

Statistical dependence measure based on characteristic functions

povilas.daniusis, povilasd@neurotechnology.com

November 2021

Abstract

In this paper we propose multivariate statistical dependence measure based on the difference between joint and product of marginal characteristic functions. We discuss simulated examples, and applications for feature selection/extraction, causal inference and conduct corresponding experiments with diverse collection of multivariate data sets.

1 Introduction

Estimation of statistical dependence, both qualitatively and quantitatively, plays important role in various statistical and machine learning methods (e.g. hypothesis testing, feature selection and extraction [?], information bottleneck methods [?], cost function / reinforcement learning reward design, causal inference [?], among others). Therefore, earliest statistical dependence estimation ideas (e.g. conditional probability) likely share nearly-common origin with the beginning of formal statistical reasoning itself. During last two centuries ideas of correlation and (relative) entropy (including various generalizations) were proposed and became very popular in numerous applications and theoretical developments. However, with the increasing popularity of statistical machine learning, new statistical dependence estimation methods, that are robust, applicable to noisy, high-dimensional data, and can be efficiently integrated with modern machine learning methods are helpful for the development both of the theory and application.

In this study we will begin with the short review of some important previous dependence estimation approaches (Section

2 Previous work

Shannon mutual information [?] and generalizations [?], Hilbert-Schmidt independence criterion [?], distance correlation[?]. Survey [?].

2.1 Characteristic-function-based methods

[?] (...) [?] proposes dependence measures for n random variables of arbitrary dimensions *distance multivariate* and derive from it another dependence measure, called *total distance multivariate*. Both measure rely (...) [?]

3 Proposed Method

Characteristic function of random vector X is defined as

$$\phi_X(\alpha) = \mathbb{E}_{\mathbb{X}} e^{i\alpha^T X}, \quad (1)$$

where $i = \sqrt{-1}$.

We will derive independence measure relying on property of characteristic functions (also known as Kac theorem [?]), that independence of two random vectors $X \in R^{d_x}$ and $Y \in R^{d_y}$ is equivalent to $\forall \alpha \in R^{d_x}, \beta \in R^{d_y}$,

$$\phi_{X,Y}(\alpha + \beta) = \phi_X(\alpha)\phi_Y(\beta), \quad (2)$$

where d_y are dimensions of X and Y , respectively.

This motivates the construction of a novel statistical measure, which we further refer to as Kac independence measure (KacIM):

$$\kappa(X, Y) = \max_{\alpha, \beta} |\phi_{X,Y}(\alpha + \beta) - \phi_X(\alpha)\phi_Y(\beta)| \quad (3)$$

It is easy to see that $0 \leq \kappa(X, Y) \leq 1$, $\kappa(X, Y) = \kappa(Y, X)$.

3.1 Estimation

Having i.i.d. data (x_j, y_j) , $j = 1, 2, \dots, n$ an empirical estimator of KacIM (is defined via corresponding empirical characteristic functions:

$$\hat{\kappa}(X, Y) = \max_{\|\alpha\|=\|\beta\|=1} \left| \frac{1}{n} \sum_{j=1}^n e^{i(\langle \alpha, x_j \rangle + \langle \beta, y_j \rangle)} - \frac{1}{n^2} \sum_{j=1}^n e^{i\langle \alpha, x_j \rangle} \sum_{k=1}^n e^{i\langle \beta, y_k \rangle} \right|. \quad (4)$$

Empirical estimator also admits $0 \leq \hat{\kappa}(X, Y) \leq 1$. Normalisation of parameters α and β on to unit sphere is included due to stability issues. The estimator (can be calculated by using Algorithm

Algorithm 1 KacIM estimator computation algorithm

Require: data batch (x, y) , gradient-based optimiser $GradOpt(loss)$

Calculate KacIM estimator $\hat{\kappa}(x, y)$, without maximization step (i.e. using current α, β).

Perform one maximization iteration of computed $\hat{\kappa}(x, y)$ via $\alpha, \beta \rightarrow GradOpt(\hat{\kappa}(x, y))$.

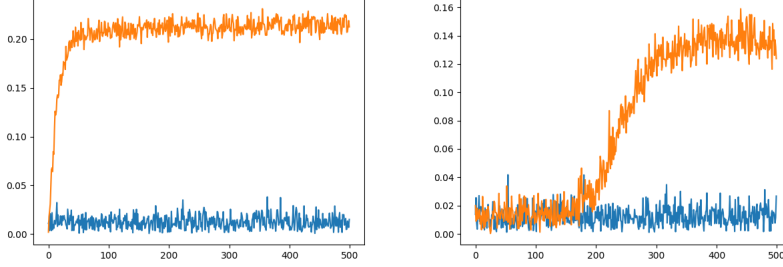


Figure 1: Dependence detection in additive (left) and multiplicative (right) noise scenario.

4 Experiments

Dependence measures have broad area of applications. For example, regularization [?, ?], feature selection and extraction [?], information bottleneck methods [?], causal inference [?], among others. Further we will conduct empirical investigation of KacIM. Starting with simple illustrative simulations, we will reformulate some key ideas in aforementioned topics for KacIM, and experimentally investigate corresponding empirical scenarios.

4.1 Generated data

We begin with simple example, which demonstrates the efficiency of KacIM for simulated multivariate data with additive and multiplicative noise.

In Figure

When data is independent (blue graph), both in additive and multiplicative cases, due to independence, estimator (is resistant to maximization, and oscillates near zero. On the other hand, when data is not independent (orange graph), the condition of Kac theorem is violated and maximization of estimator (is possible.

4.2 Influence of noise level λ to KacIM estimator value

In this simulation we use the same additive noise setting as in previous paragraph, but evaluate all noise levels $\lambda \in [0.0, 2.4]$, with step 0.1. Figure

4.3 Feature Extraction

We conduct linear feature extraction by seeking

$$W^* = \arg \max_W \kappa(Wx, y). \quad (5)$$

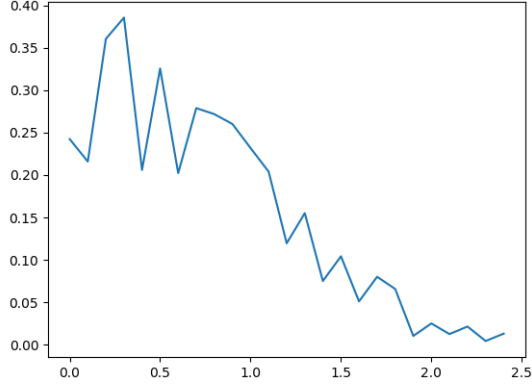


Figure 2: Noise level (x axis) vs final iteration KacIM value (y axis). KacIM values for larger noise levels saturates as in tail of graph.

Afterwards, feature extraction is conducted by $f = W^*x$ and k -nearest neighbor classification with Euclidean distance is performed, comparing unmodified inputs x and features of all possible dimensions up to d_x .

5 Discussion

In this article we propose statistical dependence measure, KacIM, which relies on simple fact that statistical independence is equivalent to the decomposability of joint characteristic function into the product of marginal ones. Although we formulated and analysed KacIM for bivariate vectorial case, similarly it can be generalised for multivariate case. In addition, since characteristic functions are defined for matrices, graphs, and other objects [?], likely KacIM can be extended to those objects as well, which is potential direction of future research of KacIM.

Empirical analysis show, that KacIM can detect and measure statistical independence for non-linearly related, high-dimensional data. (...)

6 Acknowledgements