

Response to the reviews of the manuscript: “Concentration of
Maxima and Fundamental Limits in High-Dimensional Testing and
Inference” by Zheng Gao and Stilian Stoev submitted to
SpringerBriefs Series in Probability and Statistics

April 19, 2021

1 Summary of the revision

We sincerely thank the managing Editor and the five anonymous referees for their very helpful and constructive suggestions. We really appreciate the reviewers’ time, effort, and thoughtful suggestions, which stimulated us to substantially revise and improve the manuscript. We begin with a high-level description of our edits and in the following section we respond to some of the more major specific points raised by the reviewers. We do not list a large number of smaller edits made to the manuscript.

- **(Organization)** The manuscript has been revised substantially and now includes a brief introductory chapter and 6 substantive chapters, which amount to a total of 125 pages. As with the original submission, Chapter 2 reviews the statistical risks, estimation and testing procedures, as well as probability models for the errors used in the rest of the manuscript.

After very careful deliberation, we decided to keep a majority of the content in Chapter 3, which provides a panorama of the phase-transition results in the signal detection and support estimation problems for the additive iid Gaussian error model. There were some mixed opinions from the reviewers regarding the inclusion of a chapter that paints the landscape in broad strokes and provides references to the state-of-the-art of the literature, at the expense of delaying entirely original contributions. We think that providing an accessible account of the subject of high-dimensional multiple testing would benefit readers who are not already experts in the area, and might just narrowly outweigh the additional patience required on the part of experts before we get to the more exciting material. Details and explanations of the changes we made can be found in the response to Reviewer 2 (R2) below.

To better balance the exposition, we reorganized Chapter 4 of the previous version of the draft into two new chapters (now Ch. 4, entitled “Exact Support Recovery Under Dependence”, and Ch. 5 “Bayes and Minimax Optimality”). This change, in addition to making all chapters more manageable and of about equal lengths, exposes more clearly our contributions. The results contained in Chapters 4 and 5 as well as the more probabilistic Chapter 6 and the more applied Chapter 7 are original contributions to the literature.

Following the suggestion of R2, we have placed most proofs in the text after the corresponding claims (theorems, propositions, lemmas). In the interest of space and to conform with the format of the SpringerBriefs series, we no longer include “Exercises” sections.

- **(Typos, grammar, and style)** Prompted by the suggestions of R2, we have revised the language of the manuscript. There have been extensive re-writes throughout the manuscript in order to make the content more connected and style more consistent. In trying to make the text more accessible, we re-formatted the statements of most theorems to make them more modular. We have also corrected typos and missing articles.
- **(An improvement)** We implemented a small improvement, which broadens the scope of our exact support recovery results. Specifically, we noticed that the size of the signal support $|S_p|$ (previously $|S_p| = \lfloor p^{1-\beta} \rfloor$) can be parameterized as

$$|S_p| \sim \ell(p)p^{1-\beta}, \quad \text{as } p \rightarrow \infty,$$

where $0 < \beta \leq 1$ and where $\ell(\cdot)$ is a slowly varying function. This more general parameterization required only minor, cosmetic changes to some proofs in Chapters 4 and 5.

- **(Length and online supplement)** We have shortened the manuscript to just under 125 pages. The Appendix appearing after the References is meant to become an online supplement when the manuscript becomes ready for publication. The appendix contains some proofs of phase-transition results stated in Ch. 7, which are similar to ones appearing in Ch. 3. Thus, even without the online supplement, not much is lost by not including the Appendix in the printed version.

2 Responses to specific points raised by the reviewers

2.1 Reviewer 1 (R1)

We are grateful for the positive comments and important feedbacks from the reviewer.

- *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.*

We thank the reviewer for the favorable and encouraging assessment.

- *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.

Thank you. Stimulated by your feedback, we have attempted to provide more details and background where necessary. In the interest of space and to conform to the format of the SpringerBriefs series, we had to relegate some of the proofs in Chapter 7 (new labeling) to an online supplement. This still allowed us to treat carefully the important application

to GWAS. We believe we have also improved the treatment of the exact support recovery topic under dependence. As indicated above, we divided the phase-transitions results (under dependence) and the Bayes and minimax optimality results into two separate chapters (Ch. 4 and 5, now). We have added some more insights to the statements in Ch. 3 and Ch. 7, as well as some more background material in Ch. 2. These changes, in addition to making the more advanced material accessible, should improve the balance between the chapters and make them read better independently, thus making the text better suited for a special topics course.

2.2 Reviewer 2 (R2)

We really appreciate the reviewer's feedback and suggestions. We have taken all of them into account while revising the manuscript.

- *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.*

Thank you for the references and insights. Indeed, we use a rather basic relative stability property. Mathematically, the study of finer large deviations of maxima is of great interest. Our main focus, however, is on the situation where the variables could be dependent and possibly very strongly dependent. To the best of our knowledge, less is known about almost sure relative stability under dependence. In the statistical context of exact support recovery, it was more important for us to be able to deal with nearly arbitrary dependence rather than to have finer results on concentration of maxima. This is because we establish only asymptotic phase-transition-type results. It would be of great interest to obtain more precise finite-sample and/or rate of concentration results. This would require more advanced mathematics beyond the scope of this manuscript.

- *The manuscript is well written in general and the English is basically ok. There are some problems with articles... 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.

We have thoroughly revised the exposition and fixed hopefully all typos and grammatical errors. We have also streamlined the exposition and tried to ground it.

As indicated above, we have shortened the manuscript, eliminated the exercises, and cast it into the format of a SpringerBriefs monograph. We appreciate the reviewer's suggestions on extending the manuscript, but we felt that the work is not mature enough to be expanded into a longer textbook. In trying to balance the chapter lengths and content, we believe now

the manuscript is more appropriate for an auxiliary text in a special topics graduate course on high-dimensional statistics.

- *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.*

We really appreciate the feedback, which prompted us to thoroughly revise the presentation.

The main goal of Chapters 2 and 3 is to shed light on the relationships between typical high-dimensional problems and their respective statistical limits. In particular, Ch. 3 reviews the existing literature, and provide a unified treatment of the phase transitions under different loss functions.

The inclusion of material surrounding Theorem 3.1 seems to be of concern. Indeed, Theorem 3.1 summarizes some known results in the signal detection problem, which may be familiar to experts in the field. On the other hand, Theorem 3.1 also presents some new, previously overlooked results on the sub-optimality of certain popular statistics. These results are not the most exciting advances in either methodology or theory (and unfortunately will likely not be found in a journal publication because of their perceived lack of novelty). Nevertheless, they are natural and practically important questions that needed answers. In addition to these incremental contributions, we think that there is some value in collecting a range of related problems and results in one place. This inclusion of a bird’s-eye view, however cursory, would make the text more suitable for an introduction to a wider audience. To emphasize this role, we changed the title of this chapter to “A panorama of phase transitions”.

We agree that the core of our contribution is on the exact support recovery under dependence. To do justice to these results, as explained above, we have divided the old Chapter 4 into two new chapters. The first part (new Ch. 4) highlights the role of relative stability in exact support recovery under dependence. The new Ch. 5 more clearly presents another one of our contributions, namely the finite-sample Bayes optimality in exact support recovery. These important results were a bit obscured in the previous version of the manuscript.

We hope that the revision has resulted in a more accessible manuscript, which at the same time more clearly highlights new ideas and results in the context of existing literature.

- *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... 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...*

Chapter 6 (now Ch. 7) has been thoroughly revised. The proofs of phase-transition results similar to the ones in Ch. 3 have been omitted. They will be presented in an online supplement (appearing as the Appendix section after the References). This allowed us to emphasize more clearly the novel ideas in this chapter on the parametrization of association tests and the study of their power.

- *Personally, I do not favor the separation of proofs from the results ... 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.*

We agree. All proofs have been incorporated in the text and the manuscript has been thoroughly revised. As described in a previous point, we have made a number of changes to the organization of the manuscript, where now the chapters are better connected and balanced.

2.3 Reviewer 4 (R4)

We are very grateful to the reviewer for their exceptionally detailed and helpful feedback! We appreciate the reviewer's encouraging assessment and important suggestions on improving the manuscript, which have all been taken into account in the present revision. We respond below to only some of the major points raised but we would be happy to provide a more detailed account.

1. *Most existing works assume the constant signal strength, which essentially assume that the location parameters are drawn from a two-point distribution... 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.*

Thank you. We were not aware of the interesting and very recent work [4] of Li and Fithian when the first draft was prepared. We wholeheartedly agree that the set-up described by the reviewer is much more realistic. It is stimulating for us to learn that the Higher Criticism statistic no longer dominates the max-type Bonferonni thresholding procedure when $1/2 < \beta < 3/4$, as long as the signals are spread out, even if the noise terms remain Gaussian. We now discuss and reference this recent and fascinating development in Chapters 2 and 3 where appropriate.

One naturally wonders what happens, e.g., when both signals and noise terms are distributed with regularly-varying tails, and what can be said about the support recovery problems when signals are spread-out. Since our focus in this manuscript is weighted towards support estimation problems rather than the detection problem, we have refrained from discussing more in the direction of the more realistic setting of [4], which would require resolving some new mathematical challenges. We hope to understand this phenomenon better in the near future.

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?*

We have edited the following explanation when introducing the asymmetric statistical risks:

... In applications, however, attitudes towards false discoveries and missed signals are often asymmetric. In the example of GWAS, where the number of candidate locations p could be in the millions, and a class imbalance between the number of nulls and signals exists, researchers are typically interested in the marginal (location-wise) power of discovery, while exercising stringent (family-wise) false

discovery control. These types of asymmetric considerations, while important in applications, have not been studied theoretically. For example, the GWAS application motivates the *exact-approximate* support recovery risk, which weighs both the family-wise error rate and the marginal power of discovery: ...

In fact, we believe that power calculations for GWAS are most naturally interpreted when viewed through the exact-approximate support recovery risk.

On the other hand, we are not aware of an equally strong motivating application for the approximate-exact support recovery risk. It is harder to come up with a scenario where we want to weed out a few nulls from a large number of signals, with strong type II control and evaluate marginal type I errors.

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?*

There have been discussions on unequally weighted type I and type II errors. For example, the version of asymmetric risk suggested by the reviewer (i.e., $FDR + \lambda FNR$) appeared in Genovese and Wasserman [?] Section 6, although the asymptotic limits were not discussed there.

The asymptotic limits would not have been different for the weighted risk, so long as λ is bounded away from zero and infinity. This is because $FDR + \lambda FNR$ vanishes if and only if both FDR and FNR vanish; conversely, non-vanishing FDR and FNR is equivalent to non-vanishing weighted risk. Therefore, the only (theoretically) interesting weighting of type I and type II errors is combining family-wise error metrics with marginal errors, as in the exact-approximate support recovery risk.

We have added this comment on risk weighting to the manuscript.

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*

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

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

It can be shown that the more restrictive UDD' condition (Definition 6.2 in the new numbering) implies weak dependence as described in Fan et al [2] (the single sum in the reviewer's comment seems to be a typo),

$$\lim_{p \rightarrow \infty} \frac{1}{p^2} \sum_{i,j} |\Sigma_p(i, j)| = 0.$$

Starting with the assumed permutations l_p of $\{1, \dots, p\}$ in UDD', we have the equality

$$\sum_{i,j} |\Sigma_p(i, j)| = \sum_{i,j} |\Sigma_p(i', j')|,$$

where $i' = l_p(i)$, $j' = l_p(j)$, since permutations do not affect the sum. By the second assumption of a sequence $r_n \rightarrow 0$ that bounds the correlations from above, we have

$$\frac{1}{p^2} \sum_{i,j} |\Sigma_p(i', j')| \leq \frac{1}{p^2} \sum_{i,j} r_{|i-j|} \leq \frac{1}{p} \sum_{k=0}^{p-1} r_k \rightarrow 0.$$

Therefore UDD' implies the weak dependence condition.

We can also establish that the UDD/URS condition is no-weaker than the weak dependence condition. We use the same non-UDD example outlined in Example 6.1 (new numbering). Suppose $\mathcal{E} = (\epsilon_p(i))_{i=1}^p$ is Gaussian, and is comprised of $\lfloor p^{1-\beta} \rfloor$ blocks, each of size at least $\lfloor p^\beta \rfloor$. Let the elements within each block have correlation 1, and let the elements from different blocks be independent. It follows that

$$\frac{1}{p^2} \sum_{i,j} |\Sigma_p(i, j)| \sim \frac{1}{p^2} (p^{1-\beta})^2 p^\beta = p^{-\beta} \rightarrow 0,$$

and therefore \mathcal{E} is weakly dependent as defined in [2], but not UDD/URS.

However, it is not clear to us whether the weak dependence condition is always weaker than UDD. The comparison is complicated by the fact that the definition of UDD only cares about positive correlations, while the weak dependence condition takes the sum of absolute values of the correlations.

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?*

We thank the reviewer for pointing to the interesting methodological work of Fan, Xu and Gu [2], which aimed to estimate false discovery proportions for heavily dependent observations. We are not aware of works that discuss optimality or feasibility of support recovery (i.e., multiple testing) problems under such dependence conditions. The results in our text in their current form do not easily extend to that of Fan et al., since even the remainder term K_i 's in the factor model can be more heavily dependent than under the UDD/URS condition, as discussed in the response to the previous question. Nevertheless, Fan et al provides a glimpse into how heavy dependence (e.g., the equi-correlated case) may be handled, and how theoretical analysis may be carried out. We do hope to develop a theory that applies to the more general setting.

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.*

Thank you. Indeed, to clarify and better present the two very distinct themes – exact support recovery under dependence, on one hand, and finite-sample Bayes optimality (under independence), we have divided the previous Chapter 4 in two – the new Chapters 4 and 5. The new Chapter 5 entitled “Bayes and Minimax Optimality” now presents the results in a much more organized fashion where the independence assumptions are clearly stated.

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

Thank you for the sharp observation. The previous presentation was not sufficiently clear. We have now devoted the new Ch. 5 to the finite-sample Bayes optimality (and minimax optimality) to better present these topics. The former Proposition 4.3 appears now as Theorem 5.2. The MLR property on f_a is no longer necessary. This is because the *likelihood ratio thresholding* estimator is defined in such a way (in terms of the likelihood) that the desired monotonicity property is guaranteed. The MLR property was essential in Theorem 5.1 since there we deal with thresholding (as opposed to likelihood thresholding). The MLR property basically ensures that in the setting of Theorem 5.1, the simple thresholding estimators are in fact of *likelihood thresholding* type. We realize that Theorem 5.2 is a bit too theoretical (i.e., not directly useful in practice since we don't have the likelihood) but it provides some insights on the general Bayes (sub)optimal support estimation procedures.

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?*

To provide some intuition, consider the distances between the decision boundaries and the means of the null and signal part of the observations, respectively. Controlling the FWER requires setting the decision boundaries approximately

$$\sqrt{2 \log p} \tag{1}$$

away from zero (mean of nulls), as is well-known in the Bonferroni adjustment.

Controlling FDR, on the other hand, depends on the number of signals present. When there are fewer signals, a FDR-controlling procedure has to be more conservative – when the number of true signals is small, the number of false discoveries reported along with them has to reduce as well. Consequently, the decision boundary has to be further away from zero as $\beta \rightarrow 1$. Quantitatively, this distance varies with β approximately as

$$\sqrt{2\beta \log p}, \tag{2}$$

with sparser problems requiring a larger separation. Notice that the distance required for FDR control in an extremely sparse problem ($\beta \rightarrow 1$) is no different from that of FWER control. For very dense problems ($\beta \rightarrow 0$), the distance vanishes on the Bonferroni scale.

Similar to FWER, controlling the family-wise non-discovery rate (FWNR) requires this boundary to be approximately

$$\sqrt{2 \log (p^{1-\beta})} = \sqrt{2(1-\beta) \log p} \tag{3}$$

away from the mean of the signals.

Finally, controlling false non-discovery ratio (FNR) is similar to FDR control, except the class imbalance reverses – the null part greatly out numbers the signals. Therefore the separation between the decision boundary and the mean of signals can afford to be small. So small that it vanishes on the Bonferroni scale:

$$o\left(\sqrt{2 \log p}\right). \tag{4}$$

Informally, mixing and matching the two distances from nulls (1) and (2) with the two distances from signals (3) and (4) gives rise to the four boundaries we derive in this text. The fact that exact-approximate support recovery is strictly easier than approximate-exact support recovery is a consequence of $(1) + (4) \leq (2) + (3)$. But it is ultimately a consequence of the assumption on signal sparsity. That is, when there are more nulls than signals, the separation between the signals and the nulls has to be larger for simultaneous FDR and FWNR control.

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.

We have clarified the terminology throughout. In particular, Chapter 2 is now entitled “Risks, Procedures, and Error Models”.

We thank the reviewer for pointing a number of typos, which we have corrected. We respond below to a subset of the minor points and comments raised in the report.

Typos and other minor comments

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?

The area is detectable by the HC statistic. We have made changes to the figure and clarified this in the caption.

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 $P \in \{D, A, E, AE, EA\}$?

Agreed! We have made changes everywhere in the text.

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.

Thank you for the feedback. We have clarified and tempered the language:

More importantly, since Hamming loss decomposes into expectations on individual terms that are not affected by dependence, Hamming loss-minimax studies do not reveal the difference in probability of support recovery between independent and dependent observations.

Our analysis, and e.g. that of Butucea et al [?], consider all thresholding procedures which include data-dependent procedures that are affected by dependence.

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?

Thank you for the insightful comment. Indeed, by Slepian’s Lemma, negatively correlated Gaussian variables have larger maxima (in stochastic order) than iid maxima (with the same marginals). This difference, however, is negligible asymptotically. Intuitively, this is due to the fact that while Gaussian vectors can be equi-correlated with correlation ρ that can be

arbitrarily close to 1, this common correlation is bounded below by $-1/(p-1)$, where p is the number of the variables.

Our choice of words was imprecise. We should have said that the rate of growth of dependent maxima is no faster than in the iid case. We now clarify this point in the new Remark 2.1 (see Relation (2.41)).

This general observation is entirely due to the light-tailed (rapidly varying) nature of the marginal distributions. It applies to error-arrays with arbitrary dependence and the proof involves the union bound and the rapid variation property of the marginal distributions. Please see Proposition 2.2.

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

Thank you for your careful reading of the text! However this was not a typo – we wanted to provide an example where signals sizes straddle the boundary but exact support recovery is still achieved. There is only one signal of size $r^{(2)}$ which corresponds to the required signal size at its sparsity $\beta = 1$, i.e., $r^{(2)} = (1 + \delta)g(1)$.

2.4 Reviewer 5 (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.

We thank the reviewer for the encouraging assessment! In our revision we have strived to maintain the balance between reviewing the existing literature and presenting our own work.

References

- [1] Simeon M Berman. Limit theorems for the maximum term in stationary sequences. *The Annals of Mathematical Statistics*, pages 502–516, 1964.
- [2] Jianqing Fan, Xu Han, and Weijie Gu. Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.*, 107(499):1019–1035, 2012.
- [3] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983.
- [4] Xiao Li and William Fithian. Optimality of the max test for detecting sparse signals with Gaussian or heavier tail, 2020. <https://arxiv.org/abs/2006.12489>.
- [5] Vlasos Pipiras and Murad S. Taqqu. *Long-range dependence and self-similarity*. Cambridge Series in Statistical and Probabilistic Mathematics, [45]. Cambridge University Press, Cambridge, 2017.