Some Universal Insights on Divergences for Statistics, Machine Learning and Artificial Intelligence

REPLY TO REVIEW REPORTS

Michel Broniatowski, Wolfgang Stummer

Abstract We would like to thank the three referees for their very useful suggestions and comments. In this document, we give a detailed reply thereon. For the sake of brevity, we do not explicitly reply on suggestions on typos and minor reformulations (they are all incorporated).

1 Review Report 1 (RR1) (the one starting with "General Comments ...")

1. a. **RR1, General Comments:**
   I think the authors have given very interesting results on the topic of generalized distances and divergences between functions in a great variety of disciplines and contexts. The paper is well organized and the exposition seems ideal for a so long paper. ... The results obtained in the paper may provide a basis for interesting developments and discussions in almost all fields of science and engineering where the concept of the distance or dissimilarity or divergence is omnipresent and plays an important and universal role.
   b. **Our reply:**
   Many thanks for your very positive overall assessment.

2. a. **RR1, (cf. your terminology) Minor Comment 1:**
   P. 1, Abstract: The abbreviation AI should be fixed previously.
   b. **Our reply:**
   Done.

Michel Broniatowski
Sorbonne Universite Pierre et Marie Curie, LPSM, 4 place Jussieu, 75252 Paris, France,

Wolfgang Stummer (corresponding author)
Department of Mathematics, University of Erlangen-Nürnberg, Cauerstraße 11, 91058 Erlangen, Germany, e-mail: wolfgang.stummer@math.fau.de
as well as Affiliated Faculty Member of the School of Business and Economics, University of Erlangen-Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany.

3. a. **RR1, Minor Comment 2:**
   P. 2, Property (D2): Except of reflexivity, a suitable terminology for this property is, maybe, identity of indiscernibles, a terminology which I borrow from the paper by ...
   b. **Our reply:**
   We have added the terminology, and cited the suggested paper.

4. a. **RR1, Minor Comment 4:**
   P. 5, Last paragraph before Subsection 2.3: The following recent paper should be maybe included in the references, appeared in this paragraph: Soumik Pal and Ting-Kam Leonard Wong (2018).... I believe that, for the sake of completeness, some of such recent bibliography should be mentioned in the References and the relations with the present material should be briefly traced.
   b. **Our reply:**
   In Subsection 2.3, we have included the suggested reference and several other recent ones; concerning the latter, we were not completely sure what you meant by "such", and interpreted it as "connected to optimal transport theory and information geometry". The – only loose – relation of Pal & Wong's L-divergence with the present material is discussed in a footnote in Section 3.3.1.1; notice that the general framework of the therein cited paper of Stummer & Kißlinger 2017 (adapted to concave functions) goes even more flexible than the L-divergence (for the separable/additive setup).

5. a. **RR1, Minor Comment 5:**
   There are several points which need elaboration, during the formatting process of the manuscript. To mention a few: ....
   b. **Our reply:**
   We have done your suggested formatting changes. We have left minor exceeds to the margin in some full-line formulas, since from our experience, Springer will re-typeset formulas anyway. Also, the \enumerate-command leads (in the Springer template) automatically to some minor exceeds, and thus can not be corrected.

6. a. **RR1, Minor Comment 6:**
   Given the length of the manuscript, a final section of conclusions seems to be necessary in order to be summarized the results presented here and to be stated some insights of further development.
   b. **Our reply:**
   Done.

2 Review Report 2 (RR2) (the one starting with "The paper presents a unifying framework ...")

1. a. **RR2, General Evaluation:**
   The paper is very interesting and represents a valuable contribution to the study of general divergences and of divergence based methods. The new enlarged context is a flexible one, for instance allows to incorporate data uncertainty, such as data incompleteness, by adding a random argument to the considered functions. The presented concepts and tools are useful for applications in various fields including Machine Learning and Artificial Intelligence. ... In my opinion, the paper deserves to be published even in this form. The degree of novelty is high. However, the answers to the above remarks would be interesting to be added.
   b. **Our reply:**
   Many thanks for your very positive overall assessment. Although you indicate to accept the paper even in the current form, we have incorporated your comments as much as possible (subject to the pregiven page-limit and time-deadline constraints).

2. a. **RR2, First Paragraph of Comments:**
   In Section 4.4 two ways to circumvent the problem presented in Subsetup 2, Sect 4.3, are indicated: grouping of data and smoothing of empirical density function. When applying these techniques, the minimum divergence estimation result depends on the partitioning/smoothing method. Here, the authors mention that the corresponding robustness needs to be addressed. Some details or explanations about this idea regarding the robustness would be useful to be added.
   b. **Our reply:**
   We have already stated in the original text that "for the sake of brevity, a detailed discussion is beyond the scope of this paper" (since we have already reached the pregiven maximum amount of pages). Therefore, we basically only added some references which deal with grouping/smoothing of Csiszar-Ali-Silvey divergences; a detailed discussion (which for the desired wide audience should be comprehensive rather than short) including our new general framework has to be postponed to a forthcoming paper.

3. a. **RR2, Second Paragraph of Comments:**
   Regarding Remark 6, page 48, beside the reference Al Mohamad [4], which uses kernel density estimation in the duality formula in order to define robust estimators, Toma and Broniatowski (2011) ... consider dual $\phi-$divergence estimators (using escort parameter) under misspecification of the model. This reference could be added in text, being closely related to the idea of Remark 6.
   b. **Our reply:**
   Done.

3 Review Report 3 (RR3) (the one starting with "The paper/chapter under review presents ...")

1. a. **RR3, General Evaluation:**
   The authors have also presented the general formulation of divergence between two different types of functions, which appears to be a very useful tool. It is a very interesting and useful piece of work along the lines of the current state-of-the-art of the field. Here are some minor suggestions to improve the presentation of the paper for more general audiences which the authors may consider if they feel appropriate.
   b. **Our reply:**
   Many thanks for your very positive overall assessment. Although you implicitly indicate to accept the paper in the current form, we have incorporated your comments as much as possible (subject to the pregiven page-limit and time-deadline constraints).

2. a. **RR3, (cf. your terminology) Minor Suggestion 1:**
   Section 2 contains several ideas and concepts from different fields where the divergence measures are applied. The literature survey of this section is extremely comprehensive. However, the presentation assumes that the reader is reasonably familiar with the area. To reach the readers who are not familiar with some of these ideas, a presentation in a somewhat simpler language could have been more useful. Of course all depends on the authors and the editor, and if the current set up is what the authors intended then that is fine. After all, it is not as if the particular segment is deficient in any way.
   b. **Our reply:**
   Your inspiring comment to widen the audience (without unwanted loss of precision) could only be realized with substantially more pages; however, we have already reached the pregiven maximum page-limit. Nevertheless, since you indicate that even a wider audience is interested in our addressed issues, we shall take into account your valuable suggestion in future publications.

3. a. **RR3, Minor Suggestion 2:**
   Section 3 onwards, the paper is extensively rich in theory and hence there are a lot of notations. Frankly, I found myself lost among the notations after a few pages. I think, it would be much easier for a reader to follow the main ideas and results if they can be separated out from the notational worry. The authors could add a subsection at the beginning of Section 3 and list there all the notations to be used throughout the rest of the paper. Then, they need not to define these notations in-between text and the presentation will be much clearer. Also, it will help readers to consult this list of notation (at one place) whenever needed.
   b. **Our reply:** We have followed your suggestion and added a list of the main notation and symbols at the beginning of Section 3.

4. a. **RR3, Minor Suggestion 3:**
   For the same reason as above, I suggest that the authors move all the proofs to a separate section at the last or to an appendix.
   b. **Our reply:** Done.

5. a. **RR3, Minor Suggestion 4:**
   It will better to write σ-field, x-th, λ-a.a., etc. in place of $\sigma-$field, x$-$th, $\lambda-$a.a., etc., respectively.
   b. **Our reply:** Done.

6. a. **RR3, Minor Suggestion 5:**
   Section 4 contains different ways of solving minimum divergence problems between two different types of functions. If the authors provided some numerical comparisons in this section that might have been helpful. This actually is a relevant comment for other sections of the article also. The article is entirely theoretical and contains no data.
   b. **Our reply:** Your comment to include concrete data analyses – which for the desired wide audience should be comprehensive (rather than short) in variation, fields of application and explanations – could only be realized with substantially more pages; however, we have already reached the pregiven maximum page-limit. Accordingly, the corresponding data analyses (e.g. of Section 4) will be dealt with in a different future publication. Recalling your comments "... which the authors may consider if they feel appropriate", "... if the current set up is what the authors intended then that is fine", we hope that you accept.

7. a. **RR3, Minor Suggestion 6:**
   In some cases, existing words in English may have been extended to form new words, but I am not sure if they are legitimate English Words. E.g., flexibilizing, or abbreviatingly;.
   b. **Our reply:** We have replaced flexibilizing by "more flexible", and "write abbreviatingly" by "use the abbreviations".