Help Log out

COLT 2015 (author)

My Submissions | COLT 2015 | Help questions | Alerts | EasyChair

Update authors
Review response
Withdraw

COLT 2015 Submission 167

Click "Review response" to **see the reviews sent to you for the review response period and your response**.

If you want to **change any information** about your paper or withdraw it, use links in the upper right corner.

For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.

All **reviews sent to you** can be found at the bottom of this page.

Paper 167	
Title:	Single sample path based estimation of the mixing time in finite, reversible Markov chains
Submission	
Attachment	
Author keywords:	Markov chain mixing time spectral gap fully data-dependent confidence interval
EasyChair keyphrases:	markov chain (422), ingle sample path based estimation (232), reversible m arkov chain (200), sample path (183), mixing time (180), stationary distribution (160), spectral gap (160), ext ext (120), confidence interval (105), s ingle sample (95), fully data dependent confidence (80), s ingle sample path (80), ext diag (70), matrix tail inequality (63), data dependent confidence interval (60), operator norm (60), random matrix (60), data dependent confidence bound (60), data dependent (50), mixing coefficient (50), transition probability (50), absolute constant (50), mixing rate (50), perturbation result (50), ergodic markov chain (47), ingle sample path (47), reversible markov chain (47), second largest eigenvalue (47), data dependent confidence (42), symmetric matrix (40)
Topics:	computational learning theory, generalization bounds/statistical learning theory, unsupervised learning
Abstract:	We propose an efficient procedure for estimating the mixing time of a Markov chain. The novelty of our approach lies in several key aspects. The first is that we do not assume any a priori bounds on the mixing rate, which is in contrast with most other approaches that assume some known mixing rate. Another departure from common practice is the sufficiency of a single long sampling sequence, without access to a restart mechanism. Perhaps most importantly, we overcome an obstacle inherent in the chicken-and-egg nature of estimating mixing coefficients: to give confidence bounds one typically needs to know the mixing rate. We provide confidence guarantees that are fully data-dependent in the sense that they yield non-trivial finite-sample deviation bounds without any knowledge of the mixing coefficients.
Time:	Feb 20, 03:43 GMT

Authors

first name	last name	email	country	organization	Web site	corresponding?
Daniel	Hsu	djhsu@cs.columbia.edu	United States of America	Columbia University	http://www.cs.columbia.edu/~djhsu/	
Aryeh	Kontorovich	karyeh@cs.bgu.ac.il	Israel	Ben Gurion University	http://www.cs.bgu.ac.il/~karyeh/	✔
Csaba	Szepesvári	szepesva@cs.ualberta.ca	Canada	University of Alberta	http://www.ualberta.ca/~szepesva/	

Reviews

Review 2

Summary of the paper:	The work provides estimators for the mixing time of finite and reversible Markov chains from a single sampling sequence. The main result is a fully empirical confidence interval for the spectral gap.
Significance and scope:	The result is clean and solves a clearly defined and important problem about Markov chain estimation: that of providing confidence bound that do not depend on unknown coefficients.
Novelty and related work:	The proofs are clean and use some recent matrix perturbation theory results.
Soundness:	The proofs are detailed and the paper is generally self-consistent.
Summary of review:	A cleanly written paper solving an interesting problem of Markov chain estimation. Some comments about the necessity of the inverse dependence on pi* would be nice to have. Post-rebuttal: I share the concerns raised during the discussion phase about the correcteness of applying Bernstein's inequality to a sum of a random number of random variables. Some sloppiness in the presentation was detected as well. Nevertheless, I believe this paper will deserve publication once these issues are properly fixed.
Significance:	4: (good)
Novelty and related work:	3: (fair)
Soundness:	3: (fair)
Presentation:	4: (good)
Overall evaluation:	**6**: (marginally above acceptance threshold)
Reviewer's confidence:	**1**: (low)

Review 1

Summary of the paper:	This paper considers the problem of estimating the mixing time in Markov chains when one only has access to a single trajectory. For reversible chains, confidence intervals on the related spectral gap are provided that depend on the length of the trajectory, the spectral gap itself, the minimal value of the stationary distribution, as well as the size of the state space. This is complemented by lower bounds that show that dependence on these parameters is necessary.
Significance and scope:	The considered problem is definitely of general interest. However, the main restriction in my view is that the results only hold for reversible Markov chains. The paper mentions several possible applications, but does not discuss in which of them bounds for reversible chains would be sufficient. Moreover, the fact that the size of the confidence intervals in the main theorem depends on the value to estimate as well as some random quantities seriously confines the applicability, as there is no direct access to the size of the confidence interval and therefore to the quality of the estimate. Thus, overall the results seem to be only of theoretical interest.
Novelty and related work:	As far as I can tell, the paper makes a reasonable contribution. However, maybe the result of Theorem 1 (which the authors seem to consider not that important) may be of more interest than Theorem 4, as the confidence intervals for the latter contain some random terms.
Soundness:	There are a few issues with soundness and clarity of the paper: - First of all, it was not clear to me what is the difference between Theorem 1 and 4. For the former, it is claimed that it is not directly usable, since one cannot rule out that the value of pi* (the minimal value of the stationary distribution) is 0. However, I don't see why the computation of Algorithm 1 and the respective results of Theorem 4 are better in that respect. In general, the paper assumes that the spectral gap and the minimal stationary probability are positive, but it isn't discussed whether this is a serious restriction. - Similarly, the introduction and comparison of the two different values \hat{Delta} and \tilde{Delta} is rather confusing. If the results in Theorem 4 are for the tighter \hat{Delta}, why is \tilde{Delta} used in the algorithm? Also, I don't see why \hat{Delta} is tighter. Since \hat{Delta} has an additional factor of \sqrt{N_i} in the numerator, at least for big N_i the value of \tilde{Delta} should be the smaller one. In general, why care about a worse estimate anyway? - The confidence interval of Algorithm 1 / Theorem 4 contains the random values of N_i and hat{P}_ij, which makes the results in my view less applicable than those of Theorem 1. Moreover, the argument concerning the width of the confidence interval is rather confusing. First it is said that one considers the case that n goes to infinity, but then the bounding terms still contain n. Also, for the \hat{Delta}-term, I did not see what happened to the \sqrt{N_i}-term in the numerator when n goes to infinity. Last but not least, in the final step, it is not clear to me why the bounding term for

the product of \hat{rho}(delta) and the sqrt-term should be the sum (and not the product) of the individual bounding terms.

- Overall, the proof of Theorem 4 (I didn't check the proof of Theorem 1 in the appendix) could sometimes use a bit more information in some steps, like definition of notation, references to applied results, or some intermediate steps, cf. comments below.

- Finally, there also seems to be a problem with the estimation of the norm of the E-terms on p.10. First, in the bounds for ||E_pi,1|| and ||E_pi,2||, it seems that the two terms are mixed up, that is, for ||E_pi,1|| one should have the fraction pi/hat{pi} instead of hat{pi}/pi and vice versa for ||E_pi,2||. More seriously, it seems that the first inequality does not hold in general. If e.g. a:= \hat{pi}/pi = 1/4, then |\sqrt{a}-1| = 1/2 > 1/2*|a-1|=3/8.

Further comments and corrections

- p.2f: In the collection of the main results, some quantities (n, delta) are not explained. Also, the algorithm uses some notation like A^# that is only introduced later.

- p.6: In Theorem 3, the statement "For any initial state there exists a Markov chain ..." makes no sense, since the states usually do not exist independent of the Markov chain.

- p.6: In Theorem 4, I didn't understand what event the phrase "on the same event" refers to. Further, if n goes to infinity, it doesn't make sense to still have terms of n in the formula, so "n \tp \infty" should probably be dropped.

- p.7: In l.4 of the algorithm, 'P' should be '\hat{P}'.

- p.7: It is claimed that "the interval from Theorem 1 does not include zero", however this contradicts what has been claimed on p.5 in the discussion of Theorem 1.

- p.8: The application of the Bernstein bound should be explained in more detail.

- p.9: For the definition and the results (uniqueness) for the group inverse a reference would be appropriate.

- p.9: In eq.(6), the notation [...]_+ is used without explanation.

- p.9: A reference to Weyl's equality wouldn't hurt here.

- p.10: In the decomposition of L-\hat{L} it should be said that one also uses the definitions of L and \hat{L}. Also in the next step, it seems one uses that ||\hat{L}||=1, which should be mentioned as well.

- p.10: In the bound for ||E_P||, the exponents for hat{pi_i} and hat{pi_i} are missing.

Summary of review:	While the paper considers an interesting problem, the results only hold in the rather confined setting of reversible Markov chains and seem to be of little practical relevance. Further, there are some problems with soundness and clarity.
Significance:	3: (fair)
Novelty and related work:	4: (good)
Soundness:	2: (poor)
Presentation:	3: (fair)
Overall evaluation:	**4**: (reject: OK paper, but not good enough: issues with technical content, significance or originality.)
Reviewer's confidence:	**3**: (high)

Review 2

Summary of the paper:	The work provides estimators for the mixing time of finite and reversible Markov chains from a single sampling sequence. The main result is a fully empirical confidence interval for the spectral gap.
Significance and scope:	The result is clean and solves a clearly defined and important problem about Markov chain estimation: that of providing confidence bound that do not depend on unknown coefficients.

Novelty and related work:	The proofs are clean and use some recent matrix perturbation theory results.
Soundness:	The proofs are detailed and the paper is generally self-consistent.
Summary of review:	A cleanly written paper solving an interesting problem of Markov chain estimation. Some comments about the necessity of the inverse dependence on pi* would be nice to have.
Significance:	4: (good)
Novelty and related work:	4: (good)
Soundness:	4: (good)
Presentation:	4: (good)
Overall evaluation:	**7**: (accept: good paper)
Reviewer's confidence:	**1**: (low)

Review 3

Summary of the paper:

This paper proposes and analyzes a neat way of estimating the mixing time of a reversible (finite state) Markov chain in a fully data-dependent way by means of a single path dynamic of the chain. Upper and lower bounds are provided on the spectral gap and the minimum probability of the stationary distribution, the two quantities which are known to determine mixing rate.

Significance and scope:

Neatly presented and well-motivated paper on a relevant topic. I really enjoyed the argument that symmetrizes the matrix whose eigenvalues have to be estimated. I expect the argument of using a single run to be of practical relevance.

Comments/requests for clarification:
1) your lower bounds in Sect. 4.2 do not say anything specific about the effort to estimate \pi^*. Are the upper bouns in Thm 1 any tight ? Why do you need constant C, i.e., why is error in estimating \gamma^*
additive while the one for \pi^* multiplicative ?

2) Page 8, "by Bernstein inequality..." : which Bernstein ineq. are you using exactly here on N_{i,j} ? Pls clarify.

Minor comments (some of which are just typos) :
- page 2: "is assumes"
- page 2, last line: I'd say that n is the length of the sequence.
- page 4, def of {\hat \pi_i}, it should be "i \in [d]"
- page 6, "in he"
- page 7, Item 4 in Algorithm 1: P --> {\hat P}

Post discussion.

I still don't get how you can use the standard Bernstein ineq. for the i.i.d. setting when N_i is actually random ... The authors' rebuttal does not clarify this issue. There might be a fix to it, but I would like (have liked...) to see what's the impact on the bounds.

Novelty and related work:

From a technical standpoint, this paper is really hinging on replacing P with sym(L), then using tighter perturbation schemes, and replacing concentration bounds (obtained by standard blocking) by empirical versions thereof. Still, the result is a neat piece of work.
There are a number of papers on the web (e.g., arxiv) providing concentration inequalities for Markov chains (e.g., a' la MCDiarmid) that could perhaps provide alternative ways of estimating the mixing-relevant quantities considered here. One classic is

Samson, P.-M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes.
Ann. Probab. 28, pp. 416–461

and a more recent one is

D. Paulin, Concentration inequalities for Markov chains by Marton couplings and spectral methods. arxiv, 2015.

How are the authors positioning themselves compared to this (uncited!) literature ?

Soundness: All main claimes are well suppoted.

Summary of review:	This paper proposes and analyzes a neat way of estimating the mixing time of a reversible (finite state) Markov chain in a fully data-dependent way by means of a single path dynamic of the chain. Upper and lower bounds are provided on the spectral gap and the minimum probability of the stationary distribution, the two quantities which are known to determine mixing rate. Neat, well-motivated, and well-presented piece of work.
Significance:	5: (excellent)
Novelty and related work:	4: (good)
Soundness:	4: (good)
Presentation:	4: (good)
Overall evaluation:	**6**: (marginally above acceptance threshold)
Reviewer's confidence:	**1**: (low)

Review 3

Summary of the paper:

This paper proposes and analyzes a neat way of estimating the mixing time of a reversible (finite state) Markov chain in a fully data-dependent way by means of a single path dynamic of the chain. Upper and lower bounds are provided on the spectral gap and the minimum probability of the stationary distribution, the two quantities which are known to determine mixing rate.

Significance and scope:

Neatly presented and well-motivated paper on a relevant topic. I really enjoyed the argument that symmetrizes the matrix whose eigenvalues have to be estimated. I expect the argument of using a single run to be of practical relevance.

Comments/requests for clarification:
1) your lower bounds in Sect. 4.2 do not say anything specific about the effort to estimate \pi^*. Are the upper bouns in Thm 1 any tight ? Why do you need constant C, i.e., why is error in estimating \gamma^*
additive while the one for \pi^* multiplicative ?

2) Page 8, "by Bernstein inequality..." : which Bernstein ineq. are you using exactly here on N_{i,j} ? Pls clarify.

Minor comments (some of which are just typos) :
- page 2: "is assumes"
- page 2, last line: I'd say that n is the length of the sequence.
- page 4, def of {\hat \pi_i}, it should be "i \in [d]"
- page 6, "in he"
- page 7, Item 4 in Algorithm 1: P --> {\hat P}

Novelty and related work:

From a technical standpoint, this paper is really hinging on replacing P with sym(L), then using tighter perturbation schemes, and replacing concentration bounds (obtained by standard blocking) by empirical versions thereof. Still, the result is a neat piece of work.
There are a number of papers on the web (e.g., arxiv) providing concentration inequalities for Markov chains (e.g., a' la MCDiarmid) that could perhaps provide alternative ways of estimating the mixing-relevant quantities considered here. One classic is

Samson, P.-M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes.
Ann. Probab. 28, pp. 416–461

and a more recent one is

D. Paulin, Concentration inequalities for Markov chains by Marton couplings and spectral methods. arxiv, 2015.

How are the authors positioning themselves compared to this (uncited!) literature ?

Soundness: All main claimes are well suppoted.

Summary of review: This paper proposes and analyzes a neat way of estimating the mixing time of a reversible (finite state) Markov chain in a fully data-dependent way by means of a single path dynamic of the chain. Upper and lower bounds are provided on the spectral gap and the minimum probability of the stationary distribution, the two quantities which are known to determine mixing rate. Neat, well-motivated, and well-presented piece of work.

Significance: 5: (excellent)

Novelty and related work:	4: (good)
Soundness:	5: (excellent)
Presentation:	5: (excellent)
Overall evaluation:	**7**: (accept: good paper)
Reviewer's confidence:	**2**: (medium)