Reviews for *Concentration of Maxima and Fundamental Limits in High-Dimensional Testing and Inference*

December 2020

Dear Authors,

We contacted 7 people for your manuscript. Three of them came back to us with a detailed review. Two others commented very briefly. We will refer to them as R1, R2, R3, R4 and R5 below.

R1

I think the topic of this manuscript is excellent and timely. There are few self-contained books on high-dimensional statistics with this sort of twist.

This manuscript seems like an ideal contribution to SpringerBriefs. It is a manageable size and gives a solid, but brief overview of current methods such as FDR etc and puts them into the context of "the support recovery problem". I am not so enamored with hypothesis test for high-dimensional problem, since I think the key goal is to identify and recover sparse signals. Bringing to bear extreme value theory methods into the problem is a natural idea and this monograph goes much further in this direction than other research efforts. On the other hand, this material is not so accessible to the practitioner, but would require some strong training in probability. As such, the target audience as described in the preface, "doctoral students in the Statistics with background in measure-theoretic probability, . . . " is spot on, but the manuscript is perhaps not so well suited to "researchers in applied fields, ..." While the material is important to the latter group, it would require some hand-holding by an expert in order to implement and fully understand the methodology.

The manuscript contains sufficiently new topics especially the application in Chapter 4 to dependent data. This is an important topic and has not received as much attention in the high-dimensional literature as it should.

The manuscript is generally well-written and the issues/problems are clearly laid out. On the other hand, the writing may be a bit overly concise for which a supplement would be needed in a topics course. But the authors seem to be aware of this aspect.

I like the fact that this is not a 500 page manuscript which would be forbidding. This is a book that one could work through in a topics/reading course in statistics and would provide the necessary overview for students to pursue a research agenda in this area.

So the bottom line is that I have quite a favorable impression of the book. I hope these comments are helpful. Let me know if you have any further questions.

R2

The book project grew out of the thesis of Z. Gao and some related articles. The main idea is about concentration of maxima in the sense of a law of large numbers (which is a rather rough result in this context; it could be supplemented by finer results on large deviations of maxima, as used in the context of a.s. convergence of maxima; e.g. the cited results of Barndorff; Resnick, Tomkins. BTW the best results on a.s. stability of maxima are due to Klass) and to exploit this fact for various applications.

The manuscript is well written in general and the English is basically ok. There are some problems with articles (a, the, lacking both) and singular/plural. The style of writing changes throughout the manuscript. The first chapter uses many words and long sentences which sometimes appear to be baroque and might be difficult to be accessed by younger readers. Towards the end the sentences are getting shorter and seem to use only the most necessary vocabulary for explaining mathematical details. There are plenty of typos and problems with citations (names appear twice at various places). In my opinion, the authors exaggerate their explanations in the first chapters. There are repetitions of text and references.

It is not clear whether the authors aim at a textbook or a monograph: some chapters contain exercises, others do not. For briefs the text is already too long in the present state. The authors could easily extend the manuscript to more than 200 pages, for example by extending the exercises and reorganizing the text.

For my taste, the authors take too long to come to the point. An example is Chapter 3. The authors cite dozens of sources before they come to their own results. I think this is the wrong approach. The main line of the manuscript should be to present own results. The results of others and their relation to those in the book can be discussed anywhere else, e.g. in comments after the results, or at the end of a section or chapter. It honors the authors that they want to quote correctly and give merit to previous work, but this approach disturbs the flow.

Personally, I do not favor the separation of proofs from the results, as done in Chapter 3, but this is a question of taste. Moreover, there are too many auxiliary results which do not belong there (Section 3.3.1 and 3.3.2). They should be moved to an appendix.

Chapter 4-6 are relevant because they illustrate the idea of concentration of maxima for various advanced statistical topics. But the writing has several disadvantages. I could repeat various aforementioned points. For example, Section 6.6.1 does not belong there and should be given in an appendix. The property of rapid variation can be introduced in an appendix in a general context, and normal and chi square distributions and their properties can be treated as examples, which then appear in different chapters. The notion of stochastic monotonicity should be explained in an appendix too.

As the manuscript stands now, it reads like a collection of distinct papers. The authors should think how to organize the results in a more suitable way such that the chapters get linked in a better way.

R3

It seems like an interesting manuscript summarizing some research I know fairly well, but I will not have time to review this (short) book, sorry.

R4

This is an excellent paper that presents a substantial collection of old and new results on global testing and support recovery problems with independent and dependent test statistics in large-scale problems. I enjoy reading it a lot for two main reasons. First, the paper has a brilliant literature review. The line of work on the phase transition in large-scale inference is rich but challenging to follow for those who are not working in this area, since the results are derived in different asymptotic regimes under different assumptions on the marginal null distributions, distribution of signal strengths, dependence structure with different criteria. The paper starts by defining five criteria explicitly – detection error, exact support recovery, approximate support recovery, exact-approximate and approximate-exact support recovery, and discussing their relationships. I find this very helpful in understanding the difference between results in the

literature. Second, the new results are quite interesting. For example, the exact support recovery under arbitrary dependence or uniform relative stability is intriguing because the dependent case is more realistic in practice but relatively understudied compared to the independent case. For multivariate gaussian statistics, the equivalence between uniform relative stability and uniformly decreasing dependence is remarkable. I particularly like the smart proof that applies the Ramsey's coloring theorem to prove the existence of a growing block of positively correlated statistics when the uniformly decreasing dependence is not satisfied.

I recommend the paper for a Briefs publication because I believe it is of great interest to a broad audience and should be easy to expand into a longer book. I have some minor comments which may be helpful to further improve the manuscript. The authors do not have to address all of them.

1. Most existing works assume the constant signal strength, which essentially assume that the location parameters are drawn from a two-point distribution. It is great that (3.3) relaxes this assumption by allowing the support to be the union of $\{0\}$ and an interval with a growing gap from 0. However, it is a bit far from realistic from my perspective because the practical case often involves a mixture of strong and weak signals. A recent paper by Li 2020 considers the generic distribution of alternative means under appropriate rescaling, and shows the surprising result that the higher criticism test no longer dominates Bonferroni's method even in the sparse regime $\beta > 3/4$. It would be more instructive if the paper can start from the more general model to provide a fuller picture.

2. The statistical risks for exact-approximate and approximate-exact support recovery look artificial and uninterpretable. Could you provide further explanations of these two criteria?

3. In practice, type-I error/FWER/FDR and type-II error/FNER/FNR are typically not equally weighted. The former is often worse than the latter. Is there any work that considers the asymmetric version like FDR plus a fraction of FNR? Does the weighted risk give a fundamentally different phase transition in some regimes?

4. The uniform relative stability (URS) is type of weak dependence based on the results in Chapter 5. In FDR literature, the statistics are said to be weakly dependent if

$$\frac{1}{\mathrm{p}^2}\sum_{\mathrm{i}=1}^{\mathrm{n}}|\mathrm{Cor}(\mathrm{x_i},\mathrm{x_j})| = \mathrm{o}(1);$$

see e.g. Efron 2007 and Fan 2012. Is there any connection between these two types of weak dependence?

5. This is another line of work that studies the strong dependence structure where the covariance matrix has a factor structure (eg. Fan 2012). For instance, this includes the equi-correlated case. Are you aware of any results under strong dependence in your context?

6. The results in Chapter 4 require different dependence structure but some of them are not clearly stated. For instance, the results in Section 4.2.2 and 4.2.3 rely on the independence assumption that is stated in the last sentence of the first paragraph in page 59. It would be more clear if this assumption can be included in the Propositions and Theorems.

7. Proposition 4.3 still requires f_a to be MLR, right? If so, please state it.

8. Fig 3.2 shows that the exact-approximate support recovery is strictly easier to achieve than the approximate-exact support recovery. Do you have intuition on it?

9. The phrase "exact support recovery" is used for both the general problem (as in the title of Chapter 2), and for the specific criterion risk^{E}. This is a bit confusing.

Typos and other minor comments

1. Page xii, "add more" needs to be removed.

2. Page 3, "$= (X'X)^{-1}X'\xi$" requires $n > p$.

3. Page 11 and 34, "Candés" should be "Candès".

4. Page 20, "TO DO: Verify and add" should be removed.

5. Page 21, "Give a good hint" should be replaced by the hint

6. Page 21, "7." should be removed.

7. Page 24, the statistic "S" is not defined.

8. Page 25, what does the dark-gray area that is below the area marked by "M" and on the right of the area marked by "S, L_1, L_2" correspond to?

9. Chapter 3, the notation $f(\beta), g(\beta), h(\beta), \tilde{g}(\beta), \tilde{h}(\beta)$ is hard to memorize. What about $f_P(\beta)$ where $P \in \{D, A, E, AE, EA\}$?

10. Page 46, "these Hamming loss-minimax studies naturally amounts to the analysis of the elementary case of iid data,..." Is this true? Although the indicators can be dealt with separately but the threshold may depend on the entire dataset as in Holm, Hochberg, and BH procedures.

11. Page 50, "Intuitively, this is because the maxima of the errors grow at their fastest in the case of independence." Is it true that the maxima of negatively equi-correlated statistics grow slower than that of independent statistics?

12. Page 52, scenario 1, $r^{(2)} = (1 + \delta) \Longrightarrow r^{(2)} = (1 + \delta)g(\beta)$.

13. Page 69, Corollary 4.5, $P[Z_i \leq t] = \exp\{-x^{-\alpha}\}$ should be $P[Z_i \leq t] = \exp\{-t^{-\alpha}\}$.

14. Page 87, latex errors below (5.41).

15. Page 98, "BHprocedure" should be "BH procedure".

16. Page 99, "...with v = 1,, corresponds to..." should be "...with v = 1 correspond to..."

17. Page 102 and 105, MacArthur et al. (2016) should be in parentheses

References

Bradley Efron. Correlation and large-scale simultaneous signicance testing. Journal of the American Statistical Association, 102(477):93103, 2007.

Jianqing Fan, Xu Han, andWeijie Gu. Estimating false discovery proportion under arbitrary covariance dependence. Journal of the American Statistical Association, 107(499):10191035, 2012.

Xiao Li and William Fithian. Optimality of the max test for detecting sparse signals with gaussian or heavier tail. arXiv preprint arXiv:2006.12489, 2020.

R5

The book definitely looks interesting and extremely timely. These concentration tools are fundamental for high-dimensional statistics and we still lack good articles that survey them. This book project fills in this crucial gap.