The Annals of Applied Probability
1995, Vol. 5, No. 1, 37–43

A PROBABILITY INEQUALITY FOR THE OCCUPATION MEASURE OF A REVERSIBLE MARKOV CHAIN

BY I. H. DINWOODIE

Tulane University

A bound is given for a reversible Markov chain on the probability that the occupation measure of a set exceeds the stationary probability of the set by a positive quantity.

1. Introduction. Consider a reversible Markov chain $x = X_1, X_2, \dots$ on a finite state space S with irreducible transition matrix $\pi = (\pi_{ij})$. By reversible we mean that the stationary distribution μ and π satisfy $\mu_i \pi_{ij} = \mu_j \pi_{ji}$. Since π is irreducible, $\mu_i > 0$.

Let L_n be the occupation measure for the process, so that $L_n(A)$ is the proportion of time that the chain spends in the set $A \subset S$:

$$L_n(A) = \sum_{i=1}^{n} I_A(X_i)/n.$$

It is well known that $L_n(A) \to \mu_A$ a.s. Let $f\colon S \to \mathbb{R}$ satisfy $0 \le f_i \le 1$ and let $\mu_f = \int f\, d\mu$. We are interested in probabilities of deviations of the type $\{\int f\, dL_n - \mu_f \ge \varepsilon\}$, $\varepsilon > 0$.

As motivation for the results, we describe a few applications. Consider a connected graph with undirected edges bearing positive edge weights. Define the transition probability from vertex x to y to be the weight of the edge connecting x to y relative to the total weight of all edges attached to x. The resulting Markov chain on the vertices is reversible and irreducible. The stationary probability μ_x at vertex x for this random walk is proportional to the total weight of the edges attached to vertex x. If one is estimating the stationary probability $\mu(A)$ of a set of vertices A with $L_n(A)$, it is useful to know how long one must wait until the proportion of time spent in A is close to $\mu(A)$ with high probability. This problem was addressed by Gillman (1993), where different simulation methods were compared. Theorem 3.1 gives a way to compute sufficient waiting time. Recall that every irreducible reversible Markov chain can be described probabilistically as a random walk on a graph in this way.

A variation on the above estimation problem arises with the Metropolis algorithm. Here we are interested in simulating random elements in a very large set S with a prescribed distribution $\mu > 0$, or perhaps only with

Received April 1994; revised May 1994.

AMS 1991 *subject classification.* 60F10.

Key words and phrases. Markov chain, occupation measure, expander graphs, Metropolis algorithm.

prescribed ratios μ_y/μ_x. If we start with simply an irreducible and symmetric transition matrix K on S, then K can be modified to a reversible transition matrix π with stationary distribution μ. To define π, let $p(x, y) = \min\{\mu_y/\mu_x, 1\}$ and let π be the stochastic matrix that satisfies $\pi(x, y) = K(x, y)p(x, y)$, $x \neq y$. Thus, we only make a transition from x to y if both pointers $K(x, \cdot)$ and $p(x, \cdot)$ indicate y. Again, the question is how long one must wait until $L_n(A)$ is reliably close to $\mu(A)$. For the marginal law μ_n of the Markov chain, which is the nonrandom distribution of the coordinate process at time n, this problem was studied in Diaconis and Hanlon (1992). More specific applications of the algorithm can be found in Smith and Roberts (1993), where the convergence issues are also raised.

The large deviation result of Donsker and Varadhan (1975) says

$$\limsup \frac{1}{n} \log P_x\left\{\int f\,dL_n - \mu_f \geq \varepsilon\right\} \leq -\inf\left\{I(q) : \int f\,dq \geq \mu_f + \varepsilon\right\},$$

where I is the rate function $I(q) = \sup_{u>0} \int \log(u/\pi u)\,dq$. However, rather than asymptotic results, we are concerned with bounding the probability of the event $\{\int f\,dL_n - \mu_f \geq \varepsilon\}$. Our main result is Theorem 3.1, where such a bound is given involving computable quantities. There are almost no results of this type in the literature, and ours improves on that of Gillman (1993). The articles of Höglund (1976) and Nagaev (1961) are relevant, but do not lead readily to computable inequalities.

2. Spectral radius. Let $p = |S|$ and let M be the $p \times p$ diagonal matrix given by

$$M = \begin{pmatrix} \sqrt{\mu_1} & 0 & \cdots & \cdots & \cdots \\ 0 & \sqrt{\mu_2} & 0 & \cdots & \cdots \\ 0 & 0 & \sqrt{\mu_3} & 0 & \cdots \\ 0 & 0 & 0 & \sqrt{\mu_4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

The matrix $\pi_\mu = M\pi M^{-1}$ has entry (i, j) given by $(\sqrt{\mu_i}/\sqrt{\mu_j})\pi_{ij}$ and is symmetric by the reversibility of π. The spectrum of π_μ is that of π: $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_p \geq -1$. By symmetry, these correspond to orthonormal eigenvectors $v_1 = M\mathbf{1}, v_2, \dots, v_p$. V will be the matrix with row i equal to v_i. The matrix π_μ allows one to extend techniques for symmetric chains to reversible chains and was used for a similar purpose in Diaconis and Stroock (1991).

For $t \geq 0$ let π_t be the matrix given by $\pi_t(i, j) \doteq \exp(t(f_i - \mu_f))\pi(i, j)$. Let r_t denote the spectral radius of the matrix π_t. Let Q be the symmetric, nonnegative definite matrix given by

$$Q_{ij} = \sum_{k=2}^{p} \frac{v_k(i)v_k(j)}{1 - \lambda_k}$$

and let $\beta = 1 - \lambda_2$ denote the spectral gap of π. The column vector $(f_i - \mu_f)$ will be denoted w_f.

LEMMA 2.1. *Let* $t \in [0, \ \beta/24]$. *Then* $r_t \le 1 + bt^2 + \beta 2^7(1 + 2/\beta)^3 t^3$, *where*

$$b = w_f^T MQMw_f$$

and furthermore $b \le 1/4\beta$.

PROOF. Let $\pi_{\mu,t}$ be the symmetric matrix given by

$$\pi_{\mu,t}(i,j) = \exp\big(t(f_i - \mu_f)/2\big)\pi_\mu(i,j)\exp\big(t(f_j - \mu_f)/2\big),$$

which clearly has spectral radius r_t as well. Let D be the diagonal matrix with entry (i, i) given by $f_i - \mu_f$. Then r_t is no greater than the spectral radius a_t of the matrix A_t for $t \in [0, 1]$, given by

$$A_t = \big[I + tD/2 + t^2D^2/4\big]\pi_\mu\big[I + tD/2 + t^2D^2/4\big],$$

since $\exp(t(f_i - \mu_f)/2) \le 1 + tD_{i,i}/2 + t^2D_{i,i}^2/4$ when $t \in [0, 1]$. Now A_t is a finite expansion of symmetric matrices in the parameter t:

$$A_t = \pi_\mu + \frac{t}{2}\big(\pi_\mu D + D\pi_\mu\big) + \frac{t^2}{4}\big(\pi_\mu D^2 + D\pi_\mu D + \pi_\mu D^2\big) + \frac{t^3}{8}\big(D^2\pi_\mu D + D\pi_\mu D^2\big) + \frac{t^4}{16}D^2\pi_\mu D^2.$$

It follows that the spectral radius a_t of A_t has a power series expansion

$$a_t = 1 + b_1t + b_2t^2 + b_3t^3 + \cdots$$

and Theorem 1 of Rellich [(1940), page 360] gives a radius of convergence of at least $\beta/24$ (in the notation of Rellich, $p = 1$, $a = 1$, $b = 0$ and $C = 8 + 16/\beta \le 24/\beta$). Furthermore,

$$|1 + b_1t + b_2t^2 - a_t| \le \frac{\beta}{4}\left(8 + \frac{16}{\beta}\right)^3 t^3 = \beta 2^7\left(1 + \frac{2}{\beta}\right)^3 t^3. \tag{2.1}$$

Now A_t is conjugate to the matrix $B_t = (1 + tD/2 + t^2D^2/4)^2\pi$, which can be written

$$B_t = \left(1 + tD + \frac{3t^2}{4}D^2 + \frac{t^3}{4}D^3 + \frac{t^4}{16}D^4\right)\pi.$$

Let v_t be the eigenfunction of norm $\sqrt{p}$ for the spectral radius a_t, so $v_0 = \mathbf{1}$. Both a_t and v_t have power series representations and we let

$$v_t(i) = \mathbf{1} + c_{1,i}t + c_{2,i}t^2 + O(t^3).$$

Now we have

$$B_t(v_t)i = \mathbf{1} + t\big[f_i - \mu_f + \pi c_1(i)\big] + t^2\Big[\tfrac{3}{4}\big(f_i - \mu_f\big)^2 + \big(f_i - \mu_f\big)\pi c_1(i) + \pi c_2(i)\Big] + O(t^3),$$

$$a_tv_t = \mathbf{1} + t(b_1\mathbf{1} + c_1) + t^2(b_2\mathbf{1} + b_1c_1 + c_2) + O(t^3).$$

Set $B_t(v_t) = a_t v_t$ and integrate with respect to μ. This implies that $b_1 = 0$ using the reversibility of π, and c_1 satisfies $\pi c_1(i) + f_i - \mu_f = c_{1,i}$. Now $(3/4)(f - \mu_f)^2 + (f - \mu_f)\pi c_1 + \pi c_2 = b_2\mathbf{1} + c_2$, which with the above equation becomes

$$\tfrac{3}{4}(f - \mu_f)^2 + (f - \mu_f)(c_1 + \mu_f - f) + \pi c_2 = b_2\mathbf{1} + c_2.$$

Integrating again with respect to μ and using the reversibility of π,

$$-\tfrac{1}{4}\sum_i \mu_i(f_i - \mu_f)^2 + \sum_i \mu_i(f_i - \mu_f)c_{1,i} = b_2.$$

Let V be the $p \times p$ matrix whose rows are the orthonormal eigenvectors v_k. Then

$$(\pi - I)c_1 = -w_f,$$

$$VM(\pi - I)M^{-1}V^T VMc_1 = VM(-w_f),$$

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 - 1 & 0 & \vdots \\ 0 & 0 & \lambda_3 - 1 & 0 \\ 0 & 0 & 0 & \lambda_4 - 1 \end{pmatrix} VMc_1 = VM(-w_f),$$

which means that $0 = VM(w_f)_1 = \Sigma_i \ \mu_i(f_i - \mu_f)$, comparing the first entries. Now

$$\begin{aligned} Mc_1 - \langle v_1, Mc_1 \rangle v_1 &\\ = V^T &\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1/(\lambda_2 - 1) & 0 & \cdots \\ 0 & 0 & 1/(\lambda_3 - 1) & 0 \\ 0 & 0 & 0 & 1/(\lambda_4 - 1) \end{pmatrix} VM(-w_f) \\ = QM(w_f)&, \end{aligned}$$

$$c_1 = M^{-1}QM(w_f) + \langle v_1, Mc_1 \rangle M^{-1}v_1 = M^{-1}QM(w_f) + \langle v_1, Mc_1 \rangle \mathbf{1},$$

since $v_1 = M\mathbf{1}$. Then from the expression above for b_2, we see that

$$\begin{aligned} b_2 &= \sum_i \mu_i(f_i - \mu_f)c_{1,i} - \tfrac{1}{4}\sum_i \mu_i(f_i - \mu_f)^2 \\ &= \sum_i \mu_i(f_i - \mu_f)M^{-1}QM(w_f) + \langle Mc_1, v_1 \rangle \sum_i \mu_i(f_i - \mu_f)\mathbf{1}_i \\ &\quad - \tfrac{1}{4}\sum_i \mu_i(f_i - \mu_f)^2 \\ &= w_f^T MQMw_f + 0 - \tfrac{1}{4}\sum_i \mu_i(f_i - \mu_f)^2. \end{aligned}$$

Let $b = w_f^T MQMw_f$. It follows that

$$b \leq \| Mw_f \|^2/\beta = \sum_i (f_i - \mu_f)^2 \mu_i/\beta \leq 1/4\beta.$$

Now with (2.1) it follows that for $t \in [0, \beta/24]$,

$$a_t \le 1 + b_2 t^2 + \beta 2^7 (1 + 2/\beta)^3 t^3$$
$$\le 1 + bt^2 + \beta 2^7 (1 + 2/\beta)^3 t^3. \qquad \square$$

3. A large deviation inequality. Let $\varepsilon \in [0, (8\beta + 16)^{-3}]$ and let $t = \varepsilon\beta$. Then the estimate from Lemma 2.1 gives

$$r_t \le 1 + \beta\varepsilon^2/2 \le \exp(\beta\varepsilon^2/2).$$

THEOREM 3.1. *For* $\varepsilon \in [0, (8\beta + 16)^{-3}]$,

$$P_x\{\textstyle\int f\,dL_n - \mu_f \ge \varepsilon\} \le \left(1 + 9\varepsilon\left(\frac{\beta + 2}{\sqrt{\mu_x}}\right)\right)\exp\left(\frac{-n\beta\varepsilon^2}{2}\right). \tag{3.1}$$

PROOF. Let D_t be the diagonal matrix with entry (i, i) given by exp $t(f_i - \mu_f)/2$. Then

$$\pi_{\mu,t}(i, j) =^{\text{def}} \exp\left(\frac{t(f_i - \mu_f)}{2}\right)\pi_\mu(i, j)\exp\left(\frac{t(f_j - \mu_f)}{2}\right) = D_t \pi_\mu D_t.$$

Thus $\pi_{\mu,t}^n = D_t^{-1} M \pi_t^n M^{-1} D_t$ and $\pi_t^n = M^{-1} D_t \pi_{\mu,t}^n D_t^{-1} M \le M^{-1} D_t A_t^n D_t^{-1} M$.

Let v_t be a right eigenvector for the symmetric matrix A_t with norm $\| M\mathbf{1} \| = 1$. Then the estimate of Rellich [(1940), page 360] gives

$$\| M\mathbf{1} - v_t \| \le \| M\mathbf{1} \| \tfrac{1}{2}\varepsilon\beta\, 8(1 + 2/\beta) = 4\varepsilon(\beta + 2).$$

Let δ_x denote the row vector given by $\delta_x(i) = \delta_{xi}$, $i = 1, \dots, p$. By the Markov inequality,

$$\begin{aligned}
&P_x\left\{\int f\,dL_n - \mu_f \ge \varepsilon\right\} \\
&\quad \le E_x \exp t\left(\sum^n f(X_i) - \mu_f - \varepsilon\right)\Big] = \exp(-nt\varepsilon)\,\delta_x \pi_t^n(\mathbf{1}) \\
&\quad = \exp(-nt\varepsilon)\,\delta_x M^{-1} D_t \pi_{\mu,t}^n D_t^{-1} M(\mathbf{1}) \\
&\quad \le \exp(-nt\varepsilon)\,\delta_x M^{-1} D_t A_t^n D_t^{-1} M(\mathbf{1}) \\
&\quad \le \exp(-nt\varepsilon)\,\delta_x M^{-1} D_t A_t^n (D_t^{-1} M(\mathbf{1}) - v_t) \\
&\qquad\qquad + \exp(-nt\varepsilon)\,\delta_x D_t M^{-1} A_t^n(v_t) \\
&\quad \le \exp(-nt\varepsilon)\, r_t^n \,\| \delta_x M^{-1} D_t \| \, \| D_t^{-1} M(\mathbf{1}) - v_t \| \\
&\qquad\qquad + \exp(-nt\varepsilon)\, r_t^n \delta_x D_t M^{-1} v_t \\
&\quad \le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\left(\| \delta_x M^{-1} D_t \| \, \| D_t^{-1} M(\mathbf{1}) - v_t \| + \delta_x D_t M^{-1} v_t\right)
\end{aligned}$$

$$
\begin{aligned}
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\left(\left(\frac{\exp(t/2)}{\sqrt{\mu_x}}\right)\right. \\
&\qquad\qquad \times(\| D_t^{-1} \| \| M(\mathbf{1}) - v_t \| + \| D_t^{-1} - I \| \| v_t \|) \\
&\qquad\qquad\qquad \left. + \exp\left(\frac{t}{2}\right)\frac{v_t(x)}{\sqrt{\mu_x}}\right) \\
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\exp\left(\frac{t}{2}\right) \\
&\quad \times\left(\left(\frac{1}{\sqrt{\mu_x}}\right)\left(\exp\left(\frac{t}{2}\right) \| M(\mathbf{1}) - v_t \|\right.\right. \\
&\qquad\qquad \left.\left. + \left(\exp\left(\frac{t}{2}\right) - 1\right) \| M\mathbf{1} \|\right) + \frac{v_t(x)}{\sqrt{\mu_x}}\right) \\
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\exp\left(\frac{t}{2}\right) \\
&\quad \times\left(\left(\frac{1}{\sqrt{\mu_x}}\right)\left(\exp\frac{t}{2}\right) \| M(\mathbf{1}) - v_t \| + \left(\frac{t}{2}\right)\exp\left(\frac{t}{2}\right) + \frac{v_t(x)}{\sqrt{\mu_x}}\right) \\
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\exp(t)\left(\left(\frac{1}{\sqrt{\mu_x}}\right)\right. \\
&\qquad\qquad \left.\times\left(\left(\| M(\mathbf{1}) - v_t \| + \frac{t}{2}\right) + (M(\mathbf{1})_x + \| M(\mathbf{1}) - v_t \|)\right)\right) \\
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\exp(t)\left(\frac{2 \| M(\mathbf{1}) - v_t \|}{\sqrt{\mu_x}} + \frac{t}{\left(2\sqrt{\mu_x}\right)} + 1\right) \\
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)\exp(t)\left(8\varepsilon\frac{(\beta + 2)}{\sqrt{\mu_x}} + 1 + \frac{\varepsilon\beta}{\left(2\sqrt{\mu_x}\right)}\right) \\
&\le \exp\left(\frac{-n\beta\varepsilon^2}{2}\right)(1 + 2t)\left(8\varepsilon\frac{(\beta + 2)}{\sqrt{\mu_x}} + 1 + \frac{\varepsilon\beta}{\left(2\sqrt{\mu_x}\right)}\right) \\
&\le \left(1 + 9\varepsilon\frac{(\beta + 2)}{\sqrt{\mu_x}}\right)\exp\left(\frac{-n\beta\varepsilon^2}{2}\right). \qquad \square
\end{aligned}
$$

The technique for obtaining the constant in this paper is much cruder than the technique in Dinwoodie (1994). Here we use the Euclidean norm for the difference between a perturbed and an unperturbed eigenvector to get the constant, whereas one really wants the supremum norm of the difference. The use of the Euclidean norm accounts for the factor $1/\sqrt{\mu_x} \approx \sqrt{p}$ in the constant, as in the result of Gillman (1993), but our technique adds the factor ε to mitigate the effect of the $1/\sqrt{\mu_x}$. The bound in Theorem 3.1 is not optimal. We have tried to find a simple bound in terms of computable quantities, for an ε in an interval independent of the functional f and of a reasonable size.

Acknowledgments. Persi Diaconis suggested this problem. We also thank a referee for suggesting a useful improvement on the original result.

REFERENCES

DIACONIS, P. and HANLON, P. (1992). Eigenanalysis for some examples of the Metropolis algorithm. *Contemp. Math.* **138** 99–117.

DIACONIS, P. and STROOCK, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** 36–61.

DINWOODIE, I. H. (1994). Occupation measure for random walk on the circle. Unpublished manuscript.

DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.* **28** 1–47.

GILLMAN, D. (1993). A Chernoff bound for random walks on expander graphs. In *Proceedings of the 34th Symposium on Foundations of Computer Science.* IEEE Computer Society Press, Los Alamitos, CA.

HÖGLUND, T. (1974). Central limit theorems and statistical inference for finite Markov chains. *Z. Wahrsch. Verw. Gebeite* **29** 123–151.

NAGAEV, S. V. (1961). More exact statements of limit theorems for homogeneous Markov chains. *Theory Probab. Appl.* **6** 62–81.

RELLICH, F. (1940). Störungstheorie der Spektralzerlegung. IV. *Math. Ann.* **117** 356–382.

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55** 3–23.

TULANE UNIVERSITY
DEPARTMENT OF MATHEMATICS
NEW ORLEANS, LOUISIANA 70118