Hidden Markov Models for Multivariate Longitudinal Data

Alexa Sochaniwsky [1]and Paul D. McNicholas[1]

[1] Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada (e-mail: sochaa1@mcmaster.ca, paul@math.mcmaster.ca)

Abstract: A method for handling the unique correlation structure that can occur in longitudinal data is introduced for hidden Markov models. This approach uses a family of independent mixture models that apply a variety of constraints to the covariance matrix, which is then used in hidden Markov models, i.e, dependent mixture models.

Keywords: clustering, hidden Markov models, longitudinal data, EM algorithm.

1 Introduction

Longitudinal data is information that is collected on several subjects across several points in time. Longitudinal studies are often used in clinical or sociological research, but difficulties may arise as the correlation that can occur between subjects must be accounted for. For certain longitudinal studies, it would be useful not only to cluster the subjects but to model the transitions between states. The change in state can be modeled by hidden Markov models (HMMs). Efforts have been made in regression models, specifically AR and MA models (Hasan & Sneddon, 2009; Sutradhar, 2003), and in independent mixture models (McNicholas & Murphy, 2010) to account for the unique longitudinal correlation structure. This research modifies the EM algorithm for HMMs by using the covariance structures from the Cholesky-decomposed Gaussian mixture model (CDGMM) family (McNicholas & Murphy, 2010).

2 Background

A hidden Markov model comprises of two processes, an unobserved parameter process and an observed state-dependent process. The simplest HMM for longitudinal data can be defined as

$$P(C_{it}|\mathbf{C}^{(it-1)}) = P(C_{it}|C_{it-1}), \qquad \text{for } i = 1,...,n,\ t = 2,3,...,T$$
$$P(X_{it}|\mathbf{X}^{(it-1)},\mathbf{C}^{(t)}) = P(X_{it}|C_{it}), \qquad \text{for } i = 1,...,n,\ t = 1,...,T$$

where $\mathbf{C}^{(it)}$ represents the history of the unobserved parameter process $\{C_{it} : i = 1,...,n, t = 1,2,...,T\}$ with state space $C = 1,...,m$, and $\mathbf{X}^{(it)}$ represents the history of the state-dependent process $\{X_{it} : i = 1,...,n, t = 1,2,...,T\}$. The parameter process C_{it} satisfies the Markov property and is then used in the distribution of the state-dependent process X_{it}.

A common method for maximum likelihood estimation of an HMM is the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977). An EM for HMMs is called the Baum-Welch algorithm (Baum *et al.*, 1970, 1972; Welch, 2003). Specifically, it is an EM for a hidden Markov model whose Markov chain is homogeneous. By assuming a homogeneous HMM, the parameter estimates have closed form solutions. The parameters are derived from the complete-data log-likelihood given by

$$l(\boldsymbol{\vartheta}) = \sum_{i=1}^{n} \left\{ \sum_{g=1}^{m} u_{i1g} \log \delta_i + \sum_{t=2}^{T} \sum_{g=1}^{m} \sum_{k=1}^{m} v_{itgk} \log \gamma_{gk} \right. \\ \left. + \sum_{t=1}^{T} \sum_{g=1}^{m} u_{itg} \log f(x_{it} | S_{it} = g) \right\},$$

where $\boldsymbol{\vartheta}$ denotes the vector containing the model parameters, δ_i is the stationary distribution, γ_{gk} are the transition probabilities, the unknown labels $u_{itg} = 1$ if the observation i is in state g at time t and $u_{itg} = 0$ otherwise, and the other unknown labels $v_{itgk} = 1$ if the observation i is in state g at time $t-1$ and in state k at time t, and $v_{itgk} = 0$ otherwise.

For longitudinal data, McNicholas & Murphy (2010) use a Gaussian (independent) mixture model with a modified Cholesky decomposed covariance structure (Pourahmadi, 1999, 2000) such that the precision matrix $\boldsymbol{\Sigma}$ can be decomposed into $\boldsymbol{\Sigma}^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$, where $\mathbf{T}$ is a unique unit lower triangular matrix and $\mathbf{D}$ is a unique diagonal matrix with strictly positive diagonal entries. For a p-dimensional random variable $\mathbf{X}$, the multivariate Gaussian mixture model with the modified-Cholesky decomposition, the gth component density is given by

$$f(\mathbf{x}|\boldsymbol{\mu}_g, (\mathbf{T}_g'\mathbf{D}_g^{-1}\mathbf{T}_g)^{-1}) = \frac{1}{\sqrt{(2\pi)^p |\mathbf{D}_g|}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_g)'\mathbf{T}_g'\mathbf{D}_g^{-1}\mathbf{T}_g(\mathbf{x}-\boldsymbol{\mu}_g) \right\}.$$

A family of eight Gaussian mixture models are constructed by constraining $\mathbf{T}_g$ and/or $\mathbf{D}_g$ with the option to impose the isotropic constraint $\mathbf{D}_g = \delta_g \mathbf{I}_g$. This family is called the Cholesky-decomposed Gaussian mixture model (CDGMM) family. The nomenclature, covariance structure, and number of free covariance parameters for all models are displayed in Table 1.

Table 1. *CDGMM Family*

Model	$\mathbf{T}_g$	$\mathbf{D}_g$	$\mathbf{D}_g$	Free Cov. Parameters
EEA	Equal	Equal	Anisotropic	$p(p-1)/2+p$
VVA	Variable	Variable	Anisotropic	$m[p(p-1)/2]+mp$
VEA	Variable	Equal	Anisotropic	$m[p(p-1)/2]+p$
EVA	Equal	Variable	Anisotropic	$p(p-1)/2+mp$
VVI	Variable	Variable	Isotropic	$m[p(p-1)/2]+m$
VEI	Variable	Equal	Isotropic	$m[p(p-1)/2]+1$
EVI	Equal	Variable	Isotropic	$p(p-1)/2+m$
EEI	Equal	Equal	Isotropic	$p(p-1)/2+1$

Constraining $\mathbf{T}_g$ such that $\mathbf{T}_g = \mathbf{T}$ suggests that all states have the same correlation structure. Constraining $\mathbf{D}_g$ such that $\mathbf{D}_g = \mathbf{D}$ suggests that all states have the same variability at each time point and the isotropic constraint $\mathbf{D}_g = \delta_g \mathbf{I}_p$ suggests that the variability at each time point is the same. All models would be fitted using an EM algorithm and then based on a model selection criterion, one would be selected.

3 Methodology

We propose modifying the M-step in the EM algorithm for a Gaussian HMM by substituting the 'traditional' covariance update, i.e.,

$$\mathbf{\Sigma}_g = \frac{1}{n_g} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{u}_{itg}(x_{it} - \boldsymbol{\mu}_g)(x_{it} - \boldsymbol{\mu}_g)'$$

where $n_g = \sum_{i=1}^{n} \sum_{t=2}^{T} \hat{u}_{itg}$, with a member of the CDGMM family. This modified algorithm is outlined in Algorithm 1.

Algorithm 1 EM Algorithm for Gaussian HMM

initialize $\boldsymbol{\delta}$ and $\boldsymbol{\Gamma}$
initialize u_{itg} and v_{itgk}
while convergence criterion is not met **do**
  update $\hat{u}_{itg}$, $\hat{v}_{itgk}$
  update γ_{gk}, δ_g
  update $\hat{\boldsymbol{\mu}}_g$
  update $\hat{\mathbf{T}}_g$, $\hat{\mathbf{D}}_g$
  update $\hat{\boldsymbol{\Sigma}}_g^{-1} = \hat{\mathbf{T}}_g'\hat{\mathbf{D}}_g^{-1}\hat{\mathbf{T}}_g$
  check convergence criterion
end while

References

Baum, Leonard E, Petrie, Ted, Soules, George, & Weiss, Norman. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics*, **41**(1), 164–171.

Baum, Leonard E, *et al.* 1972. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, **3**(1), 1–8.

Dempster, Arthur P, Laird, Nan M, & Rubin, Donald B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1), 1–22.

Hasan, M Tariqul, & Sneddon, Gary. 2009. Zero-inflated Poisson regression for longitudinal data. *Communications in Statistics—Simulation and Computation®*, **38**(3), 638–653.

McNicholas, Paul D, & Murphy, T Brendan. 2010. Model-based clustering of longitudinal data. *Canadian Journal of Statistics*, **38**(1), 153–168.

Pourahmadi, Mohsen. 1999. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, **86**(3), 677–690.

Pourahmadi, Mohsen. 2000. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, **87**(2), 425–435.

Sutradhar, Brajendra C. 2003. An overview on regression models for discrete longitudinal responses. *Statistical Science*, **18**(3), 377–393.

Welch, Lloyd R. 2003. Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, **53**(4), 10–13.