ .$$

<u>Proof</u>: See, for example, Riordan [13].

<u>Theorem 3.2</u> (Main theorem): If $(n)_i = n(n-1)\ldots(n-i+1)$ (the falling factorial) then

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum_{\substack{\underset{\sim}{\gamma} \\ \{\theta(\underline{i}) \le \underline{\sigma}(\underset{\sim}{\gamma}) \le \underline{i}}} (\underline{n})_{\underline{\rho}(\underset{\sim}{\gamma})} \cdot \underset{\sim}{P}^{\underset{\sim}{\gamma}} \cdot C(\underset{\sim}{\gamma},\underline{i})$$

where $\theta(x) = 0$ if $x = 0$, and 1 otherwise (θ operates component by component in the case of a vector) and

$$C(\underset{\sim}{\gamma},\underline{i}) = \sum_{\substack{\underset{\sim}{\alpha} \\ \left\{\begin{smallmatrix} \underline{\sigma}(\underset{\sim}{\alpha}) = \underline{i} \\ \underset{\sim}{\alpha} \ge \underset{\sim}{\gamma} \end{smallmatrix}\right.}} \begin{bmatrix} \underline{i} \\ \underset{\sim}{\alpha} \end{bmatrix}_{\sigma} s_{\underset{\sim}{\alpha}}^{(\underset{\sim}{\gamma})} \quad \text{for} \quad \underline{\sigma}(\underset{\sim}{\gamma}) \le \underline{i}\ .$$

<u>Proof</u>: From Definition 3.1 and Lemma 3.1 we have:

$$G_{jk}^{q} = \sum_{i=\theta(q)}^{q} s_i^{(q)} P_{jk}^{i} D_{jk}^{i}\ , \quad G_{jk}^{o} = 1\ .$$

Since the elements of $\underset{\sim}{P}$ are treated as independent variables when they appear as operands of $\underset{\sim}{G}$ we have:

$$G_{jk}^{q}(\rho_k(\underset{\sim}{P}))^{p} = \sum_{i=\theta(q)}^{q} (p)_i P_{jk}^{i}(\rho_k(\underset{\sim}{P}))^{p-i} s_i^{(q)}$$

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \Big(\sum_{\substack{\underset{\sim}{\alpha} \\ \{\underline{\sigma}(\underset{\sim}{\alpha}) = \underline{i}}} \begin{bmatrix} \underline{i} \\ \underset{\sim}{\alpha} \end{bmatrix}_{\sigma} \cdot \underset{\sim}{G}^{\underset{\sim}{\alpha}})(\underline{\rho}(\underset{\sim}{P}))^{\underline{n}}$$

When the operators of $\underline{G}$ are assembled with each of the columns of $\underline{P}$ to which they apply then:

$$B_{\underline{i}}(\underline{n},\underline{P}) = \sum_{\{\underline{\sigma}(\underline{\alpha})=\underline{i}} \left[\frac{\underline{i}}{\underline{\alpha}}\right]_\sigma \prod_{j=1}^{\lambda} \{\underline{G}_j^{\underline{\alpha}_j}(\rho_j(\underline{P}))^{n_j}\}$$

Application of the expansion for G_{jk}^{q} yields:

$$B_{\underline{i}}(\underline{n},\underline{P}) = \sum_{\{\underline{\sigma}(\underline{\alpha})=\underline{i}} \left[\frac{\underline{i}}{\underline{\alpha}}\right]_\sigma \prod_{j=1}^{\lambda} \{ \sum_{\underline{\gamma}_j=\theta(\underline{\alpha}_j)}^{\underline{\alpha}_j} (n_j)_{\rho_j(\underline{\gamma})} \cdot S_{\underline{\alpha}_j}^{(\underline{\gamma}_j)} \cdot \underline{P}_j^{\underline{\gamma}_j}(\rho_j(\underline{P}))^{n_j-\rho_j}$$

where $\underline{\alpha}$, $\underline{\gamma}$ and $\underline{P}$ are conformable matrices. Now, for an arbitrary function $g(x)$ the following identity holds:

$$\prod_{j=1}^{\lambda} \left(\sum_{x_j=a_j}^{b_j} g(x_j)\right) \equiv \left(\prod_{j=1}^{\lambda} \sum_{x_j=a_j}^{b_j} \right)\left(\prod_{k=1}^{\lambda} g(x_k)\right)$$

Thus:

$$B_{\underline{i}}(\underline{n},\underline{P}) = \sum_{\{\underline{\sigma}(\underline{\alpha})=\underline{i}} \left[\frac{\underline{i}}{\underline{\alpha}}\right]_\sigma \left(\prod_{j=1}^{\lambda} \sum_{\underline{\gamma}_j=\theta(\underline{\alpha}_j)}^{\underline{\alpha}_j} \right)\left(\prod_{j=1}^{\lambda} \{(n_j)_{\rho_j(\underline{\gamma})} \cdot S_{\underline{\alpha}_j}^{(\underline{\gamma}_j)} \cdot \underline{P}_j^{\underline{\gamma}_j}(\rho_j(\underline{P}))^{n_j-\rho_j(\underline{\gamma})}\}\right)$$

which may be rewritten as

$$B_{\underline{i}}(\underline{n},\underline{P}) = \sum_{\{\underline{\sigma}(\underline{\alpha})=\underline{i}} \left[\frac{\underline{i}}{\underline{\alpha}}\right]_\sigma \sum_{\{\theta(\underline{\alpha})\leq\underline{\gamma}\leq\underline{\alpha}} (\underline{n})_{\underline{\rho}(\underline{\gamma})} \cdot S_{\underline{\alpha}}^{(\underline{\gamma})} \cdot \underline{P}^{\underline{\gamma}} \cdot (\underline{\rho}(\underline{P}))^{\underline{n}-\underline{\rho}(\underline{\gamma})}$$

Reversal of the order of summation and utilization of the column stochastic property of $\underset{\sim}{P}$, the latter step valid since the operator $\underset{\sim}{G}$ has been eliminated, yields:

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum_{\substack{\underset{\sim}{\gamma} \\ \{\theta(\underline{i}) \le \underline{\sigma}(\underset{\sim}{\gamma}) \le \underline{i}}} (\underline{n})_{\underline{\rho}(\underset{\sim}{\gamma})} \cdot \underset{\sim}{P}^{\underset{\sim}{\gamma}} \cdot C(\underset{\sim}{\gamma},\underline{i})$$

where the constants $C(\underset{\sim}{\gamma},\underline{i})$ are defined above.

QED

__Remark__: It follows from Theorem 3.2 that the computation time for $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ is independent of N .

The computation time for determining the constants $C(\underset{\sim}{\gamma},\underline{i})$ defined in Theorem 3.2 may be reduced by the simplification given in the following lemma.

__Lemma 3.2__: For the constants defined in Theorem 3.2 then:

$$C(\underset{\sim}{\gamma},\underline{i}) = \begin{bmatrix} \underline{\sigma}(\underset{\sim}{\gamma}) \\ \underset{\sim}{\gamma} \end{bmatrix}_{\sigma} S_{\underline{i}}^{(\underline{\sigma}(\underset{\sim}{\gamma}))}$$

__Proof__: From Theorem 3.2 we have:

$$C(\underset{\sim}{\gamma},\underline{i}) = \sum_{\substack{\underset{\sim}{\alpha} \\ \underline{\sigma}(\underset{\sim}{\alpha}) = \underline{i} \\ \underset{\sim}{\alpha} \ge \underset{\sim}{\gamma}}} \begin{bmatrix} \underline{i} \\ \underset{\sim}{\alpha} \end{bmatrix}_{\sigma} S_{\underset{\sim}{\alpha}}^{(\underset{\sim}{\gamma})} \quad \text{for } \underline{\sigma}(\underset{\sim}{\gamma}) \le \underline{i}$$

Since the summations over the rows of $\underset{\sim}{\alpha}$ are independent (no column-binding in $\underset{\sim}{\alpha}$) then $C(\underset{\sim}{\gamma},\underline{i})$ may be decomposed as follows:

$$C(\underset{\sim}{\gamma},\underline{i}) = \prod_{j=1}^{\lambda} C^*({}_j\underset{\sim}{\gamma},i_j)$$

where $C^*(\underline{a},p) = \sum_{\left\{\substack{\underline{x} \\ \sigma(\underline{x})=p \\ \underline{x}\geq\underline{a}}\right.} \begin{bmatrix} p \\ \underline{x} \end{bmatrix}_\sigma S_{\underline{x}}^{(\underline{a})}$ for $\sigma(\underline{a}) \leq p$

Now the generating function for Stirling numbers of the second kind, $S_n^{(m)}$, is:

$$(e^x - 1)^m = m! \sum_{n=m}^{\infty} S_n^{(m)} \frac{x^n}{m!} , \quad n \geq m \geq 1$$

Now $(e^x - 1)^{\sigma(\underline{a})} = \prod_{i=1}^{\lambda} (e^x - 1)^{a_i}$. Expanding both sides of this identity using the generating function, and equation coefficients of x^p, we obtain:

$$C^*(\underline{a},p) = \begin{bmatrix} \sigma(\underline{a}) \\ \underline{a} \end{bmatrix}_\sigma S_p^{(\sigma(\underline{a}))}$$

whence, $C(\underset{\sim}{\gamma},\underline{i}) = \begin{bmatrix} \sigma(\underset{\sim}{\gamma}) \\ \underset{\sim}{\gamma} \end{bmatrix}_\sigma S_{\underline{i}}^{(\sigma(\underset{\sim}{\gamma}))}$

QED

The following corollary gives a simililfied form of $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ as a multivariate polynomial in $\underline{n}$ and $\underset{\sim}{P}$ with constants which may be determined directly by table look-up, rather than by computation. A convenient tabulation of Stirling Numbers and Multinomial Coefficients is given by Abramovitz and Segun [14].

<u>Corollary 3.1</u>: $B_{\underline{i}}(\underline{n},\underline{P}) = \sum\limits_{\{\theta(\underline{i})\leq\underline{\sigma}(\underline{\gamma})\leq\underline{i}}^{\underline{\gamma}} (\underline{n})_{\underline{\rho}(\underline{\gamma})}\cdot\underline{P}^{\underline{\gamma}}\begin{bmatrix}\underline{\sigma}(\underline{\gamma})\\ \underline{\gamma}\end{bmatrix}_{\sigma}\cdot S_{\underline{i}}^{(\underline{\sigma}(\underline{\gamma}))}$

<u>Proof</u>: Direct from Theorem 3.2 and Lemma 3.2.

An expression for the i-th moment of the distribution $P(m|\underline{n})$ can be derived directly by specialization of Corollary 3.1. Denoting the i-th moment by $M_j^{(i)}(\underline{n},\underline{P})$ then:

$$M_j^{(i)}(\underline{n},\underline{P}) = \sum\limits_{\{\theta(\underline{i})\leq\sigma(\underline{\gamma})\leq i}^{\underline{\gamma}} (\underline{n})_{\underline{\gamma}}\cdot {}_j\underline{P}^{\underline{\gamma}}\begin{bmatrix}\sigma(\underline{\gamma})\\ \underline{\gamma}\end{bmatrix}_{\sigma} S_i^{(\sigma(\underline{\gamma}))}$$

Further specialization to the univariate case of a single variable distributed binomially yields:

$$m_s = \sum_{x=1}^{s} (n)_x\, p^x\, S_s^{(x)}$$

which is an expression for the s-th moment derived by Riordan [15].

4. Recurrence Relations and Methods of Expansion

A recurrence relation between the $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ is now derived with a view to decreasing the computational order of the problem yet further. The following definition is needed:

Definition 4.1:

i) The substitution convention $\mathcal{C}(x)$ is defined by

a) $\mathcal{C}(x) : x^i \to (x)_i$; i.e. $(x^i)\big|_{\mathcal{C}(x)} = (x)_i$

b) $\mathcal{C}^{-1}(x) : (x)_i \to x^i$

ii) The simultaneous application of conventions $\mathcal{C}(x)$ and $\mathcal{C}(y)$ is denoted by $\mathcal{C}(x,y)$ or $\mathcal{C}(\underline{z})$ where $\underline{z} = (x,y)$

iii) The sequential application of conventions $\mathcal{C}(x)$ and $\mathcal{C}(y)$ from left to right is denoted by $\mathcal{C}(x), \mathcal{C}(y)$.

iv) Multiplication with respect to substitution convention $\mathcal{C}(x)$ is denoted by $\times_{\mathcal{C}(x)}$. Let $g_1(x)$ and $g_2(x)$ be two arbitrary functions of x . Then

$$g_1(x) \times_{\mathcal{C}(x)} g_2(x) \equiv ((g_1(x))\big|_{\mathcal{C}^{-1}(x)} \times (g_2(x))\big|_{\mathcal{C}^{-1}(x)})\big|_{\mathcal{C}(x}$$

v) The well-known pair of inverses (see Riordan [13]):

a) $x^n = \sum_{i=1}^{n} S_i^{(n)}(x)_i$ where $S_i^{(n)}$ are Stirling numbers of the second kind;

b) $(x)_n = \sum_{i=1}^{n} s_i^{(n)} x^i$ where $s_i^{(n)}$ are Stirling numbers of the first kind;

are algebraic devices for manipulating expressions into a form suitable for transformation under one of the conventions $\mathcal{C}(x)$, $\mathcal{C}^{-1}(x)$.

<u>Theorem 4.1</u> (Recurrence relations):

i) $B_{\underline{i}}(\underline{n},\tilde{P}) = (\tilde{P}\cdot\underline{n})^{\underline{i}}\Big|_{\mathcal{C}^{-1}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}(\underline{n})}$

ii) $B_{\underline{i}+\underline{i}'}(\underline{n},\tilde{P}) = B_{\underline{i}}(\underline{n},\tilde{P}) \times_{\mathcal{C}^{-1}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}(\underline{n})} B_{\underline{i}'}(\underline{n},\tilde{P})$

<u>Proof</u>: i) From Corollary 3.1:

$$B_{\underline{i}}(\underline{n}.\tilde{P}) = \sum_{\theta(\underline{i})\le\underline{j}\le\underline{i}} s_{\underline{i}}^{(\underline{j})} \sum_{\underline{\sigma}(\gamma)=\underline{j}} (\underline{n})_{\underline{\rho}(\tilde{\gamma})} \cdot \tilde{P}^{\tilde{\gamma}}\begin{bmatrix}\underline{\sigma}(\tilde{\gamma})\\ \tilde{\gamma}\end{bmatrix}$$

Thus, from Definition 4.1(i)(a):

$$B_{\underline{i}}(\underline{n},\tilde{P}) = \sum_{\theta(\underline{i})\le\underline{j}\le\underline{i}} s_{\underline{i}}^{(\underline{j})} \Big(\sum_{\underline{\sigma}(\tilde{\gamma})=\underline{j}} (\underline{n})^{\underline{\rho}(\tilde{\gamma})} \cdot \tilde{P}^{\tilde{\gamma}} \cdot \begin{bmatrix}\underline{\sigma}(\tilde{\gamma})\\ \tilde{\gamma}\end{bmatrix}\Big)\Big|_{\mathcal{C}(\underline{n})}$$

$$= \sum_{\theta(\underline{i})\le\underline{j}\le\underline{i}} s_{\underline{i}}^{(\underline{j})}((\tilde{P}\cdot\underline{n})^{\underline{j}})\Big|_{\mathcal{C}(\underline{n})}$$

$$= \sum_{\theta(\underline{i})\le\underline{j}\le\underline{i}} s_{\underline{i}}^{(\underline{j})} (\tilde{P}\cdot\underline{n})_{\underline{j}}\Big|_{\mathcal{C}^{-1}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}(\underline{n})}$$

$$= ((\tilde{P}\cdot\underline{n})^{\underline{i}})\Big|_{\mathcal{C}^{-1}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}(\underline{n})}$$

ii) From (i): $B_{\underline{i}+\underline{i}'}(\underline{n},\tilde{P}) = ((\tilde{P}\cdot\underline{n})^{\underline{i}} \times (\tilde{P}\cdot\underline{n})^{\underline{i}'})\Big|_{\mathcal{C}^{-1}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}(\underline{n})}$

$$= ((B_{\underline{i}}(\underline{n},\tilde{P}))\Big|_{\mathcal{C}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}^{-1}(\underline{n})} \times (B_{\underline{i}'}(\underline{n},\tilde{P}))\Big|_{\mathcal{C}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}^{-1}(\underline{n})}\Big)\Big|_{\mathcal{C}^{-1}(\tilde{P}\cdot\underline{n}),\,\mathcal{C}}$$

Thus from Definition 4.1(iv):

$$B_{\underline{i}+\underline{i}'}(\underline{n},\underset{\sim}{P}) = B_{\underline{i}}(\underline{n},\underset{\sim}{P}) \times_{\mathcal{C}^{-1}(\underset{\sim}{P}\cdot\underline{n}),\,\mathcal{C}(\underline{n})} B_{\underline{i}'}(\underline{n},\underset{\sim}{P})$$

QED

The following theorem provides a means for computing $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ by a set of recurrence relations derivable from Theorem 4.1.

<u>Theorem 4.2</u> (Recurrence relations):

If $\underline{Q}^j = (0,0,\ldots,0,i_j,0,\ldots,0)$ (j-th component nonzero) where i_j is the j-th component of $\underline{i}$, and if $\underline{1}^j$ is a vector each of whose components is zero except the j-th which is one, th

i) $B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \left(\prod_{j=1}^{\lambda}{}_{\mathcal{C}(\underline{n})} B_{\underline{Q}^j}(\underline{n},\underset{\sim}{P})\right)$

ii) $B_{\underline{Q}^j}(\underline{n},\underset{\sim}{P}) = \sum_{k=1}^{i_j} S_k^{(i_j)} \left(B^k_{\underline{1}^j}(\underline{n},\underset{\sim}{P})\right)\Big|_{\mathcal{C}(\underline{n})}$

<u>Proof</u>:

i) Consider the change of variable $\sum_{q=1}^{\lambda} P_{pq} n_q = x_p$

Then from Theorem 4.1(i) we have:

$$B_{\underline{Q}^j}(\underline{n},\underset{\sim}{P}) \times_{\mathcal{C}^{-1}(\underset{\sim}{P}\cdot\underline{n}),\mathcal{C}(\underline{n})} B_{\underline{Q}^k}(\underline{n},\underset{\sim}{P}) = (x_j^{i_j} \times_{\mathcal{C}^{-1}(x_j,x_k),\mathcal{C}(\underline{n})} x_k^{i_k})$$

$$= \{(x_j^{i_j})\Big|_{\mathcal{C}^{-1}(x_j,x_k)} \times_{\mathcal{C}(\underline{n})} (x_k^{i_k})\Big|_{\mathcal{C}^{-1}(x_j,x_k)}\} \quad \text{if} \quad j \neq k$$

$$= (B_{\underline{Q}^j}(\underline{n},\underset{\sim}{P}) \times_{\mathcal{C}(\underline{n})} B_{\underline{Q}^k}(\underline{n},\underset{\sim}{P}))$$

Thus, by repeated application of this result, and from Theorem 4.1(ii), we have

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = (\prod_{j=1}^{\lambda}{}_{\mathcal{C}(\underline{n})} B_{\underline{Q}^j}(\underline{n},\underset{\sim}{P}))$$

ii) From Theorem 4.1(i):

$$B_{\underline{Q}^j}(\underline{n},\underset{\sim}{P}) = (x_j^{i_j})\Big|_{\mathcal{C}^{-1}(x_j),\mathcal{C}(\underline{n})} = (\sum_{k=0}^{i_j} s_k^{(i_j)} (x_j)_k)\Big|_{\mathcal{C}^{-1}(x_j),\mathcal{C}(\underline{n})}$$

$$= (\sum_{k=0}^{i_j} s_k^{(i_j)} (x_j)^k)\Big|_{\mathcal{C}(\underline{n})} = \sum_{k=0}^{i_j} s_k^{(i_j)} (B^k_{\underline{1}^j}(\underline{n},P))\Big|_{\mathcal{C}(\underline{n})}$$

QED

It follows from Theorem 4.2 that each $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ may be computed from the set of polynomials $\{B^k_{\underline{1}^j}(\underline{n},\underset{\sim}{P})\}$; $k = 1,2,\ldots,i_j$, $j = 1,2,\ldots,\lambda$. These polynomials are exponentiated multilinear forms. Theorem 4.3 (Permutation theorem) will show that this set may be reduced to $\{B^k_{\underline{1}}(\underline{n},\underset{\sim}{P})\}$; $k = 1,2,\ldots,\max(i_1,i_2,\ldots,i_\lambda)$ by suffix permutation. An example of the use of Theorem 4.2 is given in Section 5.

Another class of recurrence relations between the $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$'s may be derived by observing that some of the $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$'s may be obtained from others by permutation of suffixes. The results are incorporated into the following theorem.

<u>Theorem 4.3</u> (Permutation theorem):

If π_σ , π_ρ are row and column suffix permutations, respectively, for integers $(1,2,\ldots,\lambda)$ then:

i) $B_{\pi_\sigma(\underline{i})}(\underline{n},\underset{\sim}{P}) = B_{\underline{i}}(\underline{n},\pi_\sigma^{-1}(\underset{\sim}{P}))$

ii) $B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = B_{\underline{i}}(\pi_\rho(\underline{n}),\pi_\rho(\underset{\sim}{P}))$

<u>Proof</u>: i) From Theorem 4.1(i):

$$B_{\pi_\sigma(\underline{i})}(\underline{n},\underset{\sim}{P}) = (\underset{\sim}{P}\cdot\underline{n})^{\pi_\sigma(\underline{i})}\Big|_{\mathcal{C}^{-1}(\underset{\sim}{P}\cdot\underline{n}),\ (\underline{n})} = (\pi_\sigma^{-1}(\underset{\sim}{P}\cdot\underline{n}))^{\underline{i}}\Big|_{\mathcal{C}^{-1}(\underset{\sim}{P}\cdot\underline{n}),}$$

$$= (\underset{\sim}{P}\cdot\pi_\sigma^{-1}(\underline{n}))^{\underline{i}} = B_{\underline{i}}(\underline{n},\pi_\sigma^{-1}(\underset{\sim}{P}))$$

ii) From Theorem 4.1(i):

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = (\underset{\sim}{P}\cdot\underline{n})^{\underline{i}}\Big|_{\mathcal{C}^{-1}(\underset{\sim}{P}\cdot\underline{n}),\mathcal{C}(\underline{n})} = (\pi_\rho(\underset{\sim}{P})\cdot\pi_\rho(\underline{n}))^{\underline{i}}\Big|_{\mathcal{C}^{-1}(\underset{\sim}{P}\cdot\underline{n}),\mathcal{C}(\underline{n})}$$

$$= B_{\underline{i}}(\pi_\rho(\underline{n}),\pi_\rho(\underset{\sim}{P})) .$$

QED

Theorem 4.3 is of importance in the computation of the $B_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P})$'s as a set of multivariate polynomials in $\underset{\sim}{n}$ and $\underset{\sim}{P}$. Part (i) of the theorem has two consequences. First, it gives a rule for deriving $B_{\underline{j}}(\underset{\sim}{n},\underset{\sim}{P})$ from $B_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P})$ if $\underline{j}$ is a permutation of $\underline{i}$. Second, it provides a means of generating some of the terms of $B_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P})$ from other terms of $B_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P})$. Part (ii) of the theorem also accomplishes the latter. The results are incorporated into Theorem 4.4 which may be regarded as an algorithm for determining $B_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P})$.

<u>Definition 4.2</u>: A <u>basis function</u> $\Psi_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P})$ is a multivariate polynomial in $\underset{\sim}{P}$ and $\underset{\sim}{n}$ defined by:

i) $$\Psi_{\underline{i}}(\underset{\sim}{n},\underset{\sim}{P}) = \sum_{\substack{\underset{\sim}{\gamma} \\ \left\{\begin{array}{l} \underline{\sigma}(\underset{\sim}{\gamma})=\underline{i} \\ \rho_1(\underset{\sim}{\gamma})\geq\rho_2(\underset{\sim}{\gamma})\geq\cdots\geq\rho_\lambda(\underset{\sim}{\gamma})\end{array}\right.}} \left[\frac{\underline{\sigma}(\underset{\sim}{\gamma})}{\underset{\sim}{\gamma}}\right]_\sigma (\underline{n})_{\underline{\rho}(\underset{\sim}{\gamma})} \cdot \underset{\sim}{P}^{\underset{\sim}{\gamma}}$$

ii) If $\underset{\sim}{\gamma}$ and $\underset{\sim}{\gamma}'$ satisfy the summation conditions and if $\underline{\sigma}(\underset{\sim}{\gamma}) = \underline{\sigma}(\underset{\sim}{\gamma}')$ where $\underset{\sim}{\gamma}' = \pi_\sigma(\underset{\sim}{\gamma})$ or if $\underline{\rho}(\underset{\sim}{\gamma}) = \underline{\rho}(\underset{\sim}{\gamma}')$ where $\underset{\sim}{\gamma}' = \pi_\rho(\underset{\sim}{\gamma})$ then one of the matrices $\underset{\sim}{\gamma}$, $\underset{\sim}{\gamma}'$ is eliminated from the summation.

<u>Theorem 4.4</u>: Let Q_ρ be the set of all column suffix permutations π_ρ, including the identity, for the integers $(1,2,\ldots,\lambda)$. Let $Q_\sigma^{(\underline{i})}$ be the set of all row suffix permutations π_σ, including the identity, for the integers $(1,2,\ldots,\lambda)$ with the

property that $\pi_\sigma(\underline{i}) = \underline{i}$. If $i_1 \geq i_2 \geq \ldots \geq i_\lambda$, then

$$B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum^*_{\begin{cases}\pi_\sigma,\pi_\rho \\ \pi_\rho \varepsilon Q_\rho \\ \pi_\sigma \varepsilon Q_\sigma^{(\underline{i})}\end{cases}} \mathcal{B}_{\underline{i}}(\pi_\rho(\underline{n}),\pi_\rho\pi_\sigma^{-1}(\underset{\sim}{P}))$$

where $$\mathcal{B}_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum_{\begin{cases}\underline{j} \\ \theta(\underline{i}) \leq \underline{j} \leq \underline{i} \\ j_1 \geq j_2 \geq \cdots \geq j_\lambda\end{cases}} S_{\underline{i}}^{(\underline{j})} \cdot \Psi_{\underline{j}}(\underline{n},\underset{\sim}{P})$$

where $\Psi_{\underline{j}}(\underline{n},\underset{\sim}{P})$ are basis functions, and * denotes summation over algebraically distinct forms.

<u>Proof</u>: The theorem follows from Corollary 3.1 and Theorem 4.3.

Theorems 4.3 and 4.4 enable all $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ with $\sigma(\underline{i}) \leq Q$ to be computed from the set of basis functions $\Psi_{\underline{j}}(\underline{n},\underset{\sim}{P})$ with $\underline{j} \leq \underline{i}$; $j_1 \geq j_2 \geq \ldots \geq j_\lambda$; $i_1 \geq i_2 \geq \ldots \geq i_\lambda$ and $\sigma(\underline{i}) \leq Q$.

The basis functions have the following property which is of importance in the extension to higher values of λ .

<u>Theorem 4.5</u> (Extension theorem):

If $\Psi_{\underline{j}}(\underline{n},\underset{\sim}{P})$ is a basis function, then $\Psi_{\underline{j}}(\underline{n},\underset{\sim}{P})$ is independent of λ if $\lambda \geq \sigma(\underline{j})$.

<u>Proof</u>: If $\lambda \geq \sigma(\underline{j})$, then a vector of length λ can accommodate any set of column sums which sum to a number

less than or equal to λ . Addition of a row and column of zeroes to the matrix $\underline{\gamma}$ (see Definition 4.2) does not affect the basis functions.

5. Results and Summary

The statements of Theorems 4.2 and 4.4 may be incorporated into two algorithms for computing the $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$'s as polynomials in $\underline{n}$ and $\underset{\sim}{P}$. The two methods are illustrated below for the 3-variable problem with $\underline{i} = (2,1,0)$.

Method 1 (Based on Theorem 4.2):

From Theorem 4.2: $B_{210}(\underline{n},\underset{\sim}{P}) = B_{200}(\underline{n},\underset{\sim}{P}) \times_{\mathcal{C}(\underline{n})} B_{010}(\underline{n},\underset{\sim}{P})$

Using the change of variable $x_i = \sum_{j=1}^{3} P_{ij}n_j$ we have:

$$B_{200}(\underline{n},P) = (x_1^2 + x_1)\Big|_{\mathcal{C}(\underline{n})} \quad \text{and}$$

$$B_{010}(\underline{n},P) = x_2\Big|_{\mathcal{C}(\underline{n})} \quad \text{whence}$$

$$B_{210} = (x_1^2+x_1)x_2\Big|_{\mathcal{C}(\underline{n})}$$

$$= ((P_{11}n_1+P_{12}n_2+P_{13}n_3)^2+(P_{11}n_1+P_{12}n_2+P_{13}n_3)) \cdot (P_{21}n_1+P_{22}n_2+P_{23}n_3)\Big|_{\mathcal{C}(n_1n_2n_3)} \tag{5.1}$$

which can be expanded straightforwardly. The expansion of the resulting forms may be carried out by a polynomial manipulation system (Horowitz [op.cit.]). It is important to note that the application of $\mathcal{C}(\underline{n})$ at the final stage implies that the variables $\{x_i\}$ may not be allowed to assume numerical values at the start of the computation, even if $\underset{\sim}{P}$ and $\underline{n}$ are given.

Method 2 (Based on Theorem 4.4):

From Theorem 4.4: $B_{210}(\underline{n},\underset{\sim}{P}) = \sum_{\begin{cases}\pi_\rho \\ \pi_\rho \in Q_\rho\end{cases}} (\Psi_{210}(\underline{n},\underset{\sim}{P}) + \Psi_{110}(\underline{n},\underset{\sim}{P}))$

The construction of $\Psi_{210}(\underline{n},\underset{\sim}{P})$ involves 4 matrices $\underset{\sim}{\gamma}$, (see definition 4.2):

$$\underset{\sim}{\gamma}: \text{i)} \begin{array}{cc} \left[\begin{matrix}2&\cdot&\cdot\\1&\cdot&\cdot\\\cdot&\cdot&\cdot\end{matrix}\right] & \begin{matrix}2\\1\\\cdot\end{matrix} \\ \begin{matrix}3&\cdot&\cdot\end{matrix} & \end{array};\quad \text{ii)} \begin{array}{cc} \left[\begin{matrix}2&\cdot&\cdot\\\cdot&1&\cdot\\\cdot&\cdot&\cdot\end{matrix}\right] & \begin{matrix}2\\1\\\cdot\end{matrix} \\ \begin{matrix}2&1&\cdot\end{matrix} & \end{array};\quad \text{iii)} \begin{array}{cc} \left[\begin{matrix}1&1&\cdot\\1&\cdot&\cdot\\\cdot&\cdot&\cdot\end{matrix}\right] & \begin{matrix}2\\1\\\cdot\end{matrix} \\ \begin{matrix}2&1&\cdot\end{matrix} & \end{array};$$

$$\left[\frac{\sigma(\underset{\sim}{\gamma})}{\underset{\sim}{\gamma}}\right]_\sigma: \quad (1) \qquad\qquad (1) \qquad\qquad (2)$$

$$\text{iv)} \begin{array}{cc} \left[\begin{matrix}\cdot&1&1\\1&\cdot&\cdot\\\cdot&\cdot&\cdot\end{matrix}\right] & \begin{matrix}2\\1\\\cdot\end{matrix} \\ \begin{matrix}1&1&1\end{matrix} & \end{array}$$

$$(2).$$

The construction of $\Psi_{110}(\underline{n},\underset{\sim}{P})$ involves 2 matrices $\underset{\sim}{\gamma}$ (see definition 4.2):

$$\underset{\sim}{\gamma}: \text{v)} \begin{array}{cc} \left[\begin{matrix}1&\cdot&\cdot\\1&\cdot&\cdot\\\cdot&\cdot&\cdot\end{matrix}\right] & \begin{matrix}1\\1\\\cdot\end{matrix} \\ \begin{matrix}2&\cdot&\cdot\end{matrix} & \end{array};\quad \text{vi)} \begin{array}{cc} \left[\begin{matrix}1&\cdot&\cdot\\\cdot&1&\cdot\\\cdot&\cdot&\cdot\end{matrix}\right] & \begin{matrix}1\\1\\\cdot\end{matrix} \\ \begin{matrix}1&1&\cdot\end{matrix} & \end{array}$$

$$\left[\frac{\sigma(\underset{\sim}{\gamma})}{\underset{\sim}{\gamma}}\right]: \quad (1) \qquad\qquad (1)$$

Thus $$\Psi_{210}(\underline{n},\underset{\sim}{P}) = (n_1)_3P_{11}^2P_{21}+(n_1)_2(n_2)_1(P_{11}^2P_{21}+2P_{11}P_{12}P_{21}) +2(n_1)_1(n_2)_1(n_3)_1P_{12}P_{13}P_{21} \tag{5.2}$$

$$\Psi_{110}(\underline{n},\underset{\sim}{P}) = (n_1)_2P_{11}P_{21} + (n_1)_1(n_2)_1P_{11}P_{22}$$

Each term of these polynomials is generated by one of the matrices. The factorization in (5.2) may be carried out by constructing all matrices with the same column sums, for given row sums.

The expansion of (5.2) to the full form of $B_{210}(\underline{n},\underset{\sim}{P})$ is simpler than the expansion of (5.1) to $B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ since the former involves only suffix permutation while the latter involves multivariate expansion. However, Method 2 relies on an efficient technique for constructing feasible matrices of the type given in Definition 4.2. The method is of particular value in deriving polynomials for the $\lambda + 1$ variable problem from those of the λ variable problem by means of Theorem 4.5 (Extension theorem).

The polynomials given in Appendix II may be used in an important special case of the (w,k,N)-problem, namely that of determining the expectations ϕ^* of statistical functions ϕ of the distance function f for pairs of binary sequences of fixed length when the digits are susceptible to error with given probability. It has been seen that this is a 3-variable problem. Appendix II contains the basis functions for computing the expectations for all $\underline{i}$ with $\sigma(\underline{i}) \leq 4$. Extension to higher $\sigma(\underline{i})$ has not been given for the sake of brevity. The probability matrix $\tilde{P}$ for the problem is determined from the given a priori state transition probability matrix by (2.2). For an arbitrary pair of sequences, it has been proved as a special case of Theorem 3.2, that the time taken to determine the expectations of $\phi(f)$ is independent of the length of the sequences. This has made the computation of these expectations an entirely feasible proposition.

The basis functions in Appendix II were generated by hand using Method 2 since only 39 matrices were involved. The main expression for $B_{\underline{i}}(\underline{n},\tilde{P})$, given in Corollary 3.1, was then implemented with the aid of the SAC-1 system to provide a check on these results. The time taken to compute and print $B_{\underline{i}}(\underline{n},\tilde{P})$ for the ten different vectors $\underline{i}$ given in Appendix II was 57 seconds. This gave the $B_{\underline{i}}(\underline{n},\tilde{P})$ as multivariate polynomials in $\underline{n}$ and $\underline{P}$, both for arbitrary $\tilde{P}$ and for $\tilde{P}$ determined under the assumption of equiprobability of error. The abbreviation obtained by presenting the $B_{\underline{i}}(\underline{n},\tilde{P})$ in terms of the basis functions $\Psi_{\underline{i}}(\underline{n},P)$ is particularly striking.

The technique used in generating the polynomials of Appendix II may be used for the general (w,k,N)-problem.

In the comparison of Method 1 and Method 2 the former is the more readily programmable since the method contains little more than polynomial exponentiation and since languages are available which permit polynomial manipulation. The SAC-1 system is one such system. However, the data structure representation of polynomials in this system, and others, is such that all variables of a product must appear in each term of that product. The data structure for the product of two sparse polynomials contains a high proportion of zero exponents and zero coefficients. Although the polynomials $B_{Q^j}(\underline{n},\underset{\sim}{P})$ (See Theorem 4.2(ii)), are not sparse their product (see Theorem 4.2(i)) is since these polynomials involve differing sets of variables. Accordingly, Method 1 is expensive on core storage requirements, particularly as the components of $\underline{i}$ increase in value or as λ increases.

Method 2 avoids the core store problems encountered in Method 1 by enabling the construction of $B_{\underline{i}}(n,\underset{\sim}{P})$ to be carried out term by term. This requires only a nominal amount of core storage. This method requires, however, that matrices satisfying Definition 4.2(ii) must be constructed by an efficient algorithm. Although the total time for the computation of the pol nomials of Appendix II was exceedingly small since the number of m rices was small, the computing time does become an important issue larger values of λ and for larger values of the components of $\underline{i}$. The efficient generation of matrices satisfying the conditions of Definition 4.2(i) and (ii) is being investigated.

Acknowledgements

The authors thank Mr. S. K. Sahni and Mr. J.K. Wong of the Department of Computer Science, Cornell University, for their participation in this work. The work was supported by National Science Foundation Grant GN-27224.

References

1. Good, I.J. The estimation of probabilities: An essay modern Bayesian methods, Research Monograph No. 30, MIT Press, 1965.

2. Nagy, G. State of the art in pattern recognition, Proceedings of IEEE, 56, 1968, pp. 836-862.

3. Friedman, H.P. and Rubin, J. On some invariant criteria for grouping data, Journal of the American Statistical Association, 62, 1967, pp. 1159-1178.

4. Zadeh, L.A. Fuzzy sets, Information and Control, 8, 1965, pp. 338-353.

5. Cover, T.M., and P.E. Hart. Nearest Neighbour Pattern Classification, IEEE Transactions on Information Theory, IT-13, 1967, pp. 21-27.

6. Cover, T.M. Estimation by nearest neighbour rule, IEEE Transactions on Information Theory, IT-14, 1968, pp. 50-55.

7. Sokal, R.R., and P.H.A. Sneath. Principles of Numerical Taxonomy, Freeman, San Francisco, 1963.

8. Horowitz, E. User notes for the SAC-1 system at Cornell University, Department of Computer Science, Cornell University, Ithaca, New York.

9. Barton, D.E. The matching distributions: Poisson limiting forms and derived methods of approximation, Journal of the Royal Statistical Society, B, 20, 1958, pp. 73-92.

10. Jackson, D.M. An error anlaysis for functions of qualitative attributes with application to information retrieval, In: Software Engineering, Vol. 2, Academic Press, New York, 1971.

11. Jackson, D.M., and L.J. White. The weakening of taxonomic inferences by homological errors, Mathematical Biosciences, 10, 1971, pp. 63-89.

12. Jackson, D.M., and L.J. White. The combinatorial decomposition of stability problems in nonstatistical classification theory, Technical Report 70-87, Department of Computer Science, Cornell University, Ithaca, New York.

13. Riordan, J. Combinatorial Identities, Wiley, New York, 1968.

14. Abramovitz, M., and J.A. Segun. Handbook of Mathematical Functions, Dover, 1965.

15. Riordan, J. Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions, Annals of Mathematical Statistics, 8, 1937, pp. 103-111.

APPENDICES

Appendix I

Notation and Conventions

1. Vectors and matrices are underscored by $\_$ and $\sim$, respectively.

2. The elements of all vectors and matrices, with the exception of those which are probabilistic, are non-negative integers.

3. It will be assumed throughout the text that the sizes of the vectors and matrices are such that the expressions in which they occur are well defined. Under this assumption, explicit discussion of the sizes of vectors and matrices is largely avoided.

4. Inequalities between vectors operate componentwise; e.g. $\underline{x} > \underline{y} \iff x_i > y_i$ for $1 \leq i \leq \lambda$, where $\underline{x}$, $\underline{y}$ have λ component

5. Rows and Columns of matrices:

   1) ${}_i\underline{M}$, $\underline{M}_j$ are the i-th row and j-th column, respectively, of the matrix $\underset{\sim}{M}$.

   2) $\underline{\sigma}(\underset{\sim}{M}) \equiv \{\sigma_i(\underset{\sim}{M})\} \equiv \{\sum_j M_{ij}\}$ are row sums of $\underset{\sim}{M}$:

   3) $\underline{\rho}(\underset{\sim}{M}) \equiv \{\rho_j(\underset{\sim}{M})\} \equiv \{\sum_i M_{ij}\}$ are column sums of $\underset{\sim}{M}$.

6. Implied Products:

   1) For vectors and matrices

      1) $\underline{m}^{\underline{i}} \equiv \prod_k m_k^{i_k}$

      2) $\underset{\sim}{M}^{\underset{\sim}{I}} \equiv \prod_{i,j} M_{ij}^{I_{ij}}$

2) For operators:

$$\sum_{\underset{\sim}{i}=\underset{\sim}{a}}^{\underset{\sim}{I}} \equiv \prod_{p,q} \sum_{i_{pq}=a_{pq}}^{I_{pq}}$$

3) For multinomials:

1) $$\left[\frac{\underline{i}}{\underset{\sim}{\underline{M}}}\right]_{\sigma} \equiv \prod_{j=1}^{k} \left[\begin{matrix} i_j \\ {}_j\underline{M} \end{matrix}\right] \quad \text{if} \quad \underline{i} = \underline{\sigma}(\underset{\sim}{M})$$

where $$\left[\begin{matrix} i_j \\ {}_j\underline{M} \end{matrix}\right] = \frac{i_j!}{M_{j1}!M_{j2}!\ldots M_{jk}!}$$

2) $$\left[\frac{\underline{i}}{\underset{\sim}{\underline{M}}}\right]_{\rho} \equiv \prod_{j=1}^{k} \left[\begin{matrix} i_j \\ \underline{M}_j \end{matrix}\right] \quad \text{if} \quad \underline{i} = \underline{\rho}(\underset{\sim}{M})$$

where $$\left[\begin{matrix} i_j \\ \underline{M}_j \end{matrix}\right] = \frac{i_j!}{M_{1j}!M_{2j}!\ldots M_{kj}!}$$

4) For Stirling numbers:

1) $$S_{\underline{i}}^{(\underline{j})} \equiv \prod_{k} S_{i_k}^{(j_k)}$$

2) $$S_{\underset{\sim}{I}}^{(\underset{\sim}{J})} \equiv \prod_{i,j} S_{I_{ij}}^{(J_{ij})}$$

5) For falling factorials:

$$(\underline{n})_{\underline{i}} \equiv \prod_{j} (n_j)_{i_j}$$

Appendix II

List of basis functions $\Psi_{\underline{i}}(\underline{n},\underset{\sim}{P})$ for the 3-variable problem with $\sigma(\underline{i}) \leq 4$

a) If $\phi(f(\underline{n}),f(\underline{m})) = \sum_{\underline{i}} a_{\underline{i}}(\underline{n}) \cdot \underline{m}^{\underline{i}}$ (see (3.1))

then $\mathcal{E}(\phi(f(n),f(\underline{m}))|\underline{n}) \equiv \phi^*(\underline{n}) = \sum_{\underline{i}} a_{\underline{i}}(\underline{n}) \cdot B_{\underline{i}}(\underline{n},\underset{\sim}{P})$ (see (3.2))

Case 1: $i_1 \geq i_2 \geq \cdots \geq i_\lambda$.

Then $B_{\underline{i}}(\underline{n},\underset{\sim}{P}) = \sum_{\begin{cases}\pi_\sigma,\pi_\rho \\ \pi_\sigma \in Q_\sigma^{(\underline{i})} \\ \pi_\rho \in Q_\rho\end{cases}} \mathcal{B}_{\underline{i}}(\pi_\rho(\underline{n}),\pi_\rho\pi_\sigma^{-1}(\underset{\sim}{P}))$ (see Theorem 4

Case 2: For the remaining $\underline{i}$'s , excluded from case 1, we may use:

$$B_{\pi_\sigma(\underline{i})}(\underline{n},\underset{\sim}{P}) = B_{\underline{i}}(\underline{n},\pi_\sigma^{-1}(\underset{\sim}{P}))$$ (see Theorem 4

b) The quantities $\mathcal{B}_{\underline{i}}(\underline{n},\underset{\sim}{P})$ are linear functions of the basis functions as follows: (The arguments of $\mathcal{B}_{\underline{i}}(\underline{n},\underset{\sim}{P})$ are suppressed for brevity)

$$\mathcal{B}_{100} = \Psi_{100}$$
$$\mathcal{B}_{200} = \Psi_{200} + \Psi_{100}$$
$$\mathcal{B}_{300} = \Psi_{300} + 3\Psi_{200} + \Psi_{100}$$
$$\mathcal{B}_{400} = \Psi_{400} + 6\Psi_{300} + 7\Psi_{200} + \Psi_{100}$$

$$\mathcal{B}_{110} = \Psi_{110}$$
$$\mathcal{B}_{210} = \Psi_{210} + \Psi_{110}$$
$$\mathcal{B}_{220} = \Psi_{220} + \Psi_{210} + \Psi_{110}$$

$$\mathcal{B}_{310} = \Psi_{310} + \Psi_{210} + \Psi_{110}$$

$$\mathcal{B}_{111} = \Psi_{111}$$
$$\mathcal{B}_{211} = \Psi_{211} + \Psi_{111}$$

c) The basis functions $\Psi_{\underline{i}}(\underline{n},\underset{\sim}{P})$ are as follows: (The arguments of $\Psi_{\underline{i}}(\underline{n},\underset{\sim}{P})$ are suppressed for brevity).

$$\Psi_{100} = (n_1)_1 P_{11}$$
$$\Psi_{200} = (n_1)_2 P_{11}^2 + 2(n_1)(n_2)_1 P_{11} P_{12}$$
$$\Psi_{300} = (n_1)_3 P_{11}^3 + 3(n_1)_2(n_2)_1 P_{11}^2 P_{12} + 6(n_1)_1(n_3)_1 P_{11} P_{12} P_{13}$$
$$\Psi_{400} = (n_1)_4 P_{11}^4 + 4(n_1)_3(n_2)_1 P_{11}^3 P_{12} + 6(n_1)_2(n_2)_2 P_{11}^2 P_{12}^2 + 12(n_1)_2(n_2)_1(n_3)_1 P_{11}^2 P_{12} P_{13}$$

$$\Psi_{110} = (n_1)_2 P_{11} P_{21} + (n_1)_1(n_2)_1 P_{11} P_{22}$$
$$\Psi_{210} = (n_1)_3 P_{11}^2 P_{21} + (n_1)_2(n_2)_1(P_{11}^2 P_{22} + 2P_{11} P_{12} P_{21}) + 2(n_1)_1(n_2)_1(n_3)_1 P_{12} P_{13} P_{21}$$
$$\Psi_{220} = (n_1)_4 P_{11}^2 P_{21}^2 + 2(n_1)_3(n_2)_1 P_{11} P_{21}^2 P_{12} + (n_1)_2(n_2)_2(4P_{11} P_{12} P_{21} P_{22} + P_{11}^2 P_{22}^2) + (n_1)_2(n_2)_1(n_3)_1(4P_{11} P_{12} P_{21} P_{23} + 2P_{11}^2 P_{22} P_{23})$$
$$\Psi_{310} = (n_1)_4 P_{11}^3 P_{21} + (n_1)_3(n_2)_1(3P_{11}^2 P_{12} P_{21} + P_{11}^3 P_{21}) + 3(n_1)_2(n_2)_2 P_{11}^2 P_{12} P_{22} + (n_1)_2(n_2)_1(n_3)_1(6P_{11} P_{12} P_{13} P_{21} + 3P_{11}^2 P_{12} P_{23})$$

$$\Psi_{111} = (n_1)_3 P_{11}P_{21}P_{31} + (n_1)_2(n_2)_1 P_{11}P_{21}P_{32} + (n_1)_1(n_2)_1(n_3)_1 P_{11}P_{22}P_{33}$$

$$\Psi_{211} = (n_1)_4 P_{11}^2 P_{21}P_{31} + (n_1)_3(n_2)_1(P_{11}^2 P_{21}P_{32} + 2P_{11}P_{12}P_{21}P_{31}) + (n_1)_2(n_2)_1(n_3)_1(P_{11}^2 P_{22}P_{33} + 2P_{11}P_{12}P_{21}P_{33} + 2P_{12}P_{13}P_{21}P_{31} + (n_1)_2(n_2)_2(P_{11}^2 P_{22}P_{32} + 2P_{11}P_{12}P_{21}P_{32})$$