Approximating the Permanent: a Simple Approach

Lars Eilstrup Rasmussen

March 12, 1992

Abstract

In this paper, we present an efficient and very simple unbiased estimator for the Permanent of square (0-1) matrices. As the main result, we prove that our estimator, even though its worst case behaviour is provably very bad, runs in time polynomial in the size of the input matrix for almost all matrices. We also generalise our technique and apply it to another enumeration problem, that of counting Hamilton cycles in directed graphs. The ease with which this was done suggests that the basic technique may have wide applicability.

1 Introduction

1.1 The permanent

Let $A = \{a_{ij} : 1 \le i, j \le n\}$ be an $n \times n$ (0-1) matrix. The *permanent* of A, $\mathrm{per}(A)$, is defined by

$$\mathrm{per}(A) = \sum_{\pi} \prod_{i=1}^{n} a_{i,\pi(i)},$$

where the sum is over all permutations π of $[n] = \{1, \ldots, n\}$. The permanent has many combinatorial interpretations, one of which is related to bipartite graphs. For let $G = ([n], [n], E)$, where $(i, j) \in E$ iff $a_{ij} = 1$, be the bipartite graph which has A as its adjacency matrix. Then $\mathrm{per}(A)$ equals the number of perfect matchings in G.

The permanent function has been the object of study by mathematicians since first appearing in the work of Cauchy and Binet in 1812 (see [13] for background). Despite this effort and its syntactical similarity with the determinant, no efficient procedure for computing this function is known (the best known deterministic algorithm, which is due to Ryser [14], runs in time $\mathrm{O}(n2^n)$). In the late 1970's, Valiant [15] provided convincing evidence for the inherent intractability of the permanent, by demonstrating the problem to be complete for the class #P of enumeration problems. It is thus as hard to compute the permanent as it is, say, to count the number of satisfying assignments to a CNF formula (see [4] for an introduction to Complexity Theory). Interest has therefore turned to finding efficient approximation algorithms for the permanent.

1.2 Randomised approximation schemes

We shall formalise the notion of "computationally efficient approximation algorithm" as follows: let f be a function from some set Σ^* of inputs (strings) to natural numbers. A *randomised approximation scheme* [11] for f is a probabilistic algorithm which, when given an input $x \in \Sigma^*$ and a real number $0 < \epsilon < 1$, outputs a number X (a random variable) s.t.

$$\Pr\left((1-\epsilon) f(x) \le X \le (1+\epsilon) f(x)\right) \ge \frac{3}{4}.$$

The success probability may be boosted to $1 - \delta$ for any $0 < \delta < 1$ by running the algorithm $\mathrm{O}(\log \delta^{-1})$ times and taking the median of the results [9]. Such a scheme is said to be *fully polynomial* if its execution time is bounded by some polynomial in $|x|$ and ϵ^{-1}. We shall henceforth contract the phrase *fully polynomial randomised approximation scheme* to *fpras*.

Often (and throughout this report) the fpras is constructed from an *unbiased* *estimator*: Suppose we want to estimate f and have available a probabilistic algorithm whose output is a random variable X s.t. $\mathrm{E}(X) = f(x)$ and $\mathrm{E}(X^2)$ is finite. Suppose further that the algorithm runs in time polynomial in $|x|$, and that we can repeat it as many times as we wish, getting independent and identically distributed outputs. Then a straightforward application of Chebychev's inequality shows that running our algorithm $\mathrm{O}(\frac{\mathrm{E}(X^2)}{\mathrm{E}(X)^2}\epsilon^{-2})$ times and taking the mean of the results constitutes a randomised approximation scheme for f. From this it is clear that if the critical ratio $\frac{\mathrm{E}(X^2)}{\mathrm{E}(X)^2}$ is bounded by a polynomial in $|x|$, we have constructed an fpras for f.

The question of whether there exists an fpras for the permanent of a (0-1) matrix remains for the time being open. Given the apparent lack of progress on this front, it seems reasonable to weaken the requirements further, and ask whether there exists an fpras for $\mathrm{per}(A)$ that works for "almost all" inputs. This statement is made precise by the following definition: Let P be some predicate on (0-1) matrices. We say that P holds for *almost all* (0-1) matrices if the fraction of matrices for which the predicate is true tends to 1 as n tends to infinity. In probabilistic terms, this is stated as follows: Choose a matrix A u.a.r. from $\mathcal{A}(n)$, the set of all $n \times n$ (0-1) matrices. Then P holds for *almost all* (0-1) matrices if $\Pr(P(A)) \to 1$ as $n \to \infty$. This modified question has been answered affirmatively by Jerrum and Sinclair [8], and later by Jerrum [7].

In this paper, we describe an alternative estimator for the permanent. We prove that it is unbiased, and determine the worst case behaviour to be factorial in n, the size of the input matrix. We then prove the main result of this paper, that the estimator presented does indeed provide an fpras for almost all matrices. As described above, results of this kind has been obtained before by other researchers, but we present evidence that our approach compares favourably with known methods for approximating the permanent. The main benefits of our approach are its great simplicity and potential range of applicability. To illustrate the latter point, we conduct a suitable generalisation of the technique, and apply it to another popular enumeration problems, that of counting Hamilton cycles in directed graphs.

2 An alternative estimator

In this section we describe an alternative estimator for the permanent. We prove that it is unbiased, and determine the worst case behaviour to be factorial in n, the size of the input matrix. The section also contains the main result of this paper, that the estimator presented does indeed provide an fpras for almost all matrices.

2.1 The estimator

The estimator we present was originally inspired by Laplace's expansion formula for the permanent:

$$\mathrm{per}(A) = \sum_{j=1}^{n} a_{1j}\mathrm{per}(A_{1j}),$$

(where A_{1j} denotes the submatrix obtained from A by removing the 1st row and the jth column, and per(), the permanent of the empty matrix, is set to 1). This formula will be used extensively in the proofs presented. When suitably generalised, our idea becomes very similar to techniques used by Hammersley [6] and Knuth [12].

To each (0-1) $n \times n$ matrix $A = \{a_{ij} : 1 \leq i, j \leq n\}$ we assign a r.v. X_A as follows:

```
if n = 0 then
    X_A = 1
else
    W = {j : a_1j = 1}
    if W = ∅ then
        X_A = 0
    else
        Choose J u.a.r. from W
        X_A = |W| X_{A_1J}
```

Clearly this algorithm can easily be implemented to run in time $\mathrm{O}(n^2)$ (under the assumption that arithmetic operations can be carried out in constant time).

The random variable is easily proven to be an unbiased estimator of $\mathrm{per}(A)$:

Theorem 2.1 *Let $A = \{a_{ij} : 1 \leq i, j \leq n\}$ be an $n \times n$ (0-1) matrix, and let X_A be defined as above. Then*

$$\mathrm{E}(X_A) = \mathrm{per}(A).$$

Proof: Induction on n. The case $n = 0$ is trivial. Otherwise, if there are no 1's in the first row, $X_A = 0 = \mathrm{per}(A)$. For the final case, assume $W \neq \emptyset$, and set $m = |W|$. Then

$$\begin{aligned}
\mathrm{E}(X_A) &= \sum_{j \in W} \mathrm{E}(X_A \mid J = j)\Pr(J = j) \\
&= \sum_{j \in W} \mathrm{E}(mX_{A_{1j}} \mid J = j)m^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j \in W} \mathrm{E}(X_{A_{1j}}) \\
&= \sum_{j \in W} \mathrm{per}(A_{1j}) \\
&= \mathrm{per}(A). \quad \Box
\end{aligned}
$$

Equally easy to obtain is an upper bound on the performance of the estimator:

Theorem 2.2 *Let $A = \{a_{ij} : 1 \leq i, j \leq n\}$ be an $n \times n$ (0-1) matrix, and let X_A be defined as above. Then $\mathrm{E}(X_A^2) \leq \mathrm{per}(A)^2 n!$*

Proof: Induction on n. The case $n = 0$ is trivial. Otherwise, if there are no 1's in the first row, $X_A = 0 = \mathrm{per}(A)$. For the final case, assume $W \neq \emptyset$, and set $m = |W|$. Then

$$
\begin{aligned}
\mathrm{E}(X_A^2) &= \sum_{j \in W} \mathrm{E}(X_A^2 \mid J = j) \Pr(J = j) \\
&= \sum_{j \in W} \mathrm{E}(m^2 X_{A_{1j}}^2 \mid J = j) m^{-1} \\
&= \sum_{j \in W} \mathrm{E}(X_{A_{1j}}^2) m \\
&\leq \sum_{j \in W} \mathrm{per}(A_{1j})^2 (n-1)! m \\
&\leq \mathrm{per}(A)^2 n!
\end{aligned}
$$

where the last inequality follows from the fact that $m \leq n$ and $\sum_{j \in W} \mathrm{per}(A_{1j})^2 \leq (\sum_{j \in W} \mathrm{per}(A_{1j}))^2 = \mathrm{per}(A)^2$. $\Box$

This analysis may seem terribly pessimistic, but unfortunately it is not. Consider the "upper triangular unit matrix", U, of size n:

$$
U = \begin{pmatrix}
1 & 1 & \cdots & & 1 \\
& 1 & & 1 & \\
& & \ddots & & \vdots \\
& & & 1 & 1 \\
& & & & 1
\end{pmatrix}
$$

Clearly, $\mathrm{per}(U) = 1$, but $\mathrm{E}(X_U^2) = n!$. Another bad example is the "upper 2-band matrix", B, of size n:

$$
B = \begin{pmatrix}
1 & 1 & & & \\
& 1 & 1 & & \\
& & \ddots & \ddots & \\
& & & 1 & 1 \\
& & & & 1
\end{pmatrix}
$$

We here have that $\mathrm{per}(B) = 1$, but $\mathrm{E}(X_B^2) = 2^{n-1}$.

2.2 Random matrices

Though these examples show a devastatingly bad worst case performance of the estimator, one might expect a better result for a general (random) matrix. Following standard techniques (see [1] for an extensive treatment), we introduce, in addition to the space of coin-tosses related to the estimator, two new probability spaces of matrices, $\mathcal{A}(n)$, the set of all $n \times n$ (0-1) matrices, and $\mathcal{A}(n,m)$, the set of all $n \times n$ (0-1) matrices containing exactly m 1's. The probability measure on both spaces is the uniform one. We observe that the mean (as well as the second moment) of our estimator is merely a function from matrices to the real numbers, i.e., a random variable on $\mathcal{A}(n)$ and $\mathcal{A}(n,m)$ whose mean we can compute. Where confusion may arise from this, we write $\mathrm{E}_\sigma()$ for the mean over coin-tosses, $\mathrm{E}_{\mathcal{A}}()$ or $\mathrm{E}_{\mathcal{A}(n)}()$ for the mean over $\mathcal{A}(n)$, and $\mathrm{E}_{\mathcal{A}}()$ or $\mathrm{E}_{\mathcal{A}(n,m)}()$ for the mean over $\mathcal{A}(n,m)$. We also observe that per(), being a function from matrices to the reals, is a random variable whose mean we can compute:

Theorem 2.3 *Choose A_n from $\mathcal{A}(n)$. Then*

$$\mathrm{E}(\mathrm{per}(A_n)) = \frac{n!}{2^n}.$$

Proof: It is easily seen that the matrix A_n chosen from $\mathcal{A}(n)$ can be regarded as an $n \times n$ matrix of independent Bernoulli trials with parameter $p = \frac{1}{2}$. With this in mind, we get by Laplace's expansion formula and the linearity of expectation:

$$\begin{aligned}
\mathrm{E}(\mathrm{per}(A_n)) &= \sum_{j=1}^{n} \mathrm{E}(a_{1j}\mathrm{per}(A_n^{1j})) \\
&= \sum_{j=1}^{n} \mathrm{E}(a_{1j})\mathrm{E}(\mathrm{per}(A_n^{1j})) \\
&= \sum_{j=1}^{n} \frac{1}{2}\mathrm{E}(\mathrm{per}(A_{n-1})) \\
&= \frac{n!}{2^n},
\end{aligned}$$

Where A_{n-1} has been chosen from $\mathcal{A}(n-1)$, and the last equality follows by induction. $\Box$

As promised above, we can compute the expectation of $\mathrm{E}(X_{A_n})$ and $\mathrm{E}(X_{A_n}^2)$, when A_n is chosen from $\mathcal{A}(n)$.

Theorem 2.4 *Choose A_n from $\mathcal{A}(n)$, and let X_{A_n} be the corresponding estimator. Then*

$$\mathrm{E}_{\mathcal{A}}(\mathrm{E}_\sigma(X_{A_n})) = \prod_{i=1}^{n} \mathrm{E}(M_i),$$

and

$$\mathrm{E}_{\mathcal{A}}(\mathrm{E}_{\sigma}(X_{A_n}^2)) = \prod_{i=1}^{n} \mathrm{E}(M_i^2),$$

where each M_i is a binomial variable with parameters i and $p = \frac{1}{2}$.

Proof: The first equality follows directly from Theorems 2.1 and 2.3. For the second, we use induction on n. Again, the case $n = 0$ is trivial. In the case $n > 0$, note that $M = |W|$ is a binomial variable with parameters n and $p = \frac{1}{2}$ (recall that W is the set of column indices with a 1 in the first row). Now,

$$\begin{aligned}
\mathrm{E}_{\mathcal{A}}(\mathrm{E}_{\sigma}(X_{A_n}^2)) &= \sum_{m=0}^{n} \mathrm{E}_{\mathcal{A}}(\mathrm{E}_{\sigma}(X_{A_n}^2) \mid M = m) \Pr(M = m) \\
&= \sum_{m=0}^{n} \mathrm{E}_{\mathcal{A}} \left(m \sum_{j \in W} \mathrm{E}_{\sigma}(X_{A_n^{1j}}^2) \;\middle|\; M = m \right) \Pr(M = m) \\
&= \sum_{m=0}^{n} \mathrm{E}_{\mathcal{A}}(\mathrm{E}_{\sigma}(X_{A_{n-1}}^2)) m^2 \Pr(M = m) \\
&= \mathrm{E}_{\mathcal{A}}(\mathrm{E}_{\sigma}(X_{A_{n-1}}^2)) \mathrm{E}(M^2) \\
&= \prod_{i=1}^{n} \mathrm{E}(M_i^2),
\end{aligned}$$

where the last equality follows by induction. □

Unfortunately, we can not obtain the mean of the critical ratio easily from this, since the two random variables are clearly not independent. However, with a little more work, the main result (that the estimator runs in polynomial time for almost all matrices) can be established. The result rests heavily on the following technical lemma, the proof of which in turn rest on a result by Jerrum [7], which we state here without proof.

Lemma 2.5 *Suppose $m = m(n)$ satisfies $m^2 n^{-3} \to \infty$ as $n \to \infty$, and choose A from $\mathcal{A}(n, m)$. Then*

$$\mathrm{E}(\mathrm{per}(A)) = n! \left(\frac{m}{n^2}\right)^n \exp\left\{ -\frac{n^2}{2m} + \frac{1}{2} + \mathrm{O}(\frac{n^3}{m^2}) \right\},$$

and

$$\frac{\mathrm{E}(\mathrm{per}(A)^2)}{\mathrm{E}(\mathrm{per}(A))^2} = 1 + \mathrm{O}(\frac{n^3}{m^2}). \quad \square$$

Lemma 2.6 *Let $\omega = \omega(n)$ be any function satisfying $\omega \to \infty$ as $n \to \infty$. Then for almost all* (0-1) *matrices, A,*

$$\mathrm{per}(A) \geq \frac{\mu(n)}{\omega},$$

where $\mu(n) = \mathrm{E}_{\mathcal{A}(n)}(\mathrm{per}(A))$ denotes the mean of the permanent of a random matrix as computed in Theorem 2.3.

Proof: To utilise the result of Lemma 2.5, we take the following detour to choose a matrix from $\mathcal{A}(n)$:

(1) choose M from $b(n^2, p = \frac{1}{2})$, the binomial distribution with parameters n^2 and $\frac{1}{2}$.

(2) choose A from $\mathcal{A}(n, M)$.

To see that this is equivalent to choosing from $\mathcal{A}(n)$, let B be any $n \times n$ (0-1) matrix, and let m be the number of 1's in B. Then the probability of choosing B in the above experiment is

$$\begin{aligned}\Pr(M = m)\Pr(A = B \mid M = m) &= \binom{n^2}{m} 2^{-n^2} \binom{n^2}{m}^{-1} \\ &= 2^{-n^2},\end{aligned}$$

which equals the probability of choosing B from $\mathcal{A}(n)$.

Now, let $\omega' = \frac{\omega}{2e}$, and $\omega'' = \ln\sqrt{\omega'}$. Since $\omega'' \to \infty$ as $n \to \infty$, standard results from probability theory (e.g. Chernoff, see [1] p. 12-13) tells us that

$$\Pr(M < n^2/2 - \omega'' n) \overset{n\to\infty}{\to} 0$$

and Lemma 2.5 gives that

$$\Pr((\mathrm{per}(A) < \frac{1}{2}\mu(n, M)) \mid M = m) \overset{n\to\infty}{\to} 0,$$

(provided m not too small), where $\mu(n, m) = \mathrm{E}_{\mathcal{A}(n,m)}(\mathrm{per}(A))$ denotes the mean of the permanent of a random matrix with m 1's as computed in Lemma 2.5. It now suffices to prove that the conjunction of the negations of the above two events implies the wanted inequality (for sufficiently large n).

Thus, assuming that $M \geq n^2/2 - \omega'' n$ and that $\mathrm{per}(A) \geq \frac{1}{2}\mu(n, M))$, we get

$$\begin{aligned}\mu(n)^{-1}\mathrm{per}(A) &\geq \mu(n)^{-1}\frac{1}{2}\mu(n, m) \\ &\geq \mu(n)^{-1}\frac{1}{2}\mu(n, n^2/2 - \omega'' n) \\ &= \frac{1}{2}2^n\left(\frac{n^2/2 - \omega'' n}{n^2}\right)^n \exp\left\{-\frac{n^2}{n^2 - 2\omega'' n} + \frac{1}{2} + \mathrm{O}(\frac{n^3}{n^4})\right\} \\ &\geq \frac{1}{2}\left(1 - \frac{2\omega''}{n}\right)^n \exp\left\{-1 + \mathrm{O}(\frac{n^3}{n^4})\right\} \\ &\sim (2e\omega')^{-1} \\ &= \omega^{-1},\end{aligned}$$

where the penultimate step follows from the fact that

$$\omega' \left(1 - \frac{\ln(\omega')}{n}\right)^n \stackrel{n\to\infty}{\to} 1. \quad \square$$

Equipped with this lemma, we are now able to prove the main result:

Theorem 2.7 *Let $\omega = \omega(n)$ be any function satisfying $\omega \to \infty$ as $n \to \infty$. Then*

$$\frac{\mathrm{E}_\sigma(X_A^2)}{\mathrm{E}_\sigma(X_A)^2} \leq \mathrm{O}(n\omega),$$

for almost all (0-1) *matrices, A, where X_A is the estimator defined above.*

Proof: Set $\omega' = \sqrt[3]{\omega}$, and choose A from $\mathcal{A}(n)$. Then, by Markov's inequality,

$$\Pr\left(\mathrm{E}_\sigma(X_A^2) \geq \omega' \mathrm{E}_\mathcal{A}(\mathrm{E}_\sigma(X_A^2))\right) \leq \frac{1}{\omega'} \stackrel{n\to\infty}{\to} 0.$$

Combining this with Lemma 2.6 we get for almost all matrices A:

$$\begin{aligned}
\frac{\mathrm{E}_\sigma(X_A^2)}{\mathrm{E}_\sigma(X_A)^2} &\leq (\omega')^3 \frac{\mathrm{E}_\mathcal{A}(\mathrm{E}_\sigma(X_A^2))}{\mathrm{E}_\mathcal{A}(\mathrm{E}_\sigma(X_A))^2} \\
&= (\omega')^3 \prod_{i=1}^{n} \frac{i+i^2}{i^2} \\
&= (\omega')^3 (n+1) \\
&= \mathrm{O}(n\omega)
\end{aligned}$$

where the first equality follows from theorem 2.4. $\square$

To see that we can not obtain a similar result for sparse matrices, consider the probability space $\mathcal{A}(n,p)$ of $n \times n$ (0-1) matrices where each entry is chosen to be 1 with probability p. In this space, theorem 2.4 still holds (with the new parameter p as the parameter of the binomial variables). However, the ratio of the two expectations computed in that theorem is now:

$$\prod_{i=1}^{n} \left(1 + \frac{1-p}{p} i^{-1}\right)$$

which clearly tends to infinity as p tends to 0. Therefore, the proof of theorem 2.7 breaks down in the case of sparse matrices.

2.3 Known methods

As mentioned before, results similar to theorem 2.7 have been proved about other approximation schemes for the Permanent before this. An approach resting on the construction of suitable Markov Chains was suggested by Broder [2] in 1986. Jerrum and Sinclair [7] proved in 1989 that this approach *does* lead to an fpras for the permanent of all (0-1) $n \times n$ matrices with at least $n/2$ 1's in each row and column, and almost all other (0-1) matrices. They also devise an efficient method for determining whether a given matrix is amenable to approximation with this approach.

Such strong results has been proved about no other estimator for the Permanent. However, the degree of the polynomial bounding the runtime of the Jerrum/Sinclair estimator is not small, and it is not clear whether their approach could lead to a truly practical estimator for the permanent. For this reason alone, it is worth investigating alternative approaches.

More suitable for practical use is the $\mathrm{K}^2\mathrm{L}^3$ estimator proposed in 1988 by Karmarkar, Karp, Lipton, Lovász and Luby [10]. This estimator, which is an improvement of an estimator due to Godsil and Gutman [5], rests on the calculation of determinants, exactly one per run. In 1990, Jerrum [7] proved a theorem corresponding to theorem 2.7 for this estimator. The estimator thus needs a total of $\mathrm{O}(n^4)$ arithmetic operations for almost all matrices.

Each run of the estimator presented in this article can be performed using $\mathrm{O}(n^2)$ operations, giving a total of only $\mathrm{O}(n^3)$ operations for almost all matrices. Simply address the entries in the input matrix via index vectors to allow constant time generation of submatrices. Note that this implementation permutes the columns of the matrix during the course of the algorithm. However, this minor adjustment makes no difference to the proofs presented above.

3 Hamilton cycles

In this section, we first supply a generalisation of the above estimator, and apply the technique to construct an unbiased estimator for the number of Hamilton cycles in a directed graph. We then prove that this estimator yields an fpras for almost all directed graphs.

3.1 Counting the size of a set

Let S be any set s.t. $|S| < \infty$. Suppose that we have a randomised algorithm which outputs members of the set $S \cup \{\perp\}$, i.e., a (not necessarily uniform) random generator of members of S which may fail by outputing $\perp$. Suppose furthermore that

$$\forall s \in S. \Pr(\text{output is } s) > 0,$$

and that the algorithm can keep track of the probability associated with its output. We can then construct an unbiased estimator, X_S, of $|S|$ as follows:

$$X_S = \begin{cases} \Pr(\text{output is } s)^{-1} & \text{if output is } s \in S \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $\mathrm{E}(X_S) = |S|$. Also, the closer our generator is to uniform, the smaller the second moment. In particular, we wish to minimise $\Pr(\text{output is } \perp)$.

3.2 Counting Hamilton cycles

Let $G = ([n], E)$ be a directed graph (in the rest of this paper we shall use the term graph for directed graph where nothing else is mentioned), and let $\mathrm{H}(G)$ denote the number of Hamilton cycles in G. The naive approach to an estimator of the above type for $\mathrm{H}(G)$ is to start at some vertex and at random build a longer and longer Hamilton path until either there are no more reachable unused vertices, in which case the experiment has failed, or until a Hamilton cycle has been constructed. We shall, for reasons that will become clear later, choose a slightly different approach.

The idea is to fix an ordering of the vertices and then choose a successor vertex for each vertex in turn among the vertices that are still unused (i.e., have not yet been chosen as successor to some vertex). This approach is exactly equivalent to the algorithm presented in section 2. The generator may fail either because at some point there are no more reachable unused vertices, or because a premature cycle (a cycle of length less than n) has been constructed. It is, however, easy to remove the later source of failure, since for every choice of a successor vertex, at most one more edge can result in a premature cycle. This edge is then, as are all the unused edges from the current vertex and to its chosen successor, removed before choosing a successor for the next vertex. We present this algorithm in a recursive style, where we also remove a vertex at each step. Note that the potential premature cycle edge in this way always becomes a single loop when the subgraph is constructed for the next recursion. Again we assign to each graph $G = ([n], E)$ a r.v. X_G:

if $n = 1$ then

$$X_G = \begin{cases} 1 & \text{if } (1,1) \in E \\ 0 & \text{otherwise} \end{cases}$$

else

$W = \{j > 1 \ : \ (1, j) \in E\}$

if $W = \emptyset$ then

$X_G = 0$

else

Choose J u.a.r. from W

$X_G = |W| X_{G_{1J}}$

Here, the graph $G_{1J} = ([n-1], E')$ is constructed from G in the following way: simultaneously, all edges $e = (i, 1)$ ending in vertex 1 are changed to (i, J), and all edges $e = (i, J)$ ending in vertex J are changed to $(i, 1)$. Then vertex 1 and all edges incident to it are removed, and the remaining vertices are relabeled to $1, 2, \ldots, n-1$ respectively (and the edges with them). The obvious optimisation of this somewhat laborious construction is left out intentionally for the sake of the proof of theorem 3.1. As we shall see shortly, the construction can be carried out very easily when considering adjacency matrices.

With a bit of work we can now establish that the above estimator is of the kind described in section 3.1, with $\mathrm{H}(G)$ playing the role of the set S:

Theorem 3.1 *Let $G = ([n], E)$ be a graph, and let X_G be defined as above. Then $\mathrm{E}(X_G) = \mathrm{H}(G)$.*

Proof: We first prove, by induction on n, that the set of n edges chosen by the algorithm (if it does not return 0) is indeed a Hamilton cycle in G. Assume the estimator has not returned a zero. The case $n = 1$ is trivial. Otherwise, let $(1, J)$ be the first edge chosen, and let $e_1, \ldots, (i-1, J-1), (J-1, j-1), \ldots, e_{n-1}$ be the Hamilton cycle chosen by the estimator for the subgraph G_{1J}. Tracing the above subgraph construction backwards, we see that these edges were labelled $e'_1, \ldots, (i, 1), (J, j), \ldots, e'_{n-1}$ in G, where each e'_k has been obtained from e_k by adding 1 to each label. Clearly these edges combines with $(1, J)$ form a Hamilton cycle in G.
It is now easy to see that any Hamilton cycle in G can be chosen by the above algorithm, and that it indeed returns exactly the reciprocal of the probability associated with its output, or 0 if it fails by runing out of edges to choose from. This completes the proof. $\square$

In order to be able to reuse the computations from section 2, we set also this algorithm in the world of square (0-1) matrices. As it will be clear, the above algorithm is very similar to the one presented in that section. Let $A_G = \{a_{ij} : 1 \leq i, j \leq n\}$, where $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise, be the adjacency matrix of G. We also assign to each (0-1) $n \times n$ matrix A a adjacency graph G_A in the obvious way. Again we assign to each graph G a r.v. X_G as follows:

if $n = 1$ then
    $X_G = a_{11}$
else
    $W = \{j > 1 \ : \ a_{1j} = 1\}$
    if $W = \emptyset$ then
        $X_G = 0$
    else
        Choose J u.a.r. from W
        Permute column 1 and J
        $X_G = |W| X_{G_{A_{11}}}$

The reader is encouraged to work out that the two above algorithms are indeed equivalent.

We are now in a position to prove results equivalent to those in section 2 (typically, the only difference is a factor of n). Since the proofs are exactly as in section 2, we only state the main theorems here without proofs.

Theorem 3.2 *Let $G = ([n], E)$ be a graph, and let X_G be defined as above. Then* $\mathrm{E}(X_G^2) \leq \mathrm{H}(G)^2 (n-1)!$ $\square$

Again, there is a matrix, U':

$$U' = \begin{pmatrix} 0 & 1 & \cdots & & 1 \\ & 0 & 1 & 1 & \\ & & 0 & 1 & \vdots \\ & & & 0 & 1 \\ 1 & & & & 0 \end{pmatrix}$$

which yields equality, since clearly $\mathrm{H}(G_{U'}) = 1$, but $\mathrm{E}(X_{G_{U'}}) = (n-1)!$

Theorem 3.3 *Let $\omega = \omega(n)$ be any function satisfying $\omega \to \infty$ as $n \to \infty$. Then*

$$\frac{\mathrm{E}_\sigma(X_G^2)}{\mathrm{E}_\sigma(X_G)^2} \leq \mathrm{O}(n\omega),$$

for almost all graphs, G, where X_G is the estimator defined above. $\square$

This last theorem rests, as does theorem 2.7 through lemma 2.6, on a result regarding the behaviour of a random matrix chosen from the space $\mathcal{A}(n, m)$. This time, the result is borrowed from a recent paper by Frieze and Suen [3]:

Theorem 3.4 *Suppose $m = m(n)$ satisfies $m^2 n^{-3} \to \infty$ as $n \to \infty$, and choose A from $\mathcal{A}(n, m)$. Then*

$$\mathrm{E}(\mathrm{H}(G_A)) = (n-1)! \left(\frac{m}{n^2}\right)^n \exp\left\{-\frac{n^2}{2m} + \frac{1}{2} + \mathrm{O}(\frac{n^3}{m^2})\right\},$$

and

$$\frac{\mathrm{E}(\mathrm{H}(G_A)^2)}{\mathrm{E}(\mathrm{H}(G_A))^2} = 1 + \mathrm{O}(\frac{n^3}{m^2}). \quad \square$$

4 Conclusion

We have in this paper presented an unbiased estimator X_A for the permanent of a square (0-1) matrix A. We have determined that the critical ratio $\frac{\mathrm{E}(X_A^2)}{\mathrm{E}(X_A)^2}$ for that estimator is bounded above by $n!$, and that that bound is tight. More positively, we have demonstrated that the estimator yields a randomised approximation scheme for the permanent which is fully polynomial for almost all square (0-1) matrices. We have also demonstrated that our estimator compares favourably with known methods.

We then applied the same technique to another enumeration problem, that of counting the number of Hamilton cycles in directed graphs. Again, the technique results in a randomised approximation scheme which is fully polynomial for almost all directed graphs.

References

[1] Bollobás. *Random Graphs.* Academic Press, London, 1985.

[2] Broder. How hard is it to marry at random? (on the approximation of the permanent). In *Proceedings of the 18th ACM Symposium on Theory of Computing*, pages 50–58, 1986. Erratum in *Proceedings of the 18th ACM Symposium on Theory of Computing*, 1988, p.551.

[3] Frieze and Suen. Counting the number of hamilton cycles in random digraphs. To appear in Random Structures and Algorithms, 1992.

[4] Garey and Johnson. *Computers and Intractability.* W.H. Freeman and Company, 1979.

[5] Godsil and Gutman. On the matching polynomial of a graph. *Algebraic Methods in Graph Theory*, pages 241–249, 1981.

[6] Hammersley. Existence theorems and monte carlo methods for the monomer-dimer problem. *Research Papers in Statistics*, pages 125–146, 19??

[7] Jerrum. An analysis of a monte carlo algorithm for estimating the permanent. Technical report, LFCS, 1991.

[8] Jerrum and Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18:1149–1178, 1989.

[9] Jerrum, Valiant, and Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science*, pages 169–188, 1986.

[10] Karmarkar, Karp, Lipton, Lovász, and Luby. A monte-carlo algorithm for estimating the permanent. preprint, 1988.

[11] Karp and Luby. Monte-carlo algorithms for enumeration and reliability problems. In *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, pages 56–64, 1983.

[12] Knuth. Estimating the efficiency of bactrack programs. *Mathematics of Computations*, 29(129):121–136, Jan 1975.

[13] Minc. *Permanents.* Addison-Wesley, 1978.

[14] Ryser. *Combinatorial Mathematics.* The Carus Mathematical Monographs, No.14. The mathematical Association of America, 1963.

[15] Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.