Differential Entropy on Statistical Spaces

Jacques Calmet and Xavier Calmet

*University of Karlsruhe (TH), Germany and University of North Carolina at Chapel Hill, North Carolina, USA**

We must introduce an abstract.

INTRODUCTION

Differential entropy is the entropy of a continuous random variable. It is related to the shortest description length and thus similar to the entropy of a discrete random variable. A basic introduction can be found in the book of Cover and Thomas [1]. In this paper, we are interested in the concept of shortest description length. Indeed, in a recent paper [2], we have investigated the case of spaces where points are in fact localized within a certain volume, i.e. they are statistical in nature. The motivation was the existence of a minimal length in physical theories. It was possible to introduce a concept of distance using Fisher information metric on such spaces. In this paper we show that the reasoning leading to the definition of a distance is analogous to the usual introduction of differential entropy in information theory. In our case also, care must be taken of the precise meaning of minimal distance or shortest description length.

In this extended abstract we only present an outline of our method. It is structured as follows. We remind first the introduction of differential entropy. Then, we show that the concept of distance introduced in [2] leads to a mutual information function analogous to the usual one. This is the main and new contribution of this paper.

Definition 1: The differential entropy $h(X)$ of a continuous random variable X with a density $f(x)$ is defined as

$$h(X) = -\int_S f(x) \log f(x) dx, \qquad (1)$$

where S is the support set of the random variable.

It is well-known that this integral exists if and only if the density function of the random variables is such that the integrals can be defined. This is related to the issue concerning the precise meaning of minimal distance. The next step consists in establishing the relation of differential entropy to discrete entropy. Then, one proceeds to define the joint and conditional differential entropy including the entropy of a multivariable distribution. The next step is to introduce the relative entropy and the mutual information functions.

Definition 2: The Kullback-Leibler distance or relative entropy is defined as

$$D(f||g) \;=\; \int f \log \frac{f}{g}, \qquad (2)$$

where f and g are two density functions.

Definition 3: The mutual information $I(X;Y)$ between two random variables with joint density $f(x,y)$ is defined as

$$I(X;Y) \;=\; \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy. \qquad (3)$$

The previous definitions lead to an equation for the mutual information given by:

$$I(X;Y) \;=\; D(f(x,y)||f(x)f(y)). \qquad (4)$$

An important point is that this provides a link between the discrete and continuous cases since the properties of the Kullback-Leibler distance and the mutual information are the same in both cases.

In [2] we have shown that using the concept of relative entropy, one can introduce a concept of a distance, equivalent to the Kullback-Leibler distance, on a statistical spaces. We propose the following definition

$$I(q_{\theta'^{\mu}}(x^{\mu})||p_{\theta^{\mu}}(x^{\mu})) \;=\; \int d^4x q_{\theta'^{\mu}}(x^{\mu}) \log \frac{q_{\theta'^{\mu}}(x^{\mu})}{p_{\theta^{\mu}}(x^{\mu})} \qquad (5)$$

for the distance between two "points" $p_{\theta^{\mu}}(x^{\mu})$ and $q_{\theta'^{\mu}}(x^{\mu})$. The metric on the manifold of distributions is given locally by

$$g_{\mu\nu} \;=\; \int_X d^4x p_\theta(x) \left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\mu}\right)\left(\frac{1}{p_\theta(x)}\frac{\partial p_\theta(x)}{\partial\theta^\nu}\right) \qquad (6)$$

and corresponds to the Fisher information matrix.

A new contribution in this paper is to show that this definition allows to define the same expression for the mutual information $I(X;Y)$ as given in eq. (4). In the usual case, the differential entropy gives a bound on discrete entropy. We can use a similar method to introduce bounds. However, a big advantage is that we can use Gaussian distributions as density functions which are realistic models for quantum states as used in quantum computing. It is known that finding such bounds is of great significance in quantum computing. This is work in progress.

References

1. T. M. Cover and J. A. Thomas, "Elements of Information Theory," Wiley Series in Telecommunications, 1991.

2. J. Calmet and X. Calmet, "Metric on a Statistical Space-Time," WSEAS Trans. on Circuits and Systems, issue 10, vol. 3, pp. 2267-2271, Dec. 2004.

* Electronic address: calmet@ira.uka.de,calmet@physics.unc.edu