STATISTICAL PROBLEMS INVOLVING PERMUTATIONS WITH RESTRICTED POSITIONS

PERSI DIACONIS, RONALD GRAHAM AND SUSAN P. HOLMES

Stanford University, University of California and ATT, Stanford University and INRA-Biométrie

The rich world of permutation tests can be supplemented by a variety of applications where only some permutations are permitted. We consider two examples: testing independence with truncated data and testing extra-sensory perception with feedback. We review relevant literature on permanents, rook polynomials and complexity. The statistical applications call for new limit theorems. We prove a few of these and offer an approach to the rest via Stein's method. Tools from the proof of van der Waerden's permanent conjecture are applied to prove a natural monotonicity conjecture.

AMS subject classifications: 62G09, 62G10.
Keywords and phrases: Permanents, rook polynomials, complexity, statistical test, Stein's method.

1 Introduction

Definitive work on permutation testing by Willem van Zwet, his students and collaborators, has given us a rich collection of tools for probability and statistics. We have come upon a series of variations where randomization naturally takes place over a subset of all permutations. The present paper gives two examples of sets of permutations defined by restricting positions.

Throughout, a permutation π is represented in two-line notation

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \pi(1) & \pi(2) & \pi(3) & \cdots & \pi(n) \end{pmatrix}$$

with $\pi(i)$ referred to as the label at position i. The restrictions are specified by a zero-one matrix A_{ij} of dimension n with A_{ij} equal to one if and only if label j is permitted in position i. Let S_A be the set of all permitted permutations. Succinctly put:

(1.1) $$S_A = \{\pi : \Pi_{i=1}^n A_{i\pi(i)} = 1\}$$

Thus if A is a matrix of all ones, S_A consists of all $n!$ permutations. Setting the diagonal of this A equal to zero results in derangement, permutations with no fixed points, i.e., no points i such that $\pi(i) = i$.

The literature on the enumerative aspects of such sets of permutations is reviewed in Section 2, which makes connections to permanents, rook polynomials and computational complexity.

Section 3 describes statistical problems where such restricted sets arise naturally. Consider a test of independence based on paired data $(X_1, Y_1), (X_2, Y_2) \ldots (X_n, Y_n)$. Suppose the data is truncated in the following way:

For each x there is a known set $I(x)$ such that the pair (X, Y) can be observed if and only if $Y \in I(X)$. For example, a motion detector might only be able to detect a velocity Y which is neither too slow nor too fast. Once movement is detected the object can be measured yielding X. Of course, such truncation usually induces dependence. Independence may be tested in the following form: Does there exist a probability measure μ on the space where Y is observed such that

$$(1.2) \quad P\{Y_i \in B_i, 1 \leq i \leq n | X_i, Y_i \in I(X_i), \ 1 \leq i \leq n\} = \Pi_{i=1}^{n} \frac{\mu(B_i)}{\mu(I(X_i))}$$

for all $B_i \subset I(X_i)$. Under assumption (1.2), given the unpaired data $\{X_i\}, \{Y_i\}$, any permutation π with $\{Y_{\pi(i)} \in I(X_i), 1 \leq i \leq n\}$ is equally likely for the paired data $(X_i, Y_{\pi(i)}), 1 \leq i \leq n$. This allows any standard test of independence to be quantified by its permutation distribution using S_A of (1.1) with A defined by

$$A_{ij} = \begin{cases} 1 & \text{if } Y_j \in I(X_i) \\ 0 & \text{else} \end{cases}$$

This example raises the problem of developing a theory of the distribution of rank statistics such as Kendall's tau or

$$\sum_{i=1}^{n} a_{i\pi(i)}$$

when π is chosen uniformly in S_A.

In Section 3 we discuss natural examples where the restriction matrix A has an interval structure (the ones in each row are contiguous). For one-sided intervals some of the standard limit theory can be pushed through, although much remains to be done. A classical contingency table with structural zeros of Karl Pearson is treated as an example.

For two-sided intervals, we offer an exchangeable pair so that Stein's method can be used. The exchangeable pair gives a Monte Carlo Markov chain for calibrating the permutation distribution. A red-shift data set of Efron and Petrosian (1998) is treated as an example.

Permutations with restricted position appear in a different guise in *skill scoring*, a technique for evaluation of taste testing and extra sensory perception experiments with feedback to subjects permitted. This application

is reviewed in Section 4. Some of the tools developed to prove the van der Waerden permanent conjecture are applied here to give a simple proof of a natural monotonicity conjecture.

We hope that these two examples of permutation testing with restricted positions will interest Bill. This part of the subject can certainly use his help.

2 Permanents

Let A be a zero-one square matrix of dimension n. The number of elements in S_A of (1.1) is determined by the permanent of A:

$$(2.3) \qquad |S_A| = Per(A) = \sum_\pi \Pi_{i=1}^n A_{i\pi(i)}$$

The sum in (2.1) is over all permutations in the symmetric group. Thus the permanent is like the determinant without signs. There is a large mathematical literature on permanents. We review some of this pertaining to matching theory (Section 2.1), rook theory (Section 2.2) and complexity theory(Section 2.3).

We believe that many of the nice developments in permutation enumeration and testing will work out nicely for the case of interval restricted permutations. An example, Fibonacci permutations, is developed in Section 2.4. It may be consulted now for motivation.

2.1 Bipartite Matchings

Let $[n] = \{1, 2, \ldots, n\}$ and $[n'] = \{1', 2', \ldots, n'\}$ be two disjoint n element sets($n = n'$). A bipartite graph G is specified by giving a set of undirected edges $\mathcal{E} = \{(i_1, i'_1), (i_2, i'_2) \ldots (i_e, i'_e)\}$. For example, when $n = 3$ the graph might appear as:
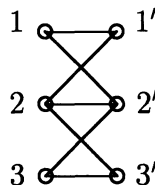
A matching in G is a set of vertex disjoint edges. Thus $(1, 1')(2, 3')$ is a matching in the figure above. A perfect matching is a matching containing n edges. There are three perfect matchings in the figure above.

There is a one-to-one correspondence between perfect matchings and the set S_A of (1.1) when A is taken as the adjacency matrix of the graph G.

Thus for the figure above:

$$A = \begin{array}{c c c c} & 1' & 2' & 3' \\ 1 & 1 & 1 & 0 \\ 2 & 1 & 1 & 1 \\ 3 & 0 & 1 & 1 \end{array}$$

and $Per(A) = 3$. This correspondence allows standard graph theory algorithms to be used for permutations with restricted positions. For example, given a restriction matrix A there is a polynomial time algorithm (order $n^{2.5}/\sqrt{\log n}$) for finding if there exists a perfect matchings using the widely available algorithm for solving the assignment problem of combinatorial optimization (see Cook (1998)).

Naively computing the permanent from its definition (2.3) takes $n.n!$ steps. Ryser's algorithm (see van Lint and Wilson (1992); theorem 11.2) improves this to order $n.2^n$. As discussed in Section 2.3 below, no substantial improvement can be expected for general restriction matrices (the problem is #-P complete). For some special cases (Sections 2.2,3.3,4.1) enumeration is feasible. For matrices A with row sums $r_1, r_2, \ldots, r_n$, van Lint and Wilson (1992 ; theorem 11.3) gives the bound

$$Per(A) \le \Pi_{i=1}^{n}(r_i)!^{\frac{1}{r_i}}.$$

Bipartite matching is the easiest case of matching theory in general graphs. Many further results applications and references are collected in the splendid book by Lovasz and Plummer (1986).

2.2   Rook Theory

This is an algebraic technique for enumerating S_A. The classical setting is a set of squares B of an $n \times n$ chess board. Let $r(B,k)$ be the number of ways of placing k non-attacking rooks on the squares of B (that is, choosing k squares in B no two in the same row or column). A board B is identified with a bipartite graph G with edge (i, i') if and only if (i, i') is in B. Thus $r(B,k)$ is the number of matchings with k edges.

The rook polynomial $r(G, x)$ is defined as

$$(2.4) \qquad r(G, x) = \sum_{k=0}^{n} (-1)^k r(B, k) x^{n-k}$$

Thus the number of perfect matchings is the value $r(G, 0)$. Rook polynomials have been extensively studied. Stanley (1986, Chapter 2) gives a useful treatment of the essentials including extensive references to the work of Goldman, Joichi, White.

Riordan (1958) reviews the extensive classical literature. Godsil (1981) and Lovasz/Plummer (1986) treat the generalization to matching polynomials of a general graph due to Heilmann-Lieb.

Rook and matching polynomials satisfy useful recurrences, have real zeroes and, for neat graphs, give rise to orthogonal polynomials.

2.3 Complexity Theory

Evaluation of the permanent of a square matrix is a celebrated problem in modern complexity theory. Indeed, Valiant (1979) used this as the first example of a #*P*-complete problem. Recall (Garey and Johnson, 1978) that NP-complete problems have "yes" or "no" answers, e.g.,"Is there some subset sum of this list of integers equal to 137?" The class of #*P*-complete problems are counting problems "How many subset sums are equal to 137?" For bipartite matching there is a fast way to check existence, but the counting problem is #*P*-complete; if a polynomial time algorithm exists then thousands of other intractable problems (e.g., computing the volume of a convex polyhedra) can be solved in polynomial time. A review of the work on the permanent from a complexity viewpoint can be found in Jerrum and Sinclair(1989) or Sinclair (1993).

Modern computer science has produced efficient randomized algorithms for approximating the permanent of a dense bipartite graph (every vertex having degree at least $\frac{n}{2}$). These algorithms perform a random walk on the set of perfect matchings and almost matchings (at most one edge missing). It has been proved that these walks converge rapidly and allow efficient selection of an essentially random perfect matching. This makes Monte Carlo quantification of the tests outlined in the introduction feasible for dense graphs.(Jerrum and Sinclair (1989), Sinclair (1993))

Unfortunately, arguments used by Jerrum and Sinclair and later workers really seem to depend on the denseness assumption. Despite extensive work over the past ten years the rate of convergence of natural random walks on the set of permutations consistent with a general restriction matrix remains open.

Among interesting recent developments we mention: Kendall, Randall and Sinclair (1996) show that the random walk algorithm works in polynomial time for the perfect matchings in bipartite graphs which are vertex transitive (e.g. d-dimensional rectangular grids with toroidal boundaries). Jerrum and Vazirani (1992) give an algorithm that approximates the permanent of any $n \times n$ zero-one matrix in time $e^{cn^{\frac{1}{2}}(logn)^2}$. This is superpolynomial but better than the $n2^n$ of Ryser's algorithm. Rasmussen (1994) gives a simple greedy approximation algorithm for the permanent which is shown to run in polynomial time for almost every graph for the usual $G(n,p)$ model of random graphs. Finally, Karmarkar et al.(1993) followed by Barvinok

(1998) and Rasmussen (1998) show how to approximate the permanent by a stochastic algorithm which replaces the ones by cube roots of unity and takes the squared modulus of the determinant. This algorithm has good average case behavior but exponential worst case behavior.

In Section 3 below we describe walks with interval restrictions where some things can be proved.

2.4   The Example of Fibonacci Permutations

This Section treats a simple example which shows that elegant theory can be developed for enumeration, random generation and the study of cycle structure. Let A_n be the $n \times n$ matrix with ones on, just above and just below the diagonal. Thus when $n = 4$

$$
A_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}
$$

From the definition (2.1), Per(A) is multilinear. Expanding by the first row, we derive a result first noted by Lehmer (1970):

$$
Per A_n = Per A_{n-1} + Per A_{n-2}.
$$

From direct computation, $Per(A_1) = 1$, $Per(A_2) = 2$, so $Per(A_n) = F_{n+1}$ is the $(n+1)st$ Fibonacci number. The Fibonacci numbers are defined as

$$
F_0 = 0, F_1 = 1, F_2 = 1, F_3 = 2, F_4 = 3, F_5 = 5, \ldots, F_{n+1} = F_n + F_{n-1}, \ldots
$$

We will call the elements of S_{A_n} Fibonacci permutations.

We first develop several bijections with combinatorial objects well known to be counted by Fibonacci numbers. Permutations counted by $Per(A_n)$ can be described in their cycle form. They are all permutations consisting of fixed points and pairwise adjacent transpositions. To see this, observe that permutations can be constructed as follows.

Place symbols $1, 2, 3, \ldots, n$ in a line, put a left parenthesis at the start and a right parenthesis at the end. Proceeding sequentially, decide to pass on or to place parentheses)(between i and $i+1$. This results in the following five permutations consistent with A_4:

$$
(1)(2)(3)(4), (1)(2)(34), (1)(23)(4), (12)(3)(4), (12)(34).
$$

From this description it is easy to see that the permutations enumerated by A_n are in one to one correspondence with the following well-known Fibonacci equivalents.

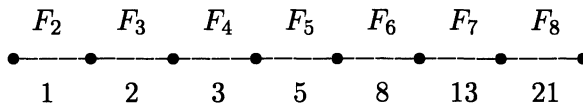**Proposition 2.1** *The set of Fibonacci permutations on n letters is in one-to-one correspondence with:*

- *Subsets of $[n-1]$ with no consecutive elements:* $\phi, \{1\}, \{2\}, \{3\}, \{1,3\}$

- *binary (n-1)-tuples without two consecutive ones:* $000; 100; 010; 001; 101$

- *Matchings in an n-path:*

  o o o o ; o—o o o ; o o—o o ; o o o—o ; o—o o—o

- *Compositions of n with all parts equal to one or two* $1111, 211, 121, 112, 22$

The next proposition gives an easy, direct method for uniformly choosing a random Fibonacci permutation. It is based on a theorem of Zeckendorf (1972) and the Fibonacci *numbering system*.

Proposition 2.2 *Any positive integer n can be uniquely expressed as $n = F_{k_1} + F_{k_2} + \cdots + F_{k_\ell}$ with F_{k_i} distinct Fibonacci numbers starting with F_2, no two adjacent.*

Thus $1 = F_2, 2 = F_3, 3 = F_4, 4 = F_2 + F_4, 5 = F_5, 6 = F_5 + F_2, 7 = F_5 + F_3, 8 = F_6, 9 = F_6 + F_2, 10^6 = F_{30} + F_{26} + F_{24} + F_{12} + F_{10}$, these topics are covered in Graham, Knuth, Patashnik(1989, pp.281-283). They show that the representation is easy to find, each time substracting off the largest possible Fibonacci number.

For our purposes, consider a path with n vertices and edges labeled with consecutive Fibonacci numbers starting with F_2;eg for n=8:

$$
\begin{array}{ccccccc}
F_2 & F_3 & F_4 & F_5 & F_6 & F_7 & F_8 \\
\bullet\!-\!-\!-\!\bullet\!-\!-\!-\!\bullet\!-\!-\!-\!\bullet\!-\!-\!-\!\bullet\!-\!-\!-\!\bullet\!-\!-\!-\!\bullet \\
1 & 2 & 3 & 5 & 8 & 13 & 21
\end{array}
$$

The weight of a matching in this graph is the sum of the weights of the edges. The largest possible sum comes from choosing the unique maximal matching. It is easy to prove that this has weight $F_{n+1} - 1$. Thus in the example above, $21 + 8 + 3 + 1 = 33 = F_9 - 1$. With this preparation, the algorithm is simple.

Proposition 2.3 *The following algorithm produces a randomly chosen Fibonacci permutation on n letters.*

Choose an integer U, uniformly between zero and $F_{n+1} - 1$. Express U in the Fibonacci numbering system and use the edges as a matching. Using the correspondence between matchings and Fibonacci permutations completes the construction.

For a randomly chosen permutation $\pi \in S_n$ the number of fixed points $F(\pi)$ has an approximate Poisson(1) distribution, the number of transpositions $T(\pi)$ has an approximate Poisson($\frac{1}{2}$) distribution, the number of cycles $C(\pi)$ has an approximate Normal distribution with mean $\log n$ and variance $\log n$, and finally the number of inversions has an approximate Normal distribution with mean $\frac{n^2}{4}$ and variance $\frac{n^3}{36}$.

The following proposition shows how these results carry over to Fibonacci permutations. In this case, the four results coalesce since $T(\pi) = I(\pi)$, $C(\pi) = T(\pi) + I(\pi)$ and $n = F(\pi) + 2T(\pi)$.

Proposition 2.4 *Let π be a randomly chosen Fibonacci permutation on n letters. Let $T(\pi)$ be the number of transpositions in π. Then*

$$(2.5) \qquad P\{T = k\} = \frac{\binom{n-k}{k}}{F_{n+1}}, \qquad 0 \leq k \leq \lfloor \frac{n}{2} \rfloor.$$

$$(2.6) \qquad E_n(T) = \frac{n(\sqrt{5}-1)}{2\sqrt{5}} + \frac{\hat{\varphi}}{5} + O(n\hat{\varphi}^{2n}),$$

$$with \ \hat{\varphi} = \frac{1-\sqrt{5}}{2},$$

$$(2.7) \qquad var_n(T) = \frac{n}{5\sqrt{5}} + O(1).$$

$$(2.8) \qquad P\{\frac{T - E_n}{\sqrt{var_n}} \leq x\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, dt.$$

Proof   A bijection between Fibonacci permutations with $T(\pi) = k$ and k sets of an $n - k$ set is easy to see; Put $n - k$ dots down in a row. Circle each element in a subset of size k. Now working from left to right, all encircled points are expanded to two points and correspond to transpositions. Thus

$$\circledcirc \quad \circledcirc \quad \bullet \longleftrightarrow (12)(34)5$$

The generating function for $\binom{n-k}{k}$ is a classical result:

$$f(z) \ = \ \sum_k \binom{n-k}{k} z^k$$

$$(2.9) \qquad = \ \frac{1}{\sqrt{1+4z}} \left\{ (\frac{1+\sqrt{1+4z}}{2})^{n+1} - (\frac{1-\sqrt{1+4z}}{2})^{n+1} \right\}$$

Setting $z = 1$ gives the classical expression

$$(2.10) \qquad F_{n+1} = \sum_k \binom{n-k}{k} = \frac{1}{\sqrt{5}} \{\varphi^{n+1} - \hat{\varphi}^{n+1}\},$$

$$\text{where } \varphi = \frac{1+\sqrt{5}}{2}, \hat{\varphi} = \frac{1-\sqrt{5}}{2}.$$

Differentiating (2.9) at $z = 1$ gives

$$\sum_k k\binom{n-k}{k} = \frac{(n+1)}{5}(\varphi^n + \hat{\varphi}^n) - \frac{2}{5\sqrt{5}}(\varphi^{n+1} - \hat{\varphi}^{n+1})$$

$$\sum_k k(k-1)\binom{n-k}{k} = \frac{\sqrt{5}}{25}n(n+1)(\varphi^{n-1} - \hat{\varphi}^{n-1})$$

$$-\frac{6(n+1)}{25}(\varphi^n + \hat{\varphi}^n) + \frac{12\sqrt{5}}{125}(\varphi^{n+1} - \hat{\varphi}^{n+1})$$

Dividing by F_{n+1} and using (2.10), routine simplifications give (2.6,2.7) .

An interesting proof of the liming normality uses the identification of Fibonacci permutations with the set of all matchings in an n-path. In this identification, the number of transpositions corresponds to the number of edges in the associated matching. Godsil(1981) has shown that the number of edges in a random matching of a general graph tends to a Normal distribution provided only that the variance tends to infinity. ∎

Remark 2.1 The limiting normality can be proved directly from (2.5) using Stirling's formula to give a local central limit theorem. Alternatively, the generating function (2.9) can be used. Godsil's proof used above itself uses Harper's method and depends on the fact that the zeros of the matching polynomial are all real. Pitman(1997) develops Harper's method and gives easily computed error bounds to the central limit theorem.

Remark 2.2 Similar results can be developed for the case where the restriction matrix has ones on the diagonal and k to the left and right of the diagonal in each row. The Fibonacci example has $k = 1$.

2.5   Other applications of permanents

The above surveys only mention features of the permanent literature directly related to the present project. There are many further applications. Mallows (1957) shows how permanents appear in computing normalizing constants in non-null ranking models. Bapat (1990) surveys a variety of appearances of permanents in statistics and further applications appear in Sections 3 and 4 below. Daley and Vere-Jones (1988) show how permanents occur for point-process theory and computing moments of complex normal variables.

One of the most active recent developments is the immanants. These are expansions of the form:

(2.11)
$$\sum_\pi \Xi(\pi)\Pi_{i=1}^n A_{i\pi(i)}$$

with Ξ a character of the symmetric group. Taking $\Xi = 1$ gives the permanent, taking $\Xi(\pi) = sgn(\pi)$ gives the determinant. There is active work giving inequalities for other immanants (see Lieb (1966) and Stembridge (1991, 1992) for surveys).

3 Testing for Independence

3.1 Introduction

Consider the classical problem of testing for independence without truncation. One observes pairs $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ drawn independently from a joint distribution $\mathcal{P}$ with $X_i \in \mathcal{X}, Y_i \in \mathcal{Y}$, suppose that $\mathcal{P}$ has margins $\mathcal{P}^1$ and $\mathcal{P}^2$. A test of the null hypothesis of independence: $\mathcal{P} = \mathcal{P}^1 \times \mathcal{P}^2$ may be based on the empirical measure $\hat{\mathcal{P}}_n$. Let δ be a metric for probabilities on $\mathcal{X} \times \mathcal{Y}$. One class of test statistics is

$$(3.12) \qquad\qquad T_n = \delta(\hat{\mathcal{P}}_n, \hat{\mathcal{P}}_n^1 \times \hat{\mathcal{P}}_n^2)$$

Extending classical work of Hoeffding (1948), Blum, Kiefer and Rosenblatt (1961), and Bickel (1969), Romano (1989) show that under very mild regularity assumptions, the permutation distribution of the test statistic T_n gives an asymptotically consistent locally most powerful test of independence.

Consider next the truncated case explained in Section 1. The hypothesis (1.2) may be called quasi-independence in direct analogy with the similar problem of testing independence in a contingency table with structural zeros. Clogg (1986) and Stigler (1992) review the literature and history of tests for quasi-independence with references to the work of Caussinus and Goodman.

While optimality results are not presently available in the truncated case, it is natural to consider the permutation distribution of statistics such as (3.12). This leads to a host of open problems in the world of permutations with restricted position.

We were led to present considerations by a series of papers in astrophysics literature dealing with the expanding universe. The red shift data that is collected for these problems suffers from heavy truncation problems.

For example, Figure 1 from Efron, Petrosian (1998) shows a scatterplot of 210 $x - y$ pairs subject to interval truncation, the x coordinate corresponds to red-shift, the y coordinate corresponds to log-luminosity. A suggested theory of 'luminosity evolution' says that early quasars were brighter. This suggests that points on the right side of the picture are higher because the high redshift corresponds to high age.

Astronomers beginning with Lynden-Bell (1971, 1992) have developed permutation type tests based on Kendall's tau for dealing with these problems. There is a growing statistical literature on regression in the presence of truncation; see Tsui et al.(1988) for a survey.
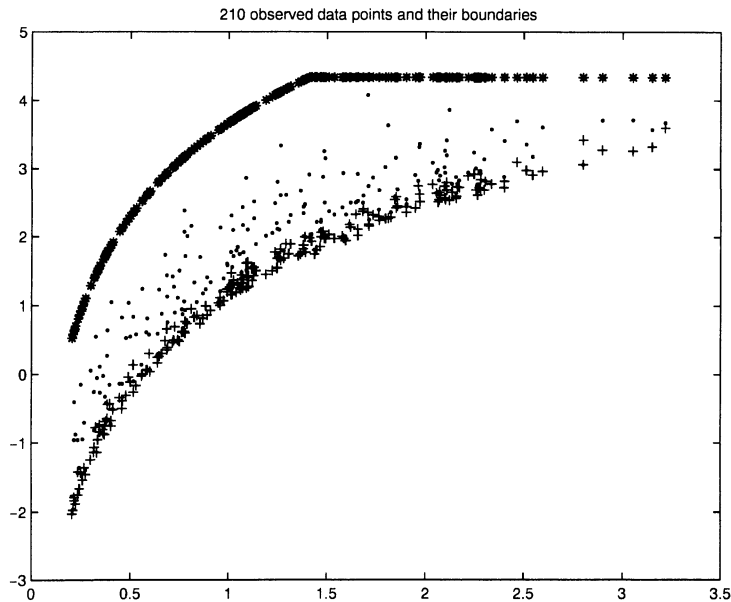
210 observed data points and their boundaries

Figure 1. Quasar Data, 210 points, upper and lower limits.

Most previous work deals with one-sided truncation of real-valued observations. The theory and practice is easier here as explained in Section 3.2 . Efron and Petrosian (1998) have recently developed tests and estimates for the case of two-sided truncation. We develop some theory for their setup in Section 3.3. The following preliminary lemma shows that interval truncation of real valued observations leads to restriction matrices with intervals of ones in each row.

Lemma 3.1 *Let $x_1, \ldots, x_n$ take values in an arbitrary set. Let $I(x_i)$ be a real interval. Let $y_1, y_2, \ldots, y_n$ be real numbers with $y_i \in I(x_i)$. Suppose the ordering is chosen so that $y_1 \leq y_2 \leq y_3 \leq \cdots \leq y_n$. Finally, define a zero-one matrix A of dimension n by*

$$A_{ij} = \begin{cases} 1 & \text{if } y_j \in I(x_i) \\ 0 & \text{else} \end{cases}$$

Then each row in A has its ones in consecutive positions with a one in position $(i, i), 1 \leq i \leq n$.

Proof The intervals may be pictured as in Figure 2, which is translated
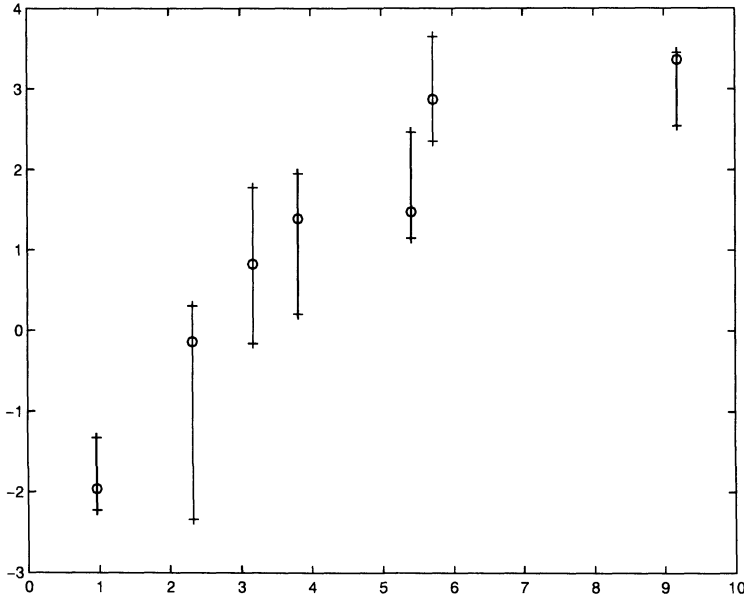
Figure 2. Seven permitted intervals and interior points

to the matrix whose transpose is

$$(3.13) \qquad \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Consider row i of A. By definition, $A_{ii} = 1$. Suppose that $A_{ij} = 1$, for some $i < j$. Thus $y_j \in I(x_i)$ and of course $y_i \in I(x_i)$. By monotonicity, $y_\ell \in I(x_i)$ for $i \le \ell \le j$. A similar argument for $j < i$ completes the proof. ∎

3.2   One-sided restriction

Paired data with (y_i, x_i) observable, if and only if, $y_i \in (a_i, \infty)$ give rise to one sided restriction matrices ($y_i \in (-\infty, a_i)$ is similarly treated). This Section shows that a neat theory emerges for one-sided truncation. Sets of permutations consistent with one-sided truncation can be described as follows. Let $b = (b_1, b_2, \ldots, b_n)$ with $1 \le b_i \le n$ be positive integers. Let

$$(3.14) \qquad S_b = \{\pi : \pi(i) \ge b_i, 1 \le i \le n\}$$

Equivalently, the matrix A has $A_{ij} = 1$ iff $j \geq b_i$. Without loss of generality, we take $b_1 \leq b_2 \leq \cdots b_i \leq \cdots \leq b_n$ in the sequel.

Some examples:

- If all $b_i = 1$, S_b contains all permutations.

- If $b_i = 1, 1 \leq i \leq a$, $b_i = 2, a + 1 \leq i \leq n$, S_b contains all permutations with 1 in the first a places.

- If $b_i = i$, S_b contains only the identity permutation.

- If $b_1 = b_2 = 1, b_i = i - 1, 1 \leq i \leq n$ S_b contains 2^{n-1} permutations.

The restriction matrices arising from sets S_b give rise to what are called Ferrers boards in the rook polynomial literature. The following lemma, due essentially to Karl Pearson in 1913 shows it is easy to choose from the uniform distribution on S_b.

Lemma 3.2 *Let* $b_1 \leq b_2 \leq \cdots \leq b_i \leq \cdots \leq b_n$ *be positive integers with* $b_i \leq i, 1 \leq i \leq n$. *The following algorithm results in a uniform choice from* S_b. *Begin with a list containing* $1, 2, \ldots, n$.

- *Choose* $\pi(n)$ *uniformly from* $J = \{j : j \geq b_n\}$. *Delete* $\pi(n)$ *from the list of choices.*

- *Choose* $\pi(n-1)$ *from the elements* j *in the current list* J *with* $j \geq b_{n-1}$. *Delete* $\pi(n-1)$ *from the list.*

- $\vdots$ *and so on...*

Proof The algorithm produces an element of S_b without getting stuck because of the restriction $b_i \leq i$. Conversely, any π in S_b is possible through a unique set of choices. ∎

The following is a classical corollary from the combinatorics literature. It shows there is a clean formula for $|S_b|$ and gives a representation of the permutation distribution of Kendall's tau as a sum of independent uniform variates.

Corollary 3.1 *Let* $I(\pi)$ *be the number of inversions in* π *and* $C(\pi)$ *be the number of cycles in* π. *Then*

(1) $|S_b| = \Pi_{i=1}^n (1 + (i - b_i))$

(2) $\sum_{x \in S_b} x^{I(\pi)} = \Pi_{i=1}^n (1 + x + x^2 + \cdots x^{i - b_i})$

(3) $\sum_{x \in S_b} x^{C(\pi)} = \Pi_{i=1}^n (x + (i - b_i))$

Corollary 3.2 *Suppose that X_i is uniform on $\{0, 1, \ldots, i - b_i\}$. Let Y_i be independent binary variables with $P\{Y_i = 0\} = \frac{i-b_i}{i}$, $P\{Y_i = 1\} = \frac{b_i}{i}$. Let π be chosen uniformly from S_n. Then*

(1) $P(I(\pi) = j) = P(X_1 + X_2 + \cdots X_n = j)$

(2) $P(C(\pi) = j) = P(Y_1 + Y_2 + \cdots Y_n = j)$

Remark 3.1 The statistics $I(\pi)$ and $n - C(\pi)$ are standard distances on the permutation group used as measures of disarray (see Diaconis and Graham(1977) or Diaconis(1988, chapter 8)). Indeed, $I(\pi)$ is the minimum number of pairwise adjacent transpositions required to bring π to the identity and $n - C(\pi)$ is the minimum number of transpositions required to bring π to the identity. Further, $I(\pi)$ is affinely related to Kendall's tau, a non parametric measure of association. The corollary represents these statistics as sums of simple independent random variables. This allows easy calculation of means and variances and a proof of the central limit theorem with Edgeworth corrections. For further details see Feller (1968).

Remark 3.2 One natural notion of rank test involves statistics of form $\sum_{i=1}^{n} m(i, \pi(i))$, with π uniformly chosen in S_b. Efron and Petrosian (1992) have introduced an apparently different notion of rank test with a simple distribution theory based on lemma 3.2. The relation between these rank tests and the distribution theory above are open problems.

   While Corollary 3.1 is well known in the combinatorial literature, it is often rediscovered by statisticians, see Tsai (1990).

3.3   An example with historical insights

Karl Pearson considered a natural source of censored observations in his work on what is now called quasi independence. He considered families with one or more imbecilic children, cross tabulating the family size versus the birth order of the first such child. Clearly a family of j children can only have its first born special child in a position between one and j and consequently, $T_{ij} = 0$ for $i > j$. Pearson carried out a test of independence with this truncated dataset in 1913! A historical report on Pearson's work and its later impact is given by Stigler (1992).

   It is worth beginning with an exact quote of Pearson's procedure from the article by Elderton et al. (1913).

   "Lastly we considered the correlation between the imbecile's place in the family and the gross size of that family. Clearly the size of the family must always be as great or greater than the imbecile's place in it, and the correlation table is accordingly one cut off at the diagonal, and there would certainly be correlation, if we proceeded to find it by the usual product

moment method, but such correlation is, or clearly may be, wholly spurious. Such tables often occur and are of considerable interest for a number of reasons. They have been treated in the Laboratory recently by the following method: one variate x is greater than or equal to the other y; let us construct a table with the same marginal tables, such that y is always equal to or less than x, but let its value be distributed according to an "urn-drawing" law, i.e. purely at random. This can be done. We now have two tables, one the actual table, the other one with the same marginal frequencies, would arrive if x and y were distributed by pure chance but subject to the condition that y is equal or less than x, this table we call the independent probability table. Now assume it to be the theoretical table, which is to be sampled to obtain the observed table, and to measure by χ^2 and P the probability that the observed result should arise as a sample from the independent probability table."

We find this paragraph remarkable as an early clear example of the conditional permutation interpretation of the chi-square test. A careful reading reveals that Pearson is not explicit about the "urn drawing" commenting only that *this can be done*. In the rest of this Section we give an explicit algorithm by translating the problem into that of generating a random permutation with restrictions of the one-sided type and showing that lemma 3.2 achieves a particularly simple form.

To begin with, it may be useful to give the classical justification for Fisher's exact test of independence in an uncensored table. Let T_{ij} be a table with row sums r_i and column sums c_j . Under independence the conditional distribution of T_{ij} given r_i, c_j is the multiple hypergeometric. This may be obtained and motivated as a permutation test as follows. Suppose the n individuals counted by the table have row and column indicators $(X_i, Y_i)_{1 \leq i \leq n}$ with $1 \leq X_i \leq I, 1 \leq Y_i \leq J$. The usual permutation test chooses $\pi \in S_n$ at random forming a new data set $(X_i, Y_{\pi(i)})$. The $I \times J$ table formed by this dataset has the multiple hypergeometric distribution. Said another way, here is a simple algorithm for generating a random table drawn from the multiple hypergeometric distribution :

- Place N balls in an urn with r_i of color i.

- Draw c_1 balls without replacement and set T_{i1} to be the number of balls of color i among these, $1 \leq i \leq I$.

- Draw c_2 balls from the remainder without replacement and set T_{i2} to be the number of balls color i among these...

- $\vdots$

We can now mimic these computations for the triangular table considered by Pearson. Call the table entries n_{ij} with $n_{ij} = 0$ for $j > i$. Suppose $n = \sum_{i \geq j} n_{ij}$. The original data can be regarded as $(X_k, Y_k)_{1 \leq k \leq n}$ with $Y_k < X_k$. This falls into the truncated data pattern with $I(x_i) = [1, x_i]$. Following the prescription of Lemma 3.1, choose the labels i so that $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ and let $A_{ij} = \begin{cases} 1 & \textit{if } Y_j \in I(X_i) \\ 0 & \textit{else} \end{cases}$. The algorithm of Lemma 3.2 translates to the following algorithm to generate the triangular table will row sums $r_1, \ldots r_I$, column sums $c_1, \ldots c_J$, and $T_{ij} = 0$ if $j > i$.

- Place c_1 balls labeled 1 in an urn and sample r_1 of these without replacement. Let T_{11} be the number of balls in the sample labeled 1.

- Add c_2 balls labeled 2 to the urn. Sample r_2 from the urn without replacement. Let T_{2i} be the number of balls labeled i in the sample $i = 1, 2$.

- $\vdots$

Remark 3.3 It is clear from Pearson's discussion following the quote above that he was aware of essentially this algorithm. He used it to give a closed form expression for the maximum likelihood estimates.

Very similar upper triangular tables arise in genetics in testing goodness of fit of the Hardy-Weinberg equilibrium model. The analogous exact sampling scheme is well-known. Recently, Markov chain Monte Carlo techniques have been used to to do the sampling; Guo and Thompson (1992) derive such an algorithm which is a further studied in Diaconis, Graham and Holmes(1999). Lazzeroni and Lange(1997) give a stopping time approach which is an early example of exact sampling. All of these ideas can be extended to the one-sided censoring case.

3.4   Two-sided restrictions

All of the neat factorizations and sampling schemes in Section 3.3 disappear in the case of two-sided truncation. In this Section we introduce a graph structure on permutations in S_A. This gives a reversible Markov chain on S_A which can be run to calibrate permutation tests. Further, the graph structure gives an exchangeable pair so that Stein's method may be used to approximate the distribution of rank statistics as in Stein (1986), Bolthausen (1984), and Bai, Chao, Liang and Zhao(1997).

Lemma 3.3 *Let A be the zero-one matrix of dimension n. Suppose that for all i, $A_{ii} = 1$ and that the ones in each row of A lie in an interval. Define a graph G with vertex set S_A and edges between σ and τ for σ and τ that differ by a transposition of labels. This graph G is connected.*

Proof   By construction, $Id \in S_A$. Connectedness is proved by showing that any $\sigma \in S_A$ is connected to Id. This in turn is proved by regarding σ as a product of disjoint cycles and showing that any cycle can be broken into two resulting in another permutation still in S_A. Towards this end, it is useful to picture the permutations superimposed on the matrix A coloring the places in A corresponding to the permutation representation of σ. For example $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 4 & 3 & 2 \end{pmatrix}$ appears as

$$(3.15) \qquad \begin{pmatrix} 1 & 1 & 1 & 1 & \boxed{1} \\ \boxed{1} & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & \boxed{1} & 0 \\ 0 & 1 & \boxed{1} & 1 & 1 \\ 0 & \boxed{1} & 1 & 1 & 1 \end{pmatrix}$$

The rows correspond to positions, the columns to labels. An allowable transition can be pictured on A as well. For example, transposing labels 2 and 5 gives $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{pmatrix} \in S_A$. The transition is pictured via the parentheses in display (3.16):

$$(3.16) \qquad \begin{pmatrix} 1 & \boxed{1} & 1 & 1 & \boxed{(1)} \\ \boxed{1} & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & \boxed{1} & 0 \\ 0 & 1 & \boxed{1} & 1 & 1 \\ 0 & \boxed{(1)} & 1 & 1 & \boxed{1} \end{pmatrix}$$

In general, transposing two labels is admissible if and only if there are two ones in the two available places in A. In the example, σ is a product of two cycles $(1, 5, 2)$ and $(3, 4)$. It will be useful to picture moving along the cycle on the picture of σ on A. From a boxed square at (i, j) move to diagonal (j, j) and then to the unique box in row j. Finally, observe that given a cycle $(i_1, i_2, \ldots, i_\ell)$ with i_1 smallest, the submatrix $\tilde{A}$ of A formed by rows $\{i_1, i_2, \ldots, i_\ell\}$ and columns $\{i_1, i_2, \ldots, i_\ell\}$ has the row interval property with ones on the diagonal. For example, the cycle $(1, 5, 2)$ above the gives $\tilde{A}$ as

$$(3.17) \qquad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Beginning the cycle with its smallest elements results in $\tilde{A}$ having the leftmost boxed element in the first column.

We may thus study the submatrix corresponding to the cycle $(i_1, i_2, \ldots, i_\ell)$ in a matrix A with ones on the diagonal and having the ones in each row in an interval use $\{i_1, i_2, \ldots, i_\ell\}$ to label the rows and columns. Thus the matrix appears as

$$
\text{(3.18)} \qquad
\begin{array}{c}
i_1 \\ \\ \\ \\ \\ \\ i_2 \\ \\
\end{array}
\begin{bmatrix}
1 & 1 & \cdots & \cdots & \cdots & \cdots & \boxed{1} & 1 \\
1 & \cdots & \cdots & \cdots & 1 & 1 & \cdots & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & \cdots & \cdots & \cdots & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & \cdots & \cdots & \cdots & \cdots & 1 & 1 \\
\boxed{1} & 1 & 1 & 1 & \cdots & \cdots & \cdots & 1 \\
1 & \cdots & \cdots & & \cdots & 1 & 1 & 1 \\
\end{bmatrix}
$$

The proof proceeds in two cases:

Case 1 $i_\ell > i_2$ Then, by the row interval property there is a 1 in positions (i_1, i_1) and (i_ℓ, i_2). Thus labels i_2 and i_ℓ can be transposed.

Case 2 $i_\ell < i_2$ Following around the cycle starting with the box in the first row leads to the diagonal (i_2, i_2). Suppose the box at position (i_2, i_3) in row i_2 is to the left of (i_2, i_2) then by the row interval property labels i_3 and i_3 can be transposed. Finally, consider the case where the box in row i_2 is to the right of (i_2, i_2) the position is thus pictured:

$$
\text{(3.19)} \qquad
\begin{bmatrix}
1 & 1 & \cdots & \cdots & \cdots & \cdots & 1 & 1 \\
1 & \cdots & \cdots & \cdots & 1 & 1 & \cdots & 1 \\
\boxed{1} & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 \\
1 & 1 & \cdots & \cdots & \cdots & 1 & 1 & 1 \\
1 & \cdots & \cdots & \cdots & (i_2, i_2) & \cdots & 1 & \vdots \\
1 & 1 & \cdots & \cdots & \cdots & \cdots & 1 & \vdots \\
1 & 1 & 1 & 1 & \cdots & \cdots & \cdots & 1 \\
\boxed{1} & \cdots & \cdots & \cdots & 1 & 1 & 1 \\
\end{bmatrix}
$$

The reader may picture the horizontal ℓ_H and the vertical line ℓ_V through (i_2, i_2). The path along the cycle next goes down to (i_3, i_2) and continues. Eventually, the cycle path must hit the box in position (i_ℓ, i_1). To do this, it must cross the vertical line ℓ_V. Consider the first time this happens, the path must cross the line from right to left winding up in a box at position (i_r, i_c) with $i_r > i_2, i_c < i_2$. By the row interval property there are ones between (i_r, i_c) and (i_r, i_r). Thus $A_{i_1 i_2} = 1$. Further, in the first row $A_{i_1 i_c} = 1$. It follows that labels i_2 and i_c can be transposed.

This covers all cases and completes the proof. ∎

Remark 3.4 The graph of Lemma 3.3 can be used to run a Markov chain on S_A. From $\sigma \in S_A$ choose one of $\binom{n}{2}$ transpositions uniformly at random and transform σ by switching the two chosen labels. If this new permutation is in S_A, the walk moves there. If not, the walk stays at σ. This results in a symmetric connected Markov chain which has uniform stationary distribution on S_A. The chain is aperiodic whenever A is not the matrix entirely filled with ones.

This chain allows calibration of any test statistic. Its proper use needs an estimate of the relaxation time. We have carried out some preliminary work which suggests that order $n^2 log n$ steps are sufficient in the case of interval structure for A, this remains a conjecture. We remark that Hanlon(1996) has determined all the eigenvalues of this Markov chain for the case of one-sided truncation.

Remark 3.5 For more general zero patterns, the graph based on transpositions need not be connected. For example, consider the derangements in S_3. The matrix is $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. The two derangements are $(1,3,2)$ and $(1,2,3)$ in cycle notation. These are even permutations and cannot be connected by transpositions.

The 3×3 matrix for derangements (call it D) can be used to construct larger examples which show the difficulty of making a general theory. For example, construct a $3n \times 3n$ matrix A by placing copies of D down the diagonal, zeros in the remaining upper triangular part and ones in the remaining lower triangular part. It is easy to see that there are 2^n permutations in S_A and that **none** of these can be connected to any others by transpositions. As a second example, construct an $n \times n$ matrix A with a single copy of D in the upper left hand corner. Place zeros in the remaining places of the first three rows and ones everywhere else. Here, there are two giant components in S_A which cannot be connected by transpositions.

We have shown that for $n \geq 4$ the set of all derangements is connected by transpositions. The argument proceeds by showing that any derangement can be brought to the n-cycle $(1,2,3,4,\ldots,n)$ in at most $n-1$ transpositions.

While not exploited in the current paper, there is a large class of examples of sets of permutations which are connected by transpositions; these are the linear extensions of a given partial order. See Brightwell and Winkler (1991) for an overview and references.

Remark 3.6 The Markov chain of Remark 1 above gives an exchangeable pair of random permutations (X, X'), with X chosen from the uniform distri-

bution on S_A and X' one step of the chain away from X. Such an exchangeable pair forms the basis of Stein's approach to the study of Hoeffding's combinatorial limit theorem. Bolthausen (1984) and Schneller (1989) used extensions of Stein's method to get the right Berry-Esseen bound and Edgeworth corrections. Zhao, Bai, Chao and Liang (1997) give limit theorems for double indexed statistics (à la Daniels) of form $\sum a(i, j, \pi(i), \pi(j))$ using Stein's method. Finally, Mann(1995) and Reinert(1998) have used Stein's method of exchangeable pairs to show that the chi-square test for independence in contingency tables has an approximate χ^2 distribution, conditional on the margins.

We have used the exchangeable pair described above to prove normal and Poisson limit theorems for the number of fixed points in a permutation chosen randomly from the set S_A. There is a lot more work to be done. We note in particular that the limiting distribution of linear rank statistics is an open problem with even one-sided truncation. The distribution of Kendall's tau is an open problem in the case of two-sided truncation.

We close this Section with a statistical comment and a useful lemma. The widely used non parametric measure of association Kendall's tau applied to paired data $\{(x_i, y_i)\}$ can be described combinatorially as follows: Sort the pairs by increasing values of x_i. Then calculate the minimum number of pairwise adjacent transpositions required to bring $\{y_i\}$ into the same order as $\{x_i\}$. When working with restricted positions, it is natural to ask if any admissible permutation can be brought to any other by pairwise adjacent transpositions.

The following example shows that this is not so. For $n = 3$ consider the matrix $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$. There are two admissible permutations $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. No pairwise adjacent transpositions of the labels is allowable. The matrix has the row interval property and all transpositions connect.

It is not hard to see that pairwise adjacent transpositions connect all admissible permutations in the one-sided case.The following lemma proves connectedness in the monotone case: The intervals $I(x_i) = (a_i, b_i)$ can be arranged so that $a_1 \leq a_2 \leq a_3 \leq \cdots \leq a_n; b_1 \leq b_2 \leq b_3 \leq \cdots \leq b_n$. For the monotone case, order i by a_i increasing. Then, it is easy to see that the restriction matrix $A_{ij} = \begin{cases} 1 & if \ y_j \in I(x_i) \\ 0 & else \end{cases}$ has both row interval property and column interval property, and $A_{ii} = 1$. Call such a restriction matrix monotone.

Lemma 3.4 *Let A be a n-dimensional monotone restriction matrix. Form a graph with vertex set S_A and an edge between two permutations if and only if they differ by a pairwise adjacent transposition of labels. This graph is connected.*

Proof By construction, S_A contains the identity. The argument proceeds by showing that any permutation $\pi \in S_A$ can be brought to the identity by pairwise adjacent transpositions. It is useful to picture π on A by boxing entry (i, j) if $\pi(i) = j$.

$$(3.20) \qquad \begin{pmatrix} 1 & 1 & \boxed{1} & 0 \\ 1 & 1 & 1 & \boxed{1} \\ \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{1} & 1 & 1 \end{pmatrix} \qquad \text{represents} \qquad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$$

A pairwise adjacent transposition corresponds to a basic move in two adjacent rows

$$\boxed{} \cdot \; \cdot \; \cdot \; \cdot \; \cdot \qquad \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \boxed{}$$
$$\cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \boxed{} \quad \longrightarrow \quad \boxed{} \cdot \; \cdot \; \cdot \; \cdot \; \cdot \quad .$$

Recall that π has an inversion at (i, j) if $i < j$ and $\pi(i) > \pi(j)$. Only the identity has no inversions. Consider two consecutive rows $i, i + 1$, if $\pi(i) > \pi(i + 1)$ the picture appears as

$$\begin{array}{ll} i & \qquad\qquad\quad \boxed{} \\ i+1 & \quad \boxed{} \end{array} \; .$$

Consider the first row for which $\pi(i) > i$, if $\pi \neq identity$, such a row must exist. By the row and column interval property a basic move can be made. This reduces the number of inversions by one. Continue until no pair of adjacent rows has an inversion. This gives the identity. ∎

Remark 3.7 While lemma 3.3 shows the graph is connected we have found simple examples of monotone restriction matrices where the distance between pairs and agreeing permutations is not given by Kendall's tau. For example, let A be a 4×4 matrix with ones everywhere except in the upper right and lower left corners. The two permutations $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$ and $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$ have Kendall's distance 5, but their graph distance is 6.

Remark 3.8 In the case of discrete data(contingency tables) arbitrary patterns of truncation can be handled using the moves in Diaconis and Sturmfels (1998). For this case, the moves are easy to specify. Suppose the x-variable takes on I levels and the y variable takes on J levels. Form the complete bi-partite graph on the set $\{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\}$. If cell (i, j) is unobservable, delete this edge. The circuits in the remaining graph form a connecting set of moves successively adding and deleting one while traversing the circuit. This suggests two lines of generalization. First, continuous data can be treated by discretization. Second, truncated multivariate data can be approached using the multiway table moves from Diaconis and Sturmfels (1998).

4   An application to ESP guessing experiments

This final Section gives a different set of applications for permutations with restricted positions.

A classical test of parapsychology involves a deck of 25 cards made up of the following five symbols of

Each repeated five times. Under ideal conditions the deck is well shuffled and a guessing subject attempts to guess at the cards in order. Under the natural chance model, each guess has chance $\frac{1}{5}$ of being correct and so the expected number of correct guesses is five. Of course the distribution of the number of correct guesses depends on the guessing sequence. If the subject always guesses the same symbol, there is no variability. It is not hard to show that the variance of the number of correct guesses is largest if the subject guesses some permutation of the values. In Diaconis and Graham(1981), we studied variations where feedback was given to the subject after each guess. For example, suppose the subset is shown the card at position i after guess i (complete feedback). Then the optimal strategy is to guess a card with highest frequency among those remaining. Read (1962) shows in this case the expected number of guesses is 8.65. This is of some practical interest since many early experiments were done with feedback and 8.5[+] is reported as the highest of average trials in actual trials.

The most interesting type of feedback is *yes/no* feedback: if a subject guesses correctly, they are told so. If they are incorrect they are only given that information. Now, the subjects optimal strategy is **not** obvious. We have shown in Diaconis and Graham (1981), that the greedy strategy (guess the most likely value), is only close to optimal. We also determined that the

expected number of correct guesses under the optimal strategy is 6.63.

The reason for discussing these matters here is twofold. First; the evaluation of the probabilities involved uses permanents. Second; a natural monotonicity conjecture proved by a longish combinatorial argument in Chung, Diaconis, Graham and Mallows (1981) follows from one tool developed to prove the van der Waerden conjecture. To define things, let $N(a_1, a_2, \ldots, a_r; b_1, b_2, \ldots, b_r)$ be the number of arrangements of a deck of $a_1 + a_2 + \cdots a_r = n$ cards with a_i of type i, such that symbol one does not appear in the first b_1 places, symbol 2 does not appear in places $b_1 + 1, b_1 + 2, \ldots, b_1 + b_2$, and so on.

This quantity allows evaluation of the probabilities of events like: The next card is type i given b_j "no" responses on type j. From the definition, $N(\mathbf{a}, \mathbf{b}) = Per(M)$ where M is the $n \times n$ zero/one matrix of the form:

$$M = \begin{bmatrix} [m_1] & 1 & 1 & \cdots & \cdots & \cdots & 1 \\ 1 & [m_2] & 1 & \cdots & \cdots & \cdots & 1 \\ 1 & 1 & [m_3] & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & \ddots & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & \ddots & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & \ddots & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & [m_r] \end{bmatrix}$$

with m_i an $a_i \times b_i$ block of zeros and the rest all ones. The inequality to be proved is:

Proposition 4.1 *For any* $\mathbf{a}, \mathbf{b}$ *with* $N(\mathbf{a}, \mathbf{b}) \neq 0$, *and any* $i, 1 \leq i \leq r$

$$\frac{N(a; b + e_i)}{N(a, b)} \geq \frac{N(a; b + 2e_i)}{N(a, b + e_i)}$$

with $e_i = (0, 0, \ldots, 1, 0, \ldots, 0)$ *the usual ith basis vector.*

Remark 4.1 In the card guessing context, the inequality has the following interpretation: in a yes-no feedback experiment; the chance that the guess at the next card is of type i cannot decrease if the next card is of type i and is incorrect.

Proof The argument uses a quadratic form defined on $\mathbf{R}^n$. Given positive vectors $V_1, V_2, \ldots, V_{n-2}$ in $\mathbf{R}^n$, define

$$< x|y > = Per[V_1, V_2, \ldots, V_{n-2}, x, y]$$

This is a symmetric bilinear form on $\mathbf{R}^n$. It is used in a crucial way in Egorychev's proof of the permanent conjecture. We follow the account in

van Lint and Wilson (1992). They show (Theorem 12.6) that this form is Lorentzian having $n - 1$ negative eigenvalues and one positive eigenvalue. Such forms are easily seen to satisfy a "reverse Cauchy-Schwarz inequality"; for x positive and y arbitrary.

(4.21) $< x|y >^2 \geq < x|x >< y|y >$

By continuity, (4.21) also holds if some of the entries in V_i or x are allowed to be zero. Proposition 4.1 follows from (4.21). By symmetry of the permanent, it is enough to prove it for e_r. Consider the three matrices corresponding to $a_r \times b_r, a_r \times b_r + e_r, a_r \times b_r + 2e_r$. Move these rth blocks to the right of the full matrix. The last two columns of the full matrices with these blocks appear as

$$a_r \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \right.$$

All other columns are the same. Call the two columns from the first matrix x and y and apply (4.21). ∎

Remark 4.2 In Chung, Diaconis, Graham and Mallows (1981) it was in fact shown that $n_k = N(a, b + ke_i)$ is log-concave: $n_k^2 \geq n_{k+1}n_{k-1}$. Their proof was combinatorial and only worked for zero-one matrices. It is not clear if there is an analog of log-concavity for more general Lorentzian forms.

REFERENCES

Albers, W., Bickel, P. J,. van Zwet, W. R.(1976) Asymptotic expansions for the power of distribution-free tests in the one-sample problem (Corr: V6 p1170), *Annals of Statistics*,4, 108-156.

Bai,Z. Chao,C.C.,Liang C., Zhao L. (1997) Error bounds in a central limit theorem of doubly indexed permutation matrices, *Annals of Statistics*, 25, 2210-2227.

Bapat R. B. (1990) Permanents in probability and statistics, *Linear Algebra and its Applications*, 127, 3-25.

Barvinok,A. (1998) *A simple polynomial time algorithm to approximate the permanent within a simply exponential factor. Preprint, Mathematical Sciences Research Center, Berkeley.*

Bickel P. (1969) A distribution-free version of the Smirnov two-sample test in the *p*-variate case. *Annals of Mathematical Statistics*, 40, 1-23.

Blum, J. R. Kiefer, J. and Rosenblatt, M., (1961) Distribution free tests of independence based on the sample distribution function, *Annals of Mathematical Statistics*, 485-49.

Bolthausen E. (1984) An estimate of the remainder in the combinatorial central limit theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 66, 379-386.

Brightwell, G. and Winkler P. (1991), Counting linear extensions,*Order*, 8, 225-242.

Caussinus H. (1966) Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de Toulouse*(année 1965), 29, 77-182.

Chung F., Diaconis P., Graham R. and Mallows C. L. (1981) On the permanents of the complements of the direct sum of Identity matrices, *Advances in Applied Maths.*, 2, 121-137.

Clogg, C.C. (1986) Quasi-independence, *in Encyclopedia of Statistical Sciences (9 vols. plus Supplement)* 7, 460- 464.

Cook, W. (1998) *Combinatorial Optimization*, Wiley, NY.

Daley, D. J. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes*, Springer-Verlag,NY.

Diaconis, P. (1988) *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, California.

Diaconis, P. and Graham R. (1977), Spearman's footrule as a measure of disarray *Journal of the Royal Statistical Society*, Series B, 39, 262-268.

Diaconis, P. and Graham R. (1981) The analysis of sequential experiments with feedback to the subjects. *Annals of Statistics*, 9, 3-23.

Diaconis P. and Sturmfels B. (1998), Algebraic algorithms for sampling conditional distributions, *Annals of Statistics*, 26, 363-397.

Efron B. and Petrosian V. (1992), A simple test of independence for truncated data with applications to red-shift surveys, *Astrophysical Journal*, 399,345-352.

Efron B. and Petrosian V. (1999), Non parametric methods for doubly truncated data, *Journal of the American Statistical Association*, 94, 824-834.

Elderton E.M., Barrington A., Jones H.G., Lamotte E.M., Laski H.J., Pearson, K. (1913) On the correlation of fertility with social value: A cooperative study, *Eugenics Laboratory Memoirs XVIII*, University of London.

Feller W. (1968) *Introduction to Probability Theory and its applications*, vol I, 3rd edition, Wiley, New York.

Goodman L.A. (1968), The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries, *Journal of the American Statistical Association*, 63, 1091-1131.

Garey M.R. and Johnson D.S. (1979) *Computers and intractability, a guide to the theory of NP-completeness*, Freeman and co., San Francisco.

Godsil C.D. (1981) Matching behavior is asymptotically normal. *Combinatorica*,1,369-376.

Graham R.L., Knuth D., Patashnik, (1989) *Concrete Mathematics*, Addison Welsley, Reading, MA.

Guo, S. W., Thompson, E. A.,(1992), Performing the exact test of Hardy-Weinberg proportion for multiple alleles, *Biometrics*, 48, 361-372

Hanlon P. (1996), A random walk on the rook placements on a Ferrers board. *Electronic Journal of Combinatorics*, vol 3,
        `http://math34.gatech.edu:8080/Journal/journalhome.html`

Hoeffding, W. (1948), A non-parametric test of independence, *Annals of Mathematical Statistics*, 19, 546-557.

Jerrum, M.R. and Sinclair,A.J.(1989), Approximating the permanent. *SIAM Journal on Computing* **18** , 1149-1178.

Jerrum, M.R. and Vazirani,U. (1992) A mildly exponential approximation algorithm for the permanent. *Proceedings of the 33rd Annual IEEE Conference on Foundations of Computer Science*, 320-326.

Karmarkar,N.,Karp, R.,Lipton,R., Lovász,L. and Luby,M., (1993) A Monte-Carlo algorithm for estimating the permanent. *SIAM Journal on Computing* **22**, 284-293.

Lai, T.L. Ying, Z. (1991), Rank regression methods for left-truncated and right-censored data, *Annals of Statistics*, 19, 531-556.

Lazeronni L. and Lange K. (1997), Markov chains for Monte Carlo tests of geneteic equilibrium in multidimensional contingency tables, *Annals of Statistics*, 138-168.

Lehmer D.H. (1970) Permutations with strongly restricted displacements, in *Combinatorial Theory and its Applications II*, Eds. Erdős P., Renyi A., Sós V., 755-770 North Holland, Amsterdam.

Lieb E.H. (1966) Proofs of some conjectures on Permanents, *Journal of Mathematics and Mechanics*, 16, 127-134.

Lovasz L. and Plummer (1986) *Matching Theory*, North Holland, Amsterdam.

Lynden-Bell, D. (1971) A method of allowing for a known observational selection in small samples applied to 3CR quasars, *Monographs of the National Royal Astrophysical Society*, **155**, 95-118.

Lynden-Bell, D. (1993) Eddington-Malmquist bias, streaming motions, and the distribution of galaxies (Disc: p217-220) in *Statistical Challenges in Modern Astronomy*, ed. Babu J. and Feigelson E., 201- 216.

Mallows,C. (1957) Non-null ranking models I, *Biometrika*, 44, 114-130.

Mann B.(1995) Ph D thesis, Harvard University, Cambridge,MA.

Pitman J. (1997) Probabilistic bounds on the coefficients of polynomials with only real zeros, *Journal of Combinatorial Theory*,A, 77, 279-303.

Rasmussen,L.E.(1998) *On approximating the permanent and other #P-complete problems.* PhD Thesis, University of California at Berkeley.

Rasmussen,L.E.(1994) Approximating the permanent: a simple approach. *Random Structures and Algorithms* 5, 349-361.

Read R. (1962) Card guessing with information-a problem in probability. *American Mathematics Monthly*, 69, 506-511.

Reinert G. (1998) Stein's method for the Chi-2 statistic, *Technical Report.*

Riordan J. (1958) *An introduction to combinatorial theory*, Wiley, N.Y.

Romano, J. P. (1989) Bootstrap and randomization tests of some nonparametric hypotheses, *Annals of Statistics*, 17, 141-159

Schneller,W. (1989), Edgeworth expansions for linear rank statistics, A*nnals of Statistics*, 17, 1103-1123.

Sinclair A. (1992) *Algorithms for random generation and counting. A Markov chain approach.* Birkhaeuser.

Stein C. (1986) *Approximate computation of expectations* Institute of Mathematical Statistics, Lecture Notes and Monographs, Hayward, California.

Stembridge J. (1991) Immanants of totally positive matrices are nonnegative, *Bulletin of the London Mathematics society*,23, 422-428.

Stembridge J. (1992) Some conjectures for Immanants, *Canadian Journal of Mathematics*, 44, 1079-1099.

Stanley R. (1986) *Enumerative Combinatorics*, volume I, Wadsworth/Brooks, Cole, Monterey, CA.

Stigler S. (1992) Studies in the history of probability and statistics XLIII. Karl Pearson and quasi-independence, *Biometrika*,79,563-575.

Tsai,W.Y. (1990), Testing the assumption of independence of truncation time and failure time, *Biometrika*, 77, 169-177.

Tsui, K.L., Jewell, N. P. Wu, C. F. J. (1988) A nonparametric approach to the truncated regression problem, *Journal of the American Statistical Association*,83, 785-792.

Valiant L.,(1979) The complexity of computing the permanent, *Theoretical Computer Science*, 8, 189-201.

Van Lint J. and Wilson,R. (1992) *A course in combinatorics* , Cambridge University Press, Cambridge.

Zeckendorf,E. (1972) Representation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas, *Bulletin de la Société Royale Scientifique de Liége*, 41,179-182.

PERSI DIACONIS
MATHEMATICS AND STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
CA 94305-4065

RONALD GRAHAM
COMPUTER SCIENCE
UNIVERSITY OF CALIFORNIA AT SAN DIEGO
AND ATT, FLORHAM PARK, NJ.
graham@ucsd.edu

SUSAN HOLMES
STATISTICS
STANFORD UNIVERSITY
AND UNITÉ DE BIOMÉTRIE
INRA-MONTPELLIER, FRANCE
susan@stat.stanford.edu