Computational Statistics and Data Analysis 93 (2016) 131–145

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

General framework and model building in the class of Hidden Mixture Transition Distribution models

Danilo Bolano [a,b,∗], André Berchtold [a,c]

[a] National Center of Competence in Research LIVES, Switzerland
[b] Institute of Demographic and Life Course Studies, University of Geneva, Switzerland
[c] Institute of Social Sciences, University of Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 29 April 2014
Received in revised form 29 August 2014
Accepted 12 September 2014
Available online 22 September 2014

Keywords:
Mixture model
Model selection
Hidden Markov model
Mixture Transition Distribution model
BIC
Panel data

ABSTRACT

Modeling time series that present non-Gaussian features plays as central role in many fields, including finance, seismology, psychological, and life course studies. The Hidden Mixture Transition Distribution model is an answer to the complexity of such series. The observed heterogeneity can be induced by one or several latent factors, and each level of these factors is related to a different component of the observed process. The time series is then treated as a mixture and the relation between the components is governed by a Markovian latent transition process. This framework generalizes several specifications that appear separately in related literature. Both the expectation and the standard deviation of each component are allowed to be functions of the past of the process. The latent process can be of any order, and can be modeled using a discrete Mixture Transition Distribution. The effects of covariates at the visible and hidden levels are also investigated. One of the main difficulties lies in correctly specifying the structure of the model. Therefore, we propose a hierarchical model selection procedure that exploits the multilevel structure of our approach. Finally, we illustrate the model and the model selection procedure through a real application in social science.

1. Introduction

Real data are often a combination of many different, possibly non-observed causes that lead to apparently unpredictable behaviors. For example, in the context of longitudinal data, time series may show non-homogeneous behaviors, can switch between alternative regimes characterized by a low or high variance, can contain extreme values, and the distribution of future values can take complex multimodal shapes.

The Hidden Mixture Transition Distribution (HMTD) model considered in this study is a general framework to study time series. The model can be used to describe and analyze the evolution of any continuous variable observed on a set of M independent sequences that can also vary in length. The model integrates several refinements of the mixture model and the hidden Markov model. More specifically, it can be used for different purposes, including describing observed data, searching for a generalizable model, testing hypotheses, prediction, and classifying time series.

Mixture models are a popular and efficient approach, both for cross-sectional data and time series, to describe multimodal distributions that do not correspond to any specific statistical family. Related literature provides many examples of the usefulness of such models since the work of Weldon and Pearson at the end of the 19th century. Historically, mixture models

∗ Correspondence to: University of Geneva (CH), bd. du Pont d'Arve 40, Switzerland. Tel.: +41 22 379 98 74.
E-mail addresses: Danilo.Bolano@unige.ch, d_bolano@yahoo.it (D. Bolano), Andre.Berchtold@unil.ch (A. Berchtold).

http://dx.doi.org/10.1016/j.csda.2014.09.011

for continuous-valued time series were introduced in Le et al. (1996) as the Gaussian Mixture Transition Distribution (GMTD) model, building upon the earlier work of Raftery (1985) in the discrete case. See Berchtold and Raftery (2002) for a complete review of the basic principles of Mixture Transition Distribution (MTD) models. The general principle of all MTD-like models for count data is to combine different Gaussian distributions (called *components*) using a mixture model in which the mean of each distribution is a function of the past observed process. The weights associated to each component are interpreted as the probability of that specific component generating the next value of the process. This class of models has been expanded on in several ways. For example, allowing detailed specifications for the mean of each component (e.g., Wong and Li, 2001), allowing the variance of each component to depend on its past (Wong and Li, 2001; Berchtold, 2003), and replacing the fixed probabilities associated with each component by a Markov chain (Bartolucci and Farcomeni, 2010). Then, covariates have been included (Chariatte et al., 2008; Luo and Qiu, 2009) used non-Gaussian distributions, and extensions to bivariate data have been suggested (Hassan and Lii, 2006; Hassan and El-Bassiouni, 2013).

Hidden Markov Models (HMMs) form another class of stochastic processes often used to represent and analyze complex time series in the presence of over-dispersion. They are particularly well suited to analyzing data switching between several regimes. In its traditional formulation, a HMM combines a hidden first-order Markov chain with different conditionally independent distributions for the observed process. Here each distribution is related to one of the states of the Hidden Markov chain. Many developments to the basic HMM have been proposed, including using high-order Markov chains and removing the conditional independence hypothesis. The latter change allows the observed process to depend on both its past and the hidden process (Wellekens, 1987; Berchtold, 1999, 2002). Under the hypothesis of stationarity, the marginal distribution of the observations in a HMM is a finite mixture, with the number of components being equal to the number of states in the unobserved Markov process. Therefore, we may consider the two models as one unique approach. When the transitions between components are driven by a Markov chain, the mixture transition model becomes a HMM. In addition, when the observations are allowed to follow an autoregressive process, the HMM becomes a mixture transition model. Some authors even refer to HMMs as Markov-dependent mixture models (as coined by Leroux, 1992).

It is worth noting that mixture transition and Markovian models are distributional stochastic models, in contrast to other well-known and widespread approaches that are essentially point processes. The latter group includes the ARMA and ARIMA models, which are based on autoregressive equations, and the ARCH and GARCH models, which explicitly consider the variance of the process. See, for example, Box et al. (1994), Hamilton (1994), and Bollerslev et al. (1992) for a complete review of these models, and Kon (1984) and Kim and Kon (1994) for a comparison of the different models. When trying to predict the next value of a series, the advantage of the point approach is that the model will provide a clear answer as one numerical value (associated with a confidence interval). However, the drawback is that, in most cases, the answer is either inaccurate or completely wrong. Given the high variability of many time series, the expectation of the model is not a good estimator of the value of the next observation. Even when modeling the variance of the process, there is only a small probability of accurately predicting the next value. However, in the probabilistic approach, instead of trying to determine *the* next value of the series, a non-null probability is associated with values (discrete case) or intervals (continuous case), which are then possible candidates to be *the* next data. Then, an adequate probabilistic model does not provide a single value, but rather leads to a complete representation of possible futures through a (possibly multimodal) distribution. In that sense, the answer given by this approach generally has a higher probability of helping to make the right decisions, because it shows all possibilities rather than one (probably) wrong value.

Hidden Markov models were historically used for speech recognition (Rabiner, 1989; Baum and Petrie, 1966), but many applications have since been found in other fields, including econometrics (e.g., Elliott et al., 1998; Hayashi, 2004; Netzer et al., 2008) and the biosciences (e.g., Le Strat and Carrat, 1999; Shirley et al., 2010). Mixture models for count data are also quite common in finance and biomedical studies (Sclattmann, 2009) and behavioral studies (under the name of growth mixture modeling, e.g., Muthen, 2001). However, Mixture Transition Models seem to be used quite exclusively in economics and finance (e.g., Wong and Chan, 2005; Frydman and Schuermann, 2008). In fact, despite their unanimously recognized advantages, mixture transition and hidden models are still only sparsely used in social sciences. This is unfortunate, since the current trend in this field is clearly to switch from cross-sectional to longitudinal surveys, hence the need for advanced methods for modeling longitudinal data showing non-Gaussian distributions.

In addition, even though many developments have been proposed during the past few decades on the basic MTD and HMM models, there is still a need for a more general framework that integrates these refinements. As a result, this study has three objectives. First, we define a general framework that integrates the many extensions presented previously separately in the literature (see Section 2), and then we discuss the estimation procedure (see Section 3). Second, as noted by Rabiner (1989), the number of possibilities offered by hidden models is so large that it becomes difficult to identify an adequate model structure for a particular research question. Therefore, in Section 4, we outline a search strategy similar to that used with other multilevel models. Finally, we illustrate the model and the proposed procedure using a real dataset, the US Panel Study of Income Dynamics (see Section 5).

2. The HMTD model

The Hidden Mixture Transition Distribution (HMTD) model developed in this paper combines a hidden and an observed level. It can be used to describe and analyze the evolution of any continuous variable observed on a set of M independent sequences, which may vary in length.