

**Silvia Pandolfi**, Università degli Studi di Perugia (Italy)

**Title.** Hidden Markov models for longitudinal data: advances to deal with missing data, dropout, and variable selection

**Abstract.** Hidden Markov (HM) models may be usefully applied for the analysis of longitudinal data, as they deal with time dependence in a flexible way and allow us to perform a dynamic model-based clustering. We illustrate methodological and inferential advances to address the problem of missing responses and dropout, which typically arise in the context of longitudinal data. Additionally, we perform variable selection to reduce model complexity. The proposed approach is based on an HM model for multivariate continuous responses, according to which, given the latent state, the response variables are assumed to follow a multivariate Gaussian distribution with state-specific parameters. The method explicitly accounts for different patterns of missing data, that is, the intermittent pattern, treated under the missing at random assumption, and the monotone pattern also known as informative dropout. To address the latter, we introduce an extra absorbing state in the latent state space. Furthermore, we propose an approach for simultaneously perform variable and model selection to choose the optimal number of informative variables and clusters (or states) by relying on a greedy search algorithm. The objective is that of obtaining a parsimonious model that provides more stable parameter estimates and enhances interpretability, particularly for high-dimensional data. We illustrate the proposed methodology through two applications: one based on historical data about primary biliary cholangitis and another using macroeconomic data referred to socioeconomic indicators.

**Reza Skandari**, Imperial College Business School (UK)