# Report on the manuscript
# "Bernstein's inequalities for general Markov chains"
# by Bai Jiang, Qiang Sun and Jianqing Fan

The paper under review presents Bernstein inequalities for functions of general Markov chains, imposing in particular no restrictions regarding reversibility of the chain or finiteness of the state space. Theorem 1 provides an estimate for the sum of possibly time-dependent bounded functions, with the upper bound involving the absolute spectral gap of the Markov operator induced by the transition kernel of the Markov chain. Similarly, Theorem 2 makes a statement for the mean of a constant bounded function, with upper bound depending on the right spectral gap of the Markov operator. For the special case of independent observations, the theorems in particular yield the classical Bernstein estimates.

Key ideas for the proof go back to Lezaud ([2]) who developed them in the context of finite state-space, reversible Markov chains. As detailed in the discussion in Section 6, a crucial step of Lezaud's approach cannot be transferred to the case of general state-space Markov chains. By introducing an extremal version of the original Markov operator $P$ (the so-called León–Perron operator $\widehat{P}$), this difficulty can be circumvented. The same approach has already been used by the authors to derive Hoeffding-type inequalities for Markov chains. The main contribution of the submitted paper is the extension of the analysis to the derivation of Bernstein inequalities.

**Conclusion** It is interesting to note that sharp Bernstein inequalities can be proved with the proposed method, and the presentation of the proof and its ideas is instructive and complete. In its current form, however, the manuscript offers too little substance to justify its publication in EJP. Given the authors' prior work on Hoeffding inequalities for general Markov chains in [1], one might first ask what benefit the Bernstein inequalities presented in Section 2 actually offer. The general statement that Bernstein inequalities can be sharper by taking the variance into account should be substantiated by a concrete application (e.g., to the examples from the context of statistical learning considered in [1]). Section 3 consists solely of definitions and a collection of existing results, before the proofs follow in Sections 4 (proofs of theorems) and 5 (technical proofs - why the separation?). The concluding discussion emphasises the technical novelties of the paper, but does not address the central issues of applicability and utility of the proven results.

**Further remarks**

- Section 1 consists of a short technical introduction and an overview of concentration inequalities for general Markov chains. In order to awaken interest among non-experts, it would be advantageous to describe at least one concrete application and to name it in the introduction. It is of course true that concentration inequalities "have found enormous applications"; however, this commonplace does not yet prove the usefulness of the versions presented in the submission.

- The discussion of the main results seems incomplete:
  - What statements can be expected, for example, for unbounded functions and the non-stationary case? You address all these questions in [1] - what about the Bernstein case?
  - The upper bounds in the two main theorems involve the absolute spectral gap and right spectral gap, respectively. With a view to statistical applications, concrete assumptions should be made about the underlying Markov chain from which a concrete evaluation of these spectral gaps can be derived. Is this possible? It is of course beneficial that the results presented do not assume reversibility of the Markov chain, as this is often difficult to verify in concrete examples. However, the applicability of the results presented in the submission is also not obvious to me.
  - Please demonstrate the benefit of your Bernstein inequalities compared to the Hoeffding-type inequalities in [1] by giving concrete examples.
- I don't really understand the point of the detailed listing of variance proxies given in Tables 1 and 2. It's nice to get precise results, but the derivation of concentration inequalities is certainly not an end in itself. Rather, the results should be measured by their applicability, and the paper makes no statements at all in this regard.

## References

[1] Fan, J., Jiang, B. and Sun, Q. (2018). Hoeffding's lemma for Markov chains and its applications to statistical learning. *arXiv preprint*

[2] Lezaud, P. (1998). Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*