Rho-Tau Embedding of Statistical Models

Jan Naudts and Jun Zhang

Overview

The paper gives an informative and clear overview of the research project stated in the abstract. The introduction relates the new research with the relevant literature. The following five sections offer a discussion of different topics in the project. The main aim is to relate general embeddings and statistical manifolds with deformed exponential families through the systematic use of a special non-parametric class of divergences. The style is partly rigorous and partly conceptual. Two specific results are given, namely Th. 2 and Th. 3.

Moving from the introduction to the Sec. 2 on the generalities about rho-tau-divergence introduced by Jun Zhang in [18] I am faced with a the following problem. Most of the literature quoted in the introduction consider a manifold of probability densities while the divergences are defined as functions on random variables. If the random variables are non-negative it is a set of unnormalized densities otherwise I do not know. I feel that some word of explanation would improve the clarity of the argument.

A similar remark applies to Sec. 3, where the tangent space of the manifold is defined. In Eq. (10) the logarithm of a (stricly positive) density is considered as a prototype of a general case to be discussed. In contrast to that, the definition of directional derivative top of page 5 is not really compatible with any positivity assumption. Apparently, Eq. (11) is intended to overcome the difficulty with the introduction of the τ function as a generalized embedding. In the classical case, the embedding $p \mapsto \sqrt{p}$ maps densities to the L^2 sphere, and the push-back of the sphere geometry give a geometry on densities. The corresponding set-up in the generalization should be specified in order the argument to be clear. The statement following Eq. (11) of this section should be clarified, at least to explain the general set-up, before referring the reader to [20]. I can see that formal equivalence of Eq. (11) with the following equation can be derived from Eq. (5).

Detailed comments

The following comments are intended to be suggestions to the Authors to improve the the presentation of the initial sections.

- **page 2, line 11:** A measure space is usually written with brackets, $(\mathcal{X}, \mu)$.
- **page 2, line 12:** The notation for the expectation is not consistent through the paper. Sometimes there are brackets (of two types) and in other cases the brackets are missing.

- **page 2, Eq. (2):** As ρ is differentiable, it holds $d\rho(x) = \rho'(x)dx$. What is the advantage of the Stieltjes Integral notation? In fact, both ρ' and τ' are used in Eq. (12).
- **page 2, line -3:** An "open convex domain of $\mathbb{R}$" is an open interval, correct? The statement as it is makes the reader wonder if it would be possible to define the rho-tau-divergence with ρ and τ defined on $\mathbb{R}^d$, $d > 1$.
- **page 3, line 8:** According to the following Eq. (6), $p(\zeta, \eta)$ is joint *density* function. Is this a special case or the existence of a joint density is a general assumption? As it is stated, it is not clear. However, no special assumption is needed to prove inequality (6).
- **page 3, Eq. (6):** The last inequality is Cauchy–Schwarz inequality. Why is the result relevant? To give sufficient condition for the divergence to be finite? Why not to use the most general Hölder's inequality?
- **page 4, line -14:** The statement "The Frechét derivative of a random variable …" is confusing. The random variables are the points in the manifold, while the Frechét derivative applies to function on some domain.
- **page 5, line 5:** Please explain "to deform the logarithmic function". It is a deformation or a replacement? If τ is an embedding, the space into which the embedding is done should be specified.
- **page 5, Eq. (11):** Please explain the relation between Eq. (11) and Eq. (10). Also, how the inner product defined in the equation following (11) relates with the metric in Eq. (12)?
- **page 5, line 17:** Please check "…in the form of for the non-parametric …".
- **page 5, line -12:** Reference "From (3)" does not seem correct. Is it (12)?