

Markov Chains and Polynomial time Algorithms

Ravi Kannan*

Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA, 15213

Abstract

This paper outlines the use of rapidly mixing Markov Chains in randomized polynomial time algorithms to solve approximately certain counting problems. They fall into two classes : combinatorial problems like counting the number of perfect matchings in certain graphs and geometric ones like computing the volumes of convex sets.

1 Introduction

We consider the problems of determining approximately the number of elements in certain sets; for example, the number of perfect matchings in a graph, the number of connected subgraphs of a given graph, the number of lattice points in a given polytope, or the number of $m \times n$ matrices of nonnegative integers with given row and column sums . In these and other “counting problems”, it turns out that the exact version is $\# P$ hard. The approximate version is polynomial time reducible to another problem which we may call the “(random) sampling problem” : output an element of the desired set randomly so that the probability of any particular element being output is approximately 1 over the cardinality of the set. [For the precise result, see Jerrum, Valiant and Vazirani [51].] This we refer to as (nearly) uniform sampling. A more general (“weighted”) version requires the probability of x being output to be approximately proportional to $F(x)$ where $F(\cdot)$ is a positive valued function on the set; we may call this “sampling according to” F . In the paper, we also consider the weighted version.

Our purpose here is to describe the recent success in developing polynomial time algorithms for some sampling problems using the following general scheme (for simplicity, consider the uniform sampling case): devise a Markov Chain whose states are the elements of the set and whose steady state probabilities are the same for all states. [This is often the easy part.] Then show that after a polynomial number of steps, the probabilities are “close” to the steady state probabilities [this is the hard part], so that running the Markov Chain for this many steps solves the sampling problem.

Let us illustrate with a toy example : suppose we wish to sample from the set of lattice points in an n -dimensional cube each of whose sides goes from 1 to a positive integer d (and have forgotten the obvious method). Here is a Markov Chain whose steady state

probabilities are (easily shown to be) $1/d^n$ for each state : from the current lattice point (in the cube), choose one of the $2n$ coordinate directions, each with probability $1/(2n)$; then go to the lattice point neighbor in that direction if it is in the cube, otherwise, stay at the current point. [This describes one move of the chain.] It can be shown that if we start the Markov Chain at any state, after $O(n^2 d^2 \log d)$ steps, the probability of each state is between $1/(2d^n)$ and $3/(2d^n)$. This is thus an example of a “rapidly mixing Markov Chain” where the number of steps needed so that the probabilities are close to the steady state is small (polynomial in n, d) whereas the number of states is d^n .

The problems for which this approach has yielded polynomial time algorithms, fall into two main categories : combinatorial problems like our first two examples or geometric problems, like our last two examples. [The algorithms are in general fully polynomial randomized approximation schemes (fpras) which take also as input a desired upper bound ϵ on the relative error; in this paper, we do not mention this ϵ anymore to avoid obscuring the main points.]

The rate of convergence of general Markov Chains was related to a combinatorial quantity called “conductance” by Sinclair and Jerrum [82] and Alon [7]. Broder [19] introduced the Markov Chain approach to combinatorial algorithms. Jerrum and Sinclair [48] pioneered the algorithmic approach by showing lower bounds on the conductance of certain Markov Chains, leading to a proof of fast convergence to the steady state.

In the geometric setting, the conductance is related to a natural geometric relation called isoperimetry. The proof of isoperimetric inequalities forms the central part of the proof of fast convergence of geometric Markov chains. The initial motivation came from the classical problem of computing the volume of convex sets (a problem obviously related to our third example above), but the approach has since been applied to other problems like multivariate statistical sampling, as well as to certain Stochastic Optimization problems.

Techniques other than the Markov chains approach have also been used to solve counting problems efficiently. Closed form formulas as one encounters in Combinatorics, the matrix-tree Theorem of Kirchoff that counts among others the number of trees in a graph, the algorithm of Karp and Luby to count the

*Partially supported by NSF-Grant CCR9208597

number of solutions to a disjunctive normal form formula [59] and the algorithms of Karpinski and Luby [60] are some examples of this. Recently also, there has been an important breakthrough by Barvinok [15] who has a polynomial time deterministic algorithm to count (exactly) the number of integer points in a fixed dimensional polytope using exponential sums. These results do not form the subject of this paper. Another important area that we do not cover here is the use of Markov Chains for graph traversal problems [6].

2 General Tools

2.1 Preliminaries, Notation

The reader is assumed to be familiar with the basic theory of Markov Chains. If P is the transition probability matrix with P_{xy} denoting the probability of transition from state x to state y , for any natural number t , the matrix power P^t denotes the t -step transition probabilities, i.e., $P_{x,y}^t$ is the probability of going from x to y in t steps. All our chains will be connected and aperiodic and thus have steady state probabilities, i.e.,

$$\pi(y) = \lim_{t \rightarrow \infty} P_{x,y}^t$$

exists and is independent of the start state x . [The notation $\pi(\cdot)$ will be used throughout for steady state probabilities.] We let the vector $p^{(t)} = p^{(0)}P^t$ denote the probabilities at time t where we start with the initial distribution $p^{(0)}$. All our chains will be “time-reversible”, i.e.,

$$\pi(x)P_{xy} = \pi(y)P_{yx}$$

will be valid for all pairs x, y . Symmetric Markov chains are ones for which $P_{xy} = P_{yx}$ for all x, y ; symmetric chains have uniform steady state probabilities. So they form a subclass of time-reversible ones. In fact, often our chains will be symmetric. We will henceforth assume that our chains are all time-reversible.

Linear algebra yields the following results [80]. The eigenvalues of P are $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \dots \lambda_N \geq -1$ (where N is of course the number of states). Viewing $p^{(t)}$ and π as N -vectors, one can quantify the convergence to the steady state by inequality of the form

$$\text{distance}(p^{(t)}, \pi) \leq s_0 \text{distance}(p^{(0)}, \pi) [\max(|\lambda_2|, |\lambda_N|)]^t.$$

Such an inequality holds for several standard definitions of distance like the L_∞ or L_1 distance (with a suitable definition of s_0). Perhaps the most useful distance to choose is the so-called “chi-squared” distance studied by Fill [37] following the work of Mihail [74] (the sum below is over all states x ; note that connectedness implies that $\pi(x) > 0$):

$$\chi^2(p^{(t)}, \pi) = \sum_x \frac{(p^{(t)}(x) - \pi(x))^2}{\pi(x)}$$

for which he essentially derives the following nice inequality with no s_0 . [The inequality is not explicitly stated in Fill’s paper.]

Theorem 1 For time-reversible Markov Chains, we have for any natural numbers t, T .

$$\chi^2(p^{(T+t)}, \pi) \leq [\max(|\lambda_2|, |\lambda_N|)]^t \chi^2(p^{(T)}, \pi).$$

So to prove fast convergence, it suffices to prove an upper bound on $\max(|\lambda_2|, |\lambda_N|)$.

More often, one encounters the so-called total variation distance. This distance between two probabilities p, q on a discrete set X is defined as $\sum_{x \in X} |p(x) - q(x)|$. The above theorem also gives a bound on the total variation distance between $p^{(t)}$ and π (as Fill notes) because this distance is easily shown to be at most $\chi^2(p^{(t)}, \pi)$ by using the Cauchy-Schwartz inequality.

[For the chains we consider, it is often easy to prove that the above maximum is attained by $|\lambda_2|$. But, even more simply, modifying a Markov Chain by making it stay at the current state with probability 1/2 and move according to its transition function with probability 1/2 ensures that $\lambda_N > 0$ while only increasing the (expected) running time by a factor of 2. In the Markov chains we describe below, we implicitly assume that this modification is made without stating it.]

The quantity

$$\frac{1}{1 - \max(|\lambda_2|, |\lambda_N|)}$$

is of interest because the inequality above guarantees that in this many steps, the chi-squared distance to the steady state is cut by a constant factor. We call this quantity the “mixing time”. In Markov Chains we deal with, the number of states will be exponential whereas the mixing time will be proved to be polynomial.

2.2 Conductance

Alon [7] and Sinclair and Jerrum [82] related $\max(|\lambda_2|, |\lambda_N|)$ to a combinatorial quantity called “conductance”. This has turned out to be of great use in practice; almost all known proofs of polynomial time convergence use conductance. Actually, this kind of relation between the second largest absolute value eigenvalue and conductance was proved in the setting of Riemannian manifolds by Cheeger in 1970 [20] and is known as Cheeger’s inequality there.

The conductance is defined as follows: for any subset S of states of the Markov Chain with $0 < \pi(S) = \sum_{x \in S} \pi(x) < 1$, the *conductance* $\Phi(S)$ of S is defined by

$$\Phi(S) = \frac{\sum_{x \in S, y \notin S} P_{xy} \pi(x)}{\min(\pi(S), 1 - \pi(S))}.$$

The conductance Φ of the whole chain is defined by (the minimum taken over all such subsets S of states)

$$\Phi = \min_S \Phi(S).$$

Restricting attention first to sets S with $\pi(S) \leq 1/2$, we see that the conductance of S measures what

we might call the “conditional escape probability” from S to its complement. For sets with $\pi(S) > 1/2$, we may use time-reversibility to rewrite $\Phi(S)$ as the conditional escape probability from the complementary set \bar{S} of states. Further for singleton sets $\{x\}$ with $\pi(x) \leq 1/2$, the conductance of $\{x\}$ is just the probability of moving in one step from x to some $y \neq x$. This quantity will later be called the “local conductance” and will play an important role.

To give some more intuition about conductance, we look at random walks on undirected graphs. In this setting, we have an undirected graph $G(V, E)$; the Markov Chain is on the vertices of the graph. From the current vertex, we choose an adjacent vertex uniformly at random and move to it. Such Markov Chains have been widely studied. For such chains, it is easy to show that the steady state probabilities are proportional to the degrees. Thus for sets S with $\pi(S) \leq 1/2$, we have

$$\phi(S) = \frac{\text{number of edges in the } (S, \bar{S}) \text{ cut}}{\text{sum of the degrees of the vertices in } S}.$$

Specializing further to regular graphs of degree r , we see that the denominator is replaced by $r|S|$. Hence, a lower bound on the conductance of the chain says that the graph possesses a certain weak expansion property (in the sense used in the literature on expanders).

It is intuitively easy to see that if some set S has “low” conductance, the mixing time is high: observe that (i) among all probability distributions that put 0 probability on all states of \bar{S} , the lowest χ^2 distance to steady state is achieved by the one that puts $\pi(x)/\pi(S)$ on each state x of S ; (ii) so, if we start in this distribution, the χ^2 distance can only improve after some positive probability is attached to some state in \bar{S} and (iii) this takes (expected) time at least $1/\Phi(S)$, (iv) so the mixing time has to be at least this amount, by the definition of conductance of S . The left inequality of the Theorem below quantifies this.

More interestingly, Alon and Milman and Sinclair and Jerrum showed a converse: that if the conductance of the chain is high, then the mixing time is low. Here is a precise theorem (see also [25],[37]).

Theorem 2: Let $\beta = \max(|\lambda_2|, |\lambda_N|)$ be the second largest absolute value eigenvalue of a time-reversible Markov chain and Φ its conductance. Then we have

$$1 - 2\Phi \leq \beta \leq 1 - \frac{1}{2}\Phi^2.$$

As mentioned above, with this theorem on hand, we can show rapid mixing by showing a lower bound on conductance. There are, in addition, several other techniques to prove rapid mixing, see for example Aldous [2], Diaconis and Strook [25], and Fill [37]. [There is also some recent work of Chung and Yau [21] on eigenvalues of certain types of graphs that may improve the running times of some algorithms.] There is a direct proof that combines the results of Theorems 1 and 2 (i.e., proofs bounds on the χ^2 distance in terms of conductance without going through eigenvalues [74].

In the combinatorial examples, the technique using the so-called “canonical paths” to prove lower bounds on conductance pioneered by Jerrum and Sinclair [48] has been the tool used in nearly every successful rapid mixing proof. We outline this technique on the toy example of the discrete cube from the introduction.

Again, let

$$C = \{x \in \mathbf{R}^n : x_i \in \{1, 2, \dots, d\} \text{ for } i = 1, 2, \dots, n\}.$$

Consider the Markov Chain in the introduction. For each pair $x, y \in C$, we define a path (called the canonical $x - y$ path) $\text{CP}(x, y)$ in the directed graph underlying the Markov Chain. $\text{CP}(x, y)$ is obtained by “fixing” in turn first, second, ..., n th coordinate of x , so it becomes y . [I.e., we first increase or decrease x_1 one step at a time until it becomes y_1 , then do the same for x_2 etc.] Define the **congestion** of an edge of the graph to be the number of x, y pairs whose $\text{CP}(\cdot, \cdot)$ uses the edge.

We claim that the congestion of each edge is at most $d|C|$. To see this, consider an edge e from z to z' where $z = (z_1, z_2, \dots, z_i, \dots, z_n)$ and $z' = (z_1, z_2, \dots, z_{i-1}, z_i + 1, z_{i+1}, \dots, z_n)$. The idea of the argument is that the $\text{CP}(x, y)$ that use the edge must satisfy $y_1 = z_1, y_2 = z_2, \dots, y_{i-1} = z_{i-1}$ and $x_{i+1} = z_{i+1}, \dots, x_n = z_n$ because we are modifying the coordinates in order $1, 2, \dots, n$. More precisely, corresponding to any $\text{CP}(x, y)$ that uses the edge, we may define a “complementary” point $w \in C$ as follows:

$$w_1 = x_1, w_2 = x_2, \dots, w_i = x_i \quad w_{i+1} = y_{i+1}, \dots, w_n = y_n.$$

Then, given w and y_i (and z, z'), it is clear that x and y are uniquely determined. Thus corresponding to each $w \in C$, there are at most d $\text{CP}(x, y)$ that use the edge which proves the claim.

Now suppose $S \subseteq C$. There are $|S||\bar{S}|$ canonical paths going from S to \bar{S} . By the congestion result, there must be at least

$$\frac{|S||\bar{S}|}{d|C|} \quad (\geq \frac{\min(|S|, |\bar{S}|)}{2d})$$

edges crossing from S to \bar{S} , each of which contributes a probability of $1/(2n)$. So,

$$\Phi \geq 1/(4nd).$$

This only gives a mixing time of $O(n^2 d^2)$ and working through Theorem 1 we see that the closeness to the steady state claimed in the introduction can be achieved in time $O(n^3 d^2 \log d)$ steps. The actual mixing time is $\Theta(nd^2)$ as can be shown by another technique for proving rapid mixing called “coupling” (see for example [1], [2]). It seems to be the case that coupling, and other techniques yield sharper results than does conductance for certain simple Markov Chains; but apparently for more complex chains needed to get polynomial time algorithms, conductance seems to be the tool of choice. There are some limited exceptions - see for example [46] and [39].

In the toy example, the chain was symmetric; there are weighted versions of the above canonical paths argument in case that the steady state distribution is not uniform. (for example [50].

The question arises as to how much we might loose by going through this method to prove a lower bound on conductance. A result of Leighton and Rao [65] on multi commodity flows can be used to show that if we generalize the canonical flows argument to allow flows of unit value (rather than just paths) between each pair of nodes, then the argument can loose at most a factor of $O(\log N)$ where N is the number of states of the Markov Chain. See [81] for a discussion of this connection.

By Theorem 1, an upper bound on the number of steps we need to run the Markov Chain before we can get one sample (from a distribution close to the desired one) is given by the mixing time times a quantity depending on $\chi^2(p^{(0)}, \pi)$. For estimating numerical quantities, we need many such samples; standard analysis assumes that the samples are independent and using standard probability techniques, one then derives error bounds. But it turns out that we do not need total independence of samples. Aldous [3] has shown that we can run the Markov Chain for as many steps as dictated by the Theorem 1 say to get the first sample, but then we can start the run of the chain at this sample to get the second and so on (the point being that the sample is already drawn from a distribution close to the desired one, so its χ^2 distance to the steady state may be less than that of the original starting distribution). He shows that we can essentially draw second and subsequent samples after every mixing time number of steps.

There is more work on how to exploit a long run of the Markov Chain to estimate quantities we need [44],[53].

3 Combinatorial examples

We will describe in some detail the first successful application of the method to matchings. Other applications will then be outlined.

3.1 Matchings

The first problem for which this approach yielded a polynomial time algorithm was that of sampling uniformly from the set of perfect matchings in a dense bipartite graph (dense means each vertex is adjacent to at least half of all vertices); this result is due to Jerrum and Sinclair [48]. It is still an open problem as to whether the restriction of dense can be removed; if so, this would of course give a fpras for approximating the 0-1 permanent. Computing the permanent exactly was the first problem shown to be $\#P$ hard by Valiant [87]. For this and other practical reasons, approximating the permanent remains a very interesting open problem. The best known randomized approximation algorithm for this problem runs in sub-exponential ($e^{O(\sqrt{n})}$) time (see Jerrum and Vazirani [52]). There is also an older exponential time algorithm [58]. A recent result [30] shows that an approach based on the above paper actually succeeds in polynomial time for certain special cases including the

“strictly” dense case (where the degree of each vertex is at least αn for some constant $\alpha > 1/2$.) This approach is not based on Markov Chains. In the following, we describe only the Markov Chain approach of Jerrum and Sinclair.

Broder [19] made an important contribution by first introducing the Markov Chain approach to polynomial time algorithms. He proposed the Markov Chain used by Jerrum and Sinclair and claimed using a coupling argument that it mixes rapidly for the dense case. However, there was a fatal flaw in the proof. [75].

The Markov Chain has as its set of states the set \mathcal{M}_n of perfect matchings as well as the set \mathcal{M}_{n-1} of near-perfect matchings (a near perfect matching is one that has one less edge than a perfect one) of the given bipartite graph $G(V, E)$ on $2n$ vertices. Transitions of the Markov Chain are as follows :

In any current state, M , we pick an edge $e = (u, v)$ of the graph uniformly at random (all edges are equally likely) and

- if $M \in \mathcal{M}_n$ and $e \in M$, move to $M' = M - e$.
- If $M \in \mathcal{M}_{n-1}$ and u and v are both unmatched in M , then move to $M' = M + e$.
- $M \in \mathcal{M}_{n-1}$, u is matched to w in M and v unmatched, then move to $M' = (M + e) - (u, w)$; make a symmetric move if v is matched and u unmatched.
- In all other cases, stay at M .

The proof of high conductance is based on a canonical paths argument. Here is a very quick outline of this proof. The symmetric difference of any two matchings is the union of disjoint cycles and paths. We order at the outset all cycles and paths in the graph and also order the edges along each cycle or path. A canonical path $CP(M, M')$ from a matching $M \in \mathcal{M}_n \cup \mathcal{M}_{n-1}$ to a matching M' in the same set is obtained by “fixing” in this order each cycle or path (recall the toy example) : i.e., we switch pairs of edges along each cycle or path (in order) so that we transform M into M' . For any transition $Z - Z'$ of the chain, and any $CP(M, M')$ that uses the transition, we may again define a complementary matching W which agrees with M on the cycles and paths (and part thereof) which are already fixed and agrees with M' on those not yet fixed. There is some delicacy in handling the edges that the transition is in the process of fixing. Then, one argues that W, Z, Z' determine M, M' uniquely. The reader is of course referred to the paper for details.

The algorithm only yields a perfect or near-perfect matching. To sample from perfect matchings, we just reject if we get a near-perfect matching and repeat. That we do not (expect to) repeat too many times before getting a perfect matching follows from the fact (proven by Broder [19]) that for dense graphs,

$$\frac{|\mathcal{M}_{n-1}|}{|\mathcal{M}_n|} \leq \text{poly.}$$

In fact, Jerrum and Sinclair show that we can sample nearly uniformly from the perfect matchings in the class of graphs that satisfy the above inequality.

Another contribution of their paper is a polynomial time algorithm to sample nearly uniformly from the set of all matchings in **any** graph. This follows as a special case of a problem called the “monomer-dimer” problem which arises in Physics. Computationally this problem is the following : given a graph $G(V, E)$ with nonnegative weights $c(e)$ on the edges of the graph, generate a (not necessarily perfect) matching in G such that the probability of generating any matching M is proportional to the “weight” $W(\cdot)$ of the matching defined as the product of the $c(\cdot)$ on the edges of the matching. The first task as usual is to come up with a Markov Chain whose steady state probabilities are proportional to the weights of the matchings. This is accomplished by an old general method often called the Metropolis method described below. [73].

Suppose we have a symmetric Markov Chain (whose steady state probabilities are uniform) and we wish to modify the chain so that the modified chain has steady state probabilities proportional to a function F on the states. Here is one simple modification that works : from the current state M , choose using the old chain the next state M' . However, now we do not definitely go to M' , but only go there with probability $\min(1, \frac{F(M')}{F(M)})$; i.e., the new transition probability from M to $M' \neq M$ is

$$P_{M,M'} \min\left(1, \frac{F(M')}{F(M)}\right),$$

where P was the transition probabilities for the chain with uniform steady state. We call the modified Markov Chain the **Metropolis version** of the original random walk. Actually, other modifications will also work; for example we may also let the new probabilities be

$$P_{M,M'} \left(\frac{F(M')}{F(M) + F(M')} \right) \quad \text{or} \quad P_{M,M'} \frac{\sqrt{F(M')}}{c\sqrt{F(M)}}.$$

For the chains we will encounter all these chains will have mixing times that are within a constant of each other.

[More generally, we need not start with a symmetric chain. We may modify any convenient time-reversible “base” chain in a manner that is easy to work out.]

The sub-exponential algorithm of [52] for counting approximately the number of perfect matchings in a general bipartite graph is based on a result they prove which roughly says that if the chain above does not mix rapidly for a certain graph, then there is a “small” ($O(\sqrt{n})$ - size) “obstruction”. They devise an algorithm to find such an obstruction and then deal with the remaining graph recursively.

3.2 The Ising model

The Ising model is a model arising in Physics. (see [50] and references there). The computational problem arising from that is the following : we are given a

real symmetric $n \times n$ matrix V (the entries of V arise as pairwise interaction energies), a real number B (the external field) and a positive real number β (the temperature). The Ising partition function is defined as

$$Z = Z(V_{ij}, B, \beta) = \sum_{\sigma \in \{-1, +1\}^n} e^{-\beta H(\sigma)},$$

where the “Hamiltonian” $H(\sigma)$ is given by

$$H(\sigma) = - \sum_{i,j} V_{ij} \sigma_i \sigma_j - B \sum_k \sigma_k.$$

We do not discuss in any detail the physical importance of the model. However, it is apt to point out that the rough idea is the following : there are n objects (possibly electrons) that mutually interact and V_{ij} quantifies the level of interaction between object i and j . Each object is in one of two states - $+1$ or -1 (the state may represent spin for example). $H(\sigma)$ represents the energy of a particular configuration σ of states. Each configuration σ occurs with probability proportional to $e^{-\beta H(\sigma)}$; Z is the normalization constant for the probabilities.

Jerrum and Sinclair [50] presented a polynomial time approximation algorithm to compute Z in the case when all V_{ij} are nonnegative (called the ferromagnetic case). They also show that a randomized polynomial time approximation algorithm in the general case would imply NP=RP.

Their algorithm for the ferromagnetic case first reduces the problem to the corresponding sampling problem and then more interestingly reduces this sampling problem to another one concerning subgraphs of a graph, namely the following :

Given a graph $G(V, E)$ with positive edge weights $\lambda(e)$, pick a subset of edges of G at random such that the probability of picking a particular subset T is proportional to

$$w(T) = \mu^{|\text{odd}(T)|} \prod_{e \in T} \lambda(e),$$

where μ is a given positive number and $\text{odd}(T)$ denotes the set of odd degree vertices in T .

Here is the random walk they use. The states of the Markov Chain are the subsets of E . Their chain is the Metropolis version of the following simple Markov Chain whose steady state probabilities are uniform over all subsets of the edges, namely : at any current subset T of E , pick uniformly at random an edge $e \in E$; if $e \in T$, then go to $T' = T - e$, otherwise go to $T' = T + e$. [On top of taking the Metropolis version, they also modify the chain to stay at the current state with probability $1/2$ to handle the least eigenvalue.]

The proof of a lower bound on conductance relies on a canonical paths argument. To define the canonical path from a subset T to T' , first they note that the symmetric difference S between T and T' can be decomposed into a disjoint union of cycles and trails.

These cycles and trails are “fixed” in some order chosen at the outset to get from T to T' . Details are found in the paper.

The algorithm that is preferred by physicists is the one due to Swendsen and Wang [86]. This algorithm switches the signs on large blocks of vertices of the graph at once. But while this seems to work well in practice, no proof of rapid mixing is known.

3.3 Other Combinatorial problems

Several other approximate counting problems have been reduced to approximate counting of perfect matchings in dense graphs. These include counting Hamilton cycles in dense or random graphs (Dyer, Frieze and Jerrum [30]) and dense digraphs (Frieze and Suen [40]), counting graphs with a given degree sequence (Jerrum and Sinclair [49]) and counting Eulerian orientations of an undirected graph (Mihail and Winkler [76]) (an orientation is Eulerian if the in-degree of each vertex equals its out-degree).

Aldous [4] and Broder [18] show that the following simple idea samples perfectly uniformly from the set of trees of an undirected graph: do a random walk on the graph and keep any edge that is used for visiting a vertex for the first time.

There has been very limited success on certain other problems. There is a sub-exponential time (dependence on n is $e^{O(\sqrt{n}(\log n)^3)}$) algorithm to count approximately the number of 0-1 solutions to an n variable linear inequality, i.e., the number of feasible solutions to a 0-1 knapsack [32]. The Markov chain is the obvious one - at a current feasible solution x , pick at random one of the n coordinates each with probability $1/n$ and change it to its complement if the change preserves feasibility; otherwise stay at x . They generalize to more than one inequality.

Alon, Frieze and Welsh [8] have an algorithm to evaluate the Tutte polynomial of a dense graph at several values of the parameters; this among things for example counts approximately the number of forests in such graphs, a result due to Annan [10] earlier.

Feder and Mihail [38] show that a natural random walk for sampling from the bases of what they call a balanced matroid (see for example Welsh [88] for a treatment of matroids) mixes in polynomial time. A balanced matroid is one where the event that an edge e is in a basis is negatively correlated with the event that some other edge f is in the basis. The class of regular matroids is a subset of the class of balanced matroids. The problem of sampling from the set of bases of a general matroid is an important open problem. The problem is open for even cographic matroids; this special case includes the problem of sampling uniformly from the set of connected subgraphs of a given graph (or counting their number) and arises in reliability computations (see open problems section).

One other interesting open problem would be to come up with a rapidly mixing chain to sample from the strings in a regular language of a certain length. There has been some recent work on this [57].

4 Geometric Markov Chains

The initial motivation for geometric random walks in convex sets that mix in polynomial time came from the classical problem of computing the volume of convex sets in Euclidean n -space where n is variable. For fixed n , this can be done by decomposing the convex set (approximately) into simplices. However, the number of simplices needed grows exponentially with the dimension. In fact, Dyer and Frieze [28] showed that the problem of (exactly) computing the volume of a polytope is # P-hard. Independently, Khachiyan [62] had also shown the NP-hardness of this problem. Even a special case - that of computing the volume of a so-called “order polytope” (one defined by a set of linear inequalities of the form $x_i \leq x_j$ and $0 \leq x_i \leq 1$), which is equivalent to the problem of computing the number of linear extensions of a partial order has been shown to be # P hard by Brightwell and Winkler [17].

There are also absolute hardness results for general convex sets. A general convex set is described by an oracle Grötschel, Lovász and Schrijver [45]. A “membership oracle” is a black box description of a convex set which for any queried point x says whether or not x is in the convex set. Using the ellipsoid algorithm of Khachiyan [61], it is shown in [45] that given a convex set by a membership oracle plus the center and radius of a ball contained in the set and the center and radius of a ball containing the set, we can solve many interesting problem regarding the convex set, most notably, we can approximately minimize a linear function over the set in polynomial number of calls to the oracle plus polynomial additional time. We will assume this model of the convex set. [The two balls are technical pieces of information. It is best to think of the model as describing the convex set essentially by a membership oracle, i.e., in a sense, we can have local information about the set, but no global or closed form description is given.] In this model, the Elekes [35] and Bárány and Füredi [14] have shown that we cannot compute the volume of convex sets approximately even upto exponential (in n) factors. This is an absolute result (not a hardness one), but of course in this model where the input is given by an oracle. See also [69].

We assume that a convex set is given as in this model.

By proving that a natural Markov Chain mixes rapidly, Dyer, Frieze and Kannan [31] were able to solve the problem of sampling with nearly uniform density from convex sets; they reduced the problem of computing approximately the volume of convex sets to this sampling problem and thus came up with the first fully polynomial randomized approximation scheme for the volume of convex sets.

The classical Monte Carlo approach to, say, computing the area of a 2-dimensional figure is one of throwing darts: enclose the figure in, say, a rectangle whose area is known, throw darts (sample with uniform density) from the rectangle; then the ratio of the number of darts that land in the figure to the total is a good approximation to the ratio of the area of the figure to that of the rectangle. The basic idea would be to apply the same technique. The first problem,

however is that if we wish to enclose an arbitrary convex set in a nice set (whose volume is known in closed form) like a ball, the ratio of the volume of the ball to that of the convex set cannot be bounded; since this ratio is the expected number of darts to be thrown before one lands in the set, we will essentially have no information until we throw this many darts. This problem is somewhat mitigated by the following : A result of Grötschel, Lovász and Schrijver [45] finds a linear transformation τ in polynomial time so that τ transforms the given convex set K into a well-rounded one, namely :

$$B(0, 1) \subseteq \tau K \subseteq B(0, n^{3/2}),$$

where $B(p, r)$ denotes a ball of radius r with p as center, i.e., $B(p, r) = \{x : |x - p| \leq r\}$. We may replace K by τK and assume that $B(0, 1) \subseteq K \subseteq B(0, n^{3/2})$. [Note that linear transformations preserve the ratio of the volumes of the ball to the convex set.]

Now we may throw darts at the outer ball (it is well-known how we may pick a random point with uniform density from a ball) and see the proportion that land in the convex set.

But the volume of the outer ball could be as much as $n^{3n/2}$ times the volume of the convex set, so this does not give a polynomial time algorithm. Instead, we write a similar ratio as the (telescopic) product of several ratios :

$$\frac{\text{volume}(B(0, 1))}{\text{volume}(K)} = \prod_{i=0}^k \frac{\text{volume}(K \cap B(0, \lambda^i))}{\text{volume}(K \cap B(0, \lambda^{i+1}))},$$

where $\lambda = (1 + \frac{1}{n})$ and k is the least positive integer such that $\lambda^k \geq n^{3/2}$.

Each ratio in the telescopic product is the ratio of volume of a convex set P to the volume of a convex set Q containing it, but it is easy to see that Q has volume at most $\lambda^n \leq O(1)$ times the volume of P . So if we can throw darts at Q , we could approximate the ratio well. Thus we have reduced the volume of convex sets problem to that of sampling with uniform density from a general convex set.

The sampling problem can be solved by one of several random walks. The original Markov Chain used in [31] was the following coordinate walk : the states of the Chain are all the points in K with coordinates which are integer multiples of a certain positive real number δ . The choice of δ will be addressed below. We refer to the states as grid points. At any current state x , we choose one of the $2n$ coordinate directions uniformly at random and go to the coordinate neighbor point (state) in the chosen direction provided it is in K ; if it is not in K , we stay at x . [Note that the membership oracle for K is enough to determine the move.] It turns out that this random walk does not converge very fast (in fact is not even connected) for a general convex set because the set could have a sharp corner so that a grid point has no coordinate grid neighbor in the set. However, the authors proved that if sharp corners are avoided, we do get fast convergence. Here is a statement of their main theorem.

Theorem 3 : Suppose K is a well-rounded convex set and for every point $x \in K$, there is a ball of radius $n^{-3/2}$ containing x and wholly contained in K , then with $\delta \leq n^{-5/2}$, the above coordinate random walk has mixing time polynomially (in n alone) bounded.

For any well-rounded convex set K , the set $K(\alpha)$ of points at (Euclidean) distance at most α is also a convex set for any positive α ; choosing $\alpha = n^{-(3/2)}$, and replacing K by $K(\alpha)$ does not change the volume too much, but ensures the hypothesis of the Theorem and thus leads to an algorithm for the volume of a general convex set.

The current best known results for the mixing time of coordinate random walk (including the weighted case of log-concave measures- see section 4.2) is found in Frieze, Kannan and Polson [39].

While the mixing time of the original algorithm was polynomial, it was a very high polynomial leading to a complexity dependence on n of $O(n^{27})$ for approximating the volume. There have been a sequence of improvements of this complexity; the best currently known algorithm makes only $O^*(n^5)$ calls to the membership oracle and does an additional $O^*(n^6)$ work to approximate the volume (Kannan, Lovász and Simonovits [54]. [The * obscures some $\log n$ factors as well as the dependence on ϵ the relative error allowed. We use this convention throughout.] These improvements have been based generally on new mathematical ideas and the more recent algorithms are in fact fairly simple in spirit. We will describe the progress below, but here is first a brief summary of it. Lovász and Simonovits [67] did away with the artificial device of going to $K(\alpha)$ and instead excluded small exceptional sets of points (like those near sharp corners) from the definition of conductance (and handled them differently). More importantly, this paper proved a strong version of an isoperimetric inequality proposed in [31]. Using these two ideas, they were able to bring down the number of oracle calls to $O^*(n^{16})$.

The isoperimetric inequality was also independently proved by Khachiyan and Karzanov [63]. The inequality and its generalization by Applegate and Kannan [11] play a crucial role in all the algorithms. [11] improved the running time to $O^*(n^{10})$ by using sampling according to log-concave densities, a problem of independent interest in Statistics [12]. Dyer and Frieze [27] improved this further to $O^*(n^8)$; they saved a factor of n by arguing that independence of the samples in the evaluation of different ratios in the telescopic product implies lower variance and also by incorporating the handling of small sets as in [67].

Lovász and Simonovits [68] improved this to $O^*(n^7)$ and $O^*(n^8)$ additional time by introducing and analyzing a random walk with a continuous state space, namely : from the current point x in the convex set, choose with uniform density a point y in a ball of radius δ around x and go to y if it is in the set, else, stay at x . We call this the "ball walk". (with radius δ). The ball walk generally seems to save a factor of n from the number of oracle calls needed. However, since each move needs to update n coordinates, there is no

saving in the overall work needed. The ball walk also has the advantage that proofs are a little simpler; this is because for the coordinate walk, the known proofs first transform the discrete problem to a “continuous” property involving surface areas and volumes, whereas for the ball walk, the transformation is unnecessary.

4.1 Conductance of the Ball Walk and Isoperimetric inequality

Consider the ball walk in a convex set K in \mathbf{R}^n with balls of radius δ . This is a continuous state space random walk; so some rigorous extension of the concepts related to Markov Chains (which have finite state space) is necessary for carrying out an analysis. Such an extension is possible and the interested reader is referred to standard texts or [68]. However, for the discussion here, it suffices to note that there are no surprises in the extension, so we may use our intuition to replace the discrete concepts by their corresponding natural continuous analogs (replace probability by probability density etc.). All sets over which we have integrals in the following are assumed to be measurable.

We use the notation P_{xy} for the transition probability density from x to y ; for the ball walk, this is given by

$$\frac{1}{\text{volume}(B(0, \delta))} \quad \text{for } x \neq y, x, y \in K, |x - y| \leq \delta$$

and zero for other pairs $x \neq y$; also the probability of staying at x is $1 - \int_y P_{xy} dy$. For a subset S of K and $x \in K$, we denote by $P(x, S)$ the probability of going from x to S in one step, i.e., $P(x, S) = \int_{y \in S} P_{xy} dy$. The steady state probability density $\pi(\cdot)$ is of course uniform (everywhere equal to $1/\text{volume}(K)$) for the ball walk.

The conductance of a (measurable) subset S of K is now defined as

$$\frac{\int_{x \in S} \pi(x) P(x, K \setminus S)}{\min(\pi(S), 1 - \pi(S))}.$$

The above expression is stated in general not just for the ball walk.

Now we introduce isoperimetry. The term comes from the classical fact that among all curves of the same perimeter, the circle (in 2-space) has the smallest area. Such a fact is also true in higher dimensions. This fact may be restated as a lower bound on the surface area of an n dimensional body in terms of its volume. The isoperimetric inequality we will need involves instead a lower bound on the surface area of a surface dividing a convex set into two parts.

More precisely, we define the **isoperimetric constant** of a convex set K as infimum over all measurable subsets S of K (with the denominator below not equal to zero) of the

$$\frac{\text{surface area of } (\partial S \cap \text{interior}(K))}{\min(\text{volume}(S), \text{volume}(K \setminus S))}.$$

[∂S is the $n - 1$ dimensional boundary of S ; it is assumed above that this is measurable, i.e., has $n - 1$ dimensional measure, which is referred to as surface area.]

If for example, we have a cylinder of length d cut into S and \bar{S} by a surface perpendicular to the axis in the middle, then the above ratio for this S is obviously, $2/d$. The central theorem we now have about the isoperimetric constant is that this is the worst among all convex sets. More precisely,

Theorem 4 : The isoperimetric inequality
The isoperimetric constant of a convex set of diameter d is at least $2/d$.

The basic theorem was proved by Lovász and Simonovits [67] and independently also by Khachiyan and Karzanov [63] with $1/d$ instead of $2/d$. The proof of the first reference is simpler (but the other proof using differential geometry gives more information about the extreme S). We will indicate an outline of the proof below. The result was generalized to the case of log-concave measures (instead of the Lebesgue measure) by Applegate and Kannan [11] (See section 4.2 below). The improvement to $2/d$ is by Dyer and Frieze [27].

[The term isoperimetric inequality is used in the literature to refer generally to such inequalities.]

We denote by $\text{Iso}(K)$ the isoperimetric constant of K .

Here is an outline of the basic proof : assume wlg that $\text{volume} S \leq \text{volume}(K \setminus S)$. Using the Borsuk-Ulam or the so-called ham-sandwich Theorem we can cut the convex set K by a hyperplane into two parts K_1, K_2 such that

$$\text{volume}(K_1 \cap S) = \text{volume}(K_2 \cap S) \quad \text{and,}$$

$$\text{volume}(K_1 \setminus S) = \text{volume}(K_2 \setminus S),$$

whence we have that if K violates the Theorem, then so does either K_1 or K_2 . They show that by repeating this process carefully, we arrive essentially at a one-dimensional counterexample to the Theorem, actually, we arrive at something called a needle-like set : a needle-like set is a convex set in \mathbf{R}^n which has the property that it has an “axis” (a line) L such that every cross section of it perpendicular to L is contained in $(n - 1)$ dimensional ball of radius τ where $\tau > 0$ is allowed to tend to zero. Such a reduction to the

one-dimensional case was used in a similar context by Payne and Weinberger [78].

Then it only remains to argue the result for needles which a direct calculation yields. The reduction to the needle like case is fairly general - it also works for other measures than the uniform. This is observed and used in [11], see [55] for a general statement of the reduction to the needle-like case. [68] uses the reduction to the needle like case to prove other properties of convex sets as well, many of which had longer proofs originally.

What [11] showed was that for log-concave measures, the result holds for the needle-like case and thus in general. Radcliffe [79] has generalized this further by showing that this is true for measures whose reciprocals are of the form (convex function)ⁿ⁻¹. More recently, [55] shows that the isoperimetric constant is in fact at least c/D where $c > 0$ is an absolute constant independent of n , K and D is now the “effective diameter”, which is defined as the expected distance from the center of gravity of points in K . [This is also true for log-concave measures where the above expectation is also taken with respect to this measure.]

We now give the argument that the lower bound on the isoperimetric constant implies essentially high conductance and therefore low mixing time. This argument is essentially due to Lovász and Simonovits [68]. Since it is not quite stated and proved in this form, we give a proof here. We need first one other concept called local conductance. We define the local conductance of a Markov Chain at a point x to be the probability that we move to $y \neq x$ from x in one move. Clearly a necessary condition for low mixing time is that the local conductance of any point be high. Here is a precise theorem.

Theorem 5 Consider the ball walk with balls of radius δ in a convex set K which has local conductance at least $\nu > 0$ at every $x \in K$. Then the conductance of the walk is at least

$$\frac{\text{Iso}(K)\nu^2\delta}{32\sqrt{n}}.$$

Suppose S, \bar{S} is a partition of K (and we wish to prove that the conductance of S is high). If a point $x \in S$ and a point $y \in \bar{S}$ satisfy

$$|x - y| \leq \frac{\nu\delta}{4\sqrt{n}},$$

then from elementary geometry, the balls of radius δ with x and y as centers intersect “a lot”, i.e.,

$$\text{volume}((B(x, \delta) \setminus B(y, \delta))) \leq (\nu/4)\text{volume}(B(0, \delta)).$$

Letting $\ell(\cdot)$ denote the local conductance, we see then that with probability at least $3\ell(x)/4$, the chain moves from x in one step to a point other than x in the set

$$C = B(x, \delta) \cap B(y, \delta) \cap K.$$

We also get a symmetric statement for y . From this, one can show that either $P(x, C \cap \bar{S})$ or $P(y, C \cap S)$ is at

least $\nu/4$. Let B (for boundary) consist of the set of points in S at distance at most $\frac{\nu\delta}{4\sqrt{n}}$ from \bar{S} as well as the set of points in \bar{S} at at most the same distance from S . Then, we have, using time-reversibility (actually, here we have a symmetric Markov Chain)

$$\begin{aligned} 2\Phi_S &= \frac{\int_{x \in S} P(x, \bar{S})\pi(x)dx + \int_{y \in \bar{S}} P(y, S)\pi(y)dy}{\min(\pi(S), \pi(\bar{S}))} \\ &\geq \frac{\text{volume}B}{\min(\text{volume}(S), \text{volume}(\bar{S}))} \frac{\nu}{4} \geq \text{Iso}(K) \frac{\nu\delta}{4\sqrt{n}} \frac{\nu}{4} \end{aligned}$$

where the last inequality uses basic geometry.

4.2 Log-concave Sampling

Applegate and Kannan [11] observed that the previous algorithms to approximate the volume of convex sets sample uniformly in a convex set, i.e., sample according to the characteristic function of the set K which is 1 on the set and zero outside. This function is terribly discontinuous and if K has “sharp corners”, the discontinuity is severe. In fact, the high complexity (with exponents of n in the double digits) of the previous algorithms was in part due to this. They rectified this by damping the fall from 1 to 0 exponentially. The idea is then that we sample according to a function which falls towards 0 as the distance from the convex set increases. Here is a suitable notion of distance (called the “gauge function”) which they used.

$$\phi_K(x) = \begin{cases} 0 & \text{if } x \in K \\ \inf\{\lambda > 0 : \lambda x \in K\} - 1 & \text{if } x \notin K \end{cases}$$

[For example, if K is the unit ball, ϕ_K is zero in the ball and is $|x| - 1$ outside.] The function F used was

$$F(x) = e^{-\phi_K(x)}.$$

For this, a simple integration over all of space establishes that

$$\int_{\mathbf{R}^n} F(x)dx = \text{volume}(K) s,$$

where s is a constant ($= 1 + n \int_1^\infty t^{n-1} e^{-t} dt$). Thus, it suffices to integrate this log-concave function. [A function is log-concave if it is positive real valued and its logarithm is concave.] They reduced the integration problem to the problem of sampling according to $F(\cdot)$ in a cube. This reduction is a natural divide and conquer algorithm which to integrate the function over a cube, draws samples, figures out the (approximate) ratio of the whole integral to the integral over a part obtained by cutting one of the sides into two equal parts and repeats this process until the cube is small enough so that the the function varies little over the remaining cube.

The problem of sampling according to a function is a standard and basic problem in Statistics. In fact, most standard probability densities like the normal

and the exponential family are log-concave functions. There are natural random walks that have been used to solve the sampling problem in Statistics. (See for example Gelfand and Smith [42] and Geman and Geman [43].) But, there were no provable bounds on the number of steps needed to make say the chi-squared distance to the steady state (which is proportional to F) smaller than a given ϵ . [11] supplied such bounds for a natural coordinate random walk whose steady state is proportional to F . The random walk is a Metropolis version of the coordinate random walk we described earlier to sample according to the uniform. To spell out the walk, we choose first a grid of side δ where $\delta > 0$ as we see later will be so that F varies by at most a constant factor over any cube of side δ . Then the states are the grid points in the cube we want to sample from and the transition probability matrix is given by

$$P_{xy} = \frac{1}{2n} \min\left(1, \frac{F(y)}{F(x)}\right) \forall \text{ adjacent grid points } x, y \in C$$

$$\text{and } P_{xx} = 1 - \sum_y P_{xy}.$$

The steady state probabilities satisfy $\pi(x) = F(x) / \sum_z F(z)$, where the last sum is over all states z . They relate the convergence rate of the chain via conductance to an isoperimetric constant which is now weighted by F . More precisely, the isoperimetric constant of K with respect to a measurable positive valued function F is defined as the infimum over all (measurable) subsets S of K with $0 < \int_S F < \int_K F$ of the ratio :

$$\frac{\int_{\partial S \cap \text{interior}(K)} F}{\min(\int_S F, \int_{K \setminus S} F)}.$$

They showed that using essentially the same proof, that the isoperimetric constant of any convex set with respect to any log-concave function is still at least $1/d$ where d is the diameter of the set. The improvement of Dyer and Frieze [27] of this to $2/d$ applies to this case as well.

Frieze, Kannan and Polson [39] analyze the coordinate random walk to sample according to a log-concave function in detail. They prove bounds on the eigenvalue (paying regard to the constants as well as the order of magnitude) under weaker assumptions than [11]. They also suggest a way of handling small sets near the boundary of the convex set that is different from the one in [67].

Lovász and Simonovits [68] develop a Metropolis version of the ball walk to sample according to a log-concave function. They use this to derive a volume algorithm. We give a description of it here : Let

$$F(\cdot) = e^{-\phi_K(\cdot)},$$

where ϕ_K is the gauge function of the convex set K . We assume that the convex set is well-rounded. Actually, we will assume that

$$B(0, r) \subseteq K \subseteq B(0, R),$$

where we use r, R which later will assume values other than 1, $n^{3/2}$. $B(0, r) \subseteq K$ implies that the function F does not vary by more than a constant factor in any ball of radius r . We will do the ball walk in a ball of radius r . It is also easy to see that the integral of F outside of the ball $B(0, 10Rn \log n)$ is negligible, so we may restrict our attention to this ball. We wish to appeal to Theorem 5 for proving fast convergence. To this end, first from basic geometry, we get that for any point x in $B(0, 10Rn \log n)$, we have $\text{volume}(B(x, r) \cap B(0, 10Rn \log n)) \geq (0.4) \text{volume}(B(x, r))$ and so from x , we will pick with probability at least 0.4, a point in the first set to go to. The Metropolis version then accepts with probability at least $1/\epsilon$. Thus the local conductance of every point is at least $0.4/\epsilon = \Omega(1)$. Thus, we will have a mixing time of (noting that $\text{Iso}(K) \geq 1/d = 1/Rn \log n$)

$$O(n^3 (\log n)^2 \left(\frac{R}{r}\right)^2).$$

The ratio R/r in fact plays a role in the complexity of most algorithms. This has been called the sandwiching ratio.

4.3 The ratio of the outer ball to the inner

All algorithms work by “sandwiching” the convex set after a linear transformation between two balls or as in [11] two cubes, the outer ball (or cube) which contains the convex set is obtained by dilating the inner ball (or cube) which is contained in the convex set about its center by a factor of $f(n)$. In the original papers, the sandwiching ratio $f(n)$ was $n^{3/2}$, it was observed in [11] that it could be improved to $O(n)$ by an algorithm of Lenstra’s [66] if one used cubes instead of balls. [68] (and independently [36]) showed algorithmically that if we do not insist that the outer ball contain the body, but only that it contain most of the set (for concreteness, think of it as containing $3/4$ of the volume of the set), then $f(n)$ can be made $O(n)$ even with balls. [68] show that if a ball of radius u about the origin contains $3/4$ th the volume of a convex set, then for any $1 > \epsilon > 0$, a ball of radius $O(u \log(1/\epsilon))$ contains at least $1 - \epsilon$ of the volume of convex set. For algorithms to approximate the volume, it clearly suffices to have the outer ball contain most of the volume.

More recently, Kannan, Lovász and Simonovits [55] has shown that we can make $f(n)$ $O(\sqrt{n})$ which is the best possible. The linear transformation that accomplishes this is the one that puts the body in the so-called “inertial position” (or also “isotropic position”). We describe this now.

The matrix $I(K)$ of inertias of a set K in \mathbf{R}^n is the $n \times n$ matrix whose i, j th entry is the expectation of $x_i x_j$ over K , i.e.,

$$I(K)_{ij} = \frac{1}{\text{volume}(K)} \int_{x \in K} x_i x_j.$$

[Statisticians call such a matrix the variance-covariance matrix corresponding to the probability

distribution which is uniform on K .] If this matrix is the identity, then we say that K is in inertial position. In general, $I(K)$ is a positive semi definite matrix; we assume it is nonsingular, and so positive definite for our cases. Therefore, by linear algebra, we can write $I(K) = BB^T$ (B^T is the transpose of B .) Applying the linear transformation B^{-1} to the set K puts it in inertial position, as is easy to show. For a convex set K in inertial position, they argue that the unit ball is fully contained in K by resorting to a reduction to the needle-like case and also that $3/4$ th the volume of K is contained in a ball of radius $O(\sqrt{n})$. The latter fact is simple : just observe that by the definition of inertial position, we have that the expected length squared of a vector in K is $O(n)$ and then use Markov inequality.

That $O(\sqrt{n})$ is the best possible follows from the example of a cube. The question of how to find $I(K)$ accurately enough so that we have approximately the right sandwiching ratio is in a sense open : they indicate how to do this with $O(n^2)$ samples drawn from K , but conjecture that $O^*(n)$ samples would do. The reader may be puzzled by the fact that we are suggesting using the inertial position to solve the sampling problem fast, but seem also to need samples to get the matrix $I(K)$. The point is that the quality of samples (actually, the chi-squared distance to the uniform of the density we are actually drawing samples from) needed to find the inertial position approximately need not be as high as that of the samples to approximate the volume. This is exploited by [54].

4.4 Contingency Tables

There is an interesting application of sampling from convex sets to a problem that arises in Combinatorics and Statistics (see Diaconis and Gangolli [24]) The problem is the following : given a set of $m + n$ positive integers $r_1, r_2, \dots, r_m, c_1, c_2, \dots, c_n$, count approximately the number of $m \times n$ matrices of nonnegative integers with row sums r_1, r_2, \dots, r_m respectively and column sums c_1, c_2, \dots, c_n respectively. This is polynomial time reducible to the problem of sampling (nearly) uniformly from the set of such matrices. There have been several algorithms to tackle these problems, see for example Mehta and Patel [72], Baglivo, Oliver and Pagano [13]; and Gangolli [41] for a survey. Most recently, Dyer, Kannan and Mount [34] have developed a provably polynomial time algorithm using the technique for sampling from convex sets for a subcase of the general problem. We describe their approach after discussing the motivation for the problem.

The importance of this problem to the analysis of contingency tables in Statistics was developed by the influential paper of Diaconis and Efron [23]. We will only say a few words about this : suppose we have two properties, for example, eye and hair color (of people) and want to test the hypothesis that these two properties are correlated. Let us suppose there are m hair colors and n eye colors and by sampling from the population in question we have data consisting an $m \times n$ matrix T whose i, j th entry T_{ij} is the number of people in the sample that have the i th hair color and the j th eye color. If the row sums in the table are

r_1, r_2, \dots, r_m and column sums c_1, c_2, \dots, c_n , then if the properties were totally independent, the i, j th entry of the table would be $r_i c_j / N$ where $N = \sum r_i = \sum c_j$ is the total population. We may measure the deviation of the actual T from this set of entries in one of many ways to obtain a numerical value of the deviation. The problem is that a priori, we do not have any way to tell whether this numerical value is “high”, so the properties are correlated or the value is “low” and the properties are uncorrelated because there is no benchmark to compare the value against. Diaconis and Efron [23] proposed the following : sample uniformly from the set of possible tables with the same row and column sums (i.e., sample from the set of possible populations with the same number of people with each hair and eye color) and see what proportion of these tables have a higher numerical value than the current table and base the decision on this. Clearly this leaves us with the random sampling problem we described.

The sampling problem can be posed as sampling uniformly integer points from the polytope $P(r, c)$ in mn dimensional space (where we index the coordinates of a point x in this space by x_{ij}) :

$$P(r, c) = \{x : \sum_i x_{ij} = c_j \forall j; \sum_j x_{ij} = r_i \forall i; x_{ij} \geq 0\}.$$

We of course know how to sample from the polytope with uniform density. We then get a continuous point and the hope that “rounding” it to an integer point will solve our problem. The rounding has to be done in a suitable way so as not to violate row and column sums. Also, we have to enlarge the polytope so that each integer point has the same volume of the region from which it will be the rounded point.

Under the condition that the row and column sums are each at least $(n + m)nm$, the paper shows that we can sample from the integer points in $P(r, c)$ in polynomial time. These lower bounds seem to be satisfied for several practical situations and the algorithm in its preliminary implementation seems to be useful. (See Diaconis and Gangolli for a discussion [24].) The central part is to sample with uniform density from the polytope $P(r, c)$ for which the paper studies the detailed structure of the polytope and derives a better algorithm than the general one of sampling from any convex set. The algorithm uses a coordinate random walk which is more suitable here, since updates are very quick.

The general case without the restrictions is still open. See a related open problem.

The paper also shows the #P hardness of the problem of counting exactly the number of tables of nonnegative integers with given row and column sums, even in the case when there are only two rows.

There is also some work on algorithms and implementations for the exact counting problem for fixed dimensions (fixed m, n). In this case a polynomial time algorithm exists as a special case of Barvinok’s algorithm for counting the number of integer point in any polytope. It is also possible to decompose the space of possible row and column sums into “chambers” so

that in each chamber, the number of tables with the row and column sums is a polynomial function of the row and column sums [22, 85].

One may also consider 3 and higher dimensional tables and certain generalizations. For these problems, it is nontrivial even to come up with Markov Chains that have the correct steady state probabilities; Diaconis and Sturmfels [26] have done that by appealing to some algebraic techniques. No proofs of rapid mixing are known in this general setting.

4.5 Optimization

The ability to sample randomly has been used by Kannan, Mount and Tayur [56] to solve certain convex programming problems, namely the approximate minimization of linear functions over up-monotone convex sets described only by a membership oracle. We explain the terms in the sentence above. A set is up-monotone if x is in the set implies that any y which is coordinate wise greater than or equal to x is also in the set. In the case of polytopes described by $\{x : Ax \geq b\}$, this happens iff the matrix A has only nonnegative entries. We assume that we are given a membership oracle for the set as well as a point in the set. This plus up-monotonicity can be used by the methods of Yudin and Nemirowski (see [45]) to produce in polynomial time what is called a separation oracle. This last is a black box which given a point x in space, either says it is in the convex set or gives a separating hyperplane which has x on one side and the convex set on the other. However, each call to the separation oracles causes $\Omega(n^4)$ calls to the membership oracle. Since most optimization methods like the ellipsoid method as well as gradient descent require a separation oracle, this is expensive if what is readily available is only a membership oracle. Such a situation arises in Stochastic Optimization problems (see for example Wets [90]).

The algorithm of [56] is in a sense akin to simulated annealing. We impose a "weight" function

$$F(x) = e^{-M \cdot c \cdot x} \quad M \text{ a positive real number,}$$

on the feasible set K . [Other functions may also be used.] This function obviously weighs the near optimal region heavily. If we then sample according to this log-concave function, and if M is sufficiently large, then we hope to get a nearly optimal point with high probability. However, calculations show that the choice of M has to be based on knowing the optimal value for this simple scheme to work. So instead, the paper uses an adaptive algorithm that works in stages. At the beginning of each stage, we have a lower bound L on the optimal value and a feasible solution x^f which of course gives an upper bound on the optimal value. The stage cuts down the difference between these bounds (called the gap) by a factor of 2/3 rd, thus making geometric progress. This is done by choosing

$$M = O\left(\frac{1}{c \cdot x^f - L}\right),$$

whence an argument shows that if the optimal value is at most $L + (1/3)(c \cdot x^f - L)$, then the probability measure (according to F) of the set of feasible points with

$c \cdot x \leq L + (2/3)(c \cdot x^f - L)$ is at least 1/2 of the measure of the set K ; thus under this hypothesis, we would find a feasible point x with $c \cdot x \leq L + (2/3)(c \cdot x^f - L)$ (called the "tip" in the paper) with high probability after a certain number of samples according to F are drawn. If we fail to find such a point, we may with high confidence increase L to $L + (1/3)(c \cdot x^f - L)$. In any case, the gap is narrowed.

The complexity of this algorithm has lately been further improved by the author by the following idea which may be of more general use in simulated annealing type situations : if we increase M by a constant factor more than what the previous paragraph calls for, then we can ensure that the tip has measure $1 - e^{-n}$ times the measure of K . This then raises the question of whether we can argue a faster convergence time for this problem of visiting a "haystack" starting at a "needle" attached to it. The answer is yes, and this is based on a strengthening of the isoperimetric lemma which yields sharper bounds when the minimum measure of the two sets in the partition is small. Here is the precise new theorem.

Theorem 6 : Suppose K is a convex set in \mathbf{R}^n of diameter d and $F : K \rightarrow \mathbf{R}_+$ is a log concave function. For any measurable $S \subseteq K$ with $\int_S F \leq \int_K F/2$, we have,

$$\int_{\partial S \cap \text{interior } K} F \geq \frac{1}{d} \log_2 \left(\frac{\int_K F}{\int_S F} \right) \int_S F.$$

4.6 Other geometric problems, algorithms

Khachiyan and Karzanov [63] develop a faster (than the general algorithm) for computing the volume of order polytopes (discussed above). Matthews [70] gives a coupling argument for the rapid mixing of their chain.

Dyer and Frieze [29] show by proving a polynomial upper bound on the time taken by a certain randomized algorithm that polytopes defined by totally unimodular constraint matrices have combinatorial diameter bounded above by a polynomial.

There are other random walks that can be used to sample from convex sets. One class of them called the hit and run algorithms work as follows ([16],[84]) : from the current point, pick randomly a line going through the point (really pick a point of the unit ball centered at the point and take the line joining the two) and pick a random point on the line segment obtained by intersecting this line (extending in both directions) with the convex set. This has as the steady state the uniform density. We can also have Metropolis versions of this. Also, we can have a coordinate version where we pick a coordinate line each with probability $1/n$. These walks have been studied. It is not difficult to establish that these walks are no slower (really at most a constant factor slower) than the coordinate walk described in this paper (see [33],[12]). However, one might suspect that they are a lot faster, because they cover large distances in one step. It would be very nice to prove that in fact they converge faster. The key seems to be to come up with some nice geometric

property like isoperimetry that reflects the convergence rate.

The problem of choosing at random a graph with weights satisfying the triangle inequality has been studied empirically. (see [77], [71])

5 Open Problems

5.1 Combinatorial problems

1) The problem of developing an approximation algorithm for the general (not necessarily dense) 0-1 permanent discussed in the text is an important one.

2) The reliability problem is the following : given a graph and a probability $p(e)$ of failure of each edge (independently), determine the probabilities of there being an s-t path and the probability of the graph being connected. The problem is of great practical importance and there has been a lot of work on it. [The exact version is #P hard.] The approximate version is reducible to the corresponding approximate sampling problem. A proof of rapid mixing (for a Markov Chain to do the sampling) even for the case when all $p(e) = 1/2$ (when the problem reduces to sampling uniformly from the set of s-t connected, respectively connected) subgraphs would of interest.

3) Counting the number of feasible solutions to a 0-1 knapsack problem approximately in polynomial time is open.

5.2 Geometric Problems

1) The following conjecture on isoperimetry is formulated in [55]. There is an absolute constant $c > 0$ independent of n such that for any convex set K , there is a hyperplane H (dividing space into two parts H^+ , H^- such that

$$\frac{\text{volume}_{n-1}(H \cap K)}{\min(\text{volume}(H^+ \cap K), \text{volume}(H^- \cap K))} \leq c \text{ Iso}(K).$$

In words, the conjecture says that among all partitions of a convex set into two parts, there is a hyperplane partition which (within a constant) is the worst for isoperimetry. While, this seems to be an interesting geometry problem in itself, see below for the algorithmic significance.

2) The problem of how many samples suffice to find the inertial matrix accurately enough so that its inverse can be approximately found is open. [The inverse is needed to put the convex set in inertial position.] The conjecture is that $O^*(n)$ samples would do.

The previous conjecture on isoperimetry and this one together give an improvement of $O^*(n)$ in the volume algorithm.

3) The problem of proving faster mixing times for hit and run algorithms (see section 4.6) is theoretically and practically interesting.

4) Faster algorithms for the contingency tables problem would be of great practical interest. For the optimization problem, it would be good to extend the method to general linear programming (i.e., not assume up-monotonicity) as well as improve the running times.

5) It would be nice to drop the conditions on the row and column sums of contingency tables under which we now know polynomial time sampling algorithms. Much of this may be accomplished if we can develop a unary (in the row and column sums) polynomial time algorithm. There is a nice Markov Chain which is conjectured to mix in unary polynomial time for this (see for example [24]) : from any current table, X , pick (uniformly) at random two rows $i < j$ and two columns $k < l$ and go to X' obtained by changing 4 entries of X as given below if it has nonnegative entries; otherwise, stay at X :

$$X'_{i,k} = X_{i,k} + 1; X'_{i,l} = X_{i,l} - 1; X'_{j,k} = X_{j,k} - 1; X'_{j,l} = X_{j,l} + 1.$$

Acknowledgement : Many thanks to Alan Frieze and John Mount for reading the manuscript and suggesting many improvements.

References

- [1] D. Aldous, "Random walks on finite groups and rapidly mixing Markov chains," *Séminaire de Probabilités XVII*, Springer Lecture Notes in Mathematics 986, pp. 243-297, 1981/82.
- [2] D. Aldous, "Some inequalities for reversible Markov chains," *Journal of the London Mathematical Society (2)*25, pp. 564-576, 1982.
- [3] D. Aldous, "On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing," *Probability in the Engineering and Informational Sciences* 1, pp. 33-46, 1987.
- [4] D. Aldous, "The random walk construction for spanning trees and uniform labeled trees," *SIAM Journal on Discrete Mathematics* 3, pp. 450-465, 1990.
- [5] D. Aldous, P. Diaconis, "Shuffling cards and stopping times," *American Mathematical Monthly* 93, pp. 333-348, 1986.
- [6] R. Aleliunas, R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff, "Random walks, universal traversal sequences and the complexity of the maze problem", *Proceedings of the 20 th Foundations of Computer Science*, pp 218-223, 1979.
- [7] N. Alon, "Eigenvalues and expanders", *Combinatorica*, 6 pp83-96, (1986).
- [8] N. Alon, A. M. Frieze, and D.J.A. Welsh, "Polynomial time randomised approximation schemes for the Tutte polynomial of dense graphs", In these proceedings.
- [9] N. Alon, V. d. Milman, " λ_1 , isoperimetric inequalities for graphs and superconcentrators," *Journal of Combinatorial Theory Series B* 38, pp. 73-88, 1985.

- [10] J. D. Annan, "A randomised approximation algorithm for counting the number of forests in dense graphs," *Combinatorics, Probability and Computing*, (to appear)(1994) .
- [11] D. Applegate, R. Kannan, "Sampling and integration of near log-concave functions," *Proceedings of the 23rd ACM Symposium on Theory of Computing*, ACM Press, pp. 156-163, 1991.
- [12] D. Applegate, R. Kannan and N. Polson, "Random polynomial time algorithms for sampling from joint distributions", Technical Report # 500, Department of Statistics, Carnegie-Mellon University, 1990.
- [13] J. Baglivo, D. Oliver, M. Pagano, "Methods for Exact Goodness-of-Fit Tests," *Journal of the American Statistical Association* 87, pp. 464-469, 1992.
- [14] I. Bárány and Z. Füredi, "Computing the volume is difficult" *Proc. of the 18 th ACM Symposium on the Theory of Computing*, pp. 442-447, 1986.
- [15] A. I. Barvinok, "A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed", *Proceedings of the Symposium on the Foundations of Computer Science*, 1993.
- [16] H.C.P. Berbie, C.G.E. Boender, A.H.G. Rinooy Kan, C.L. Schaffer and R.L. Smith and J. Telgen, "Hit and run algorithms for the identification of nonredundent linear inequalities", *Math. Programming*, **37** , (1987) pp 184-207.
- [17] G. Brightwell, P. Winkler, "Counting linear extensions is $\sharp P$ -complete," *Proc. 23rd Symp. Th. Comp.*, pp. 175-182, 1991.
- [18] A.Z. Broder, "Generating random spanning trees," *30th IEEE Symposium on Foundations of Computer Science*, pp. 442-447, 1989.
- [19] 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*, pp. 50-88, 1986 (Erratum in *Proceedings of the 20th ACM Symposium on Theory of Computing*, 1988, p. 551).
- [20] J. Cheeger, "A lower bound for the smallest eigenvalue for the Laplacian," *Problems in Analysis* (R. C. Gunning, ed.), Princeton University Press, pp. 195-199, 1970.
- [21] F. R. K. Chung and Yau, Manuscript, 1994.
- [22] W. Dahmen, C.A. Micchelli, "The number of solutions to linear diophantine equations and multivariate splines," *Transactions Amer. Math. Soc.* **308**, 509-532, 1988.
- [23] P. Diaconis, B. Efron, "Testing for independence in a two-way table", *Annals of Statistics*, **13**, pp 845-913, 1985.
- [24] P. Diaconis, and A. Gangolli, "Rectangular arrays with fixed margins", To appear in the *Proceedings of the workshop on Markov Chains*, Institute of Mathematics and its Applications, University of Minnesota, (1994).
- [25] P. Diaconis, D. Stroock, "Geometric bound for eigenvalues of Markov chains," *Annals of Applied Probability* 1, pp. 36-61, 1991.
- [26] P. Diaconis, and B. Sturmfels, "Algebraic algorithms for sampling from conditional distributions", Technical Report, Department of Statistics, Stanford University (1993)
- [27] M. E. Dyer, A. Frieze, "Computing the volume of convex bodies: a case where randomness provably helps", *Probabilistic Combinatorics and its Applications*, Proceedings of AMS Symposia in Applied Mathematics 44, pp. 123-170, 1991.
- [28] M. E. Dyer, A. Frieze, "On the complexity of computing the volume of a polyhedron," *SIAM J. Comput.* 17, 967-974.
- [29] M. E. Dyer, A. Frieze, "Random walks, totally unimodular matrices and a randomised dual simplex algorithm," *Mathematical Programming* 64, pp. 1-16.
- [30] M. E. Dyer, A. Frieze, M. Jerrum, *Approximately Counting Hamilton Cycles in Dense Graphs*, Report ECS-LFCS-93-259, Department of Computer Science, University of Edinburgh, April 1993.
- [31] M. E. Dyer, A. Frieze, R. Kannan, "A random polynomial time algorithm for approximating the volume of convex bodies," *Journal of the ACM* 38, pp. 1-17, 1991.
- [32] M. E. Dyer, A. Frieze, R. Kannan, A. Kapoor, L. Perkovic, U. Vazirani, "A mildly exponential time algorithm for approximating the number of solutions to a multidimensional knapsack problem," *Combinatorics, Probability and Computing*, (to appear).
- [33] M. E. Dyer, A. M. Frieze and L. Stoogie, "Convergence of Hit and Run algorithms", Private Communication.
- [34] M. Dyer, R. Kannan and J. Mount, "Generating random contingency tables", Manuscript.
- [35] G. Elekes, "A geometric inequality and the complexity of computing the volume", *Discrete and Computational Geometry*, **1**, pp 289-292, 1986.
- [36] U. Faigle, N. Gademann and W. Kern, "Randomized rounding", Preprint.
- [37] J. A. Fill, "Eigenvalue bounds on the convergence to stationarity for nonreversible Markov Chains, with an application to the exclusion process", *The Annals of Applied Probability*, Vol. 1, No. 1, pp 62-87, 1991.

- [38] T. Feder, M. Mihail, "Balanced matroids," *Proceedings of the 24th ACM Symposium on Theory of Computing*, pp. 26-38, 1992.
- [39] A. Frieze, R. Kannan, and N. Polson, "Sampling from log-concave distributions" To appear in *Annals of Applied Probability*.
- [40] A. Frieze, S. Suen, "Counting the number of Hamiltonian cycles in random digraphs," *Random Structures and Algorithms* 3, pp. 235-241, 1992.
- [41] A. R. Gangolli, *Convergence Bounds for Markov Chains and Applications to Sampling*, Phd. Thesis, Stanford STAN-CS-91-1361, 1991.
- [42] A. E. Gelfand and A. F. M. Smith, "Sampling based approaches to calculating marginal densities", *J. of the American Statistical Association*, **85**, pp 398-409, 1990.
- [43] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp 721-741 , 1984.
- [44] D. Gillman, "A Chernoff bound for random walks on expander graphs," *Proceedings of the 34th Annual IEEE Conference on Foundations of Computer Science*, 1993.
- [45] M. Grötschel, L. Lovász and A. Schrijver, "Geometric Algorithms and Combinatorial Optimization", Springer-Verlag, New York, 1988.
- [46] M. Jerrum, "A very simple algorithm for counting the number of k -colorings of low degree graphs", Manuscript.
- [47] M. Jerrum, "An analysis of a Monte Carlo algorithm for estimating the permanent," *Proceedings of the 3rd Conference on Integer Programming and Combinatorial Optimization*, CORE, Lovain-la-Neuve, Belgium, pp. 171-182, 1993.
- [48] M. R. Jerrum, A. Sinclair, "Approximating the permanent," *SIAM J. Comp.* 18, pp. 1149-1178, 1989.
- [49] M. R. Jerrum, A. Sinclair, "Fast Uniform Generation of Regular Graphs," *Theoretical Computer Science* 73, pp. 91-100, 1990.
- [50] M. R. Jerrum, A. Sinclair, "Polynomial-time approximation algorithms for the Ising model," *SIAM Journal on Computing* (to appear).
- [51] M. R. Jerrum, L.G. Valiant, V.V. Vazirani, "Random generation of combinatorial structures from a uniform distribution," *Theor. Comp. Sci.* 43, pp. 169-188, 1986.
- [52] M. Jerrum, U. Vazirani, "A mildly exponential approximation algorithm for the permanent," *Proceedings of the 33rd Annual IEEE Conference on Foundations of Computer Science*, pp. 320-326, 1992.
- [53] N. Kahale, "Large deviation bounds on Markov Chains", DIMACS Technical Report, 94-39, 1994.
- [54] R. Kannan, L. Lovász and M. Simonovits, "Random walks and a faster algorithm for volumes of convex sets", Manuscript.
- [55] R. Kannan, L. Lovász and M. Simonovits, "Isoperimetric problems for convex set", To appear in *Discrete and Computational Geometry*.
- [56] R. Kannan, J. Mount and S. Tayur, "A randomized algorithm to optimize over certain convex sets", To appear in *Mathematics of Operations Research*.
- [57] S. Kannan, Z. Sweedyk, S. Mahaney, "Counting and Random Generation of Strings in Regular Languages," *preprint*, 1994.
- [58] N. Karmarkar, R. Karp, R. Lipton, L. Lovász, M. Luby, "A Monte-Carlo algorithm for estimating the permanent," *SIAM Journal on Computing* 22, pp. 284-293, 1993.
- [59] R. M. Karp, M. Luby, "Monte-Carlo algorithms for enumeration and reliability problems," *Proceedings of the 24th IEEE Symposium on Foundations of Computer Science*, pp. 56-64, 1983.
- [60] M. Karpinski, M. Luby, "Approximating the number of solutions of a GF(2) polynomial" *preprint*, 1989.
- [61] L. G. Khachiyan, "A polynomial time algorithm in Linear Programming", (in Russian) *Doklady Akademii Nauk SSSR* 244 pp 1093-1096 1979. [English translation : Soviet Mathematics Doklady 20 pp 191-194 1979].
- [62] L. G. Khachiyan, "The problem of computing the volume of polytopes is NP-hard", *Uspekhi Mat. Nauk* 44 pp 199-200, 1989.
- [63] L. Khachiyan, and A. Karzanov, *On the conductance of order markov chains*, Technical Report DCS 268, Rutgers University, June 1990.
- [64] G. F. Lawler, A. D. Sokal, "Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality," *Transactions of the American Mathematical Society* 309, pp. 557-580, 1988.
- [65] T. Leighton and S. Rao, "An approximate max-flow min-cut theorem for the uniform multicommodity flow problems with applications to approximation algorithms", *Proceedings of the 29th Symposium on Foundations of Computer Science* pp422-431 , 1988.

- [66] H. W. Lenstra, "Integer programming with a fixed number of variables", *Operations Research*, **8**, pp 538-548 9183.
- [67] L. Lovász and M. Simonovits "Mixing rate of Markov chains, an isoperimetric inequality, and computing the volume", *Proc. 31st Symposium on Foundations of Computer Science*, pp 346-355 (1990).
- [68] L. Lovász and M. Simonovits, "Random walks in a convex body and an improved volume algorithm" *Random Structures and Alg.* **4**, pp 359-412 1993.
- [69] L. Lovász and M. Simonovits "On the randomized complexity of volume and diameter", *Proc. 33rd FOCS*, pp 482-491 1992.
- [70] P. Matthews, "Generating random linear extensions of a partial order," *The Annals of Probability* **19**, pp.
- [71] C. McGeoch, "Hit and run algorithms for generating clustered distributions", Manuscript
- [72] C. R. Mehta, N. R. Patel, "A Network Algorithm for Performing Fisher's Exact Test in $r \times c$ Contingency Tables," *Journal of the American Statistical Association* **78**, pp. 427-434, 1983.
- [73] N. Metropolis, A.W. Rosenbluth, M. N. Rosenbluth, A.H. Teller and E. Teller, "Equation of state calculation by fast computing machines", *Journal of Chemical Physics*, **21**, pp 1087-1091, 1953.
- [74] M. Mihail, "Combinatorial aspects of expanders", Ph.D. Dissertation, Dept. of Computer Science, Harvard University, 1989.
- [75] M. Mihail, "On coupling and the approximation of the permanent," *Information Processing Letters* **30**, pp. 91-95, 1989.
- [76] M. Mihail, P. Winkler, *On the number of Eulerian orientations of a graph*, Bellcore Technical Memorandum TM-ARH-018829, 1991.
- [77] M. J. Muller, "Generating random graphs that obey the triangle inequality : a random graph approach", Department of Mathematics and computer Science, Amherst College, April 1990.
- [78] L. E. Payne and H. F. Weinberger, "An optimal Poincaré inequality for convex domains", *Archives of Rational Mechanics and Analysis*, **5** , pp 286-292 1960.
- [79] J. Radcliffe, "A class of measures admitting an isoperimetric inequality" Private Communication, 1989.
- [80] E. Seneta, "Non-negative matrices and Markov Chains" *Springer-Verlag* , New York, 1973.
- [81] A. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatorics, Probability and Computing* **1**, pp. 351-370, 1992.
- [82] A. J. Sinclair, M.R. Jerrum, "Approximate counting, uniform generation and rapidly mixing Markov chains," *Information and Computation* **82**, pp. 93-133, 1989.
- [83] M. Sohoni and H. Naarayanan, "On rapid mixing of matroids", Manuscript.
- [84] R.L. Smith, "Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions", *Operations Research*, **32** (1984) pp 1296-1308.
- [85] B. Sturmfels, "On Vector Partition Functions," *manuscript*, Feb. 1994.
- [86] R. H. Swendsen, J.S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical Review Letters* **58**, pp. 86-88, 1987.
- [87] L. G. Valiant, "The complexity of computing the permanent," *Theoretical Computer Science* **8**, pp. 189-201, 1979.
- [88] D.J.A. Welsh, "Matroid Theory", *Academic Press*, London, 1976.
- [89] D.J.A Welsh, *Complexity: Knots, Colourings and Counting*, London Mathematical Society Lecture Note Series **186**, Cambridge University Press, 1993.
- [90] R. Wets, *Stochastic Programming* in Handbooks in Operations Research and Management Science, Vol. 1, North Holland, 1989