*Combinatorics, Probability and Computing* (1997) 6, 465–474. Printed in the United Kingdom
© 1997 Cambridge University Press

# Large Deviation Bounds for Markov Chains

NABIL KAHALE†

AT&T Labs–Research, Florham Park, NJ 07932, USA
(e-mail: kahale@research.att.com)

*Received 1 June 1994; revised 31 October 1996*

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 for a certain class of Markov chains. It beats the best previously known results in this direction. We present an application to the leader election problem.

## 1. Introduction

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

$$\pi(u)P(u,v) = \pi(v)P(v,u), \text{ for all } u,v \in V. \tag{1.1}$$

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 [2] that, for any initial distribution, the fraction $S_l/l$ converges almost surely to $\pi(A)$ as $l$ goes to infinity. This led Aldous [3] 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 [3] and [14]. An exponential bound on the tail of $S_l/l$ was established in [5] and [11] in a special case, and in a more general setting in [7].

In this paper, we establish a bound on the tail of the distribution of $S_l/l$ that beats the previously known bounds. Like the previous bounds, ours depends on the second largest eigenvalue of $P$. We also show that for given values of $\beta$, $\pi(A)$ and of the second largest

† This work was done while the author was at the Massachusetts Institute of Technology and at DIMACS.

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

eigenvalue of $P$, the decay rate in our bound is optimal for a certain class of Markov chains. 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. As discussed in Section 4, the decay rate in our bound is better than the one in [7] by at least a constant factor. When $\pi(A)$ is small, which is the case in several applications such as approximating the dense permanent [12], it beats it by a factor of $\Theta(\pi(A)^{-1})$.

As in [4], [7] and [10], 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 $P$, except the largest one, are equal. This case can be further 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 [12, 17] (see also [14], and references therein), complexity and cryptography [1, 8, 11], and distributed computing [15].

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. We establish more explicit but weaker bounds in Section 4, and present an application to the leader election problem. We also give a bound on the probability that the Markov chain stays inside $A$ for $l$ steps, generalizing a result in [13]. 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 $\langle\ ,\ \rangle_\pi$:

$$\langle f, g\rangle_\pi = \sum_{x\in V} \pi(x) f(x) g(x).$$

For $f \in L^2(\pi)$, let $||f||_\pi = \sqrt{\langle f, f\rangle_\pi}$. In matrix notation, $\langle f, g\rangle_\pi = f^T \Pi g$, where $\Pi$ 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

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

Equation (1.1) shows that $P$ is a self-adjoint operator in $L^2(\pi)$, *i.e.* $\langle Pf, g\rangle_\pi = \langle f, Pg\rangle_\pi$, for all $f, g \in L^2(\pi)$. Hence $P$ is diagonalizable in an orthonormal (with respect to the scalar product $\langle\ ,\ \rangle_\pi$) 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 [16] shows that 1 is the largest eigenvalue of $P$ in absolute value. Denote by $\lambda_i$ the eigenvalue corresponding to $e_i$, for $1 \leqslant i \leqslant n-1$, and assume that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{n-1}$. Since the chain is irreducible, $\lambda_1 < 1$. We will assume throughout the paper that $\lambda_1 \geqslant 0$. 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, if $X_t = v$, then $X_{t+1}$ is a uniformly random neighbour of $v$. This chain is reversible and $\pi(v) = d(v)/2|E|$, where $d(v)$ is the degree of $v$.

Note that we do not impose any condition on $\lambda_{n-1}$ for our results to be valid. In

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

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 $\beta$ 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 initial distribution $q$, integer $l \geqslant 1$ and $1 < \beta < 1/\pi(A)$,*

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

*where*

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

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

Before proving Theorem 3.1, we prove the following three lemmas. The first lemma gives a bound on the generating function of $S_l$ in terms of the largest eigenvalue of a certain matrix. The second lemma gives a bound on the largest eigenvalue of this matrix in terms of the largest eigenvalue of a simpler matrix. The third lemma reduces the calculation of the largest eigenvalue of the latter matrix to the largest eigenvalue of a $2 \times 2$ matrix.

Let $D_\theta$ 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, $\sqrt{d_\theta} = \sqrt{D_\theta}\mathbf{1}$. From the theory of non-negative matrices [16, Theorem 1.1], for any non-negative matrix $M$, $\lambda_0(M)$ is the largest eigenvalue of $M$ in *absolute value*. Therefore, the largest eigenvalue of $M^{l-1}$ is equal to the $(l-1)$st power $\lambda_0(M)^{l-1}$ of $\lambda_0(M)$.

**Lemma 3.2.** *For any $\theta \geqslant 0$,*

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

**Proof.**

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

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

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 $\langle\ ,\ \rangle_\pi$. The sixth equation follows from the fact that $\sqrt{D_\theta}P\sqrt{D_\theta}$ and $D_\theta P$ are similar matrices, and thus have the same eigenvalues. □

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

**Proof.** For the proof, we need the following definition. For two $n \times n$ matrices $A$ and $B$, we say that $A \leqslant_\pi B$ if $\langle f, Af\rangle_\pi \leqslant \langle f, Bf\rangle_\pi$ for any $f \in L^2(\pi)$.

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

$$\begin{aligned} \langle f, Pf\rangle_\pi &= \sum_{i=0}^{n-1} \lambda_i {x_i}^2 \qquad (3.1)\\ &\leqslant {x_0}^2 + \lambda_1 \sum_{i=1}^{n-1} {x_i}^2 \\ &= \langle f, P'f\rangle_\pi. \end{aligned}$$

Thus $P \leqslant_\pi P'$ which, since $\sqrt{D_\theta}$ is self-adjoint in $L^2(\pi)$, implies that $\sqrt{D_\theta}P\sqrt{D_\theta} \leqslant_\pi \sqrt{D_\theta}P'\sqrt{D_\theta}$. Since $\lambda_0(A) = \max_{f\neq 0}\langle f, Af\rangle_\pi / ||f||_\pi^2$, for any self-adjoint operator $A$, we conclude that $\lambda_0(\sqrt{D_\theta}P\sqrt{D_\theta}) \leqslant \lambda_0(\sqrt{D_\theta}P'\sqrt{D_\theta})$. This is equivalent to the inequality $\lambda_0(D_\theta P) \leqslant \lambda_0(D_\theta P')$. □

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 largest eigenvalue is 1 and 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 modulus $\mu$ is a real positive number [16, 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 $\mu$. Since the matrix $D_\theta P'$ is non-negative irreducible and $x\chi_A + y\chi_{V-A}$ is strictly positive, this implies [16, Theorem 1.6] that $\mu$ is the largest eigenvalue of $D_\theta P'$. □

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

We are now ready to prove Theorem 3.1. Lemmas 3.3 and 3.4 show that $\lambda_0(D_\theta P) \leqslant \mu(\theta)$, 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,

$$\begin{aligned} \Pr[S_l \geqslant \beta\pi(A)l] &\leqslant \Pr\left[e^{\theta S_l} \geqslant e^{\theta\beta\pi(A)l}\right] \\ &\leqslant \frac{\mathrm{E}\left[e^{\theta S_l}\right]}{e^{\theta\beta\pi(A)l}} \\ &\leqslant \frac{e^\theta}{\mu(\theta)} ||\Pi^{-1}q||_\pi \left(\frac{\mu(\theta)}{e^{\theta\beta\pi(A)}}\right)^l, \end{aligned} \tag{3.2}$$

for any $\theta \geqslant 0$. But $e^\theta = (\mu(\theta)^2 - b\mu(\theta))/(a\mu(\theta) - \lambda_1)$ since $r_\theta(\mu(\theta)) = 0$. Thus, (3.2) reduces to

$$\Pr[S_l \geqslant \beta\pi(A)l] \leqslant \frac{e^\theta}{\mu(\theta)} ||\Pi^{-1}q||_\pi \phi(\mu(\theta))^l, \tag{3.3}$$

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 $\mu$ goes to infinity, $\phi(\mu)$ attains its minimum in $[1, \infty)$ at a point $\mu_0$. Let $\theta_0 = \ln\left(\frac{\mu_0^2 - b\mu_0}{a\mu_0 - \lambda_1}\right)$. Note that $\theta_0 \geqslant 0$, since $\mu^2 - b\mu \geqslant a\mu - \lambda_1$ for $\mu \geqslant 1$. It is readily verified that $r_{\theta_0}(\mu_0) = 0$. Moreover, $\mu_0$ is the largest zero of $r_{\theta_0}$, since $r_{\theta_0}(1) = (1-b)(1-e^{\theta_0}) \leqslant 0$. Thus $\mu_0 = \mu(\theta_0)$, and (3.3) implies that

$$\Pr[S_l \geqslant \beta\pi(A)l] \leqslant \frac{e^{\theta_0}}{\mu_0} ||\Pi^{-1}q||_\pi \phi(\mu_0)^l. \tag{3.4}$$

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 $\mu_0$ 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 $\mu_0$ is a zero of the polynomial $\psi$ 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 $\psi(b) = -\beta\pi(A)b(1-a)(1-b) \leqslant 0$, $\mu_0$ is the largest zero of $\psi$. Thus $\alpha$ can be calculated explicitly in terms of $\beta$, $\pi(A)$ and $\lambda_1$.

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 [9], up

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

to a multiplicative constant. An easy continuity argument shows that if $\beta$ 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 $\lambda_1$ goes to 0, and $c_{\beta,\pi(A),\lambda_1}$ stays bounded.

## 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. These bounds come from choosing a $\mu$ in the statement of Theorem 3.1, which may not minimize $\phi(\mu)$. 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 $l \geqslant 1$ and $1 < \beta < 1/\pi(A)$,*

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

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

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

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

$$\phi(\mu_1) \leqslant \exp\left(\left(\epsilon(1-\beta\pi(A)) - \frac{(1-\pi(A))\epsilon\beta\pi(A)}{\pi(A)+\epsilon}\right)(1-\lambda_1)\right). \tag{4.1}$$

The coefficient of $1-\lambda_1$ in the right-hand side of (4.1) 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 $\epsilon$. Thus, by (3.4),

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

as desired. □

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

**Theorem 4.2.** *For any integer $l \geqslant 1$ and $1 < \beta < 1/\pi(A)$,*

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

**Proof.** We replace the subscript 0 by 2 in (3.4), 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. Equation (4.1) shows that $\phi(\mu_2) \leqslant \exp\left(-(\beta-1)^2\pi(A)^2(1-\lambda_1)\right)$. 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 theorem follows. □

The bound in [7] can be stated as follows:

$$\Pr\,[S_l \geqslant \beta\pi(A)l] \leqslant 2\sqrt{\sum_{v\in V}\frac{q(v)^2}{\pi(v)}}\exp\left(-(\beta-1)^2\pi(A)^2(1-\lambda_1)l/(20v)\right), \tag{4.2}$$

where $v = \max_{x,y}\pi(x)/\pi(y)$. Gillman [6] has subsequently shown that $v$ can be replaced by 1 in (4.2). Theorem 4.2 shows that the exponent in our bound always beats the one in (4.2) by at least a constant factor. For fixed $\beta$, Theorem 4.1 shows that it beats it by a factor of $\Theta(\pi(A)^{-1})$ when $\pi(A)$ is small.

Following [1, 8], an improved bound on the probability that a random walk on a regular graph stays inside a given set was given in [13]. The following theorem generalizes this result to any reversible reversible Markov chain on a finite state space. The proof we use is similar.

**Theorem 4.3.** *The probability that $X_i \in A$, for $1 \leqslant i \leqslant l$, 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 (3.1) and noting that $\langle f, P_A f\rangle_\pi = \langle f_A, Pf_A\rangle_\pi$ and $\langle f, P'_A f\rangle_\pi = \langle f_A, P'f_A\rangle_\pi$ for any $f \in L^2(\pi)$, we see that $P_A \leqslant_\pi P'_A$. Since $P_A$ and $P'_A$ are self-adjoint in $L^2(\pi)$, it follows that $\lambda_0(P_A) \leqslant \lambda_0(P'_A)$. The largest eigenvalue of $a^{-1}P'_A$ is equal to the largest eigenvalue of its restriction to $A$, which is 1 since this matrix is stochastic. Therefore $\lambda_0(P_A) \leqslant a$. Now,

$$\begin{aligned}
\Pr\,[X_i \in A, 1 \leqslant i \leqslant l] &= q^T {P_A}^{l-1}\chi_A \\
&= \langle \Pi^{-1}q, {P_A}^{l-1}\chi_A\rangle_\pi \\
&\leqslant ||\Pi^{-1}q||_\pi ||\chi_A||_\pi \lambda_0(P_A)^{l-1} \\
&\leqslant ||\Pi^{-1}q||_\pi \sqrt{\pi(A)}\;a^{l-1}.
\end{aligned}$$

□

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

### 4.1. 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 behaviour of the dishonest players. Recently, Ostrovsky, Rajagopalan and Vazirani [15] 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 [7]. Since the bound in Theorem 3.1 is sharper than the one in [7], Theorem 3.1 implies an improved bound on the fraction of dishonest players that the protocol can tolerate.

## 5. Tightness of bounds

In this section, we prove that the exponent $\alpha$ in Theorem 3.1 is optimal in the following sense. For any integer $n \geqslant 2$ and any real numbers $0 < p < 1$ and $0 < \lambda_1 < 1$, there exists an irreducible reversible Markov chain on a state space of size $n$ whose second largest eigenvalue equals $\lambda_1$, a subset $A$ of states with stationary distribution $\pi(A) = p$, such that the exponent in Theorem 3.1 is optimal for any initial distribution. We start with the case $n = 2$. Consider the Markov chain on the state space $\{1, 2\}$ whose transition matrix is $Q_{a,b}$, where $a = \lambda_1 + p - \lambda_1 p$ and $b = 1 - p + \lambda_1 p$. This Markov chain is irreducible and its second largest eigenvalue equals $\lambda_1$. 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 $l \geqslant 1$, $1 < \beta < 1/p$, and any initial distribution,*

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

*where the constant behind $\Omega$ is a function of $\lambda_1$, $p$ and $\beta$.*

**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 disjoint events for which $S_l = \beta p l$ 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 non-negative 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 $1 \leqslant i \leqslant xl$, 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 $(m_1 + 1) + (m_2 + 1) + \cdots + (m_{xl} + 1) = \beta p l$, 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 $xl$

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

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 $(n_1, n_2, \ldots, n_{xl})$,

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

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 $\mu_0$ is a zero of the polynomial $\psi$ 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. □

The case for integer $n \geqslant 2$ can be reduced to the case $n = 2$ by taking $A$ to be any proper subset of states, $\pi$ being uniform on $A$ and on its complement with $\pi(A) = p$, and the transition matrix being $P'$.

## 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 a certain class of Markov chains, 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 times.

## References

[1] Ajtai, M., Komlós, J. and Szemerédi, E. (1987) Deterministic simulation in logspace. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, ACM Press, pp. 132–140.

[2] Aldous, D. and Fill, J. (1997) *Reversible Markov Chains and random walks on graphs.* Monograph in preparation.

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

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press

[4] Chernoff, H. (1952) A measure for asymptotic efficiency of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.
[5] Cohen, A. and Wigderson, A. (1989) Dispersers, deterministic amplification, and weak random sources. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, pp. 14–19.
[6] Gillman, D. (1994) Personal communication.
[7] Gillman, D. (1993) A Chernoff bound for random walks on expander graphs. In *Proceedings of the 34th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, pp. 680–691.
[8] Goldreich, O., Impagliazzo, R., Levin, L., Venkatesan, R. and Zuckerman, D. (1990) Security preserving amplification of hardness. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, pp. 318–326.
[9] Hagerup, T., and Rub, C. (1989/90) A guided tour of Chernoff bounds. *Inform. Process. Lett.* **33** 305–308.
[10] Hoglund, T. (1974) Central limit theorems and statistical inference for finite Markov Chains. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **29** 123–151.
[11] Impagliazzo, R. and Zuckerman, D. (1989) How to recycle random bits. In *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, pp. 248–253.
[12] Jerrum, M. and Sinclair, A. (1989) Approximating the permanent. *SIAM J. Comput.* **18** 1149–1178.
[13] Kahale, N. (1991) Better expansion for Ramanujan graphs. In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, pp. 296–303.
[14] Lovász, L. and Simonovits, M. (1993) Random walks in a convex body and an improved volume algorithm. *Random Strutures & Algorithms* **4** 359–412.
[15] Ostrovsky, R., Rajagopalan, S. and Vazirani, U. (1994) Simple and efficient leader election in the full information model. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, ACM Press, pp. 234–242.
[16] Seneta, E. (1981) *Non-negative matrices and Markov Chains.* Springer.
[17] Sinclair, A. and Jerrum, M. (1989) Approximate counting, uniform generation, and rapidly mixing Markov Chains. *Information and Computation* **82** 93–113.

https://doi.org/10.1017/S0963548397003209 Published online by Cambridge University Press