Threshold Limits for Cover Times

David J Aldous *
Department of Statistics
University of California
Berkeley CA 94720

Abstract

Under a natural hypothesis, the cover time for a finite Markov chain can be approximated by its expectation, as the size of state space tends to infinity. This result is deduced from an abstract result concerning covering an unstructured set by i.i.d. arbitrarily-distributed random subsets.

running head: Threshold limits for cover times.
key words. cover time, Markov chain, random walk on graph, threshold limit.

J. Theoretical Probability **4** (1991) 197-211.

*Research supported by N.S.F. Grant MCS87-01426

1 Results

Our results are easier to state than to motivate: so we give brisk statements in this section, followed by a more leisurely discussion in the next section.

Let S be a finite set. Let $\mathcal{S}$ be a random subset of S, whose distribution is arbitrary subject to the requirement

$$P(x \in \mathcal{S}) > 0 \text{ for each } x \in S. \tag{1}$$

Let $\mathcal{S}_1, \mathcal{S}_2, \ldots$ be independent random subsets distributed as $\mathcal{S}$. Let $\mathcal{R}_n$ be the *range* of this process:

$$\mathcal{R}_n = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \ldots \cup \mathcal{S}_n$$

and let C be the *cover time*

$$C = \min\{n : \mathcal{R}_n = S\} \geq 1.$$

Clearly $C < \infty$ a.s., and C has the submultiplicity (or "new better than used") property

$$P(C > m + n) \leq P(C > m)P(C > n); \;\; m, n > 0,$$

which implies that all moments of C are finite. We are interested in the distribution of C in settings where C is large, and a natural question is whether C/EC is concentrated near 1. It transpires that an answer involves the *terminal set* $\mathcal{T}$, that is the last uncovered portion of S:

$$\mathcal{T} = S \setminus \mathcal{R}_{C-1}.$$

For any $B \subseteq S$ let $C(B)$ be the cover time of B:

$$C(B) = \min\{n : \mathcal{R}_n \supseteq B\}$$

and let

$$c(B) = EC(B).$$

Here is our "abstract" result, proved in section 3.

Theorem 1 *(a)* $P(|\frac{C}{EC}-1|>\varepsilon)\leq\frac{4Ec(\mathcal{T})}{\varepsilon^2 EC};\ 0<\varepsilon\leq 1.$

(b) $var(\frac{C}{EC})\leq\kappa\frac{Ec(\mathcal{T})}{EC}$ *for an absolute constant* κ.

(c) $Ec(\mathcal{T})\leq 2+4s.d.(C)$.

Here "s.d." means standard deviation. The Theorem implies that C/EC is close to 1 iff $Ec(\mathcal{T})/EC$ is small. Think of $Ec(\mathcal{T})$ as the expected time for $\mathcal{T}$ to be covered by an *independent* sequence $(\mathcal{S}'_i)$.

Part (a) looks like Chebyshev's inequality applied to part (b) with $\kappa=4$, but the proof doesn't go that way: in fact we use (a) to prove (b) for some $\kappa>4$. It is conceivable that the variance in (b) has upper bound $O(Ec(\mathcal{T})/EC)^2$, which is the order of the lower bound implied by (c).

The distribution of C is determined by $(S,\mathcal{S})$. One can consider limit results for sequences $C^{(K)}$ derived from sequences $(S^{(K)},\mathcal{S}^{(K)})$: we shall omit the superscripts. For such a sequence, the submultiplicity property shows that convergence in probability

$$\frac{C}{EC}\xrightarrow{p}1$$

is equivalent to convergence of moments

$$E(\frac{C}{EC})^m\to 1;\ \text{each } m<\infty$$

and to convergence of exponential moments

$$E\exp(\alpha\frac{C}{EC})\to e^\alpha;\ \text{each } \alpha<\infty.$$

When these hold, say $C^{(K)}$ has a *threshold*. Theorem 1 implies

Corollary 2 *Suppose* $EC\to\infty$. *Then* C *has a threshold iff* $\frac{Ec(\mathcal{T})}{EC}\to 0$.

Without any structure being imposed on $(S,\mathcal{S})$ it is not clear how to estimate $Ec(\mathcal{T})$ in order to use these results; of course, without any structure it is unreasonable to expect any explicit results. When extra structure exists, one can seek to use Corollary 2 as a starting point. The following result is deduced from Theorem 1 in section 4.

Consider a Markov chain $(X_m;m\geq 0)$ on state-space S with initial state s and transition matrix Q. Let T_j be the first hitting time on j, let $V=\max_j T_j$ be the cover time (for X started at s), and let $\bar{t}=\max_{i,j}E_iT_j$. Here V is not exactly a cover time in the sense of Theorem 1, but is closely related to such a cover time defined in terms of the i.i.d. s-blocks of the chain.

Theorem 3 *Suppose* (S, Q, s) *vary in such a way that* $EV/\bar{t} \to \infty$. *Then* $V/EV \xrightarrow{p} 1$.

2 Motivation

This section is informal discussion.

2.1. There has been recent literature on the subject of the time V taken by a finite Markov chain to cover (i.e. to visit all its states), in particular in the special case of simple symmetric random walk on a finite graph. For a survey of the latter case see [**?**] and the papers following that paper. As one would expect, the existing literature deals with calculating asymptotics for EV (and, where possible, further distributional information) in concrete examples, and with finding general bounds for EV in some class of processes.

Our Theorem 3 has a novel form: one can show that $V/EV \stackrel{p}{\to} 1$ under conditions *weaker* than those necessary to actually calculate EV. This result is similar in spirit to martingale concentration inequalities which have recently been of great interest in probabilistic computer science/combinatorics settings [**?**]. It is designed for use in cases which are "critical" (in a sense explained below): here are 2 examples in the random-walk-on-graph setting.

Example: Balanced b-ary tree. The balanced b-ary tree of height K has asymptotically ($K \to \infty, b$ fixed) $b^{K+1}/(b-1)$ vertices. It is shown in [**?**] that the cover time V_K satisfies $EV_K \sim 2K^2 b^{K+1}(\log b)/(b-1)$, and that Theorem 3 applies to show that $V_K/EV_K \stackrel{p}{\to} 1$. (Order-of-magnitude bounds on EV_K were previously established in [**?**, **?**]).

Example: Discrete 2-dimensional torus. On the $K \times K$ discrete torus (i.e. Z_K^2) it is known only [**?**] that the cover time V_K satisfies asymptotically

$$a_1 K^2 (\log K)^2 \leq EV_K \leq a_2 K^2 (\log K)^2$$

for certain constants $a_1 < a_2$. Theorem 3 applies to show $V_K/EV_K \stackrel{p}{\to} 1$.

2.2. In the abstract setting of Theorem 1, two qualitatively different behaviors are possible. First, if $p(x) \equiv P(x \in \mathcal{S})$ is not constant, then the cover time C may be essentially just the time until $\mathcal{S}$ covers a particular point x_0 (or finite set of points) for which $p(x_0)$ is small. In this setting one naturally expects C/EC to have a non-degenerate limit distribution. Second, in a symmetric setting where $p(x) \equiv p$ it is easy to see that

$$c = \log(|S|)/p$$

is a natural upper bound in the following sense: if $p \to 0$ and $|S| \to \infty$ then

$$E\left(\frac{C}{c} - 1\right)^+ \to 0; \text{ and so } P(\left(\frac{C}{c} > 1 + \varepsilon\right) \to 0.$$

Here it may or may not be true that

$$EC/c \to 1; \; C/c \xrightarrow{p} 1. \tag{2}$$

Loosely, one expects this when the events $\{x \in \mathcal{S}\}$ are not too dependent as x varies. To prove (2) it suffices to obtain the lower estimate on EC:

$$\liminf EC/c \geq 1.$$

We call this the "natural method" in symmetric problems. Our results are aimed at critical cases on the borderline between these two behaviors (loosely, symmetric but highly- dependent cases). In the graph setting, the "critical" graphs turn out to be those like the examples above. In the first example, the limit is not the natural bound. In the second example, the best known upper bound is just the natural bound.

2.3. Another application is to covering with i.i.d. blocks. Let (ξ_i) be i.i.d. with distribution θ on a finite set I. Given K, consider the process $X_n = (\xi_n, \xi_{n+1}, \ldots, \xi_{n+K-1})$ as a Markov chain on I^K, and let V_K be its cover time. Where θ is uniform, it is not very difficult to show that as $K \to \infty$

$$EV_K \sim |I|^K \log |I|^K$$
$$V_K/EV_K \xrightarrow{p} 1. \tag{3}$$

See e.g. [**?**, **?**] for technical treatments, and [**?**] Chapter F for informal discussion. Consider the non-uniform case, and write $\theta_0 = \theta(i_0)$, say, for the minimum probability. In this case it is rather harder to obtain explicit sharp asymptotics for EV_K, although it is easy to get the weaker estimate $(EV_K)^{1/K} \to 1/\theta_0$. Nevertheless one can show that (3) still holds, by verifying the hypotheses of Theorem 3. To outline the argument briefly, one can show

$$\bar{t}_K \sim \theta_0^{-K}(1-\theta_0)^{-1} \tag{4}$$

the right side being asymptotically the mean hitting time on the string of all i_0's. Now consider the $K-1$ strings of length K, say $v_{K,1}, \ldots, v_{K,K-1}$, which contain $K-2$ occurences of i_0 and 2 occurences of some specified $i_1 \neq i_0$, with the first element of the string being i_1. The mean hitting time to such a string is asymptotically $\theta_0^{-(K-2)}\theta_1^{-2}$. It can be shown that these hitting times $(T_{v_{K,i}}; 1 \leq i \leq K-1)$ are asymptotically independent, so that

$$EV_K \geq E(\max_i T_{v_{K,i}}) \sim \theta_0^{-(K-2)}\theta_1^{-2} \log K. \tag{5}$$

Now (4,5) establish the hypothesis of Theorem 3.

2.4. The hypothesis of Theorem 3 is not quite a necessary condition. One issue is the cover time depends on the starting state. It is possible that mean cover times are asymptotically different for different starting states, or that a "threshold limit" holds for some starting states but not for others. Even to get the stronger result that the *same* threshold limit occurs for all starting states, the hypothesis is not necessary: consider the chain which cycles deterministically. However, the random-walk-on-graph setting forbids that example.

Proposition 4 *For simple symmetric random walk on a sequence of graphs G, suppose there exist constants $v = v(G) \to \infty$ such that $V/v \xrightarrow{p} 1$ for all starting states s. Then $\bar{t}/v \to 0$.*

Sketch of proof. It is enough to prove the result for the associated continuous-time random walks. Write T_{iji} for the time taken by the walk started at i to hit j and return to i. The key fact is: for each $x < \infty$ there exists $\delta(x) > 0$ such that for any continuous-time reversible Markov chain

$$P(T_{iji}/ET_{iji} > x) \geq \delta(x). \tag{6}$$

Under the hypothesis of the Proposition, by considering V for the walks started at i and j, we see

$$P(T_{iji} > 3v) \to 0.$$

The conclusion now follows from (6).

The proof of (6) involves the complete monotonicity properties of hitting times associated with continuous-time reversible Markov chains. Proposition 16 of [**?**] gives the related assertion

$$s.d.(T_{iji})/ET_{iji} \geq \sqrt{\frac{e-2}{2e-1}}$$

and similar arguments can be used to prove (6).

2.5. Let S be a compact metric space and let $(\mathcal{S}_i)$ be i.i.d. open subsets of S. A simple (not quite trivial) argument shows that (1) is still sufficient to imply that the cover time C is a.s. finite. Theorem 1 extends unchanged to this continuous setting. A variety of well-studied stochastic geometry problems fit into this continuous set-up. Here is a typical example. On a $K \times K$ square, throw down randomly (uniformly) centered discs of unit radius: let C_K be the number of discs needed to completely cover the square. Then

$$EC_K \sim (2/\pi) K^2 \log K$$

$$C_K/EC_K \xrightarrow{p} 1.$$

and there are sharper limits of the form (7). See [**?**] for technical treatment and [**?**] Chapter H for further discussion.

One original motivation for this study was to try to unify these "stochastic geometry" and "Markov chain" covering problems. In a superficial sense, we succeed: Theorem 3 is a Markov chain result which is proved from the abstracted stochastic geometry result, Theorem 1. Unfortunately, studying the usual stochastic geometry problems via Theorem 1 gives results weaker than those obtainable by the usual methods. It is conceivable that the theorem may be useful in more difficult stochastic geometry problems, but we do not have a convincing example.

2.6. Of course the classical (uniform) coupon-collector's problem is the case of Theorem 1 where $\mathcal{S}$ is a (uniform) random singleton. Asymptotics for this and related simple combinatorial models are easy: in uniform and near-uniform settings one gets $C_K/EC_K \xrightarrow{p} 1$ and the stronger "extreme value distribution" limit: there exist constants (b_K, c_K) such that $b_K \to \infty$ and

$$b_K(C_K - c_K) \xrightarrow{d} \xi; \quad P(\xi \leq x) = \exp(-e^{-x}). \tag{7}$$

Such limits also occur in stochastic geometry models, i.i.d. block models [**?**] and random walk on highly-symmetric graphs [**?**]. It would be interesting to give abstract hypotheses, analogous to those of Theorem 1, which implied the stronger conclusion (7).

3 Proof of Theorem 1.

Let $\mathcal{F}_i = \sigma(\mathcal{S}_1, \ldots, \mathcal{S}_i)$. Write $C_i = E(C|\mathcal{F}_i)$. In proving part (a), the central idea is the following lemma, whose proof is deferred.

Lemma 5 *$var(C_1) \leq 2Ec(\mathcal{T})$.*

Fix i, and consider an event of the form

$$F = \{\mathcal{S}_1 = A_1, \ldots, \mathcal{S}_i = A_i\}; \ \bigcup_{j=1}^{i} A_j \neq S.$$

Applying Lemma 5 to $S' = S \setminus \cup A_j$ and the random subsets $\mathcal{S}_{i+1}, \mathcal{S}_{i+2}, \ldots$, we deduce

$$\mathrm{var}((C_{i+1} - i)|F) \leq 2E(c(\mathcal{T})|F).$$

Since this holds for each atom F of $\mathcal{F}_i$,

$$\mathrm{var}(C_{i+1}|\mathcal{F}_i) \leq 2E(c(\mathcal{T})|\mathcal{F}_i)I(C > i) \tag{8}$$

where $I(\cdot)$ denotes an indicator r.v.. Now martingale orthogonality implies

$$\mathrm{var}(C_j) = \sum_{i=0}^{j-1} E\mathrm{var}(C_{i+1}|\mathcal{F}_i).$$

So, summing (8) over i,

$$\begin{aligned} \mathrm{var}(C_j) &\leq 2Ec(\mathcal{T}) \min(C, j) && (9) \\ &\leq 2jEc(\mathcal{T}). && (10) \end{aligned}$$

Now $EC_j = EC$ and

$$\begin{aligned} C_j &= C \text{ on}\{C \leq j\} \\ &> j \text{ on}\{C > j\} \end{aligned}$$

Put $j = \lfloor 2EC \rfloor$. Then for any $0 < \varepsilon \leq 1$,

$$\begin{aligned} P(|\frac{C}{EC} - 1| > \varepsilon) &= P(|\frac{C_j}{EC_j} - 1| > \varepsilon) \\ &\leq \frac{\mathrm{var}(C_j)}{\varepsilon^2 (EC_j)^2} \text{ by Chebyshev's inequality.} \end{aligned}$$

Part (a) of the Theorem now follows from (10).

To prove part (b), set $\alpha = \frac{4Ec(\mathcal{T})}{EC}$. Submultiplicity implies an absolute bound for $\mathrm{var}(C/EC)$, so we may suppose $\alpha \le 1/2$. Then part (a) implies

$$P(C > j) \le \alpha.$$

Then

$$\begin{aligned} EC^2 I(C > j) &= \sum_{n=1}^{\infty} EC^2 I(nj < C \le (n+1)j) \\ &\le \sum_{n=1}^{\infty} j^2 (n+1)^2 P(C > nj) \\ &\le j^2 \sum_{n=1}^{\infty} (n+1)^2 \alpha^n \text{ by submultiplicity} \\ &\le j^2 \alpha\beta \qquad (11) \end{aligned}$$

for some constant β, since $\alpha \le 1/2$. From the definitions of α and j, we obtain

$$EC^2 I(C > j) \le 4\beta (EC) Ec(\mathcal{T}). \qquad (12)$$

For C_j as above,

$$\begin{aligned} EC^2 I(C \le j) &= EC_j^2 I(C_j \le j) \\ &\le EC_j^2 \\ &= (EC)^2 + \mathrm{var}(C_j) \\ &\le (EC)^2 + 4(EC)Ec(\mathcal{T}) \text{ by (10)} \end{aligned}$$

Combining with (12) we get

$$\mathrm{var}(C) \le 4(1+\beta)(EC)Ec(\mathcal{T})$$

which gives part (b) of the Theorem.

Proof of Lemma 5. Because $C_1 = 1 + c(S \setminus \mathcal{S}_1)$,

$$\begin{aligned} \mathrm{var}(C_1) &= E(C_1 - EC)^2 \\ &= \sum_B P(\mathcal{S} = B)(1 + c(S \setminus B) - EC)^2 \\ &\le \sum_B P(\mathcal{S} = B) E(1 + C(S \setminus B) - C)^2 \qquad (13) \end{aligned}$$

using the inequality $(EX_1 - EX_2)^2 \leq E(X_1 - X_2)^2$. But $C(S \setminus B) \leq C$, and so

$$E(1 + C(S \setminus B) - C)^2 \leq 1 + \sum_{d=2}^{\infty} (d-1)^2 P(C(S \setminus B) = C - d).$$

Combining with (13) gives

$$\mathrm{var}(C_1) \leq 1 + \sum_B P(\mathcal{S} = B) \sum_{d=2}^{\infty} (d-1)^2 P(C(S \setminus B) = C - d). \quad (14)$$

We now consider $Ec(\mathcal{T})$. Let $\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2, \ldots$ be an independent sequence of random subsets with distribution $\mathcal{S}$, and write $\hat{C}(B)$ for cover times for this process. Then $Ec(\mathcal{T}) = E\hat{C}(\mathcal{T})$. We can write

$$\hat{C}(\mathcal{T}) - 1 = \sum_{j=1}^{\infty} I(\mathcal{R}_{C-1} \cup \hat{\mathcal{R}}_j \neq S).$$

Break over $\{C = i, \mathcal{S}_i = B\}$ and sum over i:

$$\begin{aligned}(\hat{C}(\mathcal{T}) - 1) I(\mathcal{S}_C = B) &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} I(C = i, \mathcal{S}_i = B, \mathcal{R}_{i-1} \cup \hat{\mathcal{R}}_j \neq S) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} I(\mathcal{S}_i = B) I(\mathcal{R}_{i-1} \cup \hat{\mathcal{R}}_j \neq S, \mathcal{R}_{i-1} \cup B = S).\end{aligned}$$

Take expectations, using the fact that for each (i,j) the indicator r.v.'s are independent.

$$E(\hat{C}(\mathcal{T}) - 1) I(\mathcal{S}_C = B) = P(\mathcal{S} = B) \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(\mathcal{R}_{i-1} \cup \hat{\mathcal{R}}_j \neq S, \mathcal{R}_{i-1} \cup B = S). \quad (15)$$

I now claim

$$P(\mathcal{R}_{i-1} \cup \hat{\mathcal{R}}_j \neq S, \mathcal{R}_{i-1} \cup B = S) = P(C > i + j - 1, C(S \setminus B) \leq i - 1). \quad (16)$$

For the left is the probability of a certain event defined in terms of $\mathcal{S}_1, \ldots, \mathcal{S}_{i-1}, \hat{\mathcal{S}}_1, \ldots, \hat{\mathcal{S}}_j$, and the right is the probability of the corresponding event defined in terms of $\mathcal{S}_1, \ldots, \mathcal{S}_{i-1}, \mathcal{S}_i, \ldots, \mathcal{S}_{i+j-1}$.

Since $C \geq C(S \setminus B)$, algebra gives

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(C > i + j - 1, C(S \setminus B) \leq i - 1) = \frac{1}{2} \sum_{d=2}^{\infty} d(d-1) P(C(S \setminus B) = C - d). \quad (17)$$

Putting together (17,16,15), summing over B and comparing with (14), we find

$$\begin{aligned} \text{var}(C_1) &\leq 1 + 2\sum_B E(\hat{C}(\mathcal{T}) - 1)I(\mathcal{S}_C = B) \\ &= 2E\hat{C}(\mathcal{T}) - 1 \\ &= 2Ec(\mathcal{T}) - 1. \end{aligned}$$

This establishes Lemma 5.

We now start the proof of part (c) of Theorem 1. Extend $(\mathcal{S}_i)$ to a doubly-infinite i.i.d. sequence $(\mathcal{S}_i : -\infty < i < \infty)$. For $i \geq 0$ let D_i be the cover time for the sequence $(\mathcal{S}_{1-i}, \mathcal{S}_{2-i}, \mathcal{S}_{3-i}, \ldots)$. Thus $D_0 = C$ and $D_i \stackrel{d}{=} C$ for each i. Now $D_i \leq C + i$, and $D_i = C + i$ iff $\mathcal{S}_0 \cup \mathcal{S}_{-1} \cup \ldots \cup \mathcal{S}_{1-i}$ does not cover $\mathcal{T}$. In other words,

$$P(D_i = C + i) = P(\hat{C}(\mathcal{T}) > i)$$

where $\hat{C}(\mathcal{T})$ is the number of terms of $\mathcal{S}_0, \mathcal{S}_{-1}, \ldots$ required to cover $\mathcal{T}$. So

$$\begin{aligned} P(\hat{C}(\mathcal{T}) > i) &= P(D_i = C + i) \\ &\leq E(D_i - C)^2/i^2 \\ &\leq 4 \text{ var}(C) \,/i^2 \end{aligned}$$

using the fact that $D_i \stackrel{d}{=} C$. Now put $z = 1 + [\sqrt{4 \text{ var}(C)}]$. Then

$$\begin{aligned} E\hat{C}(\mathcal{T}) &= \sum_{i=0}^{\infty} P(\hat{C}(\mathcal{T}) > i) \\ &\leq \sum_{i=0}^{\infty} \min(1, 4 \text{ var}(C)/i^2) \\ &\leq 1 + z + \sum_{i=z+1}^{\infty} 4 \text{ var}(C)/i^2 \\ &\leq 1 + z + 4 \text{ var}(C)/z \\ &\leq 2 + 2\sqrt{4 \text{ var}(C)} \\ &= 2 + 4 s.d.(C). \end{aligned}$$

Since $Ec(\mathcal{T}) = E\hat{C}(\mathcal{T})$, part (c) is established.

4 Proof of Theorem 3.

Let $(X_m; m \geq 0)$ be the chain started at s. Write ξ_i for the time of the i'th return to s. Then the random sets $\mathcal{S}_i$

$$\mathcal{S}_i = \{X_m : \xi_{i-1} \leq m < \xi_i\}$$

are i.i.d.. As in Theorem 1, let C be the cover time for these random subsets, and let $\mathcal{T}$ be the terminal set. The cover time V for the chain satisfies

$$\xi_{C-1} < V < \xi_C. \tag{18}$$

Define

$$\eta = \min_{j \in S} P(T_j < \xi_1). \tag{19}$$

The central idea of the proof is

Proposition 6 $Ec(\mathcal{T}) \leq 4(EC)^{3/4}\eta^{-1/4}$.

Under the limit hypotheses of Theorem 3 we shall show

$$\eta EC \to \infty \tag{20}$$

$$\frac{\xi_m}{mE\xi} \stackrel{p}{\to} 1 \text{ whenever } m = \Theta(EC) \tag{21}$$

for deterministic m depending on (S, Q, s). Then (20), Proposition 6 and Theorem 1 imply

$$C/EC \stackrel{p}{\to} 1.$$

It is easy to see that this, combined with (21) and (18), establishes Theorem 3.

To establish (20) and (21), first recall the hypothesis $EV/\bar{t} \to \infty$. Since Wald's identity and (18) give $EV \leq (EC)(E\xi_1)$, we obtain

$$\frac{\bar{t}}{(EC)(E\xi_1)} \to 0. \tag{22}$$

To establish (21), recall the renewal-theory relationship between inter-renewal times and equilibrium waiting times: applied to the renewal process (ξ_i), we obtain

$$E_\pi T_s = (\frac{E\xi^2}{E\xi} - 1)/2$$

where π is the stationary distribution. Thus

$$\bar{t} \geq E_\pi T_s \geq \frac{\text{var}(\xi)}{2E\xi}.$$

Using (22),

$$\frac{\text{var}(\xi)}{EC(E\xi)^2} \to 0.$$

Now (21) follows from Chebyshev's inequality.

To establish (20), let j achieve the minimum in the definition (19) of η. Then

$$\begin{aligned} 2\bar{t} &\geq E_s T_j + E_j T_s \\ &= E_s(\text{ time to hit } j \text{ and return to } s) \\ &= E\xi_1/\eta \end{aligned}$$

the last line using Wald's identity. Now (22) implies (20).

It remains to prove Proposition 6. From here, the arguments involve only a fixed chain (X_m). Let $\mathcal{F}_m$ be the natural σ-fields associated with the chain. We quote a martingale lemma, proved by the obvious stopping time argument: note the inequality is in the opposite direction from the usual L^1 maximal inequality.

Lemma 7 *Consider a martingale $M_n = P(A|\mathcal{F}_n)$, where $A \in \mathcal{F}_\infty$ and $M_0 = P(A)$. Let $P(A) \leq b < 1$. Then*

$$P(\ P(M_{n+1} \geq b|\mathcal{F}_n) > 0 \text{ for some } n) \geq P(A)/b.$$

For $k \neq s$ and $B \subseteq S$ let $\rho_k(B)$ be the chance that the chain, started from k and run until it first hits s, visits all states in B. Set $\rho_s(B) = 0$.

Lemma 8 *Let $P(\mathcal{S}_1 = S) \leq b < 1$. Then*

$$P(\rho_k(S \setminus \mathcal{S}_1) \geq b \text{ for some } k) \geq P(\mathcal{S} = S)/b.$$

Proof. Write

$$M_n = P(\mathcal{S}_1 = S|\mathcal{F}_n)$$
$$\Omega_n = \{P(M_{n+1} \geq b|\mathcal{F}_n) > 0\}.$$

Writing $\mathcal{B}_n = \{X_0, X_1, \ldots, X_n\}$, we have

$$M_{n+1} = \rho_{X_{n+1}}(S \setminus \mathcal{B}_n) \text{ on } \{\xi_1 > n\}.$$

So on $\{\xi_1 > n\}$,

$$\Omega_n \subseteq \{\rho_k(S \setminus \mathcal{B}_n) \geq b \text{ for some } k\}$$

and therefore

$$\cup_n \Omega_n \subseteq \{\rho_k(S \setminus \mathcal{B}_n) \geq b \text{ for some } k, \text{ for some } n < \xi_1\}.$$

But $\rho_k(S \setminus B) \uparrow$ as $B \uparrow$, so

$$\cup_n \Omega_n \subseteq \{\rho_k(S \setminus \mathcal{S}_1) \geq b \text{ for some } k\}.$$

The result now follows from Lemma 7.

To prove the Proposition, consider $k \in S$, $B \subseteq S$ and $b > 0$ such that $\rho_k(B) \geq b$. Then $P(B \subseteq \mathcal{S}_i) \geq \eta b$ since the chain has chance at least η to visit k during a s-block. So

$$W = \min\{i : B \subseteq \mathcal{S}_i\}$$

has $EW \leq 1/(\eta b)$. Now we can write

$$C \leq W_1 + W_2$$

where W_1 is the first i for which

$$\rho_k(S \setminus \mathcal{S}_i) \geq b \text{ for some } k),$$

and where W_2 is conditionally distributed as some W above. Lemma 8 implies $EW_1 \leq b/P(\mathcal{S}_1 = S)$, and hence

$$EC \leq \frac{b}{P(\mathcal{S}_1 = S)} + \frac{1}{\eta b}.$$

Minimizing over b gives

$$EC \leq 2/(\eta P(\mathcal{S}_1 = S))^{1/2},$$

and rearranging gives

$$P(\mathcal{S}_1 = S) \leq \frac{4}{\eta (EC)^2}. \tag{23}$$

Now the argument for (23) uses no assumptions other than the fact that $\mathcal{S}$ is the random set of states visited during an s-block of a Markov chain. Given $B \subseteq S$, we can consider a new Markov chain $\hat{X}$ which is the original chain watched only when it is in $B \cup \{s\}$. Applying (23) to this chain, we see that

$$P(B \subseteq \mathcal{S}_1) \leq \frac{4}{\eta(c(B))^2}. \tag{24}$$

This is enough for a crude bound on $c(\mathcal{T})$. Take $0 < \varepsilon < 1$ and an integer $K \geq 1$.

$$\begin{aligned}
P(c(\mathcal{T}) \geq \varepsilon EC, C \leq K) &= \sum_{i=1}^{K} P(C = i, c(\mathcal{T}) \geq \varepsilon EC) \\
&= \sum_{i=1}^{K} P(S \setminus \mathcal{R}_{i-1} \subseteq \mathcal{S}_i, c(S \setminus \mathcal{R}_{i-1}) \geq \varepsilon EC) \\
&\leq K \ \max\{P(B \subseteq \mathcal{S}_1) : c(B) \geq \varepsilon EC\} \\
&\leq \frac{4K}{\eta \varepsilon^2 (EC)^2} \text{ by (24).}
\end{aligned}$$

Bounding $P(C > K)$ by Markov's inequality,

$$\begin{aligned}
P(c(\mathcal{T}) \geq \varepsilon EC) &\leq \frac{4K}{\eta \varepsilon^2 (EC)^2} + \frac{EC}{K+1} \\
&\leq \frac{4}{\varepsilon (\eta EC)^{1/2}}
\end{aligned}$$

the latter by minimizing over K. Now $c(\mathcal{T}) \leq EC$, so

$$\begin{aligned}
E(\frac{c(\mathcal{T})}{EC}) &\leq \varepsilon + P(c(\mathcal{T}) \geq \varepsilon EC) \\
&\leq \varepsilon + \frac{4}{\varepsilon (\eta EC)^{1/2}} \\
&\leq \frac{4}{(\eta EC)^{1/4}}
\end{aligned}$$

the latter by minimizing over ε. This is Proposition 6.