Statistica Sinica **17**(2007), 857-872

SEQUENTIAL MONTE CARLO METHODS FOR PERMUTATION TESTS ON TRUNCATED DATA

Yuguo Chen and Jun S. Liu

University of Illinois at Urbana-Champaign and Harvard University

Abstract: The permutation test is one of the oldest techniques for making statistical inferences. Monte Carlo methods and asymptotic formulas have been used to approximate the associated p-values. When data are truncated, however, the permutation null distribution is difficult to handle. We describe here an efficient sequential importance sampling strategy for generating permutations with restricted positions, which provides accurate p-value approximations in all examples we have tested. The algorithm also provides good estimates of permanents of zero-one matrices, which by itself is a challenging problem. The key to our strategy is a connection between allowable permutations and zero-one tables with structural zeros.

Key words and phrases: Importance sampling, Markov chain Monte Carlo, permanent, permutation test, structural zero, zero-one table.

1. Introduction

Randomization and permutation are at the heart of statistical inference. Using permutations to generate the null distribution of a test statistic is one of the oldest and perhaps still the most popular statistical techniques. When assuming a parametric distribution for the underlying data is unsafe, permutation tests provide a solid probabilistic basis for making inference. One "inconvenience" of using permutation tests is that their null distributions are usually not in closed form, and some asymptotic or numerical approximations are needed to provide valid p-values. Still, except for some special cases, no good approximations are available for permutation tests with truncated data. The goal of this paper is to develop efficient Monte Carlo methods to approximate the p-values of permutation tests.

According to Stigler (1992) and Diaconis, Graham and Holmes (2001), Karl Pearson in 1913 already considered the problem of testing independence with truncated data. More specifically, he was interested in whether, among those families with one or more mentally disabled children, the birth order Y of the first such child is related to the family size X. Apparently, the birth order of the first such child is truncated up by the family size, i.e., $1 \leq Y \leq X$ by design. Thus, the usual correlation coefficient measure will only reveal a spurious relationship between X and Y. To obtain the null distribution of the chi-square statistic under the truncation constraint, Karl Pearson suggested an "urn-drawing" randomization law under the constraint $Y \leq X$, which is just the uniform distribution on all permutations of the Y that satisfy the constraint (Diaconis et al. (2001)).

A more challenging case was studied in Efron and Petrosian (1999), in which they were interested in whether the redshift X and the logarithm of luminosity Y for quasars are independent. The answer to this question can shed light on the suggested theory that earlier quasars were brighter. Due to experimental constraints, we can observe Y if and only if $Y \in [l(X), u(X)]$. Figure 1 shows the quasar data of Efron and Petrosian (1999), in which $n = 210$. The "apparent" relationship between X and Y may be due to the truncation effect. Truncated data also arise in many other situations, such as survival analysis (McLaren, Wagstaff, Brittegram and Jacobs (1991) and Bilker and Wang (1996)). Astronomers have developed permutation tests based on Kendall's tau for dealing with these problems (Lynden-Bell (1971)).

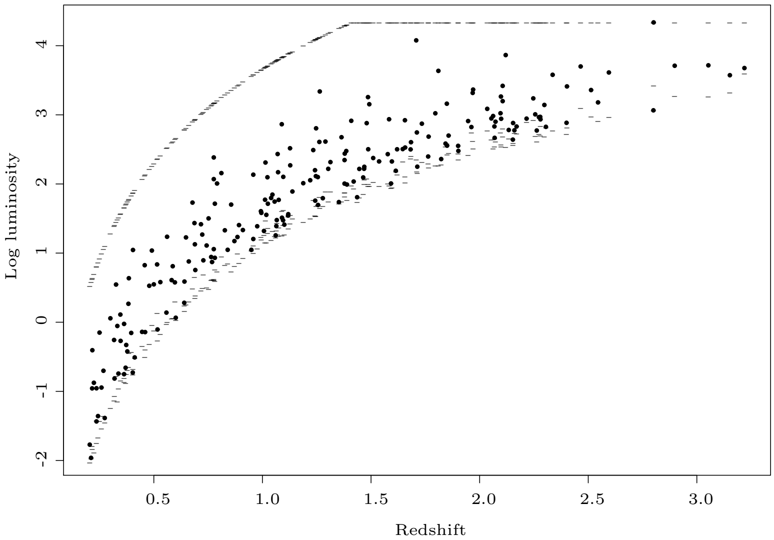

Figure 1. Doubly truncated data from an astronomical study of Efron and Petrosian (1999). Redshifts and log-luminosities for 210 quasars are denoted by points. Upper and lower truncations for log-luminosities are indicated by "-".

A general setting of the problem is as follows. Let (X, Y) be a pair of random variables, and observe n independent realizations, $(x_1, y_1), \ldots, (x_n, y_n)$, under the constraint that we can observe (X, Y) if and only if $Y \in S(X)$, where $S(X)$ is a set dependent on X. Of interest is whether X and Y are independent. Intuitively, if X and Y are indeed independent, then the independent realizations $(x_1, y_1), \ldots, (x_n, y_n)$ are as likely as any permuted version: $(x_1, y_{\sigma(1)}), \ldots, (x_n, y_{\sigma(n)})$, provided that

$$y_{\sigma(i)} \in S(x_i), \quad \text{for all} \;\; i = 1, \ldots, n. \tag{1}$$

Here we let $\sigma = (\sigma(1), \ldots, \sigma(n))$ denote a permutation of $(1, \ldots, n)$.

Permutation tests are natural choices for testing independence nonparametrically because of their optimality properties for data without truncation (Romano (1989)). Efron and Tibshirani (1993) emphasize that "when they (permutation methods) apply, …… …, they give gratifying exact answers without parametric assumptions. …… When there is something to permute, ……, it is a good idea to do so, even if other methods like the bootstrap are brought to bear." When truncation occurs, a main difficulty is the lack of reliable asymptotic distributions of test statistics under the permutation distribution, except for the case with one-sided truncation (Diaconis et al. (2001)). For the quasar data, Efron and Petrosian ((1999), p.218) employed a simple Markov chain Monte Carlo (MCMC) strategy to estimate the p-value. However, more complicated MCMC moves are needed to handle data with more complex truncation constraints because the Markov transitions proposed in Efron and Petrosian (1999) may not be irreducible (Diaconis et al. (2001)). The mixing rate of the resulting Markov chain can be too slow to give a satisfactory estimation accuracy.

As noted in Diaconis et al. (2001), one can record the constraints (1) by a *restriction matrix* A, in which a "1" at its (i, j)-th entry indicates that $y_j \in S(x_i)$, and a "0" otherwise. Then the total number of allowable permutations (i.e., all σ's that satisfy (1)) is equal to the *permanent* of the zero-one matrix A. In general, the permanent of a matrix $A = (a_{ij})_{n \times n}$ is defined as

$$\text{perm}(A) = \sum_{\sigma \in \mathcal{P}} \prod_{i=1}^{n} a_{i\sigma(i)}, \tag{2}$$

where $\mathcal{P}$ is the set of all permutations of $\{1, \ldots, n\}$.

Approximating permanents is one of the most challenging and well-studied problems in theoretical computer science and applied mathematics. It is known that the exact calculation of the permanent is a #P-complete problem (Valiant (1979)). #P-complete problems are a class of intractable counting problems; if one of them is solved in polynomial time, then all other problems in the class can be solved in polynomial time. So far there is no polynomial time algorithm available for solving #P-complete problems. Much effort has been devoted to approximating the permanent by randomized algorithms (Jerrum and Sinclair (1989); Karmarkar, Karp, Lipton, Lovász and Luby (1993), Rasmussen (1994), Jerrum and Sinclair (1996) and Barvinok (1999)), some of which are based on MCMC. Several algorithms based on the importance sampling principle have also been developed (Kuznetsov (1996), Beichl and Sullivan (1999) and Smith and Dawkins (2001)). However these algorithms are not efficient for matrices as large as the 210 × 210 constraint matrix of the quasar data.

In this paper, we propose an efficient sequential importance sampling (SIS) approach to implement permutation tests for truncated data. Here we use the term "truncation" to refer to the general restriction set $S(X)$ instead of only intervals. As a byproduct, our algorithm can also give accurate estimates of permanents of zero-one matrices. We show in both data and simulations that our algorithm significantly outperforms the existing ones in both computing speed and estimation accuracy. In Section 2 we set up the formal problem. In Section 3 we describe the connection of the permutation test problem with the counting of zero-one tables with structural zeros, which gives rise to the two simple SIS strategies to be discussed in Section 4 for generating allowable permutations. We demonstrate the power of the new approaches by several examples in Section 5.

2. Problem Setting

Suppose we have n independent realizations of a pair of random variables, $(x_1, y_1), \ldots, (x_n, y_n)$. For each x, we can observe the pair (x, y) if and only if $y \in S(x)$. Following Diaconis et al. (2001), we use the restriction matrix $A = (a_{ij})_{n\times n}$ to represent the constraints, where a_{ij} equals 1 if $y_j \in S(x_i)$, and 0 otherwise. For example, A_4 is the restriction matrix for the artificial data in Table 1.

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

Table 1. An artificial example of truncated data.

x_i	y_i	$S(x_i)$
1.7	3.2	[2.5, 4.1]
2.2	2.7	[2.0, 3.0]∪[4.5, 6.0]
3.6	5.5	[2.9, 5.7]
4.9	4.6	[3.1, 6.0]

Let $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, and let $\Sigma_{\mathbf{xy}}$ be the set of permutations satisfying the truncation constraints, i.e.,

$$\Sigma_{\mathbf{xy}} = \{\sigma = (\sigma(1), \ldots, \sigma(n)) : \sigma \text{ is a permutation of } (1, \ldots, n) \text{ and } y_{\sigma(i)} \in S(x_i),\ i = 1, \ldots, n\}.$$

We call $(a_{1,\sigma(1)},\ldots,a_{n,\sigma(n)})$ a *permutation sequence* of A if $\sigma=(\sigma(1),\ldots,\sigma(n))$ is an element of $\Sigma_{\mathbf{xy}}$. Clearly, the total number of elements in $\Sigma_{\mathbf{xy}}$ is equal to perm(A).

The p-value for the permutation test can be written as

$$\mu = E_p f(\mathbf{x},\mathbf{y}_\sigma) = \sum_{\sigma\in\Sigma_{\mathbf{xy}}} f(\mathbf{x},\mathbf{y}_\sigma)p(\sigma), \tag{3}$$

where $p(\sigma)$ is the uniform distribution on $\Sigma_{\mathbf{xy}}$, and $f(\mathbf{x},\mathbf{y}_\sigma)$ is a function of the test statistic. One popular way of testing independence is to use Kendall's tau statistic

$$\tau(\mathbf{x},\mathbf{y}_\sigma) = \sum_{(i,j)\in\mathcal{C}(\mathbf{x},\mathbf{y}_\sigma)} \frac{\mathrm{sign}[(y_{\sigma(i)}-y_{\sigma(j)})(x_i-x_j)]}{|\mathcal{C}(\mathbf{x},\mathbf{y}_\sigma)|}, \tag{4}$$

where

$$\mathcal{C}(\mathbf{x},\mathbf{y}_\sigma) = \{(i,j):\ y_{\sigma(i)}\in S(x_j), y_{\sigma(j)}\in S(x_i)\}. \tag{5}$$

In this case,

$$f(\mathbf{x},\mathbf{y}_\sigma) = 1_{\{\tau(\mathbf{x},\mathbf{y}_\sigma)\geq\tau(\mathbf{x},\mathbf{y})\}}. \tag{6}$$

Tsai (1990) and Efron and Petrosian (1999) used Kendall's tau statistic for one-sided truncation and doubly truncated data.

Since sampling from $p(\sigma)$, i.e., drawing a permutation uniformly from $\Sigma_{\mathbf{xy}}$, is difficult, we can use the importance sampling approach: sample $\sigma\in\Sigma_{\mathbf{xy}}$ from a different distribution $q(\sigma)$ with $q(\sigma)>0$ for all $\sigma\in\Sigma_{\mathbf{xy}}$, and estimate μ by

$$\hat{\mu} = \frac{\sum_{i=1}^N f(\mathbf{x},\mathbf{y}_{\sigma_i})\frac{p(\sigma_i)}{q(\sigma_i)}}{\sum_{i=1}^N \frac{p(\sigma_i)}{q(\sigma_i)}} = \frac{\sum_{i=1}^N f(\mathbf{x},\mathbf{y}_{\sigma_i})\frac{1_{\{\sigma_i\in\Sigma_{\mathbf{xy}}\}}}{q(\sigma_i)}}{\sum_{i=1}^N \frac{1_{\{\sigma_i\in\Sigma_{\mathbf{xy}}\}}}{q(\sigma_i)}}, \tag{7}$$

where $\sigma_1,\ldots,\sigma_N$ are i.i.d. samples from $q(\sigma)$. We can also estimate the total number of elements in $\Sigma_{\mathbf{xy}}$, i.e., the permanent of A, by

$$\widehat{\mathrm{perm}}(A) = \frac{1}{N}\sum_{i=1}^N \frac{1_{\{\sigma_i\in\Sigma_{\mathbf{xy}}\}}}{q(\sigma_i)}, \tag{8}$$

because of the simple identity $|\Sigma_{\mathbf{xy}}| = \sum_{\sigma\in\Sigma_{\mathbf{xy}}}[q(\sigma)]^{-1}q(\sigma)$.

A key to successful importance sampling is the design of the sampling distribution $q(\sigma)$, which should be both easy to sample from, and close enough to the target distribution $p(\sigma)$. Since our goal here is to provide accurate estimates for the expected values of a variety of functions under $p(\sigma)$, we use the square of the *coefficient of variation* (cv)

$$cv^2 = \frac{\mathrm{var}_q\{\frac{p(\sigma)}{q(\sigma)}\}}{E_q^2\{\frac{p(\sigma)}{q(\sigma)}\}} \tag{9}$$

to evaluate the efficiency of our importance sampling algorithms (Kong, Liu and Wong, (1994)). A low cv^2 indicates that $q(\sigma)$ is close to $p(\sigma)$.

3. Connection Between Permutations and Zero-One Tables

In the search of an efficient importance sampling distribution, we notice the following connection between permutation sequences (see definition in Section 2) and zero-one tables with structural zeros. Let Ω be a subset of $\{(i,j) : i = 1,\dots,n;\ j = 1,\dots,n\}$, and let $\Sigma_n(\Omega)$ be the set of $n \times n$ zero-one tables with all row and column sums equal to 1 and the set of structural zeros Ω. That is, a zero-one table $Z = (z_{ij})_{n\times n} \in \Sigma_n(\Omega)$ if $z_{ij} = 0$ for all $(i,j) \in \Omega$, and $\sum_{l=1}^n z_{lj} = \sum_{l=1}^n z_{il} = 1$, for $i = 1,\dots,n;\ j = 1,\dots,n$.

Theorem 1. *Let $A = (a_{ij})_{n\times n}$ be a restriction matrix and*

$$\Omega = \{(i,j) : \ a_{ij} = 0, \quad 1 \le i \le n; \quad 1 \le j \le n\}. \tag{10}$$

Then there is a one-to-one correspondence between the set of permutation sequences of A and the set of zero-one tables in $\Sigma_n(\Omega)$.

Proof. For a permutation sequence $(a_{1,\sigma(1)},\dots,a_{n,\sigma(n)})$, we can construct a zero-one table $T = (t_{ij})_{n\times n}$ such that $t_{k,\sigma(k)} = 1$, $k = 1,\dots,n$, and all other entries equal to 0. Since $(\sigma(1),\dots,\sigma(n))$ is a permutation of $(1,\dots,n)$, all row and column sums of the zero-one table T are 1. Since $\prod_{k=1}^n a_{k,\sigma(k)} = 1$, we have $(k,\sigma(k)) \notin \Omega$, $k = 1,\dots,n$. Therefore $T \in \Sigma_n(\Omega)$. On the other hand, for any $T \in \Sigma_n(\Omega)$, the non-zero entries of T can be written as $(t_{1,\sigma(1)},\dots,t_{n,\sigma(n)})$, where $(\sigma(1),\dots,\sigma(n))$ is a permutation of $(1,\dots,n)$. Since $(k,\sigma(k)) \notin \Omega$, $k = 1,\dots,n$, it follows that $\prod_{k=1}^n a_{k,\sigma(k)} = 1$. Therefore $(a_{1,\sigma(1)},\dots,a_{n,\sigma(n)})$ is a permutation sequence of A. We thus established a one-to-one mapping between permutation sequences of A and zero-one tables in $\Sigma_n(\Omega)$.

The connection established in Theorem 1 transforms the problem from the uniform sampling of permutation sequences of A to the uniform sampling of zero-one tables in $\Sigma_n(\Omega)$. If we sample a table T from $\Sigma_n(\Omega)$ and denote those non-zero entries in T as $(t_{1,\sigma(1)},\dots,t_{n,\sigma(n)})$, then $(a_{1,\sigma(1)},\dots,a_{n,\sigma(n)})$ is a permutation sequence of A, and thus $(\sigma(1),\dots,\sigma(n))$ is a permutation satisfying the truncation constraint. This observation is the key to the development of new SIS algorithms for sampling allowable permutations (see Section 4), which makes use of the general SIS strategies for handling zero-one tables with fixed marginal sums and a given set of structural zeros (Chen (2007)).

An immediate application of Theorem 1 is an asymptotic formula for the permanent of a zero-one matrix. Bender (1974) provides an asymptotic approximation to the number of tables in $\Sigma_{\mathbf{rc}}(\Omega^*)$: the set of zero-one tables with row sums $\mathbf{r} = (r_1, \ldots, r_m)$, column sums $\mathbf{c} = (c_1, \ldots, c_n)$, and a set of structural zeros Ω^*:

$$|\Sigma_{\mathbf{rc}}(\Omega^*)| \approx \Delta_{\mathbf{rc}}(\Omega^*) \equiv \frac{M!}{\prod_{i=1}^m r_i! \prod_{j=1}^n c_j!} \times \exp\left\{ -\frac{\left(\sum_{i=1}^m r_i^2 - M\right)\left(\sum_{j=1}^n c_j^2 - M\right)}{2M^2} - \sum_{(i,j)\in\Omega^*} \frac{r_i c_j}{M} \right\}, \tag{11}$$

where $M = \sum_{i=1}^m r_i = \sum_{j=1}^n c_j$. This approximation works well for large sparse zero-one tables with bounded row and column sums and bounded total number of structural zeros in each row and column.

Corollary 1. *An asymptotic formula for the permanent of an* $n \times n$ *restriction matrix* A *is:*

$$\text{perm}(A) = |\Sigma_{\mathbf{11}}(\Omega)| \approx \Delta_{\mathbf{11}}(\Omega) \equiv n! \exp\left\{ -\frac{1}{n} \sum_{i=1}^n (n - C_i) \right\}, \tag{12}$$

where $C_1, \ldots, C_n$ *are the column sums of* A*, and* Ω *is defined in* (10).

Proof. Theorem 1 shows that perm(A) equals to the total number of $n \times n$ zero-one tables with row and column sums equal to 1 and a set of structural zeros Ω, defined in (10). Replacing $\mathbf{r}$ and $\mathbf{c}$ in (11) by $\mathbf{1} = (1, \ldots, 1)$ and noticing that the total number of structural zeros in Ω is $\sum_{i=1}^n (n - C_i)$, we immediately have the approximation (12).

4. Importance Sampling for Permutations with Constraints

4.1. Sampling strategies

Let $R_1, \ldots, R_n$ and $C_1, \ldots, C_n$ be the row and column sums of the constraint matrix A. To sample a table from $\Sigma_n(\Omega)$, we need to choose n cells $(t_{1,\sigma(1)}, \ldots, t_{n,\sigma(n)})$ to put 1's in, where $(\sigma(1), \ldots, \sigma(n))$ is a permutation of $(1, \ldots, n)$. We can start by sampling $\sigma(1)$ from all available positions in row 1, and, conditional on the realization of $\sigma(1)$, we continue to sample $\sigma(2)$, and so on. More precisely, we have to choose $\sigma(1) = j$ from one of the R_1 columns such that $(1, j) \notin \Omega$, i.e., $a_{1,j} = 1$. The true marginal distribution of $\sigma(1)$ is

$$p(\sigma(1) = j) \propto |\Sigma_{n-1}(\Omega_{-j})|, \qquad j \in \{k : (1, k) \notin \Omega\}, \tag{13}$$

where Ω_{-j} is the set of structural zeros in the restriction matrix A with the first row and the j-th column removed, and $\Sigma_{n-1}(\Omega_{-j})$ is the set of $(n-1) \times (n-1)$ zero-one tables with all marginal sums equal to 1 and a set of structural zeros Ω_{-j}.

The approximation (12) works well for zero-one tables with all marginal sums equal to 1, because such tables are sparse with bounded row and column sums. Applying (12) to (13), we have

$$\begin{aligned} p(\sigma(1)=j) &\propto |\Sigma_{n-1}(\Omega_{-j})| \\ &\approx (n-1)!\exp\Big\{-\frac{1}{n-1}\Big[\sum_{i=1}^{n}(n-C_i)-(n-C_j)\Big]\Big\} \\ &\propto \exp\Big\{\frac{n-C_j}{n-1}\Big\}, \end{aligned} \qquad (14)$$

for $j \in \{k : (1,k) \notin \Omega\}$. Thus this approximation suggests that we should sample $\sigma(1)$ according to the multinomial distribution on $\{j : (1,j) \notin \Omega\}$ with probabilities proportional to $\exp\{(n-C_j)/(n-1)\}$. After we sample a number j for $\sigma(1)$, we remove the first row and the j-th column of the table, and then sample $\sigma(2)$ from the remaining $(n-1)\times(n-1)$ subtable in the same way. Thus, our sampling distribution can be written as

$$q(\sigma=(\sigma(1),\ldots,\sigma(n))) = q(\sigma(1))q(\sigma(2)|\sigma(1))\cdots q(\sigma(n)|\sigma(1),\ldots,\sigma(n-1)),$$

which gives rise to the name *sequential importance sampling.* We refer to the sequential sampling strategy based on (14) as SIS-1. The following theorem, which is a special case of Theorem 1 in Chen (2007), provides the rationale for another SIS strategy.

Theorem 2. *Consider the uniform distribution over all* $n \times n$ *zero-one tables with given column sums* $1,\ldots,1$*, first row sum* 1*, and the set of structural zeros* Ω*. Let* C_j *be the number of non-structural zeros in the* j*-th column. Assume* $C_j > 1$ *for* $j \in \{k : (1,k) \notin \Omega\}$*, then*

$$p(\sigma(1)=j) \propto \frac{1}{C_j-1}, \quad j \in \{k :\ (1,k) \notin \Omega\}, \qquad (15)$$

where $\sigma(1)$ *is the column index of the nonzero element in the first row.*

Proof. We note that, because there are no constraints on the row sums, $p(\sigma(1)=j)$ is proportional to the number of ways to pick an arbitrary position not in Ω for each column (to put a "1" in) after the first row and the j-th column are removed. Therefore

$$p(\sigma(1)=j) \propto \frac{\prod_{i=1}^{n}(C_i - 1_{\{(1,i)\notin\Omega\}})}{C_j-1} \propto \frac{1}{C_j-1},$$

and the theorem is proved.

The simple result of Theorem 2 is derived by ignoring the constraints on the row sums for row 2 to row n. It suggests that we can sample $\sigma(1)$ according to the multinomial distribution on $\{j : (1, j) \notin \Omega\}$ with probabilities proportional to $1/(C_j - 1)$. If $C_j = 1$ for some $\{j : (1, j) \notin \Omega\}$, then $\sigma(1)$ should be set to j (if C_k also equals to 1 for some $k \neq j$, then there is no valid permutation sequences of the zero-one table, and we refer to it as an invalid table). After we sample a number j for $\sigma(1)$, we remove the first row and the j-th column from the table, and then sample $\sigma(2)$ from the remaining $(n-1) \times (n-1)$ subtable in the same way. We refer to this strategy as SIS-2.

Although it is not clear if the approximation based on (14) or (15) is accurate for a given table, the importance weights in the SIS procedure can correct for the biases and therefore can still perform well even when the approximation accuracy is inadequate. We observed that the importance sampling based on approximation (14) (SIS-1) tends to perform better than that based on approximation (15) (SIS-2) in most examples, especially when the marginal sums of the constraint matrix vary a lot. SIS-2 can sometimes give slightly better results than SIS-1 when the marginal sums of the constraint matrix do not vary much.

4.2. Further improvements and complications

If the rows of the constraint matrix A are re-ordered before implementing the sequential sampling, we can easily transform the sampled permutation back to the original order. We found that arranging the rows in an ascending order of the row sum R_i usually can improve the efficiency greatly. In fact, the best strategy we found is to re-arrange the rows at every step of sequential sampling so that the row to be sampled always has the smallest updated row sum. However, after arranging the rows in the ascending order of the row sum R_i at the beginning of the sampling, there are often very few re-arrangements needed in later steps. Such re-ordering usually leads to more valid tables and smaller variation of the importance weights.

One explanation of the efficiency improvement of the above ordering strategy is to look at the monotone case, i.e., the case that $S_{i_1} \subseteq \cdots \subseteq S_{i_n}$, where $S_{i_l} = \{j : a_{i_l j} = 1\}$, for $l = 1, \ldots, n$, and $i_1, \ldots, i_n$ is a permutation of $1, \ldots, n$. This implies, of course, that $R_{i_1} \leq \cdots \leq R_{i_n}$. Both SIS-1 and SIS-2 produce *perfect* samples from the uniform distribution of all allowable permutations provided that we first rearrange the rows in the ascending order of the row sum. This fact is easy to see since the conditional sampling distribution implied by both formulas (14) and (15) is uniform. Thus, all the importance weights are equal.

In the above SIS algorithms, we proposed to sample row by row to generate a table from $\Sigma_n(\Omega)$. It is also possible to sample column by column, and the columns can be sampled in an arbitrary order.

Neither SIS-1 nor SIS-2 can guarantee to always sample a valid table from $\Sigma_n(\Omega)$. Sometimes the sequential sampling cannot proceed after a few rows have been generated because no valid zero-one table can be produced. When this happens, we can simply throw away this partially generated table and start sampling a new one. This is equivalent to assigning a zero weight to this bad sample. If we want to always generate good tables, we need to check the existence of subtables with updated constraints, which is equivalent to checking the existence of a perfect matching for a bipartite graph (Diaconis et al. (2001)). Cook (1998, pp.134-143) gives several algorithms for finding perfect matchings in such graphs. However, these algorithms are too slow to implement. Sampling tables without checking the existence of subtables is usually more efficient, as long as the percentage of invalid tables is not too large.

4.3. Some existing procedures

Several importance sampling strategies have been proposed in the literature (Kuznetsov (1996), Beichl and Sullivan (1999) and Smith and Dawkins (2001)). The goal of these algorithms is to give an accurate estimate within a fixed running time. This is different from the algorithms developed by Jerrum and Sinclair (1996), Karmarkar et al. (1993), and Barvinok (1999), which are randomized polynomial time algorithms. In general it is hard to compare these two types of algorithms, and there is no proof that the algorithms based on importance sampling are randomized polynomial time algorithms. However, Smith and Dawkins (2001) compared their importance sampling method with Barvinok's (1999) and Karmarkar, et al.'s (1993) algorithms on many different types of matrices, and showed that the importance sampling method gives more precise estimates for a fixed amount of computing time. Another difference is that algorithms based on importance sampling can estimate both the permanent and the p-value for permutation tests simultaneously by using (8) and (7), whereas many randomized algorithms are designed to estimate only the permanent.

Kuznetsov (1996)'s algorithm is similar to SIS-2, but it chooses $\sigma(1)$ uniformly from all possible values $\{j : (1, j) \notin \Omega\}$, instead of using (15). Beichl and Sullivan (1999) use Sinkhorn balancing to derive the proposal distribution. Smith and Dawkins (2001) suggest several proposal distributions, and the most sophisticated and efficient one is equivalent to sampling $p(\sigma(1) = j) \propto 1/C_j$. This is only slightly different from (15), and its performance is similar to SIS-2 (based on (15)). Therefore in the next section, we only compare SIS-1, SIS-2, Kuznetsov's algorithm, and Beichl and Sullivan's algorithm.

5. Numerical Examples

In this section, we apply the SIS strategies developed in the previous section to approximate permanents and p-values of the Kendall's tau statistic for permutation tests with restricted positions. All algorithms were coded in C language except Beichl and Sullivan's algorithm, which was implemented in their Matlab program. Their code used a handy Matlab internal function and it is not obvious how to rewrite it in C. All algorithms were run on a Pentium 4 computer with 2.4 GHz processor.

5.1. Approximating permanents

Permanents have connections with many interesting problems, such as the perfect matching in a bipartite graph (Diaconis et al. (2001)), the dimer covering problem in physics (Beichl and Sullivan (1999)), and the performance bound of certain digital mobile radio systems in communications theory (Smith and Dawkins (2001)). Bapat (1990) and Diaconis et al. (2001) provide more examples in statistics and probability where permanents play an important role. Here we show that our methods can estimate the permanent of a zero-one matrix very accurately.

The first example is an $n \times n$ matrix with all diagonal elements equal to 0 and all non-diagonal elements equal to 1. A restriction matrix of this type is called a derangement because it represents permutations with $\sigma(i) \neq i$ for all $i = 1, \ldots, n$. The true value of the permanent for an $n \times n$ derangement matrix is known to be $n![1 - (1/2!) + (1/3!) - \cdots + (-1)^{n-1}(1/n!)]$. For $n = 100$, the true value is 3.43328×10^{157} (only the first six digits are given here). Table 2 summarizes the estimates based on 1,000 samples using different methods, where the number after ± is the standard error of the estimate. All 1,000 samples are valid tables for the four methods. SIS-1 gave the smallest cv^2, indicating that the proposal distribution is closest to the target distribution among the four methods. From the standard error and running time, we can see that SIS-1 is about four times faster (to produce the same standard error) than SIS-2, about 500,000 times faster than Kuznetsov's method. SIS-1 appears to be 300,000 times faster than Beichl and Sullivan's method, but this is not a fair comparison because they were written in different languages.

The second example is an $n \times n$ tri-diagonal zero-one matrix, i.e., $A = (a_{ij})_{n \times n}$ with $a_{ij} = 1$ if $|i - j| \leq 1$, and 0 otherwise. The theoretical value of perm(A) is the $(n + 1)$-th Fibonacci number F_{n+1} (Diaconis et al. (2001)). For

Table 2. Comparison of different methods on the 100 × 100 derangement matrix.

	Estimate	cv^2	Time
SIS-1	$(3.43329 \pm 0.00002) \times 10^{157}$	6.5×10^{-8}	2 sec
SIS-2	$(3.43332 \pm 0.00004) \times 10^{157}$	1.7×10^{-7}	2 sec
Kuznetsov	$(3.42 \pm 0.02) \times 10^{157}$	3.0×10^{-2}	1 sec
Beichl & Sullivan	$(3.43 \pm 0.0004) \times 10^{157}$	1.2×10^{-5}	25 min

$n = 100$, this value is 5.73148×10^{20} (only the first six digits are given here). Table 3 summarizes the estimates based on 1,000 samples using different methods. All 1,000 samples are valid tables for all methods. In this case, SIS-2 is a little better than SIS-1. Both SIS-1 and SIS-2 are more efficient than Kuznetsov's method and Beichl and Sullivan's algorithm (coded in Matlab).

Table 3. Comparison of different methods on the 100 × 100 matrix of Fibonacci permutations.

	Estimate	cv^2	Time
SIS-1	$(5.9 \pm 0.8) \times 10^{20}$	8.0	0.5 sec
SIS-2	$(5.6 \pm 0.2) \times 10^{20}$	1.1	0.5 sec
Kuznetsov	$(5.1 \pm 1.2) \times 10^{20}$	27.8	0.5 sec
Beichl & Sullivan	$(5.6 \pm 0.08) \times 10^{20}$	0.21	7.4 min

For Efron and Petrosian's (1999) quasar data, the restriction matrix is a 210×210 matrix. The permanent of this matrix is unknown. Another feature of the restriction matrix is that the row and column sums vary a lot. Kuznetsov's method was not able to generate any allowable permutations in 2,000 samples. Therefore we only compared SIS-1, SIS-2, and Beichl and Sullivan's method. Table 4 summarizes the estimates based on 1,000 samples, which took about the same amount of time for SIS-1 and SIS-2. However, the cv^2 of the tables from SIS-2 ranged from 100 to 1,000, compared to a cv^2 for SIS-1 of about 3. The standard errors indicate that SIS-1 is about 300 times more efficient than SIS-2 for this example. SIS-1 in C appears to be 250,000 times faster than Beichl and Sullivan's algorithm in Matlab. In this example, we arranged the columns of the constraint matrix in an ascending order of the column sums and sampled the permutations column by column.

To further challenge our method, we randomly generated a $1,000 \times 1,000$ table for which the probability for each cell to be 1 is 0.1. SIS-1 gave a cv^2 of 23.8; and SIS-2 gave a cv^2 of 1.4. This shows that both of our SIS algorithms are still very efficient in this case.

Table 4. Comparison of SIS-1 and SIS-2 on the restriction matrix of the quasar data.

	Estimate	cv^2	Time
SIS-1	$(1.53 \pm .08) \times 10^{280}$	3	5 sec
SIS-2	$(3.6 \pm 1.4) \times 10^{280}$	514	5 sec
Beichl & Sullivan	$(3.0 \pm .9) \times 10^{280}$	85.7	134 min

5.2. Permutation test on Efron and Petrosian's quasar data

Efron and Petrosian (1999) implemented a permutation test with Kendall's tau statistic (defined in (4)) for the independence between log-luminosity and redshift. To approximate the p-value of the statistic, they designed the following simple MCMC algorithm to generate allowable permutations. Starting with the identity permutation $(\sigma(1), \ldots, \sigma(n)) = (1, \ldots n)$, at each iteration step one randomly picks a pair of elements, $\sigma(i)$ and $\sigma(j)$, and proposes to transpose them. If the new permutation resulting from the transposition is an allowable one, i.e., $y_{\sigma(j)} \in [l(x_i), u(x_i)]$ and $y_{\sigma(i)} \in [l(x_j), u(x_j)]$, then the new permutation is accepted; otherwise, one stays put. Diaconis et al. (2001) showed that the proposed moves do result in an irreducible Markov chain with the uniform distribution as its invariant distribution. However, for truncations more complex than the interval ones, they also showed that more complicated Markov moves need to be designed in order to guarantee irreducibility.

The SIS strategy we developed is a more efficient way to estimate the p-value using (7). SIS-1 took about 10 seconds to generate 2,000 tables (permutations), based on which we estimated the p-value to be 0.016 ± 0.002. Among the 2,000 tables, only two tables were invalid. The MCMC algorithm took about 12 minutes to generate 1,100,000 samples (with 100,000 samples as burn-in period), and estimated a p-value of 0.015 ± 0.002. Thus, for this example SIS-1 appeared to be about 70 times more efficient than the MCMC algorithm.

5.3. Permutation test on Vidmar's data on juror verdicts

Vidmar (1972) did an experiment to study the effects of limiting the number of available alternatives on the verdicts reached by jurors. A murder trial was presented to students in an introductory social psychology course, they were asked to act as individual jurors. Then the students were divided into seven groups with 24 students each, and they were asked to give their verdicts under the set of decision alternatives available to their group. For example, students in the first group could choose between first-degree murder and not guilty; students in the second group could choose between second-degree murder and not guilty. Table 5 shows the number of times each alternative was chosen under different conditions. Based on the data he collected, Vidmar concluded "that under conditions of restricted decision alternatives, the more severe the degree of guilt associated with the least severe guilt alternative, the greater were the chances of obtaining a not guilty verdict." Later this conclusion was challenged by Larntz (1975).

From a statistical point of view, we may use the test of (quasi) independence to test whether the number and nature of the available alternatives has an effect on decision making. This was suggested in an exercise of Bishop, Fienberg and

Holland (1975, pp.205-206). Instead of using the standard chi-square goodness of fit test which relies on asymptotics, here we treat the data in Table 5 as 168 realizations of the pair (X, Y) and implement the permutation test.

Table 5. Decision alternatives and frequency of verdicts (Vidmar (1972); Columns are reordered). Here "–" means that decision alternative was not available to subjects in that condition.

	Condition (X)						
Alternative (Y)	1	2	3	4	5	6	7
First degree (Y=4)	11	–	2	–	–	7	2
Second degree (Y=3)	–	20	22	–	11	–	15
Manslaughter (Y=2)	–	–	–	22	13	16	5
Not guilty (Y=1)	13	4	0	2	0	1	2

We treat condition X and alternative Y as ordinal variables and use Kendall's tau statistic. The seven levels $1, 2, \ldots, 7$ for X are ordered according to the chances of obtaining a not guilty verdict based on Vidmar's conclusion. The four levels 1, 2, 3 and 4 for Y are ordered according to the severity of the alternative. For each condition X, only certain alternatives of Y are allowed due to the design of the experiment. For example, if $X = 2$, then Y can only be 1 and 3. Note that the permutation test for this example is a type of exact test for a two-way contingency table, because each allowed permutation leads to a two-way table with fixed marginal sums and given structural constraints (i.e., the entries with "–" are structural zeros).

The truncation pattern in Table 5 is more complicated than one-sided or two-sided truncations. Therefore, Efron and Petrosian's (1999) MCMC method cannot be applied directly here. Based on 2,000 sampled tables using SIS-1, which took about 6 seconds, we estimated that the p-value is 0.097 ± 0.007. The cv^2 is about 0.24. All 2,000 samples were valid tables.

Acknowledgement

This work was partially supported by the National Science Foundation grants DMS-0203762, DMS-0503981, and DMS-0244638. The authors thank Persi Diaconis for helpful discussions and suggesting two numerical examples in Section 5.1, and Francis Sullivan for sharing his Matlab code. We are very grateful to the two referees for helpful comments.

References

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

Barvinok, A. (1999). Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor. *Random Structures and Algorithms* **14**, 29-61,

Beichl, I. and Sullivan, F. (1999). Approximating the permanent via importance sampling with application to the dimer covering problem. *J. Comput. Phys.* **149**, 128-147.

Bender, E. A. (1974). The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Math.* **10**, 217-223.

Bilker, W. B. and Wang, M.-C. (1996). A semiparametric extension of the Mann-Whitney Test for Randomly Truncated Data. *Biometrics* **52**, 10-20.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* The MIT Press, Cambridge.

Chen, Y. (2007). Conditional inference on tables with structural zeros. *J. Comput. Graph. Statist.* **16**, 445-467.

Cook, W. J., Cunningham, W. H., Pulleyblank, W. R. and Schrijve, A. (1998). *Combinatorial Optimization.* Wiley, New York.

Diaconis, P., Graham, R. and Holmes S. (2001). Statistical problems involving permutations with restricted positions. In *State of the Art in Probability and Statistics* (Festschrift in Honor of William van Zwet), **36** of IMS Lecture Notes, Beachwood, OH, 195-222.

Efron, B. and Petrosian, V. (1999). Nonparametric methods for doubly truncated data. *J. Amer. Statist. Assoc.* **94**, 824-834.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, Boca Raton.

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

Jerrum, M. and Sinclair, A. (1996). The Markov chain Monte Carlo method: an approach to approximate counting and integration. In *Approximation Algorithms for NP-hard Problems* (Edited by D. S. Hochbaum). PWS Publishing, Boston.

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

Kong, A., Liu, J. S. and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89**, 278-288.

Kuznetsov, N. Y. (1996). Computing the permanent by importance sampling method. *Cybernet. Systems Anal.* **32**, 749-755.

Larntz, K. (1975). Reanalysis of Vidmar's data on the Effects of decision alternatives on verdicts of simulated jurors. *J. Personality and Social Psychology* **31**, 123-125.

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.

McLaren, C., Wagstaff, M., Brittegram, G. and Jacobs, A. (1991). Detection of two-component mixtures of lognormal distributions in grouped, doubly truncated data analysis of red blood cell volume distributions. *Biometrics* **47**, 607-622.

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

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

Smith, P. and Dawkins, B. (2001). Estimating the permanent by importance sampling from a finite population. *J. Statist. Comput. Simulation* **70**, 197-214.

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

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

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

Vidmar, N. (1972). Effects of decision alternatives on the verdicts and social perceptions of simulated jurors. *J. Personality and Social Psychology* **22**, 211-218.

Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 U.S.A.

E-mail: yuguo@uiuc.edu

Department of Statistics and Department of Biostatistics, Harvard University, Cambridge, MA 02138 U.S.A.

E-mail: jliu@stat.harvard.edu

(Received May 2005; accepted February 2006)