Health Services and Outcomes Research Methodology
https://doi.org/10.1007/s10742-020-00217-5

Mixed location scale hidden Markov model for the analysis of intensive longitudinal data

Xiaolei Lin[1] · Robin Mermelstein[2] · Donald Hedeker[3]

Received: 24 March 2020 / Revised: 26 July 2020 / Accepted: 5 August 2020

Abstract
Hidden Markov models (HMM) presents an attractive analytical framework for capturing the state-switching process for auto-correlated data. These models have been extended to longitudinal data setting where simultaneous multiple processes are observed by including subject specific random effects. However, application of HMMs for intensive longitudinal data, where each subject gets measured intensively over relatively short period of time, has not been widely studied. In this paper, we extend the mixed hidden Markov model and allow subject heterogeneity with respect to the mean and within subject variance by including subject random effects in both perspectives. We focus on the application of this model to intensive longitudinal studies in psychological and behavioral research where individual's latent states and state-switching process are of interest. Models are estimated using forward–backward algorithm via Bayesian sampling approach. Advantages over regular HMM and mixed HMM that only accounts for the subjects' mean heterogeneity are illustrated through a series of simulation studies. Finally, models are applied to an adolescent mood study data set and results show that the proposed mixed location scale HMM achieves better model fit and more interpretative mood state identification in terms of state specific covariate effects compared to regular HMM and mixed HMM.

Keywords Intensive longitudinal data · Mixed effect models · Latent class classification

1 Introduction

Hidden Markov models (HMMs) are an ubiquitous tool for modeling time dependent data. They are used to represent the probability distributions over sequences of observations and include two stochastic processes: an underlying hidden (latent) process that is assumed to follow a Markov chain, and an observed process that are modeled as independent distributions conditional on the hidden process. The hidden process is specified by an initial

✉ Xiaolei Lin
xiaoleilin@fudan.edu.cn

[1] Fudan University, Shanghai, China

[2] University of Illinois at Chicago, Chicago, USA

[3] University of Chicago, Chicago, USA

probability distribution over discrete states and a transition probability matrix that represents the probability of moving from one state to another. As with first-order Markov chain, the hidden process assumes that the probability of a particular state depends on its history only through the previous state. The observed process is specified by the emission probabilities or conditional probabilities, representing the likelihood of an observation being generated from a specific state. It assumes that the observations depend on their history only through their current state instead of any other states or observations. Due to the flexibility in the model structure, Hidden Markov Models are widely used in speech recognition, computational biology, data compression and other areas of artificial intelligence (Gales and Young 2008; Choo et al. 2004; Ghahramani 2001).

In longitudinal data settings, individual differences are to be expected in addition to time serial dependence, and thus HMMs with random effects, also referred to as mixed HMMs, offer a flexible alternative in analyzing longitudinal data when a person alternates between discrete states in a wide range of situations (Altman 2007). In mixed HMMs, subject can exhibit heterogeneity in both the transition patterns across time and the probability distributions given a specific state, leading to the potential inclusion of random effects in both the observed and hidden process. Using mixed HMMs, researchers are able to identify discrete affective states (e.g., pleasant-unpleasant mood, calm-tense mood), cognition (e.g., appraisals, self-esteem) and behaviors (e.g., treatment compliance, drinking behaviors) as well as how these latent states fluctuate over time, in the presence of subject heterogeneity (Crayen et al. 2012). For psychological and behavioral research, subjects are often measured intensively over relatively short period of time, resulting in a burst of longitudinal data. For this type of intensive longitudinal data, subjects often exhibit strong heterogeneity in terms of both their mean and within subject variability. Including random effects in the within subject variance model can significantly increase the model fit yet provide information about how subjects influence their variability in addition to measured covariates (Hedeker et al. 2008; Rast et al. 2012; Leckie et al. 2014; Goldstein et al. 2018). de Haan-Rietdijk et al. applied mixed hidden Markov model to intensive longitudinal data where subject specific transition patterns are allowed (de Haan-Rietdijk et al. 2017). However, few mixed HMMs have been extended in the intensive longitudinal setting where outcome mean (location) and within subject variability (scale) can jointly provide insights about distinct states identification and differential state-dependent covariate effects (Shiffman et al. 2008).

To account for individual heterogeneity in HMMs, both observed covariates and subject specific random effects are incorporated in the conditional distributions for the observed process. The inclusion of random effects are crucial in identifying subjects' hidden states at each time point since it reduces the bias arising from unobserved subject level variables that could have been attributed to systematic state differences. The addition of random effects in the error variance is a natural extension of the regular mixed HMMs and provides an effective alternative for statistical inference and latent state identification with regard to within subject variance, which is of great importance from psychological and behavioral research perspective (Lin et al. 2018a).

In this article, we focus on extending the mixed HMMs to include additional random effects in the error variance of the conditional distribution, which applies in the context of intensive longitudinal data and provides a general framework for working with intensive longitudinal data. The augmented model allows subject heterogeneity in both the mean and within subject variance of the outcomes in the observed process. The article is organized as follows. In Sect. 2, an adolescent mood study example is introduced and described to motivate the proposed model. In Sect. 3, both mixed HMM and the proposed Mixed Location

Scale HMM are explained in the context of the motivating example, and Bayesian estimation methods are briefly mentioned. In Sect. 4, we compare the performance of regular HMM, mixed HMM and mixed location scale HMM using a series of simulation studies. Finally, the article concludes with the results of applying the models to the motivating example and a discussion.

2 Motivating example

This research is motivated by an intensive longitudinal study which aims to investigate the role of psycho-social factors on mood regulation among adolescents. The entire study was conducted across 6 waves: baseline, 6 months, 15 months, 2 years, 5 years and 6 years from the baseline. For models considered in this paper, we will only focus on data from the baseline wave. Detailed description of the data can be found in a previous study (Lin et al. 2018b).

At baseline, 339 adolescents (with complete longitudinal measurements) from 9th and 10th grade (average age 15.6, with minimum 14.4 and maximum 16.7) were asked to carry electronic devices and answer questions when randomly prompted during a 7 day study period. Each individual was prompted multiple times within a single day. Questions included location, activities, companionship, mood and other psychological assessments. The primary outcomes of interest are positive affect (PA) as well as negative affect (NA), which consist of the average of several mood items rated from 1 to 10 that are proxies for subject's positive/negative mood measurements. For PA, questions include: I felt happy, I felt relaxed, I felt cheerful, I felt confident, and I felt accepted by others; for NA, questions include: I felt sad, I felt stressed, I felt angry, I felt frustrated, and I felt irritable. Higher PA levels indicates better mood while higher NA indicates worse mood. For the analyses presented here, we will be investigating the effects of subject level variable "smoking status" (indicator of whether a person smokes during the study period) and occasion level variable "alone" (indicator of whether a person is alone or with others at the time of the prompt) on individuals' PA assessment. In order to adapt the discrete time HMM which requires roughly equal time interval, observations as well as observational level covariate (alone) are averaged within a day to produce average PA assessment and proportions of time occasions that the individual is alone on that particular day.

A person's mood is regulated by a variety of biological, psychological and social factors, and thus it is possible to fluctuate frequently between distinct states (e.g., pleasant–unpleasant). The pattern of mood fluctuation partly reflects subjects' mood regulation behavior. Research demonstrated that adolescents differ considerably in their mood regulation ability and effectiveness, and even for the same individual, mood regulation activity can change quite a lot over time (Rocke et al. 2009). For instance, the effect of smoking in regulating an individual's mood could change from positive at an earlier time point to negative at later time points, while the direction can be totally opposite for a different individual, depending on the amount of unmeasured covariates and surrounding social environment. Mixed Hidden Markov Models are well suited to investigate both these inter-individual and intra-individual variability since these models take into account the time serial dependence of mood states for the same individual while adjusting for the individual differences (MacDonald and Zucchini 1997). That is, an adolescent is more likely to be in a pleasant mood state if he or she has a pleasant mood state at the previous time point; some adolescents have the tendency to have more pleasant mood compared to others though they might have the same measured characteristics.

In what follows, we propose a more comprehensive mixed HMM that takes into account individual differences in both the mean and within subject variance of the outcomes for the observed process. For demonstration purposes, we will illustrate the model framework using a normally distributed outcome, one subject level covariate (smoking status) and one occasion level covariate (alone) in the context of the example intensive longitudinal study.

3 Mixed location scale hidden Markov model

In this section, we propose a mixed location scale HMM on top of the existing mixed HMM. We start by introducing the mixed HMM developed in longitudinal data settings (Altman 2007). We then extend the mixed HMM in the context of quantitative psychological and behavioral studies where subjects are intensively measured longitudinally over time and the primary research interests are centered around how subjects change in terms of their within subject variability.

3.1 Mixed HMM

Let $Z_i = (z_{i1}, z_{i2}, \ldots, z_{in_i})$ be an unobserved binary random vector whose elements follow a two-state Markov chain (state 1 represents a distinct mood state compared to state 2) with unknown transition probability matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and initial probability distribution $(\pi_0, 1 - \pi_0)$, with $a_{11} + a_{12} = 1$ and $a_{21} + a_{22} = 1$, as shown in Eq. 1.

$$Pr(z_1 = 1) = \pi_0, \qquad Pr(z_t \mid z_{t-1}) = \prod_{j=1}^{2} \prod_{k=1}^{2} a_{jk}^{\mathbb{1}_{(z_{t-1}=j)} \mathbb{1}_{(z_t=k)}}, \quad t > 1 \tag{1}$$

The (first order) Markov assumption states that the current state is independent of the history states given the state at the previous time point, i.e., $Pr(z_t \mid z_0, z_1, \ldots, z_{t-1}) = Pr(z_t \mid z_{t-1})$ for $t \geq 2$. For most real data applications, the first order Markov property is assumed to simplify the probability models. The benefit of the Markov property would be diminished when higher order Markov process is added due to quickly increase model complexity (Langeheine and van de Pol 2000). While it is theoretically feasible to additionally model the transition probability a_{11} and a_{22} using collected covariates and incorporate subject level random effect to account for individual differences in the evolution pattern, these models can easily become intractable due to difficulties in model estimation procedure. Therefore, we choose to investigate the potential subject heterogeneity exhibited in the observed process. We model the positive mood assessment PA_{it} for the i_{th} individual measured at the t_{th} time point, conditional on his / her hidden mood state z_{it} using a mixed effect regression model shown in Eq. 2.

$$PA_{it} \mid z_{it} = k, X_{it}, v_i, \beta^k, \sigma_\epsilon^k \quad \sim \quad \mathcal{N}(X_{it}\beta^k + v_i, \sigma_\epsilon^k) \tag{2}$$

where X_{it} are the covariates for subject i measured at time t that can potentially affect the state-specific mean of the outcome PA. X_{it} can be either time varying, such as being alone, or time invariant, such as smoking status. The parameter β^k assesses the effect of smoking and being alone on mood assessment PA averaged over all time points when in mood state k. σ_{ϵ}^{k} reflects the within subject stability of the repeated mood assessments for the k_{th} mood states. v_i is the subject specific random effect with $v_i \sim \mathcal{N}(0, \sigma_v^2)$, reflecting the influence of subject i on his / her own mood assessment. σ_v^2, the variance of the random effects distribution, indicates the amount of between subject variation, accounts for the variation not explained by the covariates X_{it} and induces a correlation structure among the mood outcomes. Therefore, in addition to the time serial dependence, this model also incorporates heterogeneity among subjects due to unmeasured confounding by introducing random effects in the mean structure. However, mixed HMM assumes that individuals have common within subject variability in the same state, represented by σ_{ϵ}^{k}, which is often violated in intensive longitudinal studies due to the fact that psychological measurements such as positive affect usually take on high volatility and observed covariates are far from enough in explaining the total variability.

3.2 Mixed location scale HMM: extending the regular mixed HMM

Here we extend the mixed HMM and allow individuals to differ in terms of the stability of PA repeated measurements within subjects, by replacing the univariate subject random effect in Sect. 3.1 by a set of bivariate random effects in the observed process. Using the same notation as in Sect. 3.1, let $\left(v_i, \omega_i\right)$ denote the bivariate random location and scale effect vector for subject i. We assume that the hidden processes are homogeneous among subjects with transition probability matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and initial probability distribution $\pi = (\pi_0, 1 - \pi_0)$. Borrowing from mixed effect location scale model developed for intensive longitudinal data and conditional on the random effect vector (v, ω) and the hidden states z_i, the augmented mixed location scale HMM assumes that PA_{it} are independent Gaussian as expressed in Eq. 3:

$$PA_{it} \mid z_{it} = 0, X_{it}, W_{it}, v_i, \omega_i, \beta^k, \alpha^k \quad \sim \quad \mathcal{N}(X_{it}\beta^k + v_i, \exp(W_{it}\alpha^k + \omega_i)) \tag{3}$$

where W_{it} are the covariates for subject i measured at time t that can potentially affect the state-specific within subject variance of the outcome PA. It can be either time varying, such as being alone, or time invariant, such as smoking status, and is usually a subset of X_{it}. The parameter α^k assesses the effect of smoking and being alone on the within subject variability of mood assessment PA across all time points when in mood state k. As in Sect. 3.1, v_i is the subject specific random effect that reflects the influence of subject i on his / her mood assessment averaged over all time points, which we call subject random location effect. In addition, another subject specific random effect ω_i is included to reflect the influence of subject i on his / her mood stability represented by the within subject variance across all time points, which is called subject random scale effect. The random effect vector (v_i, ω_i) follows a bivariate normal distribution $\mathcal{N}\left(\mathbf{0}, \Sigma_{v,\omega}\right)$ and correlation between v and ω is allowed by specifying the structure of $\Sigma_{v,\omega}$. In psychometrics, this is of particular importance due to floor / ceiling effects that extreme measurements are often accompanied by smaller variance. The augmented model incorporates both random location and scale effects and is able to explain the subject heterogeneity exhibited in the outcome mean and within subject variability, beyond the covariates.

Let ϕ denote the probability density function for a normally distributed variable and $\theta = \left(\beta^k, \alpha^k, \Sigma_{v,\omega}\right)$ denote the parameter vector, the likelihood of the mixed location scale HMM is

$$
\begin{aligned}
\mathcal{L}(\theta;y) &= \int_{\nu,\omega} \sum_{z} \phi(y;z,\nu,\omega,\theta) f(z;\theta)\phi(\nu,\omega;\theta)d(\nu,\omega) \\
&= \int_{\nu,\omega} \sum_{z} \left\{ \prod_{i=1}^{N} \prod_{t=1}^{n_i} \phi(y_{it};z_{it},\nu_i,\omega_i,\theta) \right\} \times \left\{ \prod_{i=1}^{N} \pi_{z_{i1}} \prod_{t=2}^{n_i} a_{z_{i,t-1},z_{it}} \right\} \phi(\nu,\omega;\theta)d(\nu,\omega) \\
&= \int_{\nu,\omega} \prod_{i=1}^{N} \left\{ \sum_{z_i} \pi_{z_{i1}} \phi(y_{i1};\nu_i,\omega_i,\theta) \times \prod_{t=2}^{n_i} a_{z_{i,t-1},z_{it}} \phi(y_{it};\nu_i,\omega_i,\theta) \right\} \phi(\nu,\omega;\theta)d(\nu,\omega)
\end{aligned} \tag{4}
$$

The summation inside the integral is associated with the individual likelihood of the observed PA outcome given each mood state and subject random location/scale effects. The integrand then computes the total likelihood of the observed PA vector for all N subjects given their subject random effects. They are integrated with respect to the bivariate random effect distribution to give the total marginal likelihood. Similar to mixed HMM, the conditional model in the mixed location scale HMM indicates that, at time t, one's mood assessment depends on both his/her location and scale effects (ν_i, ω_i) and the latent mood state he/she is in. In the adolescent mood study example, individuals are suspected to have different response to smoking and being alone in regard to their mood regulation. In addition to the state-specific response, there could be significant subject specific effects such that some individuals tend to have better and more consistent mood compared to others despite the fact that they might have the same set of covariates and covariate effects. Strong subject heterogeneity is usually the result of unmeasured confounding as most models assume that unmeasured covariates are independent of the outcome. The inclusion of random subject scale effect is the novelty of this model and omitting the random scale effects would lead to incorrect state identification in the within subject variance, as the true state category will be confounded by the individual differences introduced by unmeasured subject level variables.

3.3 Model estimation

Parameters in HMM are often estimated by the expectation–maximization (EM) algorithm (Cappe 2011). Altman proposed to maximize the likelihood directly by evaluating the likelihood as a product of matrices before using quasi-Newton method to locate the maximum, in the case of regular mixed HMM (Altman 2007). Seltman claimed that likelihood-based methods are intractable and took a Bayesian approach instead, in the special case of one random effect (Seltman 2002).

The estimation of mixed location scale HMM poses a more challenging problem due to the addition of multiple random effects in both the mean and within subject variance model. Here we implement the forward-backward algorithm using a Bayesian approach. Specifically, for any individual, the conditional joint probability can be written as a product of the forward and backward quantities

$$
\begin{aligned}
Pr(z_{it}, y_i \mid \nu_i, \omega_i) &= Pr(z_{it}, y_{i1}, \dots, y_{it} \mid \nu, \omega) \times Pr(z_{it}, y_{i,t+1}, \dots, y_{in_i} \mid \nu, \omega) \\
&= p_{it} \times q_{it}
\end{aligned} \tag{5}
$$

where p_{it} and q_{it} are the forward and backward quantities, and are defined to be $p_{it} = Pr(y_{i2}, \dots, y_{it}, z_{it} \mid \nu_i, \omega_i)$ when $t \geqslant 2$ and $p_{i1} = Pr(z_{i1})Pr(y_{i1} \mid z_{i1}, \nu_i, \omega_i)$ when $t = 1$;

$q_{it} = Pr(y_{i,t+1}, \dots, y_{i,n_i}, z_{it} \mid v_i, \omega_i)$ when $t \leqslant n_i - 1$ and $q_{i,n_i} = 1$ when $t = n_i$, which can be computed recursively

$$
\begin{aligned}
p_{it} &= Pr(y_{i1}, \dots, y_{it}, z_{it} \mid v_i, \omega_i) \\
&= Pr(y_{it} \mid z_{it}, v_i, \omega_i) \sum_{z_{i,t-1}} Pr(z_{i,t-1}, y_{i1}, \dots, y_{i,t-1} \mid v_i, \omega_i) \times Pr(z_{it} \mid z_{i,t-1}) \\
&= \phi(y_{it} \mid z_{it}, v_i, \omega_i) \sum_{z_{i,t-1}} \alpha_{i,t-1} a_{z_{i,t-1}, z_{it}}
\end{aligned}
\tag{6}
$$

Similarly we can derive the backward variable q_{it} recursively in terms of $q_{i,t+1}$ and $Pr(z_{i,t+1} \mid z_{i,t})$. For a mixed location scale HMM, the conditional likelihood for an individual can be expressed as a function of the forward probabilities

$$
\begin{aligned}
\mathcal{L}(y_i \mid v_i, \omega_i) &= Pr(y_i, z_{in_i} = 0 \mid v_i, \omega_i) + Pr(y_i, z_{in_i} = 1 \mid v_i, \omega_i) \\
&= \sum_{k=1}^{2} p_{i,n_i}^k
\end{aligned}
\tag{7}
$$

where $p_{in_i}^k$ is the forward probabilities for the i_{th} individual when in state k at the last observation time point. The individual marginal likelihood can then be evaluated by integrating the conditional likelihood with respect to the distribution of (v, ω). Since the random effect vector (v, ω) is incorporated in the mean and within subject variance model which makes it difficult for adaptive Gaussian quadrature, we propose to evaluate the conditional likelihood with respect to the posterior distribution of (v, ω, θ) given the prior distributions and the data likelihood evaluated at the previous iteration, which can be implemented using Bayesian Monte Carlo sampler (Stan-Development-Team 2014). The detailed derivation and implementation is available upon request.

4 Simulation study

To validate the proposed mixed location scale HMM using the estimation procedure, we conducted a series of simulation studies under two scenarios and present the results here. Because of the computational burden (each iteration takes approximately 10 minutes), we limited the number of simulations to 100 under each scenario.

For each simulation, an intensively measured longitudinal outcome Y was generated according to (1) mixed HMM and (2) mixed location scale HMM for 100 individuals at a total of 10 equally spaced time points: for both scenarios, subjects have equal chance to be in state 0 or state 1 at t_0 ($\pi_0 = 0.5$); for $t \geq 2$, the transition probability matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$; Under scenario 1, subjects in different mood states exhibit state-specific heterogeneity in terms of the mean ($\mu_1 = 1, \mu_2 = 2$) and subject-specific heterogeneity in terms of the location ($v \sim \mathcal{N}(0, \sigma_v^2 = 1^2)$); while under scenario 2, subjects in different mood states exhibit state-specific heterogeneity in terms of both the mean and within subject variance ($\mu_1 = 1, \sigma_{\epsilon,1} = 2; \mu_2 = 2, \sigma_{\epsilon,2} = 1$) and subject-specific heterogeneity in terms of both the location and scale ($\begin{bmatrix} v \\ \omega \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_{v,\omega} = \begin{bmatrix} 1 & -0.3\sqrt{3} \\ -0.3\sqrt{3} & 0.3 \end{bmatrix} \right)$). (The true values of simulation parameters are chosen to mimic the real-world cases.) The simulated data are then analyzed by three candidate models: (1) HMM with no random effects, (2) regular mixed HMM (HMM with random subject location effects in the conditional model), (3) Mixed location scale HMM (HMM with both random subject location and scale effects in the conditional model). For each candidate model analysis, point estimate as well as 2.5% and 97.5% quantiles are recorded for π_0, a_{11}, a_{22}, μ_1, μ_2, σ^2_ϵ ($\sigma^2_{\epsilon,1}$, $\sigma^2_{\epsilon,2}$), σ^2_ν and σ^2_ω. Bias, AIW (average interval width) and coverage rate are calculated for the above parameters using posterior samples out of the 100 simulation runs.

Results are summarized in Table 1. π_0 is the initial probability of being in mood state 0. a_{11} and a_{22} are the probabilities of being in the same mood state given subjects already in that state at the previous time point. (μ_1, μ_2) and σ^2_ϵ or $(\sigma^2_{\epsilon,1}, \sigma^2_{\epsilon,2})$ are the mean and within subject variance for subjects the two mood states. σ^2_ν and σ_ω are the variance for random subject location and scale effects, respectively. Bias is computed for each parameter as the average point deviation from the true value: $Bias = \sum_{k=1}^{100}(\hat{\theta}_k - \theta)/100$, where $\hat{\theta}_k$ denotes the posterior mean for $\left(\pi_0, a_{11}, a_{22}, \mu_0, \mu_1\right)$ and mode for $\left(\sigma^2_\epsilon, \sigma^2_{\epsilon,0}, \sigma^2_{\epsilon,1}, \sigma^2_\nu, \sigma^2_\omega\right)$ from the k_{th} simulation. AIW (average interval width) is computed as the average range between the 97.5% and 2.5% quantile of the posterior: $AIW = \sum_{k=1}^{100}(\theta^U_k - \theta^L_k)/100$, where θ^U_k and θ^L_k are the 97.5% and 2.5% quantile of the posterior distribution from the k_{th} simulation. For each parameter, we also calculate the number of times out of 100 that the 95% credible interval contains its true value, thus providing the coverage rate as $COV = \sum_{k=1}^{100} \mathbb{1}\left\{\theta^L_k \leqslant \theta \leqslant \theta^U_k\right\}/100$.

Table 1 indicates that, under the first scenario where data come from a regular mixed HMM process and subjects exhibit heterogeneity in terms of the location in addition to the over all state-specific mean difference, the latter two candidate models - mixed HMM and mixed location scale HMM behave similar in terms of the bias and coverage rate while standard HMM produces large bias and insufficient coverage. This can also be seen by Fig. 1, where true parameter values together with estimates from all three candidate models are plotted against each other. Under the second scenario where subjects are allowed to exhibit heterogeneity in both location and scale in addition to the overall state-specific mean and within subject variance, only mixed location scale HMM provide valid parameter estimates and correct coverage rate. Neither standard HMM nor mixed HMM perform well as can be shown by Fig. 2. Regardless of the underlying data generating mechanism, mixed location scale HMM provides a more comprehensive modeling framework and at the same time achieves good estimation accuracy, reasonable interval width and asymptotic coverage rate. In addition, mixed location scale HMM provides insightful information about individual differences in both location and scale perspectives that cannot be obtained using simpler models, as can be indicated by next section.

5 Application to adolescent mood study example

In this section, we revisit the example introduced in Sect. 2. One question of interest is to investigate the effect of smoking and being alone in regulating adolescents mood activity, including how adolescents' mood change over time and how much of the total variation can be explained by the between subject portion (compare different subjects, averaged over all time points) and within subject portion (compare the same subject at different time points). By adopting a hidden Markov model structure, the between subject variability can be explained by the state-specific mood mean modeled by the covariates and subject specific location effects; the within subject variability can be explained by the time serial

Table 1 Simulation results under two scenarios: location process/location and scale process

Parameter	True value	HMM			Regular mixed HMM			Mixed LS HMM		
		Bias	AIW	COV (%)	Bias	AIW	COV (%)	Bias	AIW	COV (%)
Scenario 1:										
π_0	0.50	− 0.016	0.249	83	− 0.070	0.748	96	− 0.076	0.739	97
a_{11}	0.90	0.082	0.041	0	− 0.143	0.612	99	− 0.151	0.610	97
a_{22}	0.80	0.173	0.050	0	− 0.126	0.669	99	− 0.126	0.660	100
μ_1	1.0	− 0.545	0.317	0	− 0.057	0.957	96	− 0.074	0.958	95
μ_2	2.0	0.304	0.326	22	− 0.053	1.122	98	− 0.089	1.087	99
σ_ϵ	1.0	0.159	0.115	0	− 0.042	0.196	98	− 0.057	0.210	98
σ_ν	1.0	–	–	–	0.019	0.352	94	0.024	0.353	94
Scenario 2:										
π_0	0.50	− 0.036	0.323	85	− 0.122	0.371	78	− 0.072	0.659	94
a_{11}	0.90	0.063	0.080	31	− 0.019	0.203	96	− 0.144	0.571	90
a_{22}	0.80	0.152	0.082	1	0.085	0.160	51	− 0.071	0.514	98
μ_1	1.0	− 0.549	0.684	21	− 0.208	0.896	86	− 0.215	1.585	95
μ_2	2.0	0.327	0.493	33	− 0.154	0.597	76	− 0.082	0.885	95
$\sigma_{\epsilon,1}$	2.0	0.254	0.405	32	0.450	0.582	9	− 0.019	0.882	93
$\sigma_{\epsilon,2}$	1.0	0.466	0.322	2	0.142	0.367	65	0.190	0.809	92
σ_ν	1.0	–	–	–	− 0.038	0.381	92	0.036	0.432	94
σ_ω	0.548	–	–	–	–	–	–	0.079	0.469	94

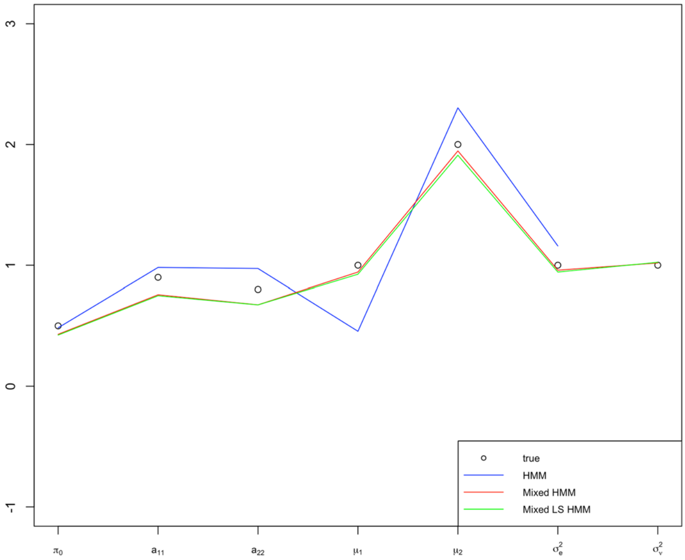

Fig. 1 Scenario 1: Comparison of parameter estimates from regular HMM, mixed HMM and mixed location scale HMM

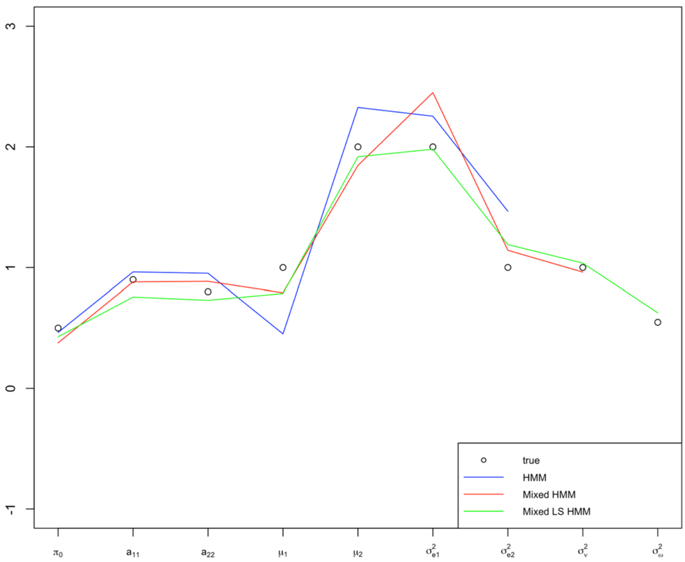

Fig. 2 Scenario 2: Comparison of parameter estimates from regular HMM, mixed HMM and mixed location scale HMM

dependence of the mood assessments modeled by a first order Markov chain, together with subject specific scale effects. As a result, the effects of smoking and being alone on mood regulation are likely to differ for the same subjects at different time points or for different subjects at the same time point, depending on their mood state and subject-specific traits. Two sources of individual heterogeneity-represented by random subject location and scale effects-are needed to account for the intra-subject correlation introduced by intensively repeated mood assessments. The detailed modeling strategy are described below.

Let Y_{it} denote the intensively measured longitudinal outcome-positive effect (PA), smk_i (1 for smoker and 0 for non-smoker) and $alone_{it}$ (1 if alone and 0 if with others) denote the subject and occasion level covariates, and let Z_{it} be the latent mood state associated with subject i, $i = 1, ..., N$, at time t, $t = 1, ..., n_i$. Here we pre-set the number of distinct mood states associated with each observation to be 2 based on domain knowledge and previous studies (Piasecki et al. 2016): π_0 denotes the initial probability for subject to be in mood state 1 ($1 - \pi_0$ for state 2), A denotes the transition probability matrix for the hidden process, and its element a_{jk} denotes the probability of transition from state j at time point $t-1$ to state k at time t.

Three candidate models are applied to the example adolescent mood data: (1) standard HMM (without random subject effects), (2) mixed HMM (with subject location random effect), and (3) mixed location scale HMM (with both subject random location and scale effects). They all have the same model specifications for the hidden process, which can be expressed as

$$\begin{aligned} P(z_t|z_{t-1}, A) &= \prod_{k=1}^{2}\prod_{j=1}^{2} A_{jk}^{z_{t-1,j}, z_{t,k}}, \quad t \geqslant 2 \\ P(z_1|\pi_0) &= \pi_0 \end{aligned} \tag{8}$$

For the observed process, let $X_{it} = (1, smk_i, alone_{it})$ denote the covariate vector including the intercept. Standard HMM model has

$$\begin{aligned} Y_{it}|z_{it} = k, X_{it}, \beta, \sigma_\epsilon^2 &\sim \mathcal{N}(\mu_{itk}, \sigma_\epsilon^2) \\ \mu_{itk} &= X_{it}\beta_k \end{aligned} \tag{9}$$

On top of standard HMM, mixed HMM includes the subject location random effects

$$\begin{aligned} Y_{it}|z_{it} = k, X_{it}, \beta, v, \sigma_\epsilon^2, \sigma_v^2 &\sim \mathcal{N}(\mu_{itk}, \sigma_\epsilon^2) \\ \mu_{itk} = X_{it}\beta_k + v_i, &\quad v_i \sim \mathcal{N}(0, \sigma_v^2) \end{aligned} \tag{10}$$

Mixed location scale HMM additionally differentiates the state-specific within variance and includes the scale random effects

$$\begin{aligned} &Y_{it}|z_{it} = k, X_{it}, \beta, \alpha, \Sigma_{v\omega} \sim \mathcal{N}(\mu_{itk}, \sigma_{ik}^2) \\ &\mu_{itk} = X_{it}\beta_k + v_i, \quad \sigma_{ik}^2 = \exp(X_{it}\alpha_k + \omega_i), \quad (v, \omega)^\top \sim \mathcal{N}\big(0, \Sigma_{v\omega}\big) \end{aligned} \tag{11}$$

Here β and α are the regression coefficients in the mean and within subject variance model. The two states are restricted to have the same β_0 and α_0 for model identifiability reason, but different β_1, α_1, β_2 and α_2, indicating that individuals in distinct states are expected to have the same population averaged mood at baseline, but different smoking and alone effects on mood regulation. v_i and ω_i are the subject random location and scale effects, indicating the influence of subject i on his or her mood mean and consistency. The variance σ_v^2 and σ_ω^2 reflects the amount of subject heterogeneity on baseline mood and mood stability that cannot be explained by the observed covariates, and $\rho_{v\omega}$ as the correlation between mood assessments and their within subject variability due to repeated measurements.

For comparison purposes, results from the three candidate models are summarized in Table 2. The point estimates are obtained as the posterior mean for regression coefficients (β and α) and probabilities (π_0, a_{11} and a_{22}) since their posteriors are approximately symmetric, and as the mode for random effect covariance elements σ_ν^2, σ_ω^2 and $\rho_{\nu\omega}$ since their posteriors are relatively skewed. The 95% credible intervals (CI) are obtained as the 2.5% and 97.5% posterior quantiles for all parameters. The model selection criteria $elpd_{LOO}$, proposed by Vehtari et al., estimates the pointwise leave one out (LOO) prediction accuracy from a fitted Bayesian model by evaluating the log likelihood over the posterior samples (Vehtari et al. 2017). It is preferred over the deviance information criterion (DIC) since it accounts for the entire posterior distribution, works for singular models and is invariant to parametrization. Higher $elpd_{LOO}$ indicates better model fit adjusting for the model complexity.

The three models provide different estimates and credible intervals for the probability parameters (π_0, a_{11} and a_{22}), indicating that they have identified different state mixture at baseline as well as different transition patterns over time. Mixed location scale HMM has the lowest model selection criteria $elpd_{LOO}$, indicating that although it has the largest number of parameters in the model, it still fits the data significantly better compared to other reduced models. Therefore, we will focus on the results and conclusions from the mixed location scale HMM in the example data.

Mixed location scale HMM has identified two distinct states with different smoking and alone effects on mood regulation. Specifically (1) for state one, being a smoker (versus non-smoker) and being alone (versus with others) both increase a subject's susceptibility to unpleasant and erratic mood as shown by their negative estimates (and credible intervals) of β_1^1 and β_2^1, and positive estimates of α_1^1; while (2) for state two, being a smoker (versus non-smoker) and being alone (versus with others) both make a subject more likely to have stable mood

Table 2 Comparison of parameter estimates and credible intervals between HMM, mixed HMM, and mixed location scale HMM

Par	HMM		Mixed HMM		Mixed location scale HMM	
	Estimate	CI	Estimate	CI	Estimate	CI
π_0	0.426	(0.349, 0.507)	0.472	(0.281, 0.644)	0.678	(0.472, 0.877)
a_{11}	0.982	(0.957, 0.999)	0.823	(0.718, 0.897)	0.837	(0.745, 0.907)
a_{22}	0.988	(0.973, 0.997)	0.908	(0.862, 0.947)	0.860	(0.758, 0.934)
β_0	7.135	(7.040, 7.227)	7.143	(6.968, 7.308)	7.136	(6.958, 7.312)
β_{smk}^1	− 1.044	(− 1.239, − 0.854)	− 1.028	(− 1.653, − 0.591)	− 0.635	(− 1.011, − 0.292)
β_{smk}^2	0.213	(0.063, 0.361)	0.142	(− 0.132, 0.417)	0.094	(− 0.209, 0.415)
β_{alone}^1	− 1.668	(− 1.942, − 1.398)	− 0.900	(− 1.362, − 0.485)	− 0.946	(− 1.244, − 0.676)
β_{alone}^2	0.480	(0.252, 0.707)	− 0.129	(− 0.384, 0.121)	0.207	(− 0.049, 0.469)
α_0	0.477	(0.416, 0.537)	− 0.124	(− 0.238, − 0.008)	− 0.390	(− 0.570, − 0.207)
α_{smk}^1	–	–	0.585	(0.270, 0.847)	0.476	(0.173, 0.793)
α_{smk}^2	–	–	− 0.683	(− 0.987, − 0.367)	− 0.449	(− 0.923, − 0.013)
α_{alone}^1	–	–	0.268	(− 0.133, 0.752)	− 0.107	(− 0.532, 0.288)
α_{alone}^2	–	–	− 0.821	(− 1.303, − 0.356)	− 1.001	(− 2.012, − 0.247)
σ_ν^2	–	–	1.095	(1.000, 1.190)	1.173	(1.057, 1.279)
σ_ω^2	–	–	–	–	0.709	(0.594, 0.834)
$\rho_{\nu,\omega}$	–	–	–	–	− 0.345	(− 0.475, − 0.294)
$elpd_{LOO}$	− 4119	–	− 3471	–	− 3359	–

assessments with no obvious effect on mood mean, as shown by the negative estimates of α_1^2 and α_2^2. Since state 1 is associated with significant smoke and alone effect mostly on mood mean, we call it mean-responding state; state 2 is associated with significant smoke and alone effect on mood variability, and is thus called variance-responding state. This model also estimates significant subject random effects in location and scale, which are negatively correlated as is often the case in psychological and behavioral studies due to floor or ceiling effects.

Both mixed HMM and mixed locations scale HMM successfully identified the two distinct states with β^1 and α^1 separated well from β^2 and α^2. However, mixed HMM generally have more separated estimates compared to mixed locations scale HMM, especially in terms of the variance coefficients α, which often leads to over classification and possibly increased false discovery rate. When subject heterogeneity does exist in the within subject variability, models that ignore it when performing the latent state classification would attribute more of the total variability to state specific differences while rather it can be explained by the individual specific differences. This is also true if we compare mixed HMM and HMM in terms of the separation between β_1^1 and β_1^2, or β_2^1 and β_2^2. Therefore, including both subject location and scale random effects in the conditional model increases the latent state classification efficiency and accuracy in the adolescent mood study example.

Once the hidden mood state for each subject at each time point is estimated, we can further explore the dominant mood state at each time point (i.e., which mood state has the majority of subjects at each time point) and the dominant mood state for each subject across all time pints (i.e., which mood state is more frequent across all time points for a certain subject). To achieve this goal, one can do post-hoc clustering at the subject level or observation level given the results from mixed location scale HMM. For subject level clustering, one can predict the state associated probability for each subject at each time point, obtain the state label associated with that observation according to a pre-set threshold (such as 0.5), and cluster subjects into different categories based on their proportions of certain states; for observation level clustering, observations can be clustered across all time points ignoring their subject label, or clustering can be done at each time point adjusting for subject affiliation. In the example data set, at subject level, 179 out of 339 subjects haven't changed their state through the study, among which 177 remain in state 1 (mean-responder) and 2 remain in state 2 (variance-responder); among the rest 160 subjects, 106 have more than 50% time points that are associated with state 1. At observation level, there are 1901 out of total 2373 observations belonging to state 1 overall; while at each day, there are 321, 301, 273, 265, 259, 243, 239 out of total 339 observations belonging to state 1 on each study day, respectively. The large number of state 1 (mean-responding state) observations at day 1 and its decreasing trend along the study period match the initial probability π_0 and transition matrix A estimate. These results indicate that the majority of subjects belong to mean-responders where smoking and being alone both increase susceptibility to unpleasant and erratic mood.

6 Discussion

In this paper, we extend the regular mixed HMM into a mixed location scale HMM and show that inclusion of both subject random effects in the mean and within subject variance of the observed process yields a model that can be well interpreted and estimated, yet provides a better fit for data in the context of intensive longitudinal studies. The application of this model to the adolescent mood study example demonstrates the advantages of such an approach over standard HMM or regular mixed HMM in extracting subject specific information as well as providing accurate and efficient state identification regarding the effects of smoking and being alone on mood regulation.

In intensive longitudinal studies where outcomes get intensively measured over a relatively short period of time, research interests are often surrounded around both the effects of certain behaviors or interventions on subjects' mood and mood stability trajectory evolving over time (Stone and Shiffman 1994). Therefore, mixed location scale approach provides a rigorous way to address these questions by introducing subject random location and scale effects which represent subjects' specific influence on their own mean and within subject variability, up or below the population average. The two distinct mood states, mean-responding state and variance-responding state, identified by our model make intuitive sense as they further divide individuals based on the effects of smoking and being alone on mood regulation, adjusting for the unobserved subject level covariates confounding. Following the model, one can further perform overall clustering at subject level by summarizing each subject's state specific proportion and transition pattern, as well as time-dependent clustering at observation level by focusing on the state mixture at a specific time point. Incorporating the location scale approach within hidden Markov model structure presents a novel yet efficient way for latent state modeling for intensive longitudinal data.

In this paper, we limited the number of simulations to 100 under each scenario, since results had become fairly stable and we did not see much need to go beyond. However, in practice, this number can be greatly increased using computing cluster or other advanced computing tools. For simplicity, we have assumed a constant transition probability matrix throughout this paper. However, it is also possible that the transition probability matrix in the hidden process can be covariate-dependent or subject specific (Jackson et al. 2015). But such models would be more difficult to estimate as one generally has less information about the hidden process compared to the observed process. Therefore, whether random effects need to be included in the hidden process depend on the research interests and abundance of the data. In carrying out the mixed location scale HMM, we have assumed equally spaced time intervals. However, this might not be true in practice as intensive longitudinal studies often produce measurements that are randomly spaced during the entire study span. In this case, a better alternative is to switch to a continuous-time domain in the Markov model where the temporal structure of the state-switching process is captured in a transition intensity matrix instead of a probability matrix (Liu et al. 2015). The original data from the example adolescent mood study have hierarchical structure - measurements were taken at random time intervals during a day and days are then nested within a week - thus future work could focus on exploring separate transition patterns within a day as well as across the days, followed by a link between these two transition matrices (Shirley et al. 2010).

Acknowledgements The project described was supported by National Heart Lung and Blood Institute (Grant Number R01 HL121330), National Cancer Institute (Grant Number 5P01CA09862) and Shanghai Commission of Science and Technology (Grant Number 19YF1402900). The content is solely the responsibility of the authors and does not necessarily represent the official views of National Heart Lung and Blood Institute, National Cancer Institute, or Shanghai Commission of Science and Technology. The authors thank editors, associate editors and referees for their valuable feedback.

Compliance with ethical standards

Conflict of interest The authors declare no potential conflict of interests.

References

Altman, R.M.: Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. J. Am. Stat. Assoc. **102**(477), 201–210 (2007)

Cappe, O.: Online EM algorithm for hidden Markov models. J. Comput. Gr. Stat. **20**(3), 728–749 (2011)

Choo, K.H., Tong, J.C., Zhang, L.: Recent applications of hidden Markov models in computational biology. Genom. Proteom. Bioinform. **2**(2), 84–96 (2004)

Crayen, C., Eid, M., Lischetzke, T., Courvoisier, D.S., Vermunt, J.K.: Exploring dynamics in mood rgulation—mixture latent markov modeling of ambulatory assessment data. Psychosom. Med. **74**(4), 366–376 (2012)

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C.S., Sheeber, L.B., Allen, N.B., Hamaker, E.L.: On the use of mixed Markov models for intensive longitudinal data. Multivar. Behav. Res. **52**(6), 747–767 (2017)

Gales, M., Young, S.: The application of hidden Markov models in speech recognition. Found. Trends Signal Process. **1**(3), 195–304 (2008)

Ghahramani, Z.: An introduction to hidden Markov models and bayesian networks. Int. J. Pattern Recogn. Artif. Intell. **15**(1), 9–42 (2001)

Goldstein, H., Leckie, G., Charlton, C., Tilling, K., Browne, W.J.: Multilevel growth curve models that incorporate a random coefficient model for the level 1 variance function. Stat. Methods Med. Res. **27**(11), 3478–3491 (2018)

Hedeker, D., Mermelstein, R.J., Demirtas, H.: An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. Biometrics **64**(2), 627–634 (2008)

Jackson, J.C., Albert, P.S., Zhang, Z.: A two-state mixed hidden Markov model for risky teenage driving behavior. Ann. Appl. Stat. **9**(2), 849–865 (2015)

Langeheine, R., van de Pol, F.: Fitting higher order Markov chains. Methods Psychol. Res. **5**(1), 33–54 (2000)

Leckie, G., French, R., Charlton, C., Browne, W.: Modeling heterogeneous variance-covariance components in two-level models. J. Educ. Behav. Stat. **39**(5), 307–332 (2014)

Lin, X., Mermelstein, R.J., Hedeker, D.: A 3-level bayesian mixed effects location scale model with an application to ecological momentary assessment data. Stat. Med. **37**(13), 2108–2119 (2018a)

Lin, X., Mermelstein, R.J., Hedeker, D.: A shared parameter location scale mixed effect model for ema data subject to informative missing. Health Serv. Outcomes Res. Methodol. **18**(4), 227–243 (2018b)

Liu, Y., Li, S., Li, F., Song, L., Rehg, J.M.: Efficient learning of continuous-time hidden Markov models for disease progression. Adv. Neural Inf. Process. Syst. **28**, 3599–3607 (2015)

MacDonald, I.L., Zucchini, W.: Hidden Markov Models and Other Models for Discrete-Value Time Series. Chapman & Hall, Boca Raton (1997)

Piasecki, T.M., Hedeker, D., Dierker, L.C., Mermelstein, R.J.: Progression of nicotine dependence, mood level, and mood variability in adolescent smokers. Psychol. Addict. Behav. **30**(4), 484–493 (2016)

Rast, P., Hofer, S.M., Sparks, C.: Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. Multivar. Behav. Res. **47**(2), 177–200 (2012)

Rocke, C., Li, S.C., Smith, J.: Intraindividual variability in positive and negative affect over 45 days: do older adults fluctuate less than young adults? Psychol. Aging **24**(4), 863–878 (2009)

Seltman, H.J.: Hidden Markov models for analysis of biological rhythm data. In: Gatsonis, C., Kass, R.E., Carriquiry, A., Gelman, A., Higdon, D., Pauler, D.K., Verdinelli, I. (eds.) Case Studies in Bayesian Statistics, pp. 397–405. Springer, New York (2002)

Shiffman, S., Stone, A.A., Hufford, M.R.: Ecological momentary assessment. Ann. Rev. Clin. Psychol. **4**, 1–32 (2008)

Shirley, K.E., Small, D.S., Lynch, K.G., Maisto, S.A., Oslin, D.W.: Hidden Markov models for alcoholism treatment trial data. Ann. Appl. Stat. **4**(1), 366–395 (2010)

Stan-Development-Team.: Stan Modeling Language: User's Guide and Reference Manual. Technical report, Bell Labs. (2014)

Stone, A.A., Shiffman, S.: Ecological momentary assessment (EMA) in behavioral medicine. Ann. Behav. Med. **16**(3), 199–202 (1994)

Vehtari, A., Gelman, A., Gabry, J.: Practical bayesian model evaluation using leave-one-out cross-validation and waic. Stat. Comput. **27**(5), 1413–1432 (2017)