

Report on

“Bernstein’s inequalities for general Markov chains”

by Bai Jiang, Qiang Sun, and Jianqing Fan

The submitted paper deals with Bernstein-type inequalities. Undeniably, there already exists a vast body of concentration inequalities, both for dependent and independent samples. The paper promises to contribute in that

- (1) the results can be applied to a more general class of Markov chains (general-state-space, possibly non-reversible), and
- (2) the Hoeffding-type inequality proved in a previous paper by the same authors is complemented by a Bernstein variant, potentially giving sharper results in specific applications.

In this revision, the points of criticism of the previous report were taken up, whereby the crucial question of practicability of the proposed inequalities could only be answered unsatisfactorily. As the authors state themselves, “there is no effective methodology so far in the literature to compute or estimate spectral gaps for general-state-space Markov chains”. So, in what way should the extension to general-state-space Markov chains be a contribution if the concentration inequalities are formulated in terms of these non-computable and non-estimable spectral gaps? Would it not be appropriate to state that the proposed framework is not practicable in the general case?

Given the current version, which addresses issues previously somewhat swept under the carpet, the probabilistic contribution is small, all the more so for a submission to a high-class probability journal. The applications to robust mean estimation in Section 5 demonstrate that the paper makes a contribution with regard to question (2). For a submission to EJP, however, issue (1), aiming at the novel probabilistic contribution, is crucial, and this question was not answered convincingly. In view of this, the current version seems somewhat inflate; after reading the 31-page paper in detail, I found little substance. There is a moderate probabilistic contribution, which could make the paper interesting for ECP, but it should be clearly elaborated.

In any case, from my point of view, the following cuts would be appropriate:

- Section 2.2: Does this section offer any new insight? There should be a clear statement of what the specific contribution of the submitted work is, and what cases can now be addressed that were previously out of reach. In this respect, the entire section provides no information at all; it should therefore be omitted or at least very significantly shortened.
- Section 2.3: The heading “The Extendable and Non-extendable” seems inappropriate. Moreover, the results could easily be integrated into Section 2.1. Why are Theorem 1 and Theorem 3 formulated as separate results? The more general case described in Theorem 3 arises by means of standard arguments; even if the non-stationary case is proved directly, the proof structure is still very clear, so that the decomposition into separate results does not make sense. Theorem 4, which is not even proved separately, could be formulated more adequately as a remark.
- Section 3 should be substantially streamlined and condensed. Section 3.1, for example, merely introduces standard definitions, and it is inadequate to present them in such detail (with all the displays) in the main body of the paper. This is a submission to EJP, so it is inappropriate to reserve so much space for standard concepts. In the same spirit, it is not necessary to use the `{definition}`-environment for defining the

absolute/right spectral gap. It fits much better if the concepts are recalled just before stating the main result. The same remark applies to the definition of the León–Perron operator. Finally, Lemma 5 can be omitted; the corresponding result can simply be used with a precise reference to Paulin’s paper when it is needed.

- Section 5.1: The application to MCMC integral estimation feels out of place. My question regarding applicability was aimed at the new probabilistic contribution. It is a very weak argument to support the new inequalities when you, for a specific function f and a *reversible* Markov chain, use experimental results to infer that the variance proxy is “about 30% sharper” than previous versions. In my view, the section should be omitted.
- Section 6: The discussion provides a very long summary and recap of the proofs and does not really help to “highlight the technical novelty” of the paper. The section should therefore be reduced to one or two sentences. By the way, even when the paper was originally submitted to EJP, the reference Paulin (2015) was no longer really “recent” (as referred to at the beginning of the second-to-last paragraph on p.28). The authors would do well to do a more thorough research on the actual state-of-the-art and to update the references accordingly.

Minor comments/typos

- title and abstract: Bernstein’s inequalities belong to Bernstein. The term “Bernstein-type inequalities”, as it is also used in most of the paper, therefore seems more appropriate.
- p.1, line -3: Replace Frenchel by Fenchel.
- p.2, line -2: When submitting to EJP, a phrase like “the Markov dependence, an important dependence structure” seems out of place.
- p.4, l.1: If you mention it, it should be stated what kind of “impossibility result” you are showing, otherwise the remark is meaningless.
- p.8, Theorem 3, lhs of the second and third display: Replace $\sum_{i=1}^n f(X_i)$ with $\sum_{i=1}^n f_i(X_i)$.
- p.13, Lemma 6: Specify the range for k .
- Please edit your references. So far, many proper names in the titles are written in lower case (markov, metropolis-hastings, ...).