

## Chernoff-type bound for finite Markov chains

Pascal Lezaud

► **To cite this version:**

Pascal Lezaud. Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, Institute of Mathematical Statistics (IMS), 1998, 8 (3), pp 849-867. hal-00940907

**HAL Id: hal-00940907**

**<https://hal-enac.archives-ouvertes.fr/hal-00940907>**

Submitted on 1 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## CHERNOFF-TYPE BOUND FOR FINITE MARKOV CHAINS

BY PASCAL LEZAUD

*CENA and Université Paul Sabatier*

This paper develops bounds on the distribution function of the empirical mean for irreducible finite-state Markov chains. One approach, explored by Gillman, reduces this problem to bounding the largest eigenvalue of a perturbation of the transition matrix for the Markov chain. By using estimates on eigenvalues given in Kato's book *Perturbation Theory for Linear Operators*, we simplify the proof of Gillman and extend it to nonreversible finite-state Markov chains and continuous time. We also set out another method, directly applicable to some general ergodic Markov kernels having a spectral gap.

**1. Introduction.** Let  $(X_n)$  be an irreducible Markov chain on a finite set  $G$  with transition matrix  $P$  and stationary distribution  $\pi$ . Then, for any function  $f$  and any initial distribution  $q$ , the weak law of large numbers states that for almost every trajectory of the Markov chain, the empirical mean  $n^{-1} \sum_{i=1}^n f(X_i)$  converges to  $\pi f = \sum_y \pi(y) f(y)$ . This result is the basis of the Markov chain simulation method to evaluate the mean of the function  $f$ . In this paper, we will quantify this rate of convergence by studying the probability

$$P_q \left[ n^{-1} \sum_{i=1}^n f(X_i) - \pi f \geq \gamma \right],$$

where  $P_q$  denotes the probability measure of the chain with the initial distribution  $q$ , and the size of deviation,  $\gamma$ , is a small number, such as  $\pi f/10$  or  $\pi f/100$ .

The corresponding rate of convergence for sums of independent random variables has been given by Kolmogorov (1929), Cramér (1938), Chernoff (1952), Bahadur and Ranga Rao (1960) and Bennett (1962), but the samples in Markov chains are generally correlated with each other, even if this correlation decreases exponentially with the number of steps. In that case, the theory of large deviations gives an asymptotic rate of convergence, since the empirical mean satisfies the large deviation principle with the good rate function  $I(z) = \sup_{r \in \mathbb{R}} \{rz - \log \beta_0(r)\}$ , where  $\beta_0(r)$  is the largest eigenvalue (i.e., Perron–Frobenius eigenvalue) of a perturbation of the transition matrix  $P$  [see Dembo and Zeitouni (1993)]. This asymptotic result is not satisfactory if one wants to achieve bounds that are useful for fixed  $n$ . Using perturbation theory, Gillman (1993) estimated the rate of convergence for reversible finite-state Markov chains by bounding the eigenvalue  $\beta_0(r)$ . He got a bound in terms of the spectral gap of the original transition matrix  $P$ , defined

---

Received December 1996; revised June 1997.

AMS 1991 subject classification. 60F10.

Key words and phrases. Markov chain, Chernoff bound, eigenvalues, perturbation theory.

by  $\varepsilon(P) = 1 - \beta_1(P)$ , where  $\beta_1(P)$  denotes the second largest eigenvalue of  $P$ . Using a technique adapted from Rellich, Dinwoodie (1995) improved the bound obtained by Gillman, but only for  $\gamma$  small. Improvements of such bounds are motivated by their wide use in simulation.

The aim of this work, based on Kato's perturbation theory, is to obtain a bound which depends on the variance of  $f$  and which is applicable for all  $\gamma$ . More precisely, this bound provides a Gaussian behavior for the small values of  $\gamma$  and a Poissonian behavior for the large values. Moreover, the result we achieved can also be easily extended to nonreversible Markov chains and continuous time. For instance, we get in the reversible case the following theorem, proved in Section 3.1, where  $\|\cdot\|_2$  denotes the  $l^2(\pi)$ -norm.

**THEOREM 1.1.** *Let  $(P, \pi)$  be an irreducible and reversible Markov chain on a finite set  $G$ . Let  $f: G \rightarrow \mathbb{R}$  be such that  $\pi f = 0$ ,  $\|f\|_\infty \leq 1$  and  $0 < \|f\|_2^2 \leq b^2$ . Then, for any initial distribution  $q$ , any positive integer  $n$  and all  $0 < \gamma \leq 1$ ,*

$$(1) \quad P_q \left[ n^{-1} \sum_{i=1}^n f(X_i) \geq \gamma \right] \leq e^{\varepsilon(P)/5} N_q \exp \left[ -\frac{n\gamma^2 \varepsilon(P)}{4b^2(1 + h(5\gamma/b^2))} \right],$$

where  $N_q = \|q/\pi\|_2$  and

$$h(x) = \frac{1}{2}(\sqrt{1+x} - (1-x/2)).$$

Therefore, if  $\gamma \ll b^2$  and  $\varepsilon(P) \ll 1$ , the bound is

$$(1 + o(1))N_q \exp \left[ -\frac{n\gamma^2 \varepsilon(P)}{4b^2(1 + o(1))} \right].$$

**REMARK 1.** For  $\gamma > 1$ , the probability of deviation is zero; thus, we can assume that  $\gamma \leq 1$ . Moreover, applying Theorem 1.1 to  $-f$  gives the same bound on  $P_q[n^{-1} \sum_{i=1}^n f(X_i) \leq -\gamma]$ . Thus, inequality (1) implies bounds on all moments of the positive random variable  $|n^{-1} \sum_i^n f(X_i)|$ , and quantifies convergence in each  $l^p$  norm,  $1 \leq p < \infty$ .

**REMARK 2.**  $h(x)$  is an increasing function of  $x \geq 0$  such that  $h(x) \leq x/2$  for all  $x$ . Thus, for  $\gamma \leq 2b^2/5$ , we get

$$P_q \left[ n^{-1} \sum_{i=1}^n f(X_i) \geq \gamma \right] \leq e^{\varepsilon(P)/5} N_q \exp \left[ -\frac{n\gamma^2 \varepsilon(P)}{4b^2} \left( 1 - \frac{5\gamma}{2b^2} \right) \right],$$

whereas for  $\gamma > 2b^2/5$ ,

$$P_q \left[ n^{-1} \sum_{i=1}^n f(X_i) \geq \gamma \right] \leq e^{\varepsilon(P)/5} N_q \exp \left[ -\frac{n\gamma \varepsilon(P)}{10} \left( 1 - \frac{2b^2}{5\gamma} \right) \right].$$

**REMARK 3.** Obviously, we can take  $b^2 = 1$  in inequality (1), which gives

$$(2) \quad P_q \left[ n^{-1} \sum_{i=1}^n f(X_i) \geq \gamma \right] \leq e^{\varepsilon(P)/5} N_q \exp[-n\gamma^2 \varepsilon(P)/12].$$

The constant  $N_q$  is the  $l_2(\pi)$ -norm of the density of  $q$  related to the stationary distribution  $\pi$ . This norm can be large but is always bounded by  $\pi_*^{-1/2}$ , where  $\pi_* = \min\{\pi(y) : y \in G\}$ . However, it is worth noting that  $N_q = 1$  when the initial distribution is the stationary one. This fact will be used to improve the convergence, by splitting the simulation in two steps, as suggested in Aldous (1987). The first step, the initial transient period, may be defined as the period of time which must be discarded before the chain comes close to  $\pi$ , whereas the second step, the observation period, may be defined as the period of time where observations must be collected to reach the desired precision.

A short description of the paper is as follows. Section 2 of this paper sets out preliminaries on the perturbation theory of linear operators. The results taken from Kato (1966) are based on a function-theoretical study of the resolvent, in particular on the expression of the eigenprojections as contour integrals of the resolvent. However, the specific results needed here are easier than in Kato (1966), since we restrict our study to the case of operators having an eigenvalue whose algebraic multiplicity is 1.

Section 3.1 proves Theorem 1.1. The bound is extended to nonreversible Markov chains in Section 3.2, by considering the multiplicative symmetrization of the operator  $P$ , namely,  $K = P^*P$ , where  $P^*$  is the adjoint of  $P$  on  $l_2(\pi)$ . Explicitly,

$$P^*(x, y) = \pi(y)P(y, x)/\pi(x).$$

Dinwoodie (1995) also obtained a bound for nonreversible Markov chain, but his bound depended on the random covering time  $T$  defined by  $T = \inf\{n \geq 1 : \{X_0, X_1, \dots, X_n\} = G\}$ . Bounding this covering time can be difficult.

Section 3.3 shows how the previous bounds extend to continuous-time Markov chains, by employing a method suggested by Fill (1991). We thus get an exponential bound in terms of the smallest nonzero eigenvalue of  $-(\Lambda + \Lambda^*)/2$ , where  $\Lambda$  is the generator of the semi-group  $P_t$ .

Perturbation theory of linear operators is a fruitful method to achieve explicit bounds on finite Markov chains. Mann (1996) used it to obtain a Berry–Esseen bound for Markov chains with finite or countably infinite state space, such that

$$(3) \quad \beta := \sup \left\{ \frac{\|Pg\|_2}{\|g\|_2} \right\} < 1,$$

where the sup is over all complex-valued functions on the state space  $G$  with  $\pi g = 0$ . Nagaev (1957, 1961) used also the perturbation method for studying some limit theorems for Markov chains under the Doeblin condition. Following the method used in a setting of the finite-state space, we recently have been able to achieve a Chernoff-type bound for general Markov operator  $P$  under the assumptions that 1 is an isolated eigenvalue of  $P$  and the dimension of the eigenspace for the eigenvalue 1 is finite. In particular, these assumptions are fulfilled for Markov chains satisfying the Doeblin condition or the relation (3). We will publish this result in a forthcoming paper. Instead of presenting these results, we would rather set out another method, directly applicable to a gen-

eral irreducible Markov kernel assuming condition (3). Section 4 introduces this method, due to Bakry and Ledoux, which gives the following inequality for  $\gamma \leq b^2$ :

$$P_\pi \left[ n^{-1} \sum_{i=1}^n f(X_i) - \pi f \geq \gamma \right] \leq \exp \left[ -n \frac{\gamma^2(1 - \beta)}{21b^2} \right].$$

Note that this last inequality involves  $\beta$  and not  $\varepsilon(P)$ . For instance, if the state space is finite and the Markov chain is periodic, then  $\beta = 1$ . Thus, this method is not applicable in this case, unlike the perturbation method.

Finally, Section 5 shows how the running time of simulation can be improved by considering an initial burning period whose effect is to decrease  $N_q$ .

**2. Perturbation theory of linear operators.** Let  $T$  be a linear operator on some finite-dimensional vector space  $X$ . We denote the spectrum of  $T$  by  $\Sigma(T)$ . The resolvent of  $T$ , defined by  $R(\zeta) = (T - \zeta)^{-1}$ , is an analytic operator-valued function on the domain  $\mathbb{C} \setminus \Sigma(T)$ , called the resolvent set. Furthermore, the only singularities of  $R(\zeta)$  are the eigenvalues  $\lambda_h, h = 1, \dots, s$ , of  $T$ .

The Laurent series of the resolvent at a simple eigenvalue  $\lambda_h$  takes the form [Kato (1966), page 38]

$$R(\zeta) = -(\zeta - \lambda_h)^{-1} P_h + \sum_{n=0}^{\infty} (\zeta - \lambda_h)^n S_h^{n+1},$$

where  $P_h$  is the eigenprojection operator for the eigenvalue  $\lambda_h$  and  $S_h$  is called the reduced resolvent of  $T$  with respect to the eigenvalue  $\lambda_h$ . More precisely,  $S_h$  is the inverse of the restriction of  $T - \lambda_h$  in the subspace  $(I - P_h)X$ . We deduce that  $P_h$  is the residue of  $-R(\zeta)$  in  $\lambda_h$ , and

$$P_h = -\frac{1}{2\pi i} \int_{\Gamma_h} R(\zeta) d\zeta,$$

where  $\Gamma_h$  is a positively oriented small circle enclosing  $\lambda_h$ , but excluding other eigenvalues of  $T$ .

Consider a family of operator-valued functions with the form

$$T(\chi) = T + \chi T^{(1)} + \chi^2 T^{(2)} + \dots$$

Then the resolvent  $R(\zeta, \chi) = (T(\chi) - \zeta)^{-1}$  of  $T(\chi)$  is analytic in the two variables  $\zeta, \chi$  in each domain in which  $\zeta$  is not equal to any of the eigenvalues of  $T(\chi)$  [Kato (1966), page 66]. So it can be expanded into the following power series in  $\chi$  with coefficients depending on  $\zeta$ :

$$(4) \quad R(\zeta, \chi) = R(\zeta) + \sum_{n=1}^{\infty} \chi^n R^{(n)}(\zeta),$$

where each  $R^{(n)}$  is an operator-valued function. This series is uniformly convergent if

$$(5) \quad \sum_{n=1}^{\infty} |\chi|^n \|T^{(n)} R(\zeta)\| < 1.$$

Let  $r(\zeta)$  be the value of  $|\chi|$  such that the left member of (5) is equal to 1. Then (5) is satisfied for  $|\chi| < r(\zeta)$ . Let  $\lambda$  be one of the eigenvalues of  $T = T(0)$  with multiplicity  $m = 1$ , and  $\Gamma$  be a positively oriented circle, in the resolvent set of  $T$ , enclosing  $\lambda$  but no other eigenvalues of  $T$ . The series (4) is then uniformly convergent for  $\zeta \in \Gamma$  if

$$(6) \quad |\chi| < r_0 = \min_{\zeta \in \Gamma} r(\zeta).$$

In the special case in which  $X$  is a Hilbert space and  $T$  is normal (i.e.,  $T^*T = TT^*$ ), we get

$$\|R(\zeta)\| = 1/\text{dist}(\zeta, \Sigma(T)),$$

$$r_0 = \min_{\zeta \in \Gamma} \left( \frac{a}{\text{dist}(\zeta, \Sigma(T))} + c \right)^{-1}$$

for every  $\zeta$  in the resolvent set of  $T$ , where  $a = \|T^{(1)}\|$  and  $c$  is such that  $\|T^{(n)}\| \leq ac^{n-1}$ . If we choose as  $\Gamma$  the circle  $|\zeta - \lambda| = d/2$ , where

$$d = \min_{\mu \in \Sigma(T) \setminus \{\lambda\}} |\lambda - \mu|,$$

we obtain  $r_0 = (2ad^{-1} + c)^{-1}$ .

The existence of the resolvent  $R(\zeta, \chi)$  for  $\zeta \in \Gamma$  implies that there are no eigenvalues of  $T(\chi)$  on  $\Gamma$ . The operator

$$(7) \quad P(\chi) = -\frac{1}{2\pi i} \int_{\Gamma} R(\zeta, \chi) d\zeta$$

is the eigenprojection for all the eigenvalues of  $T(\chi)$  lying inside  $\Gamma$ . In particular [Kato (1966), page 68], for all  $|\chi|$  sufficiently small, we have

$$\dim P(\chi)X = \dim PX = 1.$$

Therefore, only the eigenvalue  $\lambda(\chi)$  of  $T(\chi)$  lies inside  $\Gamma$ , and  $P(\chi)$  is the eigenprojection for this eigenvalue.

As the only eigenvalues of  $T(\chi)P(\chi)$  are 0 and  $\lambda(\chi)$ , we will consider

$$\lambda(\chi) - \lambda = \text{tr}((T(\chi) - \lambda)P(\chi)).$$

Combining (7) and substitution for  $R(\zeta, \chi)$  from (4) give the Taylor series expansion [Kato (1966), page 79]

$$\lambda(\chi) - \lambda = -\frac{1}{2\pi i} \text{tr} \int_{\Gamma} (\zeta - \lambda) R(\zeta, \chi) d\zeta = \sum_{n=1}^{\infty} \chi^n \lambda^{(n)},$$

where

$$(8) \quad \lambda^{(n)} = \sum_{p=1}^n \frac{(-1)^p}{p} \sum_{\substack{\nu_1 + \dots + \nu_p = n \\ k_1 + \dots + k_p = p-1 \\ \nu_i \geq 1, k_j \geq 0}} \text{tr}[T^{\nu_1} S^{(k_1)} \dots T^{\nu_p} S^{(k_p)}],$$

with  $S^{(0)} = -P$ ,  $S^{(n)} = S^n$ . Here  $S$  is the reduced resolvent of  $T$  with respect to the eigenvalue  $\lambda$ .

**3. Chernoff-type bound for Markov chains.** In this section, we consider an irreducible finite Markov chain with transition matrix  $P$  and stationary distribution  $\pi$ . This transition matrix defines an operator acting on functions by

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

We are interested in bounding the rate of convergence of

$$P_q[n^{-1}t_n \geq \gamma], \quad \text{where } t_n = \sum_{i=1}^n f(X_i) - \pi f.$$

If the Markov chain is reversible, the operator  $P$  is self-adjoint on  $l_2(\pi)$  endowed with the inner product

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

This operator  $P$  has largest eigenvalue 1 and because of irreducibility, the constant functions are the only eigenfunctions with eigenvalue 1. These eigenvalues are denoted by

$$\beta_0(P) = 1 > \beta_1(P) \geq \beta_2(P) \geq \dots \geq \beta_{|G|-1}(P) \geq -1,$$

and the spectral gap by  $\varepsilon(P) = 1 - \beta_1(P)$ .

If  $(P, \pi)$  is not reversible, we will consider the multiplicative symmetrization  $K = P^*P$  of  $P$ . This operator  $K$  is a reversible Markov kernel with same stationary distribution  $\pi$ . We will assume that  $K$  is irreducible. For instance, this will be the case as soon as  $P(x, x) > 0$  for every  $x \in G$ .

Since we will always work with a self-adjoint operator on  $l_2(\pi)$ , every norm considered will be the  $l_2(\pi)$ -norm. Moreover, except for constant functions without interest, replacing  $f$  by  $(f - \pi f)/\|f - \pi f\|_\infty$  shows that there is no loss of generality in assuming

$$\pi f = 0, \quad \|f\|_\infty \leq 1, \quad \|f\|_2^2 \leq b^2 \leq 1.$$

**3.1. Reversible Markov chains.** This section proves Theorem 1.1 stated in the Introduction, by using the method of Gillman (1993). We first prove the following lemma.

**LEMMA 3.1.** *Referring to the setting of Theorem 1.1, let  $r > 0$ . Then, for any positive integer  $n$ ,*

$$P_q[t_n/n \geq \gamma] \leq e^{-rn\gamma} \mathbf{q}^T P(r)^n \mathbf{1},$$

where  $P(r) = (e^{rf(y)} P(x, y))$ .

**PROOF.** By Markov's inequality,

$$P_q[t_n \geq n\gamma] \leq e^{-rn\gamma} E_{\mathbf{q}} \exp(rt_n),$$

where  $E_q$  denotes the expectation given the initial distribution  $q$ . Observe that

$$E_q \exp(rt_n) = \sum_{x_0, x_1, \dots, x_n} \exp(rt_n) q(x_0) \prod_{i=1}^n P(x_{i-1}, x_i),$$

where the summation is evaluated over all possible trajectories  $x_0, x_1, \dots, x_n$ . By introducing the operator  $P(r) = (e^{rf(y)}P(x, y))$ , we obtain the expression

$$E_q \exp(rt_n) = \mathbf{q}^T P(r)^n \mathbf{1}.$$

LEMMA 3.2. *Let  $r > 0$  and  $N_q = \|q/\pi\|_2$ . Then*

$$\mathbf{q}^T P(r)^n \mathbf{1} \leq e^r N_q \beta_0^n(r),$$

where  $\beta_0(P(r))$  is the largest eigenvalue of  $P(r)$ .

PROOF. Let us recall that  $P$  is a self-adjoint operator in  $l_2(\pi)$ , and introduce the diagonal matrix  $E_r = \text{diag}(e^{rf(y)})$ , so that  $P(r) = PE_r$ . Then  $P(r) = \sqrt{E_r^{-1}}\sqrt{E_r}P\sqrt{E_r}\sqrt{E_r}$  is similar to the self-adjoint matrix  $S(r) = \sqrt{E_r}P\sqrt{E_r}$ , so its eigenvalues are real values for  $r \geq 0$ . In this case, the Perron–Frobenius theorem says that  $\|S(r)\|_{2 \rightarrow 2} = \beta_0(P(r))$ , where  $\beta_0(P(r))$  is the largest eigenvalue of  $P(r)$ . Applying the Cauchy–Schwarz inequality in  $l_2(\pi)$  gives

$$\begin{aligned} \mathbf{q}^T P(r)^n \mathbf{1} &= \left\langle \frac{q}{\pi}, P(r)^n \mathbf{1} \right\rangle_{\pi} \\ &\leq N_q \|\mathbf{1}\|_2 \|E_{r/2}\|_{2 \rightarrow 2} \|E_{-r/2}\|_{2 \rightarrow 2} \|S(r)^n\|_{2 \rightarrow 2} \\ &\leq e^r N_q \beta_0^n(P(r)). \end{aligned}$$

Thus, Lemma 3.2 gives the inequality

$$(9) \quad P_q[t_n/n \geq \gamma] \leq e^r N_q \exp\{-n[r\gamma - \log \beta_0(P(r))]\}.$$

Let  $D = \text{diag}(f(x))$ ; then we can write

$$P(r) = P + \sum_{i=1}^{\infty} \frac{r^i}{i!} PD^i.$$

Since  $P \leq P(r) \leq e^r P$ , we get  $1 \leq \beta_0(P(r)) \leq e^r$  [see Kato (1966), Theorem I-6.44]. It follows that  $\beta_0(P(r))$  is the perturbation of the eigenvalue 1 of  $P$ . The irreducibility of  $P$  implies that 1 is a simple eigenvalue and the eigenprojection for the eigenvalue 1 is the operator  $\pi$  defined by

$$\pi f(x) := \sum_x f(x) \pi(x).$$

Thus (8) can be used to obtain the coefficients  $\beta^{(n)}$  in the Taylor series of  $\beta_0(P(r)) - 1$ , so long as  $r < r_0$ , where  $r_0$  is the convergence radius defined by



(6). We obtain

$$(10) \quad \beta^{(n)} = \sum_{p=1}^n \frac{(-1)^p}{p} \sum_{\substack{\nu_1+\dots+\nu_p=n \\ k_1+\dots+k_p=p-1 \\ \nu_i \geq 1, k_j \geq 0}} \frac{1}{\nu_1! \dots \nu_p!} \langle f^{\nu_1}, S^{(k_1)} P D^{\nu_2} \dots S^{(k_{p-1})} P f^{\nu_p} \rangle,$$

where we used the following relation whatever the matrix  $M$ :

$$\text{tr}(\pi P D^{\nu_1} M D^{\nu_p}) = \text{tr}(\pi D^{\nu_1} M D^{\nu_p}) = \langle f^{\nu_1}, M f^{\nu_p} \rangle_{\pi}.$$

For instance,

$$\beta^{(1)} = \langle f, 1 \rangle_{\pi} = 0, \quad \beta^{(2)} = \langle f, -Sf \rangle_{\pi} - \frac{1}{2} \langle f, f \rangle_{\pi}.$$

An explicit calculation gives  $\beta^{(2)} = \lim_{n \rightarrow \infty} E_{\pi} t_n^2 / n$ , which is precisely the asymptotic variance of  $t_n$ .

The number of terms in the formula (10) is

$$C(n) := \sum_{p=1}^n \binom{n-1}{p-1} \binom{2(p-1)}{p-1} \frac{1}{p}.$$

Furthermore, as  $(1/n!) \leq 2^{1-n}$  and  $\binom{2k}{k} < 2^{2k} (k\pi)^{-1/2}$ , we obtain the following inequalities for  $n \geq 3$ :

$$C(n) < 1 + \pi^{-1/2} \sum_{p=1}^{n-1} \binom{n-1}{p} \frac{1}{p+1} 4^p \leq \left(\frac{25}{4n}\right) 5^{n-2}.$$

Thus, for  $n \geq 7$ , we get  $C(n) \leq 5^{n-2}$ , but direct computations show this bound is also valid for  $n \geq 3$ . Finally, the Cauchy–Schwarz inequality gives

$$\beta^{(n)} \leq (b^2/5)(5/\varepsilon(P))^{n-1}, \quad n \geq 2.$$

Hence, provided that  $r < \varepsilon(P)/5$ , we obtain

$$(11) \quad \beta_0(P(r)) \leq 1 + \frac{b^2}{\varepsilon(P)} r^2 \left(1 - \frac{5r}{\varepsilon(P)}\right)^{-1}.$$

Combining inequalities (9) with (11) and  $\log(1+x) \leq x$  gives, for  $r < \varepsilon(P)/5$ ,

$$P_q[t_n/n \geq \gamma] \leq e^{\varepsilon(P)/5} N_q \exp(-nQ(r)),$$

where  $Q(r) = \gamma r - (b^2/\varepsilon(P))(1 - 5r/\varepsilon(P))^{-1} r^2$  is maximized when

$$r = \frac{\gamma \varepsilon(P)}{b^2(1 + 5\gamma/b^2 + \sqrt{1 + 5\gamma/b^2})}.$$

We can check that  $r < \varepsilon(P)/5$ , and simple computations yield inequality (1).  $\square$

3.2. *Nonreversible Markov chains.* This section extends the previous bound to nonreversible Markov chains. We now consider the multiplicative symmetrization of  $P$ , namely,  $K = P^*P$ . Set  $K(r) = P^*(r)P(r)$  and denote its largest eigenvalue by  $\beta_0(r)$ . As in the previous section, Markov's inequality gives

$$\begin{aligned} P_q[t_n/n \geq \gamma] &\leq e^{-rn\gamma} \mathbf{q}^T P(r)^n \mathbf{1} \\ &\leq e^{-rn\gamma} N_q \|\mathbf{1}\|_2 \|P(r)^n\|_{2 \rightarrow 2} \\ &= e^{-rn\gamma} N_q \|(P(r)^*)^n P(r)^n\|_{2 \rightarrow 2}^{1/2} \\ &= e^{-rn\gamma} N_q \sqrt{\beta_{0,n}(r)}, \end{aligned}$$

where  $\beta_{0,n}(r)$  denotes the largest eigenvalue of  $(P(r)^*)^n P(r)^n$ . Result analogous to Lemma 3.2 requires the use of Marcus' Theorem [see Marshall and Olkin (1979), Theorem 9.H.2.a]. This theorem states that  $\beta_{0,n}(r) \leq \beta_0(r)^n$ , whence the following inequality:

$$P_q[t_n/n \geq \gamma] \leq N_q \beta_0^{n/2}(r) e^{-rn\gamma}.$$

Here again, the operator  $K(r)$  will be considered as a perturbation of  $K$ , since we have

$$K(r) = K + \sum_{i=1}^{\infty} r^i \left( \sum_{j=0}^i \frac{1}{j!(i-j)!} D^j K D^{i-j} \right).$$

In the remainder of this section, we assume that  $K$  is irreducible. From here, the arguments are exactly the same as the ones used in the proof of Theorem 1.1. The coefficients  $\beta^{(n)}$  in the Taylor series of  $\beta_0(r) - 1$  are bounded by

$$\beta^{(n)} \leq 2b^2(2/\varepsilon(K))^{n-1} C(n) \leq (2/5)b^2(10/\varepsilon(K))^{n-1}.$$

Therefore, provided that  $r < \varepsilon(K)/10$ , we obtain

$$\beta_0(r) \leq 1 + \frac{4b^2}{\varepsilon(K)} r^2 \left( 1 - \frac{10r}{\varepsilon(K)} \right)^{-1};$$

hence,

$$P_q[t_n/n \geq \gamma] \leq N_q \exp \left[ -n \left( r\gamma - \frac{4b^2}{\varepsilon(K)} r^2 (1 - (10r)/\varepsilon(K))^{-1} \right) \right].$$

Optimizing in  $r < \varepsilon(K)/10$ , we may therefore state the following.

**THEOREM 3.3.** *Let  $(P, \pi)$  be an irreducible Markov chain on a finite set  $G$ , such that its multiplicative symmetrization  $K = P^*P$  is irreducible. Let  $f: G \rightarrow \mathbb{R}$  be such that  $\pi f = 0$ ,  $\|f\|_{\infty} \leq 1$  and  $0 < \|f\|_2^2 \leq b^2$ . Then, for any initial distribution  $q$ , any positive integer  $n$  and all  $0 < \gamma \leq 1$ ,*

$$P_q[t_n/n \geq \gamma] \leq N_q \exp \left[ -\frac{n\gamma^2 \varepsilon(K)}{8b^2(1 + h(5\gamma/b^2))} \right],$$

where  $\varepsilon(K)$  is the spectral gap of  $K$  and

$$h(x) = \frac{1}{2}(\sqrt{1+x} - (1-x/2)).$$

Obviously, with some modifications, the remarks stated in the Introduction are also valid.

**3.3. Continuous-time Markov chain.** In this section, we consider an irreducible continuous-time Markov chain on a finite-state space. If  $\pi$  denotes its unique stationary distribution, the weak law of large numbers states that, for any functions  $f$ ,

$$P\left[\frac{1}{t} \int_0^t f(X_s) ds - \pi f \geq \gamma\right] \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

We can express the previous integral as the following limit:

$$\frac{1}{t} \int_0^t f(X_s) ds = \lim_{k \rightarrow \infty} k^{-1} \sum_{i=1}^k f(X_{it/k}).$$

Fix  $k$  and assume that  $\pi f = 0$ ,  $\|f\|_\infty \leq 1$  and  $0 < \|f\|_2^2 \leq b^2$ . The Markov kernel  $P(t/k)$  satisfies

$$P(t/k)(x, y) > 0 \quad \forall t > 0 \quad \forall (x, y) \in G^2.$$

It follows that the Markov kernel  $K(t/k) = P^*(t/k)P(t/k)$  is irreducible. Let  $\beta_1(t/k)$  be the second largest eigenvalue of  $K(t/k)$  and denote its spectral gap by  $\varepsilon(t/k) = 1 - \beta_1(t/k)$ . Hence, for all  $0 \leq \gamma \leq 1$ , Theorem 3.3 implies

$$(12) \quad P_q\left[k^{-1} \sum_{i=1}^k f(X_{it/k}) \geq \gamma\right] \leq N_q \exp\left[-\frac{k\gamma^2\varepsilon(t/k)}{8b^2(1+h(5\gamma/b^2))}\right].$$

If  $\Lambda$  denotes the generator of the semigroup  $(P_t)$ , the generator of  $(P_t^*)$  is  $\Lambda^*$ . So, we deduce that, for  $k \rightarrow \infty$ ,  $K(t/k) = I + (t/k)(\Lambda + \Lambda^*) + o(1/k^2)$  and  $\varepsilon(t/k) = 2(t/k)\lambda_1 + o(1/k^2)$ , where  $\lambda_1$  denotes the smallest positive eigenvalue of  $-(\Lambda + \Lambda^*)/2$ . Now applying Fatou's lemma for  $k \rightarrow \infty$  to inequality (12), gives the following theorem.

**THEOREM 3.4.** *Let  $(P_t, \pi)$  be an irreducible continuous-time Markov chain on a finite set  $G$ , and  $\Lambda$  its infinitesimal generator. Let  $f: G \rightarrow \mathbb{R}$  be such that  $\pi f = 0$ ,  $\|f\|_\infty \leq 1$  and  $0 < \|f\|_2^2 \leq b^2$ . Then, for any initial distribution  $q$ , all  $t > 0$  and all  $0 < \gamma \leq 1$ ,*

$$P_q\left[\frac{1}{t} \int_0^t f(X_s) ds \geq \gamma\right] \leq N_q \exp\left[-\frac{\gamma^2\lambda_1 t}{4b^2(1+h(5\gamma/b^2))}\right],$$

where  $\lambda_1$  is the smallest positive eigenvalue of  $-(\Lambda + \Lambda^*)/2$  and

$$h(x) = \frac{1}{2}(\sqrt{1+x} - (1-x/2)).$$

Finally, bounding  $b^2$  by 1 gives the bound  $N_q \exp(-\gamma^2 \lambda_1 t / 12)$ , which is valid for all  $\gamma \leq 1$ .

REMARK. Applying Lemma 3.1 with the transition matrix  $P(t/k)$  and using the following formula valid for all matrices  $A, B$ :

$$\lim_{k \rightarrow \infty} (\exp(A/k) \exp(B/k))^k = \exp(A + B),$$

give

$$P_q \left[ \frac{1}{t} \int_0^t f(X_s) ds \geq \gamma \right] \leq e^{-r t \gamma} N_q \| \exp(\Lambda(r)t) \|_{2 \rightarrow 2},$$

where  $\Lambda(r) = \Lambda + r \operatorname{diag}(f(x))$ . Thus, we get a linear perturbation of the infinitesimal generator  $\Lambda$ . Moreover,  $\| \exp(\Lambda(r)t) \|_{2 \rightarrow 2} \leq \exp(-\lambda_0(r)t)$ , where  $\lambda_0(r)$  is the smallest eigenvalue of  $-(\Lambda + \Lambda^*)/2 - rD$ . By using the Trotter product formula [see Trotter (1959)], this method can be extended to general state space ergodic Markov processes.

**4. Direct method.** This section develops another method to obtain upper bound on the distribution function of the empirical measure. This method does not require the perturbation theory of linear operators and works easily in a more general setting.

From now on, we consider an ergodic Markov kernel  $P(x, dy)$  defined on a general state space  $G$ , with the stationary distribution  $\pi$ . This kernel defines an operator  $P$  acting on  $L^2(\pi)$  by

$$Pg(x) = \int_G g(y)P(x, dy).$$

Assume now that the condition (3) is fulfilled by  $P$ . Our goal is to obtain an upper bound for  $E_\pi \exp(r \sum_{i=1}^n f(X_i))$ , where  $0 \leq r$  and  $f$  is such that  $\|f\|_2^2 \leq b^2$ ,  $\pi f = 0$  and  $\|f\|_\infty \leq 1$ . Write  $g = \exp(rf)$ ; then

$$E_\pi \exp\left(r \sum_{i=1}^n f(X_i)\right) = \int Q^{(n)} g d\pi,$$

where, for all  $h \in L^2(\pi)$ ,

$$Q^{(1)}h(x) = \int P(x, dy)h(y),$$

$$Q^{(n+1)} = \int P(x, dy)g(y)Q^{(n)}h(y).$$

The principle of the proof, due to Bakry and Ledoux, consists of centering each successive term on which the operator  $P$  acts. Introduce some notation:

$$\alpha_0 = 1, \quad \alpha_i = E_\pi \exp\left(r \sum_{j=1}^i f(X_j)\right), \quad 1 \leq i \leq n,$$

$$g_1 = g - \int g d\pi, \quad g_i = gPg_{i-1} - \int gPg_{i-1} d\pi, \quad 2 \leq i \leq n,$$

$$b_1 = \int g d\pi, \quad b_i = \int gPg_{i-1} d\pi = \int g_1Pg_{i-1} d\pi, \quad 2 \leq i \leq n;$$

then

$$\begin{aligned} a_n &= a_{n-1}b_1 + \int Q^{(n)} g_1 d\pi \\ &= a_{n-1}b_1 + a_{n-2}b_2 + \int Q^{(n-1)} g_2 d\pi \\ &\vdots \\ &= \sum_{i=1}^n b_i a_{n-i}. \end{aligned}$$

Now, each  $b_i$  can be bounded by using (3), since the operator  $P$  acts on the centered functions  $g_{i-1}$ . Actually, let  $\alpha = \beta e^r$ ,  $a = b^2 e^r / 2$ ; then

$$\begin{aligned} b_1 &= \int \exp(rf) d\pi = 1 + \frac{r^2}{2} E_\pi f^2 + \frac{r^3}{3!} E_\pi f^3 + \dots \leq 1 + ar^2, \\ b_i &\leq \beta \alpha^{i-2} \|g_1\|_2^2, \quad 2 \leq i < n, \end{aligned}$$

where we used the inequality  $\|g_i\|_2 \leq \alpha^{i-1} \|g_1\|_2$ ,  $1 \leq i \leq n$ . Furthermore, by Jensen's inequality, we get  $\int \exp(rf) d\pi \geq 1$ , so

$$\begin{aligned} \|g_1\|_2^2 &= \int \exp(2rf) d\pi - \left( \int \exp(rf) d\pi \right)^2 \\ &\leq 1 + 2r^2 e^{2r} E_\pi f^2 - 1 \leq 4ar^2 e^r. \end{aligned}$$

Choose  $r < 1 - \beta$ , so  $\alpha < 1$ , since  $\log \beta \leq -(1 - \beta)$ . Now, as induction hypothesis, suppose that  $a_k \leq \phi^k(r)$ , where  $\phi(r) = 1 + Cr^2$  and  $C \geq 4a$  is independent of  $k$ . As

$$a_1 \leq \|g\|_2 \leq (1 + 4ar^2)^{1/2} \leq (1 + 2ar^2),$$

the hypothesis is true for  $k = 1$  and the induction hypothesis gives

$$\begin{aligned} a_k &\leq \left\{ (1 + ar^2) \phi^{k-1}(r) + 4ar^2 \alpha^{k-1} \sum_{i=0}^{k-2} (\phi(r)/\alpha)^i \right\} \\ &\leq \left\{ (1 + ar^2) \phi^{k-1}(r) + \frac{4ar^2 \alpha \phi^{k-1}(r)}{\phi(r) - \alpha} \right\} \\ &\leq \phi^{k-1}(r) \{1 + 4ar^2(1 - \alpha)^{-1}\}. \end{aligned}$$

Choosing  $C = 4a(1 - \alpha)^{-1}$ , it follows that the induction hypothesis is valid. Hence, for every  $0 \leq \gamma < 4b^2$  and  $r = \gamma(1 - \beta)/(4b^2)$ , we obtain, by Markov's inequality,

$$\begin{aligned} P_\pi[t_n/n \geq \gamma] &\leq \exp\{-n(r\gamma - 2b^2 r^2 e^r (1 - \beta e^r)^{-1})\} \\ &= \exp\left\{-\frac{n(1 - \beta)\gamma^2}{8b^2} \left(2 - \frac{(1 - \beta) \exp([\gamma(1 - \beta)]/4b^2)}{1 - \beta \exp([\gamma(1 - \beta)]/4b^2)}\right)\right\}. \end{aligned}$$

Assume now that  $\gamma \leq 2b^2$ . As  $e^x \leq 1 + 4x/3$ , for  $x \leq 1/2$ , we get the following:

$$(13) \quad P_\pi[t_n/n \geq \gamma] \leq \exp\left\{-\frac{n(1-\beta)\gamma^2}{8b^2}\left(1 - \frac{\gamma}{2b^2}\right)\right\}.$$

This method is directly applicable to general Markov chains and gives the same type of asymptotic bound as the one obtained by the perturbation method for finite nonreversible Markov chains. Observe, however, that for finite reversible Markov chains the bound (13) involves the constant  $\beta$ , instead of the second largest eigenvalue of the kernel  $P$ .

Inequality (13) can be extended to continuous time by assuming that  $f$  is continuous and bounded and

$$(14) \quad \sup\{\|P_t g\|_2/\|g\|_2, \pi g = 0\} \leq e^{-t\lambda} \quad \forall t \geq 0 \text{ and } \lambda > 0.$$

Arguing as in the proof of Theorem 3.4, we get that, for every  $\gamma \leq 2b^2$ ,

$$P_\pi\left[\frac{1}{t} \int_0^t f(X_s) ds \geq \gamma\right] \leq \exp\left\{-\frac{\gamma^2 t \lambda}{8b^2}\left(1 - \frac{\gamma}{2b^2}\right)\right\}.$$

An exponential inequality applicable for all  $0 \leq \gamma \leq 1$  can be obtained by following the proof of the Prokhorov's inequality for independent random variables given in Stout (1974). We obtain the following inequality, when  $q\pi, E_\pi f = 0, \|f\|_\infty \leq a$  and  $\|f\|_2^2 \leq b^2$ :

$$P_\pi[t_n/n \geq \gamma] \leq \exp\left\{-\frac{\gamma}{2a} \operatorname{arcsinh}\left(\frac{n\gamma a(1-\beta)}{4b^2}\right)\right\}.$$

PROOF. For every  $x \geq 0, e^x \geq ex$ . Therefore,  $E_\pi \exp(rt_n) \leq \exp[E_\pi \exp(rt_n) - 1]$ . Let  $Q(x) = P_\pi[t_n \leq x]$ ; we get

$$E_\pi \exp(rt_n) \leq \exp\left[\int (\exp(rx) - 1) dQ(x)\right].$$

Moreover,  $Q$  assigns mass 1 to the interval  $[-na, na]$  and

$$\int x dQ(x) = 0, \quad \int x^2 dQ(x) = E_\pi(t_n^2) \leq 2nb^2(1-\beta)^{-1}.$$

Thus,

$$\begin{aligned} E_\pi \exp(rt_n) &\leq \exp\left[\int (\exp(rx) - 1 - rx) dQ(x)\right] \\ &\leq \exp\left[2 \int (\cosh(rx) - 1) dQ(x)\right]. \end{aligned}$$

As the function  $h(x) = \cosh(rx) - 1$  is even and convex, we obtain

$$\int h(x) dQ(x) \leq \int h(x) dQ_1(x),$$

where  $Q_1$  assigns mass only to 0 and to  $na$ , with as much mass as possible assigned to  $na$  without violating

$$\int x^2 dQ_1(x) \leq 2nb^2(1 - \beta)^{-1};$$

thus,  $Q_1(na) \leq (2b^2)(n(1 - \beta)a^2)^{-1}$ . This yields

$$E_\pi \exp(rt_n) \leq \exp\left\{\frac{4b^2}{na^2(1 - \beta)}(\cosh(rna) - 1)\right\}.$$

Hence, for any  $r \geq 0$ ,

$$P_\pi[t_n/n \geq \gamma] \leq \exp\left\{-n\gamma r + \frac{4b^2}{na^2(1 - \beta)}(\cosh(rna) - 1)\right\}.$$

Differentiation shows that

$$r = (na)^{-1} \operatorname{arcsinh}(na\gamma(1 - \beta)(4b^2)^{-1}) := (na)^{-1}\theta$$

minimizes the right-hand side of the above inequality. It suffices to show  $\cosh \theta - 1 \leq (\theta \sinh \theta)/2$ . But the above inequality can easily be deduced from

$$\forall x \in \mathbb{R}, \quad e^x(2 - x) + e^{-x}(2 + x) \leq 4.$$

Another proof, due to Ledoux, can be developed for continuous-time Markov chain under the weaker assumption that the function  $f$  is bounded, measurable and hypothesis (14) is fulfilled. Let  $Y_t = \int_0^t f(X_s) ds$ ; then, for all  $k \geq 1$ ,

$$\begin{aligned} Y_t^k &= k! \int_0^t ds_1 \int_{s_1}^t ds_2 \cdots \int_{s_{k-1}}^t f(X_{s_1})f(X_{s_2}) \cdots f(X_{s_k}) ds_k, \\ E_\pi Y_t^k &= k! \int_0^t du_1 \int_0^{t-u_1} du_2 \cdots \\ &\quad \int_0^{t-(u_1+\cdots+u_{k-1})} du_k \int f P_{u_2}(f \cdots (f P_{u_k} f) \cdots) d\pi. \end{aligned}$$

Let  $a_{2,\dots,k} = |\int f P_{u_2}(f \cdots (f P_{u_k} f) \cdots) d\pi|$ ; then, using (14) and the centering method gives the inequality

$$\begin{aligned} a_{2,\dots,k} \leq \|f\|_2^2 &\left(\exp(-(u_2 + \cdots + u_k)\lambda) + a_2 \exp(-(u_4 + \cdots + u_k)\lambda) \right. \\ &\quad \left. + \cdots + a_{2,\dots,k-2} \exp(-u_k \lambda)\right). \end{aligned}$$

An iteration of this inequality shows that  $a_{2,\dots,k}$  is bounded by a sum of at most  $2^{k-1}$  of such terms:  $(b^2)^{s+1} \exp(-(m_2 + \cdots + m_k)\lambda)$ , where  $m_i = 0$  or  $u_i$  and  $s$  is the number of  $m_i \neq u_i$  with  $0 \leq s \leq \lceil k/2 \rceil - 1$ . This implies

$$\begin{aligned} E_\pi \exp(rY_t) &\leq 1 + \sum_{k=2}^\infty r^k 2^{k-1} \sum_{s=1}^{\lceil k/2 \rceil} \frac{(b^2 t)^s}{s!} \lambda^{s-k} \\ &\leq 1 + (2 - 4r/\lambda)^{-1} \sum_{s=1}^\infty \frac{(4b^2 r^2 t/\lambda)^s}{s!}, \end{aligned}$$

so long as  $2r/\lambda < 1$ . For  $2r/\lambda \leq 1/2$ , this yields  $E_\pi \exp(rY_t) \leq \exp[4b^2tr^2\lambda^{-1}]$ . Optimizing in  $r \geq \lambda/4$  together with Markov's inequality, it follows that, for every  $0 \leq \gamma \leq 2b^2$ ,

$$P_\pi \left[ \frac{1}{t} \int_0^t f(X_s) ds \geq \gamma \right] \leq \exp \left[ -\frac{\gamma^2 t \lambda}{16b^2} \right].$$

**5. Some reflections on simulation.** This last section shows how Theorem 1.1 can be combined with quantification of the closeness to stationarity, in order to compute the simulation time required to obtain a sufficiently accurate estimate. More precisely, we will consider the algorithm suggested by Aldous (1987).

THE CONSTANT  $N_q$ . As it has already been said before, the coefficient  $N_q = \|q/\pi\|_2$  may be large but always bounded by  $\pi_*^{-1/2}$ , where  $\pi_* = \min\{\pi(y) : y \in G\}$ . Furthermore,  $N_q = 1$  when  $q = \pi$ . This fact suggests that it might be better to begin the computation of  $t_n$  only when the distribution of the Markov chain is close to  $\pi$ . For instance, if the Markov chain is ergodic (i.e., irreducible and aperiodic), the distribution  $\pi_n$  of  $X_n$  converges to  $\pi$  exponentially as  $n$  tends to infinity. In that case, the simulation can be split into two phases. During the first one, the Markov chain approaches the stationary distribution sufficiently close to eliminate the "bias" from the initial position  $X_0$ , while the second one is the period of time necessary before  $t_n$  is a reasonable estimate of  $\pi f$ .

Let  $\pi_n$  be the distribution of  $X_n$  for the initial distribution  $q$ , and

$$N_n^2 = \|\pi_n/\pi\|_2^2 = \sum_x \pi(x) \left( \sum_y q(y) P^n(y, x) / \pi(x) \right)^2.$$

Then, the time until the distribution is close to stationary can be formalized by a parameter  $\tau$ , defined as

$$\tau = \min\{n : N_n \leq \sqrt{1 + 1/e^2}\} = \min\{n : \|(\pi_n/\pi) - 1\|_2 \leq 1/e\}.$$

From now on, we consider reversible and aperiodic Markov chains. Instead of working with  $N_n$ , we will use the chi-square distance  $\|(\pi_n/\pi) - 1\|_2$ . We have

$$\begin{aligned} \|(\pi_n/\pi) - 1\|_2^2 &\leq \sum_{x, y} \pi(x) q(y) (P^n(y, x) / \pi(x) - 1)^2 \\ &\leq \sup_y \|P^n(y, \cdot) / \pi(\cdot) - 1\|^2 = \|P^n - \pi\|_{2 \rightarrow \infty}^2 \\ &\leq \|P^{n_1}\|_{2 \rightarrow \infty}^2 \|P^{n_2} - E_\pi\|_{2 \rightarrow 2}^2, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality and  $n_1 + n_2 = n$ . From the last inequality, we deduce

$$\begin{aligned} \|(\pi_n/\pi) - 1\|_2 &\leq \min_{n_1+n_2=n} D(n_1) \|P^{n_2} - E_\pi\|_{2 \rightarrow 2} \\ (15) \qquad \qquad \qquad &\leq (1/\pi_* - 1)^{1/2} \mu(n), \end{aligned}$$



where  $D(n_1) = \|P^{n_1}\|_{2 \rightarrow \infty}$  and  $\mu(n) = \|P^n - E_\pi\|_{2 \rightarrow 2}$ . Finally, an upper bound for  $\mu(n)$  is accomplished by using the inequality  $\mu(n) \leq \beta_*(P)^n$ , where

$$\beta_*(P) = \max\{|\beta_{|G|-1}(P)|, \beta_1(P)\}.$$

See Diaconis and Saloff-Coste (1993) and Fill (1991) for more details. Combining this inequality with (15) gives

$$(16) \quad \|(\pi_n/\pi) - 1\|_2 \leq (1/\pi_* - 1)^{1/2} \beta_*(P)^n.$$

We are interested in determining the sufficient total number of steps for the simulation,  $n_e = \tau + n$ , such that

$$(17) \quad P_q \left[ \left| n^{-1} \sum_{i=\tau+1}^{\tau+n} f(X_i) \right| \geq \gamma \right] \leq \alpha,$$

where  $\alpha < 1$  is a small constant such as 0.05, for instance. Then with the setup of Theorem 1.1, the inequality (2) stated in Remark 3 gives

$$(18) \quad P_q \left[ \left| n^{-1} \sum_{i=\tau+1}^{\tau+n} f(X_i) \right| \geq \gamma \right] \leq 2 \exp[-n\gamma^2 \varepsilon(P)/12].$$

Moreover, different techniques can be used to get a bound on  $\tau$ , such as eigenvalues and geometric techniques or coupling [see Aldous and Fill, Fill (1991), Diaconis and Saloff-Coste (1996a, b)]. For instance, Poincaré inequality gives upper bound for  $\mu(n)$  [see Diaconis and Saloff-Coste (1996a)], whereas, when applicable, Nash and Log-Sobolev inequalities allow us to bound  $D(n)$  [see Diaconis and Saloff-Coste (1996b)].

EXAMPLES. The following examples deal with the Metropolis algorithm, a widely used tool in simulation. Consider  $G = \{0, 1, \dots, N\}$  and  $\pi$  a fixed distribution. We would like to construct an irreducible Markov chain  $M(x, y)$  whose stationary distribution is  $\pi$ . The Metropolis algorithm begins with a base chain  $P(x, y)$  on  $G$  which is modified by an auxiliary randomization. We will assume that  $P$  is irreducible and aperiodic and that  $P(x, y) > 0$  implies  $P(y, x) > 0$ . Let

$$A(x, y) = \begin{cases} \pi(y)P(y, x)/(\pi(x)P(x, y)), & \text{if } P(x, y) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Formally, the Markov kernel  $M$  is

$$M(x, y) = \begin{cases} P(x, y), & \text{if } A(x, y) \geq 1, \ x \neq y, \\ P(x, y)A(x, y), & \text{if } A(x, y) < 1, \\ P(x, y) + \sum_{z: A(x, z) < 1} P(x, z)(1 - A(x, z)), & \text{if } x = y. \end{cases}$$

We can easily prove that  $M$  is reversible and aperiodic. For the present examples, we take the base chain to be the nearest-neighbor random walk

$$P(x, x + 1) = P(x, x - 1) = 1/2, \quad 1 \leq x \leq N - 1,$$

$$P(0, 1) = P(0, 0) = P(N, N - 1) = P(N, N) = 1/2.$$

**BINOMIAL DISTRIBUTION.** The stationary distribution is  $\pi(x) = 2^{-N} \binom{N}{x}$ . The standard Metropolis construction gives [see Diaconis and Saloff-Coste (1996a)]

$$M(x, y) = \begin{cases} 1/2, & \text{if } \begin{cases} y = x + 1, & 0 \leq x \leq (N - 1)/2, \\ y = x - 1, & (N + 1)/2 \leq x \leq N, \end{cases} \\ x/2(N - x + 1), & \text{if } y = x - 1, 1 \leq x \leq (N + 1)/2, \\ (N - x)/2(x + 1), & \text{if } y = x + 1, (N - 1)/2 \leq x \leq N - 1, \\ (N - 2x + 1)/2(N - x + 1), & \text{if } y = x, 0 \leq x \leq (N + 1)/2, \\ (2x - N + 1)/2(x + 1), & \text{if } y = x, (N - 1)/2 \leq x \leq N. \end{cases}$$

Here  $1/N \leq 1 - \beta_1(M) \leq 2/N$ ,  $\pi_* = 2^{-N}$ , and  $\beta_N(M) \geq -1 + 2 \min M(x, x) \geq -1/2$ . Furthermore, it is shown in Diaconis and Saloff-Coste (1996a), by using the Log-Sobolev inequality, that

$$\|M^n(x, \cdot)/\pi - 1\| \leq (1 + 2e^2)^{1/2} e^{-c} \quad \text{for } n \geq (N/2)(\log N + 2c) + 1.$$

In other words, we find that approximate equilibrium is reached for  $\tau$  of the order of  $N \log N$ , whereas inequality (16) asserts approximate randomness for  $\tau$  of order  $N^2$ . We thus see, using inequality (2) (i.e., with  $\tau = 0$ ), that order  $(N/\gamma)^2$  steps are sufficient for condition (17) to hold, whereas inequality (18) improves the running time, since we get  $n_e$  of the order of  $N(\log N + 1/\gamma^2)$ .

**EXPONENTIAL FALL-OFF.** In this example, given in Diaconis and Saloff-Coste (1995), the stationary distribution is

$$\pi(i) = z(a) a^{h(i)}, \quad 0 < a < 1, \quad z \text{ the normalizing constant.}$$

Then, assuming

$$h(i + 1) - h(i) \geq c \geq 1, \quad 0 \leq i \leq N - 1,$$

the Metropolis chain is given by

$$M(i, i - 1) = 1/2, \quad M(i, i) = 2^{-1}(1 - a^{h(i+1)-h(i)}),$$

$$M(i, i + 1) = 2^{-1} a^{h(i+1)-h(i)}, \quad 1 \leq i \leq N - 1,$$

$$M(0, 0) = 1 - 2^{-1} a^{h(1)-h(0)}, \quad M(0, 1) = 2^{-1} a^{h(1)-h(0)},$$

$$M(N, N - 1) = M(N, N) = 1/2.$$

Here, the second eigenvalue of this chain satisfies

$$\beta_1 \leq 1 - 2^{-1}(1 - \alpha^{c/2})^2.$$

See Diaconis and Saloff-Coste (1995) for the proof based on a Poincaré inequality. Moreover,  $\beta_N \geq -1 + 2 \min M(x, x) \geq -1 + 2(1/2 - \alpha^c/2) = -\alpha^c$ . Thus,

$$(19) \quad \beta_* \leq \max(1 - 2^{-1}(1 - \alpha^{c/2})^2, \alpha^c).$$

Moreover, inequality (19) says order  $N$  steps are sufficient to reach stationarity. It follows that, using inequality (18), condition (17) holds for  $n_e$  of the order of  $N$ , whereas applying inequality (2) gives  $n_e$  of the order of  $N/\gamma^2$ .

**Acknowledgments.** Dominique Bakry, Laurent Saloff-Coste, Persi Diaconis and Michel Ledoux were exceptionally helpful with this paper. I would like to thank also the referee and Associate Editor for suggesting a number of improvements.

#### REFERENCES

- ALDOUS, D. (1987). On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Prob. Engrg. Inform. Sci.* **1** 33–46.
- ALDOUS, D. and FILL, J. Reversible Markov chains and random walks on graphs. Unpublished manuscript.
- BAHADUR, R. R. and RANGA RAO, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1033.
- BENNETT, G. (1962). Probability inequalities for sums of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.
- CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.
- CRAMÉR, H. (1938). Sur un nouveau théorème de la théorie des probabilités. *Actualités Sci. Indust.* **736**.
- DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.
- DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730.
- DIACONIS, P. and SALOFF-COSTE, L. (1995). What do we know about the Metropolis algorithm? Unpublished manuscript.
- DIACONIS, P. and SALOFF-COSTE, L. (1996a). Logarithmic Sobolev inequalities and finite Markov chains. *Ann. Appl. Probab.* **6** 695–750.
- DIACONIS, P. and SALOFF-COSTE, L. (1996b). Nash inequalities for finite Markov chains. *J. Appl. Probab.* **9** 459–510.
- DINWOODIE, I. (1995). A probability inequality for the occupation measure of a reversible Markov chain. *Ann. Appl. Probab.* **5** 37–43.
- FILL, J. (1991). Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.* **1** 62–87.
- GILLMAN, D. (1993). Hidden Markov chains: rates of convergence and the complexity of inference. Ph.D. dissertation, MIT.
- KATO, T. (1966). *Perturbation Theory for Linear Operators*. Springer, New York.
- KOLMOGOROV, A. (1929). Über das Gesetz des iterierten Logarithmus. *Math. Ann.* **99** 309–319.
- MANN, B. (1996). Berry–Esseen central limit theorem for Markov chains. Ph.D. dissertation, Harvard Univ.

- MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.
- NAGAEV, S. V. (1957). Some limit theorems for stationary Markov chains. *Theory Probab. Appl.* **2** 378–406.
- NAGAEV, S. V. (1961). More exact statements of limits theorems for homogeneous Markov chains. *Theory Probab. Appl.* **6** 62–81.
- STOUT, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.
- TROTTER, H. F. (1959). On the product of semi-groups of operators. *Proc. Amer. Math. Soc.* **10** 545–551.

CENTRE D'ETUDES DE LA NAVIGATION AÉRIENNE  
31055 TOULOUSE CEDEX  
AND  
LAB. STATISTIQUE ET PROBABILITÉ  
UNIVERSITÉ PAUL SABATIER  
31055 TOULOUSE CEDEX  
FRANCE  
E-MAIL: lezaud@cenatls.cena.dgac.fr