

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

March 18, 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.

- **(Brief description)** The revised manuscript has now a brief introductory chapter and 6 substantive chapters, which amount to a total of 124 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. It contains also a high-level summary of our results in the context of the existing literature and some auxiliary facts from probability and extreme value theory. For convenience of the reader, we have added statements of Slepian’s Lemma and the Sudakov-Fernique inequality (without proofs).

Chapter 3 provides a panorama of the phase-transition results in the signal detection and support estimation problems for the additive iid Gaussian error model. Its purpose is to emphasize ideas and provide an accessible account, including references to the state-of-the-art in the literature. The chapter contains also some new results on the (sub)optimality of some popular statistics, which appear to have been overlooked in the literature (Theorem 3.1). One main idea behind the exposition is to systematically present the different types of phase transition results as a function of the chosen statistical risk.

Chapters 4 through 7 contain the core of our original contributions.

Chapter 4 establishes phase-transitions results for the exact support recovery problem in the additive error model. The main themes of this chapter are *error-dependence* and *concentration of maxima*. We establish that the notion of uniform relative stability – a type of concentration of maxima phenomenon – is key to characterizing the phase-transition in exact support recovery for the general class of thresholding estimators.

Chapter 5 begins with a study of the finite-sample (Bayes) optimality and sub-optimality of the threshold-based support estimators. Criteria for the latter are established and certain theoretical results indicate that *likelihood thresholding* (as opposed to data thresholding) is in general optimal, when the thresholding estimators are sub-optimal. In the regime when the thresholding estimators are optimal, i.e., for log-concave error densities, our results entail a universal statistical limit in the exact support recovery problem, valid for all estimators (Theorem 5.3). On the other hand, for the general class of thresholding estimators, we obtain minimax characterizations of the phase-transition in exact support recovery, for every fixed error-dependence structure that satisfies mild uniform relative stability condition (Corollary 5.3).

Chapter 6 is nearly unchanged. It is more probabilistic in nature and it provides a complete characterization of the uniform relatively stable (URS) Gaussian triangular arrays. This URS condition is the key property, in the lower bound on exact support recovery under dependence in Chapter 4. These probabilistic results as well as their methods of proof may of independent interest and are accessible to a wider audience of graduate students in statistics.

Chapter 7 focuses on an application to genome-wide association studies (GWAS) in statistical genetics. It is shown that all phase-transition results in the additive error models of Chapter 3, have their close counterparts for the chi-square models arising naturally in GWAS. To connect these asymptotic results to the practical notions in statistical genetics, we establish the connection between the odds-ratio (effect size) in multinomial 2×2 models and the signal-size parameter of the corresponding chi-square association test. This allows us to quantify the statistical power and optimal design questions as well as ultimately explain the role of phase-transitions in the fundamental statistical limits of GWAS.

- **(Organization)** Following a suggestion of a reviewer (**R2**), we have included most proofs in the text right 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 “Exercise” sections.

Last but not least, to better balance the exposition, we decided to divide the previous Chapter 4 into two new chapters (now Ch. 4, entitled “Exact Support Recovery Under Dependence”, and Ch. 5 “Bayes and Minimax Optimality”). In addition to making all chapters more manageable and of about equal length, it exposes more clearly our contributions. The results contained in Chapters 4 and 5 as well as the more probabilistic Chapter 6 and more applied Chapter 7 are original and new contributions to the literature.

- **(Typos, grammar, and style)** Prompted by criticism of a referee (**R2**), we have revised the language of the manuscript. We have corrected a large number of typos, missing articles as well as attempted to make the language more consistent. For example, we re-formatted the statements of a large number theorems so that they are more clear and modular.
- **(Length and online supplement)** We have softened the manuscript to just under 125 pages. The Appendix section appearing after the References is meant to become an online supplement should the manuscript be considered for publication. This Appendix section 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)

- *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. In addition to making the more advanced material better accessible, this case resulted in a balanced chapters, which can be read independently and are 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 most (and 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, but we felt that the work is not mature enough to extend to a textbook. At the same time, by balancing 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 made us thoroughly revise the presentation!

Some results of Chapter 3 are in fact new. The main philosophy behind Chapters 2 and 3 is to emphasize the (perhaps overlooked) impact the choice the statistical loss has on the statistical limits in support recovery. Ch. 3 serves the purpose of reviewing the existing literature, while at the same time providing a unified treatment of phase transitions, suitable for a wider audience. To emphasize this role, we have now changed the title of this chapter to "A panorama of phase transitions". Theorem 3.1 both summarizes known as well as presents some new *overlooked* results on the sub-optimality of certain popular statistics. The other results in Chapter 3 are also closely connected to the phase-transition results in Chapter 7 on GWAS, which are new. This allowed us to omit the proofs of the results in Ch. 7 (presented in an online supplement) and focus on the discussion of statistical power and optimal design issues, which are very relevant in practice and also novel.

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 previous 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 important contribution on 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 shows our new ideas and results in the context of the 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 and 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 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 only to some of the major points raised but 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. We wholeheartedly agree that the set-up proposed by the reviewer is much more realistic. It is quite interesting and stimulating for us to learn that the Higher Criticism statistic no longer dominates the Bonferroni thresholding procedure. This changes our general understanding about optimality of the max-statistic. Although, we are not prepared to contribute more to the subject from the more general perspective of [4], we have included brief comments about this very recent development in Chapters 2 and 3. Since our focus in this manuscript is mostly on the support estimation problem (rather than on the detection problem), we decided not to pursue this new and important line of research, which may require resolving some new mathematical challenges. We hope to understand this phenomenon better in the near future.

Zheng, PLEASE CHECK how if I quoted the work of Li and Fithian in Ch 2 and 3, appropriately. Thanks

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?

Zheng, need help here?

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?

Help.

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?

Indeed, the notion of *weak dependence* in its various forms of mixing coefficients and dependence functionals has a long and fruitful history in the study of large-sample limit theorems. We appreciate this natural question! There is a fundamental difference between the notions of weak dependence required to establish the asymptotic behavior of sums and those for maxima. As a rule, the maxima of dependent variables behave very similarly to the maxima of independent variables even under *severe dependence*. This is in stark contrast with the large-sample behavior of dependent sums (averages), which have been much more extensively studied in the literature. We have now clarified and emphasized these differences in several places in the text.

We provide below some brief comments on the way we understand the differences between the asymptotic behaviors of sums and maxima under dependence. This is certainly a big and important subject and the following discussion is just on the level of ideas. The weak dependence condition stated by the referee appears to be tailored to the study of sums rather than maxima. For example, a stationary (finite-variance) time-series $\{X_k, k \in \mathbb{Z}\}$ is said to be short-range dependent if

$$\sum_{k=1}^{\infty} |\text{Cov}(X_k, X_0)| < \infty.$$

If the latter sum is infinite, the time series is referred to as long-range dependent. For example, consider the Gaussian case, where the time series has a power-law auto-covariance $\text{Cov}(X_k, X_0) \propto k^{-\gamma}$, $k \rightarrow \infty$, for $\gamma \in (0, 1)$. In this long-range dependent regime, it can be shown that, as $n \rightarrow \infty$,

$$\left\{ \frac{1}{n^H} \sum_{k=1}^{\lfloor nt \rfloor} (X_k - \mathbb{E}X_k) \right\}_{t \geq 0} \xrightarrow{f.d.d.} B_H = \{B_H(t), t \geq 0\},$$

where B_H is the fractional Brownian motion process – a Gaussian, stationary increment and self-similar process with exponent $H = 1 - \gamma/2 \in (1/2, 1)$. This is in stark contrast with the

usual Brownian motion limit which arises under the *weak dependence*. The behaviors of many sum-based statistics will change dramatically when the dependence switches from short- to long-range. For more details, see e.g. [5] and the references therein.

Shockingly, when considering even most heavily dependent (say Gaussian) stationary time series $\{X_k\}$, the limit behavior of the maxima is no different than in the iid case. Indeed, the Berman condition (cf [1, 3])

$$\text{Cov}(X_k, X_0) = o(\log k), \quad \text{as } k \rightarrow \infty,$$

guarantees in this setting that, as $n \rightarrow \infty$,

$$\frac{1}{a_n} \max_{k=1, \dots, n} X_k - b_n \xrightarrow{d} Z \sim \text{Gumbel}, \quad (1)$$

where a_n and b_n are *the same* scaling and shift sequences as in the case of iid data. That is, if X_k^* are iid with the same marginal distribution as the X_k 's, then

$$\frac{1}{a_n} \max_{k=1, \dots, n} X_k^* - b_n \xrightarrow{d} Z,$$

where the normalizing constants a_n and b_n as well as the limit Z are the same as in (1).

In particular, the Berman condition holds for Gaussian long-range dependent models! Thus, even when the traditional notions of weak dependence for sums no longer hold, the dependence is weak enough for the maxima to behave as in the iid setting. Please note that in our monograph we need an even weaker notion of dependence since we need only have relative stability, i.e.,

$$\frac{1}{a_n} \max_{k=1, \dots, n} X_k \xrightarrow{\mathbb{P}} 1, \quad (2)$$

as $n \rightarrow \infty$. As shown in the seminal work of Berman [1] and in a more general situation for triangular arrays in our Chapter 6), for the latter convergence (2) to hold one merely needs the auto-covariance to vanish, i.e.,

$$\text{Cov}(X_k, X_0) = o(1), \quad \text{as } k \rightarrow \infty.$$

We find this to be an important and perhaps somewhat overlooked universality phenomenon showing that maxima of Gaussians (and in fact, general light-tailed variables) continue to concentrate under drastically stronger dependence than we are accustomed to deal with in the case of sums/averages.

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 the important reference to the work of Fan, Xu and Gu [2], which we had overlooked. It can be seen that the *weak dependence* condition in their Definition 1, implies our UDD condition (for a correlation matrix). Therefore, the Gaussian error arrays of the K_i -variables in Relation (10) in the latter reference, are URS (uniformly relatively stable). This suggests that the PFA methodology in [2] can perhaps be extended to apply to

the exact support recovery problem. That is, one can project on the orthogonal complement of the space spanned by the top- k eigenvectors of the error-covariance. Then, perhaps under standard in-coherence conditions, between the signal and the covariance, the simple Bonferroni thresholding estimator is optimal for exact support recovery. We are not aware of many more examples, but hope that the developed general theory applies in many more setting. We now include a remark on this important reference.

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?*

Help, please! Is it related to the fact that $p \gg p^{1-\beta}$.

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

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.*

You are right that our statement was a bit exaggerated and thus perhaps misleading. We have clarified and tempered the language.

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 smaller 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.

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] Vladas Pipiras and Murad S. Taqqu. *Long-range dependence and self-similarity*. Cambridge Series in Statistical and Probabilistic Mathematics, [45]. Cambridge University Press, Cambridge, 2017.