Re.: *JCIS, paper ID: 53* *18 Mars 2010*
Authors: ***Jugurta Montalvão, Romis Attux, Daniel Silva***
Title: ***A pragmatic entropy and differential entropy estimator for small datasets***

Dear Editor,

Thank you very much for your interest in our paper. Valuable comments from the reviewers guided us throughout our revision of the paper. All comments and related changes are detailed below:

Reviewer A:

Comment: "Although the proposed method seems to be useful, the significance of the paper is weakened by a lack of evaluations. The method should be evaluated in a wider range of scenarios, e. g., considering other distributions (Laplace, Exponential,…)."

> Response: In Section 'Experiments with continuous variables', we included experiments with Laplace and exponential random variables in 1D, 2D and 3D, along with experiments for all scenarios with the adapted Ma's method, for the sake of further comparisons.

Comment: "The paper is an extension of a previous work of the authors. Deeper considerations with respect to other papers in the field should be provided."

> Response: The following references were included:

- Bonachela, J. A., Hinrichsen, H., Muñoz, M. A., 2008. Entropy estimates of small data sets. J. Phys. A: Math. Theor. 41, 1—9.
- Ma, S.-K., 1980. Calculation of Entropy from Data of Motion. Journal of Statistieal Physics, Vol. 26, No. 2, 221—240.
- Nemenman, I., 2011. Coincidences and Estimation of Entropies of Random Variables with Large Cardinalities. Entropy. 13, 2013--2023.

As well ss the following texts that were inserted in the Introduction:

- "However, using K coincidence detectors can be problematic. For instance, if the number of available samples is less than K, histogram-based estimators are expected to perform badly, since at least one coincidence counter is not incremented at all, thus inducing strong estimator bias and variance. For small data sets and discrete random variables, \citet{Bonachela2008} propose a method to balance estimator bias and variance, along with a very interesting point of view that elegantly links existing methods such as Miller's and Grassberger's to their own approach''
- "by \citet{Ma1980}, in a journal paper, and re-explained in a book by the same author"
- "\citet{Nemenman2011} further analyses this estimator previously proposed by himself and collaborators, in 2002. His analysis, to a certain extent, bridges the gap between entropy and differential entropy estimation through their coincidence counting approach, by considering random variables with large cardinalities, and thus coming to the conclusion that the {\em a priori} knowledge of the cardinality of the alphabet size is not necessary. It is noteworthy

that it allows for the estimation of differential entropies, where cardinalities tends to infinity. Unfortunately, in spite of this open possibility, the analysed method was not adapted to continuous random variables. By following the same path, in \citep{Montalvao2012} we briefly proposed a simpler method (for discrete variables only) which can be used without any knowledge of the cardinality of the alphabet size, and is simple enough to be easily employed even by experimenters unfamiliar with the theoretical bases of statistical estimation.''

Comment: The claims seem to be convincing. This reviewer strongly recommends the evaluation of the method using other distributions, e.g. Laplace and exponential.

Response: Point taken.

Comment: Considerations with respect to other methods in the literature are expected.

Response: Point taken. As explained above, we included more references and more experimental results from the Ma's method (for comparison).

Comment: The paper has English problems but in general the idea is well explained. Some problems regarding the manuscript.

1. Third line of Section 1: Renyi, 1961. References: Rényi, 1960.

Response: Typos fixed.

2. End of first paragraph of Section 1: Beirlant, 1997 (only one author). References: Berlaint, Dudewicz, Gyorfi, Van Der Meulen (four authors).

Response: Typos fixed.

3. “Ma-like methods can also be attractive for problems belonging to a variety of domains” à There is a lack references and the affirmation of the authors is vague.

Response: We replaced the former sentence with the following: 'we conjecture that Ma-like methods can also be attractive for problems belonging to a variety of domains, wherever phenomena with a huge number of reachable states are observed'.

4. responses(Nemenman à responses (Nemenman

Response: Typos fixed.

5. Is Figure 6 adequately explained? References should be improved.

Response: References were improved and caption of Figure 6 was replaced with the following: "Incremental estimation of the average number of symbols until coincidence detection, for continuously valued samples. Numerical samples are sequentially compared to all other samples, from the latest ``start'' or ``resume'' position, until a new sample falls inside a region already marked by a former sample. When it does occur for a pre-defined region size (Δ), a coincidence is detected, the corresponding delay (D) is recorded, and this process is resumed from next sample."

Reviewer B:

Comment: The proposed strategy is elegant and computationally efficient. The authors could improve the quality of the work by using other distributions to test the conjectures of the paper (e.g., exponential and Laplace distributions).

> Response: In Section 'Experiments with continuous variables', we included experiments with Laplace and exponential random variables in 1D, 2D and 3D, along with experiments for all scenarios with the adapted Ma's method, for the sake of further comparisons.

Comment: As pontual suggestions, we can cite:
1) The number of references should be increased in order to support some parts of the paper (e.g. second paragraph of page 2).

> Response: Point taken. We included more references and more experimental results from the Ma's method (for comparison), and the second paragraph of page 2 was rewritten.

2) Expression in Eq. (5) is an approximation. How could it be transformed in the equality of Eq. (6)?

> Response: The typo in Eq. (6) was fixed.

3) By the end of Section 2 the authors explains the relation between N and K. This should be explained also in the abstract.

> Response: We appended the following to the abstract: "(…) can provide useful estimates even when the number of samples, N, is less than K, for discrete variables, whereas plug-in methods typically demand $N>>K$ for a proper approximation of probability mass functions. Experiments done with both discrete and continuous random variables illustrate the simplicity of use of the proposed method, whereas numerical comparisons to other methods show that, in spite of its simplicity, useful results are yielded."

4) Just after Eq. (9) the authors should explain why K does not goes to infinity when \Delta goes to zero.

> Response: The following text was included: "Please note that the induced cardinality, K, diverges to infinity as Δ goes to zero, as well as $H(\delta)$, but the difference between $H(\delta)$ and $R(\delta)$ converges to a finite value."

5) Section 5 should include experiments with distributions different of Uniform and Gaussian.

> Response: Point taken. In Section 'Experiments with continuous variables', we included experiments with Laplace and exponential random variables in 1D, 2D and 3D, along with experiments for all scenarios with the adapted Ma's method, for the sake of further comparisons.

We would like to thank again the reviewers for their comments and suggestions, and we sincerely hope the above changes are in accordance with their remarks.

Sincerely yours,

The authors