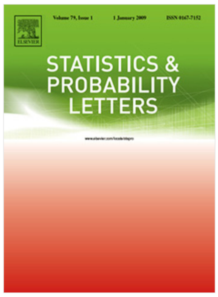
Accepted Manuscript



On the role of latent variable models in the era of big data

Francesco Bartolucci, Silvia Bacci, Antonietta Mira

PII: S0167-7152(18)30068-3
DOI: https://doi.org/10.1016/j.spl.2018.02.023
Reference: STAPRO 8145

To appear in: *Statistics and Probability Letters*

On the role of latent variable models in the era of big data

Francesco Bartolucci*

Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Perugia (IT)

Silvia Bacci

Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Perugia (IT)

Antonietta Mira

Institute of Computational Science, Università della Svizzera italiana, Via Buffi 13, Lugano (CH) and Università dell'Insubria, Como (IT)

Abstract

We discuss how latent variable models are useful to deal with the complexities of big data from different perspectives: simplification of data structure; flexible representation of dependence between variables; reduction of selection bias. Problems involved in parameter estimation are also discussed.

Keywords: Bayesian inference, Complex data, Maximum likelihood estimation, Parallel computing, Selection bias

2010 MSC: 00-01, 99-00

1. Introduction

Big data present different types of complexity that go well beyond their mere size and make their use problematic, possibly leading to biased conclusions on the main investigated effects. We discuss the role of latent variable

*Corresponding author
Email addresses: francesco.bartolucci@unipg.it (Francesco Bartolucci), silvia.bacci@unipg.it (Silvia Bacci), antonietta.mira@usi.ch (Antonietta Mira)