SciencePG
Science Publishing Group

Mixed Hidden Markov Models for Clinical Research with Discrete Repeated Measurements

Yosuke Inaba[1, 2, *], Asanao Shimokawa[3], Etsuo Miyaoka[3]

[1]Department of Data Science, National Center for Global Health and Medicine, Tokyo, Japan

[2]Department of Mathematical Science for Information Science, Tokyo University of Science, Tokyo, Japan

[3]Department of Mathematics, Tokyo University of Science, Tokyo, Japan

Email Address:

yinaba@hosp.ncgm.go.jp (Y. Inaba), shimokawa@rs.tus.ac.jp (A. Shimokawa), miyaoka@rs.kagu.tus.ac.jp (E. Miyaoka)
[*]Corresponding author

To cite this article:

Yosuke Inaba, Asanao Shimokawa, Etsuo Miyaoka. Mixed Hidden Markov Models for Clinical Research with Discrete Repeated Measurements. *American Journal of Theoretical and Applied Statistics*. Vol. 6, No. 6, 2017, pp. 290-296. doi: 10.11648/j.ajtas.20170606.15

Abstract: A hidden Markov model (HMM) is a method for analyzing a sequence of transitions for a set of data by considering the outcomes Y to be output from latent state X, which has the Markov property. The HMM has been widely applied, with applications that include speech recognition, genomic analysis, and finance forecasting. The HMM was originally a method for dealing with single-process data. Thus, it is a natural extension to apply it to data with a repeated measure structure by incorporating random effects in it. This is called the mixed hidden Markov model (MHMM). With this extension, the MHMM was recently applied to clinical research data with repeated measurements, e.g. multiple sclerosis, alcohol consumption, and primary biliary cirrhosis. In relation to parameter inference, because regular HMM methods can be used in an MHMM framework, some legacy knowledge is applicable. The likelihood can be obtained by simply adding a random effect parameter to a single process HMM, and the conventional maximum-likelihood method can be used for parameter estimation. On the other hand, much work must still be performed. For instance, the mathematical property of the maximum likelihood estimator has not yet been thoroughly examined. In this study, the asymptotic normality and consistency of the maximum likelihood estimator of the MHMM concerned with time points are examined via simulation, and found that these properties were almost fine. These methods are applied to actual study data, and future perspectives are provided in the conclusion.

Keywords: Hidden Markov Models, Random Effects, Gaussian Quadrature, Newton–Raphson Method, Epilepsy Data, Poisson Distribution, Count Data

1. Introduction

A hidden Markov model (HMM) is a method used to analyze sequential data with characteristic transitions by considering the sequence of the outcomes to be output from a hidden state with the Markov property, for each sequence point [1]. The main benefit of this model is its ability to estimate the parameters of hidden states from the outcome. The history of the model began in the 1960s and 1970s, when Baum et al. investigated the mathematical property of the probabilistic function of Markov chains [2, 3, 4, 5]. In the 1980s and 1990s, the technology expanded to wider applications, including speech recognition [6], genomic analysis [7], finance forecasting [8], and online signature recognition [9] In the life science field, the HMM is mainly used for bioinformatics, including genetic linkage maps [10] distinguishing coding regions in DNA [11], protein modeling [12], and the estimation of the genotype and copy number [13]. Other implementations include cancer segmentation in MRI data [14] and the analysis of kidney disease data [15].

The HMM was originally a method for dealing with single-process sequential data. Thus, it is natural to extend its application to repeated measure data. Wang et al. used a hidden Markov Poisson regression (MPR) model [16], and Crespi et al. used a hidden semi-Markov model [17]. Altman made a breakthrough for this model in 2007 [18]. In this paper, she incorporated random effect into the HMM likelihood function, and applied it to multiple sclerosis data.

With this extension, the mixed hidden Markov model (MHMM) could be applied to clinical research data with repeated measurements. Kenneth et al. used this model to analyze alcohol consumption data [19]. Bartolucci et al. expanded it to develop shared parameter models to apply missing data [20], but it has room for improvement. For example, the distribution which random effect follows is limited to discrete distribution. Marino et al. proposed a mixed hidden markov quantile regression model in conjunction with MHMM and quantile regression and applied to CD4 count data in HIV patients [28]. They also examined multiple integral problems in Gaussian quadrature method in MHMM [29]. De Ruiter et al. analyzed blue whale behavior using multivariate MHMM [30].

The properties of the single-process HMM parameters have also been examined. Leroux proved the consistency of the maximum-likelihood estimator, under certain regularity conditions [21]. Bickel et al. showed the asymptotic normality of the maximum-likelihood estimator (MLE) of the HMM [22]. However, in relation to a multi-process HMM like an MHMM, almost no conclusions about the estimator's property have been reported. This paper reports some simulation results for this problem. In this paper, a maximum-likelihood estimation via the direct maximization mentioned in Altman [18] is performed, and proposed an approximated likelihood function that reduces computational resources. This approximated likelihood function is refered in section 3.

This paper consists of six sections. Section 2 gives some notations for the models and symbols used throughout this paper, section 3 describes the inference method for the HMM, and section 4 reports some simulation results. Section 5 gives descriptions of the actual data analyzed, and shows the analysis results. In section 6, a global discussion and some challenges for the future are provided.

2. Notations

2.1. Markov Chain

This section provides definitions for some basic ideas. First, throughout this paper, Bold letters are used to denote vectors, and small letters for scalars, for example, $\{A_n\} := \{A_1, A_2, \dots, A_n\}$. Then, a Markov chain is defined as follows. Let $T, s \in \mathbb{N}$ and $1 \le t \le T$ represent sequence points, and $\{X_T | 1 \le X_i \le s\}$ be a series of discrete random variables. Then, $\{X_T\}$ is called a Markov chain with state size s when it satisfies $\Pr(X_t|X_{t-1}) = \Pr(X_t|X_{t-1})$ for $1 \le \forall t \le T$, which means the current state X_t depends only on the previous state, and is naturally independent of $X_{t-2}, X_{t-3}, \dots, X_1$. The Markov chain is called *homogeneous* when it satisfies $Pr(X_{t+1} = k|X_t = j) = Pr(X_2 = k|X_1 = j)$ for $1 \le \forall t \le T$. The Markov chain is distinguished by some parameters. One of these is the transition probability $\gamma_{jk} := Pr(X_v = k|X_u = j)$ for $1 \le u \le v \le T$ and $1 \le j, k \le s$. This is sometimes specified in the matrix form $\Gamma := (\gamma_{jk})$, which is denoted as a transition probability matrix (TPM). Another parameter is a probability indicating the initial state for the Markov chain, called the initial distribution

$\delta := \{\delta_1, \dots, \delta_s | \delta_j := \Pr(X_1 = j)$ for $1 \le j \le s\}$. For $t = 1$, the Markov chain probability cannot be calculated because no previous state exists. Thus, parameters need to be specified. The initial probability can also be set by solving simultaneous equations $= \delta\Gamma, \sum_{k=1}^s \delta_k = 1$. In this case, the HMM is called *stationary*. In this paper, homogeneous Markov chains are considered.

2.2. HMM

Now, the HMM is defined. Outcome Y and latent state X are already defined as satisfying the Markov property. It is assumed that Y is discrete. The pair of sequence $\{Y_t, X_t\}_{t=1}^T$ is called an HMM when it satisfies $Pr(Y_t|X_t, Y_{t-1}) = Pr(Y_t|X_t)$. The defining characteristics of an HMM consist of three kinds of parameters: 1) state-dependent density parameters such as α for the mean of a Poisson distribution, 2) a TPM, and 3) the initial distribution, which represents the probability of the initial state for the Markov chain.

2.3. MHMM

The MHMM is an extension of the HMM that incorporates random effects to explain the between-cluster variability. To define a new model, Altman used generalized mixed effect models within an exponential family framework, which treats $\{Y_{it}\}$, where $1 \le i \le n$ represents certain clusters (e.g., patients, sites), as a conditional distribution: let $\psi := ((\tau_k)_{k=1}^s, (\beta`_k)_{k=1}^s, \phi)$ represent all of the model parameters, and for certain functions a, c, d, $f(y_{it}|X_{it} = k, u, \psi) = \exp(\frac{y_{it}\eta_{itk} - c(\eta_{itk})}{a(\phi)} + d(y_{it}, \phi))$ and $\eta_{itk} = \tau_k + x`_{it}\beta_k + w`_{itk}u$, where $\tau_k$ is a fixed effect when $X_{it} = k$, with $x`_{it}$ a covariate for i and t, let u be a random effect [18]. The MHMM is denoted as $\{Y_{it}, X_{it}\}$.

2.4. Discrete Repeated Measure Data

Discrete repeated measure data are data that are discrete versions of repeated measurements. Repeated measurements are longitudinal data collected over time for a certain cluster [1, 23, 24]. The cluster could be a subject, but could also easily be a rat, tree, field, country, or site. One typical example is the data taken from longitudinal studies. The main purpose of collecting this kind of data is to answer a scientific question such as the efficacy of a drug.

3. Inference Method

This section describes the inference method for the MHMM parameters in detail. First, a likelihood function is proposed. Let $\{Y_{it}, X_{it}\}$ be the HMM as previously defined, and let's be the number of states of a latent Markov chain. Let Γ be a transition probability matrix, and δ be the initial probability.

Various methods could be used to estimate the HMM parameters. Generally, when estimating mixed effects model parameters, Restricted Maximum Likelihood (REML) [25] are used, but its application to the MHMM has not been fully investigated. Altman proposed some methods to estimate

HMM parameters [18], including 1. the expectation-maximization algorithm, 2. direct maximization, 3. Monte-Carlo expectation-maximization algorithm (MCEM), and 4. simulated maximum likelihood. In this paper, the direct maximization method are used. Even though the EM algorithm is said to be "notoriously slow" [18], it is a very common method for estimating single-process HMM parameters. The difference is that when it comes to repeated measurement data, the random effect variables will increase in conjunction with an increase in the number of clusters. In the direct maximization method, random effect variables are marginalized using numerical integration, but this may cause an underestimation of the variance parameter value. This can be considered in the future. Because the Bayesian method allows freer modeling, it may provide more accurate estimates. The application of hierarchical Bayesian modeling can also be considered in the future.

3.1. Likelihood Function

Let the density function for each Y_{it} be f. Let the density for the hidden state be g, and the density for the random effect be h. Then, the likelihood function for the MHMM can be written in the following form:

$$L(\psi; y) = \int_u \sum_z f(y|x, u, \psi) g(x; \psi) h(u; \psi) \, du$$

$$= \int_u \sum_z \left\{ \prod_{i=1}^N \prod_{t=1}^{n_i} f(y_{it}|x_{it}, u, \psi) \right\} \times \left\{ \prod_{i=1}^N \pi_{x_{i1}} \prod_{t=2}^{n_i} \gamma_{x_{i,t-1}, x_{it}} \right\} h(u; \psi) du$$

$$= \int_u \prod_{i=1}^N \left\{ \sum_{x_i} x_{i1} f(y_i | x_{i1}, u, \psi) \times \prod_{t=2}^{n_i} \gamma_{x_{i,t-1}, x_{it}} f(y_{it}, u, \psi) \right\} f(u; \psi) du$$

The summation part for each i is a regular HMM likelihood function. Thus, this can be more simply expressed as a product of matrices. More specifically, for each u, $A^{i1} := (\delta_1 f(y_{i1} | X_{i1} = 1, u), \ldots, \delta_s f(y_{i1} | X_{i1} = s, u)), A^{it} := (A^{it}_{kl})$, where $A^{it}_{kl} := \gamma_{kl} f(y_{it} | X_{it} = l, u), t > 1$. Then, it can be described as follows:

$$L(\psi; y) = \int_u \prod_{i=1}^N \left\{ (A^{i1})' \left(\prod_{t=2}^T A^{it} \right) 1' \right\} f(u; \psi) du$$

In particular, suppose that a random effect is unique per subject and id. Then, the equation will be as follows:

$$L(\psi; y) = \prod_{i=1}^N \int_{u_i} \left\{ (A^{i1})' \left(\prod_{t=2}^T A^{it} \right) 1' \right\} f(u_i; \psi) du_i$$

3.2. Approximated Likelihood Function

Integrating a function is sometimes difficult using computer calculations. The simplest way to deal with this problem is to use a numerical integration technique. Let f be a function and from definition a one-dimensional integral, and let (a, b) be a certain interval where $\int_a^b f(v_i; \psi) dv_i \doteq 1, L(\psi; y) \doteq \prod_{i=1}^N \sum_{k=0}^M \frac{b-a}{M} [\{(A^{i1})'(\prod_{t=2}^T A^{it})1'\} f(v_k; \psi)]$, where M is an appropriately large positive integer and $v_k = a + \frac{(b-a)k}{M}, A^{i1} := (\delta_1 f(y_{i1} | X_{i1} = 1, v_k), \ldots, \delta_s f(y_{i1} | X_{i1} = s, v_k)), A^{it} := (A^{it}_{kl})$, where $A^{it}_{kl} := \gamma_{kl} f(y_{it} | X_{it} = l, v_k), t > 1$. This form is called an approximated likelihood function and is quite easy to compute.

3.3. Maximum Likelihood Estimation

For a single-process HMM, the most common method for computing the estimator of the MLE seems to be the Baum–Welch algorithm [2], which is the HMM version of the usual EM algorithm. However, for the MHMM, the EM algorithm converges very slowly, because this method attempts to infer each random effect individually, which causes the number of parameters to increase with the number of cluster elements. In this study, the direct maximization via Newton–Raphson method is selected. This method is very reasonable because a variance-covariance matrix of the parameter estimator is obtained at the same time, by inversing the observed Fisher information matrix.

3.4. Asymptotic Normality

The biggest issue with this method is the asymptotic normality of the estimator, which has not yet been thoroughly examined. For a single-process HMM, Leroux proved the consistency of the MLE [21], and Bickel showed asymptotic normality [22] under sine regularity conditions. In a later section, the asymptotic normality will be examined via simulation results, showing that this assumption is almost fine.

3.5. Statistical Test

Sometimes there is a need to test some statistics and clarify whether a certain parameter estimation value is significant, such as in the case of an observational study, like a drug efficacy investigation. After confirming the asymptotic normality of the MLE for the MHMM, conducting an approximate Wald-type test [23] will be very reasonable. This can be denoted as θ/SE, where the SE value of the maximum likelihood estimator can be given by $1/\sqrt{I(MLE)}$, and $I(MLE)$ is the Fisher information of the parameter replaced by the observed Fisher information.

4. Simulation

This section reports simulation results that confirm the asymptotic normality and consistency of MLEs as the number of time points increases.

4.1. Model

First, the simulation details are described. The assumptions of the simulation were as follows. It was a small clinical trial to assess the efficacy of an investigational drug. The target disease was chronic and the condition of a patient transitioned between relapse and remission, satisfying the Markov property, with the outcome following a Poisson distribution. The number of states of the Markov chain S is set to 2, and the Markov chain is supposed to be stationary. The number of time points T is 5, 10, 20, 30, or 40, and $h \in \{A, P\}$. The number of subjects is set to 10, which is denoted as $i = 1,... 10$. In addition, $i = 1$ to 5 is set as group A, $i = 6$ to 10 is group P, $\{X_{it}\}$ is the latent hidden state, $x_{i1} = \begin{cases} 0, h_i = P \\ 1, h_i = A \end{cases}, x_{i2} \in \mathbb{R}$ are the covariates of each i, and (β_1, β_2) are the coefficients for the covariates. $\{Y_{it}\}$ follows a Poisson distribution, as specified below.

$$Y_{it}|X_{it}, u_i \sim \text{Poisson}(e^{\lambda_s + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i})$$

Here, $u_i \sim N(0, e^{2\sigma_{rand}})$, and (τ_1, τ_2) are supposed to fixed effects corresponding to (x_{i1}, x_{i2}). The TPM for $\{X_{it}\}$ is $\Gamma = \begin{pmatrix} \grave{\gamma}_{11} & 1 - \grave{\gamma}_{11} \\ 1 - \grave{\gamma}_{22} & \grave{\gamma}_{22} \end{pmatrix}$, where $\grave{\gamma}_{11} := \frac{1}{1+exp(-\gamma_{11})}$, $\grave{\gamma}_{22} =$

$\frac{1}{1+exp(-\gamma_{22})}$, to avoid setting the boundary constraint to a parameter during the estimation. This Markov chain is already supposed to stationary. Thus, no initial distribution parameters are needed.

The simulation is executed as follows:
1. Construct a system that generates sample data following the MHMM, as previously mentioned.
2. Set the true values for the parameters in the system.
3. Execute the MHMM parameter estimation program for the sample data.
4. Record the estimated value.
5. Return to step 1 and iterate until the iteration count reaches a pre-specified value.

The true values are specified in Table 1. After execution, the results and calculated the sample mean of the MLEs for each parameter value are summarized, with the 5 percentile point and 95 percentile point for the 90% confidence interval.

SAS 9.4 are used for the computation, and executed the direct maximization using the Newton–Raphson method via the NLPNRA subroutines of SAS/IML. To integrate the random effect, the quadrature points of the Gaussian quadrature were set to 60, and range was set to (-3, 3), because the likelihood function had a value of zero outside of the range. Each iteration number is set to 300.

4.2. Simulation Results

Table 1 lists the simulation results.

***Table 1.** Sample Mean (5 Percentile, 95 Percentile) of the Parameter Estimates under T=5, 10, 20, 30, 40.*

Parameters	True values	Number of Timepoints				
		5	10	20	30	40
λ_1	1	1.406 (-1.199,4.923)	1.292 (-0.846,3.914)	1.092 (-0.448,2.91)	1.097 (-0.041,2.708)	0.866 (0.209,1.697)
λ_2	2	2.434 (-0.149,6.054)	2.17 (0.119,5.061)	2.108 (0.573,3.918)	2.084 (0.953,3.579)	1.863 (1.195,2.603)
γ_{11}	1.5	1.443 (0.128,2.885)	1.517 (0.765,2.325)	1.524 (0.938,2.051)	1.535 (1.083,2.015)	1.513 (1.167,1.826)
γ_{22}	0.5	-0.091 (-3.799,1.851)	0.451 (-0.514,1.466)	0.491 (-0.029,0.984)	0.456 (0.072,0.955)	0.458 (0.105,0.863)
β_1	1	1.166 (-0.12,2.846)	1.018 (-0.237,2.266)	1.149 (0.247,2.021)	1.193 (0.486,1.987)	1.115 (0.851,1.706)
β_2	0.5	0.643 (-1.386,2.448)	1.157 (-0.676,2.738)	0.781 (-0.13,3.041)	0.831 (0.217,1.855)	0.756 (0.31,1.077)
σ_{rand}1	0	-0.204 (-0.523,0.318)	-0.169 (-0.457,0.35)	-0.052 (-0.346,0.335)	-0.103 (-0.313,0.367)	-0.032 (-0.295,0.298)

For T = 5, some parameters were not well estimated. For example, the sample means of τ_1, τ_2 were overestimated. However, with T = 10, these values were improved, and with T = 20 and 30, the sample means almost coincided with the true values. For T = 40, the sample means of τ_1, τ_2 were a little bit far from the true value, but other parameters were precisely estimated. In terms of the confidential interval, the width became narrow as T increased for all the parameters. In addition, for T = 5, some skewness of the sample distribution are observed, with the interval between the 5 percentile to sample mean smaller than the range between the sample mean and 95 percentile point. However, as T increases the disproportion tends to improve. On the other hand, the variance component of the random effect tends to be underestimated for all values of T. It is considered that this is because, if less between-patient variability is observed, the estimated value tends to be smaller. Overall, the consistency and asymptotic normality of the MHMM parameters in relation to the time points are very reasonable. In addition, the confidential interval of β_1 straddles zero at T = 5, but the range decreases as T increases. At T = 10, the 5 percentile point (lower value of confidential interval) is almost equal to zero, but at T = 20, 30, and 40, the 5 percentile value is obviously greater than zero. Thus, the drug efficacy can be detected, and it is concluded that as far as the situation of this simulation, a precise estimation can be performed if the number of time points is greater than 20.

5. Application

5.1. Background

This section gives a detailed description of an actual data application. The study was a randomized, double-blind, parallel group multi-center study to compare a placebo with a new anti-epileptic drug (AED), in combination with one or two other AEDs. The datasets were obtained from the supplemental materials of Molenberghs et al. [26].

The study is described in full detail in Faught et al. [27]. The outcome of interest was the number of epileptic seizures experienced during the last week, i. e., since the last time the outcome was measured. The key research question was whether or not the additional new treatment reduced the number of epileptic seizures.

Table 2 lists the descriptive statistics for the demographic variables. There was no significant difference between the placebo arm and treatment arm.

Table 2. Descriptive statistics for demographics of epilepsy data.

		Plcebo	Treatment
N		**45**	**44**
Sex	Male	36	38
	Female	9	6
Age	Mean	36.22	33.77
	Min	19	20

		Plcebo	Treatment
N		**45**	**44**
Weight	Max	68	58
	Std. Dev	11.56	8.189
	Mean	169.48	180.90
	Min	88	97
	Max	270	264
Height	Std. Dev	36.53	36.44
	Mean	68.51	69.00
	Min	59	60
	Max	75	76
Baseline Sererate	Std. Dev	3.82	3.45
	Mean	15.98	23.65
	Min	4.00	4.30
	Max	73.70	198.30
	Std. Dev	16.10	34.10

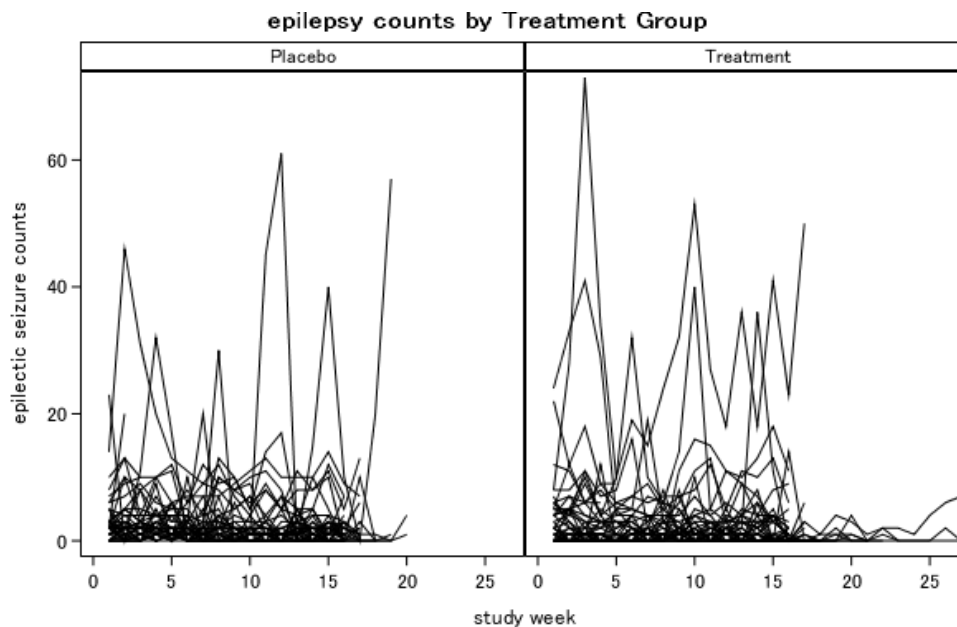The spaghetti plot of the epileptic seizure counts for each treatment group is shown below.

Figure 1. Epilepsy counts by Treatment Group.

Verbeke et al. applied a generalized linear mixed effect model with a Poisson regression to these data [23], which include a random intercept as well as a random slope. The problem with this method is that the model assumes that the data have a tendency to increase or decrease over time. However, in Figure 1 the transitions of the outcomes are more stationary. Thus, the assumption does not seem reasonable. Furthermore, there are some outstanding values, and the usual Poisson regression model with identical random effects within patient data is difficult to use to describe that phenomenon. The MHMM can deal with this problem, and effectively explain the stationarity and transition of the disease state. The treatment effects are evaluated using p-values based on Wald tests, with the level of significance set to 0.05.

5.2. Application

This section describes how to apply the MHMM model to epilepsy data. Let Y_{hit} be the epileptic seizure counts for subject i, time point t, and treatment arm $h \in \{P, D\}$, where P = placebo and D = drug, and $\{Z_{it}\}$ is the hidden state. It is supposed that $Z_{it} \in \{1, 2\}$, which represents the remission and relapse of the disease state. The distribution for Y_{hit} is supposed to follow a Poison process, which has the mean μ_{hit}, with $\mu_{hik} = exp(\tau_k + \beta_{hk}x_{hk} + u_i)$, where $x_{hk} = \begin{cases} 0 \ if \ h = P \\ 1 \ if \ h = D \end{cases}$. The parameters are set using the same method as used in the simulation. Random effect $u_i \sim N(0, e^{2\sigma_{rand}})$, and (τ_1, τ_2) are fixed effects corresponding to the hidden state. The TPM for $\{Z_{it}\}$ is $\Gamma = \begin{pmatrix} \gamma_{11} & 1 - \gamma_{11} \\ 1 - \gamma_{22} & \gamma_{22} \end{pmatrix}$, where $\gamma_{11} := \frac{1}{1+exp(-\gamma_{11})}$, $\gamma_{22} = \frac{1}{1+exp(-\gamma_{22})}$, and Z_{it} is assumed to be stationary.

5.3. Result

The results are listed in Table 3.

***Table 3.** Analysis results using MHMM.*

Parameters	Estimate	Observed Standard Error	P-Value
τ_1	-0.615806	0.1795157	-
τ_2	1.130064	0.1521669	-
γ_{11}	1.515724	0.2614786	-
γ_{22}	2.085650	0.1902266	-
β_{D1}	-0.074923	0.2218215	>0.05
β_{D2}	-0.127366	0.1720933	>0.05
σ_{rand}	0.052522	0.0421172	-

It is also denoted that the TPM $\Gamma = \begin{pmatrix} 0.8199079 & 0.1800920 \\ 0.1104994 & 0.8895006 \end{pmatrix}$ and $u_i \sim N(0, 1.1107595)$. The results show that the treatment effects are managed to decrease the outcome, but no significant differences are detected. In relation to the transition probability, as it can be seen from the TPM, it was relatively difficult for the state transition to occur, so the data were almost in the remission state. Thus, an inference on the parameter associated with the relapse state may not have worked well. This point will need further investigation. Overall, the analysis results effectively explained the phenomenon.

6. Conclusion

To summarize, MHMMs are imprementedand examined the asymptotic normality and consistency in relation to the time points via a simulation study. The results showed that both properties were reasonable, and if the time points are set to more than 20, good inference values are obtained in the situation that are set. An application to actual study data are investigated and found that it effectively interpreted the data. Our future tasks include the application of random effects to the transition probability matrix, which will cause the Markov chain to become non-homogenous and produce some difficulty explaining the correlation between random effects. It is easy to assume that all of the effects are independent, but this causes divergence from reality. Second, in this simulation study missing outcomes and subject dropout are not considered, but these are unavoidable in an actual study, especially long-term observational trials. Hence, the authers are planning to handle these problems by using a shared parameter model (SPM). An SPM is a model that is used to treat missing data in a longitudinal study, in which the missing process is non-informative.

References

[1] Zucchini W, MacDonald IL. Hidden Markov models for time series: an introduction using R: CRC press Boca Raton; 2009.

[2] Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The annals of mathematical statistics. 1970; 41 (1): 164-71.

[3] Baum LE. An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequalities. 1972; 3: 1-8.

[4] Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains. The annals of mathematical statistics. 1966; 37 (6): 1554-63.

[5] Baum LE. An Inequality with Applications to Statistical Estimation for Probabilistic Functions of a Markov Process and to a Model for Ecology. Bulletin of the American Mathematical Society. 1967; 73: 360-3.

[6] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989; 77 (2): 257-86.

[7] Durbin R, Eddy SR, Krogh A, Mitchison G. Biological sequence analysis: probabilistic models of proteins and nucleic acids: Cambridge university press; 1998.

[8] Rydén T, Teräsvirta T, Åsbrink S. Stylized facts of daily return series and the hidden Markov model. Journal of applied econometrics. 1998: 217-44.

[9] Faundez-Zanuy M. On-line signature recognition based on VQ-DTW. Pattern Recognition. 2007; 40 (3): 981-92.

[10] Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. Proceedings of the National Academy of Sciences. 1987; 84 (8): 2363-7.

[11] Churchill GA. Stochastic models for heterogeneous DNA sequences. Bulletin of mathematical biology. 1989; 51 (1): 79-94.

[12] Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. Hidden Markov models in computational biology: Applications to protein modeling. Journal of molecular biology. 1994; 235 (5): 1501-31.

[13] Scharpf RB, Parmigiani G, Pevnser J, Ruczinski I. A hidden Markov model for joint estimation of genotype and copy number in high-throughput SNP chips. 2007.

[14] Solomon J, Butman JA, Sood A. Segmentation of brain tumors in 4D MR images using the hidden Markov model. Computer methods and programs in biomedicine. 2006; 84 (2): 76-85.

[15] Marshall AH, Donaghy R, editors. Intelligent Patient Management using Dynamic Models of Clinical Variables. Computer-Based Medical Systems, 2006 CBMS 2006 19th IEEE International Symposium on; 2006: IEEE.

[16] Diggle P. Analysis of longitudinal data: Oxford University Press; 2002.

[17] Crespi CM, Cumberland WG, Blower S. A queueing model for chronic recurrent conditions under panel observation. Biometrics. 2005; 61 (1): 193-8.

[18] Altman RM. Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. Journal of the American Statistical Association. 2007; 102 (477): 201-10.

[19] Shirley KE, Small DS, Lynch KG, Maisto SA, Oslin DW. Hidden Markov models for alcoholism treatment trial data. The Annals of Applied Statistics. 2010: 366-95.

[20] Bartolucci F, Farcomeni A. A discrete time event‐history approach to informative drop‐out in mixed latent Markov models with covariates. Biometrics. 2015; 71 (1): 80-9.

[21] Leroux BG. Maximum-likelihood estimation for hidden Markov models. Stochastic processes and their applications. 1992; 40 (1): 127-43.

[22] Bickel PJ, Ritov Ya, Ryden T. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. The Annals of Statistics. 1998; 26 (4): 1614-35.

[23] Geert M, Geert V. Models for discrete longitudinal data. New York: Spriner. 2005.

[24] Wang P, Puterman ML. Analysis of Longitudinal Data of Epileptic Seizure Counts–A Two‐State Hidden Markov Regression Approach. Biometrical journal. 2001; 43 (8): 941-62.

[25] Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association. 1977; 72 (358): 320-38.

[26] Lee Y, Nelder JA, Pawitan Y. Generalized linear models with random effects: unified analysis via H-likelihood: CRC Press; 2006.

[27] Faught E, Wilder B, Ramsay R, Reife R, Kramer L, Pledger G, et al. Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. Neurology. 1996; 46 (6): 1684-90.

[28] Marino, Maria Francesca, Nikos Tzavidis, and Marco Alfò. Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences. Statistical methods in medical research 2016: 0962280216678433.

[29] Marino, Maria Francesca, and Marco Alfó. Gaussian quadrature approximations in mixed hidden Markov models for longitudinal data: A simulation study. Computational Statistics & Data Analysis 94 2016: 193-209.

[30] DeRuiter, Stacy L., et al. "A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure." The Annals of Applied Statistics 11.1 2017: 362-392.