Report on the manuscript entitled

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

by Michel Broniatowski and Wolfgang Stummer

The paper presents a unifying framework of divergences between real-valued functions, as a toolkit useful for various applications in Statistics, Machine Learning and Artificial Intelligence. For real valued functions, including here nonnegative density functions, densities with possible negative values, or even functions whose arguments are real-valued continuous function defined on a time interval, pointwise divergences (for a fixed argument of the functions), as well as aggregate divergences are defined. Corresponding properties, such as nonnegativity, reflexivity and symmetry are discussed and proved under different hypotheses. The new divergences are also connected with geometrical issues. By adapting and widering the concept of Bregman divergence to the context of arbitrary measurable functions, a flexible system of aggregate divergences is introduced. Detailed discussions regarding the elements defining these new divergences, as well as conditions to obtain the nonnegativity and the reflexivity of these divergences are provided. In the new context, minimum divergence methods, including decomposability methods and generalized subdivergence methods, are presented.

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 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.

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) *Minimum dual divergence estimators and tests: robustness results, Journal of Multivariate Analysis, 102, 20-36* consider dual ϕ-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.

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.