Combinatorics, Probability and Computing (2008) 17, 307–316. © 2007 Cambridge University Press
doi:10.1017/S0963548307008772 First published online 7 December 2007 Printed in the United Kingdom

Tail Estimates for Sums of Variables Sampled by a Random Walk

ROY WAGNER

Computer Science Department, Academic College of Tel Aviv–Yaffa,
2 Rabenu Yerham St, Jaffa 68182, Israel
and
School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel
(e-mail: rwagner@mta.ac.il)

Received 13 November 2006; revised 8 October 2007

We prove tail estimates for variables of the form $\sum_i f(X_i)$, where $(X_i)_i$ is a sequence of states drawn from a reversible Markov chain, or, equivalently, from a random walk on an undirected graph. The estimates are in terms of the range of the function f, its variance, and the spectrum of the graph. The purpose of our estimates is to determine the number of chain/walk samples which are required for approximating the expectation of a distribution on vertices of a graph, especially an expander. The estimates must therefore provide information for fixed number of samples (as in Gillman's [4]) rather than just asymptotic information. Our proofs are more elementary than other proofs in the literature, and our results are sharper. We obtain Bernstein- and Bennett-type inequalities, as well as an inequality for sub-Gaussian variables.

1. Introduction

One of the basic concerns of sampling theory is economizing on the 'cost' and quantity of samples required to estimate the expectation of random variables. Drawing states by implementing a reversible Markov chain or, equivalently, by conducting a random walk is often considerably 'cheaper' than the standard Monte Carlo procedure of drawing independent random states. Independence is indeed lost when sampling by a Markov chain; the empirical average, however, may converge to the actual average at a comparable rate to the rate of convergence for independent sampling. This form of sampling is especially useful in the context of random walks on expander graphs.

This approach plays an important role in statistical physics and in computer science (a concise summary of applications is provided in [5]). Results concerning the rate of convergence of empirical averages sampled by a random walk, which hold for a fixed number of samples (rather than just asymptotically), have been obtained by several authors starting with Gillman [4], followed by [3], [8] and [7] (for vector-valued functions consult [6]). Of these, only [8] and [6] allowed the variance to play a role in their estimates, as is the case in this paper.

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

This paper is a further step in this direction. We improve known Bernstein-type inequalities, and prove a new Bennett-type inequality and a new inequality for sub-Gaussian variables. Our methods are much more elementary than those prevailing in the literature, as we do not apply Kato's perturbation theory to estimate eigenvalues.

Our results were motivated by applications relating to graphs with large spectral gaps (expanders) and tails which go far beyond the variance (large deviations), such as the recent [1]. Accordingly, our results are stated for the reversible discrete setting. Analogues for the continuous and non-reversible settings can be derived using the simple reduction techniques presented in Sections 3.2 and 3.3 of [8].

2. The results

Let G be a finite undirected, possibly weighted, connected graph with N vertices (random walks on such graphs can represent any finite irreducible reversible Markov chain). Denote by s the stationary distribution of the random walk on the graph or the equivalent Markov chain. Let f be a function on the vertices of G, normalized to have absolute maximum 1 and mean 0 relative to the stationary distribution, namely $\sum_i f(i)s(i) = 0$. Let $V = \sum_i f^2(i)s(i)$ denote the variance of f with respect to the stationary distribution. We will think of functions on G as vectors in $\mathbb{R}^N$ and *vice versa*, so where u and v are vectors, expressions such as e^u and uv will stand for coordinatewise operations.

Denote by P the transition matrix of the Markov chain/random walk, such that P_{ij} is the probability of moving from node/state j to node/state i. By the Perron–Frobenius theorem the eigenvalues of this matrix are all real, the top eigenvalue is 1 (with s as the only corresponding eigenvector up to scalar multiplication), and the absolute value of all other eigenvalues is smaller than or equal to 1. Let $\alpha < 1$ be the maximum between the second-largest eigenvalue of P and zero, and $\beta \leqslant 1$ the second-largest absolute value of an eigenvalue of P.

Given a starting distribution q, the random variables $X_0, X_1, \ldots$ will denote the trajectory of the random walk or, equivalently, the states drawn from the Markov chain. $\mathbb{P}_q$ and $\mathbb{E}_q$ will stand for the probability and expectation of events related to this walk respectively. Let $S_n = \sum_{i=1}^n f(X_i)$. Our concern in this paper is tail estimates for the distribution of S_n.

We will prove inequalities in terms of both α and β. Note that inequalities in terms of α 'cost' an additional multiplicative factor outside the exponent, whereas inequalities in terms of β are useless in the case of $\beta = 1$ (*i.e.*, bipartite graphs), and become relatively poor if α is small and β is large, which may be the case.

Theorem 2.1. *Define*

$$\delta(x, r) = x(e^{2r} + V(e^r - 1)^2)$$

and

$$\Delta(x, r) = \frac{4xe^{2r}(e^r - 1)^2}{1 - \delta(x, r)}.$$

In the above setting we get

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant \min_{1 \geqslant \delta(\beta^2, r),\ r \geqslant 0} \left\| \frac{q}{\sqrt{s}} \right\|_2 e^{-\frac{n}{2}\left[2\gamma r - V(e^{2r} - 1 - 2r + \Delta(\beta^2, r))\right]}, \tag{2.1}$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

and

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant \min_{1\geqslant\delta(\alpha,r),\ r\geqslant 0}\left\|\frac{q}{\sqrt{s}}\right\|_2 e^{2r}e^{-n\left[2\gamma r - V(e^{2r}-1-2r+\Delta(\alpha,r))\right]}. \tag{2.2}$$

Remark. Note that the results are the same up to the factor $\frac{1}{2}$ in the exponent, the multiplicative factor e^{2r} and the replacement of β^2 by α.

Note also that when α goes to 0, which is effectively almost the case of independent sampling, $\Delta(\alpha, r)$ also vanishes, and the term we get inside the exponent is the same as the term appearing in standard proofs of Bennett and Bernstein inequalities for independent variables.

The infimum is hard to compute, so we must optimize separately for different parameter regimes.

First we use the above result to derive a Bennett-type inequality (*cf.* [2]).

Corollary 2.2. *In the above setting,*

$$\begin{aligned}\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) &\leqslant \left\|\frac{q}{\sqrt{s}}\right\|_2 \exp\left(-\frac{n}{2}tC_\beta V\left[\left(1+\frac{\gamma}{tC_\beta V}\right)\log\left(1+\frac{\gamma}{tC_\beta V}\right)-\frac{\gamma}{tC_\beta V}\right]\right)\\ &\leqslant \left\|\frac{q}{\sqrt{s}}\right\|_2 \exp\left(-\frac{n}{2}\gamma\log\frac{\gamma}{etC_\beta V}\right),\end{aligned}$$

where $t \geqslant 1, C_\beta = 2\frac{1+\beta^2}{1-\beta^2}$ and provided that $\gamma \geqslant (t-1)\frac{1+\beta^2}{\beta^2}V$.

$$\begin{aligned}\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) &\leqslant \left(1+\frac{\gamma}{tC_\alpha V}\right)\left\|\frac{q}{\sqrt{s}}\right\|_2 \exp\left(-ntC_\alpha V\left[\left(1+\frac{\gamma}{tC_\alpha V}\right)\log\left(1+\frac{\gamma}{tC_\alpha V}\right)-\frac{\gamma}{tC_\alpha V}\right]\right)\\ &\leqslant \left(1+\frac{\gamma}{tC_\alpha V}\right)\left\|\frac{q}{\sqrt{s}}\right\|_2 \exp\left(-n\gamma\log\frac{\gamma}{etC_\alpha V}\right)\end{aligned}$$

where $t \geqslant 1, C_\alpha = 2\frac{1+\alpha}{1-\alpha}$ and provided that $\gamma \geqslant (t-1)\frac{1+\alpha}{\alpha}V$.

Our theorem also allows us to reproduce Lezaud's estimates from [8] with improved constants.

Corollary 2.3.

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant \left\|\frac{q}{\sqrt{s}}\right\|_2 e^{-n\frac{1-\beta^2}{1+\beta^2}\frac{\gamma^2}{4(V+\gamma)}}$$

and

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant e^{\frac{1-\alpha}{1+\alpha}\frac{\gamma^2}{V+\gamma}}\left\|\frac{q}{\sqrt{s}}\right\|_2 e^{-n\frac{1-\alpha}{1+\alpha}\frac{\gamma^2}{2(V+\gamma)}}.$$

Remark. These inequalities imply

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant e^{\frac{1}{4V}(1-\alpha)\gamma^2}\left\|\frac{q}{\sqrt{s}}\right\|_2 e^{-\frac{n}{8V}(1-\alpha)\gamma^2}$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

for $\gamma \leqslant V$, and

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant e^{\frac{1}{4}(1-\alpha)\gamma}\left\|\frac{q}{\sqrt{s}}\right\|_2 e^{-\frac{n}{8}(1-\alpha)\gamma}$$

for $\gamma \geqslant V$. If γ is much larger or much smaller than V, the constant 8 can be decreased towards 4. Our method allows us to improve the constant multiplying γ in the denominator, but we will not include the details because the modification to the proof is cumbersome and straightforward.

The Bennett-type bound improves upon this Bernstein-type result for $\gamma \gg V$, provided β is small enough. This allows us to see how a smaller β reduces the number of required samples.

Finally, our technique can be adapted to situations where we have additional information on the distribution of f, such as sub-Gaussian tails. Let s denote here, by abuse of notation, the measure on the vertices of the graph which corresponds to the stationary distribution.

Theorem 2.4. *In the above setting assume also that* $s(f \geqslant t) \leqslant Ce^{-Kt^2}$ *for positive* t*, and remove the assumption* $|f| \leqslant 1$*. Then*

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant \left\|\frac{q}{\sqrt{s}}\right\|_2 e^{-\frac{n}{2}\left(\gamma^2 K - \log\left(C\sqrt{\pi K}\gamma + 2\right)\right)},$$

as long as $\gamma \leqslant \log\left(\frac{1}{2\beta} + \frac{1}{2}\right)/2K\|f\|_\infty$.

We also have

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant \left\|\frac{q}{\sqrt{s}}\right\|_2 e^{2\gamma K} e^{-n\left(\gamma^2 K - \log\left(C\sqrt{\pi K}\gamma + 2\right)\right)},$$

as long as $\gamma \leqslant \log\left(\frac{1}{2\sqrt{\alpha}} + \frac{1}{2}\right)/2K\|f\|_\infty$.

For some parameter regimes Theorem 2.4 asymptotically improves upon Theorem 12 from [1].

3. Proofs of results in terms of β

In this section we will prove the inequalities involving β. Sketches of proofs for inequalities involving α are deferred to the next section. Before we begin the proofs we introduce some notation. We let $\|u\|_{1/s} = \sum_i \frac{u(i)^2}{s(i)}$ denote the $\frac{1}{s}$-weighted ℓ_2-norm on $\mathbb{R}^N$. The inner product associated with this norm is $\langle u, v\rangle = \sum_i \frac{u(i)v(i)}{s(i)}$. When we refer to the standard ℓ_2-norm we will use the notation $\|\cdot\|_2$.

The transition matrix P is not necessarily symmetric, and so its eigenvectors need not be orthogonal (this would be the case only if G were a regular graph). Reversibility, however, promises that $s_j P_{ij} = s_i P_{ji}$, and so P is self-adjoint and its eigenvectors are mutually orthogonal with respect to the $\frac{1}{s}$-weighted Euclidean structure. Therefore the $\|\cdot\|_{1/s}$ norm of P restricted to the subspace orthogonal to s is β, the second-largest absolute value of the eigenvalues of P.

Proof of Theorem 2.1. The beginning of our proof is identical to that of Gillman and to those which follow its reasoning. Take $r \geqslant 0$. By Markov's inequality,

$$\mathbb{P}_q\left(\frac{1}{n}S_n > \gamma\right) \leqslant e^{-rn\gamma}\mathbb{E}_q e^{rS_n},$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

where the expectation can be directly expressed and estimated as

$$\sum_{(x_0,\ldots,x_n)\in G^{n+1}} \left(e^{rS_n} q(x_0) \prod_{i=0}^{n-1} (P^T)_{x_i,x_{i+1}} \right) = \langle s, (e^{rf} P)^n q \rangle \leqslant \|q\|_{1/s} \|P e^{rf}\|_{1/s}^n.$$

Here e^{rf} stands for the diagonal matrix with $e^{rf(i)}$ as diagonal entries, and the inner product is, we recall, the inner product associated with the $\frac{1}{s}$-weighted ℓ_2-norm.

At this point Gillman's proof and its variations symmetrize the operator so that its norm will equal its top eigenvalue, and use Kato's spectral perturbation theory to estimate this eigenvalue. Our proof, on the other hand, will proceed to simply estimate the norm directly. To do that we will use the equality

$$\|P e^{rf}\|_{1/s}^2 = \max_{\|u\|_{1/s}=1} \langle P e^{rf} u, P e^{rf} u \rangle.$$

In order to perform the computation we split the vector u into stationary and orthogonal components, $u = as + b\rho$, where ρ is normalized and orthogonal to s in the weighted Euclidean structure. Applying similar decompositions $e^{rf} s = xs + z\sigma$ and $e^{rf}\rho = ys + w\tau$ we get

$$\|P e^{rf}\|_{1/s}^2 = \max_{a^2+b^2=1,\ \rho,\sigma,\tau} \langle a(xs + zP\sigma) + b(ys + wP\tau), a(xs + zP\sigma) + b(ys + wP\tau) \rangle.$$

We open the inner product and obtain

$$\|P e^{rf}\|_{1/s}^2 = \max_{a^2+b^2=1,\ \rho,\sigma,\tau} a^2(x^2 + z^2\|P\sigma\|^2) + b^2(y^2 + w^2\|P\tau\|^2) + 2ab(xy + zw\langle P\sigma, P\tau\rangle).$$

Denote $p_\sigma = \|P\sigma\|^2$, $p_\tau = \|P\tau\|^2$ and $p_{\sigma,\tau} = \langle P\sigma, P\tau\rangle$. Our task is reduced to computing the ℓ_2-norm of the following 2 by 2 symmetric bilinear form:

$$\begin{pmatrix} x^2 + z^2 p_\sigma & xy + zwp_{\sigma,\tau} \\ xy + zwp_{\sigma,\tau} & y^2 + w^2 p_\tau \end{pmatrix}.$$

Applying standard computations to derive the norm, we get

$$\begin{aligned} \|P e^{rf}\|_{1/s}^2 &= \frac{1}{2}\Big[(x^2 + y^2 + z^2 p_\sigma + w^2 p_\tau) \\ &\quad + \big[(x^2 + y^2 - z^2 p_\sigma - w^2 p_\tau)^2 + 4z^2 w^2 (p_{\sigma,\tau}^2 - p_\sigma p_\tau) \\ &\quad + 4x^2 z^2 p_\sigma + 4y^2 w^2 p_\tau + 8xyzwp_{\sigma,\tau} \big]^{1/2} \Big] \\ &\leqslant \frac{1}{2}\Big[(x^2 + y^2 + z^2 p_\sigma + w^2 p_\tau) \\ &\quad + \big[(x^2 + y^2 - z^2 p_\sigma - w^2 p_\tau)^2 + 4(|xz\sqrt{p_\sigma}| + |yw\sqrt{p_\tau}|)^2 \big]^{1/2} \Big], \end{aligned}$$

where we used the Cauchy–Schwarz inequality $p_{\sigma,\tau}^2 \leqslant p_\sigma p_\tau$.

To estimate the square root we use the inequality $\sqrt{1 + X^2} \leqslant 1 + \frac{X^2}{2}$. This leads us to

$$\|P e^{rf}\|_{1/s}^2 \leqslant x^2 + y^2 + \frac{(|xz\sqrt{p_\sigma}| + |yw\sqrt{p_\tau}|)^2}{x^2 + y^2 - z^2 p_\sigma - w^2 p_\tau} \tag{3.1}$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

Note that this result depends on assuming that $x^2 + y^2 \geqslant z^2 p_\sigma + w^2 p_\tau$. (For the purposes of the proof of Theorem 2.4 we require the inequality

$$\|Pe^{rf}\|_{1/s}^2 \leqslant x^2 + y^2 + |xz\sqrt{p_\sigma}| + |yw\sqrt{p_\tau}|, \tag{3.2}$$

which is obtained by using $\sqrt{1+X^2} \leqslant 1 + |X|$, and depends on the same inequality.)

Let us now estimate the components of our formula. We recall that f has mean 0 with respect to the stationary distribution s and absolute maximum 1. We obtain

$$\begin{aligned} x = \langle e^{rf} s, s\rangle &= 1 + \frac{\langle fs, s\rangle r}{1!} + \frac{\langle f^2 s, s\rangle r^2}{2!} + \frac{\langle f^3 s, s\rangle r^3}{3!} + \cdots \\ &\leqslant 1 + V\left(\frac{r^2}{2!} + \frac{r^3}{3!} + \frac{r^4}{4!} + \cdots\right) \leqslant 1 + V(e^r - 1 - r). \end{aligned}$$

Note also that $|f| \leqslant 1$ implies that $x \leqslant e^r$, and that by the arithmetic–geometric mean

$$x = \sum_i s(i) e^{rf(i)} \geqslant e^{r\sum_i s(i) f(i)} = 1.$$

To estimate $y = \langle e^{rf}\rho, s\rangle = \langle e^{rf} s, \rho\rangle$, recall that ρ is normalized and orthogonal to s, and that $\langle fs, \rho\rangle \leqslant \|fs\|_{1/s} = \sqrt{V}$. We get

$$\begin{aligned} |y| = |\langle e^{rf} s, \rho\rangle| &= \langle s, \rho\rangle + \frac{\langle fs, \rho\rangle r}{1!} + \frac{\langle f^2 s, \rho\rangle r^2}{2!} + \cdots \\ &\leqslant \sqrt{V}\left(r + \frac{r^2}{2!} + \frac{r^3}{3!} + \cdots\right) \leqslant \sqrt{V}(e^r - 1) \end{aligned}$$

Note that $x^2 + y^2 \leqslant \|e^{rf} s\|_{1/s}^2 = \langle e^{2rf} s, s\rangle$, which, as in the computation of x above, is bounded by $1 + V(e^{2r} - 1 - 2r)$.

Next, using the same estimate as for y, we get $|z| = |\langle e^{rf} s, \sigma\rangle| \leqslant \sqrt{V}(e^r - 1)$. For $w = \langle e^{rf}\rho, \tau\rangle$ we use the estimate $|w| \leqslant e^r$, which also applies to x.

Finally, since the norm of P restricted to the subspace orthogonal to s is β, we deduce that $p_\sigma, p_\tau, p_{\sigma,\tau} \leqslant \beta^2$.

Now we plug our estimates into inequality (3.1), and derive

$$\begin{aligned} \|Pe^{rf}\|_{1/s}^2 &\leqslant 1 + V(e^{2r} - 1 - 2r) + \frac{\left(2e^r\sqrt{V}(e^r - 1)\beta\right)^2}{1 - \beta^2 e^{2r} - \beta^2 V(e^r - 1)^2} \\ &\leqslant 1 + V\left(e^{2r} - 1 - 2r + \frac{4\beta^2 e^{2r}(e^r - 1)^2}{1 - \beta^2(e^{2r} + V(e^r - 1)^2)}\right) \\ &\leqslant \exp\left(V\left(e^{2r} - 1 - 2r + \frac{4\beta^2 e^{2r}(e^r - 1)^2}{1 - \beta^2(e^{2r} + V(e^r - 1)^2)}\right)\right), \end{aligned}$$

as long as $1 \geqslant \beta^2(e^{2r} + V(e^r - 1)^2)$. To conclude, recall that

$$\mathbb{P}_q\left(\frac{1}{n} S_n > \gamma\right) \leqslant e^{-n\gamma r} \|q\|_{1/s} \|Pe^{rf}\|_{1/s}^n,$$

so we finally obtain

$$\mathbb{P}_q\left(\frac{1}{n} S_n > \gamma\right) \leqslant \min_{1 \geqslant \delta(\beta^2, r),\ r \geqslant 0} \left\| \frac{q}{\sqrt{s}} \right\|_2 e^{-n\left[\gamma r - \frac{1}{2} V(e^{2r} - 1 - 2r + \Delta(\beta^2, r))\right]}. \qquad \square$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

To derive the corollaries and Theorem 2.4, we only need to assign suitable values to r. We will restrict to the case $q = s$ in order not to have to carry the $\|\frac{q}{\sqrt{s}}\|_2$ term.

Proof of Corollary 2.2. Using the inequalities $(e^r-1)^2 \leqslant e^{2r}-1-2r$ and $e^{2r}+V(e^r-1)^2 \leqslant 2e^{2r}-1$ we bound the expression inside the exponent in inequality (2.1) by

$$-\frac{n}{2}\left[2\gamma r - V(e^{2r}-1-2r)\left(1+\frac{4\beta^2 e^{2r}}{1-\beta^2(2e^{2r}-1)}\right)\right]$$
$$\leqslant -\frac{n}{2}\left[2\gamma r - V(e^{2r}-1-2r)\frac{1+\beta^2(2e^{2r}+1)}{1-\beta^2(2e^{2r}-1)}\right].$$

Let $t \geqslant 1, C_\beta = 2\frac{1+\beta^2}{1-\beta^2}, \gamma \leqslant (t-1)\frac{1+\beta^2}{\beta^2}V$ and $2r = \log\left(1+\frac{\gamma}{tC_\beta V}\right)$. It is easy to verify that

$$C_\beta \geqslant \frac{1+\beta^2(2e^{2r}+1)}{1-\beta^2(2e^{2r}-1)} \leqslant tC_\beta,$$

so that the above expression will be bounded by

$$-\frac{n}{2}\left[2\gamma r - tC_\beta V(e^{2r}-1-2r)\right].$$

Substituting for r yields

$$-\frac{n}{2}\left[\gamma\log\left(1+\frac{\gamma}{tC_\beta V}\right) - tC_\beta V\left(1+\frac{\gamma}{tC_\beta V}-1-\log\left(1+\frac{\gamma}{tC_\beta V}\right)\right)\right]$$
$$\leqslant -\frac{n}{2}tC_\beta V\left[\left(1+\frac{\gamma}{tC_\beta V}\right)\log\left(1+\frac{\gamma}{tC_\beta V}\right) - \frac{\gamma}{tC_\beta V}\right] \leqslant -\frac{n}{2}\gamma\log\frac{\gamma}{etC_\beta V}.$$

□

Proof of Corollary 2.3. First, we will apply the inequalities $e^{2r}-1-2r \leqslant 2r^2e^{2r}$, $e^r-1 \leqslant re^r$, and $e^{2r}+V(e^r-1)^2 \leqslant 2e^{2r}-1$ to the exponent in inequality (2.1). The exponent then turns into

$$-\frac{n}{2}\left[2\gamma r - Vr^2e^{2r}\left(2+\frac{4\beta^2e^{2r}}{1-\beta^2(2e^{2r}-1)}\right)\right] = -n\left[\gamma r - Vr^2\frac{1+\beta^2}{(1+\beta^2)e^{-2r}-2\beta^2}\right]$$
$$\leqslant -n\left[\gamma r - Vr^2\frac{1+\beta^2}{(1+\beta^2)(1-2r)-2\beta^2}\right].$$

Now we set

$$r = \frac{1-\beta^2}{1+\beta^2}\frac{\gamma}{2(\gamma+V)}$$

and obtain the desired result. Note that using more careful estimates can lead to a sharper constant multiplying γ in the denominator. □

Proof of Theorem 2.4. In this proof we will not assume that $|f| \leqslant 1$. For the purpose of this proof we offer a different analysis of the bound $x^2+y^2 \leqslant \sum_i e^{2rf(i)}s(i)$. This is simply the expectation of e^{2rf} according to the measure s. We can now evaluate this quantity using the

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

sub-Gaussian information. We get

$$x^2 + y^2 \leqslant \int_{-\infty}^{\infty} e^{2rt} d(-s(f \geqslant t)) = \int_{-\infty}^{\infty} 2re^{2rt} s(f \geqslant t) dt$$
$$\leqslant 1 + \int_0^{\infty} 2re^{2rt} Ce^{-Kt^2} dt = 1 + C\sqrt{\frac{\pi}{K}} re^{r^2/K}.$$

Plugging this estimate into inequality (3.2) together with the simple estimates $x, w \leqslant e^{r\|f\|_\infty}$ and $y, z \leqslant e^{r\|f\|_\infty} - 1$ we obtain

$$\|Pe^{rf}\|_{1/s}^2 \leqslant 1 + C\sqrt{\frac{\pi}{K}} re^{r^2/K} + 2\beta(e^{2r\|f\|_\infty} - 1).$$

As noted, inequality (3.2) depends on taking $x^2 + y^2 \geqslant z^2\beta^2 + w^2\beta^2$, which is guaranteed as long as $\beta^2(2e^{2r\|f\|_\infty} - 1) \leqslant 1$. We will make the stronger assumption $\beta(2e^{2r\|f\|_\infty} - 1) \leqslant 1$, and obtain the bound

$$\|Pe^{rf}\|_{1/s}^2 \leqslant \left(C\sqrt{\frac{\pi}{K}} r + 2\right) e^{r^2/K}.$$

Recalling that

$$\mathbb{P}_q\left(\frac{1}{n} S_n > \gamma\right) \leqslant e^{-n\gamma r} \|q\|_{1/s} \|Pe^{rf}\|_{1/s}^n,$$

and setting $r = \gamma K$, we conclude the required

$$\mathbb{P}_q\left(\frac{1}{n} S_n > \gamma\right) \leqslant \|q\|_{1/s} e^{-n\frac{\gamma^2 K}{2}} (C\sqrt{\pi K}\gamma + 2)^{n/2}$$
$$= \left\| \frac{q}{\sqrt{s}} \right\|_2 e^{-\frac{n}{2}(\gamma^2 K - \log(C\sqrt{\pi K}\gamma + 2))}.$$

The condition $\beta(2e^{2r\|f\|_\infty} - 1) \leqslant 1$ now reduces to $\gamma \leqslant \log\left(\frac{1}{2\beta} + \frac{1}{2}\right)/2K\|f\|_\infty$. □

Remark. Note that our method allows us to increase γ as far as $\log\left(\frac{1}{2\beta^2} + \frac{1}{2}\right)/2K\|f\|_\infty$, where our estimate becomes trivial.

4. Proofs of results in terms of α

The differences between the proofs of results in terms of β and α are mostly computational, so I will only sketch the relevant differences.

Proof of Theorem 2.1. As above, our task is to estimate $\|(Pe^{rf})^n\|_{1/s}$. We will use the simple identity $Pe^{rf} = e^{-\frac{1}{2}rf} e^{\frac{1}{2}rf} Pe^{\frac{1}{2}rf} e^{\frac{1}{2}rf}$ to obtain

$$\|(e^{rf}P)^n\|_{1/s} \leqslant e^r \|\left(e^{\frac{1}{2}rf} Pe^{\frac{1}{2}rf}\right)^n\|_{1/s} \leqslant e^r \|e^{\frac{1}{2}rf} Pe^{\frac{1}{2}rf}\|_{1/s}^n.$$

Since the operator $e^{\frac{1}{2}rf} Pe^{\frac{1}{2}rf}$ is self-adjoint with respect to the weighted Euclidean structure, we have

$$\|e^{\frac{1}{2}rf} Pe^{\frac{1}{2}rf}\|_{1/s} = \max_{\|u\|_{1/s}=1} \langle e^{\frac{1}{2}rf} Pe^{\frac{1}{2}rf} u, u\rangle = \max_{\|u\|_{1/s}=1} \langle Pe^{\frac{1}{2}rf} u, e^{\frac{1}{2}rf} u\rangle.$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

Decomposing the vectors as in the β-case (with $\frac{1}{2}r$ replacing r) we get

$$\|e^{\frac{1}{2}rf}Pe^{\frac{1}{2}rf}\|_{1/s} = \max_{a^2+b^2=1,\ \rho,\sigma,\tau} \langle a(xs+zP\sigma)+b(ys+wP\tau), a(xs+z\sigma)+b(ys+w\tau)\rangle.$$

We open the inner product and obtain

$$\|e^{\frac{1}{2}rf}Pe^{\frac{1}{2}rf}\|_{1/s} = \max_{a^2+b^2=1,\ \rho,\sigma,\tau} a^2\big(x^2+z^2\langle P\sigma,\sigma\rangle\big)+b^2\big(y^2+w^2\langle P\tau,\tau\rangle\big)+2ab\big(xy+zw\langle P\sigma,\tau\rangle\big).$$

Our task is reduced to computing the ℓ_2-norm of the same 2 by 2 symmetric bilinear form as in the β-case, except that r is replaced by $\frac{1}{2}r$, and the definitions of the p's are now $p_\sigma = \langle P\sigma,\sigma\rangle$, $p_\tau = \langle P\tau,\tau\rangle$ and $p_{\sigma,\tau} = \langle P\sigma,\tau\rangle$.

The following identity still holds:

$$\begin{aligned}\|e^{\frac{1}{2}rf}Pe^{\frac{1}{2}rf}\|_{1/s} = \frac{1}{2}\Big[&(x^2+y^2+z^2p_\sigma+w^2p_\tau)\\ &+\big[(x^2+y^2-z^2p_\sigma-w^2p_\tau)^2+4z^2w^2(p_{\sigma,\tau}^2-p_\sigma p_\tau)\\ &+4x^2z^2p_\sigma+4y^2w^2p_\tau+8xyzwp_{\sigma,\tau}\big]^{1/2}\Big].\end{aligned}$$

This time, however, the treatment of the terms inside the square root is slightly more delicate. Let λ_i be the eigenvalues of P in descending order, and let $(\sigma^i)_i$ and $(\tau^i)_i$ be the coordinates of σ and τ respectively in terms of the associated orthonormal system. Define

$$p_\sigma^+ = \sum_{1>\lambda_i>0}\lambda_i(\sigma^i)^2 \quad\text{and}\quad p_\sigma^- = -\sum_{\lambda_i<0}\lambda_i(\sigma^i)^2,$$

and decompose p_τ and $p_{\sigma,\tau}$ analogously. By Cauchy–Schwarz $|p_{\sigma,\tau}^+| \leqslant \sqrt{p_\sigma^+p_\tau^+}$, and the same goes for the p^-'s.

All this yields

$$\begin{aligned}p_{\sigma,\tau}^2-p_\sigma p_\tau &= (p_{\sigma,\tau}^+-p_{\sigma,\tau}^-)^2-(p_\sigma^+-p_\sigma^-)(p_\tau^+-p_\tau^-)\\ &= ((p_{\sigma,\tau}^+)^2-p_\sigma^+p_\tau^+)+((p_{\sigma,\tau}^-)^2-p_\sigma^-p_\tau^-)+(p_\sigma^+p_\tau^-+p_\sigma^-p_\tau^+-2p_{\sigma,\tau}^-p_{\sigma,\tau}^+)\\ &\leqslant \Big(\sqrt{p_\sigma^+p_\tau^-}+\sqrt{p_\sigma^-p_\tau^+}\Big)^2\end{aligned}$$

and

$$\begin{aligned}&x^2z^2p_\sigma+y^2w^2p_\tau+2xyzwp_{\sigma,\tau}\\ &\quad= (x^2z^2p_\sigma^++y^2w^2p_\tau^++2xyzwp_{\sigma,\tau}^+)-(x^2z^2p_\sigma^-+y^2w^2p_\tau^-+2xyzwp_{\sigma,\tau}^-)\\ &\quad\leqslant \Big(|xz\sqrt{p_\sigma^+}|+|yw\sqrt{p_\tau^+}|\Big)^2.\end{aligned}$$

We now combine the two estimates to get

$$\begin{aligned}&4z^2w^2(p_{\sigma,\tau}^2-p_\sigma p_\tau)+4x^2z^2p_\sigma+4y^2w^2p_\tau+8xyzwp_{\sigma,\tau}\\ &\quad\leqslant 4\max(|xz|^2,|yw|^2,|zw|^2)\Big(\Big(\sqrt{p_\sigma^+p_\tau^-}+\sqrt{p_\sigma^-p_\tau^+}\Big)^2+\Big(\sqrt{p_\sigma^+}+\sqrt{p_\tau^+}\Big)^2\Big).\end{aligned}$$

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press

Since $\lambda_2 = \alpha$, all p^+'s are bounded by α. Note also that $p_\sigma^+ + \alpha p_\sigma^- \leqslant \alpha \|\sigma\|_{1/s}^2 = \alpha$, and the same goes for τ. So, in fact, the above is bounded by the expression

$$4\max(|xz|^2, |yw|^2, |zw|^2)\Big(\Big(\sqrt{p_\sigma^+(1-p_\tau^+/\alpha)} + \sqrt{p_\tau^+(1-p_\sigma^+/\alpha)}\Big)^2 + \Big(\sqrt{p_\sigma^+} + \sqrt{p_\tau^+}\Big)^2\Big).$$

Rearranging terms and using Cauchy–Schwarz, we get

$$\begin{aligned}&\Big(\sqrt{p_\sigma^+(1-p_\tau^+/\alpha)} + \sqrt{p_\tau^+(1-p_\sigma^+/\alpha)}\Big)^2 + \Big(\sqrt{p_\sigma^+} + \sqrt{p_\tau^+}\Big)^2\\ &\qquad\leqslant 2\Big(p_\sigma^+(1-p_\tau^+/\alpha) + p_\tau^+ + \sqrt{\big(p_\sigma^+ + p_\tau^+(1-p_\sigma^+/\alpha)\big)\big(p_\tau^+ + p_\sigma^+(1-p_\tau^+/\alpha)\big)}\Big) \leqslant 4\alpha.\end{aligned}$$

So we finally obtain

$$\begin{aligned}\|e^{\frac{1}{2}rf} P e^{\frac{1}{2}rf}\|_{1/s} = \frac{1}{2}\Big[&(x^2 + y^2 + z^2 p_\sigma + w^2 p_\tau)\\ &+ \big[(x^2 + y^2 - z^2 p_\sigma - w^2 p_\tau)^2 + 16\alpha \max(|xz|^2, |yw|^2, |zw|^2)\big]^{1/2}\Big].\end{aligned}$$

Using the same estimates as in the β-case, replacing r by $\frac{1}{2}r$ in the estimates of x, y, z and w, recalling that $p_\sigma, p_\tau \leqslant \alpha$, and finally changing the bound variable r into $2r$ we obtain the desired results. □

The other proofs derive from the remark following Theorem 2.1, which applies to the proof of Theorem 2.4 as well.

References

[1] Artstein-Avidan, S. and Milman, V. (2006) Logarithmic reduction of the level of randomness in some probabilistic geometric constructions. *J. Funct. Anal.* **235** 297–329.
[2] Bennett, G. (1962) Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.
[3] Dinwoodie, I. H. (1995) A probability inequality for the occupation measure of a reversible Markov chain. *Ann. Appl. Probab.* **5** 37–43.
[4] Gillman, D. (1998) A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.* **27** 1203–1220.
[5] Hoory, S., Linial, N. and Wigderson, A. (2006) Expander graphs and their applications. *Bulletin of the AMS* **43** 439–561.
[6] Kargin, V. (2007) A large deviation inequality for vector functions on finite reversible Markov Chains (2007). *Ann. Appl. Probab.* **17**(4) 1202–1221.
[7] Leon, C. A. and Perron, F. (2004) Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.* **14** 958–970.
[8] Lezaud, P. (1998) Chernoff type bound for finite Markov chains. *Ann. Appl. Probab.* **8** 849–867.

https://doi.org/10.1017/S0963548307008772 Published online by Cambridge University Press