

Spectral Analysis for Discrete Longitudinal Data

LAUREL BECKETT*

*Center for Research on Health and Aging; and Department of Medicine,
Rush–Presbyterian St. Luke's, Chicago, Illinois 60612*

AND

PERSI DIACONIS†

*Department of Mathematics, Harvard University,
Cambridge, Massachusetts 02138*

We develop a Fourier type analysis for data consisting of many short strings X_1, X_2, \dots, X_n , with $X_i = (X_{i1}, \dots, X_{ip})$. The paper offers an approach to testing and residual analysis based on a group theoretic decomposition of the sample space. This is illustrated on a physical chance type data set where it is natural to allow each string to have its own parameter (i.i.d. coin-tossing with parameter varying from string to string). The usual model is rejected. © 1994 Academic Press, Inc.

1. INTRODUCTION

It is a familiar scientific activity to expand a function in an orthogonal series and look at the relative size of the coefficients. The most common example is the spectral analysis of time series. James [19], Hannan [15], and Tukey [32] have pointed out that the usual analysis of variance of a designed experiment fits into this mold as well. This paper develops some non-standard examples of spectral analysis using the Fourier analysis of the permutation group.

Our methods are developed for longitudinal or panel study data. These consist of many short strings X_1, \dots, X_n with $X_i = (X_{i1}, \dots, X_{ip})$. For example, p might be 12 and $X_{ij} = 0$ or 1 as the i th person was employed in month j . Such data are often analyzed using an individual specific parametrization; each string might be modeled as a Markov chain with parameters that vary from string to string. We develop analysis of variance like decompositions for residuals from such models.

* Research supported in part by National Institute of Health Grants AG-006789 and AG-05362.

† Research supported in part by National Science Foundation Grant DMS89-05874.

Section 2 presents an analysis for a binomial example. It introduces a data set used throughout the rest of the paper. Section 3 develops the tools needed to analyze multinomial models. Sections 4 and 5 treat Markov models. These decompose into products of multinomial models but also give rise to data indexed by trees. The mathematical tools developed have application to formulas for the stationary distribution of Markov chains and to genetics algorithms.

2. AN EXAMPLE

A. Introduction. In this section we analyze strings from a typical "physical chance" setting that shows a breakdown in binomial variation.

The example involves repeated rolls of a common thumbtack. A one was recorded if the tack landed point up and a zero was recorded if the tack landed point down. All tacks started point down. Each tack was flicked or hit with the fingers from where it last rested. A fixed tack was flicked 9 times. The data are recorded in Table I. There are 320 9-tuples. These arose from 16 different tacks, 2 "flickers," and 10 surfaces. The tacks vary considerably in shape and in proportion of ones. The surfaces varied from rugs through tablecloths through bathroom floors.

These data are analyzed below and in Section 4F of this paper.

A binomial model for these data is

$$X_i = (X_{i1}, \dots, X_{ip}), \quad X_{ij} \sim \text{Bernoulli}(\theta^i), \quad 1 \leq i \leq n.$$

In the example, $p = 9$ and $n = 320$.

The first step of analysis uses sufficiency to eliminate the string to string variation as suggested by Frydman and Singer [12]. Let $X_i = X_{i1} + \dots + X_{ip}$. Under the binomial model, X_i is sufficient for θ^i . Given $X_i = j$, the law of (X_{i1}, \dots, X_{ip}) is uniform over all p -tuples with j ones. This uniformity will be used as a baseline for analyzing these data. A test based on this idea is carried out in Section 4. At the moment, we develop a spectral analysis which is the main focus of this paper.

B. Spectral Analysis. The approach suggested here uses the uniformity as a baseline for data. Let $X_{p-j,j}$ be the set of binary strings of length p having j ones. Then $|X_{p-j,j}| = \binom{p}{j}$. Let $M^{p-j,j}$ be the set of all functions from $X_{p-j,j}$ into the real numbers. The data set gives rise to functions $f_j \in M^{p-j,j}$, the number of observed sequences with a given pattern under the binomial model. These f_j should be a random function given n_j : the number of data strings with j ones. The space $M^{p-j,j}$ is a vector space and decomposes into a direct sum of subspaces:

$$M^{p-j,j} = S^p \oplus S^{p-1,1} \oplus \dots \oplus S^{p-j,j}. \quad (2.1)$$

The subspaces $S^{p-i,i}$ have a natural description in terms of invariance. This is discussed carefully in the next section. Hopefully the following intuitive description will suffice for now.

The space S^p consists of the constant functions. It is one-dimensional. The space $S^{p-1,1}$ is spanned by the functions

$$\delta_i(x) = \begin{cases} 1 & \text{if } x_i = 1 \\ -1 & \text{if } x_i = 0, 1 \leq i \leq p. \end{cases} \quad (2.2)$$

This has dimension $p-1$. The projection of the data function f_j onto $S^{p-1,1}$ amounts to looking at the number of strings having j ones with a 1 in position i , $1 \leq i \leq p$. Such projections are computed below.

The space $S^{p-2,2}$ can be described as being spanned by the functions

$$\delta_{kl}(x) = \begin{cases} 1 & \text{if } x_k = x_l = 1 \\ 0 & \text{otherwise,} \end{cases} \quad (2.3)$$

but with the first order statistics taken out, so that functions in $S^{p-2,2}$ are orthogonal to functions in $S^p \oplus S^{p-1,1}$. The spaces $S^{p-j,j}$ have a similar description in terms of j -tuples; because they will not be used further here, a careful definition is postponed until the next section.

Spectral analysis consists of the computation of the projection of f_j onto the various subspaces and the approximation of f_j by as many pieces as are required to give a satisfactory fit. Under the null hypothesis, f_j is uniformly distributed over all non-negative integer valued functions with total mass n_j . This gives a way to calibrate the analysis.

First Order Analysis. The projection of f_j onto $S^{p-1,1}$ can be found by computing the inner product of f_j with δ_i in (2.2). This has mean zero and

TABLE II
First-Order Statistics

Number of ones	Position									n_j
	1	2	3	4	5	6	7	8	9	
1	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	1.2	-0.6	3.1	3
2	-0.6	0.1	0.7	0.1	1.4	-0.6	-0.6	0.1	-0.6	13
3	-2.5	1.0	1.0	-1.0	1.0	1.5	0.5	-1.5	0	18
4	-2.5	-1.0	0.2	0.2	0.1	-0.4	1.1	2.0	0.5	48
5	-2.4	-1.6	0.8	-0.9	0.9	1.4	0.9	0	0.9	47
6	-3.5	-0.2	-0.2	0.1	0.6	1.1	1.4	0.9	-0.2	67
7	-2.6	0.3	0.3	0.6	0.6	1.3	0.6	-0.6	-0.6	54
8	-5.6	2.1	1.2	-0.2	-0.2	0.8	0.8	1.2	-0.2	51

Note. Results are standardized so each entry is approximately $N(0, 1)$ under H_0 .

variance $n_j(k/p)(1-k/p)$. These inner products, divided by their standard deviation, are reported in Table II for $j=1, 2, \dots, 8$ and $i=1, 2, \dots, 9$.

Remarks. (1) Observe that, roughly, the largest numbers in Table II are in the first column, and they all are negative. Recall that the tacks were all started out in zero position. The large negative entries show that the proportion of ones after the first flip is far smaller than it should be under the binomial model. The rest of the table appears unstructured.

(2) Except for $j=1$, entries in the tables are approximately standard normal variables under the null hypothesis. Different rows are independent, conditionally on the n_j , and entries within a row are independent save for the constraint that the rows sum to zero.

Second Order Analysis. The data vector f_j projected onto $S^{p-2,2}$ can be viewed by giving its inner product with the basis δ_{kl} of (2.3). This measures the number of strings with j ones, having ones in positions k and l , adjusted for the marginal frequency of ones in these positions. As shown in the Lemma of Section 3, the inner product of f with δ_{kl} , the projection of δ_{kl} onto $S^{p-2,2}$ is approximately normal with mean 0 and variance $n_j \|\delta_{kl}\|^2 / (p(p-3)/2)$. These variances were computed numerically (they are the same for each pair (k, l)).

Table III shows, for $j=4$ and all possible values of (k, l) , the inner product between f_j and the projection δ_{kl} standardized by the standard deviation.

Remarks. (1) The large numbers in the table tend to be positive and to occur at adjacent coordinates $(k, k+1)$. In these data we recognize a

TABLE III
Second-Order Statistics for Sequences with 4 Ones ($n_4 = 48$)

Positions	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9
Residuals	12.5	2.7	-1.5	-4.8	2.4	-4.2	-6.8	-0.3
Positions	2,3	2,4	2,5	2,6	2,7	2,8	2,9	
Residuals	8.6	-1.8	-5.1	2.1	-2.4	-5.1	-0.5	
Positions	3,4	3,5	3,6	3,7	3,8	3,9		
Residuals	9.2	-0.3	-3.6	-1.8	-10.7	-4.2		
Positions	4,5	4,6	4,7	4,8	4,9			
Residuals	10.1	-5.7	-8.0	-0.3	-2.1			
Positions	5,6	5,7	5,8	5,9		6,7	6,8	6,9
Residuals	-4.8	-0.9	0.6	5.1		8.3	3.6	-2.4
Positions	7,8	7,9		8,9				
Residuals	7.4	1.5		11.3				

tendency for pairwise adjacent coordinates to be the same. This has a simple physical interpretation. In the tack experiment, tacks were not reset between successive trials in a sequence. Often a tack would slide or "skittle" without turning over. This suggests that each string X_i may be distributed as a short Markov chain with its own parameters. This model is investigated in Section 4.

Similar patterns occur for sequences with 5 and 6 ones, respectively (not shown here). Residuals for sequences with other values of j appeared unstructured.

(2) There is a fair amount of dependence within a set of entries with n_j fixed; the 36 numbers shown in Table III have 27 degrees of freedom. The complete table of second-order statistics contains 252 entries based on 320 nine-tuples. Each entry is roughly standard normal, and entries from sequences with different numbers of ones are independent.

(3) The entries are based on the projection of f_4 into the space of second-order effects orthogonal to the constants and first-order effects. In other words, the second-order effects have been adjusted for the marginal (coordinate-wise) popularity. Then simple-to-interpret functions δ_{kl} which are 1 if $x_k = x_l = 1$ and zero elsewhere are projected into the space of second-order functions. If f_4 is particularly close to δ_{kl} , this is interpreted as a tendency for the kl coordinates of f_4 to be zero.

In the next section, we develop the spectral analysis for residuals from a multinomial mixture model more formally. Following that, we develop the spectral analysis for the Markov model. We will then return to the tack data and examine the adequacy of the Markov model for this example.

A summary of the analysis of this section is as follows. The binomial model seems inadequate. The position of the tack on the previous flip seems to really matter.

C. Related Literature. The literature on panel studies is vast. Hsiao [18] reviews the literature in economics, Kasprzyk *et al.* [21] review the sample survey literature. Ware, Lipsitz, and Speizer [35] survey problems with categorical data. Their article is in a special issue of *Statistics in Medicine* devoted to longitudinal methodology.

Mixtures of Markov chains are discussed by Markus [27], Korn and Whittemore [22], Muenz and Rubinstein [28] and by Frydman and Singer [12]. This last article introduced ideas which contributed crucially to the present paper. Langeheine and van der Pol [23] describe many further variants used in the social sciences.

A development of spectral analysis along present lines appears in Diaconis [6, 7]. This gives extensive references to work of James, Hannan, Tukey, and others.

The analysis above leaned heavily on the conditional distribution given a sufficient statistic. This is broadly applicable. Thus suppose that $P_{\theta_i}(X_{i1}, \dots, X_{ip})$ is a probability distribution for the i th string. Let $T_i(X_{i1}, \dots, X_{ip})$ be a sufficient statistic for θ_i so the law X_{i1}, \dots, X_{ip} given $T_i = t_i$ does not depend on θ_i . This can be used as the basis of a data analysis to investigate model adequacy.

For example, suppose the X_{ij} are real valued observations and the model specifies the i th string as independent and identically distributed normal variates with mean μ_i and variance σ_i^2 depending on i . A sufficient statistic is \bar{X}_i, S_i . The standardized residuals

$$(X_{i1} - \bar{X}_i, \dots, X_{ip} - \bar{X}_i)/S_i$$

should be uniformly distributed on the $p - 1$ -dimensional unit sphere. With many strings, a test or data analysis can be carried out. Spectral analysis would involve expanding in "spherical harmonics."

Fisher [11] often used such conditional distributions as tests of model adequacy. Cox [3] applied the idea for binary regression. Frydman and Singer [12] developed tests for a binomial mixture model using this idea.

Petrov [30] and Volondin [34] applied the idea in a continuous setting. One of their findings is discouraging. It can take a great deal of data to have reasonable power in pre-specified directions. For example, to test the location-scale family of uniform random variables against the location-scale family of normal random variables, with strings of length 5, several thousand strings were needed. The examples in Frydman and Singer [12] and those of the present paper suggest nonetheless that gross departures from models can be identified and understood.

The main new contribution of the present paper is the development of explicit tools for a data analysis. As will emerge, a fair amount of combinatorial analysis is required to extend the tools to multinomial or Markov models.

3. MULTINOMIAL DATA

A. Spectral Analysis. Consider data X_1, X_2, \dots, X_n , with $X_i = (X_{i1}, \dots, X_{ip})$ and each X_{ij} taking values in $\{1, 2, \dots, c\}$. We are interested in studying deviations from the multinomial model. Here, each X_i has a multinomial distribution— p observations in c categories with parameter $\theta^i = (\theta_1^i, \dots, \theta_c^i)$. Thus

$$P(X_{i1} = x_1, \dots, X_{ip} = x_p) = \binom{p}{\lambda_1 \dots \lambda_c} \prod_{j=1}^c (\theta_j^i)^{\lambda_j}$$

with $\lambda_j(x_1, \dots, x_p)$ the number of x_i that equal j . Different X_i are independent and may have different parameters.

The statistic $(\lambda_1, \dots, \lambda_c)$ is sufficient for θ^i and conditional on λ , (X_{i1}, \dots, X_{ip}) are uniform over all p -tuples having λ_j values of j , $1 \leq j \leq c$.

A single string X can be specified by saying where the λ_i values of i occurred. For example, $X = (1, 2, 1, 2, 2, 3)$ gives rise to

$$\{1, 3\}, \{2, 4, 5\}, \{6\} \quad (3.1)$$

because the ones are in position 1 and 3, the twos in position 2, 4, and 5, and the three is in position 6. This type of specification is called a *tabloid* of shape $\lambda_1 \cdots \lambda_c$. In (3.1) the shape is 2, 3, 1. Let $X(\lambda)$ denote the set of all tabloids of shape λ . Thus $|X(\lambda)| = p!/\lambda_1! \cdots \lambda_c!$. Individual tabloids are denoted $\{t\}$.

The symmetric group S_p operates transitively on $X(\lambda)$ as in the following example. If $\{t\}$ is given by (3.1),

$$\pi\{t\} = \{\pi(1), \pi(3)\}\{\pi(2), \pi(4), \pi(5)\}\{\pi(6)\}.$$

For a fixed shape $\lambda = (\lambda_1 \cdots \lambda_c)$, let M^λ be the set of all functions from $X(\lambda)$ into the real numbers R .

The space M^λ decomposes into an orthogonal direct sum of isotypic pieces

$$M^\lambda = \bigoplus_{\mu} V_{\mu}^{\lambda}, \quad (3.2)$$

where μ runs over partitions of p that are larger than λ in the order of majorization. Each V_{μ}^{λ} is a space of functions that is invariant under the permutation group: $f \in V_{\mu}^{\lambda}$ implies $\pi f \in V_{\mu}^{\lambda}$ with $\pi f\{t\} = f\pi\{t\}$. In particular, V_p^{λ} is the 1-dimensional space of constant functions.

The symmetric group S_n has an irreducible representation S^{μ} for each partition μ of n . The basic theorems of group representations say that any real vector space on which S_n acts decomposes into a direct sum of invariant, irreducible subspaces, each isomorphic to S^{μ} for some μ . Of course S^{μ} can occur several times in the decomposition. Grouping together the different copies of S^{μ} into one invariant subspace V_{μ} gives the isotypic decomposition. For the underlying space M^{λ} , the isotypic pieces are called V_{μ}^{λ} . Young's rule (James [20]) gives a simple description of which partition μ appears in M^{λ} and how many copies of S^{μ} are in a given V_{μ}^{λ} .

For binary data the partitions are all into two parts, $(p-j, j)$. The splitting turns out to be multiplicity free, so

$$M^{p-j, j} = S^p \oplus S^{p-1, 1} \oplus \dots \oplus S^{p-j, j}$$

each subspace $S^{p-i, i}$ occurs only once.

The subspaces V_μ^λ have reasonably simple combinatorial descriptions which enable the projections to be interpreted in a given subject matter context. Diaconis [6, 7] discusses this at far greater length.

Data X_1, \dots, X_n give rise to a collection of functions $f_\lambda(\cdot)$ for λ a partition of p into at most c parts. Here $f_\lambda\{t\}$ denotes the number of X_i having λ as sufficient statistic and $\{t\}$ as ancillary. Each f_λ can be decomposed into its projection into the isotypic subspace V_μ^λ . These projections can be further decomposed by taking their inner products with known simple functions. This is illustrated in Section 2, and further described and illustrated in Diaconis [7]. These projections and inner products constitute the promised spectral analysis.

B. Some Probability. Under the multinomial null hypothesis, $f_\lambda\{t\}$ should be a "random function." If there are N_λ strings with sufficient statistic λ , the values $f_\lambda\{t\}$ should be like N_λ balls dropped uniformly into $p!/\lambda_1! \dots \lambda_c!$ boxes. If N_λ is large, the projection of f onto the various V_μ^λ should result in new functions with lengths that are approximately chi-squared distributed. The inner products of a projection with an individual function should have an approximate normal distribution. The following lemma records some standard facts which allow assessment of variability under the null hypothesis. In the lemma, the set of all functions from $\{1, 2, \dots, n\}$ into the real numbers is regarded as an inner product space with $\langle f|g \rangle = \sum_{j=1}^n f(j)g(j)$.

LEMMA. Drop N balls into n boxes at random. Let $f\{j\}$ be the number of balls in box j .

(a) Let V be a subspace of the space of all real-valued functions orthogonal to the constants on $\{1, 2, \dots, n\}$. Let Π be orthogonal projection onto V . Then, as N gets large, $\|\Pi f\|^2 \sim C\chi_d^2$, with $d = \dim V$, $C = N/n$.

(b) Lengths of projections onto orthogonal subspaces are asymptotically independent.

(c) With V as in part (a), and $g \in V$ a fixed function; $\langle f|g \rangle \sim N(0, \|g^2\| N/n)$.

As an example, in Section 2, Table III presents the inner product of $f_4 \in M^{5,4}$ with the projection of δ_{kl} onto the subspace $S^{7,2}$ of $M^{5,4}$. These inner products are standardized by dividing by their standard deviations. From (c), the standard deviation of the projection of δ_{kl} onto the subspace $S^{7,2}$ of $M^{9-j,j}$ is

$$\sigma_j = \left\{ \frac{n_j}{\binom{p}{j} - \binom{p}{j-1}} \sum \hat{\delta}_{kl}(x)^2 \right\}^{1/2}$$

with n_j the number of strings having j ones and δ_{kl} the projection. The sum is over $X_{9-j,j}$. The projections were computed using character theory as explained in Diaconis [7]. The sums were done numerically. By symmetry $\sigma_j = \sigma_{9-j}$, $j = 2, 3, \dots, 7$. Numerically $\sigma_2 = 0.020833$, $\sigma_3 = 0.011116$, $\sigma_4 = 0.059524$. Thus σ_4 is small compared to the differences observed in Table III.

Efficient computation of such group-theoretic projections is closely related to recent work on the non-commutative Fast Fourier Transform. See Diaconis and Rockmore [9].

4. BINARY MARKOV CHAINS

A. *Introduction.* Consider data X_1, X_2, \dots, X_n with $X_i = (X_{i0}, \dots, X_{ip})$ and each X_{ij} taking values in $\{0, 1\}$. Assume that X_i are independent realizations from a Markov chain with transition matrix

$$\begin{pmatrix} \theta_{00}^i & \theta_{01}^i \\ \theta_{10}^i & \theta_{11}^i \end{pmatrix}$$

and arbitrary starting distribution $\theta^i = P(X_{i0} = 1)$.

A sufficient statistic for the i th string is the starting state X_{i0} , and the 2×2 transition count matrix T^i . This has j, k entry the number of j to k transitions in the i th string. For example, if

$$X_i = (0110111111), \quad T^i = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0 & 2 \\ 1 & 6 \end{pmatrix} \end{matrix}.$$

The conditional distribution of X_i given X_{i0} and T^i is uniform over all strings with these statistics. For example, there are 7 strings starting at 0 and having $\begin{pmatrix} 1 & 3 \\ 3 & 3 \end{pmatrix}$ as transition counts

$$\begin{array}{cccc} 0101111111 & 0110111111 & 0111011111 & 0111101111 \\ 0111110111 & 0111111011 & 0111111101 & \end{array}$$

The first use to be made of this sufficiency is a test of the binomial null hypothesis $\theta_{00}^i = \theta_{10}^i$. Consider all strings in the tack data set with j ones. Under the binomial model, each of the $\binom{9}{j}$ possible patterns has equal probability. As shown in Table IV, conditioning on j leaves only a moderate amount of data. It is natural to lump the j strings into equivalence classes of equal probability under the Markovian alternative.

TABLE IV
Number of Strings in Table I Having j Ones

j	0	1	2	3	4	5	6	7	8	9
n_j	0	3	13	18	48	47	67	54	51	19
$\binom{9}{j}$	1	9	36	84	126	126	84	36	9	1

Table V illustrates for strings with 8 ones in the tack example. There are three possible transition matrices. Under the binomial mixture model, each string is equally likely conditional on having 8 ones. Thus the possible transition matrices have probability $1/9$, $7/9$, $1/9$, respectively.

The standard chi-square statistic for this example gives $\chi_2^2 = 31.0$. This provides a striking rejection of the binomial null hypothesis. Table VI shows the corresponding test values and degrees of freedom for the other values of j .

Remarks. (1) All of the test statistics having moderate sample sizes are large. This allows a decisive rejection of the binomial null hypothesis.

(2) No test is possible for $j=0$ or 9 . For $j=1$, $n_1=3$ is too small to carry out a test. In Table VI, all categories having expected value < 5 have been pooled.

(3) The tests carried out here were derived before the data were collected and were not derived from the data analysis of Section 2.

B. A Parametrization of Markov Strings. Binary strings having j ones are in 1-1 correspondence with j sets of $\{1, 2, \dots, p\}$. This was useful in

TABLE V
Binary Strings of Length 10, Beginning with Zero and
Having 8 Ones, Grouped by Transition Counts

Transition matrix	Example of string	Number of strings	
		Possible	Observed
$\begin{pmatrix} 1 & 1 \\ 0 & 7 \end{pmatrix}$	0011111111	1	18
$\begin{pmatrix} 0 & 2 \\ 1 & 6 \end{pmatrix}$	0110111111	7	27
$\begin{pmatrix} 0 & 1 \\ 1 & 7 \end{pmatrix}$	0111111110	1	6

TABLE VI
Chi-Squared Tests for Binomial versus Markov Mixture Model
for Strings with j Ones

j	2	3	4	5	6	7	8
n_j	13	18	48	47	67	54	51
χ^2	0.11	0.08	31.3	17.8	77.5	18.7	31.0
df	1	1	4	5	5	3	2

setting up the spectral analysis for binomial residuals in Section 2. This section shows how the set of all binary strings with a given starting state and transition count matrix can be parametrized by a pair of subsets.

Fix a starting state $k \in \{0, 1\}$ and a transition count matrix T . These determine the final state l of all compatible strings: $l = k$ if and only if $T_{01} + T_{10}$ is even. The length of all compatible strings is one more than the sum of the entries of T .

The set of all strings compatible with k and T can be parametrized by a pair of subsets s, t with $|s| = T_{00}$, $|t| = T_{10} + l - 1$. Here $s \subset \{1, 2, \dots, T_{00} + T_{01} - l\}$ and $t \subset \{1, 2, \dots, T_{10} + T_{11} + l - 1\}$. The elements of s specify which transition from 0 go 0-0. The elements of t specify which transitions from 1 go 1-0. These uniquely determine the rest of the string since s^c and t^c specify where transitions from 0 to 1 and 1 to 1 go.

For example, consider $k = 0$, $T = \begin{pmatrix} 1 & 4 \\ 3 & 3 \end{pmatrix}$. From $T_{01} + T_{10} = 7$, all compatible strings end in $l = 1$ and all have length 12. Consider $s = \{3\}$, $t = \{1, 2, 5\}$, $s^c = \{1, 2, 4\}$, $t^c = \{3, 4, 6\}$. The string starts at 0. The first transition out of 0 is to 1, so the string starts 01. The first transition out of 1 is to 0, so the string starts 011. Continuing, gives

010100111011.

This parametrization will be justified in Section 5. It implies that the number of compatible strings is

$$\binom{T_{00} + T_{01} - l}{T_{00}} \binom{T_{10} + T_{11} + l - 1}{T_{11}}.$$

Under the Markov mixture model, each string, and so each pair of subsets, is equally likely given k and T . Given a set of strings, X_1, X_2, \dots, X_n , the subset of strings with a fixed value of k and T gives rise to a function $f_{k,T}(s, t)$ —the number of compatible strings with parameters s and t .

As an example, in the tack data set, all strings start at $k = 0$. Consider $T = \begin{pmatrix} 2 & 2 \\ 1 & 4 \end{pmatrix}$. The final state l must be 1. Now s runs over subsets of $\{1, 2, 3\}$

of size 2 and t runs over subset of $\{1, 2, 3, 4, 5\}$ of size 1. The data in Table I yield 24 compatible strings which can be displayed in a 3×5 table

$$\begin{matrix} & \{1\} & \{2\} & \{3\} & \{4\} & \{5\} \\ \begin{matrix} \{1, 2\} \\ \{1, 3\} \\ \{2, 3\} \end{matrix} & \begin{pmatrix} 2 & 3 & 3 & 2 & 2 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 5 & 1 & 0 & 3 \end{pmatrix} \end{matrix} \quad (4.1)$$

Under the Markov mixture hypothesis this table should appear like 24 balls dropped randomly into 15 boxes. Tests based on this property are carried out in subsection D.

C. *Spectral Analysis.* In general, using the notation of Sections 2 and 3, the function $f_{k,T}$ is an element of $M^\lambda \otimes M^\mu$ with

$$\begin{aligned} \lambda &= (\lambda_1, \lambda_2) & \text{and} & & \lambda_1 &= T_{01} - l, & \lambda_2 &= T_{00}, \\ \mu &= (\mu_1, \mu_2) & \text{and} & & \mu_1 &= T_{01} + l - 1, & \mu_2 &= T_{11}. \end{aligned}$$

The product of symmetric groups $S_{\lambda_1 + \lambda_2} \times S_{\mu_1 + \mu_2}$ operates on pairs of subsets and has a representation in $M^\lambda \otimes M^\mu$ which splits into irreducibles as

$$\sum_{i=0}^{\lambda_2} \sum_{j=0}^{\mu_2} S^{\lambda_1 + \lambda_2 - i, i} \otimes S^{\mu_1 + \mu_2 - j, j}.$$

TABLE VII

T	81	71	71	61	62	52	61	51	52	42	43	33
	00	10	01	11	10	20	02	12	11	21	20	30
Possible	1	8	1	7	7	21	1	6	12	30	15	20
Observed	2	1	0	4	2	7	0	5	5	7	1	0
T	51	41	42	32	23	23	24	14	41	31	32	22
	03	13	12	22	21	31	30	40	04	14	13	23
Possible	1	5	15	30	30	30	10	5	1	4	16	24
Observed	3	7	16	14	3	4	1	0	1	4	13	10
T	23	13	04	04	05	31	21	22	12	13	03	04
	22	32	31	41	40	05	15	14	24	23	33	32
Possible	36	24	16	4	1	1	3	15	15	30	10	10
Observed	14	3	1	1	0	6	10	24	10	13	3	1
T	21	11	12	02	03	11	11	02	01			
	06	16	15	25	24	07	17	16	08			
Possible	1	2	12	6	15	1	1	6	1			
Observed	4	8	23	6	13	18	6	27	19			

From this, a spectral analysis can be undertaken. Such an analysis would only be warranted for values of (k, T) having a substantial number of compatible strings.

For the tack data, Table VII shows, for all possible transition count matrices T , the number of cells in the corresponding table along with the number of compatible strings. Here, the data are sufficiently sparse that collapsing seems mandatory. This is carried out in the next section.

D. Testing the Markov Mixture Model. Under the Markov mixture model, strings with the same transition counts are conditionally uniform. An empty cells test was carried out on this basis. The 320 strings were sorted by transition count matrix. For each transition count matrix with 10 or more strings the expected number of empty cells is

$$E(X) = c \left(\frac{c-1}{c} \right)^n,$$

where c is the number of cells and n is the frequency of strings with the fixed transition count matrix. If $E(X) > 1$, the test statistic $(X - E(X))^2 / \text{Var}(X)$ was computed. This yielded 10 test statistics. A stem and leaf of the p -values is shown in Table VIII. The p -values appear like 10 uniform variates.

A variety of other tests were carried out along the following lines: consider (4, 1) in subsection B above. Projecting onto rows and columns gives

$$\begin{array}{cccccccc} 12 & 3 & 9 & 3 & 9 & 4 & 2 & 6 \\ \{1, 2\} & \{1, 3\} & \{2, 3\} & \{1\} & \{2\} & \{3\} & \{4\} & \{5\}. \end{array}$$

Under the Markov mixture model the row and column projections are conditionally independent. The row sums should appear as 24 balls

TABLE VIII
Stem and Leaf Plot of 10 p Values

0.0	9
0.1	0
0.2	
0.3	4
0.4	
0.5	
0.6	2
0.7	8490
0.8	2
0.9	1

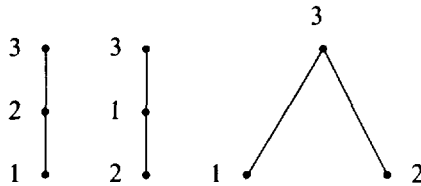
dropped at random into 3 boxes. The columns sums should appear as 24 balls dropped at random into 5 boxes. The row sums show the frequency of 0 to 0 transitions among all transitions out of 0. Thus the big count at $\{1, 2\}$ means that 0 to 0 transitions tend to occur at the earliest position, before 0 to 1 transitions.

None of the marginal multinomial distributions with sample size large enough to permit testing showed appreciable deviations from the uniform distribution. Combining the test statistics using Fisher's log p method also failed to produce evidence against the markov mixture model. Further discussion of the testing problems raised here is in D'Agostino and Stephens [4], Marden [26], and Rosenthal [31].

5. MARKOV CHAINS AND TREES

A. Introduction. This section can be read independently of what comes before. It offers a parametrization of the set of all strings with a given starting state and a given set of transitions, for strings taking values in a c state alphabet. Three uses are offered for this parametrization. The use of broadest appeal may be the following simple formula for the stationary distribution of a Markov chain.

All depends on the notion of a rooted labeled tree on c vertices. For example, there are three labeled trees on 3 vertices rooted at 3:



It is a familiar combinatorial fact that there are c^{c-2} labeled, rooted trees on c vertices. See, e.g., Lovasz [25, Chap. 4] or Moon [29].

Let $P(i, j)$ be a Markov chain on c states, assumed for simplicity to be irreducible and aperiodic. Let J be a labeled tree rooted at l . Define the tree product $\prod_J = \prod_{(i,j) \in J} P(i, j)$ where the product is over edges in J directed toward the root. Thus if $c = 3$ and

$$\begin{array}{c} 3 \\ | \\ J=2 \\ | \\ 1 \end{array} , \quad \prod_J = P(1, 2) P(2, 3)$$

THEOREM 5.1. *Let $P(i, j)$ be an ergodic Markov chain on c states with stationary distribution π . Then π can be given in terms of the entries of P as*

$$\pi(l) = \gamma \sum_J \prod_J,$$

where the sum is over all trees rooted at l and the normalizing constant γ is the l, k cofactor of the matrix $(I - P)$ (it is independent of k).

Remarks. (1) This theorem is essentially due to Kirkhoff. The result has been popularized by T. Hill [16, 17]. Peter Doyle [10] has given an elegant proof involving a random walk on trees. A new proof of the theorem is given in subsection C as an application of the description of strings consistent with a given transition count matrix. Leighton and Rivest [24] give a formula for non-ergodic chains and an interesting application to finite memory estimation. Related references are Dawson and Good [5], Goodman [13], and Haken [14].

(2) For example, in the 3×3 case, $\pi(3)$ is proportional to

$$P_{12}P_{23} + P_{21}P_{13} + P_{13}P_{23}.$$

The other applications offered are an efficient algorithm for generating a random string with a given start and transition count matrix (with applications to DNA string matching) and a description of such strings suitable for a spectral analysis along the lines of Sections 1-4.

B. A Parametrization of Markov Strings. Consider a finite string made up of letters from $\{1, 2, \dots, c\}$. Let T be a $c \times c$ transition count matrix with j, k entry the number of transitions from state j to state k . Thus 21211323 yields

$$T = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix}. \quad (5.1)$$

There are 8 strings starting with 2 and having this T as transition count matrix:

		21232113
21211323	21321123	21123213
21121323	21132123	23212113
		23211213.

The object of the present section is to give a parametrization for the set of all such strings for a general T .

Two classical results will be useful. Both can be found in Billingsley [2]. Let T be a $c \times c$ matrix of non-negative integers. The first result says that, apart from end effects, the number of transitions into i must equal the number of transitions out of i : A necessary and sufficient condition for T to be the transition count matrix for a string of length P is

$$\begin{aligned}
 \text{(a)} \quad & \sum_{ij} T_{ij} = p - 1 \\
 \text{(b)} \quad & T_{i.} - \delta_k(i) = T_{.i} - \delta_l(i) \text{ for all } i \text{ and some fixed } k, l.
 \end{aligned}
 \tag{5.2}$$

In (b), $\delta_a(b)$ is 1 or 0 as $a = b$ or not. If $k = l$ works in (b), then any k can be chosen as the starting state and the final state must also be this k . If $k \neq l$, then these are unique and all compatible strings start at k and end at l .

The second result gives a closed form expression for the number of compatible strings.

THEOREM (Whittle, 1958). *Let T satisfy conditions (a) and (b) in 5.2 above. The number of strings that start at k and end at l and have T as transition count matrix is the l, k cofactor of the $c \times c$ matrix with (i, j) entry $\delta_{ij} = T_{ij}/T_{i.}$*

Suppose now that T satisfies conditions (a) and (b) and that an admissible starting state k has been specified. This determines a unique final state l . Let J be any labeled tree rooted at l . Define a matrix T^J by

$$T^J_{ij} = T_{ij} - 1 \quad \text{if } (i, j) \text{ is an edge of } J \text{ directed toward } l. \tag{5.3}$$

For example, for the matrix T of (5.1) with starting state 2, the final state must be 3. If

$$\begin{array}{c}
 3 \bullet \\
 | \\
 J=2 \bullet \\
 | \\
 1 \bullet
 \end{array}
 \quad \text{then} \quad
 T^J = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5.4}$$

Define J to be *compatible* with T if T^J has non-negative entries.

The rows of T^J specify c partitions $(\lambda^1, \lambda^2, \dots, \lambda^c)$. For the example above $\lambda^1 = 1, 0, 1$. If λ is a partition of n , let X_λ be the set of all partitions of

$\{1, 2, \dots, n\}$ into disjoint subsets of size the parts of λ . Thus if $\lambda = (\lambda_1, \dots, \lambda_c)$ with $\lambda_i \geq 0$, $\sum_{i=1}^c \lambda_i = n$,

$$|X_\lambda| = \frac{n!}{\prod \lambda_i!}. \quad (5.5)$$

In the applications, λ^i represents the exits from i and $x \in X_{\lambda^i}$ codes where the exits go. If $x = \{1, 2, 5\}\{3, 6\}\{4\}$, the first and second exits are to symbol 1, the third exit is to symbol 2, the fourth exit is to symbol 3, and so on.

THEOREM. *Let T be a $c \times c$ array of non-negative integers satisfying (5.2) with k and l admissible starting and final states. The set of strings compatible with k , l , and T is the disjoint union $\cup A_J$ where J ranges over tabled trees on $\{1, \dots, c\}$ rooted at l and compatible with T , and*

$$A_J = \prod_{i=1}^c X_{\lambda^i}$$

with λ^i the partitions generated by the rows of T^J and X_{λ^i} defined at (5.5).

Proof. The proof proceeds by a constructive algorithm. Given k and T , the final state l is determined; let J be a labeled tree on c vertices rooted at l . A string will be constructed with last exit from i going to j if and only if (i, j) is an edge in J directed toward l . Form T^J as at (5.2). Let $\lambda^1, \lambda^2, \dots, \lambda^c$ be the partitions determined by the rows of T^J . For each i , $1 \leq i \leq c$, let $x_i \in X_{\lambda^i}$ be a partition of the numbers between 1 and $\sum_{j=1}^c \lambda_j^i$. These partitions determine a string by the following rule.

Begin by writing down the starting state k . Next look at x_k and write down the part of the partition containing the number 1. Say this part is i , so the string starts ki . Delete the 1 from x_k . Next look at x_i and write down the part of the partition containing the lowest number. Delete this and continue. If the algorithm gives directions to an x_i which has had all symbols deleted, write down the symbol above i in the tree J . Stop when all symbols in J have been used.

An example appears following the rest of the proof. To show that the algorithm works, that is, does not get stuck and produces each compatible string once and only once, observe that because of (5.2) the number of transitions remaining to go out of a state i is always at least as large as the number into i . These numbers become equal to 0 when (and only when) the last exit from i is determined from the tree. There can be no more calls to i , so that the algorithm cannot get stuck.

It is clear that strings produced by different last exit trees J or by $x \neq y$, in the same A_J are distinct. Finally, every compatible string is thus

encoded, for such a string has a last exit tree J and recording the successive exits from 1, 2, ..., c gives an element of A_J . ■

The proof just given is a slight rewording of de Bruijn–Ehrenfest–Smith–Tuttle theorem (BEST theorem). See, for example, Van Lint [33] or Zaman [36].

EXAMPLE. Consider the matrix (5.1) with J and T^J as in (5.3). The tree partitions are $\lambda^1 = (1, 0, 1)$, $\lambda^2 = (2, 0, 0)$, $\lambda^3 = (0, 1, 0)$. Take the corresponding x_i to be

$$x_1 = \{2\}, \{\phi\}, \{1\} \quad \{x_2\} = \{1, 2\}, \{\phi\}, \{\phi\} \quad x_3 = \{\phi\}, \{1\}, \{\phi\}.$$

The associated string begins with 2. In x_2 , the lowest symbol is in the first block, so the string starts 21 (and the lowest symbol is deleted from x_2). In x_1 , the lowest symbol is in the third block so the string starts 213. From x_3 , the string starts 2132. Going back to x_2 , the lowest symbol is in block 1, so the string starts 21321. From x_1 the string starts 213211. Now, all symbols in x_1 have been deleted. From the tree J , the next symbol is 2, so the string starts 2132112. Similarly, x_2 has all symbols deleted so the final string is 21321123.

COROLLARY. Let T be a $c \times c$ array of non-negative integers satisfying (5.2). Let k and l be admissible starting and ending states. If a compatible string is chosen uniformly, the chance that the last exit tree of this string is J equals

$$\theta \prod_{(i,j) \in J} T_{ij}.$$

The product is over edges in J directed toward the root. The normalizing constant θ is given by

$$\theta = \frac{1}{T_{lk}^*} \prod_{i \neq l} T_i.$$

where T_{lk}^* is the l, k co-factor of the $c \times c$ matrix with (i, j) entry $\delta_{i,j} - T_{ij}/T_i$. If any $T_i = 0$, set the (i, j) entry to be δ_{ik} and delete such i 's from the product defining θ .

Proof of the Corollary. The theorem shows that the number of strings with last exit tree J equals

$$T_l \cdot \frac{\prod_{i=1}^c (T_i - 1)!}{\prod_{i,j} T_{ij}!} \prod_{(i,j) \in J} T_{ij}.$$

Whittle's formula given at the start of their section gives the total number of compatible strings as

$$\frac{\prod_i T_i!}{\prod_{ij} T_{ij}!} T_{ik}^*$$

Dividing gives the result. If some row sums T_i are zero, no symbol i can appear and the problem is reduced to the case where $T_i > 0$.

Remarks. (1) The corollary gives an easy algorithm for choosing a random string with a fixed starting state and transition matrix T : determine the final state l , choose a tree rooted at l with probability proportional to the tree product as in the corollary. Form T^J , choose x_1, x_2, \dots, x_c uniformly at random in $\prod X_{i^j}$, and assemble the final string. This algorithm is a better version of earlier work of Diaconis and Freedman [8] and Zaman [36] who developed it as a means of proving finite versions of de Finetti's theorem for Markov chains. The older algorithms sampled repeatedly until a tree occurred. Biochemists Altschul and Erickson [1] used the older algorithms to generate random strings with fixed transitions to calibrate DNA string matching algorithms.

(2) The argument above proves Theorem (5.1): Let $X_0 = k, X_1, \dots, X_n$ be a path of the Markov chain. Let T_n be the transition count matrix of X_0, \dots, X_n . Corollary 1 showed that the last exit tree of X_0, \dots, X_n given X_0 and T_n has law proportional to the tree product. Further, \hat{P}_n , the matrix of observed transition proportions $T_{ij}/T_{i\cdot}$, converges almost surely to P as $n \rightarrow \infty$. It follows that the law of the last exit tree given $X_0 = k$ converges to a distribution proportional to the tree product based on P and that the normalizing constant converges to γ . Finally, for large n , the ergodic theorem yields

$$\pi(i) \sim P_k \{X_n = i\} = \sum_{\tau \text{ rooted at } i} P_k \{\text{last exit tree at time } n = \tau\}.$$

This proves Theorem 5.1.

REFERENCES

1. S. ALTSCHUL AND B. ERICKSON, Significance of nucleotide sequence alignments: A method for random sequence permutation that preserves dinucleotide and codon usage, *Molecular Biol. Evol.* **2** (1985), 526-538.
2. P. BILLINGSLEY, Statistical methods in Markov chains, *Ann. Statist.* **32** (1961), 12-40; correction, p. 1343.
3. D. COX, The regression analysis of binary sequences, *J. Roy. Statist. Soc. Ser. B* **20** (1958), 212-232.

4. R. D'AGOSTINO AND M. A. STEPHENS, "Goodness-of-Fit Techniques," Dekker, New York, 1986.
5. R. DAWSON AND I. J. GOOD, Exact Markov probabilities from oriented linear graphs, *Ann. Math. Statist.* **28** (1957), 946–956.
6. P. DIACONIS, "Group Representations in Probability and Statistics," Institute of Mathematical Statistics, Hayward, CA, 1988.
7. P. DIACONIS, Generalization of spectral analysis with application to ranked data, *Ann. Statist.*, in press.
8. P. DIACONIS AND D. FREEDMAN, Finite forms of de Finetti's theorem on exchangeability, *Synthese* **36** (1977), 271–281.
9. P. DIACONIS AND D. ROCKMORE, Efficient computation of the Fourier transform on finite groups, *J. Amer. Math. Soc.* (1990), 217–332.
10. P. DOYLE, Unpublished manuscript, 1981.
11. R. A. FISHER, "Statistical Methods for Research Workers," Hafner, New York, 1950.
12. H. FRYDMAN AND B. SINGER, Assessing qualitative features of longitudinal data, in "Longitudinal Analysis of Labor Market Data" (J. Heckman and B. Singer, Eds.), Cambridge Univ. Press, London/New York, 1985.
13. L. GOODMAN, Exact probabilities and asymptotic relationships for some statistics for m th order Markov chains, *Ann. Math. Statist.* **29** (1958), 476–490.
14. H. HAKEN, "Synergetics," 2nd ed., pp. 95–97, Springer-Verlag, New York, 1978.
15. E. J. HANNAN, Group representations and applied probability, *J. Appl. Probab.* **2** (1965), 1–68.
16. T. L. HILL, The stationary distribution of a Markov chain, *Theoret. Biol.* **10** (1966), 442–452.
17. T. L. HILL, Discrete time random walks on diagrams, *Proc. Natl. Acad. Sci.* **85** (1988), 5345–5349.
18. C. HSIAO, "Analysis of Panel Data," Cambridge Univ. Press, Cambridge, 1986.
19. A. T. JAMES, The relationship algebra of an experimental design, *Ann. Math. Statist.* **27** (1957), 993–1002.
20. G. D. JAMES, The representation theory of the symmetric groups, in "Lecture Notes in Math.," Vol. 682, Springer-Verlag, New York/Berlin, 1978.
21. D. KASPRZYK, G. DUNCAN, G. KULTON, AND M. P. SINGH, "Panel Surveys," Wiley, New York, 1989.
22. E. L. KORN AND A. WHITTEMORE, Methods for analyzing panel studies of acute health effects of air pollution, *Biometrics* **35** (1979), 795–802.
23. R. LANGEHEINE AND F. VAN DER POL, "Discrete Time Mixed Markov Latent Class Models," Technical Report, Netherlands Central Bureau of Statistics, Department of Statistical Methods, 1990.
24. T. LEIGHTON AND R. L. RIVEST, Estimating a probability using finite memory, *IEEE Trans. Inform. Theory* **32** (1986), 733–742.
25. L. LOVASZ, "Problems in Combinatorics," North-Holland, New York, 1983.
26. J. MARDEN, Combining independent noncentral chi-squared or F tests, *Ann. Statist.* **10** (1982), 266–277.
27. G. B. MARKUS, "Analyzing Panel Data," Sage Univ. Press, Beverly Hills, CA, 1979.
28. L. R. MUENZ AND L. V. RUBINSTEIN, Markov models for covariate dependence of binary sequences, *Biometrics* **41** (1985), 91–101.
29. J. W. MOON, Counting labeled trees, *Canad. Math. Comp.* (1970).
30. A. A. PETROV, Verification of statistical hypothesis on the type of distribution based on small samples, *Theory Probab. Appl.* **1** (1956), 223–245.
31. R. ROSENTHAL, Combining results of independent studies, *Psychol. Bull.* **85** (1978), 185–193.

32. J. TUKEY, Discussion, emphasizing the connection between analysis of variance and spectrum analysis, *Technometrics* **3** (1961), 191–219.
33. VAN LINT, Lectures in combinatorics, in “Lecture Notes in Math.,” Vol. 629, pp. 82–92, Springer-Verlag, New York/Berlin, 1968.
34. I. N. VOLONDIN, Testing of statistical hypothesis on the type of distribution by small samples, *Kazan Gos Univ. Ucen.* **125** (1965), 3–23, translation *Select. Transl. Math. Statist. Probab.* **11** (1973), 1–26.
35. J. H. WARE, S. LIPSITZ, AND F. E. SPEIZER, Issues in the analysis of repeated categorical outcomes, *Statist. Medicine* **7** (1988), 95–107.
36. A. ZAMAN, Urn models for Markov exchangeability, *Ann. Probab.* **12** (1984), 223–229.