

# Measuring statistical dependency with optimal transport

Thursday, 4 August 2022 15:30 (20 minutes)

In this talk, we introduce a novel framework for measuring statistical dependency between two random variables  $X$  and  $Y$ , the *transport dependency*  $\tau(X, Y) \geq 0$ . This coefficient relies on the notion of optimal transport and is applicable to random variables, taking values in general Polish spaces. It can be estimated consistently via the corresponding empirical measure, is versatile and adaptable to various scenarios by proper choices of the cost function, and intrinsically respects metric properties of the ground spaces. Notably, statistical independence is characterized by  $\tau(X, Y) = 0$ , while large values of  $\tau(X, Y)$  indicate highly regular relations between  $X$  and  $Y$ . Indeed, for suitable base costs,  $\tau(X, Y)$  is maximized if and only if  $Y$  can be expressed as 1-Lipschitz function of  $X$  or vice versa.

We exploit this characterization and define a class of dependency coefficients with values in  $[0, 1]$ , which can emphasize different functional relations. In particular, for suitable costs the *transport correlations* is symmetric and attains the value 1 if and only if  $Y = f(X)$  where  $f$  is a multiple of an isometry, which makes it comparable to the distance correlation.

Finally we illustrate how the transport dependency can be used in practice to explore dependencies between random variables, in a gene expression study.

**Primary authors:** NIES, Thomas Giacomo (University of Göttingen); Mr STAUDT, Thomas (University of Göttingen); Prof. MUNK, Axel (University of Göttingen)

**Session Classification:** Session B5 Statistics

**Track Classification:** Statistics