Advances in Applied Mathematics 126 (2021) 101916

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

Permanental generating functions and sequential importance sampling

Fan Chung[a], Persi Diaconis[b], Ron Graham[c,*]

[a] Dept. of Mathematics, UCSD, United States of America
[b] Dept. of Mathematics, Stanford University, United States of America
[c] Dept. of Computer Science and Engineering, UCSD, United States of America

A R T I C L E   I N F O

Article history:
Received 17 December 2018
Received in revised form 15 May 2019
Accepted 18 May 2019
Available online 29 May 2019

Dedicated to Joseph Kung

MSC:
62D99

A B S T R A C T

We introduce techniques for deriving closed form generating functions for enumerating permutations with restricted positions keeping track of various statistics. The method involves evaluating permanents with variables as entries. These are applied to determine the sample size required for a novel sequential importance sampling algorithm for generating random perfect matchings in classes of bipartite graphs.

1. Introduction

This paper has two motivations. The first is a novel technique for deriving 'nice' generating functions for permutations with restricted positions via various statistics.

Example. Let $\mathcal{F}_{t,1}(n)$ be the set of permutations $\sigma \in S_n$ with $i - t \leq \sigma(i) \leq i + 1$. Thus when $t = 1$,

$$\mathcal{F}_{1,1}(n) = \{\sigma : |\sigma(i) - i| \leq 1\}.$$

* Corresponding author.
E-mail addresses: fan@ucsd.edu (F. Chung), graham@ucsd.edu (R. Graham).

https://doi.org/10.1016/j.aam.2019.05.004

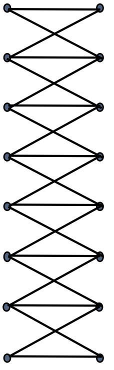

Fig. 1. Graph corresponding to $\mathcal{F}_{1,1}(8)$.

A *cycle index* for $\mathcal{F}_{t,1}(n)$ is

$$f_n = \sum_{\sigma \in \mathcal{F}_{t,1}(n)} \prod x_i^{a_i(\sigma)}$$

where σ has a_i i-cycles.

In Section 6, we show

Theorem 1.

$$\sum_{n=0}^{\infty} f_n z^n = \frac{1}{1 - x_1 z - x_2 z^2 - \ldots - x_t z^t}. \tag{1}$$

We are led to study such things via a novel importance sampling algorithm for generating random permutations with restricted positions. Suppose $B(n)$ is a bipartite graph with vertex sets $U_n = \{u_1, u_2, \ldots, u_n\}$ and $V_n = \{v_1, v_2, \ldots, v_n\}$ and various edges $\{u_i, v_j\}$ (Fig. 1).

Let $\mathcal{M}(n)$ be the set of perfect matchings in $B(n)$. Throughout we suppose that $\mathcal{M}(n)$ is nonempty. In a variety of statistical problems arising with truncated or censored data it is important to be able to study the distribution of various statistics of uniformly random elements of $\mathcal{M}(n)$. For example, as explained in [7], Lyndon Bells' test for correlation on truncated data leads to 'if you pick $\sigma \in \mathcal{M}(n)$ at random, what is the distribution of the number of involutions (i.e., 2-cycles) in σ?' A variety of techniques, reviewed in Section 2(A) are available to give approximations. This paper studies *sequential importance sampling*: Order U_n in some way, say $(u_1, u_2, \ldots, u_n)$. Consider u_1 having edges to various v_j. Some of these can be completed to a perfect matching. Call these J_1. Pick $\{u_1, v_j\}$ uniformly in J_1. Then delete $\{u_1, v_j\}$ and their incident edges. Proceed to u_2, choosing $\{u_2, v_j\}$ uniformly in $J_2, \ldots$. This always results in a perfect matching σ and the various available sets are (reasonably) efficiently computable (see the Wikipedia entry on 'matching (graph theory)'). The chance

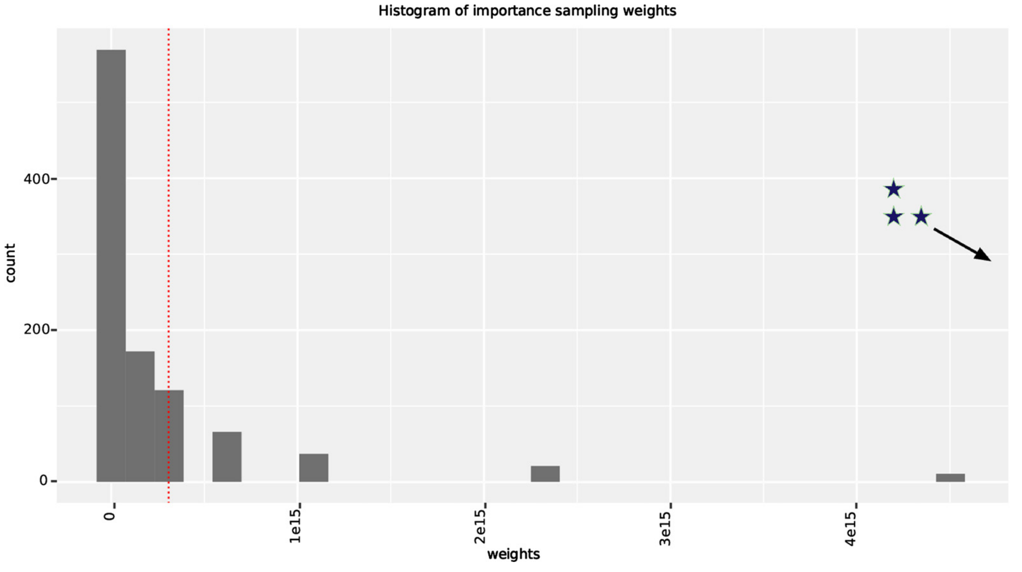

Fig. 2. Histogram of $T(\sigma_i)$ for $N = 1000$ samples for $\mathcal{F}_{70,1}$. The three starred values are 9.01×10^{15} (twice) and 3.60×10^{16}.

$$Pr(\sigma) = \prod_{i=1}^{n-1} |J_i|^{-1}, \tag{2}$$

is easy to compute. Let $T(\sigma) = \prod_{i=1}^{n-1} |J_i|$. Then

$$E(T) = \sum_{\sigma \in \mathcal{M}(n)} Pr(\sigma) T(\sigma) = |\mathcal{M}(n)|, \tag{3}$$

gives an unbiased estimate of $|\mathcal{M}(n)|$. If $Q(\sigma)$ is a statistic (e.g., the number of involutions or the number of fixed points in σ), $T(\sigma)$ allows estimating

$$P_u\{Q(\sigma) \le x\} = \frac{1}{N} \sum_{i=1}^{N} \delta(Q(\sigma_i) \le x) T(\sigma_i).$$

On the left P_u is the uniform distribution on $\mathcal{M}(n)$. On the right, σ_i is an independent sample from the sequential importance sampling algorithm.

As an example, the following plot shows a histogram of the importance weights $Pr(\sigma)^{-1}$ for a sample of size $N = 1000$ for estimating the number of $(1, 1)$ permutations when $n = 70$ (Fig. 2).

Here, the right answer is $F_{71} = 302, 061, 521, 170, 409$. The estimate is the sample mean of these weights, here $3.08 \ldots \times 10^{14}$, which is reasonably accurate. However, the estimated standard deviation from these data is $1.33 \ldots \times 10^{15}$ which is useless. Not surprisingly, the weights are all over the place; $min = 5.49 \ldots \times 10^{11}$, $max = 3.60 \ldots \times 10^{16}$.

Table 1
$\mathcal{F}_{4,1}$ with sequential sampling probabilities.

σ	1234	2134	1324	1243	2143
$Pr(\sigma)$	1/8	1/4	1/4	1/8	1/4

Importance sampling is very widely used and has resisted theoretical understanding. See [3] for an overview. The present paper gives a class of test cases permitting careful analysis.

Let us make the connection between generating functions and sequential importance sampling. We begin with a simple example. When $t = 1$, $|\mathcal{F}_{1,1}(n)| = F_{n+1}$, the Fibonacci number. For example, when $n = 4$, we have the following results, see Table 1.

The $Pr(\sigma)$ are shown in the second row of Table 1. For example, for sequential building up 1234: 1 can be placed in two places. If it is matched with 1, then there are two choices for 2. If it is matched with 2, then there are two choices for 3 and 4 is forced. This results in $Pr(1234) = 1/8$. For 2134, if 1 is matched with 2, then 2 is forced to be matched to 1. Then there are two choices for 3 and 4 is forced. This results in $Pr(2134) = 1/4$. Notice that always

$$Pr(\sigma) = \frac{1}{2^{k(\sigma)}} \tag{4}$$

where $k(\sigma)$ denotes the number of times that there were two possible choices for an edge in σ. This remains true for sequential importance sampling on $\mathcal{F}_{t,1}(n)$ (for the initial order $(1, 2, 3, \ldots, n)$). To study the variance and needed sample size of the estimator in (3) requires understanding

$$f_n(x) = \sum_{\sigma \in \mathcal{F}_{t,1}(n)} x^{k(\sigma)}. \tag{5}$$

Our techniques give

Theorem 2. *For enumerating $\mathcal{F}_{t,1}(n)$,*

$$\sum_{n=0}^{\infty} f_n(x) z^n = \frac{1 + z(1-x)(1 + xz + x^2z^2 + \ldots + x^{t-1}z^{t-1})}{1 - xz - x^2z^2 - \ldots - x^tz^t - x^tz^{t+1}}. \tag{6}$$

We are able to use this to give a sharp asymptotics for the variance of T and sample size required.

Section 2 below gives background on matchings and sequential importance sampling. Section 3 derives results for $\mathcal{F}_{t,1}(n)$. Section 4 derives results for $\{\sigma : |\sigma(i) - i| \leq 2\}$.

Section 5 studies two important variations of the basic algorithm. First, instead of 'working from the top', a variant beginning with 'working with every third' shows real improvement (much smaller sample size required). A most interesting second variant is to make the choices with non-uniform probabilities. In Section 5.4 this is shown to allow near perfect estimation with bounded sample size (as opposed to the exponential sample sizes required in all other variations.

Section 6 proves Theorem 1 for the cycle generating function.

Section 7 uses these generating functions to evaluate required sample sizes for sequential importance sampling. The result is surprising. For $\mathcal{F}_{t,1}(n)$, the sample sizes required are exponential but with tiny exponents. When $t = 1$:

$$N \asymp e^{(0.0204\ldots)n}.$$

For example, when $n = 200$, this indicates a sample of size about 60 is adequate.

Section 8 has remarks about using our permanental techniques for '(t, s) matchings'.

This may begin to explain why and when importance sampling works. It has its own introduction and may be consulted now for further motivation.

Acknowledgments

We thank Nima Anari, Sourav Chatterjee, Brett Kolesnik and Paulo Orenstein for their help, described below. We particularly thank Andy Tsao who found an error in our earlier version of Theorem 5 and his own proof of a correct result.

2. Background

This section gives background and history on: (A)—matchings, permanents and enumeration, (B)—importance sampling and (C)—a different determination of sample size.

(A) Matching theory. This classical subject has been treated wonderfully in the account of Lovász and Plummer [14]. It can be phrased as evaluating the permanent of an $n \times n$ (0/1) matrix M. This is

$$\mathrm{Per}(M) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} M_{\sigma(i),i}.$$

Thus, if M is a matrix with 0 on the main diagonal and is 1's elsewhere, $\mathrm{Per}(M)$ counts the number of derangements ($\sigma(i) \neq i$) [Monmort (1708)]. It can also be phrased as 'rook theory' (putting non-attacking rooks on a chessboard determined by M [18]). Applications in statistics are surveyed by Bapat [1] or Diaconis-Graham-Holmes [7].

Exact computation of the permanent is #P complete [20]. A variety of approximation schemes, some quite sophisticated, are developed in Chapter 3 of [2]. A celebrated achievement of Jerrum-Sinclair-Vigoda [12] gives a Markov chain Monte Carlo algorithm with a uniform stationary distribution and polynomial running time. Alas, the running time is order n^7 and as far as we know, this algorithm has never been used.

Of course for 'nice' restricted matrices, exact enumeration is possible. Examples include matrices arising as adjacency matrices of planar graphs [2] where Pfaffians and hence determinants enter. The matchings corresponding to $\mathcal{F}_{t,1}(n)$ above $(1 \leq t \leq \infty)$ allow exact enumeration as do $\{\sigma : \sigma(i) \geq b(i)\}$ (Ferrers boards) where $b(1) \leq b(2) \leq \ldots \leq b(n)$ is a fixed set of numbers (see [11]). We have been involved in other 'nice' cases [4]. Similarly, 'nice' classes of restriction matrices may allow the natural 'switch' Markov chain to mix in order $O^*(n^7)$ time. A wonderful paper of Dyer-Jerrum-Muller [9] does just this for classes of restriction matrices with bands of consecutive ones (of varying sizes) down the diagonal. Since these are typical in the censored data literature (and include our examples), this is real progress.

We have **not** seen much development around the natural theme: Fix a restriction matrix M. Pick σ consistent with M uniformly at random. What does σ 'look like'? How many cycles, inversions, fixed points, ... This has been some work for derangements, surveyed in [6]. Ozel [16] proves central limit theorems for the number of cycles of σ uniform in $\mathcal{F}_{\infty,1}(n)$.

Permanents with general entries are also of interest. The algorithms discussed here can be easily adapted. For example, instead of choosing $j \in J(1)$ uniformly, one can choose with probability proportional to the absolute value of the $(1, j)$ entry. See [2,21] for results and an overview.

(B) Importance sampling. Importance sampling is a very widely used simulation technique. Briefly, there is a space $\mathcal{X}$ with a probability measure ν specified. For a real-valued function f on $\mathcal{X}$ with finite mean

$$I(f) = \int f(x)\nu(dx)$$

one wants to estimate $I(f)$. In applications, this is intractable. But there is an auxiliary probability measure μ that is 'easy to sample from'. If μ is positive whenever ν is positive $(\nu << \mu)$ with

$$\rho = \frac{d\nu}{d\mu},$$

then $I(f) = \int f(x)\rho(x)\mu(dx)$. So we may sample $x_1, x_2, \ldots, x_N$ from μ and estimate $I(f)$ by

$$\hat{I}(f) = \frac{1}{N}\sum_{i=1}^{N} f(x_i)\rho(x_i). \tag{7}$$

The huge variety of applications and variations are surveyed in [3] and [13].

In our examples $\mathcal{X} = M$, the set of perfect matchings in a bipartite graph, ν is the uniform distribution on M and μ is the sequential importance sampling measure. Then $\rho(x) = \prod_{i=1}^{n} |J_i|/|M|$.

As explained in [3], the variance of $\hat{I}(f)$ can be an unreliable measure of accuracy for the long-tailed distributions that occur in importance sampling. The present setup gives a class of cases where this can be quantified.

Example *(Fibonacci permutations $\mathcal{F}_{1,1}(n)$).* Consider the sequential importance sampling algorithm for $\mathcal{F}_{1,1}(n)$. Let $T(\sigma) = 1/Pr(\sigma)$. So,

$$\mathbb{E}(T) = |\mathcal{F}_{1,1}(n)| = F_{n+1}.$$

$Var(T) = \mathbb{E}(T^2) - \mathbb{E}(T)^2$. We compute

$$\mathbb{E}(T^2) = \sum_{\sigma \in \mathcal{F}_{1,1}(n)} 2^{k(\sigma)} = f_n(2)$$

for f_n defined in (5).

Using (6) for $t = 1$, we have

$$\sum_{n=0}^{\infty} f_n(2) z^n = \frac{1-z}{1-2(z+z^2)}. \tag{8}$$

The denominator $1 - 2z - 2z^2 = (1 - zr_1^{-1})(1 - zr_2^{-1})$, for $r_1 = (-1+\sqrt{3})/2 = .3660\ldots$ and $r_2 = (-1-\sqrt{3})/2 = -1.3660\ldots$. Routine analysis shows

$$f_n(2) = (1/2)(r_1^{-n} + r_2^{-n}).$$

Here, r_1 is the root of minimum modulus so the dominant term is the one with r_1^n in the denominator. Since $\mathbb{E}(T) = |\mathcal{F}_{1,1}(n)| = F_{n+1}$ (the $(n+1)^{st}$ Fibonacci number) and $\phi := (1/2)(1+\sqrt{5})$, then

$$\mathbb{E}(T)^2 = \frac{1}{5}\phi^{2(n+1)} + o(1)$$

and

$$\mathrm{Var}(T) \sim (1/2)(1+\sqrt{3})^n. \tag{9}$$

Putting in numerical values, we find

$$\mathbb{E}(T) \sim \frac{1}{\sqrt{5}}\phi^{n+1} = (0.7236\ldots)(1.618\ldots)^n, \quad S.D.(T) \sim (.7071\ldots)(1.6528\ldots)^n.$$

Thus, the variability is large compared to the mean. This suggests large sample sizes are required to get an accurate estimate. For a sample of size N,

$$\frac{S.D.(T)}{\sqrt{N}} \ll \mathbb{E}(T) \iff N \gg \frac{\mathrm{Var}(T)}{\mathbb{E}(T)^2}.$$

Using our approximations,

$$N \gg \frac{5}{2\phi^2}\left(\frac{1+\sqrt{3}}{\phi^2}\right)^n \sim (.9549\ldots)(1.0435\ldots)^n \sim (.9549\ldots)e^{(0.0426\ldots)n}.$$

Thus, while the required sample size (to make the standard deviation of $(\hat{I}_n)$ small compared to I_n) is exponential in n, the constant is small. For $n = 200$, this suggests a sample size of at least $N = 4,788$ is needed. The following considerations show this is an over-estimate. Many further explicit examples are given in Section 7. The main point for now is that this is the first example where such calculations can be pushed through. They make full use of our explicit generating functions.

(C) A different determination of sample size. As explained above, the variance can be a poor measure of accuracy for such long-tailed distributions. In [3] (see also [19]), a theory is developed for the sample size required to have $|\mathbb{E}(\hat{I}_n) - \mathbb{E}(I_n)| < \epsilon$. It gives necessary and sufficient conditions. To state the result, define the Kullback-Liebler divergence by

$$L = D(\nu|\mu) := \int \rho \ \ln \rho d\mu = \int \ln \rho \ d\nu = \mathbb{E}_\nu(\ln \rho Y), \tag{10}$$

where Y has probability distribution ν. The main result shows that "$N = e^L$ steps are necessary and sufficient for accuracy".

Theorem 3.

(a) *If* $\|f\|_{2,\nu} < \infty$ *and* $N = e^{L+t}$ *for* $t > 0$ *then*

$$\mathbb{E}|\hat{I}_N(f) - I(f)| \le \|f\|_{2,\nu}[e^{-t/4} + 2P_\nu^{1/2}(\ln \rho(Y) > L + t/2)].$$

(b) *Conversely, if* $f \equiv 1$ *and* $N = e^{L-t}$, $t > 0$, *then for any* $\delta > 0$ *we have*

$$P_\nu\{\hat{I}_N(f) > (1-\delta)\} \le e^{-t/2} + P_\nu(\ln \rho(Y) \le L - t/2)/(1-\delta).$$

Remarks. To help parse this, suppose that $\|f\|_{2,\nu} \le 1$, e.g., f is the indicator function of a set. Part (a) says that if $N > e^{L+t}$ and $\ln(\rho(Y))$ is concentrated around its mean ($\mathbb{E}_\nu(\ln(\rho(Y)) = L$) then $\hat{I}_N(f)$ is close to $I(f)$ with high probability (use Markov's inequality with (a)).

Conversely, part (b) shows if $N < e^{L-t}$ and $\ln(\rho(Y))$ is concentrated about its mean then $I(1) = 1$ but there is only a small probability that $\hat{I}_N(1)$ is correct.

In the case of perfect matchings, ν is the uniform distribution, μ is the sequential distribution $P(\sigma)$, ρ is $\frac{Pr^{-1}(\sigma)}{|M|}$ and

$$L = \frac{1}{|M|}\sum_{\sigma} \ln(\rho(\sigma)) = -\mathbb{E}_{\nu} \ln Pr(\sigma) - \ln |M|.$$

The main results of this paper give sharp estimates of L and a method of proving $\ln(\rho(Y))$ is concentrated about its mean for a class of problems. As a numerical example, when $t = 1$ and $n = 200$, the use of $L + 1$ s.d. suggests that a sample size of at least $4,058$ is sufficient (compared with the $4,788$ using the variance criterion above).

In [5], Bregman's inequality is used to prove a result for general graphs. It is shown that if the set $\{u_1, u_2, \ldots, u_n\}$ is randomly ordered then

$$N_{Breg} = e^L \le \frac{1}{|\mathcal{X}|}\prod_{i=1}^{n}(d_i!)^{1/d_i} \tag{11}$$

where d_i denotes the degree of u_i. That paper was unable to prove concentration but making the reasonable assumption of concentration of $\ln(\rho(Y))$, it is of interest to compare the bound with the right answer (for our special class). As a numerical example, when $t = 1$ and $n = 200$, the Bregman bound gives N_{Breg} at least 1.004×10^{10} (!). Thus, while elegant and general, it is useless for this example.

3. *t*-Fibonacci graphs

3.1. Introduction

This section enumerates matchings and the relevant probabilities for sequential importance sampling for bipartite graphs $B_{t,1}(n)$ that we call $(t, 1)$-graphs. These are bipartite graphs with vertex sets $U_n = \{u_1, u_2, \ldots, u_n\}$ and $V_n = (v_1, v_2, \ldots, v_n)$ and having as edges all pairs $\{u_i, v_j\}$ with $-1 \le i - j \le t$ (when the indices are well-defined).

The perfect matchings in this graph are in bijection with t-Fibonacci sequences satisfying

$$F^{(t)}_{n+1} = F^{(t)}_n + F^{(t)}_{n-1} + \ldots + F^{(t)}_{n-t}, \quad F^{(t)}_1 = 1,\ F^{(t)}_j = 0,\ j \le 0.$$

Thus, $F^{(1)}_n$ starts $(0, 1, 1, 2, 3, 5, \ldots)$, the usual Fibonacci sequence (see OEIS #A000045), $F^{(2)}_n$ starts $(0, 1, 1, 2, 4, 7, \ldots)$ (a translation of OEIS #A000073), and so on.

These sequences have a fair-sized enumerative literature (see the Wikipedia entry for generalized Fibonacci sequences). We have not found previous study of the associated matchings.

As usual, for sequential importance sampling, we form random matchings by starting with u_1 and choosing a random edge incident to it (there are two choices) and thereafter, proceeding in order $u_2, u_3, u_4, \ldots$ always randomly selecting an incident edge, as long as after selecting that edge (and all the previous ones), we can still complete these choices to a perfect matching.

Fact 1. The only time there is a unique choice is if either, for some vertex u_k, no edge has yet been assigned to v_{k-t}, or the vertex is u_n. In the first case we must put in the edge $\{u_k, v_{k-t}\}$. In the second case, u_n must be matched with the unique vertex v_k which has no edge yet assigned to it. In all other cases, there are always exactly two choices.

Proof. By induction on k.

Remark. Fact 1 implies that the permutations arising from the t-Fibonacci graphs have a simple structure. The only cycles they have are of the form $(i, i+1, i+2, \ldots, i+k)$ for some k between 0 and $t+1$. Thus, all the results of this section can be interpreted as the enumerative theory of such cycles.

A generalization of Fact 1 showing that this 'from the top greedy algorithm' never gets stuck holds for (t, s) permutations.

Before we dig into details, there is one further example which provides a useful limiting case.

Example *($t = \infty$).* Here the bipartite graph has edges $\{u_i, v_j\}$ for $-1 \leq i - j$. The associated perfect matchings are $\{\sigma : \sigma(i) \leq i+1\}$. By an easy induction, there are 2^{n-1} such matchings.

Fact 2. The usual algorithm for generating a random $t = \infty$ matching, starting in order $u_1, u_2, u_3, \ldots$ generates an exactly uniform perfect matching.

Proof. Induction on n.

It follows that sequential importance matching is exact with $t = \infty$: $P(\sigma) = \frac{1}{2^{n-1}}$ for all σ and the variance is zero. This suggests that sequential importance sampling should be good for 'large t'. The results below show 'large' is $t = \ln_2 n + c \ln_2 \ln_2 n$. More generally, the variances and estimated sample sizes are decreasing in t.

3.2. The matrix $M_n^{(t)}(x)$

We define

$$M_n^{(t)}(x)[i,j] = \begin{cases} x & \text{if } -1 \leq j - i \leq t-1, \\ 1 & \text{if } i - j = t, \text{ or } i = n,\ max(1, n-t) \leq j \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

For example,

$$M_8^{(2)}(x) = \begin{bmatrix} x & x & 0 & 0 & 0 & 0 & 0 & 0 \\ x & x & x & 0 & 0 & 0 & 0 & 0 \\ 1 & x & x & x & 0 & 0 & 0 & 0 \\ 0 & 1 & x & x & x & 0 & 0 & 0 \\ 0 & 0 & 1 & x & x & x & 0 & 0 \\ 0 & 0 & 0 & 1 & x & x & x & 0 \\ 0 & 0 & 0 & 0 & 1 & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Let $a(n) = a^{(t)}(n)$ denote the permanent $Per(M_n^{(t)}(x))$. Thus, for $t = 2$ we have

$$a(0) = a(1) = 1, a(2) = 2x, a(3) = 4x^2, a(4) = 6x^3 + x^2, a(5) = 10z^4 + 3x^3, \dots.$$

Let us write

$$a(n) = \sum_{0 \le k \le n} f_n(k) x^k,$$
$$b(n) = a(n)\big|_{x=1}$$

where $f_n(k) = f_n^{(t)}(k)$ is just the number of perfect matchings in which there were k random choices each with probability $1/2$ (i.e., there were $n - k$ vertices which had only one choice).

This is immediate from the definition of the permanent.

We will often suppress the exponent t when it is understood.

Fact 3. $a(n)$ satisfies the following recurrence:

$$a(n) = xa(n-1) + x^2 a(n-2) + x^3 a(n-3) + \dots$$
$$+ x^t a(n-t) + x^t a(n-t-1),\ n \ge t+2,$$

with $a(m) = 0,\ m < 0,\ a(0) = a(1) = 1$ and $a(j) = 2^{j-1} x^{j-1}$ for $2 \le j \le t+1$.

Proof. For $2 \le j \le t+1$, the first two columns are identical and therefore $a(j) = 2xa(j-1)$. For $j \ge t+2$, we use induction on n, recursively expanding the permanent by the top row. The fact that the last power of x is only t and not $t+1$ comes from the fact that the last permanent expansion has a 1 in the top corner (and not an x).

We consider the generating function $G^{(t)}(x, z)$ defined by

$$G^{(t)}(x, z) = \sum_{n \ge 0} a(n) z^n.$$

From now on, we consider a fixed t and denote $G^{(t)}(x, z) = G(x, z)$ if there is no confusion.

Fact 4. The generating function $G^{(t)}(x,z) = G(x,z)$ is given by

$$G(x,z) = \sum_{n\geq 0} a(n)z^n = \frac{1 + z(1-x)\big(1 + xz + x^2z^2 + \ldots + x^{t-1}z^{t-1}\big)}{1 - xz - x^2z^2 - x^3z^3 - \ldots - x^tz^t - x^tz^{t+1}}.$$

Proof. We use the recurrence for $a(n)$ in Fact 3. By multiplying $a(n)$ by z^n and summing over all $n \geq 2$, we have

$$\begin{aligned} G(x,z) - 1 - \sum_{j=1}^{t+1} 2^{j-1}x^{j-1}z^j &= \sum_{n\geq t+2} a(n)z^n \\ &= \sum_{n\geq t+2} \Big(xa(n-1) + x^2a(n-2) + x^3a(n-3) + \ldots \\ &\quad + x^ta(n-t) + x^ta(n-t-1)\Big)z^n \\ &= G(x,z)\big(xz + x^2z^2 + x^3z^3 + \ldots + x^tz^t + x^tz^{t+1}\big) \\ &\quad - \sum_{j=1}^{t+1}(2^{j-1}-1)x^{j-1}z^j - \sum_{j=1}^{t} x^jz^j - x^tz^{t+1}. \end{aligned}$$

By collecting terms, we have

$$\begin{aligned} &(1 - xz - x^2z^2 - \ldots - x^tz^t - x^tz^{t+1})G(x,z) \\ &= 1 + z(1-x)\big(1 + xz + \ldots + x^{t-1}z^{t-1}\big) \end{aligned}$$

and Fact 4 is proved. □

Let $G(1,z) = \sum_{n\geq 0} b(n)z^n = \sum_{n\geq 0}\sum_{0\leq k\leq n} f_n(k)z^n$.
Thus, $\sum_{0\leq k\leq n} f_n(k) = Per(M_n(1))$ is the total number of perfect matchings in $B_t(n)$.

Fact 5. The generating function $F(z)$ for the $b(n) = \sum_k f_n(k)$ is

$$F(z) = \frac{1}{1 - z - z^2 - z^3 - \ldots - z^t - z^{t+1}}.$$

Proof. Plug in $x = 1$ in Fact 4.

3.3. Analyzing $F(z)$

In order to estimate the number $b(n)$ of perfect matchings in $B_{t,1}(n)$, we use the generating function $F(z)$ in Fact 5. Of particular interest is the root of minimum modulus in the denominator of F. (See [10] for more details.)

Fact 6. For $t \geq 2$, the polynomial $P(z) = 1 - z - \ldots - z^{t+1}$ has a unique real root ρ of minimum modulus which lies in the range given below:

$$\frac{1}{2} + \frac{1}{2^{t+3}} < \rho < \frac{1}{2} + \frac{1}{2^{t+3}} + \frac{t}{2^{2t+3}}.$$

Wolfram [22] (Cor. 3.4) gives a proof that all the zeros of $P(z)$ are simple and references to earlier work. Our inequality for ρ sharpens his result which gave

$$\rho < \frac{1}{2} + \frac{1}{2^{t+3}} + \frac{t}{2^{2t+3}} < \frac{1}{2} + \frac{1}{2^{t+1}}.$$

Proof. We note that

$$P(z) - zP(z) = 1 - 2z + z^{t+2}.$$

Since $P(z)$ does not have 1 as a root, all roots of $P(z)$ are roots of $Q(z) = 1 - 2z + z^{t+2}$ and all roots of $Q(z)$ except for 1 are roots of $P(z)$. To derive the range for ρ, it simply suffices to check that

$$\begin{aligned} Q(z) > 0, \text{ for } \ z &< \frac{1}{2} + \frac{1}{2^{t+3}} \\ Q(z) < 0, \text{ for } \ z &= \frac{1}{2} + \frac{1}{2^{t+3}} + \frac{t}{2^{2t+3}}. \end{aligned}$$

We are now ready to estimate $b(n) = \sum_k f_n^{(t)}(k)$.

Theorem 4. *The number $b(n)$ of perfect matchings in $B_{t,1}(n)$ satisfies*

$$b(n) \sim c_1 \rho^{-n}$$

where ρ is the unique real root of $P(z) = 1 - z - \ldots - z^{t+1}$ in $(0, 1)$ (see Fact 6), and c_1 satisfies

$$c_1 = \frac{1}{\rho\big((t+1)\rho^t + \ldots + 3\rho^2 + 2\rho + 1\big)}.$$

Numerical values and asymptotic approximations for ρ and c_1 are given in Table 2 and (21).

Proof. We define two useful functions:

$$\begin{aligned} S_j(w) &= 1 + w + w^2 + \ldots + w^j, \\ Q_j(w) &= 1 + 2w + 3w^2 + \ldots + (j+1)w^j. \end{aligned}$$

Table 2
Table of small values of $\rho^{(t)}$ and $c_i^{(t)}$.

t	$\rho^{(t)}$	$c_1^{(t)}$	$c_2^{(t)}$	$c_3^{(t)}$	$c_4^{(t)}$	$c_5^{(t)}$	$c_7^{(t)}$
1	.61803 . . .	.72360 . . .	.52360 . . .	−.88137 . . .	0.7577 . . .	−2.1133 . . .	.08944 . . .
2	.54369 . . .	.61841 . . .	.55695 . . .	−1.0121 . . .	1.0032 . . .	−2.8449 . . .	.05950 . . .
3	.51879 . . .	.56634 . . .	.54310 . . .	−1.0282 . . .	1.0416 . . .	−3.0183 . . .	.03138 . . .
4	.50866 . . .	.53792 . . .	.52807 . . .	−1.0230 . . .	1.0367 . . .	−3.0466 . . .	.01580 . . .
5	.50413 . . .	.52177 . . .	.51730 . . .	−1.0156 . . .	1.0257 . . .	−3.0398 . . .	.00788 . . .
10	.50012 . . .	.50122 . . .	.50110 . . .	−1.0010 . . .	1.0019 . . .	−3.0036 . . .	.00024 . . .

It can be easily verified that $P(z)$ satisfies

$$P(z) = 1 - z - \ldots - z^{t+1} = (\rho - z)R_t(z)$$

where

$$\begin{aligned} R_t(z) &= z^t + (\rho+1)z^{t-1} + \ldots + (\rho^{t-1} + \ldots + 1)z + \rho^t + \ldots + 1 \\ &= \sum_{j=0}^{t} S_j(\rho) z^{t-j}. \end{aligned}$$

Note that $R_t(\rho) = Q_t(\rho)$.

We consider the following partial fraction decomposition of the generating function $F(z)$ of $f(n)$ which can be directly verified using the fact that $S_t(\rho) + \rho Q_{t-1}(\rho) = Q_t(\rho)$ and $R_t(z) + (\rho - z)\sum_{j=0}^{t-1} Q_j(\rho) z^{t-j-1} = Q_t(\rho)$.

$$F(z) = \frac{1}{P(z)} = \frac{1}{(\rho - z)R_t(z)} = \frac{\alpha}{\rho - z} + \frac{\beta(z)}{R_t(z)} \tag{12}$$

where

$$\begin{aligned} \alpha &= \frac{1}{1 + 2\rho + \ldots + (t+1)\rho^t} = \frac{1}{Q_t(\rho)} \\ \beta(z) &= \frac{z^{t-1} + (1+2\rho)z^{t-2} + \ldots + (1 + 2\rho + \ldots + t\rho^{t-1})}{Q_t(\rho)} = \frac{\sum_{j=0}^{t-1} Q_j(\rho) z^{t-j-1}}{Q_t(\rho)}. \end{aligned}$$

Therefore we have

$$c_1 = \frac{\alpha}{\rho} = \frac{1}{\rho Q_t(\rho)} \tag{13}$$

and

$$b(n) \sim \frac{c_1}{\rho^n}.$$

Theorem 4 is proved. □

3.4. The first moment

Our next goal is to estimate the first moment for a fixed t:

$$H^{(t)}(z) = \sum_{n\geq 0}\sum_{0<k\leq n} k f_n^{(t)}(k) z^n.$$

To obtain the generating function $H^{(t)}(z)$, we simply differentiate $G^{(t)}(x,z)$ with respect to x and then set $x = 1$. For example, for the case of $t = 3$, we have

$$H^{(3)}(z) = \frac{z^2(z^5 + 2z^4 + 3z^3 + 6z^2 + 4z + 2)}{(z^4 + z^3 + z^2 + z - 1)^2}.$$

In the remainder of this section, we consider a fixed t and we abbreviate $H(z) = H^{(t)}(z)$ and suppress (t) in various expressions if there is no confusion.

We note that

$$\begin{aligned}\frac{\partial}{\partial x}G(x,z) &= \frac{-\sum_{k=0}^{t-1}\left(ix^{i-1} - (i+1)x^i\right)z^{i+1}}{1 - xz - x^2z^2 - x^3z^3 - \ldots - x^tz^t - x^tz^{t+1}} \\ &\quad + \frac{(1 + z - xz\sum_{i=0}^{t-1} x^iz^i)(z + 2xz^2 + \ldots + tx^{t-1}z^t + tx^{t-1}z^{t+1})}{\left(1 - xz - x^2z^2 - x^3z^3 - \ldots - x^tz^t - x^tz^{t+1}\right)^2}.\end{aligned}$$

Hence,

$$\begin{aligned}H(z) &= \frac{\partial}{\partial x}G(x,z)\Big|_{x=1} \\ &= \frac{-z - z^2 - \ldots - z^t}{1 - z - z^2 - z^3 - \ldots - z^t - z^{t+1}} + \frac{z + 2z^2 + \ldots + tz^t + tz^{t+1}}{\left(1 - z - z^2 - z^3 - \ldots - z^t - z^{t+1}\right)^2} \\ &= \frac{-zS_{t-1}(z)}{P(z)} + \frac{zQ_t(z) - z^{t+1}}{P(z)^2}.\end{aligned}$$

Thus we have

$$\sum_k k f_n(k) \sim c_2(n+1)\rho^{-n}$$

where

$$\begin{aligned}c_2 &= \frac{\rho + 2\rho^2 + \ldots + t\rho^t + t\rho^{t+1}}{\rho^2\left((t+1)\rho^t + \ldots + 3\rho^2 + 2\rho + 1\right)^2} \\ &= \frac{Q_t(\rho) - \rho^t}{\rho Q_t(\rho)^2}.\end{aligned}$$

We have so far shown

$$\sum_k k f_n(k) \sim c_2(n+1)\rho^{-n}.$$

In order to compute the variance, we need a sharper estimate for $\sum_k k f_n(k)$. We will show the following:

Theorem 5. *For a fixed t, we have*

$$\sum_k k f_n(k) \sim \big(c_2(n+1) + c_3\big)\rho^{-n}$$

where

$$c_2 = \frac{Q_t(\rho) - \rho^t}{\rho Q_t(\rho)^2} \tag{14}$$

$$c_3 = \frac{2\big(Q_t(\rho) - \rho^t\big)\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{Q_t(\rho)^3} - \frac{\cdot\sum_{j=0}^{t}(j+1)^2\rho^j - (t+1)\rho^t}{\rho Q_t(\rho)^2} - \frac{\sum_{j=0}^{t-1}\rho^j}{Q_t(\rho)}. \tag{15}$$

To complete the proof of Theorem 5, we will derive c_3. We need to consider the residue of $(\rho - z)^{-1}$ and in particular, such a residue which appears in the term of $H(z)$ involving $P(z)^{-2}$. We consider

$$\begin{aligned}
\frac{1}{P(z)^2} &= \frac{1}{\big(1 - z - z^2 - z^3 - \ldots - z^t - z^{t+1}\big)^2} \\
&= \Big(\frac{\alpha}{\rho - z} + \frac{\beta(z)}{R_t(z)}\Big)^2 \\
&\sim \frac{\alpha^2}{(\rho - z)^2} + \frac{2\alpha\beta(\rho)}{(\rho - z)R_t(\rho)}
\end{aligned}$$

where α, β were defined in (12). Note that we can simplify $\beta(\rho)$ as:

$$\beta(\rho) = \frac{\sum_{j=0}^{t-1} Q_j(\rho)\rho^{t-j-1}}{Q_t(\rho)} = \frac{\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{Q_t(\rho)}.$$

For the polynomial $L(z) = zQ_t(z) - z^{t+1}$, we consider its Taylor series at $z = \rho$ and we write

$$L(z) = L(\rho) + \gamma(\rho - z) + (\rho - z)^2 L_1(z)$$

for some polynomial L_1 where γ satisfies

$$\gamma = -L'(\rho) \tag{16}$$

$$= -\sum_{j=0}^{t}(j+1)^2\rho^j + (t+1)\rho^t. \tag{17}$$

We can now compute the residues of $(\rho - z)^{-1}$ and $(\rho - z)^{-2}$ by putting everything together:

$$\begin{aligned}
H(z) &\sim \frac{zQ_t(z) - z^{t+1}}{P(z)^2} - \frac{\sum_{j=1}^{t} z^j}{P(z)} \\
&\sim \big(\rho Q_t(\rho) - \rho^{t+1} + \gamma(\rho - z)\big)\Big(\frac{\alpha^2}{(\rho-z)^2} + \frac{2\alpha\beta(\rho)}{(\rho-z)R_t(\rho)}\Big) - \frac{\sum_{j=1}^{t}\rho^j}{(\rho-z)R_t(\rho)} \\
&\sim \frac{\big(\rho Q_t(\rho) - \rho^{t+1}\big)\alpha^2}{(\rho-z)^2} + \frac{\alpha^2\gamma}{\rho - z} + \frac{\big(\rho Q_t(\rho) - \rho^{t+1}\big)2\alpha\beta(\rho)}{(\rho-z)R_t(\rho)} - \frac{\sum_{j=1}^{t}\rho^j}{(\rho-z)R_t(\rho)} \\
&= \frac{\rho Q_t(\rho) - \rho^{t+1}}{Q_t(\rho)^2(\rho-z)^2} - \frac{\cdot\sum_{j=0}^{t}(j+1)^2\rho^j - (t+1)\rho^t}{Q_t(\rho)^2(\rho - z)} \\
&\quad + \frac{2\big(\rho Q_t(\rho) - \rho^{t+1}\big)\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{Q_t(\rho)^3(\rho-z)} - \frac{\sum_{j=1}^{t}\rho^j}{(\rho-z)Q_t(\rho)} \\
&= \frac{Q_t(\rho) - \rho^t}{\rho Q_t(\rho)^2(1-\frac{z}{\rho})^2} + \frac{2\big(Q_t(\rho) - \rho^t\big)\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{Q_t(\rho)^3(1-\frac{z}{\rho})} \\
&\quad - \frac{\cdot\sum_{j=0}^{t}(j+1)^2\rho^j - (t+1)\rho^t}{\rho Q_t(\rho)^2(1-\frac{z}{\rho})} - \frac{\sum_{j=0}^{t-1}\rho^j}{Q_t(\rho)(1-\frac{z}{\rho})} \\
&= \sum_n \left(c_2(n+1)\Big(\frac{z}{\rho}\Big)^n + c_3\Big(\frac{z}{\rho}\Big)^n\right)
\end{aligned}$$

where

$$c_2 = \frac{Q_t(\rho) - \rho^t}{\rho Q_t(\rho)^2}$$

$$c_3 = \frac{2\big(Q_t(\rho) - \rho^t\big)\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{Q_t(\rho)^3} - \frac{\cdot\sum_{j=0}^{t}(j+1)^2\rho^j - (t+1)\rho^t}{\rho Q_t(\rho)^2} - \frac{\sum_{j=0}^{t-1}\rho^j}{Q_t(\rho)}.$$

Numerical and asymptotic values for c_2 and c_3 are given in Table 2 and (21). They rapidly approach $\frac{1}{2}$ and -1, respectively. Values for the mean are in Section 7.

3.5. The second moment

In this section we estimate the second moment for a fixed t:

$$\sum_{n\geq 0}\sum_{0<k\leq n} k^2 f_n^{(t)}(k)z^n.$$

We first consider

$$K^{(t)}(z)=\sum_{n\geq 0}\sum_{0<k\leq n} k(k-1)f_n^{(t)}(k)z^n$$

since the generating function for $K^{(t)}(z)$ can be obtained by taking the second derivative of $G^{(t)}(x,z)$ with respect to x and then setting $x=1$.

$$\begin{aligned}\frac{\partial^2}{\partial x^2}G(x,z) &= \frac{\partial^2}{\partial x^2}\frac{1+z(1-x)\sum_{j=0}^{t-1}x^jz^j}{1-\sum_{j=1}^{t}x^jz^j-x^tz^{t+1}}\\ &= \frac{\partial}{\partial x}\Big(\frac{\sum_{j=1}^{t}\big((j-1)x^{j-2}-jx^{j-1}\big)z^j}{1-\sum_{j=1}^{t}x^jz^j-x^tz^{t+1}}\\ &\quad+\frac{(1+z-xz\sum_{j=0}^{t-1}x^jz^j)\big(\sum_{j=1}^{t}jx^{j-1}z^j+tx^{t-1}z^{t+1}\big)}{\big(1-\sum_{j=1}^{t}x^jz^j-x^tz^{t+1}\big)^2}\Big).\end{aligned}$$

By substituting $x=1$, we have

$$\begin{aligned}K(z) &= \frac{\partial^2}{\partial x^2}G(x,z)\Big|_{x=1}\\ &= \frac{-\sum_{j=2}^{t}2(j-1)z^j}{P(z)}\\ &\quad+\frac{-2\sum_{j=1}^{t}z^j(zQ_t(z)-z^{t+1})+\sum_{j=2}^{t}j(j-1)z^j+t(t-1)z^{t+1}}{P(z)^2}\\ &\quad+\frac{2(zQ_t(z)-z^{t+1})^2}{P(z)^3}.\end{aligned}$$

To estimate the variance, we only need to estimate the contributions from the terms with denominators $P(z)^2$ and $P(z)^3$ since the contributions from the terms with denominator $P(z)$ are of lower order. So, we will focus on the terms involving $P(z)^{-3}$ and $P(z)^{-2}$ and ignore the terms involving $P(z)^{-1}$.

We need the following useful fact:

$$\frac{1}{P(z)^3}=\Big(\frac{\alpha}{\rho-z}+\frac{\beta(z)}{R_t(z)}\Big)^3\sim\frac{\alpha^3}{(\rho-z)^3}+\frac{3\alpha^2\beta(\rho)}{(\rho-z)^2R_t(\rho)}$$

for α, β defined in (13) and R_t from the proof of Theorem 4. We consider the Taylor series expansion of $(zQ_t(z) - z^{t+1})^2$ at $z = \rho$.

$$\begin{aligned}(zQ_t(z) - z^{t+1})^2 &= \big(L(z)\big)^2 \\ &= L(\rho)^2 - 2L'(\rho)L(\rho)(\rho - z) + L_2(z)(\rho - z)^2 \\ &= (zQ_t(\rho) - \rho^{t+1})^2 + 2(zQ_t(\rho) - \rho^{t+1})\gamma(\rho - z) + L_2(z)(\rho - z)^2\end{aligned}$$

for some polynomial $L_2(z)$ where γ is as defined in (16).

Together we have

$$\begin{aligned}K(z) \sim\ & \frac{2(\rho Q_t(\rho) - \rho^{t+1})^2\alpha^3}{(\rho - z)^3} + \frac{4(zQ_t(\rho) - \rho^{t+1})\gamma\alpha^3}{(\rho - z)^2} + \frac{6(\rho Q_t(\rho) - \rho^{t+1})^2\alpha^2\beta}{(\rho - z)^2 R_t(\rho)} \\ &+ \frac{-2\sum_{j=1}^{t}\rho^j(\rho Q_t(\rho) - \rho^{t+1}) + \sum_{j=2}^{t} j(j-1)\rho^j + t(t-1)\rho^{t+1}}{(\rho - z)^2 R_t(\rho)^2} \\ =\ & \frac{2(\rho Q_t(\rho) - \rho^{t+1})^2}{(\rho - z)^3 Q_t(\rho)^3} + \frac{4(\rho Q_t(\rho) - \rho^{t+1})\big(-\sum_{j=0}^{t}(j+1)^2\rho^j + (t+1)\rho^t\big)}{(\rho - z)^2 Q_t(\rho)^3} \\ &+ \frac{6(\rho Q_t(\rho) - \rho^{t+1})^2\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{(\rho - z)^2 Q_t(\rho)^4} \\ &+ \frac{-2\sum_{j=1}^{t}\rho^j(\rho Q_t(\rho) - \rho^{t+1}) + \sum_{j=2}^{t} j(j-1)\rho^j + t(t-1)\rho^{t+1}}{(\rho - z)^2 Q_t(\rho)^2}.\end{aligned}$$

Hence, we have

$$H(z) \sim c_4\binom{n+2}{2}\Big(\frac{z}{\rho}\Big)^n + c_5(n+1)\Big(\frac{z}{\rho}\Big)^n + O(1)\Big(\frac{z}{\rho}\Big)^n$$

where

$$c_4 = \frac{2(Q_t(\rho) - \rho^t)^2}{\rho Q_t(\rho)^3} \tag{18}$$

$$\begin{aligned}c_5 =\ & \frac{4(Q_t(\rho) - \rho^t)\big(-\sum_{j=0}^{t}(j+1)^2\rho^j + (t+1)\rho^t\big)}{\rho Q_t(\rho)^3} + \frac{6(Q_t(\rho) - \rho^t)^2\sum_{j=1}^{t}\binom{j+1}{2}\rho^{j-1}}{Q_t(\rho)^4} \\ &+ \frac{-2\sum_{j=0}^{t-1}\rho^j(Q_t(\rho) - \rho^t) + \sum_{j=2}^{t} j(j-1)\rho^{j-2} + t(t-1)\rho^{t-1}}{Q_t(\rho)^2}.\end{aligned} \tag{19}$$

Together, we have proved the following:

Theorem 6. *For a fixed t, we have*

$$\sum_k k^2 f_n(k) = \left(c_4\binom{n+2}{2} + (c_2 + c_5)(n+1) + O(1)\right)\rho^{-n}$$

where c_2, c_4 and c_5 are defined in (14), (18) and (19).

Numerical and asymptotic values for c_2, c_4 and c_5 are given in Table 2 and (21). They rapidly approach $\frac{1}{2}, 1$ and -3, respectively.

3.6. The variance

We can finally evaluate the variance:

$$\mathrm{Var}(n) \sim \frac{c_4\binom{n+2}{2} + (c_2+c_5)(n+1)}{c_1} - \Big(\frac{c_2(n+1)+c_3}{c_1}\Big)^2.$$

We note that the coefficient of n^2 vanishes since $c_4c_1 = 2c_2^2$. Thus, we have

$$\mathrm{Var}(n) = \frac{(\frac{3c_4}{2} + c_2 + c_5)c_1 - 2c_2(c_2+c_3)}{c_1^2} n + O(1) = c_7 n + O(1). \tag{20}$$

Although the exact expressions for c_i's are rather complicated, there is a great deal of cancellation for c_7 and in particular, for $c_5c_1 - 2c_2c_3$. By substituting the c_i's into (20), we have

Theorem 7.

$$\begin{aligned} Var^{(t)}(n) &= c_7 n + O(1) \\ &= \left(\frac{(Q_t(\rho) - (1+2t)\rho^t)\rho^t}{Q_t(\rho)^2} + \frac{\rho^{2t}\sum_{j=1}^t j(j+1)\rho^j}{Q_t(\rho)^3}\right) n + O(1) \\ &= \Bigg(\frac{\rho^t(1+2\rho+\ldots+t\rho^{t-1} - (t+1)\rho^t)}{\big(1+2\rho+\ldots+t\rho^{t-1}+(t+1)\rho^t\big)^2} \\ &\qquad + \frac{\rho^{2t}\sum_{j=1}^t j(j+1)\rho^j}{\big(1+2\rho+\ldots+t\rho^{t-1}+(t+1)\rho^t\big)^3}\Bigg) n + O(1). \end{aligned}$$

In Table 2, we list some values of $\rho^{(t)}$ and $c_i^{(t)}$ for small values of t.

Rough asymptotic estimates are relatively simple. Using Fact 5, with some straightforward computation, we have

$$\begin{aligned} \rho^{(t)} &= \frac{1}{2} + O(2^{-t}) \\ Q_t(\rho^{(t)}) &= 4 + O(t2^{-t}) \\ c_1^{(t)} &= \frac{1}{2} + O(t2^{-t}) \\ c_2^{(t)} &= \frac{1}{2} + O(t2^{-t}) \end{aligned} \tag{21}$$

Table 3
Table of small values of $c_8^{(t)}$ and $c_9^{(t)}$.

t	$c_8^{(t)}$	$c_9^{(t)}$
1	.95491 . . .	.04262 . . .
2	.94165 . . .	.02822 . . .
3	.95441 . . .	.01505 . . .
4	.96927 . . .	.00768 . . .
5	.98069 . . .	.00387 . . .
10	.99878 . . .	.00012 . . .

$$c_4^{(t)} = 1 + O(t2^{-t})$$
$$c_7^{(t)} = \frac{1}{2^{t+2}} + O(t2^{-2t}).$$

The table above is consistent with this asymptotic behavior. The results in this section are used to give accurate estimates of the variance and sample size required in Section 7 below.

To conclude this section, the generating function of Fact 4 will be used to get the asymptotic behavior of the variance of the naive importance sampling estimator $T(\sigma) = \frac{1}{Pr(\sigma)}$ as in the example for $\mathcal{F}_{1,1}(n)$ in Section 2. The argument is very similar to the facts above so we will be brief. We want to compute

$$K_n^{(t)} := \frac{1}{S_n^2}\sum_{\sigma}\frac{1}{Pr(\sigma)} \sim c_8^{(t)} e^{c_9^{(t)} n}. \tag{22}$$

From Fact 4, we have $G(2, z) = P(z)/Q(z)$ with

$$Q(z) = 1 - 2z - (2z)^2 - \ldots - (2z)^t - (2z)^{t+1}.$$

Arguing as in Fact 6, the largest real root ρ satisfies

$$\frac{1}{4} + \frac{1}{4^{t+3}} < \rho < \frac{1}{4} + \frac{1}{4^{t+3}} + \frac{t}{4^{2t+3}}.$$

We carry out the same analysis as above for the cases $t = 1, 2, 3, 4, 5$ and 10, namely, we derive estimates for the asymptotic behavior of the coefficients of the corresponding generating functions (we omit the details). This results in Table 3.

In particular, it follows that

$$c_9^{(1)} = \ln\left(\frac{3-\sqrt{5}}{\sqrt{3}-1}\right) = 0.0426288\ldots.$$

It is not hard to show that $c_8^{(t)}$ tends to 1 and $c_9^{(t)}$ decreases like 2^{-t} as $t \to \infty$. These asymptotics are used in Section 7 to obtain variance-based estimates of the required sample size.

4. The $(2,2)$-graph $B_{2,2}(n)$

4.1. Introduction

In this section we will consider our usual procedure for selecting a random perfect matching from the graph $B_{2,2}(n)$. This is a bipartite graph with vertex sets $U_n = \{u_1, u_2, \ldots, u_n\}$ and $V_n = \{v_1, v_2, \ldots v_n\}$ and with edges $\{u_i, v_j\}$ for all i, j satisfying $|i - j| \leq 2$. Thus, starting with u_1, we select an incident random edge $\{u_1, v_j\}$, repeating this process with vertices $u_2, u_3, \ldots, u_n$ but always making sure that at any point, the edges chosen so far are part of a perfect matching in $B_{2,2}(n)$. The number of perfect matchings in G_n is given by $Per(M_n)$, the permanent of the $n \times n$ matrix M_n where $M_n[i,j] = 1$ if and only if $|i-j| \leq 2$. If S_n denotes $Per(M_n)$ then it is known (see [15]) that S_n satisfies the linear recurrence

$$S_n = 2S_{n-1} + 2S_{n-3} - S_{n-5} \tag{23}$$

with the initial values $S_0 = 1, S_1 = 1, S_2 = 2, S_3 = 6, S_4 = 14$. In particular, the (ordinary) generating function for S_n is given by (see [15])

$$F(z) = \sum_{n \geq 0} S_n z^n = \frac{1-z}{1 - 2z - 2z^3 + z^5}. \tag{24}$$

It then follows by standard techniques (see [17]) that

$$S_n \sim c_1 \rho^{-n} \tag{25}$$

where

$$c_1 = \frac{1-\rho}{\rho(2 + 6\rho^2 - 5\rho^4)} = 0.45463889\ldots \tag{26}$$

and $\rho = 0.428530860\ldots$ is the smallest real root of the polynomial $p(z) = 1-2z-2z^3+z^5$.

When $n = 200$, S_n is approximately 1.825×10^{73}. Two of our main goals for this section are to estimate the quantities:

$$K_n := \frac{1}{S_n^2} \sum_{\sigma} \frac{1}{Pr(\sigma)} \quad \text{and} \quad L_n := \frac{1}{S_n} \sum_{\sigma} \ln\left(\frac{1}{Pr(\sigma)}\right) - \ln(S(n)) \tag{27}$$

where the sums are over all perfect matchings σ of $B_{2,2}(n)$.

It is not hard to see that $B_{2,2}(n)$ has no cycles of length greater than 3. In fact, there are only two types of 2-cycles $((i, i+1), (i, i+2))$ and two types of 3-cycles $((i, i+1, i+2), (i, i+2, i+1))$.

4.2. The matrix $M_n(x,y)$

It is easy to see that in sequentially selecting the edges in our random perfect matching, it is ordinarily the case that an edge is selected with probability 1/3. However, if when choosing an edge for u_k, it happens that no edge has yet been chosen for v_{k-2}, then we must choose the edge $\{u_k, v_{k-2}\}$, i.e., this edge is chosen with probability 1. As usual, when we reach the final vertex u_n, there will be only one unoccupied vertex v_k and so the edge for u_n is forced. But there is one more situation to consider, namely for the vertex u_{n-1}. At this point, there are only two choices for an edge for u_{n-1} so that the probability of choosing either one of them is 1/2. The exception to this statement is when v_{n-3} is unoccupied in which case the edge from u_{n-1} is forced.

All of this information can be summarized by computing the permanent of a matrix $M_n(x,y)$, defined as follows. We start with the standard $n \times n$ matrix which has $x's$ on the four diagonals with $-2 \le i-j \le 2$ and which has the entry 1 when $i-j=2$, and of course, $0's$ everywhere else. However, to form $M_n(x,y)$ we modify the bottom two rows. Namely, we replace the last three entries of the next to the last row by y, and we replace the last three entries of the last row by 1. (This description is only meaningful when $n \ge 3$.) We show $M_8(x,y)$ below.

$$M_8(x,y) = \begin{bmatrix} x & x & x & 0 & 0 & 0 & 0 & 0 \\ x & x & x & x & 0 & 0 & 0 & 0 \\ 1 & x & x & x & x & 0 & 0 & 0 \\ 0 & 1 & x & x & x & x & 0 & 0 \\ 0 & 0 & 1 & x & x & x & x & 0 \\ 0 & 0 & 0 & 1 & x & x & x & x \\ 0 & 0 & 0 & 0 & 1 & y & y & y \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

From our discussion above, it follows that the permanent $PerM_n = PerM_n(x,y)$ captures all the information needed to compute the probabilities that a particular perfect matching will be chosen. For example, for $n = 8$ we find that

$$PerM_n = 114x^6y + 156x^5y + 48x^4y + 36x^6 + 38x^5 + 8x^4.$$

This tells us that there are 114 perfect matchings in $B_{2,2}$ that occur with probability $(\frac{1}{3})^6(\frac{1}{2})$, 156 that occur with probability $(\frac{1}{3})^5(\frac{1}{2})$, ..., and finally 8 that occur with probability $(\frac{1}{3})^4$. Of course, substituting $x = y = 1$ in $PerM_n$ yields the value 400 which is the total number of perfect matchings of $B_{2,2}(8)$.

4.3. A recurrence for $PerM_n$

Let us write

$$PerM_n = F_n(x,y) = \sum_{k,l} f_n(k,l)x^k y^l$$

where $f_n(k,l)$ denotes the number of perfect matchings which have k edges chosen with probability $\frac{1}{3}$ and l edges chosen with probability $\frac{1}{2}$. Our first goal is to find a recurrence for $F_n = F_n(x,y)$. Assuming $n \geq 6$, we will expand the permanent of M_n along the top row, and then recursively do the same thing for the resulting smaller matrices. This will create three different types of matrices which we show below.

$$A_8 = Per \begin{bmatrix} x & x & x & 0 & 0 & 0 & 0 & 0 \\ x & x & x & x & 0 & 0 & 0 & 0 \\ 1 & x & x & x & x & 0 & 0 & 0 \\ 0 & 1 & x & x & x & x & 0 & 0 \\ 0 & 0 & 1 & x & x & x & x & 0 \\ 0 & 0 & 0 & 1 & x & x & x & x \\ 0 & 0 & 0 & 0 & 1 & y & y & y \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

$$B_8 = Per \begin{bmatrix} x & x & x & 0 & 0 & 0 & 0 & 0 \\ 1 & x & x & x & 0 & 0 & 0 & 0 \\ 0 & x & x & x & x & 0 & 0 & 0 \\ 0 & 1 & x & x & x & x & 0 & 0 \\ 0 & 0 & 1 & x & x & x & x & 0 \\ 0 & 0 & 0 & 1 & x & x & x & x \\ 0 & 0 & 0 & 0 & 1 & y & y & y \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

$$C_8 = Per \begin{bmatrix} x & x & x & 0 & 0 & 0 & 0 & 0 \\ 1 & x & x & x & 0 & 0 & 0 & 0 \\ 0 & 1 & x & x & x & 0 & 0 & 0 \\ 0 & 0 & x & x & x & x & 0 & 0 \\ 0 & 0 & 1 & x & x & x & x & 0 \\ 0 & 0 & 0 & 1 & x & x & x & x \\ 0 & 0 & 0 & 0 & 1 & y & y & y \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

We only show the three matrix types for $n = 8$ but the corresponding cases for general n should be clear. We will usually suppress the variables x and y, and write A_n instead of $A_n(x,y)$, etc. Now, by expanding these permanents along their top rows, we find the following relations:

$$A_n = xA_{n-1} + xB_{n-1} + xC_{n-1}, \tag{28}$$

$$B_n = xA_{n-1} + xA_{n-2} + xB_{n-2}, \tag{29}$$

$$C_n = xB_{n-1} + xA_{n-2} + xA_{n-3}. \tag{30}$$

To solve this system of recurrences, first substitute the value of C_n into (28) to obtain

$$A_n = xA_{n-1} + xB_{n-1} + x^2B_{n-2} + x^2A_{n-3} + x^2A_{n-4} \tag{31}$$

which implies

$$X_n := A_n - xA_{n-1} - x^2 A_{n-3} - x^2 A_{n-4} = xB_{n-1} + x^2 B_{n-2},$$
$$Y_n := xA_{n-1} + xA_{n-2} = B_n - xB_{n-2}, \tag{32}$$

from (30) and (31). Thus,

$$X_{n+1} = xB_n + x^2 B_{n-1}, \tag{33}$$
$$X_n + xY_n = xB_n + xB_{n-1}. \tag{34}$$

Subtracting (34) from (33) we obtain

$$X_{n+1} - X_n - xY_n = (x^2 - x)B_{n-1}. \tag{35}$$

Now, solving (35) for B_{n-1}, we can substitute into (32) and get

$$(x-1)(X_n + xY_n) = (x^2 - x)B_n + (x^2 - x)B_{n-1},$$
$$xX_n - X_n + x^2 Y_n - xY_n = X_{n+2} - X_{n+1} - xY_{n+1} + X_{n+1} - X_n - xY_n,$$
$$xX_n + x^2 Y_n = X_{n+2} - xY_{n+1}. \tag{36}$$

Finally, replacing X_n and Y_n by their expressions in terms of A_n and simplifying, we have

$$A_n = xA_{n-1} + x(x+1)A_{n-2} + x^2(x+1)A_{n-3}$$
$$+ x^2(x+1)A_{n-4} - x^3 A_{n-5} - x^3 A_{n-6}, \tag{37}$$

for $n \geq 7$. Thus, (37) together with the initial values:

$$A_0 = 1,$$
$$A_1 = 1,$$
$$A_2 = 2y,$$
$$A_3 = 6xy,$$
$$A_4 = 10x^2 y + 4x^2,$$
$$A_5 = 18x^3 y + 6x^3 + 6x^2 y + x^2,$$
$$A_6 = 34x^4 y + 10x^4 + 24x^3 y + 3x^3 + 2x^2 y,$$
$$A_7 = 62x^5 y + 20x^5 + 64x^4 y + 14x^4 + 12x^3 y,$$

determine the recurrence for $A_n = PerM_n = F_n(x, y)$.

Notice that if we substitute $x = y = 1$ in (37), we get a recurrence for $S_n = PerM_n$:

$$S_n = S_{n-1} + 2S_{n-2} + 2S_{n-3} + 2S_{n-4} - S_{n-5} - S_{n-6} \tag{38}$$

which has the characteristic polynomial

$$1 - z - 2z^2 - 2z^3 - 2z^4 + z^5 + z^6 = (1+z)(1 - 2z - 2z^3 + z^5). \tag{39}$$

Of course, this must have the factor $p(z) = 1 - 2z - 2z^3 + z^5$ which occurs in (24).

4.4. A generating function for F_n

Our next step will be to determine the generating function

$$G(x,y,z) = \sum_{n\geq 0} F_n(x,y)z^n = \sum_n \sum_{k,l} f_n(k,l)x^k y^l z^n. \tag{40}$$

Theorem 8.

$$G(x,y,z) = \frac{P}{Q}$$

where

$$P = 1 + 2x^2(x-y)z^5 - 2x(x-1)(x-y)z^4 - x(x^2+2x-4y+1)z^3 \tag{41}$$
$$+ (-x^2 - 2x + 2y)z^2 + (-x+1)z$$

and

$$Q = 1 - xz - x(x+1)z^2 - x^2(x+1)z^4 + x^3z^5 + x^3z^6. \tag{42}$$

Proof. We start with the recurrence for A_n in (36). By multiplying A_n by z^n and summing over all $n \geq 6$, we have

$$\begin{aligned}
G(x,y,z) - \sum_{j=0}^{5} A_j z^j &= \sum_{n\geq 6} A_n z^n \\
&= \sum_{n\geq 6} \Big(xA_{n-1} + x(x+1)A_{n-2} + x^2(x+1)A_{n-3} \\
&\quad + x^2(x+1)A_{n-4} - x^3A_{n-5} - x^3A_{n-6} \Big) \\
&= G(x,y,z)\big(xz + x(x+1)z^2 + x^2(x+1)z^3 \\
&\quad + x^2(x+1)z^4 - x^3z^5 - x^3z^6\big) - xz\sum_{j=0}^{4} A_j z^j
\end{aligned}$$

$$-x(x+1)z^2\sum_{j=0}^{3}A_jz^j - x^2(x+1)z^3\sum_{j=0}^{2}A_jz^j$$
$$- x^2(x+1)z^4\sum_{j=0}^{1}A_jz^j + x^3z^5.$$

After substituting for A_j for $0 \le j \le 5$, we have

$$\begin{aligned}&(1 - xz - x(x+1)z^2 - x^2(x+1)z^3 - x^2(x+1)z^4 + x^3z^5 + x^3z^6)G(x,z)\\&= 1 + 2x^2(x-y)z^5 - 2x(x-1)(x-y)z^4 - x(x^2+2x-4y+1)z^3\\&\quad + (-x^2-2x+2y)z^2 + (-x+1)z.\end{aligned}$$

Theorem 8 is proved. □

4.5. Analyzing $G(x, y, z)$

We will find it convenient to split $G = G(x, y, z)$ into two parts: one for matchings that don't use y, and one for matchings that do. Thus, $G = G_0 + yG_1$ where

$$G_0 = \frac{1 + 2x^3z^5 + (-2x^3+2x^2)z^4 - x(x+1)^2z^3 + (-x^2-2x)z^2 + (-x+1)z}{Q} \tag{43}$$

$$G_1 = \frac{2z^2 + 4xz^3 + (2x^2-2x)z^4 - 2x^2z^5}{Q}. \tag{44}$$

Substituting $x = 3$ in the above yields

$$H_0 = G_0\big|_{x=3} = \frac{54z^5 - 36z^4 - 48z^3 - 15z^2 - 2z + 1}{27z^6 + 27z^5 - 36z^4 - 36z^3 - 12z^2 - 3z + 1} \tag{45}$$

$$= 1 + z + 4x^2z^4 + (6x^3+x^2)z^5 + (10x^4+3x^3)z^6 + \ldots \tag{46}$$

$$H_1 = G_1\big|_{x=3} = \frac{2z^2(xz+1)(xz^2-xz-1)}{27z^6 + 27z^5 - 36z^4 - 36z^3 - 12z^2 - 3z + 1} \tag{47}$$

$$= 2z^2 + 6xz^3 + 10x^2z^4 + (18x^3+6x^2)z^5 + (34x^4+24x^3+2x^2)z^6 + \ldots. \tag{48}$$

We need to get asymptotic estimates for the coefficients of H_0 and H_1. Let $\rho_1 = 0.164399\ldots$ denote the (real) root of minimum modulus of $27z^6 + 27z^5 - 36z^4 - 36z^3 - 12z^2 - 3z + 1$. Using the usual techniques for doing this, we decompose H_0 and H_1 into partial fractions where the only terms that matter for us are those with denominators having factors of the form $(z - \rho_1)^k$ for some $k \ge 1$. For H_0 it turns out to be (courtesy of Maple):

$$\frac{54\rho_1^5 - 36\rho_1^4 - 48\rho_1^3 - 15\rho_1^2 - 2\rho_1 + 1}{(z-\rho_1)(162\rho_1^5 + 135\rho_1^4 - 144\rho_1^3 - 108\rho_1^2 - 24\rho_1 - 3}$$

while for H_1 it is:

$$\frac{2\rho_1^2(-9\rho_1^3+6\rho_1^2+6\rho_1+1)}{(z-\rho_1)(162\rho_1^5+135\rho_1^4-144\rho_1^3-108\rho_1^2-24\rho_1-3}.$$

Thus, we find:

$$[z^k]H_0 = f_n(k,0) \sim C_0\rho_1^{-k}, \qquad [z^k]H_1 = f_n(k,1) \sim C_1\rho_1^{-k}$$

where

$$C_0 = \frac{54\rho_1^5-36\rho_1^4-48\rho_1^3-15\rho_1^2-2\rho_1+1}{(-\rho_1)(162\rho_1^5+135\rho_1^4-144\rho_1^3-108\rho_1^2-24\rho_1-3} = 0.19155\ldots,$$
$$C_1 = \frac{2\rho_1^2(-9\rho_1^3+6\rho_1^2+6\rho_1+1)}{(-\rho_1)(162\rho_1^5+135\rho_1^4-144\rho_1^3-108\rho_1^2-24\rho_1-3} = 0.66751\ldots.$$

Therefore, recalling the definition of the relative variance from (27),

$$K_n := \frac{1}{S_n^2}\sum_\sigma \frac{1}{Pr(\sigma)}$$

we have

$$K_n \sim \frac{(C_0+2C_1)\rho_1^{-n}}{c_1^2\rho^{-2n}} = \left(\frac{C_0+2C_1}{c_1^2}\right)\left(\frac{\rho^2}{\rho_1}\right)^n \tag{49}$$

where

$$\frac{C_0+2C_1}{c_1^2} = 0.73856\ldots, \qquad \frac{\rho^2}{\rho_1} = 1.11702\ldots, \qquad \ln\left(\frac{\rho^2}{\rho_1}\right) = 0.11067\ldots.$$

Our next goal is to estimate the sum in

$$L_n := \frac{1}{S_n}\sum_\sigma \ln\left(\frac{1}{Pr(\sigma)}\right) - \ln(S(n)).$$

Recall from (40) that

$$G(x,y,z) = \sum_{n\ge 0} F_n(x,y)z^n = \sum_n\sum_{k,l} f_n(k,l)x^k y^l z^n,$$

where $f_n(k,l)$ is the number of matchings σ which have k edges chosen with probability 1/3 and l edges chosen with probability 1/2. Hence, the probability that this σ occurs is just $Pr(\sigma) = (1/3)^k(1/2)^l$, and so,

$$\ln\left(\frac{1}{Pr(\sigma)}\right) = k\ln 3 + l\ln 2.$$

Thus, the sum in L_n can be split into two sums; Σ_0 which is a sum over all $\sigma = \sigma_0$ which have no probability 1/2 edges, and Σ_1 which is a sum over all $\sigma = \sigma_1$ which have a single probability 1/2 edge. In other words,

$$\Sigma_0 = \sum_{\sigma_0} \ln\left(\frac{1}{Pr(\sigma_0)}\right) = \sum_k f_n(k,0)(k\ln 3),$$

$$\Sigma_1 = \sum_{\sigma_1} \ln\left(\frac{1}{Pr(\sigma_1)}\right) = \sum_k f_n(k,1)(k\ln 3 + \ln 2).$$

To compute the contributions of terms involving $\ln 3$ in both Σ_0 and Σ_1, we differentiate G with respect to x, and set $x = y = 1$, resulting in

$$J_0 = \frac{z^3(2z^6 - 2z^5 - 6z^4 + 5z^3 - 2z^2 + 4z + 6)}{(z^5 - 2z^3 - 2z + 1)^2}.$$

To compute Σ_1, we differentiate G with respect to y, and set $x = 1$, resulting in

$$J_1 = \frac{-2(z^2 - z - 1)z^2}{(z^5 - 2z^3 - 2z + 1)}.$$

Expanding J_1 into partial fractions, with $\rho = 0.42853\ldots$ being the real root of minimum modulus of $z^5 - 2z^3 - 2z + 1$, we find the only relevant term in the expansion is:

$$\frac{-2\rho^2(\rho^2 - \rho - 1)}{(5\rho^4 - 6\rho^2 - 2)(z - \rho)}.$$

Hence,

$$[z^n]J_1 \sim K_1\rho^{-n}$$

where

$$K_1 = \frac{-2\rho^2(\rho^2 - \rho - 1)}{(5\rho^4 - 6\rho^2 - 2)(-\rho)} = 0.36374.$$

Expanding J_0 into partial fractions, we find that there are two terms of interest in the expansion (since the denominator of J_0 has a repeated root). They are:

$$\frac{\rho^3(2\rho^6 - 2\rho^5 - 6\rho^4 + 5\rho^3 - 2\rho^2 + 4\rho + 6)}{(5\rho^4 - 6\rho^2 - 2)^2(z - \rho)^2}$$

and

$$\frac{-2\rho^2(25\rho^{10}-20\rho^9-87\rho^8+61\rho^7+67\rho^6-44\rho^5+45\rho^4-54\rho^3-8\rho^2-16\rho-18)}{(5\rho^4-6\rho^2-2)^3(z-\rho)}.$$

This implies that

$$[z^n]J_0 \sim (K_0(n+1)+K_0')\rho^{-n}$$

with $K_0 = 0.37462\ldots$ and $K_0' = -0.99234\ldots$ (obtained by evaluating the above expressions at $z = 0$). Therefore, we have

$$\begin{aligned}\frac{1}{S_n}\sum_\sigma \ln\left(\frac{1}{Pr(\sigma)}\right) &= \frac{(K_0(n+1)+K_0')\ln 3 + K_1 \ln 2}{c_1} + o(1)\\ &= (0.90526\ldots)n - 0.93811\cdots + o(1).\end{aligned}$$

Thus,

$$\begin{aligned}\frac{1}{S_n}\sum_\sigma \ln\left(\frac{1}{Pr(\sigma)}\right) - \ln S_n &= \left(\left(\frac{K_0\ln 3}{c_1}\right) - \ln(\rho^{-1})\right)n\\ &\quad + \frac{(K_0+K_0')\ln 3 + K_1\ln 2}{c_1} - \ln c_1 + o(1) \qquad (50)\\ &= (0.05786\ldots)n - 0.14987\ldots + o(1).\end{aligned}$$

These asymptotics are used to compare algorithms in Section 7.

5. Two different models for generating perfect matchings in the Fibonacci graph

5.1. Introduction

All of the sections above have used a sequential importance algorithm that used the 'from the top' ordering of $(u_1, u_2, \ldots, u_n)$. It is natural to wonder if changing the order helps or hinders. In this section we will consider matchings in the Fibonacci graph $B = B_{1,1}(n)$. We will assume that $n = 3m$. We will designate the m vertices $\{u_2, u_5, \ldots, u_{3k+2}, \ldots, u_{3m-1}\}$ as *distinguished* vertices in U_{3m}. We can think of the vertices of B as partitioned into m blocks $D_k, 1 \le k \le m$, where D_k consists of the six vertices $\{u_{3k-2}, u_{3k-1}, u_{3k}, v_{3k-2}, v_{3k-1}, v_{3k}\}$. We will select our random matching in B in two phases.

For Phase I, for each distinguished vertex u_{3k-2} we independently choose a random edge incident to it in B. For Phase II, we then randomly select edges for the remaining vertices in U_{3m} (now the order doesn't matter), always making sure that each edge chosen is part of a perfect matching in B. In Phase I, there are three choices for each u_{3k-2}, namely $\{u_{3k-2}, v_{3k-3}\}$, $\{u_{3k-2}, v_{3k-2}\}$ or $\{u_{3k-2}, v_{3k-1}\}$. We will call the first choice an *up* edge, the second choice a *level* edge and the third choice a *down* edge. Let us consider what happens between two consecutive block D_k and D_{k+1}. There are nine total choices for the edges incident to the corresponding distinguished vertices u_{3k-2} and u_{3k+1}. If u_{3k-2} makes an 'up' or 'level' choice, **and** u_{3k+1} makes a 'level' or 'down' choice then there will be two possible choices in the second phase for the vertices u_{3k-1} and u_{3k}, namely $\{u_{3k-1}, v_{3k-1}\}$ and $\{u_{3k}, v_{3k}\}$, or $\{u_{3k-1}, v_{3k}\}$ and $\{u_{3k}, v_{3k-1}\}$. Thus, for four of the nine choices for the edges from the two distinguished vertices, there are two choices for the two vertices between them. Let us say that this is a 'good' transition from D_k to D_{k+1}. It is easy to check that for the other five choices for edges from u_{3k-1} and u_{3k}, there is only one possible choice for the edges from u_{3k-1} and u_{3k}.

Let us denote by $t_m(k)$ the number of Phase I choices which have k 'good' transitions. Define

$$T(x, y) = \sum_{m\geq 0} \sum_{0\leq k\leq m-1} t_m(k)x^k y^m.$$

Thus,

$$T(x, y) = 3 + (4x + 5)y + (4x^2 + 16x + 7)y^2 + (4x^3 + 28x^2 + 40x + 9)y^3 \dots.$$

Note that setting $x = 1$ in the coefficient for y^k, we get the value 3^{k+1}, which is just the total number of ways that the edges can be chosen for $k + 1$ distinguished vertices.

5.2. A closed form for $T(x, y)$

Our first goal will be to derive a recurrence for the $t_m(k)$. Define $a_m(j)$ to be the number of Phase I choices which have j 'good' transitions and for which the first distinguished vertex has an 'up' edge chosen. Similarly, define $b_m(j)$ to be the number of Phase I choices which have j 'good' transitions and for which the first distinguished vertex has a 'level' edge chosen, and let $c_m(j)$ be the corresponding number where a 'down' edge was chosen. It is not hard to see that the following recursive relations hold:

$$\begin{aligned} a_m(j) &= a_{m-1}(j) + b_{m-1}(j-1) + c_{m-1}(j-1), \\ b_m(j) &= a_{m-1}(j) + b_{m-1}(j-1) + c_{m-1}(j-1), \\ c_m(j) &= a_{m-1}(j) + b_{m-1}(j) + c_{m-1}(j) \end{aligned} \tag{51}$$

with $a_1(0) = b_1(0) = c_1(0) = 1$ where $m \geq 1$ and $0 \leq j \leq m - 1$. Of course, $t_m(j) = a_m(j) + b_m(j) + c_m(j)$. Eliminating the variables b_m and c_m, we end up with the recurrence:

$$a_m(j) = 2a_{m-1}(j) + a_{m-1}(j-1) - a_{m-2}(j) + a_{m-2}(j-1) \tag{52}$$

with initial conditions $a_0(0) = a_1(0) = 1$ and $a_m(j) = 0$ for $j \geq m \geq 1$ or $j < 0$ or $m < 0$. Thus, $a_m(0) = 1$ for $m \geq 0, a_2(1) = 2, a_3(1) = 6, a_3(2) = 2$, etc.

First we derive the generating function

$$A(x,y)=\sum_{\substack{m\geq 1\\ 0\leq j\leq m-1}} a_m(j)x^jy^{m-1}.$$

From the recurrence in (52), we consider

$$\begin{aligned}\sum_{\substack{m\geq 3\\ 0\leq j\leq m-1}} a_m(j)x^jy^{m-1} &= A(x,y)-1-(1+2x)y\\ &= \sum_{\substack{m\geq 3\\ 0\leq j\leq m-1}} \big(2a_{m-1}(j)+a_{m-1}(j-1)-a_{m-2}(j)+a_{m-2}(j-1)\big)\\ &= (2y+xy-y^2+xy^2)A(x,y)-2y-xy.\end{aligned}$$

Thus we have

$$A(x,y)=\frac{1-y+xy}{1-2y-y^2-xy-xy^2}.$$

Since $a_m(j)=b_m(j)$, we have

$$\begin{aligned}B(x,y) &= \sum_{\substack{m\geq 1\\ 0\leq j\leq m-1}} b_m(j)x^jy^{m-1}\\ &= A(x,y).\end{aligned}$$

To derive the generating function

$$C(x,y)=\sum_{\substack{m\geq 1\\ 0\leq j\leq m-1}} c_m(j)x^jy^{m-1},$$

we consider the following sum over the recurrence (51).

$$\begin{aligned}\sum_{\substack{m\geq 2\\ 0\leq j\leq m-1}} a_m(j)x^jy^{m-1} &= A(x,y)-1\\ &= \sum_{\substack{m\geq 2\\ 0\leq j\leq m-1}} \big(a_{m-1}(j)+b_{m-1}(j-1)+c_{m-1}(j-1)\big)\\ &= yA(x,y)+xyA(x,y)+xyC(x,y).\end{aligned}$$

Thus,

$$C(x,y) = \frac{1}{xy}\big(A(x.y)(1-y-xy)-1\big)$$
$$= \frac{1+y-xy}{1-2y-y^2-xy-xy^2}.$$

So, we have

$$T(x,y) = \sum_{\substack{m\geq 0\\ 0\leq j\leq m-1}} \big(a_m(j)+b_m(j)+c_m(j)\big)x^j y^m$$
$$= 1 + y\Big(A(x,y)+B(x,y)+C(x,y)\Big)$$
$$= 1 + \frac{y(3-y+xy)}{1-2y-y^2-xy-xy^2}$$
$$= \frac{1+y-xy}{1-2y-y^2-xy-xy^2}.$$

5.3. Completing the computations

It is clear that since for each 'good' transition there are two choices for the Phase II process, the total number of matchings you get this way is $\sum_k 2^k t_m(k) = F_{3m+1}$ (the Fibonacci number, by our previous remarks). Let us define

$$f_m(k) = 2^k t_m(k), \quad F(x,y) = \sum_{m\geq 0}\sum_{k\geq 0} f_m(k)x^k y^m.$$

Thus, the generating function for $F(x,y)$ is:

$$F(x,y) = G(2x,y) = \frac{1-(2x-1)y}{1-(2x+2)y-(2x-1)y^2}$$
$$= 1+3y+(8x+5)y^2+(16x^2+32x+7)y^3\dots \tag{53}$$

so that

$$H(y) = F(x,y)\Big|_{x=1} = \frac{1-y}{1-4y-y^2}$$
$$= 1+3y+13y^2+55y^3+\ldots+F_{3m+1}y^m+\ldots.$$

In particular, if ρ denotes the value $\sqrt{5}-2$, which is the root with minimum modulus of $1-4y-y^2$ then we find

$$F_{3m+1} = c_1\rho^{-m} + o(1)$$

where

$$c_1 = \frac{5+\sqrt{5}}{10} = 0.72360\ldots.$$

To compute the first moment $\sum_{k\geq 0} kf_m(k)$, we form

$$F_1(x,y) = \frac{\partial F}{\partial x} = \frac{y^2}{1-(2x+2)y-(2x-1)y^2,}$$
$$H_1(y) = F_1(x,y)\Big|_{x=1} = \frac{y^2}{(1-4y-y^2)^2}.$$

The standard decomposition of $H_1(y)$ into partial fraction yields the two relevant terms

$$\frac{2(-2+\sqrt{5})^2}{5(y+2-\sqrt{5})^2} + \frac{2\sqrt{5}}{25(y+2-\sqrt{5})}.$$

Therefore, defining

$$c_2 = \frac{2(-2+\sqrt{5})^2}{5(2-\sqrt{5})^2} = 0.40000\ldots$$
$$c_3 = \frac{2\sqrt{5}}{25(2-\sqrt{5})} = -0.75777\ldots$$

we find

$$[y^m]H_1(y) = \sum_{k\geq 0} kf_m(k) = (c_2(m+1)+c_3)\rho^{-m} + O(1).$$

To compute the second moment $\sum_{k\geq 0} k^2 f_m(k)$, we first form

$$F_2(x,y) = \frac{\partial F_1}{\partial x} = \frac{-32y^3(1+y)}{(2xy^2+2xy-y^2+2y-1)^3}$$

and

$$H_2(y) = F_2(x,y)\Big|_{x=1} = \frac{-32y^3(1+y)}{(1-4y-y^2)^3}.$$

Decomposing $H_2(y)$ into partial fractions yields the three relevant terms:

$$\frac{-4(-52+\sqrt{5})^3(\sqrt{(5)}-1)5\sqrt{5}}{(25(y+2-\sqrt{5})^3)} + \frac{(-382+170\sqrt{5})}{(25(y+2-\sqrt{5})^2)} - \frac{18\sqrt{5}}{(125(y+2-\sqrt{5}))}.$$

Hence, defining

$$c_4 = \frac{-4(-52+\sqrt{5})^3(\sqrt{5}-1)5\sqrt{5}}{25(2-\sqrt{5})^3} = 0.44222\ldots$$
$$c_5 = \frac{(-382+170\sqrt{5})}{25(2-\sqrt{5})^2} = -1.34111\ldots$$
$$c_6 = \frac{-18\sqrt{5}}{125(2-\sqrt{5})} = 1.36398\ldots$$

we have

$$[y^m]H_2(y) = \sum_{k\geq 0} k(k-1)f_m(k) = (c_4\frac{(m+2)(m+1)}{2} + c_5(m+1) + c_6)\rho^{-m} + O(1).$$

Finally, we compute the variance from $f_m(k)$.

$$\begin{aligned} Var(f_m(k)) &= \frac{(c_4\frac{(m+2)(m+1)}{2} + c_5(m+1) + c_6) + c_2(m+1) + c_3}{c_1} - \left(\frac{c_2(m+1)+c_3}{c_1}\right)^2 \\ &= c_7 m + O(1) \end{aligned}$$

where

$$c_7 = \frac{(-1404\sqrt{5}+3140)}{(25\sqrt{5}-3)^2(-2+\sqrt{5}))} = 0.16275\ldots.$$

Since $n = 3m$, the variance measured in terms of n is:

$$Var(f_m(k)) = (0.054251\ldots)n + O(1).$$

Two final remarks. First, using the above results, we find, for $\phi = \frac{1+\sqrt{5}}{2}$,

$$\begin{aligned} L_n &= \frac{1}{F_{3m+1}} \sum_\sigma \ln \frac{1}{Pr(\sigma)} - \ln(F_{3m+1}) \\ &= \frac{\sqrt{5}}{\phi}\left(\frac{(c_2(m+1)+c_3)\ln 2}{c_1}\right) + m\ln 3 - n\ln\phi - \ln\frac{1}{\sqrt{5}} - \ln\phi + o(1) \\ &= \left(\frac{\ln 3}{3} + \frac{c_2 \ln 2\sqrt{5}}{3c_1\phi} - \ln\phi\right)n + \frac{\sqrt{5}}{\phi}\left(\frac{(c_2+c_3)}{c_1}\ln 2\right) - \ln\frac{1}{\sqrt{5}} - \ln\phi + o(1) \\ &= (0.012713\ldots)n - 0.1501\ldots + o(1). \end{aligned} \tag{54}$$

Second, let us compute $K_n = \frac{1}{F_{n+1}^2}\sum_\sigma \frac{1}{Pr(\sigma)}$. Since there are $f_m(k)$ σ's which come from Phase I choices with k good transitions, then

$$\sum_\sigma \frac{1}{Pr(\sigma)} = 3^m \sum_k f_m(k)2^k.$$

Table 4
Table of small values of $\mathbb{E}_n$.

n	1	2	3	4	5	6	7
$\mathbb{E}_n$	0	1/2	3/4	9/8	23/16	57/32	135/64

Thus, using the generating function in (53), we find that

$$\sum_k f_m(k)2^k y^m = \frac{1-3y}{1-6y-3y^2} = 1+3y+21y^2+135y^3+\dots.$$

The usual techniques now show that the coefficient of y_m is equal to $\frac{1}{2}\big((3+2\sqrt{3})^m + (3-2\sqrt{3})^m\big)$. Therefore,

$$\begin{aligned} K_n = \frac{3^m}{F_{n+1}^2}\sum_k f_m(k)2^k &= \frac{5}{2\phi^2}\left(\frac{(9+6\sqrt{3})^{1/3}}{\phi^2}\right)^n + o(1). \\ &= (.9549\dots)(1.0262\dots)^n + o(1) \\ &= (.9549\dots)e^{(.0258\dots)n} + o(1). \end{aligned} \tag{55}$$

In particular, $K_{200} \approx 168.6$.

5.4. Still another method for choosing matchings in $B_{1,1}$

Back to the standard Fibonacci graph $B_{1,1}$, we choose edges sequentially starting from u_1 but now when there are two choices, instead of choosing each edge with probability 1/2, we choose the lower edge $\{u_k, v_{k+1}\}$ with probability p. We want to compute K, L and the variance for this model.

We begin with heuristics to motivate a suitable choice of p. Under the uniform distribution on $\mathcal{F}_{1,1}(n)$, the expected number of transpositions (see [7], Proposition 2.4) is

$$\mu_n = \frac{n}{2}\left(1-\frac{1}{\sqrt{5}}\right) + \frac{1-\sqrt{5}}{10} + o(1) = (0.2763\dots)n - 0.1236\dots + o(1). \tag{56}$$

Using our standard sampling method (working from the top), let $\mathbb{E}_n$ be the expected number of transpositions. Direct computation shows the following results, see Table 4.

It is easy to verify that $\mathbb{E}_{n+1} = \frac{1}{2}(1+\mathbb{E}_n+\mathbb{E}_{n-1})$ from which we find

$$\mathbb{E}_n = \frac{1}{9\cdot 2^{n-1}}((3n-2)2^{n-1} + (-1)^n).$$

It follows that

$$\mathbb{E}_n \sim \frac{1}{3}n = (.333\ldots)n \quad \text{while} \quad \mu_n \sim \frac{1}{2}\left(1 - \frac{1}{\sqrt{5}}\right)n = (.2763\ldots)n. \tag{57}$$

This suggests making fewer transpositions!

Let $\mathbb{E}_n(p)$ be the expected number of transpositions if, working in order $u_1, u_2, \ldots,$ at each choice point, a transposition is chosen with probability p. Thus,

$$\mathbb{E}_0(p) = \mathbb{E}_1(p) = 0,\ \mathbb{E}_2(p) = p,\ \mathbb{E}_3(p) = 2p - p^2,\ \mathbb{E}_4(p) = 3p - 2p^2 + p^3.$$

More generally, it is easy to show that

$$\mathbb{E}_n = (1-p)\mathbb{E}_{n-1} + p(1 + \mathbb{E}_{n-2}) \tag{58}$$

from which it follows that

Theorem 9.

$$E(x) := \sum_{n\geq 0} x^n \mathbb{E}_n = \frac{p}{1+p}\left(\frac{x}{(1-x)^2} - \frac{x}{1-(1-p)x-px^2}\right) = px^2 + (2p-p^2)x^3 \ldots$$

and

$$\mathbb{E}_n = \sum_{i=0}^{n-1} (-1)^i (n-i) p^{i+1}.$$

Proof. We define

$$\mathbb{E}'_n = \mathbb{E}_n - \frac{p}{1+p}n.$$

By (58), we have $\mathbb{E}'$ satisfying

$$\mathbb{E}'_n = (1-p)\mathbb{E}'_{n-1} + p\mathbb{E}'_{n-2}.$$

The generating function $E'(x) = \sum_{n\geq 0} \mathbb{E}'_n x^n$ can be easily shown to be

$$E'(x) = -\frac{\frac{p}{1+p}x}{1-(1-p)x-px^2}$$

and therefore

$$\begin{aligned} E(x) &= E'(x) + \sum_{n\geq 0} \frac{pn}{1+p} x^n \\ &= \frac{px}{(1+p)(1-x)^2} - \frac{px}{(1+p)(1-(1-p)x-px^2)}. \quad \square \end{aligned}$$

We content ourselves with the approximation

$$\mathbb{E}_n \approx \frac{pn}{1+p}.$$

It is natural to choose p so that the expected number of transpositions matches the expectation μ_n (from (57)). This gives

$$p = \frac{3-\sqrt{5}}{2} \doteq 0.3820.$$

The calculations below show this is an optimal choice.

We turn next from heuristics to a careful development. Let $q_n(k)$ denote the number of σ which have k transpositions. It is easy to see (by induction) that $q_n(k)$ satisfies the recurrence

$$q_n(k) = q_{n-1}(k) + q_{n-2}(k-1) \tag{59}$$

with $q_0(0) = q_1(0) = q_2(0) = q_2(1) = 1$ and $q_n(k) = 0$ when the indices are out of range. Standard techniques show that the generating function $Q(x, z)$ is given by

$$\begin{aligned} Q(x,z) &= \sum_{n\geq 0} \sum_{0\leq k\leq n/2} q_n(k) x^k z^n \\ &= \frac{1}{1-z-xz^2} = 1 + z + (x+1)z^2 + (2x+1)z^3 \dots . \end{aligned} \tag{60}$$

Also, let $q_n^*(k)$ denote the number of σ which have k transpositions, one of which is at the very bottom (i.e., σ has the edges $\{u_{n-1}, v_n\}$ and $\{u_n, v_{n-1}\}$). It is clear that

$$q_n^*(k) = q_{n-2}(k-1). \tag{61}$$

First we treat $t_n(p) = \sum_{\sigma\in S_n} T(\sigma)$. Recall that $T(\sigma)$ is an unbiased estimate of $|\mathcal{F}_{1,1}(n)| = F_{n+1}$. We will choose the parameter p to minimize the relative variance $T(\sigma)$. Figs. 3 and 4 plot the relative variance as a function of p (note that they are on two different scales). There is a clear minimum and Fact 7 below identifies this as occurring at ϕ^{-2}.

Theorem 10. *The generating function for $t_n(p)$ is given by*

$$G(z) = \sum_{n\geq 0} t_n(p) z^n = \frac{1 - \frac{p}{1-p}z}{1 - \frac{z}{1-p} - \frac{z^2}{p}}. \tag{62}$$

Proof. We first write

$$\sum_{\sigma\in S_n} T(\sigma) = \sum_{\sigma^*} T(\sigma^*) + \sum_{\sigma^{**}} T(\sigma^{**}) \tag{63}$$

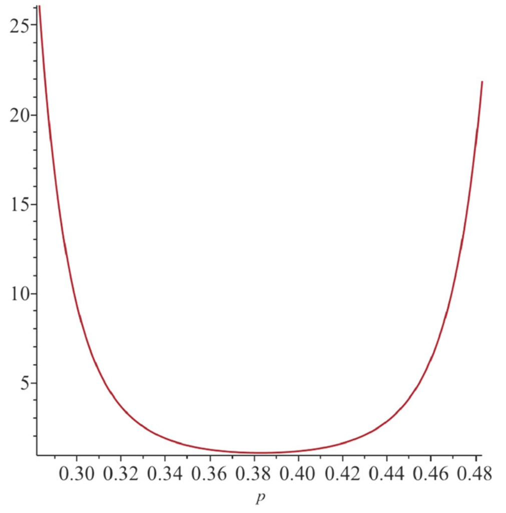

Fig. 3. $\frac{Var(T(\sigma))}{F_{n+1}^2}$ as a function of p for $n = 100$.

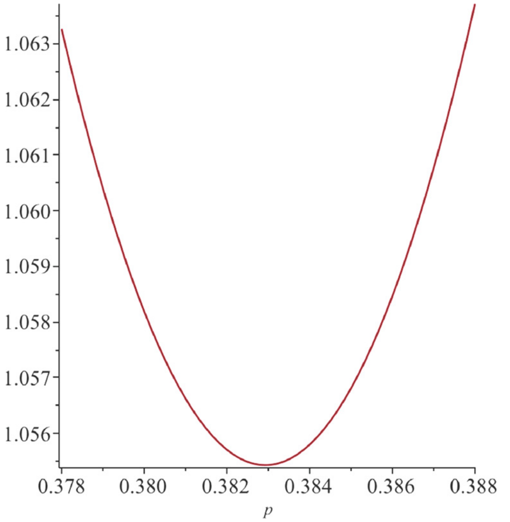

Fig. 4. $\frac{Var(T(\sigma))}{F_{n+1}^2}$ as a function of p for $n = 100$.

where σ^* ranges over all σ which have a transposition on the bottom and σ^{**} ranges over all σ which *do not* have a transposition on the bottom. Thus, by (59) and (61), we have

$$\begin{aligned} t_n(p) = \sum_{\sigma \in S_n} T(\sigma) &= \sum_k q_n^*(k)\big(p^k(1-p)^{n-2k}\big)^{-1} + \sum_k (q_n(k) - q_n^*(k))(p^k(1-p)^{n-1-2k} \\ &= \frac{1}{(1-p)^n} \sum_k q_n^*(k)\Big(\frac{(1-p)^2}{p}\Big)^k + \frac{1}{(1-p)^{n-1}} \sum_k q_{n-1}(k)\Big(\frac{(1-p)^2}{p}\Big)^k \end{aligned}$$

$$= \frac{1}{(1-p)^n}\sum_k q_{n-2}(k-1)\Big(\frac{(1-p)^2}{p}\Big)^k + \frac{1}{(1-p)^{n-1}}\sum_k q_{n-1}(k)\Big(\frac{(1-p)^2}{p}\Big)^k.$$

Thus,

$$\begin{aligned}
G(z) - 1 - z &= \sum_{n\ge 2} t_n(p) z^n \\
&= \sum_{n\ge 2}\Bigg(\frac{1}{(1-p)^n}\sum_k q_{n-2}(k-1)\Big(\frac{(1-p)^2}{p}\Big)^k \\
&\qquad + \frac{1}{(1-p)^{n-1}}\sum_k q_{n-1}(k)\Big(\frac{(1-p)^2}{p}\Big)^k\Bigg) z^n \\
&= \sum_{n\ge 2}\Bigg(\frac{1-p)^2}{p}\sum_k q_{n-2}(k-1)\Big(\frac{(1-p)^2}{p}\Big)^{k-1} \\
&\qquad + (1-p)\sum_k q_{n-1}(k)\Big(\frac{(1-p)^2}{p}\Big)^k\Bigg)\Big(\frac{z}{1-p}\Big)^n \\
&= \frac{z^2}{p} Q\Big(\frac{z}{1-p}\Big) + z\left(Q\Big(\frac{z}{1-p}\Big) - 1\right).
\end{aligned}$$

This leads to

$$G(z) = 1 + \frac{\frac{z^2}{p} + z}{1 - \frac{z}{1-p} - \frac{z^2}{p}} = \frac{1 - \frac{p}{1-p} z}{1 - \frac{z}{1-p} - \frac{z^2}{p}}$$

and Theorem 10 is proved. □

We remark that for the case of $p = 1/2$, Theorem 10 implies that the generating function in this case is

$$G(z) = \frac{1-z}{1-2z-2z^2}$$

which is consistent with the generating function in (8) through a different derivation.

To determine the asymptotic behavior of the coefficient of z^n in $G(z)$, we need to expand $G(z)$ into partial fractions. We first simplify the form of $G(z)$ to

$$G(z) = \frac{p^2 z + p^2 - p}{(1-p)z^2 + pz + p^2 - p}.$$

The roots of the denominator are:

$$r_1(p) = \frac{-p + \sqrt{4p^3 - 7p^2 + 4p}}{2(1-p)}, \qquad r_2(p) = \frac{-p - \sqrt{4p^3 - 7p^2 + 4p}}{2(1-p)}.$$

The corresponding coefficients are

$$c_1(p) = \frac{3p^3 - 4p^2 + 2p - p^2\sqrt{4p^3 - 7p^2 + 4p}}{\sqrt{4p^3 - 7p^2 + 4p}(\sqrt{4p^3 - 7p^2 + 4p} - p)},$$
$$c_2(p) = \frac{3p^3 - 4p^2 + 2p + p^2\sqrt{4p^3 - 7p^2 + 4p}}{\sqrt{4p^3 - 7p^2 + 4p}(\sqrt{4p^3 - 7p^2 + 4p} + p)}.$$

Hence, the coefficient $t_n(p)$ of z^n in the expansion of $G(z)$ is

$$t_n(p) = c_1(p)r_1(p)^{-n} + c_2(p)r_2(p)^{-n} = c_1(p)r_1(p)^{-n} + o(1).$$

When $p = 1/2$, this becomes

$$\begin{aligned} t_n(1/2) = (1/2)\left[\left(\frac{\sqrt{3}-1}{2}\right)^{-n} + \left(\frac{-\sqrt{3}-1}{2}\right)^{-n}\right] &= (1/2)\Big((1+\sqrt{3})^n + (1-\sqrt{3})^n\Big) \\ &= (1/2)\Big((1+\sqrt{3})^n\Big) + o(1) \end{aligned} \tag{64}$$

as we saw in (9).

As a consequence we have

Fact 7. The relative variance of $T_n(\sigma)$ is given by

$$Var_R(T_n(\sigma)) = \frac{1}{F_{n+1}^2}\sum_{\sigma} T_n(\sigma) - \left(\frac{F_{n+1}}{F_{n+1}}\right)^2 = \frac{5c_1}{\phi^2}\frac{1}{(r_1\phi^2)^n} - 1 + o(1) \tag{65}$$

where $c_1(p)$ and $r_1(p)$ are given above.

It is interesting to note that something special happens at the critical value $p = \phi^{-2} = .381966\ldots$ (where, as usual, $\phi = \frac{1+\sqrt{5}}{2}$). In this case, we have

$$c_1(\phi^{-2}) = 1 - \frac{1}{\sqrt{5}}, \qquad r_1(\phi^{-2}) = \phi^{-2},$$

and by (65)

$$Var_R(T_n(\sigma)) = \frac{5c_1}{\phi^2} - 1 = \frac{5-\sqrt{5}}{\phi^2} - 1 = 9 - 4\sqrt{5} = .055728\ldots,$$

independent of n!

Finally, we estimate

$$L_n = \frac{1}{F_{n+1}}\sum_{\sigma} \ln T_n(\sigma) - \ln F_{n+1}. \tag{66}$$

Fact 8.

$$L_n = \left(-\ln(1-p) + \frac{1}{\sqrt{5}\phi}\ln\left(\frac{(1-p)^2}{p}\right) - \ln\phi\right)n$$
$$-\left(\frac{\sqrt{5}-1}{10}\right)\ln\left(\frac{(1-p)^2}{p}\right) + \ln 5^{1/2} + o(1). \tag{67}$$

Note that when $p = \phi^{-2}$, the coefficient of n in (67) is 0 and the constant term is just $\ln 5^{1/2}$. (Hint: If $p = \frac{3-\sqrt{5}}{2} = \frac{1}{\phi^2}$ then $1 - p = \frac{1}{\phi}$.)

Proof. We first simplify the computation by ignoring the distinction as to whether σ has a transposition at the bottom or not. This will have no effect on the asymptotic values we obtain. Thus, if $b(\sigma)$ denotes the number of transpositions that σ has then

$$Pr(\sigma) = p^{b(\sigma)}(1-p)^{n-2b(\sigma)}.$$

Therefore,

$$\ln(T_n(\sigma)) = \ln\left(\frac{1}{Pr(\sigma)}\right) = \ln\left(\frac{1}{(1-p)^n}\left(\frac{(1-p)^2}{p}\right)^{b(\sigma)}\right)$$
$$= \ln(1-p)^{-n} + b(\sigma)\ln\rho$$

where $\rho := \frac{(1-p)^2}{p}$. Hence, with $q_n(k)$ denoting the number of σ which have k transpositions, we have

$$\sum_\sigma \ln T_n(\sigma) = F_{n+1}\ln(1-p)^{-n} + \sum_k kq_n(k)\ln\rho. \tag{68}$$

Since we know the generating function for the $q_n(k)$ from (60), we differentiate it to obtain the generating function $R(x,z)$ for the $kq_n(k)$:

$$R(x,z) = \sum_{n\geq 0}\sum_k kq_n(k)x^kz^n = \frac{z^2}{(1-z-xz^2)^2}.$$

Expanding $R(x,z)$ by the usual techniques, we find the coefficient of z^n is given by

$$\sum_k kq_n(k) = \left(\frac{1}{5}(n+1) - \frac{4}{25-5\sqrt{5}}\right)\phi^n + o(1).$$

Putting these observations together, we have from (66)

$$L_n = \ln(1-p)^{-n} + \frac{1}{5F_{n+1}} n\phi^n \ln\rho - \ln F_{n+1} + o(1)$$
$$= \left(-\ln(1-p) + \frac{1}{\sqrt{5}\phi}\ln\left(\frac{(1-p)^2}{p}\right) - \ln\phi\right) n$$
$$- \left(\frac{\sqrt{5}-1}{10}\right)\ln\left(\frac{(1-p)^2}{p}\right) + \ln 5^{1/2} + o(1)$$

as claimed. This proves Fact 8. □

6. Cycles in $B_{t,1}(n)$

Here we give a proof of Theorem 1. Recall, it is

Theorem 1.

$$\sum_{n=0}^{\infty} f_n z^n = \frac{1}{1 - x_1 z - x_2 z^2 - \ldots - x_t z^t}, \tag{69}$$

with

$$f_n(x_1, x_2, \ldots, x_t) = \sum_{\sigma \in \mathcal{F}_{t,1}(n)} \prod x_i^{a_i(\sigma)}$$

where σ has a_i i-cycles.

Before we give the proof of Theorem 1, we consider an example of the matrix $M = M_8^{(4)}$ as shown below.

$$M_8^{(4)} = \begin{bmatrix} x_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_2 & x_1 & 1 & 0 & 0 & 0 & 0 & 0 \\ x_3 & x_2 & x_1 & 1 & 0 & 0 & 0 & 0 \\ x_4 & x_3 & x_2 & x_1 & 1 & 0 & 0 & 0 \\ 0 & x_4 & x_3 & x_2 & x_1 & 1 & 0 & 0 \\ 0 & 0 & x_4 & x_3 & x_2 & x_1 & 1 & 0 \\ 0 & 0 & 0 & x_4 & x_3 & x_2 & x_1 & 1 \\ 0 & 0 & 0 & 0 & x_4 & x_3 & x_2 & x_1 \end{bmatrix}.$$

The permanent of M is:

$$PerM = x_1^8 + 7x_1^6 x_2 + 6x_1^5 x_3 x_3 + 15x_1^4 x_2^2 + 5x_1^4 x_4 + 20x_1^3 x_2 x_3 + 10x_1^2 x_2^3$$
$$+ 12x_1^2 x_2 x_4 + 6x_1^2 x_3^2 + 12x_1 x_2^2 x_3 + x_2^4 + 6x_1 x_3 x_4 + 3x_2^2 x_4 + 3x_2 x_3^2 + x_4^2.$$

The interpretation is that the bipartite graph $B_{3,1}(8)$ has one perfect matching with eight 1-cycles, seven matchings which have six 1-cycles and one 2-cycle, . . . , twenty matchings with three 1-cycles, one 2-cycle and one 3-cycle, . . . , and finally one matching with four

2-cycles. The structure of the general matrix $M_n^{(t)}$ and why the permanent counts these cycles should be clear.

Proof of Theorem 1. First we define $g_n(a_1, \ldots, a_t)$ to be the number of matchings which contain a_i i-cycles, for $1 \leq i \leq t$. We consider

$$\begin{aligned} F(x_1, x_2, \ldots, x_t, z) &= \sum_{n \geq 0} f_n z^n \\ &= \sum_{\substack{n \geq 0 \\ 0 \leq a_i \leq n-1}} g(a_1, a_2, \ldots, a_t) x_1^{a_1} x_2^{a_2} \ldots x_t^{a_t} z^n. \end{aligned}$$

We note that $g_n(a_1, a_2, \ldots, a_t)$ satisfies the following recurrence, for $n \geq 1$,

$$\begin{aligned} g_n(a_1, a_2, \ldots, a_t) = g_{n-1}(a_1 - 1, a_2, \ldots, a_t) + g_{n-2}(a_1, a_2 - 1, \ldots, a_t) + \ldots \\ + g_{n-t}(a_1 - 1, a_2, \ldots, a_t - 1) \end{aligned} \tag{70}$$

where the base case is $g_0(0, 0, \ldots, 0) = 1$ and $g(a_1, \ldots, a_t) = 0$ if $a_i < 0$ or $a_i \geq n/i$.

Now we sum $g_n(a_1, \ldots, a_t) x_1^{a_1} \ldots x_t^{a_t}$ over all $n \geq t + 1$ and all a_i's. We have

$$\begin{aligned} &F_n(x_1, x_2, \ldots, x_t) - \sum_{\substack{0 \leq n \leq t \\ 0 \leq a_i \leq n-1}} g_n(a_1, a_2, \ldots, a_t) x_1^{a_1} x_2^{a_2} \ldots x_t^{a_t} z^n \\ &= \sum_{\substack{n \geq t+1 \\ 0 \leq a_i \leq n-1}} g_n(a_1, a_2, \ldots, a_t) x_1^{a_1} x_2^{a_2} \ldots x_t^{a_t} z^n \\ &= x_1 z \sum_{\substack{n \geq t+1 \\ 0 \leq a_i \leq n-1}} g_{n-1}(a_1 - 1), a_2 \ldots, a_t) x_1^{a_1 - 1} x_2^{a_2} \ldots x_t^{a_t} z^{n-1} \\ &\quad + x_2 z^2 \sum_{\substack{n \geq t+1 \\ 0 \leq a_i \leq n-1}} g_{n-2}(a_1, a_2 - 1, \ldots, a_t) x_1^{a_1} x_2^{a_2 - 1} \ldots x_t^{a_t} z^{n-2} \\ &\quad + \ldots + x_t z^t \sum_{\substack{n \geq t+1 \\ 0 \leq a_i \leq n-1}} g_{n-t}(a_1, a_2 \ldots, a_t - 1) x_1^{a_1} \ldots x_t^{a_t - 1} z^{n-t} \\ &= \left(x_1 z + x_2 z^2 + \ldots + x_t z^t\right) F_n(x_1, \ldots, x_t, z) - W \end{aligned}$$

where

$$\begin{aligned} W = \sum_{i=1}^{t} x_i z^i \sum_{\substack{0 \leq n \leq t \\ 0 \leq a_i \leq n-1}} & g_{n-i}(a_1, \ldots, a_{i-1}, a_i - 1, a_{i+1}, \ldots, a_t) \\ & \times x_1^{a_1} \ldots x_{i-1}^{a_{i-1}} x_i^{a_i - 1} x_{i+1}^{a_{i+1}} \ldots z^{n-i}. \end{aligned}$$

For $n \le t$ we have

$$W = \sum_{i=1}^{t} \sum_{\substack{0 \le n \le t \\ 0 \le a_i \le n-1}} g_{n-i}(a_1, \ldots, a_{i-1}, a_i - 1, a_{i+1}, \ldots, a_t) x_1^{a_1} \ldots x_{i-1}^{a_{i-1}} x_i^{a_i} x_{i+1}^{a_{i+1}} \ldots z^n$$
$$= \sum_{n=1}^{t} \sum_{\substack{0 \le a_i \le n-1 \\ 1 \le i \le t}} \Big(\sum_{i=1}^{t} g_{n-i}(a_1, \ldots, a_{i-1}, a_i - 1, a_{i+1}, \ldots, a_t) \Big)$$
$$\times x_1^{a_1} \ldots x_{i-1}^{a_{i-1}} x_i^{a_i} x_{i+1}^{a_{i+1}} \ldots x^{a_t} z^n.$$

Together we have

$$\big(1 - x_1 z - x_2 z^2 - \ldots - x_t z^t\big) F_n(x_1, \ldots, x_t, z)$$
$$= 1 + \sum_{n=1}^{t} \sum_{\substack{0 \le a_i \le n-1 \\ 1 \le i \le t}} \Big(g_n(a_1, \ldots, a_t)$$
$$- \sum_{i=1}^{t} g_{n-i}(a_1, \ldots, a_{i-1} a_i - 1, a_{i+1}, \ldots, a_t) \Big) x^{a_1} \ldots x^{a_t} z^n$$
$$= 1.$$

This completes the proof of Theorem 1. □

Remarks. For fixed t and large n, it is natural to conjecture that the joint limiting distribution of the number of i-cycles has a multivariate normal distribution. The limiting means, variances and covariances are available by standard asymptotic analysis from Theorem 1. It should be possible to prove the limiting normality from results in ([17], sec. 9.6) but we have not tried to carry this out. Two special cases: setting $x_2 = x_3 = \ldots = x_t = 0$ gives the generating function for the number of fixed points; setting $x_1 = \ldots = x_t = x$ gives the generating function for the number of cycles.

7. Bringing it all together

7.1. Introduction

One motivation for the present study is to gain insight into the sample size required for accurate estimation when sequential importance sampling is used to estimate the number of perfect matchings in a bipartite graph. We have introduced several families of graphs $B_{t,s}(n)$ giving rise to matchings of the form

$$\mathcal{M}_{t,s}(n) = \{\sigma \in S_n : i - t \le \sigma(i) \le i + s\}.$$

Sequential importance sampling generates random elements of $\mathcal{M}_{t,s}(n)$ with computable probabilities $Pr(\sigma)$. Then $T(\sigma) = \frac{1}{Pr(\sigma)}$ is an unbiased estimator of

$$M = M_{t,s}(n) = |\mathcal{M}_{t,s}(n)|.$$

Two methods for estimating the sample size required for accuracy are

$$N_{var} = \frac{Var(T)}{M_{t,s}^2(n)},$$

and

$$N_{KL} = e^L \quad \text{where} \quad L = \frac{1}{M_{t,s}(n)} \sum_{\sigma \in \mathcal{M}_{t,s}(n)} \ln(Pr(\sigma)^{-1}) - \ln(M_{t,s}(n))$$

with u chosen so that $Pr\{\ln \rho(Y) \geq L + \frac{u}{2}\}$ is small. We have chosen $u = s.d. \ln \rho(Y)$. These are introduced and discussed in Section 2 above.

Further, several sampling schemes are considered:

- generating matchings from 'top down';
- generating matchings (for $B_{1,1}(n)$) from 'top down' but non-uniformly;
- generating matchings (for $B_{1,1}(n)$) in order $(2, 5, 8, \ldots)$;
- generating matchings in random order choosing the steps along the way with non-uniform probabilities.

All of these algorithms (except the last) require sample sizes exponential in n. One of our main findings is that (for these problems), the constants involved are tiny, so that sequential importance sampling can be a much more effective technique than Markov chain Monte Carlo (where available bounds give $O(n^7)$ running time estimates).

This section brings together our findings. For each scenario we present the results in two forms. First, as $N(n) \doteq ae^{bn}$ with a, b given as numerical constants. Second, as $N(200)$. We hasten to add that all our a, b values are the results of previous theorems and available exactly in terms of explicit low-degree polynomials. For example,

$$N_{var,top}(n) \sim \frac{5}{2\phi^2} \left(\frac{1+\sqrt{3}}{\phi^2} \right)^n \approx (.9549\ldots)e^{(0.0426\ldots)n}.$$

We begin with $B_{1,1}(n)$ since results are most complete here.

7.2. Fibonacci permutations

As explained in the introduction,

$$\mathcal{M}_{1,1}(n) = \{\sigma \in S_n : |\sigma(i) - 1| \leq 1\}$$

has $|\mathcal{M}_{1,1}(n))| = F_{n+1}$ with $F_n = 1, 1, 2, 3, 5, 8, \dots$ for $n = 1, 2, 3, 4, 5, 6, \dots$. When $n = 200, F_{n+1} = 4.539\dots \times 10^{41}$. Then

- $N_{var,top}(n) \doteq (.9549\dots)e^{(0.0426\dots)n}, \qquad N_{var,top}(200) \doteq 4,788$

(see the Example in Section 2). Also,

- $N_{KL,top}(n) \doteq e^{(0.0204\dots)n+(0.2989\dots)\sqrt{n}}, \qquad N_{KL,top}(200) \doteq 4,058.$

This is proved in Section 3. As explained there, the $O(\sqrt{n})$ term is required for L to be concentrated about its mean. This is guaranteed by our sharp estimates of $Var(L)$. We have neglected these $O(\sqrt{n})$ term for these numerics for the next example because it is not available.

- $N_{Breg}(n) \doteq (0.605\dots)e^{(3.301\dots)n}, \qquad N_{Breg}(200) \doteq 1.004 \times 10^{10}.$

The Bregman bound, derived in [5], given at the end of Section 2 is an upper bound for e^L based on a random order. Proof of concentration remains an open problem.

- $N_{var,3}(n) \doteq (0.9544\dots)e^{(0.02586\dots)n}, \qquad N_{var,3}(200) \doteq 168.$

These bounds follow from (49).

- $N_{KL,3}(n) \doteq (2.1295\dots)e^{(0.012\dots)n+(0.2329\dots)\sqrt{n}}, \qquad N_{KL,3}(200) \doteq 728.$

See Section 5 for details.

- $N_{var,ran}(n) \doteq e^{(0.0265\dots)n}, \qquad N_{var,ran}(200) \doteq 200.$

This and the following bounds for sampling in random order are derived in [8].

- $N_{KL,ran}(n) \doteq (2.2361\dots)e^{(0.0101\dots)n+(0.1396\dots)\sqrt{n}}, \qquad N_{KL,ran}(200) \doteq 122.$

Remarks. For this example, we see that bounds on the required sample size based on the variance can be substantial over-estimates (but not always, as for $N_{var,3}$ and $N_{KL,ran}$). The deterministic ‘from top down’, random and ‘every third’ orders are roughly comparable with the latter two slightly better. The Bregman bound, elegant and general though it may be, is useless in practice. Even worse, the celebrated FPRAS for the Markov chain Monte Carlo procedure gives an $\Omega(n^7)$ algorithm. When $n = 200$, this gives $N = (1.280\dots) \times 10^{16}$.

7.3. (2, 2)*-permutations*

As in Section 4, let

$$\mathcal{M}_{2,2}(n) = \{\sigma \in S_n : |\sigma(i) - i| \leq 2,\ 1 \leq i \leq n\}.$$

Then

$$M_n = |\mathcal{M}_{2,2}(n)| \sim c_1\rho^{-n} \quad \text{with} \quad c_1 = \frac{1-\rho}{2+6\rho-5\rho^2} = 0.45464\ldots$$

and $\rho = 0.4285\ldots$ is the root of minimum modulus of the polynomial $p(z) = 1 - 2z - 2z^2 + z^5$. When $n = 200$, $M_n = 1.851\ldots \times 10^{73}$.

Using results from Section 4 we have

- $N_{var,top}(n) \doteq (0.73856\ldots)e^{(0.11067\ldots)n}, \quad N_{var,top}(200) \doteq 3.0273 \times 10^9;$
- $N_{KL,top}(n) \doteq (0.8608...)e^{(0.05786...)n+(0.3387...)\sqrt{n}} \quad N_{KL,top}(200) \doteq 1.0985... \times 10^7;$
- $N_{Breg}(n) \doteq (0.7725\ldots)e^{0.01101\ldots)n}, \quad N_{Breg}(200) \doteq 2.828 \times 10^9.$

Again, the sample size estimates leaning on the variance and Bregman's inequality are over-estimates.

8. Concluding remarks

In principle, the same techniques can work for matrices with s diagonals above the diagonal and t diagonals below the diagonal. For example, suppose $s = 3$, $t = 4$. Let $B_{4,3}(10)$ be the resulting bipartite graph and consider the corresponding matrix $M = M_{4,3}(10)$ shown below. The form for the matrix $M_{t,s}(n)$ in the general case should be clear.

$$M_{4,3}(10) = \begin{bmatrix} x_4 & x_4 & x_4 & x_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_4 & x_4 & x_4 & x_4 & x_4 & 0 & 0 & 0 & 0 & 0 \\ x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & 0 & 0 & 0 & 0 \\ x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & 0 & 0 & 0 \\ x_1 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & 0 & 0 \\ 0 & x_1 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & 0 \\ 0 & 0 & x_1 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 & x_4 \\ 0 & 0 & 0 & x_1 & x_3 & x_3 & x_3 & x_3 & x_3 & x_3 \\ 0 & 0 & 0 & 0 & x_1 & x_2 & x_2 & x_2 & x_2 & x_2 \\ 0 & 0 & 0 & 0 & 0 & x_1 & x_1 & x_1 & x_1 & x_1 \end{bmatrix}.$$

Then

$$Per M_{4,3}(10) = 6x_1^4x_2x_3x_4^4 + \ldots + 520x_1^3x_2x_4^6 + \ldots + 24x_1^4x_3x_4x_4^5.$$

Thus, there are 6 matchings that occur with probability $(1/1)^4(1/2)(1/3)(1/4)^4$, 520 matchings that occur with probability $(1/1)^3(1/2)(1/4)^6$, etc.

To see why this is so, we first note that in general, because of the order that the edges are chosen, there are (usually) exactly 4 choices for each vertex u_k. However, if it happens that at this time, v_{k-3} is unoccupied, then the edge $\{u_k, v_{k-3}\}$ must be chosen. This observation accounts for the diagonal of x_1's. However, when k gets near the end, there are fewer choices. For example, there is only one choice for u_8. Similarly, there are only 2 choices for u_7 (which accounts for the appearance of the x_2's in row 7), except that if v_4 happened to be unoccupied then in which case there is only one choice. Similar arguments apply to the occurrences of the other x_k's in the matrix, and in fact, to the general case with arbitrary s and t. In principle, our techniques could then be used to find the appropriate recurrences, generating functions, asymptotic expansions, etc. However, even finding a general expression for the permanent of the matrix corresponding to the graph $B_{t,t}(n)$ seems formidable! For example, for $s = t = 6$, the corresponding numbers satisfy a recurrence of order 494 (see [15]). However, it is possible that with (a lot) more work, progress can be made. Be our guest!

Acknowledgment

The authors wish to acknowledge the helpful comments of a referee which increased the readability of the paper.

References

[1] R. Bapat, Permanents in Probability and Statistics, Linear Algebra and Its Appl., vol. 127, Elsevier, NY, 1990.
[2] A. Barvinok, Combinatorics and Complexity of Partition Functions, Algorithms and Combinatorics, vol. 30, Springer, NY, 2016, vi+303 pp.
[3] S. Chatterjee, P. Diaconis, The sample size required in importance sampling, Ann. Appl. Probab. 28 (2018) 1099–1135.
[4] F. Chung, P. Diaconis, R.L. Graham, C. Mallows, On the permanents of complements of the direct sum of identity matrices, Adv. in Appl. Math. 2 (1981) 121–137.
[5] P. Diaconis, Sequential importance sampling for estimating the number of perfect matchings in bipartite graphs: an ongoing conversation with Laci, preprint, 2018.
[6] P. Diaconis, J. Fulman, R. Guralnick, On fixed points of permutations, J. Algebraic Combin. 28 (2007) 189–218.
[7] P. Diaconis, R.L. Graham, S.P. Holmes, in: State of the Art in Probability and Statistics, Leiden, 1999, in: IMS Lecture Notes Monogr. Ser., vol. 36, Inst. Math. Statist., Beachwood, OH, 2001, pp. 195–222.
[8] P. Diaconis, B. Kolesnik, Analysis of some randomized algorithms for computing permanents, Stanford Univ., preprint, 2018.
[9] M. Dyer, M. Jerrum, H. Muller, On the switch Markov chain for perfect matchings, J. ACM (ISSN 0004-5411) 64 (2) (2017) 12.
[10] P. Flajolet, R. Sedgewick, Analytic Combinatorics, Cambridge Univ. Press, New York, 2009, xiii+810 pp.
[11] P. Hanlon, A random walk on the rook placements on a Ferrers board, in: The Foata Festschrift, Electron. J. Combin. 3 (2) (1996), Research Paper 26, approx. 24 pp.
[12] M. Jerrum, A. Sinclair, E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, J. ACM 51 (4) (2004) 671–697.

[13] J. Liu, Monte Carlo Strategies in Scientific Computing, Springer Series in Computing, Springer, New York, 2008, xx+357 pp.
[14] L. Lovász, M. Plummer, Matching Theory, Corrected reprint of the 1986 original MR0859549, AMS Chelsea Publishing, Providence, RI, 2009, xxxiv+554 pp.
[15] The Online Encyclopedia of Integer Sequences, Entry a002524.
[16] E. Ozel, The number of k-cycles in a family of restricted permutations, arXiv:1710.07885, 2017.
[17] R. Pemantle, M.C. Wilson, Analytic Combinatorics in Several Variables, Cambridge Univ. Press, New York, 2013, xii+380 pp.
[18] J. Riordan, An Introduction to Combinatorial Analysis, Reprint of the 1958 original [Wiley, New York; MR0096594], Dover Publications, Mineola, NY, 2002, xii+244 pp.
[19] R. Rubenstein, D. Kroese, The Cross Entropy Method, a Unified Approach to Combinatorial Optimization, Monte Carlo Simulation and Machine Learning, Springer, NY, 2004.
[20] L. Valiant, The complexity of enumeration and reliability problems, SIAM J. Comput. 8 (1979) 410–421.
[21] P. Vontobel, The Bethe permanent of a nonnegative matrix, IEEE Trans. Inform. Theory 59 (3) (2013) 1866–1901.
[22] D.A. Wolfram, Solving generalized Fibonacci recurrences, Fibonacci Quart. 36 (1998) 129–145.