DIMACS Technical Report 94-39
June 1994

Large deviation bounds for Markov chains

by

Nabil Kahale [1]

[1]DIMACS, Rutgers University, Piscataway, NJ 08855. Email: kahale@dimacs.rutgers.edu. A part of this work was done while the author was at the Massachusetts Institute of Technology.

DIMACS is a cooperative project of Rutgers University, Princeton University, AT&T Bell Laboratories and Bellcore.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

We study the fraction of time that a Markov chain spends in a given subset of states. We give an exponential bound on the probability that it exceeds its expectation by a constant factor. Our bound depends on the mixing properties of the chain, and is asymptotically optimal. It beats the best previously known results in this direction. We present an application to the leader election problem.

1 Introduction

Let $(X_m), m \geq 1$, be an irreducible Markov chain on a finite state space V with transition matrix P, initial distribution q, and stationary distribution π. We assume that P is reversible, that is

$$\pi(u)P(u,v) = \pi(v)P(v,u), \text{ for all } u, v \in V. \tag{1}$$

Let A be a proper subset of V. Denote by $\pi(A) = \sum_{u \in A} \pi(u)$ the stationary probability of the set A, and by $S_l = \chi_A(X_1) + \chi_A(X_2) + \cdots + \chi_A(X_l)$ the number of steps the Markov chain is inside A. It is known [3] that, for any initial distribution, the fraction S_l/l converges almost surely to $\pi(A)$ as l goes to infinity. This lead Aldous [2] to propose the following sampling technique: $\pi(A)$ can be estimated by simulating the Markov chain for l steps and computing the fraction S_l/l of steps it spends in A. Typically, the size of A is exponential in the input size (e.g. A is the set of matchings of a given size of a graph), and thus $\pi(A)$ cannot be computed directly in polynomial time. It is therefore important to establish a bound on the probability that S_l/l exceeds $\pi(A)$ by a given amount. A bound on the variance of S_l in terms of the mixing properties of the chain was established in [2, 12]. An exponential bound on the tail of S_l/l was established in [9] in a special case, and in a more general setting in [5].

In this paper, we establish a bound on the tail of the distribution of S_l/l that beats the previously known bounds. As the previous bounds, ours depends on the second largest eigenvalue of P. We also show that for given values of β, $\pi(A)$ and of the second largest eigenvalue of P, the decay rate in our bound is optimal. In particular, as the second largest eigenvalue of P approaches 0, the decay rate in our bound tends towards the decay rate in the standard Chernoff bound for i.i.d. Bernoulli random variables. Our bound is substantially better than the one in [5] when $\pi(A)$ is small, which is the case in several applications.

As in [4, 5, 8], the proof of our main result relies on computing an upper bound on the generating function of S_l. The main new idea in our paper is to reduce the analysis to the case where all the eigenvalues of P, but the largest one, are equal. This case can further be reduced to the case where the state space consists of two elements.

Random walks have been used in many areas of Computer Science, such as approximation algorithms [10, 15] (See also [12], and references therein), complexity and cryptography [1, 6, 9], and distributed computing [13].

The rest of the paper is organized as follows. Section 2 contains basic results and definitions. Section 3 contains the proof of our main result and an application to the leader election problem. We establish more explicit but weaker bounds in Section 4. We also give a bound on the probability that the Markov chain stays inside A for l steps, generalizing a result in [11]. In Section 5, we show the tightness of our main result.

2 Preliminaries

Let $L^2(\pi)$ be the set of real valued functions on V, with the scalar product $<,>_\pi$:

$$< f, g >_\pi = \sum_{x \in V} \pi(x) f(x) g(x).$$

For $f \in L^2(\pi)$, let $||f||_\pi = \sqrt{< f, f >_\pi}$. In matrix notation, $< f, g >_\pi = f^t \Pi g$, where Π is the diagonal matrix $(\pi(x))_{x \in V}$. We will identify the matrix P with the operator in $L^2(\pi)$ that associates to each function $f \in L^2(\pi)$ the function $Pf \in L^2(\pi)$, where

$$(Pf)(x) = \sum_{y \in V} P(x, y) f(y).$$

Eq. 1 shows that P is a self-adjoint operator in $L^2(\pi)$, i.e. $< Pf, g >_\pi = < f, Pg >_\pi$, for all $f, g \in L^2(\pi)$. Hence P is diagonalizable in an orthonormal (with respect to the scalar product $<,>_\pi$) basis $(e_0, e_1, \ldots, e_{n-1})$, where $n = |V|$. The first element e_0 of the basis can be chosen to be the all ones column vector $\mathbf{1}$, since $P\mathbf{1} = \mathbf{1}$. The theory of non-negative matrices [14] shows that 1 is the largest eigenvalue of P is absolute value. Denote by λ_i the eigenvalue corresponding to e_i, for $1 \leq i \leq n-1$. Assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1}$. Since the chain is irreducible, $\lambda_1 < 1$. We will assume throughout the paper that $\lambda_1 \geq 0$. Although this assumption does not seem to be indispensable, it simplifies the proof and statement of the results and it holds in most applications.

A Markov chain that arises in many applications is the random walk on a graph. At each step, given that X_t is at node v of degree $d(v)$, X_{t+1} goes to any neighbor of v with probability $1/d(v)$. This chain is reversible and $\pi(v) = d(v)/2|E|$.

Note that we do not impose any condition on λ_{n-1} for our results to be valid. In particular, the chain can be a random walk on a bipartite graph, in which case $\lambda_{n-1} = -1$. We denote by $\lambda_0(M)$ the largest eigenvalue of any matrix M whose eigenvalues are real.

3 Proof of main result

In this section, we give an upper bound on the probability that S_l exceeds $\beta\pi(A)l$, for any real number β such that $1 < \beta < 1/\pi(A)$. Note that if $\beta > 1/\pi(A)$ then S_l never exceeds $\beta\pi(A)l$. The case $\beta = 1/\pi(A)$ is treated in Theorem 4.3. Let $a = \lambda_1 + \pi(A) - \lambda_1\pi(A)$ and $b = 1 - \pi(A) + \lambda_1\pi(A)$.

Theorem 3.1 *For any integer $l \geq 1$ and $1 < \beta < 1/\pi(A)$,*

$$Pr[S_l \geq \beta\pi(A)l] \leq c_{\beta,\pi(A),\lambda_1} \sqrt{\sum_{v \in V} \frac{q(v)^2}{\pi(v)}}\ \alpha^l,$$

where

$$\alpha = \min_{\mu \geq 1} \mu \left(\frac{a\mu - \lambda_1}{\mu^2 - b\mu} \right)^{\beta\pi(A)} < 1,$$

and $c_{\beta,\pi(A),\lambda_1}$ is a constant that depends only on β, $\pi(A)$ and λ_1.

Before proving Theorem 3.1, we prove the following three lemmas. Let D_θ be the $n \times n$ diagonal matrix whose entry (v,v) equals $e^{\theta\chi_A(v)}$, and $d_\theta = D_\theta \mathbf{1}$. Similarly, $\sqrt{d_\theta} = \sqrt{D_\theta}\mathbf{1}$.

Lemma 3.2 *For any $\theta \geq 0$,*

$$E\left[e^{\theta S_l}\right] \leq e^{\theta} ||\Pi^{-1} q||_\pi \lambda_0 (D_\theta P)^{l-1}.$$

Proof

$$\begin{aligned}
\mathrm{E}\left[e^{\theta S_l}\right] &= \sum_{v_1,v_2,\ldots,v_l \in V} \Pr[X_1 = v_1] P(v_1,v_2)P(v_2,v_3)\cdots P(v_{l-1},v_l) e^{\theta\chi_A(v_1)} e^{\theta\chi_A(v_2)} \ldots e^{\theta\chi_A(v_l)} \\
&= q^t (D_\theta P)^{l-1} d_\theta \\
&= < \sqrt{D_\theta}\Pi^{-1} q, (\sqrt{D_\theta} P \sqrt{D_\theta})^{l-1} \sqrt{d_\theta} >_\pi \\
&\leq ||\sqrt{D_\theta}\Pi^{-1} q||_\pi \; ||(\sqrt{D_\theta} P \sqrt{D_\theta})^{l-1} \sqrt{d_\theta}||_\pi \\
&\leq ||\sqrt{D_\theta}\Pi^{-1} q||_\pi \; \lambda_0(\sqrt{D_\theta} P \sqrt{D_\theta})^{l-1} \; ||\sqrt{d_\theta}||_\pi \\
&\leq e^{\theta} ||\Pi^{-1} q||_\pi \; \lambda_0 (D_\theta P)^{l-1}.
\end{aligned}$$

The fifth equation follows from the fact that the operator $\sqrt{D_\theta} P \sqrt{D_\theta}$ is self-adjoint with respect to the scalar product $<,>_\pi$, and that, from the theory of non-negative matrices [14, Theorem 1.1], $\lambda_0(\sqrt{D_\theta} P \sqrt{D_\theta})$ is the largest eigenvalue of $\sqrt{D_\theta} P \sqrt{D_\theta}$ in *absolute value*. The sixth equation follows from the fact that $\sqrt{D_\theta} P \sqrt{D_\theta}$ and $D_\theta P$ are congruent matrices, and thus have same eigenvalues. □

For two $n \times n$ matrices A and B, we say that $A \leq_\pi B$ if $< f, Af >_\pi \leq < f, Bf >_\pi$, for any $f \in L^2(\pi)$.

Lemma 3.3 *Let $P' = \lambda_1 I + (1-\lambda_1)\mathbf{1}\pi^t$. Then $\lambda_0(D_\theta P) \leq \lambda_0(D_\theta P')$.*

Proof Note that $P'\mathbf{1} = \mathbf{1}$ and $P'f = \lambda_1 f$ if $< f, \mathbf{1} >_\pi = 0$. Any element f of $L^2(\pi)$ can be written in the form $\sum_{i=0}^{n-1} x_i e_i$, and so

$$\begin{aligned}
< f, Pf >_\pi &= \sum_{i=0}^{n-1} \lambda_i {x_i}^2 \qquad (2) \\
&\leq {x_0}^2 + \lambda_1 \sum_{i=1}^{n-1} {x_i}^2 \\
&= < f, P'f >_\pi .
\end{aligned}$$

Thus $P \leq_\pi P'$ which, since $\sqrt{D_\theta}$ is self-adjoint in $L^2(\pi)$, implies that $\sqrt{D_\theta} P \sqrt{D_\theta} \leq_\pi \sqrt{D_\theta} P' \sqrt{D_\theta}$. Since $\lambda_0(A) = \max_{f \neq 0} < f, Af >_\pi / ||f||_\pi^2$, for any self-adjoint operator A,

we conclude that $\lambda_0(\sqrt{D_\theta}P\sqrt{D_\theta}) \leq \lambda_0(\sqrt{D_\theta}P'\sqrt{D_\theta})$. This is equivalent to the inequality $\lambda_0(D_\theta P) \leq \lambda_0(D_\theta P')$. □

We now compute the largest eigenvalue of $D_\theta P'$ by reduction to the case where the state space consists of two states. Indeed, let $Q_{a,b} = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}$. Since $1 - a = (1 - \lambda_1)(1 - \pi(A))$, we have $0 < a < 1$. Similarly, $0 < b < 1$. Thus $Q_{a,b}$ is stochastic; its second largest eigenvalue is $a + b - 1 = \lambda_1$. The matrix $Q_{a,b}$ is the transition matrix of a reversible Markov chain on the state space $\{1, 2\}$ whose stationary distribution has weight $\pi(A)$ on 1 and $1 - \pi(A)$ on 2.

Lemma 3.4 *The matrices $D_\theta P'$ and $\begin{pmatrix} e^\theta & 0 \\ 0 & 1 \end{pmatrix} Q_{a,b}$ have the same largest eigenvalue.*

Proof Since the matrix $\begin{pmatrix} e^\theta & 0 \\ 0 & 1 \end{pmatrix} Q_{a,b}$ is non-negative irreducible, its largest eigenvalue in module μ is a real positive number [14, Theorem 1.1], and it has a corresponding eigenvector $\begin{pmatrix} x \\ y \end{pmatrix}$ with strictly positive components. A simple calculation shows that the vector $x\chi_A + y\chi_{V-A}$ is an eigenvector of $D_\theta P'$ with eigenvalue μ. Since the matrix $D_\theta P'$ is non-negative irreducible and $x\chi_A + y\chi_{V-A}$ is strictly positive, this implies [14, Theorem 1.6] that μ is the largest eigenvalue of $D_\theta P'$. □

We are now ready to prove Theorem 3.1. Lemmas 3.3 and 3.4 show that $\lambda_0(D_\theta P) \leq \mu(\theta)$, where $\mu(\theta)$ is the largest eigenvalue of $\begin{pmatrix} e^\theta & 0 \\ 0 & 1 \end{pmatrix} Q_{a,b}$. Thus $\mu(\theta)$ is the largest zero of the characteristic polynomial $r_\theta(\mu) = \mu^2 - (ae^\theta + b)\mu + \lambda_1 e^\theta$. By Lemma 3.2,

$$\begin{aligned} \Pr\left[S_l \geq \beta\pi(A)l\right] &\leq \Pr\left[e^{\theta S_l} \geq e^{\theta\beta\pi(A)l}\right] \\ &\leq \frac{\mathrm{E}\left[e^{\theta S_l}\right]}{e^{\theta\beta\pi(A)l}} \\ &\leq \frac{e^\theta}{\mu(\theta)}||\Pi^{-1}q||_\pi \left(\frac{\mu(\theta)}{e^{\theta\beta\pi(A)}}\right)^l, \end{aligned} \tag{3}$$

for any $\theta \geq 0$. But $e^\theta = (\mu(\theta)^2 - b\mu(\theta))/(a\mu(\theta) - \lambda_1)$ since $r_\theta(\mu(\theta)) = 0$. Thus, Eq. 3 reduces to:

$$\Pr\left[S_l \geq \beta\pi(A)l\right] \leq \frac{e^\theta}{\mu(\theta)}||\Pi^{-1}q||_\pi \phi(\mu(\theta))^l, \tag{4}$$

where

$$\phi(\mu) = \mu\left(\frac{a\mu - \lambda_1}{\mu^2 - b\mu}\right)^{\beta\pi(A)}.$$

Since $\phi(\mu)$ goes to infinity as μ goes to infinity, $\phi(\mu)$ attains its minimum in $[1,\infty)$ at a point μ_0. Let $\theta_0 = \ln\left(\frac{\mu_0^2 - b\mu_0}{a\mu_0 - \lambda_1}\right)$. Note that $\theta_0 \geq 0$ since $\mu^2 - b\mu \geq a\mu - \lambda_1$ for $\mu \geq 1$. It is readily verified that $r_{\theta_0}(\mu_0) = 0$. Moreover, μ_0 is the largest zero of r_{θ_0} since $r_{\theta_0}(1) = (1-b)(1-e^{\theta_0}) \leq 0$. Thus $\mu_0 = \mu(\theta_0)$, and Eq. 4 implies that

$$\Pr\left[S_l \geq \beta\pi(A)l\right] \leq \frac{e^{\theta_0}}{\mu_0}||\Pi^{-1}q||_\pi \phi(\mu_0)^l. \tag{5}$$

Theorem 3.1 follows with $c_{\beta,\pi(A),\lambda_1} = e^{\theta_0}/\mu_0$. Note that $\alpha = \phi(\mu_0) < 1$ since $\phi(1) = 1$ and $\phi'(1) = 1 - \beta < 0$. □

We can determine μ_0 as follows. Since $\phi(\mu_0) = \min_{\mu>1}\phi(\mu)$,

$$\frac{\phi'(\mu_0)}{\phi(\mu_0)} = \frac{1}{\mu_0} + \frac{\beta\pi(A)a}{a\mu_0 - \lambda_1} - \frac{\beta\pi(A)(2\mu_0 - b)}{{\mu_0}^2 - b\mu_0} = 0.$$

Thus μ_0 is a zero of the polynomial ψ defined by

$$\psi(\mu) = a(1-\beta\pi(A))\mu^2 + (2\lambda_1\beta\pi(A) - \lambda_1 - ab)\mu + \lambda_1 b(1-\beta\pi(A)).$$

Since $\psi(b) = -\beta\pi(A)b(1-a)(1-b) \leq 0$, μ_0 is the largest zero of ψ. Thus α can be calculated explicitly in terms of β, $\pi(A)$ and λ_1.

3.1 An application to the leader election problem

The leader election problem arises in distributed computing. A set of players, a constant fraction of which are dishonest, are trying to elect a common leader. A protocol must guarantee that there is a reasonable chance that a leader is elected among the set of honest players, regardless of the behavior of the dishonest players. Recently, Ostrovsky, Rajagopalan and Vazirani [13] designed a constructive $O(\log n)$ round protocol for the leader election problem in the full information model that tolerates a small constant fraction of dishonest players. Their protocol uses walks on an expander graph, and their analysis relies on the Chernoff bound on random walks in [5]. Since the bound in Theorem 3.1 is sharper than the one in [5], Theorem 3.1 implies an improved bound on the fraction of dishonest players that the protocol can tolerate.

4 Other bounds

We first give bounds on the tail of S_l which are more explicit but weaker than the bound in Theorem 3.1. We then establish a bound on the probability that the Markov chain stays inside A for l steps. Finally, we give an application to the leader election problem.

Theorem 4.1 *For any integer $l \geq 1$ and $1 < \beta < 1/\pi(A)$,*

$$Pr[S_l \geq \beta\pi(A)l] \leq c_{\beta,\pi(A),\lambda_1}\sqrt{\sum_{v\in V}\frac{q(v)^2}{\pi(v)}}\exp\left(-\pi(A)\left(\sqrt{\beta-\beta\pi(A)}-\sqrt{1-\beta\pi(A)}\right)^2(1-\lambda_1)l\right).$$

Proof Let $\mu_1 = 1 + \epsilon(1-\lambda_1)$, where $\epsilon \geq 0$ is to be determined later. Then

$$\begin{aligned}\frac{a\mu_1-\lambda_1}{\mu_1-b} &= 1-\frac{(1-\pi(A))\epsilon(1-\lambda_1)}{\pi(A)+\epsilon}\\ &\leq e^{-(1-\pi(A))\epsilon(1-\lambda_1)/(\pi(A)+\epsilon)},\end{aligned}$$

since $1 + x \leq e^x$ for any real number x. Similarly, ${\mu_1}^{1-\beta\pi(A)} \leq e^{\epsilon(1-\beta\pi(A))(1-\lambda_1)}$, and so

$$\phi(\mu_1) \leq \exp\left(\left(\epsilon(1-\beta\pi(A)) - \frac{(1-\pi(A))\epsilon\beta\pi(A)}{\pi(A)+\epsilon}\right)(1-\lambda_1)\right). \tag{6}$$

The coefficient of $1-\lambda_1$ in the right-hand side of Eq. 6 is minimized when

$$\epsilon = \pi(A)\left(\sqrt{\frac{\beta-\beta\pi(A)}{1-\beta\pi(A)}}-1\right),$$

and is equal to $-\pi(A)\left(\sqrt{\beta-\beta\pi(A)}-\sqrt{1-\beta\pi(A)}\right)^2$ for this value of ϵ. Thus, by Eq. 5,

$$\begin{aligned}\Pr[S_l \geq \beta\pi(A)l] &\leq c_{\beta,\pi(A),\lambda_1}||\Pi^{-1}q||_\pi\phi(\mu_0)^l\\ &\leq c_{\beta,\pi(A),\lambda_1}||\Pi^{-1}q||_\pi\phi(\mu_1)^l\\ &\leq c_{\beta,\pi(A),\lambda_1}||\Pi^{-1}q||_\pi\exp\left(-\pi(A)\left(\sqrt{\beta-\beta\pi(A)}-\sqrt{1-\beta\pi(A)}\right)^2(1-\lambda_1)l\right),\end{aligned}$$

as desired. □

Note that if $\lambda_1 = 0$ then $\mu_0 = (1-\pi(A))/(1-\beta\pi(A))$ and $\phi(\mu_0) = \left(\frac{1-\pi(A)}{1-\beta\pi(A)}\right)^{1-\beta\pi(A)}\beta^{-\beta\pi(A)}$. In this case, the bound in Theorem 3.1 coincides with the usual Chernoff bound [7], up to a multiplicative constant. An easy continuity argument shows that if β and $\pi(A)$ are fixed, then $\phi(\mu_0)$ goes to $\left(\frac{1-\pi(A)}{1-\beta\pi(A)}\right)^{1-\beta\pi(A)}\beta^{-\beta\pi(A)}$ as λ_1 goes to 0, and $c_{\beta,\pi(A),\lambda_1}$ stays bounded.

Corollary 4.2 *For any integer $l \geq 1$ and $1 < \beta < 1/\pi(A)$,*

$$Pr[S_l \geq \beta\pi(A)l] \leq \frac{\beta}{1-a+\beta a}\sqrt{\sum_{v\in V}\frac{q(v)^2}{\pi(v)}}\exp\left(-(\beta-1)^2\pi(A)^2(1-\lambda_1)l\right).$$

Proof We replace the subscript 0 by 2 in Eq. 5, where $\mu_2 = 1 + \pi(A)(\beta - 1)(1 - \lambda_1)$ and $\theta_2 = \ln\left(\frac{\mu_2^2 - b\mu_2}{a\mu_2 - \lambda_1}\right)$. This corresponds to the case $\epsilon = \pi(A)(\beta - 1)$, following the notation in Theorem 4.1. Eq. 6 shows that $\phi(\mu_2) \leq \exp\left(-(\beta-1)^2\pi(A)^2(1-\lambda_1)\right)$. A simple calculation shows that

$$\frac{e^{\theta_2}}{\mu_2} = \frac{\mu_2 - b}{a\mu_2 - \lambda_1} = \frac{\beta}{1 - a + \beta a}.$$

The corollary follows. □

Corollary 4.2 shows in particular that the exponent in our bound always beats the one in [5]. The following theorem is a generalization of a result on random walks on expanders shown in [11]. The proof we use is very similar.

Theorem 4.3 *The probability that $X_i \in A$, for $1 \leq i \leq l$, is at most $\sqrt{\pi(A)\sum_{v\in A} q(v)^2/\pi(v)}\, a^{l-1}$.*

Proof Let P_A and P'_A be the $n \times n$ matrices defined by $P_A = (P(x,y)\chi_A(x)\chi_A(y))$ and $P'_A = (P'(x,y)\chi_A(x)\chi_A(y))$. For $f \in L^2(\pi)$, define $f_A \in L^2(\pi)$ by $f_A(v) = f(v)\chi_A(v)$. By Eq. 2 and noting that $< f, P_A f >_\pi =< f_A, P f_A >_\pi$ and $< f, P'_A f >_\pi =< f_A, P' f_A >_\pi$ for any $f \in L^2(\pi)$, we see that $P_A \leq_\pi P'_A$. Since P_A and P'_A are selfadjoint in $L^2(\pi)$, it follows that $\lambda_0(P_A) \leq \lambda_0(P'_A) = a$. Now,

$$\begin{aligned}
\Pr\left[X_i \in A, 1 \leq i \leq l\right] &= q^t {P_A}^{l-1}\chi_A \\
&= < \Pi^{-1}q, {P_A}^{l-1}\chi_A >_\pi \\
&\leq ||\Pi^{-1}q||_\pi ||\chi_A||_\pi \lambda_0(P_A)^{l-1} \\
&\leq ||\Pi^{-1}q||_\pi \sqrt{\pi(A)}\, a^{l-1}.
\end{aligned}$$

□

5 Tightness of bounds

In this section, we prove that the exponent α in Theorem 3.1 is optimal as follows. For any real numbers $0 < p < 1$ and $0 < \lambda_1 < 1$, consider the Markov chain on the state space $\{1, 2\}$ whose transition matrix is $Q_{a,b}$. This Markov chain is irreducible and its second largest eigenvalue equals λ_1. Its stationary distribution has weight p on 1 and $1 - p$ on 2. Let S_l denote the number of times the Markov chain hits the set $A = \{1\}$ during the first l steps.

Theorem 5.1 *For integer $l \geq 1$, $1 < \beta < 1/p$, and any initial distribution,*

$$Pr[S_l \geq \beta p l] = \Omega(\alpha^{l(1-o(1))}),$$

where the constant behing Ω is a function of λ_1, p and β.

Proof We will assume that l is sufficiently large whenever needed, and that all multiples of l that we consider are integral. We can also assume without loss of generality that the initial distribution is concentrated on $\{1\}$ since the probablity that the Markov chain takes value 1 in the second step is at least $\min(a, 1-b) = 1-b$.

We will exhibit a set of disjoints events for which $S_l = \beta pl$ and whose probabilities sum up to $\Omega(\alpha^{l(1-o(1))})$. Fix a positive real number x less than $\min(\beta p, 1-\beta p)$. For any sequence of nonnegative integers $(m_1, m_2, \ldots, m_{xl})$ and $(n_1, n_2, \ldots, n_{xl})$ such that $m_1 + m_2 + \cdots + m_{xl} = (\beta p - x)l$ and $n_1 + n_2 + \cdots + n_{xl} = (1 - \beta p - x)l$, consider the following event that consists of xl subsequent phases, the first phase starting at step 1. For $1 \le i \le xl$, the Markov chain takes value 1 during the first m_i steps of Phase i, then takes value 2 for the $n_i + 1$ subsequent steps, then takes value 1 for one step. For each such event, the total number of steps the Markov chain takes value 1 is $(m_1 + 1) + (m_2 + 1) + \cdots + (m_{xl} + 1) = \beta pl$, as desired. Each such event consists of $m_1 + m_2 + \cdots + m_{xl} = (\beta p - x)l$ loops in position 1, $n_1 + n_2 + \cdots + n_{xl} = (1 - \beta p - x)l$ loops in position 2, xl transitions from 1 to 2 and xl transitions from 2 to 1, so it happens with probability $a^{(\beta p - x)l} b^{(1-\beta p - x)l} (1-a)^{xl} (1-b)^{xl}$. Since there are $\binom{\beta pl - 1}{xl - 1}$ ways to choose the sequence $(m_1, m_2, \ldots, m_{xl})$ and $\binom{(1-\beta p)l - 1}{xl - 1}$ ways to choose the sequence $(n_1, n_2, \ldots, n_{xl})$,

$$\begin{aligned} \Pr[S_l \ge \beta pl] &\ge \binom{\beta pl - 1}{xl - 1} \binom{(1-\beta p)l - 1}{xl - 1} a^{(\beta p - x)l} b^{(1-\beta p - x)l} (1-a)^{xl} (1-b)^{xl} \\ &= \Omega(\gamma(x)^{l(1-o(1))}), \end{aligned}$$

where

$$\gamma(x) = \frac{(\beta p)^{\beta p} (1-\beta p)^{1-\beta p}}{x^{2x} (\beta p - x)^{\beta p - x} (1 - \beta p - x)^{1 - \beta p - x}} a^{\beta p - x} b^{1 - \beta p - x} (1-a)^x (1-b)^x.$$

For the rest of the proof, let $x = (1-\beta p)(1 - b/\mu_0)$. It is clear that $0 < x < 1 - \beta p$. Since μ_0 is a zero of the polynomial ψ defined in Section 3 ($\pi(A) = p$ in this case) and $\mu_0 = b(1-\beta p)/(1 - \beta p - x)$, a simple but tedious calculation shows that $abx^2 = (1-a)(1-b)(\beta p - x)(1 - \beta p - x)$. Thus $x < \beta p$, and

$$\begin{aligned} \gamma(x) &= \frac{(a\beta p)^{\beta p} (b(1-\beta p))^{1-\beta p}}{(\beta p - x)^{\beta p} (1 - \beta p - x)^{1-\beta p}} \\ &= \frac{(a\beta p)^{\beta p} {\mu_0}^{1-\beta p}}{(\beta p - x)^{\beta p}}. \end{aligned}$$

Further calculations show that $\gamma(x) = \phi(\mu_0) = \alpha$. This concludes the proof of Theorem 5.1. □

6 Conclusion

We gave an improved bound on the tail of the distribution of S_l/l. Whether our result can be used to improve the running time of known algorithms based on sampling is a question

worth investigating. Our bound is asymptotically optimal for Markov chains on two states, but we don't know if it is tight for the examples that arise in applications. A tighter analysis adapted to these examples could also lead to an improvement in their running time.

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in logspace. In *STOC87*.

[2] D. Aldous. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probability in the Engineering and Information Sciences*, 1:33–46, 1987.

[3] D. Aldous. *Reversible Markov chains and random walks on graphs*. 1993. In preparation.

[4] H. Chernoff. A measure for asymptotic efficiency of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:493–507, 1952.

[5] D. Gillman. A Chernoff bound for random walks on expander graphs. In *FOCS93*, pages 680–691. IEEE Computer Society Press.

[6] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security preserving amplification of hardness. In *FOCS90*, pages 318–326. IEEE Computer Society Press.

[7] T. Hagerup and C. Rub. A guided tour of Chernoff bounds. *Information Processing Letters*, 33:305–308, 1989/90.

[8] T. Hoglund. Central limit theorems and statistical inference for finite markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 29:123–151, 1974.

[9] R. Impagliazzo and D. Zuckerman. How to recycle random bits. In *FOCS89*, pages 248–253. IEEE Computer Society Press.

[10] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. of Comput.*, 18:1149–1178, 1989.

[11] N. Kahale. Better expansion for Ramanujan graphs. In *FOCS91*, pages 296–303. IEEE Computer Society Press.

[12] L. Lovász and M. Simonovits. Random walks in a convex body and an improved volume algorithm. *Random Strutures & Algorithms*, 4(4):359–412, 1993.

[13] R. Ostrovsky, S. Rajagopalan, and U. Vazirani. Simple and efficient leader election in the full information model. In *STOC94*, pages 234–242. ACM Press.

[14] E. Seneta. *Non-negative matrices and Markov chains*. Springer-Verlag, 1981.

[15] A. Sinclair and Mark Jerrum. Approximate counting, uniform generation, and rapidly mixing Markov chains. *Information and Computation*, 82:93–113, 1989.