Uniform sampling modulo a group of symmetries using Markov chain simulation

Mark Jerrum

Abstract. Let Σ be a finite alphabet, and G a permutation group of degree n. The group G induces a natural action on Σ^n by permutation of the positions of symbols. This action partitions Σ^n into orbits, i.e., subsets of Σ^n which are equivalent modulo the symmetry group G. The problem addressed is that of uniform sampling from the set of all orbits. This problem is closely allied to that of estimating the cycle index polynomial of G, and can be regarded as part of the wider question "to what extent can Pólya theory be automated?" The approach taken here is to simulate a Markov chain on Σ^n that converges to a uniform distribution on the orbits. The Markov chain may be rapidly mixing for all G, but positive results to date are limited to some very simple special cases. The general question is likely to be difficult to resolve, because it includes the mixing rate of the Swendsen-Wang process as a special case.

1. Problem description

Let Σ be a finite alphabet of cardinality k, and G be a permutation group on $[n] = \{0, \dots, n-1\}$. The group G has a natural action on the set Σ^n of all words of length n over the alphabet Σ, under which the word $x = x_0 x_1 \dots x_{n-1}$ is mapped by the permutation $g \in G$ to the word $x^g = y_0 y_1 \dots y_{n-1}$ defined by $y_j = x_i$ for all $i, j \in [n]$ satisfying $i^g = j$.[1] The action of G partitions Σ^n into a number of *orbits*, these being the equivalence classes of Σ^n under the equivalence relation that identifies x and y whenever there exists $g \in G$ mapping x to y.

By way of example, let $\Sigma = \{0, 1\}$, and $n = m^2$ for some integer m. Interpret each element of Σ^n as the adjacency matrix of an m-vertex directed graph Γ. Let G be the permutation group of of degree n and order $m!$ whose elements correspond to permutations of the vertex set of Γ, i.e., to simultaneous permutations of the rows and columns of the adjacency matrix. Then the orbits of Σ^n under the action of G correspond naturally to unlabelled directed graphs on m vertices. Many other sets of unlabelled combinatorial structures can be obtained using a like construction.

1991 *Mathematics Subject Classification.* Primary 68Q25; Secondary 60J15, 60J20, 20B40.

The author was supported in part by grant GR/F 90363 of the UK Science and Engineering Research Council, and by Esprit Working Group "RAND."

A preliminary version of the paper was presented at the DIMACS *Workshop on Expanding Graphs and Applications*, Princeton NJ, USA, May 1992.

[1] Note that the permutation g is considered to act on *positions* rather than *indices*, since this is perhaps the easier to grasp of the two possible conventions.

What has just been described is the setting for *Pólya theory*, the key result of which is an expression for the number of orbits in terms of the *cycle index polynomial* of G [2]. This polynomial, in the variables $z_1, z_2, \ldots, z_n$, is defined to be

$$P_G(z_1, \ldots, z_n) = |G|^{-1} \sum_{g \in G} {z_1}^{c_1(g)} \ldots {z_n}^{c_n(g)},$$

where c_i denotes the number of cycles in g of length i. The key result just referred to is that the number of orbits of Σ^n under the action of G is P_G evaluated at the point $z_1 = \cdots = z_n = k$. For many important choices for G, this computation is feasible and leads to results concerning the number of unlabelled combinatorial structures of various kinds [8].

For general groups G it seems that P_G is hard to compute exactly. Indeed, Goldberg has shown that evaluating $P_G(2, \ldots, 2)$ is #P-complete,[2] even when G is an abelian 2-group [7]. (See also a related result of Lubiw [19].) However, the question of whether there is an efficient *approximation* algorithm for P_G remains open. This is one of a group of three related questions which form the main subject of this article. They are as follows.

(a) Is there a fully polynomial randomised approximation scheme (fpras) [16] for estimating $P_G(k, \ldots, k)$? That is to say, is there a randomised algorithm that takes as input a group G and $\varepsilon > 0$, and produces as output a number Y (a random variable) such that

$$\Pr\big((1-\varepsilon)P_G(k, \ldots, k) \le Y \le (1+\varepsilon)P_G(k, \ldots, k)\big) \ge \tfrac{3}{4},$$

and, moreover, does so within time $\mathrm{poly}(n, \varepsilon^{-1})$?

(b) Is there a polynomial-time almost uniform sampler[3] [15] for the orbits of Σ^n under the action of G? That is to say, is there a randomised algorithm that takes as input a group G and $\varepsilon > 0$, and produces as output a word $Y \in \Sigma^n$ (a random variable) such that for each orbit Δ,

$$(1-\varepsilon)N^{-1} \le \Pr(Y \in \Delta) \le (1+\varepsilon)N^{-1},$$

where $N = P_G(k, \ldots, k)$ is the total number of orbits? The execution time is required to be bounded by $\mathrm{poly}(n, \varepsilon^{-1})$.

(c) Is there a polynomial-time "almost-w" sampler for G, where the weight function $w : G \to \mathbf{Z}$ is defined by $w(g) = k^{c(g)}$, and $c(g)$ denotes the number of cycles in g. That is to say, is there a randomised algorithm that takes as input a group G and $\varepsilon > 0$, and produces as output a permutation $Y \in G$ (a random variable) such that for each $g \in G$,

$$(1-\varepsilon)w(g)Z^{-1} \le \Pr(Y = g) \le (1+\varepsilon)w(g)Z^{-1},$$

where $Z = |G|\, P_G(k, \ldots, k)$? Again, the execution time is required to be bounded by $\mathrm{poly}(n, \varepsilon^{-1})$.

---

[2] The class #P is the analogue for counting problems of the more familiar class NP which arises in the classification of decision problems. A #P-complete problem is computationally as difficult as counting the number of satisfying assignments to a Boolean formula, or the number of accepting computations of a non-deterministic Turing machine.

[3] When this concept was first introduced, "generator" was used in place of "sampler," but the latter word is more specific.

It may be observed that we have chosen to use the degree n as a simple measure of the "input size" of the permutation group G, and the reader may question whether this is reasonable. The convention may be justified by noting, as Sims did [22], that every permutation group can be specified by *strong generating set* of $\mathrm{O}(n^2)$ permutations.[4] Thus, any permutation group may be described using a number of bits which is polynomial in n. Furthermore, Furst, Hopcroft, and Luks demonstrated that many permutation group operations — of which the most basic is membership testing — can be performed in time polynomial in n, given a strong generating set for the group in question [6]. These observations confirm that the degree is a reasonable measure of the "size" of a permutation group in the current context. While we are on the subject, it should be noted also that the alphabet Σ, though arbitrary, is regarded as fixed, and does not form part of the problem instance.

It is worth recording at this point the curious fact that although question (a) as stated is open, the answer is known to be "no" (assuming $\mathrm{RP} \neq \mathrm{NP}$) if k is allowed to be fractional [7]. It remains an intriguing possibility that the particular combinatorial significance of the cycle index polynomial when k is integral may provide a path to resolving question (a) in the affirmative, while avoiding the problems that arise in the case of fractional k.

The complexity of approximate counting and of almost uniform sampling are known to be closely related,[5] which would lead one to suppose that questions (a), (b), and (c) ought to be equivalent. However, the situation here is atypical, and it is not clear, for example, whether resolving question (a) in the affirmative would immediately settle either of the others. The two known entailments are described in the following proposition, whose proof — using techniques standard in the area — is relegated to the Appendix.

PROPOSITION 1.1. *An affirmative answer to question* (c) *would entail affirmative answers to* (a) *and* (b).

This article will address questions (b) and (c), and hence indirectly (a), using the by now familiar technique of Markov chain simulation.

The remaining sections are organised as follows. Section 2 introduces two related Markov chains with state spaces Σ^n and G whose stationary distributions are precisely the output distributions specified in questions (b) and (c), respectively. These Markov chains furnish affirmative answers to their respective questions, provided the chains are rapidly mixing, that is to say, converge rapidly to their stationary distributions. Section 3 discusses techniques for analysing (non-asymptotic) rates of convergence of combinatorially defined Markov chains, and demonstrates rapid mixing of some easy special cases, namely when G is the full symmetric group or a cyclic group.

Finally, Section 4 presents a technique for encoding instances of a ferromagnetic Ising system as a permutation group in such a way that the distribution on permutations specified in question (c) corresponds to the Gibbs distribution on

[4] Indeed, $\mathrm{O}(n)$ permutations suffice [11].

[5] A rather precise statement of this relationship has been formulated by Jerrum, Valiant, and Vazirani [15].

Ising configurations. This correspondence yields an alternative development of the Swendsen-Wang process, which is a popular sampling procedure for Ising configurations according to the Gibbs distribution. At the same time, the correspondence indicates the likely difficulty in establishing the rapid mixing property for more general classes of permutation groups.

2. A Markov chain with the appropriate distribution

Our approach to sampling orbits is to simulate an appropriately defined Markov chain. This technique has proved fruitful on a number of occasions in recent years; previous applications include an algorithm of Jerrum and Sinclair for estimating the permanent of a 0,1-matrix [1, 12, 21], and one of Dyer, Frieze, and Kannan for estimating the volume of a convex body in n-dimensional space [5]. In this instance we wish to construct a Markov chain $\mathfrak{M} = \mathfrak{M}(G, \Sigma)$ whose state space is Σ^n and whose stationary distribution assigns equal probability to each orbit. In fact, we shall aim at something stronger, namely, a stationary distribution that assigns to each word $x \in \Sigma^n$ a probability inversely proportional to the size of the orbit x^G containing x.

The transition probabilities from a state $x \in \Sigma^n$ are specified by the following conceptually simple two-step experiment.

(1) Choose g uniformly at random (u.a.r.) from the point stabiliser $G_x = \{g \in G : x^g = x\}$.
(2) Choose y u.a.r. from the set $\{y \in \Sigma^n : y^g = y\}$.

The new state is y. Before analysing the stationary distribution of $\mathfrak{M}$, we should pause to consider the computational complexity of implementing the above experiment. Step (2) is computationally undemanding, and amounts to assigning, u.a.r. and independently, a symbol from Σ to each cycle of g. Step (1) is more interesting, and is equivalent (under randomised polynomial-time reducibility) to computing a setwise stabiliser in a permutation group, a task that includes deciding isomorphism of two graphs as a special case. The computational complexity of the setwise stabiliser problem is open: it is one of the very rare natural candidates for a problem that is in the class NP, but is neither in P nor NP-complete.

Although no general polynomial-time algorithm for implementing step (1) is known, and it is perfectly possible that none exists, there are significant classes of groups G for which step (1) does have an efficient implementation. Luks has shown that p-groups — groups in which every element has order a power of p for some prime p — is an example of such a class [20]. We shall see in Section 4 that the Markov chain $\mathfrak{M}$ is already interesting in the context of abelian 2-groups.

Returning to the Markov chain itself, we note immediately that $\mathfrak{M}$ is ergodic, since every state can be reached from every other in a single transition, by selecting the identity permutation in step (1). The easiest way to get at the stationary distribution is perhaps by considering a random walk on the bipartite graph B that has vertex bipartition (Σ^n, G) and edge set $\{(x, g) : x^g = x\}$. It is clear that the Markov chain $\mathfrak{M}$ can be viewed as sampling the random walk on B after every even step. Let $\pi : \Sigma^n \to [0, 1]$ denote the stationary distribution of $\mathfrak{M}$. Then $\pi(x)$ is proportional to the degree of vertex x in the graph B, which is $|G_x|$. It is an elementary group-theoretic fact that $|G_x| \times |x^G| = |G|$, and hence $\pi(x)$ is inversely proportional to $|x^G|$. We have therefore established:

THEOREM 2.1. *Let π be the stationary distribution of Markov chain $\mathfrak{M}$. Then $\pi(x) = |x^G|^{-1} P_G(k, \dots, k)^{-1}$ for all $x \in \Sigma^n$; in particular, π assigns equal probability to each orbit x^G.* $\square$

Dually we might consider the Markov chain $\mathfrak{M}'$ with state space G and transition probabilities modelled by an experiment in which steps (1) and (2) appear transposed. By relating $\mathfrak{M}'$ to the random walk on B we easily obtain:

THEOREM 2.2. *Let π' be the stationary distribution of Markov chain $\mathfrak{M}'$. Then $\pi'(g) = k^{c(g)} |G|^{-1} P_G(k, \dots, k)^{-1}$ for all $g \in G$, where $c(g)$ denotes the number of cycles in the permutation g.* $\square$

Note that the stationary distributions of the Markov chains $\mathfrak{M}$ and $\mathfrak{M}'$ match the sampling distributions specified in questions (b) and (c) of the previous section.

3. The mixing rate

We have seen that the stationary distribution of the Markov chain $\mathfrak{M}$ is appropriate for sampling u.a.r. from the set of orbits of the G-action on Σ^n. For the Markov chain simulation approach to be computationally efficient, it is necessary for $\mathfrak{M}$ to be *rapidly mixing*. In the theoretical computer science tradition, we give a precise meaning to this informal requirement by insisting that $\mathfrak{M}$ should be "close" to stationarity after a number of steps that is bounded by a polynomial in n. Since the size of the state space is exponential in n, this is a non-trivial requirement.

There are a number of ways of quantifying "closeness" to stationarity, but they are all essentially equivalent in this application. Consider an ergodic Markov chain with state space Ω. Let $x \in \Omega$ be an arbitrary state, and denote by $P^t(x, \cdot\,)$ the distribution of the state at time t given that x is the initial state. Let π be the stationary distribution of $\mathfrak{M}$. Then the *variation distance* at time t with respect to the initial state x is defined to be

$$\Delta_x(t) = \max_{S \subseteq \Omega} \left| P^t(x, S) - \pi(S) \right| = \tfrac{1}{2} \sum_{y \in \Omega} \left| P^t(x, y) - \pi(y) \right|.$$

The rate of convergence may be measured by the function

$$\tau_x(\varepsilon) = \min\{t : \Delta_x(t') \le \varepsilon \text{ for all } t' \ge t\}.$$

The Markov chain simulation technique will yield an efficient almost uniform sampler for orbits of the G-action on Σ^n provided the Markov chain $\mathfrak{M}(G, \Sigma)$ mixes rapidly. In recent years, techniques have been developed for bounding the mixing rate of combinatorially defined Markov chains, using, among other ideas, the relation between mixing rate and the expansion properties of the Markov chain viewed as a graph. Sinclair has provided a useful survey of these techniques, in addition to presenting some sharpened bounds [23]. Nevertheless, proofs of rapid mixing still tend to be technically involved.

Since no counterexamples have been identified, it remains a possibility that for any fixed alphabet Σ, the Markov chain $\mathfrak{M}(G, \Sigma)$ is rapidly mixing for all choices of G; specifically, that there is a *fixed* polynomial (in n and $\log \varepsilon^{-1}$) that uniformly bounds $\tau_x(\varepsilon)$ for all possible groups G. If this were so — and the author would not hazard a conjecture on this point — the question (b) posed in Section 1 would answered in the affirmative, at least for groups for which the Markov chain could be efficiently simulated. The following section provides evidence that even if a general result of this form is true, it is likely to be difficult to prove. A more realistic programme for the short term is to produce a catalogue of families of groups G for which rapid mixing of $\mathfrak{M}(G, \Sigma)$ can be rigorously demonstrated. The following two results represent a modest start on this programme. The second answers a question of Kannan. Their proofs are fairly routine, and are held over to the Appendix.

PROPOSITION 3.1. *Let S_n denote the symmetric group on $[n]$, and $\mathfrak{M} = \mathfrak{M}(S_n, \{0,1\}^n)$ the derived Markov chain as defined in Section 2. The convergence time of $\mathfrak{M}$ is*

$$\tau_x(\varepsilon) = \mathrm{O}(\sqrt{n} \log \varepsilon^{-1}),$$

uniformly over the choice of initial state $x \in \{0,1\}^n$.

REMARK 3.2. *A polynomial bound on $\tau_x(\varepsilon)$ can be derived for alphabets Σ other than binary, but the degree of the polynomial rises with the size of Σ.* $\square$

PROPOSITION 3.3. *Let G denote any cyclic permutation group on $[n]$, and $\mathfrak{M}' = \mathfrak{M}'(G, \Sigma)$ the derived Markov chain as defined in Section 2. The convergence time of $\mathfrak{M}'$ is*

$$\tau_g(\varepsilon) = \mathrm{O}(n \log n \log \varepsilon^{-1}),$$

uniformly over the choice of initial state $g \in G$.

4. Encoding a ferromagnetic Ising system

The *Ising model* has been the focus of much attention in the physics and mathematics communities since it was first introduced by Lenz [17] and Ising [9] in the early 1920s. A detailed account of the Ising model and its history will not be presented here; a very readable survey is given by Cipra [3], while Welsh [25] considers the Ising model from a complexity-theoretic perspective and sets it in the context of other combinatorial problems in statistical physics.

Our aim in this section to to exhibit a reduction from sampling configurations of a ferromagnetic Ising system to sampling binary words modulo an appropriate group of symmetries. The fact that the group in question is an abelian 2-group suggests that the mixing rate of $\mathfrak{M}(G, \Sigma)$ may be of substantial interest, even for restricted classes of groups. An unexpected feature of the reduction is that the dual Markov chain $\mathfrak{M}'$ which results can be directly identified with the Swendsen-Wang process, a recently developed and ingenious approach to Monte Carlo simulation of the Ising model. Although experiments seem to suggest that the mixing rate of the Swendsen-Wang process is rapid, no rigorous bounds have so far been obtained. This failure provides a salutary warning that the programme outlined in the previous section may be a difficult one. If nothing else, the reduction which follows offers an interesting (if quirky) alternative view of the Swendsen-Wang process.

Let $\Gamma = (V, E)$ be an undirected graph defining an Ising system with sites V and bonds $E \subseteq V^{(2)}$, and let J be the interaction energy, assumed positive (i.e., the system is ferromagnetic) and constant over bonds. We assume throughout that the applied field is zero. We construct a permutation group G over the set $X = \bigcup_{e\in E} X_e$, which is the disjoint union of two-element sets X_e. For $e \in E$ and $v \in V$, denote by h_e the permutation that transposes the elements of X_e and leaves everything else fixed, and denote by g_v the generator $\prod_{e\ni v} h_e$. Finally, define $G = G(\Gamma) = \langle g_v : v \in V\rangle$.

Observe that the generators of G commute and have order two, so each permutation $g \in G$ can be expressed as

$$g = g(\sigma) = \prod_{v\in V} {g_v}^{[1-\sigma(v)]/2} = \prod_{e=\{u,v\}\in E} {h_e}^{[1-\sigma(u)\sigma(v)]/2}, \tag{4.1}$$

where $\sigma : V \to \{-1, +1\}$. (Note that all the exponents are either 0 or 1.) Provided the graph G is connected, expression (4.1) is essentially canonical, in that σ is uniquely determined modulo uniform inversion of signs. We shall resolve this remaining ambiguity by imposing the condition $\sigma(v_0) = +1$ for some distinguished vertex v_0.

The function σ can be interpreted as a normalised spin configuration of the Ising system, where the convention for labelling spins by ± 1 is chosen so that $\sigma(v_0) = +1$. (Since the applied field is zero, there is nothing to choose between the two possible labellings.) Note that the function mapping g to σ is a bijection between G and the set of normalised spin configurations.

Observe that, from expression (4.1), the number of cycles in a permutation $g = g(\sigma)$ may be written

$$c(g) = \sum_{e=\{u,v\}\in E} [3 + \sigma(u)\sigma(v)]/2.$$

The cycle index evaluation at $k = 2$ is then

$$\begin{aligned} P_G(2,\dots,2) &= |G|^{-1} \sum_{\sigma:\sigma(v_0)=+1} 2^{c(g(\sigma))} \\ &= \tfrac{1}{2}|G|^{-1} \sum_{\sigma} \prod_{e=\{u,v\}} \exp_2\left\{[3+\sigma(u)\sigma(v)]/2\right\} \\ &= \tfrac{1}{2}|G|^{-1}(2\sqrt{2})^{|E|} \sum_{\sigma} \prod_{e=\{u,v\}} (\sqrt{2})^{\sigma(u)\sigma(v)}, \end{aligned}$$

which, ignoring the easily computed scaling factor $2\,|G|\,(2\sqrt{2})^{-|E|}$, is exactly the partition function of the Ising system with underlying graph Γ, and $\beta J = \ln\sqrt{2}$, where β is "inverse temperature" and J the interaction energy.

Using more exotic constructions for the group G it would be possible to obtain the partition function of the Ising system at values of βJ different from $\ln\sqrt{2}$. However, is is easier to use a two-stage reduction, in which we first reduce to an Ising system with $\beta J = \ln\sqrt{2}$, and then apply the construction described above. The first stage of the reduction is straightforward, as a weak bond can be simulated by a chain of stronger bonds placed in series, and a strong bond by a bundle of weaker bonds placed in parallel.[6] By combining these two constructions it is possible to simulate general interactions economically and with high precision.

It should be observed that there is a direct correspondence between terms of the cycle index expansion, and terms of the Ising partition function. As a result, it is possible to sample an Ising configuration from the appropriate (i.e., Gibbs) distribution by sampling a permutation $g \in G$ from the stationary distribution of the dual Markov chain $\mathfrak{M}'(G(\Gamma), \{0,1\})$, and interpreting g as an spin configuration as in (4.1). If the Markov chain were rapidly mixing, this technique would provide a direct and efficient procedure for sampling spin configurations. At present, the only known approach to constructing a polynomial-time sampler for spin configurations is very indirect, and based on the algorithm of Jerrum and Sinclair for estimating the partition function [13, 14].

Looking more closely into the Markov chain $\mathfrak{M}'(G(\Gamma), \{0,1\})$, we discover that it does not represent an entirely new approach to Monte Carlo simulation of the Ising model, but merely an alternative development of an existing approach. The *Swendsen-Wang process* is a Markov chain on spin configurations in which the transition probabilities from a configuration $\sigma : V \to \{-1, +1\}$ are modelled by the following two step experiment.

(1) Form a set $A \subseteq E$ using a sequence of $|E|$ trials. Initially, set $A = \emptyset$; then, for each edge (bond) $e = \{u, v\} \in E$ with $\sigma(u) = \sigma(v)$, add e to A with probability $1 - e^{-2\beta J}$.
(2) Compute the connected components of the subgraph (V, A) of Γ. For each connected component, choose a spin independently and u.a.r. from $\{-1, +1\}$ and assign the chosen spin to all vertices in the component.

It is not too difficult to show that the transition probabilities defined by this experiment are identical to those of the dual Markov chain $\mathfrak{M}'(G(\Gamma), \{0,1\})$. In fact the correspondence is even closer, in that steps (1) and (2) can be identified with the two steps that make up a single transition of $\mathfrak{M}'$. (Thus step (1) above corresponds to step (2) of Markov chain $\mathfrak{M}$ defined in Section 2, and vice versa.) To bring out this correspondence it is only necessary to explain the connection between the subgraphs $A \subseteq E$ and words $y \in \Sigma^n = \{0,1\}^n$ which arise as "auxiliary states" in the Swendsen-Wang process and in $\mathfrak{M}'$, respectively. (Here, $n = |X| = 2|E|$.) With each auxiliary state $y : X \to \{0,1\}$ of $\mathfrak{M}'$, associate an auxiliary state

$$A = A(y) = \{e \in E : y \text{ is not constant on } X_e\}.$$

of the Swendsen-Wang process. This rule naturally partitions Σ^n into equivalence classes and associates with each equivalence class an edge set $A \subseteq E$. Under this correspondence, the individual steps (two steps per transition) of the respective processes can be compared. It is then a matter of straightforward computation to show that the two processes are identical at the level of steps as well as transitions.

Appendix

Here we gather together proofs of the various results stated in previous sections.

[6] This is a special case of the *k-stretch* and *k-thickening* operations that Jaeger, Vertigan, and Welsh define and use in a much more general setting [10].

Proof of Proposition 1.1. We first show that (c) $\Rightarrow$ (b). Suppose there exists a polynomial-time almost-w sampler for G. The following two-step procedure constitutes an almost uniform sampler for the orbits of the G-action on Σ^n: (1) using the almost-w sampler with input parameters (G, ε), sample $g \in G$ with probability approximately proportional to the weight $w(g) = k^{c(g)}$; (2) assign independently and u.a.r. an element from Σ to each cycle of g.

The outcome of this procedure is a r.v. $Y \in \Sigma^n$ satisfying

$$(1-\varepsilon)\,|G_x|\,Z^{-1} \le \Pr(Y=x) \le (1+\varepsilon)\,|G_x|\,Z^{-1},$$

for all $x \in \Sigma^n$, where $G_x = \{g \in G : x^g = x\}$ denotes the stabiliser of x, and $Z = |G|\,P_G(k, \dots, k)$. To see this, observe that the event $Y = x$ can occur only if a permutation $g \in G_x$ is selected in step (1). Thus there are $|G_x|$ ways to obtain the outcome $Y = x$, and each occurs with probability Z^{-1}. Employing the elementary group-theoretic fact that $|G_x| \times |x^G| = |G|$ we obtain, letting $N = P_G(k, \dots, k)$,

$$\frac{1-\varepsilon}{N\,|x^G|} \le \Pr(Y=x) \le \frac{1+\varepsilon}{N\,|x^G|},$$

which demonstrates that the two-step procedure given above would provide a satisfactory affirmative answer to question (b).

We now turn to the entailment (c) $\Rightarrow$ (a). For $i = 0, \dots, n$, let $G_i \le G$ denote the pointwise stabiliser $G_i = \{g \in G : j^g = j \text{ for all } j < i\}$. Observe that $G = G_0 \ge G_1 \ge \cdots \ge G_n = \{\mathbf{1}\}$,[7] and $|G_{i-1}| \le n\,|G_i|$ for $i = 1, \dots, n$. Generating sets for the pointwise stabilisers G_i may be read off from the strong generating set for G [6], so their use creates no computational difficulties. Our approach to estimating $P_G(k, \dots, k)$ is to express the value in question as a product

$$P_G(k, \dots, k) = k^n \prod_{i=1}^{n} \frac{P_{G_{i-1}}(k, \dots, k)}{P_{G_i}(k, \dots, k)}$$

(where we have used the elementary fact that $P_{\{\mathbf{1}\}}(k, \dots, k) = k^n$), and to estimate each of the factors using the Monte Carlo method. This approach is standard in the area [5, 12, 13, 14].

To meet the requirements of an fpras for $P_G(k, \dots, k)$, it is enough to obtain estimates for each of the factors that are correct to within relative error $1 \pm \varepsilon/2n$ with probability at least $1 - (4n)^{-1}$. To avoid unnecessary complications, assume $\varepsilon < 1$. We employ an estimator for the reciprocal

$$r_i = \frac{P_{G_i}(k, \dots, k)}{P_{G_{i-1}}(k, \dots, k)}$$

of the ith factor that is defined by the following two-step experiment: (1) using the postulated almost-w sampler with input parameters $(G_{i-1}, \varepsilon/7n)$, sample $g \in G_{i-1}$ with probability approximately proportional to the weight $w(g) = k^{c(g)}$; (2) if

[7] Here, $\mathbf{1}$ denotes the identity permutation.

$g \in G_i$ then return the value $|G_{i-1}|/|G_i|$, else return 0. Letting the outcome of this experiment be the random variable Y_i, we have

$$(1 - \varepsilon/7n)r_i \le \mathrm{Exp}(Y_i) \le (1 + \varepsilon/7n)r_i.$$

Observe that $\max Y_i = |G_{i-1}|/|G_i|$ and $r_i \ge 1$, the latter inequality being a consequence of the combinatorial interpretation of $P_G(k, \dots, k)$, and the fact that G_i is a subgroup of G_{i-1}. Thus the ratio $\mathrm{Exp}(Y_i{}^2)(\mathrm{Exp}\, Y_i)^{-2}$ — which is the critical parameter governing the efficiency of the Monte Carlo method — can be bounded above as follows:

$$\frac{\mathrm{Exp}(Y_i{}^2)}{(\mathrm{Exp}\, Y_i)^2} \le \frac{\max Y_i}{\mathrm{Exp}\, Y_i} \le \frac{|G_{i-1}|}{|G_i|} \le \frac{6n}{7}.$$

It follows, by a standard result in the area of Monte Carlo estimation, that $\mathrm{O}(\varepsilon^{-2} \times n^3 \log n^{-1})$ trials suffice to obtain an estimate for $\mathrm{Exp}\, Y_i$ that is correct to within relative error $1 \pm \varepsilon/7n$ with probability at least $1 - (4n)^{-1}$, and hence an estimate for r_i that is correct to within relative error $1 \pm \varepsilon/3n$ with similar probability. A relative error of $1 \pm \varepsilon/3n$ in the reciprocals of the factors translates to a relative error of at most $1 \pm \varepsilon/2n$ in the factors themselves. □

Before tackling the proof of Proposition 3.1 we require a uniform upper bound on the terms of the hypergeometric distribution.

LEMMA A.1. *Let n, r, s, a be non-negative integers satisfying $r, s \le n$ and $a \le r$. Then*

$$\binom{s}{a}\binom{n-s}{r-a}\Big/\binom{n}{r} = \mathrm{O}\left(\frac{n^{3/2}}{\sqrt{(r+1)(s+1)(n-r+1)(n-s+1)}}\right).$$

Proof. Denote by $B(i; n, p) = \binom{n}{i} p^i (1-p)^{n-i}$ the terms of the binomial distribution. We note the following simple estimates, which follow from Stirling's approximation:

$$\begin{aligned} B(i; n, p) &= \mathrm{O}\big([p(1-p)n]^{-1/2}\big), \quad \text{for } 0 < p < 1 \text{ and } 0 \le i \le n; \\ B(i; n, p) &= \Theta\big([p(1-p)n]^{-1/2}\big), \quad \text{for } 0 < i < n \text{ and } p = i/n. \end{aligned}$$

We may suppose that $0 < r, s < n$, otherwise the result is immediate. For any $0 < p < 1$,

$$\binom{s}{a}\binom{n-s}{r-a}\Big/\binom{n}{r} = B(a; s, p) B(r-a; n-s, p)/B(r; n, p).$$

Setting $p = r/n$, and applying the estimates given earlier, we have

$$\begin{aligned} \binom{s}{a}\binom{n-s}{r-a}\Big/\binom{n}{r} &= \mathrm{O}\left(\sqrt{\frac{n}{p(1-p)s(n-s)}}\right) \\ &= \mathrm{O}\left(\frac{n^{3/2}}{\sqrt{rs(n-r)(n-s)}}\right), \end{aligned}$$

from which the result immediately follows. □

Proof of Proposition 3.1. Let S_n denote the symmetric group on $[n]$, let $\mathfrak{M} = \mathfrak{M}(S_n, \{0,1\})$ be the Markov chain defined in Section 2, and, for all $x, y \in \{0,1\}^n$, denote by $P(x,y)$ the probability of transition from state x to y in $\mathfrak{M}$. We show that, starting in any state x, the probability of being in state y after just one step is at least $\Omega(\pi(y)/\sqrt{n}\,)$, where π denotes the stationary distribution, i.e.,

$$\frac{\pi(y)}{P(x,y)} = \mathrm{O}(\sqrt{n}\,), \quad \text{for all } x, y \in \{0,1\}^n. \tag{A.1}$$

It will follow, by an induction on the time-step t, that there exists a positive constant $\alpha = 1 - \Omega(n^{-1/2})$ such that

$$P^t(x,y) \geq (1-\alpha^t)\pi(y), \quad \text{for all } x, y \in \{0,1\}^n, \tag{A.2}$$

where $P^t(x,y)$ denotes the probability of transition from state x to y in t steps. Proposition 3.1 follows immediately.

Note that inequality (A.2) holds trivially for $t = 0$. For the induction step, observe that, for $t \geq 0$,

$$\begin{aligned} P^{t+1}(x,y) &= \sum_w P^t(x,w)P(w,y) \\ &= \sum_w \left[(1-\alpha^t)\pi(w)P(w,y) + \alpha^t\pi'(w)P(w,y)\right], \end{aligned}$$

where $\pi'(w) = \alpha^{-t}[P^t(x,w) - \pi(w)] + \pi(w)$ is a probability distribution, by the induction hypothesis. Thus, by (A.1),

$$\begin{aligned} P^{t+1}(x,y) &= (1-\alpha^t)\pi(y) + \Omega(\alpha^t/\sqrt{n}\,)\,\pi(y) \\ &= \left[1 - \alpha^t(1 - \Omega(n^{-1/2}))\right]\pi(y), \end{aligned}$$

establishing (A.2), for suitably chosen $\alpha = 1 - \Omega(n^{-1/2})$.

It remains to verify (A.1). Let $x, y \in \{0,1\}^n$, and suppose that there are a positions at which x and y are both 1, b at which x is 1 and y is 0, c at which x is 0 and y is 1, and d at which they are both 0. Since y has $a+c$ occurrences of 1, the orbit containing y has size $\binom{n}{a+c}$, and hence

$$\pi(y) = \frac{1}{n+1}\binom{n}{a+c}^{-1}.$$

Now suppose that a permutation $g \in S_n$ is chosen u.a.r., and the cycles are independently assigned 0 or 1, each with probability $\frac{1}{2}$. The probability that every cycle is assigned 0 is $4^{-n}\binom{2n}{n}$.[8] Applying this result four times, the transition probability from x to y may be calculated exactly:

$$P(x,y) = \binom{a+b}{a}^{-1}\binom{c+d}{c}^{-1} 4^{-a}\binom{2a}{a}4^{-b}\binom{2b}{b}4^{-c}\binom{2c}{c}4^{-d}\binom{2d}{d}.$$

[8] See, for example, Lovász [18, Ex. 3.12], setting $x = \frac{1}{2}$.

The first two factors repesent the probability, in step (1) of the experiment described in Section 2, that a permutation g is chosen that is consistent with word y being selected in step (2) of the same experiment. The other factors represent the probability that y is selected, conditional on a suitable choice of g. Using the estimate $4^a\binom{2a}{a}^{-1} = \mathrm{O}(\sqrt{a+1}\,)$, and Lemma A.1, we have

$$\begin{aligned}
\frac{\pi(y)}{P(x,y)} &= \frac{1}{n+1}\binom{a+b}{a}\binom{c+d}{c}\binom{n}{a+c}^{-1} 4^a\binom{2a}{a}^{-1}\times\cdots\times 4^d\binom{2d}{d}^{-1}\\
&= \frac{1}{n+1}\,\mathrm{O}\bigl(\sqrt{n^3/(a+b+1)(a+c+1)(b+d+1)(c+d+1)}\,\bigr)\\
&\qquad\quad \times \mathrm{O}\bigl(\sqrt{(a+1)(b+1)(c+1)(d+1)}\,\bigr)\\
&= \mathrm{O}(\sqrt{n}\,).
\end{aligned}$$

This establishes (A.1) and completes the proof. $\square$

Proof of Proposition 3.3. Suppose that G is a cyclic group generated by the permutation u. Let $\mathfrak{M}' = \mathfrak{M}'(G,\Sigma)$ be the Markov chain defined in Section 2, and denote by P^t the t-step transition probabilities of $\mathfrak{M}'$. Our proof strategy is to show that there exists a time $t = \mathrm{O}(n\log n)$, such that $P^t(g_0,\mathbf{1})$ is bounded away from 0, uniformly over choices of initial state g_0. The result then follows by a trivial coupling argument.

First observe that $P^t(g_0,\mathbf{1}) \geq |H|^{-1}P^t(g_0,H)$, where $H \leq G$ is any subgroup of G. To see this, partition the set of all random walks of length $2t$ from vertex g_0 in the bipartite graph B of Section 2 according to the first $2t-1$ steps of the walk: each block of the partition is a maximal set of walks that agree on all but the final step. Fix attention on a particular block of the partition, and consider the possibilities for the final step. A transition to the identity permutation is always available, while the number of transitions to a permutation in H is at most $|H|$. As each of these choices for the final step is equally likely, the observation follows. As a consequence of the observation, it is enough to bound the t-step transition probability $P^t(g_0,H)$ for some subgroup H of bounded size.

Let $(X_t : t = 0,1,\dots)$ be the integer sequence in which X_t is the unique integer $0 \leq i < |G|$ satisfying $u^i = g$, where g is the state of $\mathfrak{M}'$ at time t. Clearly, (X_t) is a Markov chain isomorphic to $\mathfrak{M}'$. Let p be any prime dividing $|G|$, and let q be the highest power of p dividing $|G|$. Then

$$\text{(A.3)}\quad \Pr\bigl(X_{t+1}\equiv 0 \pmod{q} \bigm| X_t = i\bigr) \geq \begin{cases} 1-k^{-(q-q/p)}, & \text{if } i\equiv 0 \pmod{q};\\ q^{-1}, & \text{otherwise,}\end{cases}$$

where, as usual, $k = |\Sigma|$. For the first part of inequality (A.3), note that u must contain a cycle whose length is a multiple of q, and let j be a point on such a cycle. If $g = u^i$ where q divides i, then the points $j, j^u, j^{u^2},\dots,j^{u^{q-1}}$ are all on distinct cycles of g, and are independently assigned symbols from the alphabet Σ during the first step of the two-step experiment that defines the transitions of $\mathfrak{M}'$. The probability that this assignment of symbols is consistent with some permutation $g' = u^{i'}$ with $i' \not\equiv 0 \pmod{q}$ is at most $k^{-(q-q/p)}$. This gives the first part of inequality (A.3).

For the second part, let $G^q = \langle u^q \rangle \le G$ denote the group generated by u^q, and observe that for any word $x \in \Sigma^n$, the subgroup $G_x \cap G^q$ has index at most q in G_x. Thus the probability of selecting an element of G^q in the second step of the experiment defining transitions of $\mathfrak{M}'$ is at least q^{-1}.

For a natural number i, let $\delta_q(i) = 0$ if $i \equiv 0 \pmod q$, and $\delta_q(i) = 1$, otherwise. The process $\delta_q(X_t)$ is not a Markov chain, since the transition probabilities depend on a knowledge of the current state X_t, and not simply on the binary value $\delta_q(X_t)$. However, $\delta_q(X_t)$ is stochastically dominated by the Markov chain (Z_t) on state space $\{0,1\}$, whose transition probabilities are given by

$$\begin{aligned}\Pr(Z_{t+1} = 1 \mid Z_t = 0) &= a = k^{-(q-q/p)},\\ \Pr(Z_{t+1} = 0 \mid Z_t = 1) &= b = q^{-1};\end{aligned}$$

in particular, it is possible to construct a joint sample space for (X_t) and (Z_t) such that $Z_t = 0$ entails $X_t \equiv 0 \pmod q$, provided only that $Z_0 = 1$. (This construction relies on the fact that $a \le b$.) Solving the two-state Markov chain exactly, we find

$$\begin{aligned}\Pr(Z_t = 1) &= \frac{a}{a+b} + \frac{b}{a+b}(1-a-b)^t\\ &\le \frac{a}{a+b} + \mathrm{e}^{-(a+b)t}\\ &\le q2^{-q/2} + \mathrm{e}^{-t/q},\end{aligned}$$

where we have used the inequalities $a \le 2^{-q/2}$ and $a+b \le 1$. In particular, setting $t = \lceil n \ln n + cn \rceil$ where c is a positive constant to be chosen later, and noting $q \le n$, we have the bound

$$\Pr(X_t \not\equiv 0 \pmod q) \le \Pr(Z_t = 1) \le q2^{-q/2} + \mathrm{e}^{-c}/n. \tag{A.4}$$

Let the order of u be $q_1 q_2 \dots q_s$, where $q_1 < q_2 < \cdots < q_s$ are powers of distinct primes $p_1, p_2, \dots, p_s$. Since $\sum_q q2^{-q/2}$ is convergent, there exists a initial value l for q such that $\sum_{q=l}^{\infty} q2^{-q/2} < 1$; in fact it suffices to take $l = 13$. Summing inequality (A.4) over all $q_i \ge 13$ — choosing c sufficiently large and noting $s \le n$ — we have $\Pr(X_t \not\equiv 0 \pmod m) = 1 - \Omega(1)$, where $m = \prod_{i:q_i \ge 13} q_i$.[9] Since the subgroup $H = G^m = \langle u^m \rangle$ is of bounded size (less than $8 \times 9 \times 5 \times 7 \times 11$) we deduce, using the observation made at the outset of the proof, that $P^t(g_0, \mathbf{1})$ is bounded away from 0, uniformly in the choice of initial state g_0.

Finally, take two copies of the Markov chain $\mathfrak{M}'$ and construct a coupling by running the two copies independently until such time as both are in the state $\mathbf{1}$, after which point the two copies run locked together. The probability that coupling has not occurred by time t is $\exp(-\Omega(t/n \log n))$. The result now follows immediately from a fundamental result on couplings; see, for example, Diaconis [4, Chap. 4, Lemma 5]. □

[9] Note that this crude upper bound on a union of events does not rely on probabilistic independence of the events.

Acknowledgement

I thank Leslie Goldberg for comments on an early draft of this article.

References

1. A. Z. Broder, *How hard is it to marry at random? (On the approximation of the permanent)*, Proceedings of the 18th ACM Symposium on Theory of Computing, ACM Press, 1986, pp. 50–58, Erratum in Proceedings of the 20th ACM Symposium on Theory of Computing.
2. N. G. de Bruijn, *Pólya's theory of counting*, Applied Combinatorial Mathematics (E. F. Beckenbach, ed.), John Wiley and Sons, 1964, pp. 144–184.
3. B. Cipra, *An introduction to the Ising model*, American Mathematical Monthly **94** (1987), 937–959.
4. P. Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics, Hayward CA, 1988.
5. M. Dyer, A. Frieze, and R. Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*, Proceedings of the 21st ACM Symposium on Theory of Computing, ACM Press, 1989, pp. 375–381.
6. M. Furst, J. Hopcroft, and E. Luks, *Polynomial time algorithms for permutation groups*, Proceedings of the 21st IEEE Symposium on Foundations of Computer Science, Computer Society Press, 1980, pp. 36–41.
7. L. A. Goldberg, *Automating Pólya theory: the computational complexity of the cycle index polynomial*, Information and Computation (to appear).
8. F. Harary and E. M. Palmer, *Graphical enumeration*, Academic Press, 1973.
9. E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift für Physik **31** (1925), 253–258.
10. F. Jaeger, D. L. Vertigan, and D. J. A. Welsh, *On the computational complexity of the Jones and Tutte polynomials*, Mathematical Proceedings of the Cambridge Philosophical Society **108** (1990), 35–53.
11. M. Jerrum, *A compact representation for permutation groups*, Journal of Algorithms **7** (1986), 60–78.
12. M. R. Jerrum and A. J. Sinclair, *Approximating the permanent*, SIAM Journal on Computing **18** (1989), 1149–1178.
13. M. Jerrum and A. Sinclair, *Polynomial-time approximation algorithms for the Ising model (Extended Abstract)*, Proceedings of the 17th EATCS International Colloquium on Automata Languages and Programming, Springer-Verlag, 1990, pp. 462–475.
14. M. Jerrum and A. Sinclair, *Polynomial-time approximation algorithms for the Ising model*, SIAM Journal on Computing (to appear).
15. M. R. Jerrum, L. G. Valiant and V. V. Vazirani, *Random generation of combinatorial structures from a uniform distribution*, Theoretical Computer Science **43** (1986), 169–188.
16. R. M. Karp and M. Luby, *Monte-Carlo algorithms for enumeration and reliability problems*, Proceedings of the 24th IEEE Symposium on Foundations of Computer Science, Computer Society Press, 1983, pp. 56–64.
17. W. Lenz, *Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern*, Zeitschrift für Physik **21** (1920), 613–615.
18. L. Lovász, *Combinatorial Problems and Exercises*, North Holland, Amsterdam, 1979.
19. A. Lubiw, *Some NP-complete problems similar to graph isomorphism*, SIAM Journal on Computing **10** (1981), 11–21.
20. E. M. Luks, *Isomorphism of graphs of bounded valence can be tested in polynomial time*, Journal of Computer and System Sciences **25** (1982), 42–65.
21. M. Mihail, *On coupling and the approximation of the permanent*, Information Processing Letters **30** (1989), 91–95.
22. C. C. Sims, *Computational methods in the study of permutation groups*, Computational Problems in Abstract Algebra (J. Leech, ed.), Pergamon Press, New York, 1970, pp. 169–183.
23. A. Sinclair, *Improved bounds for mixing rates of Markov chains and multicommodity flow*, Combinatorics, Probability and Computing **1** (1992), 351–370.

24. R. H. Swendsen and J-S. Wang, *Nonuniversal critical dynamics in Monte Carlo simulations*, Physical Review Letters **58** (1987), 86–88.
25. D. J. A. Welsh, *The computational complexity of some classical problems from statistical physics*, Disorder in Physical Systems, Oxford University Press, 1990, pp. 307–321.

Department of Computer Science, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, United Kingdom
E-mail address: mrj@dcs.ed.ac.uk