Università Commerciale Luigi Bocconi

Department of Decision Sciences
Statistics Seminars

Efficient estimation with incomplete data via generalised ANOVA decompositions

Tom Berrett
University of Warwick

Thursday, 14th November 2024
12:00 pm **Room 3-E4-SR03** Via Roentgen 1 Milano

Abstract

In this talk I will present recent work (https://arxiv.org/abs/2409.05729) on efficient estimation with incomplete data, covering problems arising in semi-supervised learning, data fusion and missing data literatures. Our task is to estimate simple mean functionals given access to a complete dataset that is supplemented by additional incomplete datasets. In particular, we aim to use the incomplete data to reduce the variance of the naive complete-case estimator, and to characterise the minimal asymptotic risk among all estimators. Results of this type exist for monotonic missingness structures, such as those arising in semi-supervised learning and longitudinal studies, but in this work we consider more general settings. We show that the optimal variance can be expressed through the minimal value of a quadratic optimisation problem over a function space, thus establishing a fundamental link between these estimation problems and the theory of generalised ANOVA decompositions. We introduce an estimator that is proved to attain this minimal risk and to be approximately normally distributed, and use this to construct confidence intervals.

Department of Decision Sciences
Via Röntgen 1 – 20136 Milano

Tel. 02 5836.5632
Fax 02 5836.5630