Report on the resubmission

Bernstein's inequality for general Markov chains

by Bai Jiang, Qiang Sun and Jianqing Fan

The submitted paper deals with concentration inequalities of Bernstein-type for functions of general Markov chains. This is a resubmission, and the revised version provides a more concise presentation that now also addresses, as new additions, aspects such as the non-stationary case or the necessity of the boundedness assumption imposed on the functions. Much of my criticism and suggestions have been taken into account in the revision.

However, I am still very sceptical about the following points:

- The concentration inequalities stated in Theorems 1 and 2 require explicit knowledge of the absolute and right spectral gaps, respectively. The newly added remark paragraph in Section 2 provides a satisfactory overview of the relevant literature. However, according to this presentation, for general-state-space Markov chains–except for the special case cited–the corresponding quantities are neither calculable nor computable. Doesn't this mean that the new inequalities are thus only applicable and relevant for the case of Markov chains with finite state space (which has already been treated in detail in the literature)? Insofar as there is indeed added value by considering the general framework, this needs to be clarified by providing concrete examples of Markov chains (with explicitly given absolute/right spectral gaps), which can now be dealt with, but for which no comparable concentration results have yet been available in the literature. If this is not feasible, the paper promises more than it actually delivers, and the presentation should be made more transparent.
- It is helpful to find the newly added application (Section 5) to robust mean estimation under Markov dependence that illustrates the advantage of the new Bernstein inequality over the Hoeffding inequality for general Markov chains previously published by the authors. The presentation, however, seems rather uninspired. Right at the beginning of the section, a general, very brief motivation for studying *robust* mean estimation should be included. The main result (Theorem 5) should also be revised: In addition to the new Bernstein inequality, its proof substantially relies on *Assumption 1* imposed on the loss function. At this point, it is very unsatisfactory that with regard to the question of how to verify Assumption 1, there is only a reference to the i.i.d. case. Moreover, the reference is so imprecise that it actually took me a little longer to find the corresponding result in the cited article.

  In its current formulation, Theorem 5 thus offers somewhat limited insight: given abstract assumptions on the underlying Markov chain (see the point above) and the underlying loss function, an upper bound on the absolute error of the robust mean estimator holds with high probability, where the upper bound in turn contains the somewhat dubious absolute spectral gap constant. It is clear that the application is not the essence of the paper, but the value of the probabilistic results strongly hinges on their actual usability. For a paper submitted to EJP, a much more elaborated and polished version of Theorem 5 is therefore to be expected. In particular, to justify a publication in EJP, it is not sufficient to resort to the fact that, under the given abstract assumptions, the Bernstein inequality (compared to the Hoeffding inequality) provides sharper estimates. Instead, a fully worked-out result of independent interest should be stated.

At the end of the discussion in Section 6, it is noted that the example of robust mean estimation only serves as a simple showcase and that the authors expect to see more and more use cases in the future. Instead of offering such promises, it would be appropriate to be more concrete already now.

Minor remarks

p.11, l.1 Please rewrite the first sentence in the statement of Lemma 6. There are also typos in the remainder of the text (e.g. in the first sentence of Lemma 8) that need to be corrected.

p.17/18 You first state the relationship $y_i = \mu^* + \varepsilon_i$ before explaining that $\mu \in \mathbb{R}$ is the underlying mean; so something doesn't fit.

p.18 Please modify the formulation of Theorem 5 according to the formulation of the other theorems: in Theorems 1 and 2, e.g., the basic quantities are explicitly introduced or explained. In contrast, to fully understand Theorem 5, the reader has to gather the assumptions from the beginning of Section 5 and the definition of the quantities $\alpha_1(\lambda), \alpha_2(\lambda)$.